

ECONOMETRIA

Segunda edición

ALFONSO NOVALES CINCA

Catedrático del Departamento de Economía Cuantitativa
Facultad de Económicas. Universidad Complutense
Madrid

McGraw-Hill

MADRID • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA • MEXICO
NUEVA YORK • PANAMA • SAN JUAN • SANTAFE DE BOGOTA • SANTIAGO • SAO PAULO
AUCKLAND • HAMBURGO • LONDRES • MILAN • MONTREAL • NUEVA DELHI
PARIS • SAN FRANCISCO • SIDNEY • SINGAPUR • ST. LOUIS • TOKIO • TORONTO

ECONOMETRIA. Segunda edición

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.

DERECHOS RESERVADOS © 1993, respecto a la primera edición en español, por McGRAW-HILL/INTERAMERICANA DE ESPAÑA, S. A. U.

Edificio Valrealty, 1ª planta
Basauri, 17
28023 Aravaca (Madrid)

ISBN: 84-481-0128-6
Depósito legal: M. 49.970-2000

Editora: Isabel Capella
Cubierta: Félix Piñuela. Grafismo electrónico
Compuesto en: Fernández Ciudad, S. L.
Impreso en: LAVEL, Industria Gráfica, S. A.

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

A mis hermanos

*Sepamos buscar como quien
espera encontrar,
y encontrar como quien
espera seguir buscando*

(Adaptación libre de S. Agustín.)

CONTENIDO

Prefacio	xv
Introducción	xix
Capítulo 1. Análisis matricial	1
1.1. Primeras definiciones	1
1.1.a. Operaciones con matrices	2
1.2. Determinantes	4
1.3. Matriz inversa	6
1.4. Rango de una matriz	8
1.5. Valores y vectores propios de una matriz	12
Capítulo 2. Análisis estadístico	17
2.1. Introducción: variable aleatoria, distribuciones discretas	17
2.2. Distribuciones continuas. Función de densidad	21
2.3. Momentos de una distribución	23
2.3.a. Momentos poblacionales con respecto al origen	23
2.3.b. Momentos poblacionales con respecto a la media	24
2.3.c. Momentos muestrales	25
2.4. Distribuciones bivariantes	26
2.5. Momentos en una distribución bivalente	29
2.6. Propiedades de un estimador	32
2.7. Cambio de variable en distribuciones de probabilidad	33
2.8. Distribuciones derivadas	34
2.9. El estimador de máxima verosimilitud	36
2.10. Teoría asintótica	37
2.10.a. Convergencia en probabilidad	37
2.10.b. Convergencia en distribución	40
2.11. Teorema Central del Límite	42
2.12. Contrastes de hipótesis	44
2.13. Distribuciones truncadas	46
Problemas	48

Capítulo 3. El modelo lineal general	52
3.1. Introducción	52
3.1.a. Características del modelo	55
3.1.b. Descripción de los capítulos posteriores	59
3.2. El estimador de mínimos cuadrados ordinarios	61
3.3. Propiedades del estimador de mínimos cuadrados ordinarios	67
3.4. Estimación de σ_u^2	78
3.5. Contraste de Normalidad	80
3.6. El estimador de máxima verosimilitud	82
3.7. Regresión particionada	85
3.8. El modelo lineal en desviaciones con respecto a la media	86
3.9. Algunos modelos lineales sencillos	91
3.10. Cambios de escala y de origen	96
3.11. Errores de especificación	100
Problemas	103
Capítulo 4. Inferencia en el modelo lineal	113
4.1. Introducción	113
4.2. Contraste de hipótesis: un tratamiento introductorio	115
4.2.a. Interpretación del estadístico F	115
4.3. Contraste de hipótesis: tratamiento general	117
4.3.a. La formulación del problema	117
4.3.b. El estadístico F para el contraste de cualquier conjunto de hipótesis lineales	119
4.3.c. Un procedimiento alternativo	121
4.4. Aplicación a algunos casos particulares	123
4.4.a. Contraste de hipótesis acerca de un coeficiente del modelo	123
4.4.b. Contraste de hipótesis acerca de todos los coeficientes del modelo (excepto el término independiente)	123
4.4.c. El contraste de significación global del modelo econométrico	124
4.4.d. Contraste acerca de un subvector de s variables ($1 \leq s \leq k-1$)	125
4.5. Contrastes de significación mediante sumas residuales	127
4.6. Intervalos y regiones de confianza	130
4.6.a. Intervalo de confianza para un solo coeficiente	130
4.6.b. Regiones de confianza para varios coeficientes	131
4.7. Estimación bajo restricciones	132
4.8. Contraste de hipótesis lineales mediante sumas residuales	135
4.9. Contraste de hipótesis mediante sustitución de las mismas	137
4.10. Contraste de cambio estructural. Test de Chow	139
4.10.a. Test de Chow	139
4.10.b. Contrastes de estabilidad	140
4.11. Variables ficticias	142
4.12. Predicción en el modelo lineal	146
4.12.a. Cálculo de las predicciones	146
4.12.b. El error de predicción y su varianza	148
4.12.c. Intervalo de confianza para la predicción	149
4.12.d. Predicción de un vector de valores futuros de la variable endógena	150
4.12.e. Evaluación de la bondad predictiva del modelo	151
4.12.f. Predicción de la «media» de la variable y_{T+1}	152
Problemas	153

Capítulo 5. Matrices de covarianzas no escalares	161
5.1. Introducción	161
5.2. Propiedades del estimador de mínimos cuadrados ordinarios	162
5.2.a. El estimador MCO es insesgado	163
5.2.b. Matriz de covarianzas del estimador MCO	163
5.3. El estimador de mínimos cuadrados generalizados	164
5.3.a. Estimación del vector de coeficientes β . El estimador de mínimos cuadrados generalizados	165
5.3.b. Propiedades del estimador MCG	166
5.3.c. Estimación del parámetro σ_u^2	168
5.3.d. El coeficiente de determinación	169
5.4. Introducción al problema de la heteroscedasticidad	170
5.5. Introducción al problema de autocorrelación	173
5.5.a. Estimación mediante transformación de variables	176
5.5.b. Obtención del estimador MCG mediante productos matriciales ..	178
5.6. El estimador de máxima verosimilitud	179
5.7. Inferencia estadística con matrices de covarianzas no escalares	181
5.8. Predicción en un modelo con matriz de covarianzas genérica	183
5.8.a. Predicción bajo heteroscedasticidad	183
5.8.b. Predicción bajo autocorrelación	184
5.9. Ejercicios prácticos	185
Problemas	189
Capítulo 6. Heteroscedasticidad	193
6.1. Introducción	193
6.2. Posibles causas de heteroscedasticidad	193
6.3. Estimación mínimo-cuadrática en presencia en heteroscedasticidad	196
6.4. Contrastes de heteroscedasticidad	199
6.4.a. El contraste de Goldfeld y Quandt	200
6.4.b. El contraste de Breusch y Pagan	201
6.4.c. El contraste de Glesjer	203
6.4.d. El contraste de Harvey	204
6.4.e. El contraste de White	205
6.4.f. El contraste de Spearman	206
6.5. Contraste de igualdad de varianza entre submuestras	206
6.6. Transformación de Box y Cox	207
6.7. Heteroscedasticidad condicional autorregresiva (ARCH)	208
6.8. Un caso práctico: producción y empleo por comunidades autónomas ..	211
6.9. Ejercicios prácticos	214
Problemas	217
Capítulo 7. Autocorrelación	224
7.1. Introducción	224
7.2. Naturaleza y causas de la autocorrelación	225
7.3. Consecuencias de la autocorrelación	227
7.4. Contrastes de autocorrelación	228
7.4.a. El contraste de Durbin-Watson	228
7.4.b. El contraste de Breusch y Godfrey	232
7.4.c. Contrastes gráficos	233
7.4.d. Las funciones de autocorrelación	235

7.5. Estimación de modelos con autocorrelación	235
7.5.a. Estimación MCG (Procedimiento de Cochrane-Orcutt)	235
7.5.b. Estimación de máxima verosimilitud del modelo autorregresivo de primer orden	237
7.6. Análisis de dos casos prácticos: Las funciones de consumo e inversión. . .	241
7.6.a. Una función de consumo	241
7.6.b. Una función de inversión	244
7.7. Ejercicios de simulación	245
Problemas	247
Capítulo 8. Ecuaciones simultáneas con variables explicativas exógenas	250
Parte I: Una sección cruzada de series temporales	250
8.1. Introducción	250
8.2. Diversas especificaciones de interés	252
8.3. Contrastes de homogeneidad entre unidades muestrales	255
8.4. Estimaciones de modelos en el caso de igualdad de matrices de covarianzas	257
8.5. Estimación con matrices de covarianzas diferentes	265
8.6. Estimación cuando los términos de error de las distintas ecuaciones están relacionados	267
8.7. Coeficientes variando en el tiempo	268
Parte II: Regresiones aparentemente no relacionadas	273
8.8. Descripción del modelo	273
8.9. Estimación de un conjunto de regresiones aparentemente no relacionadas	274
8.10. Contraste de hipótesis	282
Problemas	283
Capítulo 9. Modelos dinámicos	296
9.1. Introducción	296
9.1.a. Primeras propiedades	297
9.1.b. Cuando todos los retardos corresponden a variables exógenas . . .	299
9.1.c. Si aparecen valores retardados de la variable endógena	299
9.2. Justificación teórica de los modelos econométricos dinámicos	301
9.2.a. El modelo de expectativas adaptativas	301
9.2.b. El modelo de ajuste parcial de Nerlove	303
9.3. Modelos de retardos infinitos	304
9.3.a. El modelo de Koyck	304
9.3.b. Estimación de máxima verosimilitud del modelo de Koyck	305
9.4. Estimación con retardos de la variable endógena	307
9.4.a. El término de error no tiene autocorrelación	307
9.4.b. Estimación cuando el término de error tiene autocorrelación	309
9.4.c. El estimador de variables instrumentales	310
9.4.d. Contraste de exogeneidad de Hausman y Wu	313
9.5. Eficiencia relativa de los estimadores de variables instrumentales	318
9.5.a. Contraste de Sargan de validez de instrumentos	321
9.6. Contrastación de hipótesis con el estimador MC2E	321
9.7. El estimador de máxima verosimilitud	322

9.8. Estimación de modelos con expectativas racionales	324
9.8.a. Expectativas racionales: primeras propiedades	326
9.8.b. Estimación por variables instrumentales	328
9.9. El estimador generalizado de momentos	330
Problemas	333
Capítulo 10. Deficiencias muestrales: Multicolinealidad y errores de medida	344
10.1. Multicolinealidad: Concepto y consecuencias	344
10.2. Multicolinealidad exacta y multicolinealidad aproximada	349
10.3. Estimación de coeficientes bajo multicolinealidad exacta	352
10.4. Detección de la multicolinealidad aproximada	354
10.4.a. Métodos basados en la correlación entre variables explicativas	355
10.4.b. Métodos basados en el tamaño de la matriz $X'X$	357
10.5. Remedios contra la multicolinealidad	358
10.5.a. Mediante estimación del modelo propuesto	358
10.5.b. Mediante exclusión de variables	360
10.5.c. Estimadores restringido y no restringido: una opción delicada	361
10.6. Errores de medida	362
10.7. Estimación por variables instrumentales	365
10.8. Un contraste de especificación para errores de medida	366
10.9. Observaciones influyentes	366
Problemas	368
Capítulo 11. Modelos no lineales	372
11.1. Introducción	372
11.1.a. Especificaciones no lineales	372
11.1.b. Una aproximación lineal al modelo no lineal	374
11.2. Mínimos cuadrados no lineales	376
11.3. El estimador de máxima verosimilitud	380
11.3.a. Condiciones necesarias	380
11.3.b. Matriz de covarianzas	381
11.4. Transformación Box-Cox	385
11.5. Contraste de restricciones	388
11.5.a. Restricciones lineales	388
11.5.b. Restricciones no lineales	388
Capítulo 12. Algoritmos numéricos de optimización	396
12.1. La estimación de modelos econométricos como solución a un problema de optimización	396
12.2. Algoritmo de búsqueda	398
12.3. Algoritmo del descenso más rápido	400
12.4. Algoritmo de Newton-Raphson	400
12.4.a. Estimación por mínimos cuadrados	402
12.4.b. Estimación por máxima verosimilitud	404
12.5. Algoritmo de «Scoring»	406
12.6. Algoritmo de Gauss-Newton	407
12.6.a. Estimación de mínimos cuadrados	407
12.6.b. Estimación de máxima verosimilitud mediante el algoritmo de Gauss-Newton	409
Problemas	409

Capítulo 13. Modelos de series temporales	413
13.1. Introducción	413
13.2. Primeras definiciones	414
13.2.a. Proceso estocástico, ruido blanco, paseo aleatorio	414
13.2.b. Estacionariedad	415
13.2.c. Estimación de las funciones de autocorrelación de un proceso estacionario	417
13.3. Modelos autorregresivos	419
13.4. Modelos de medias móviles	423
13.5. Modelos ARMA	427
13.6. Variables no estacionarias	429
13.7. Estacionariedad e invertibilidad	430
13.8. Estacionalidad	431
13.9. Predicción con modelos ARIMA	432
13.9.a. Modelos autorregresivos	432
13.9.b. Modelos de medias móviles	434
13.9.c. El modelo ARMA(1, 1)	434
13.9.d. Error de predicción	435
13.9.e. Intervalos de confianza para las predicciones	438
13.9.f. Predicción de una serie en diferencias	438
13.10. Estimación de modelos ARIMA	439
13.10.a. Estimación de modelos autorregresivos	439
13.10.b. Estimación de modelos de medias móviles	440
13.10.c. Obtención de valores iniciales para los parámetros del modelo	443
13.11. Diagnóstico del modelo	444
13.11.a. Análisis de residuos	445
13.11.b. Sobreparametrización y sobrediferenciación	447
13.11.c. Valores influyentes y anomalías. Análisis de intervención	449
13.12. Modelos de función de transferencia	450
13.12.a. Identificación del modelo de función de transferencia	453
13.12.b. Identificación con preblanqueo	454
13.12.c. Identificación de un modelo para el ruido	455
13.12.d. Estimación de un modelo de función de transferencia	456
13.12.e. Diagnóstico del modelo de función de transferencia	458
13.13. Algunos ejemplos	459
Problemas	469
Capítulo 14. Regresión con variables no estacionarias	477
14.1. Introducción	477
14.1.a. Primeras definiciones	477
14.1.b. Regresión entre procesos no estacionarios	479
14.2. Contrastes de raíz unitaria de Dickey y Fuller	481
14.3. Contrastación en modelos autorregresivos de orden superior	485
14.4. Contrastación en modelos con estructura MA	486
14.5. Contraste de k raíces unitarias	487
14.6. Integración y estacionalidad	487
14.6.a. Raíz unitaria estacional	488
14.6.b. Raíz unitaria regular, junto con raíz unitaria estacional	489
14.7. Estacionariedad y cointegración	490
14.7.a. Cointegración y el modelo de corrección de error (MCE)	492
14.7.b. Estimación de modelos de corrección de error	494

14.7.c. Contrastes de cointegración	495
14.7.d. Contrastes de cointegración estacional	498
14.8. Aplicaciones del concepto de cointegración	499
14.8.a. Una síntesis	499
14.8.b. La eficiencia de un mercado financiero	500
14.8.c. La cointegración del Consumo y el PIB españoles	502
Capítulo 15. Datos de panel	504
15.1. Descripción del problema	504
15.2. El modelo de efectos aleatorios	506
15.3. Estimación eficiente en ausencia de correlaciones entre los efectos individuales no observables y las restantes variables explicativas (El estimador de Balestra-Nerlove)	508
15.4. Estimación consistente en presencia de correlaciones entre los efectos individuales no observables y las restantes variables explicativas	511
15.4.a. El estimador intragrupos	511
15.4.b. El estimador en primeras diferencias	513
15.4.c. El estimador entre grupos	514
15.4.d. Relación entre estimadores	514
15.5. Contrastes de especificación	514
15.6. Modelos dinámicos	516
15.6.a. Estimación consistente de modelos autorregresivos	516
15.6.b. Contrastes de especificación en modelos dinámicos	518
15.6.c. Modelos dinámicos con variables predeterminadas	519
15.7. Identificación de efectos individuales en el estimador intragrupos	520
Problemas	524
Capítulo 16. Variables dependientes cualitativas y limitadas	529
Parte I: Modelos de elección discreta	529
16.1. Introducción	529
16.2. Modelos de elección binaria	530
16.3. El modelo lineal de probabilidad	531
16.3.a. Observaciones repetidas	532
16.3.b. Estimación por mínimos cuadrados generalizados	534
16.4. Las decisiones de los individuos por medio de indicadores	536
16.4.a. El modelo probit	537
16.4.b. El modelo logit	540
16.5. Inferencia en modelos de elección discreta	545
16.5.a. Interpretación de los coeficientes estimados	545
16.5.b. La bondad de ajuste del modelo	546
16.5.c. Contrastación de hipótesis	547
16.6. Modelos de alternativas múltiples	548
Parte II: Variables dependientes limitadas	550
16.7. Variables dependientes truncadas	550
16.7.a. Estimación de máxima verosimilitud	551
16.8. Variables dependientes censuradas	552
16.8.a. Estimación de mínimos cuadrados	554
16.8.b. Estimación de máxima verosimilitud	555
16.8.c. Procedimiento de Fair	556

16.9. Modelos con selección muestral	558
Problemas	560
Capítulo 17. Modelos de ecuaciones simultáneas. I. Especificación e identificación.	565
17.1. Especificación del modelo de ecuaciones simultáneas	565
17.2. Formas estructural y reducida	569
17.3. Estimación de la forma reducida	572
17.4. El problema de identificación	574
17.5. Identificación mediante restricciones de exclusión	576
17.6. Identificación con restricciones lineales homogéneas	586
17.7. Identificación con restricciones lineales no homogéneas	590
17.8. Restricciones no lineales	591
17.9. Identificación bajo restricciones entre coeficientes de distintas ecuaciones	591
17.10. Identificación con restricciones sobre la matriz de covarianzas	593
Problemas	594
Capítulo 18. Modelos de ecuaciones simultáneas. II. Estimación	599
18.1. Dificultades en la estimación por mínimos cuadrados ordinarios	599
18.2. Estimación por mínimos cuadrados indirectos	603
18.3. Estimación por variables instrumentales	608
18.4. El estimador de mínimos cuadrados en dos etapas	615
18.5. El estimador de máxima verosimilitud con información limitada	624
18.6. Estimación por mínimos cuadrados en tres etapas	625
18.7. El método de máxima verosimilitud con información completa	630
18.8. Sistemas recursivos	632
18.9. Comparación entre los distintos estimadores	634
Problemas	634
Apéndice	645
Bibliografía	663
Índice	669

PREFACIO

La presente segunda edición incorpora cambios importantes a lo largo de todo su contenido con respecto a las versiones previas de este libro. Está concebido como un manual global de Econometría, que pueda utilizarse parcialmente como texto para las asignaturas de esta materia, pero que pueda ser utilizado asimismo por los alumnos como referencia acerca de temas que, incluso no habiendo sido estudiados en la licenciatura, puedan ser de su interés en el futuro. Por ello es que en esta edición se han incorporado algunas cuestiones que, por ser de desarrollo reciente, no estaban recogidas en ediciones anteriores.

Así, el Capítulo 14 («Regresión con variables no estacionarias») recoge la teoría de contrastación de existencia de raíces unitarias en series temporales, así como acerca de las posibles relaciones de cointegración entre variables. Se ha hecho un esfuerzo por sintetizar los desarrollos acerca de la contrastación de raíces unitarias estacionales, así como el análisis de cointegración estacional, a pesar de estar aún en plena evolución. El Capítulo 15 («Datos de panel») analiza la especificación, estimación y diagnóstico de modelos para este tipo de bases de datos desde el punto de vista de las cuestiones que hoy se consideran más relevantes para el analista, como son la estimación consistente y eficiente de modelos dinámicos con paneles de datos.

El Capítulo 16 («Variables dependientes cualitativas y limitadas») está ahora estructurado en dos partes: en la primera se analizan los modelos de elección discreta, y se corresponde con el capítulo análogo de la edición anterior; en la segunda parte del capítulo se estudian los modelos con variables dependientes limitadas, es decir, truncadas o censuradas, lo que representa una adición neta a esta edición. El Capítulo 8 («Ecuaciones simultáneas con variables explicativas exógenas») está asimismo dividido en dos partes, que tratan de dar un tratamiento similar a dos clases de problemas que revisten la peculiaridad enunciada en el título del capítulo: en unos casos, el analista dispone de un conjunto de series temporales, por ejemplo, procedentes de varios países, con cada una de las cuales podría analizar una determi-

nada cuestión económica; el otro tipo de casos lo constituye el modelo conocido como de «Regresiones aparentemente no relacionadas». La diferencia que en este texto se establece reside en que, en el primer caso, el interés fundamental del investigador está en la contrastación de la homogeneidad de las diversas muestras de que se dispone, mientras que en el segundo caso se trata de obtener estimaciones eficientes de un modelo de las características mencionadas.

Este Capítulo 8 se ha adelantado en el diseño del texto con respecto a la primera edición, puesto que siendo su interés la contratación de hipótesis y la estimación eficiente del modelo, debe introducirse al alumno una vez que éste conoce los procedimientos de llevar a cabo tales cuestiones. El Capítulo 9 («Modelos dinámicos») se ha adelantado asimismo con respecto a la edición anterior, con la intención de introducir lo antes posible al alumno la clase de estimadores de variables instrumentales, que podrá utilizar tanto en este tipo de modelos como en los capítulos posteriores.

Otras cuestiones, no menos importantes, incorporadas a esta edición son:

- a) Modelos ARCH de heteroscedasticidad (Sección 6.7).
- b) Modelos de función de transferencia (Sección 13.12).
- c) Estimación de modelos con expectativas racionales. El estimador generalizado de momentos (Secciones 9.8 y 9.9).
- d) Errores de medida (Secciones 10.6, 10.7 y 10.8).
- e) Observaciones influyentes (Secciones 10.9 y 13.11).
- f) Contrastes de restricciones no lineales (Sección 11.5).
- g) Transformación Box-Cox (Secciones 6.6 y 11.4).

Una de las mejoras introducidas en esta edición consiste en la presentación de una serie de ejercicios que, a modo de ejemplo, se desarrollan a lo largo del texto. La mayoría se basan en el análisis de datos españoles que se incluyen oportunamente. Estos ejemplos se van tomando en distintos capítulos para ir discutiendo sobre ellos gradualmente los distintos conceptos y métodos que se van introduciendo. Así, por ejemplo, se especifica una sencilla función de consumo para España en el Capítulo 7, donde se discuten los posibles problemas de autocorrelación presentes en su estimación; posteriormente, se utiliza como ejemplo sobre el que discutir la multicolinealidad (Capítulo 10); se estima una versión generalizada, no lineal, de la misma (Capítulo 12); sirve de ejemplo en el análisis de modelos univariantes y de función de transferencia (Capítulo 13), y se utiliza en el análisis de cointegración (Capítulo 14). Por otra parte, la colección de problemas que aparece al final de cada capítulo se ha ampliado considerablemente con respecto a la primera edición, hallándose en estos momentos en preparación la presentación detallada de las soluciones a los mismos.

Por tratar de ser un manual relativamente completo de Econometría, es claro que no puede utilizarse en su totalidad para un curso, para lo que es preciso seleccionar los temas que se consideren más adecuados para cada grupo de alumnos. A partir de los siete primeros capítulos, que contienen la teoría básica del modelo lineal general, los restantes capítulos están diseñados

de manera que puedan utilizarse independientemente como material para un curso. Por ello es que caben diversas posibilidades, entre las que sugiero las siguientes:

1. CURSO BASICO DE ECONOMETRIA: Capítulos 1 a 7, Capítulo 8 (Parte I), Secciones 9.1 a 9.6, Secciones 10.1 a 10.6, Capítulo 17 y Secciones 18.1 a 18.5. Sería aconsejable incorporar una discusión resumida de algunas secciones de los Capítulos 13 y 14.
2. CURSO SUPERIOR DE ECONOMETRIA: Capítulos 1 a 8, Secciones 9.1 a 9.6, Capítulos 10, 13, 14, 16, 17 y Secciones 18.1 a 18.5.
3. DOS NIVELES DE METODOS ECONOMETRICOS: *Primer nivel*: Capítulos 1 a 10, Secciones 13.1 a 13.9, y 13.11. *Segundo nivel*: Capítulos 11, 12, Secciones 13.10 y 13.12; Capítulos 14 a 18 (inclusive).

Aunque nadie está mejor capacitado que el responsable de cada asignatura para decidir acerca del tiempo que debe dedicar a cada tema, así como del momento idóneo para impartirlo, espero que las anteriores sean unas pautas útiles. No puedo sino recomendar, en todo caso, que la enseñanza de esta materia se complete con la discusión de ejercicios, de tres tipos: *a*) analíticos, como son la mayor parte de los incluidos en este texto, si bien debo advertir que sólo algunos de ellos deben proponerse en un curso básico de la materia, *b*) aplicados, como los que se recogen en Aznar y García Ferrer (1980) (actualmente en reelaboración), que también cuente con ejercicios analíticos relativamente sencillos, y Pérez Amaral, Amorós y Relloso (1993), o los que puedan complementar los ejemplos que, con datos reales, se presentan en este texto, y *c*) ejercicios de simulación que, como los presentados en las Secciones 5.9, 6.9 y 7.7, traten de ilustrar numéricamente las propiedades de los distintos estimadores que, a un nivel teórico, ya se han caracterizado al alumno.

Aunque las referencias precisas, tanto en términos de cálculo matricial como de análisis estadístico, se presentan en los dos primeros capítulos, recomiendo como material complementario, tanto en estos aspectos como en los puramente econométricos, los textos de Peña (1990) y Aznar y Trivez (1993).

Para la preparación de las innovaciones que se han incorporado al contenido de esta segunda edición, he contado con las valiosas opiniones y comentarios críticos de M. Arellano, A. Aznar, J. J. Dolado, R. Flores, J. de Hevia y A. B. Treadway. P. Pardo y E. Domínguez prestaron una inestimable colaboración preparando algunos de los gráficos. No menos importante ha sido el apoyo recibido en todo momento de todos mis compañeros del Departamento de Economía Cuantitativa y, especialmente, de los que dedican su actividad docente a esta materia. Mi agradecimiento asimismo a I. Capella, J. C. Cavín y M. Norte, de McGraw-Hill, por su estímulo a lo largo de la generación de este volumen, en ésta y en sus versiones anteriores. Siempre debe estar presente mi mayor reconocimiento a C. Sims, a quien debo una gran parte de mi formación como profesor y como investigador, y a C. Margazo, por su continua y enorme generosidad.

INTRODUCCION

La Econometría se ocupa del estudio de estructuras que permitan analizar características o propiedades de una variable económica utilizando como causas explicativas otras variables económicas. Por ejemplo, podría construirse una relación para explicar el comportamiento de la inflación, utilizando como variables explicativas el ritmo de crecimiento de la oferta monetaria y algún indicador de la demanda agregada en la economía. Distintos aspectos del análisis econométrico son:

- a) la especificación de la estructura a utilizar, llamada *modelo econométrico*;
- b) el análisis de las propiedades estadísticas de dicho modelo;
- c) su estimación;
- d) la utilización de dicho modelo con fines predictivos, y
- e) la capacidad de dicho modelo para el análisis de determinadas cuestiones de política económica.

Las cuestiones de política económica a analizar (que en general dictan cuál debe ser el alcance del modelo econométrico) pueden ser de índole macroeconómica, como ocurre con cuestiones de economía monetaria o economía laboral, o bien de carácter microeconómico, como ocurre con cuestiones como la medida del grado de monopolio existente en una industria, o el análisis de los determinantes de la estructura de capital de las empresas.

La clase de modelos habitualmente utilizados en Econometría va, sin embargo, más allá de lo que acabamos de mencionar. En muchas ocasiones no se pretende explicar el comportamiento de una variable, sino el de varias variables simultáneamente. En tales casos la variable a explicar en una de las ecuaciones puede aparecer como variable explicativa en alguna otra ecuación del modelo.

Modelo económico y Bases de Datos

Tomemos como un primer ejemplo una función de consumo keynesiana:

$$C_t = \alpha + \beta Y_t, \quad t = 1, 2, 3, \dots, T$$

Este es un sencillo modelo con el que se pretende explicar la evolución temporal de los gastos de consumo por medio de una variable que indique el nivel de renta. El modelo predice que, a través del tiempo, los gastos de consumo C_t debieran haber evolucionado paralelamente a la renta Y_t . De acuerdo con esta especificación, en cada período se debería haber gastado en bienes de consumo una proporción de la renta, medida por el producto βY_t . La diferencia entre ambas cifras se supone constante en el tiempo (e igual al valor del parámetro α). Este modelo de consumo se puede utilizar a dos niveles distintos:

1. Al nivel agregado de la economía, en cuyo caso la variable C_t e Y_t serían indicadores de los gastos de consumo y del nivel de renta agregados. Para este análisis del modelo de consumo se requerirían observaciones numéricas acerca de las variables C_t e Y_t , durante un número de períodos de tiempo que denotaremos por T . La relación de las observaciones correspondientes a cada una de las variables es una *serie temporal*. Salvo en casos especiales, el investigador querría en este caso que la serie temporal de datos de la variable consumo contuviese información acerca del consumo privado, así como del público.

2. A un nivel desagregado, por ejemplo relacionando los gastos en bienes de consumo y los ingresos de un conjunto de familias. En tal caso, escribiríamos el modelo:

$$C_i = \alpha + \beta Y_i, \quad i = 1, 2, 3, \dots, N$$

donde los subíndices hacen ahora referencia al hecho de que cada observación corresponde a una familia distinta y no a un período de tiempo diferente. El tipo de información estadística que se precisa en este caso, es decir, la colección de pares de valores referentes a los gastos de consumo y nivel de renta de cada una de las familias en la muestra, se conoce como *datos de sección cruzada*.

Período	C_t	Y_t		C_i	Y_i
1960, I	130	1400	Familia 1	210	1550
1960, II	133	1420	Familia 2	133	1420
1960, III	137	1460	Familia 3	367	2010
1987, IV	421	2175	Familia 104	421	2175
Datos (trimestrales) de series temporales			Datos de sección cruzada		

Finalmente, otro tipo de datos lo constituye una muestra en que se ha recogido información acerca de un conjunto de individuos (o familias o países) a través del tiempo. En tal caso, podríamos aspirar no sólo a obtener valores numéricos de los coeficientes para cada uno de los individuos en la

muestra, sino también para cada uno de los instantes en que se dispone de información⁽¹⁾. Este tipo de muestras se conoce como *Panel de Datos*.

El nivel 1 de utilización de este modelo lleva asociado una agregación del consumo de las distintas unidades decisoras en la economía en cada instante de tiempo. Eso no quiere decir que no haya además una *agregación temporal*. De hecho, siempre la hay: El consumo es un flujo, no un stock, y si se recogen datos anuales sobre el consumo y la renta, cada una de las observaciones *acumula* los flujos de consumo y renta a lo largo del año. Esta agregación *temporal* puede producir un problema, el de que se enmascara una relación que realmente existe entre C_t e Y_t a intervalos de tiempos pequeños con respecto al intervalo con el que se han recogido los datos. Lamentablemente, la Teoría Económica no dice nada, en general, acerca del intervalo de tiempo con el que se producen las relaciones entre las variables, de modo que la elección de la frecuencia de datos con los que se trabaja (mensuales, trimestrales, anuales) queda a elección del investigador, dentro de las limitaciones existentes en cuanto a la disponibilidad de bases de datos.

Una base de datos de sección cruzada resuelve parcialmente este problema. Pudiera ser que los datos sobre gastos de consumo y renta proviniesen de una encuesta en que se ha preguntado a un grupo de familias acerca de los valores de estas variables a lo largo del último año. En tal caso, el problema de *agregación temporal* estará presente. Pero generalmente las preguntas habrán sido acerca de esas cifras durante la última semana, con lo cual las respuestas reflejarán más acertadamente la relación en el muy corto plazo entre la renta y los gastos de consumo.

Debe quedar claro de esta discusión que una ecuación matemática que describe una relación entre variables económicas no constituye por sí sola un modelo econométrico. Un modelo viene dado por dicha relación o relaciones, junto con el tipo de datos a utilizar en el trabajo empírico (si datos de series temporales o de sección cruzada), las variables concretas que van a utilizarse, es decir, si se va a analizar el gasto de consumo en todo tipo de bienes o sólo el de bienes no duraderos, la frecuencia de los datos en el caso de series temporales (si mensual, trimestral, anual u otra); si va a utilizarse el PIB u otra medida de renta agregada, etc. Sin otra limitación que la disponibilidad de datos, el economista debe tener un criterio claro acerca no sólo del modelo con el que trabajar, sino también del tipo de datos y las variables que debe utilizar de acuerdo con el tipo de cuestiones que pretende analizar.

Relaciones dinámicas

La ventaja de una base de datos de sección cruzada con respecto a una de series temporales a nivel agregado es que, en general, la teoría económica de determinación del consumo óptimo que el investigador tiene en mente al hacer el trabajo empírico (por ejemplo, la renta permanente o la teoría de la renta

⁽¹⁾ Aunque no podría estimarse un coeficiente para cada observación muestral, lo que se hace es permitir que los coeficientes varíen en el tiempo y con los individuos.

relativa) tiene implicaciones acerca del comportamiento de los gastos de consumo de las unidades decisoras (generalmente, una familia), pero escasamente tiene algo preciso que decir acerca del comportamiento de los gastos de consumo para el agregado de la economía con respecto a la renta agregada.

Por otra parte, las encuestas de las que suelen surgir los datos de sección cruzada son lo suficientemente ricas como para contener información acerca del número de hijos en la unidad familiar, el número de receptores de ingresos en dicha unidad, tamaño del municipio en que la familia vive, si vive en un medio rural o urbano, etc., variables éstas que tendrían interés incluir en un modelo que pretende explicar el comportamiento de los gastos de consumo de la unidad familiar.

Los datos de sección cruzada tienen, sin embargo, un inconveniente, y es que no permiten analizar las relaciones que puedan existir entre las variables renta y consumo a lo largo del tiempo, salvo a muy corto plazo, si se dispone de un panel de datos. Estas relaciones *dinámicas* no sólo son en general importantes para cualquier conjunto de variables, sino que además tienen sumo interés en el análisis de cuestiones de política económica.

Así, podría ser posible descubrir, mediante inspección de las series temporales de datos, variaciones grandes en la relación $\ln C_t / \ln Y_t$ a lo largo del período muestral, como por ejemplo una disminución en su valor de un 90 a un 75 por 100. Este parámetro es la *propensión marginal al consumo*, y debe esperarse una disminución en su valor numérico en el caso de una economía en desarrollo. Así, toda información acerca de la evolución de su valor numérico a través del tiempo es relevante para evaluar el grado de desarrollo de una economía. También sería muy importante para conocer el efecto de realimentación que pueda tener un determinado ritmo de expansión del PIB (que sería la variable indicadora de la renta agregada, Y_t).

Sin embargo, el análisis de cuestiones de naturaleza intrínsecamente dinámica como la citada requiere de una especificación correcta del modelo. Por ejemplo, supongamos que se está estimando con datos de series temporales el valor de la *propensión marginal al consumo* en la economía, es decir, el grado en que la demanda de consumo agregada aumentará cuando se produzca una expansión, caracterizada por un incremento en la renta agregada.

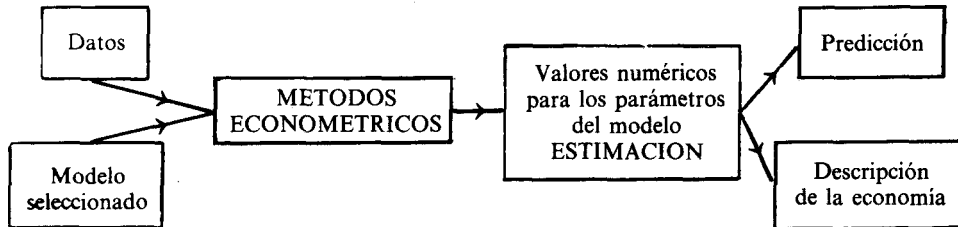
Si utilizando datos mensuales llegamos a obtener (siguiendo los métodos que desarrollaremos en capítulos sucesivos) un valor estimado de β de 0,50, ello podría interpretarse como que sólo la mitad de un incremento en renta se traduce en consumo, lo que sería una propensión marginal al consumo baja y, por tanto, una propensión al ahorro alta. Si de ello se infiere, por una parte, que a pesar de la expansión en el PIB hay que añadir más impulso a la economía con medidas fiscales o monetarias, y se diseñan medidas de política económica basadas en estas observaciones, se podrían estar cometiendo errores importantes.

Lo que podría estar ocurriendo es que la relación entre la renta o el nivel agregado de ingresos y los gastos de consumo no es exclusivamente o esencialmente contemporánea. Esto es todavía más verosímil si se utilizan datos observados frecuentemente (mensuales) que datos de menor frecuencia (anuales). En tal caso, deberíamos estar trabajando con un modelo que in-

cluyera como variables explicativas, además de Y_t , otras variables como Y_{t-1} , Y_{t-2} , etc. Con ello, considerando globalmente los coeficientes en cada una de las diferentes variables renta incluidas, se tendría una idea más acertada acerca de la magnitud de la relación entre consumo y renta, así como de la distribución de esa relación a lo largo de distintos períodos. Especificar correctamente la componente *dinámica* de la relación entre variables es una **tarea** delicada, y sin duda una de las más importantes en Econometría.

Estimación del modelo econométrico

Hasta ahora hemos respondido a dos aspectos importantes acerca del análisis de una determinada cuestión económica: ¿Qué modelo especificamos? ¿Qué tipo de datos necesitamos utilizar? Los métodos econométricos que discutimos en los capítulos siguientes ayudan a responder a una tercera pregunta: ¿Qué valores asignamos a los parámetros del modelo?



Una vez que mediante métodos econométricos hemos asignado valores numéricos a los diferentes parámetros, el modelo puede utilizarse con dos objetivos fundamentales: Predicción y/o descripción del entorno económico del que procede la información muestral. Este podría ser la economía de un determinado país, una ciudad (a cuyas familias se ha encuestado) o incluso la economía mundial.

Los valores paramétricos son necesarios para realizar ejercicios de predicción, así como para tener un conocimiento descriptivo de la economía. Los ejercicios de predicción pueden ser del tipo: ¿Si la renta crece en términos nominales un 5 por 100 anual, cuál será el gasto agregado de consumo dentro de un año? Si nos encontramos en el tercer trimestre de 1988 (1988, III), y si la renta hoy es de 100 unidades (posiblemente 10^9 pesetas), la renta prevista para (1989, III) es de 105, y los valores estimados de los parámetros son $\hat{\alpha} = 13,75$ y $\hat{\beta} = 0,78$, entonces dicha predicción será:

$$C_{1989, III} = 13,75 + (0,78) (105) = 95,65$$

Otro tipo de preguntas podría consistir en suponer un crecimiento del 1,25 por 100 trimestral para generar el nivel de renta primero y el consumo después para los próximos cinco años a partir de (1988, III). Es cierto que la fiabilidad de la predicción dependerá de:

1. El horizonte de predicción.
2. La constancia de los valores paramétricos estimados a lo largo del horizonte de predicción.
3. La calidad de nuestras estimaciones de los parámetros del modelo.
4. Que el modelo utilizado sea apropiado y en particular que esté especificado correctamente.

Un análisis de predicción es fundamental para hacer cualquier estudio de política económica. Supongamos que el banco central de un país quiere analizar el posible efecto inflacionista de una política monetaria expansiva. Normalmente, los tipos de interés aparecerán en la función que explica los gastos de consumo (especialmente si los gastos de consumo incluyen los de bienes duraderos), del mismo modo que el crecimiento de la oferta monetaria aparecerá en la ecuación de determinación de los tipos de interés. El signo y magnitud de los coeficientes correspondientes serán necesarios para responder a la cuestión propuesta, para lo que se generarían predicciones de los tipos de interés y del consumo del modo antes descrito, utilizando un determinado supuesto acerca del crecimiento de la oferta monetaria.

Los valores paramétricos son también importantes para tener un conocimiento descriptivo de la economía. Por ejemplo, ¿cuál es la propensión marginal al consumo de la economía española? ¿Cuál es la sensibilidad del empleo frente a variaciones en los salarios reales? ¿Cuánto empleo se crea o destruye si los salarios reales se mantienen constantes durante los próximos tres años.

Linealidad del modelo

Para responder a la última cuestión formulada, se podría pensar en analizar un modelo del tipo:

$$\ln U_t = \alpha + \beta \ln (w_t/p_t)$$

donde U_t denota la tasa de paro y w_t/p_t es el salario real. Esta especificación tiene la ventaja de que el valor del coeficiente β proporciona la elasticidad desempleo-salario real, puesto que:

$$\beta = \frac{\partial \ln U_t}{\partial \ln w_t/p_t} = \frac{w_t/p_t}{U_t} \cdot \frac{dU_t}{d(w_t/p_t)} = \frac{\text{Variación porcentual en } U_t}{\text{Variación porcentual en } w_t/p_t}$$

Este modelo es claramente no lineal, pero puede transformarse en otro lineal mediante un sencillo cambio de variables (lo que equivale a una simple manipulación de datos). Así, si llamamos $y_t = \ln U_t$ y $x_t = \ln (w_t/p_t)$, entonces el modelo es lineal en x_t e y_t . Casos similares son el modelo:

$$C_t = a + bY_t + dY_t^2$$

y la función de producción Cobb-Douglas: $Y_t = AK_t^\alpha L_t^\beta u_t$, que se transforma en:

$$\ln Y_t = \ln A + \alpha \ln K_t + \beta \ln L_t + u_t$$

En este último caso, los parámetros α y β son los mismos del modelo original, pero el término constante ya no es A , sino su logaritmo. Un modelo no lineal que no puede transformarse en otro lineal es:

$$C_t = \alpha + \frac{\beta Y_t}{p_t - \gamma p_t^2}$$

En otros casos, un modelo no lineal puede transformarse en un modelo lineal en las tasas de crecimiento de las variables. Por ejemplo, el tradicional modelo estático de la teoría cuantitativa de la demanda de dinero: $M_t^d V_t = Y_t p_t$, se transforma, tomando logaritmos, en: $\log M_t^d = \log p_t + \log Y_t - \log V_t$. Si escribimos el mismo modelo para el instante $t - 1$ y restamos del modelo en el instante t , se tiene: $\log (M_t^d / M_{t-1}^d) = \log (p_t / p_{t-1}) + \log (Y_t / Y_{t-1})$, suponiendo que la velocidad de circulación permanece aproximadamente constante. Pero la tasa de crecimiento de una variable puede aproximarse por un cociente como los anteriores⁽²⁾, por lo que finalmente se tiene:

$$m_t^d = \pi_t + y_t$$

donde y_t denota la tasa de crecimiento del producto en términos reales, m_t^d es la tasa de crecimiento de la demanda de dinero y π_t la tasa de inflación. Ello sugiere especificar un modelo de determinación de la demanda de dinero como:

$$m_t^d = \beta_1 \pi_t + \beta_2 y_t + u_t$$

donde el término u_t sería una variable aleatoria, llamada *término de error del modelo econométrico*, y que recoge el efecto de la posible no linealidad de la verdadera relación entre dinero, precios y renta, en cuyo caso el modelo anterior sería una mera aproximación; la posible existencia de otras variables que fuesen relevantes para explicar la evolución de la demanda de dinero, y el hecho de que una componente de la demanda de dinero tenga naturaleza puramente estocástica, imposible de prever mediante variables económicas.

Los capítulos siguientes se dedican a la discusión de los procedimientos de especificación, estimación y análisis de un modelo econométrico, considerando modelos de diversos tipos, ya sean lineales (Capítulo 3) o no lineales (Capítulo 11), estáticos o dinámicos (Capítulo 9), de una o varias ecuaciones (Capítulos 8, 17 y 18), con datos de series temporales (Capítulos 13 y 14), sección cruzada o de panel (Capítulo 15), e incluso el caso en que la variable a explicar no es de naturaleza cuantitativa (Capítulo 16). Se describen detalladamente los procedimientos de estimación dependiendo de las propiedades estadísticas del modelo (Capítulos 5, 6, 7 y 10), así como los procedimientos de contrastación de hipótesis sobre los valores de los coeficientes que contribuyan a aumentar nuestro conocimiento descriptivo de la economía (Capítulo 4).

⁽²⁾ La aproximación lineal mediante un desarrollo en serie de Taylor de la función $\ln(1+x)$ es: $\ln(1+x) = x$; ahora bien, si x_t es la tasa de crecimiento de la variable X_t , se tiene por definición: $x_t = (X_t - X_{t-1})/X_{t-1}$, por lo que: $X_t/X_{t-1} = 1 + x_t$, y por consiguiente: $\ln(1+x) = \ln(X_t/X_{t-1})$.

CAPITULO 1

ANALISIS MATRICIAL

1.1. PRIMERAS DEFINICIONES

Una *matriz* $m \times n$ es una colección de $m \cdot n$ elementos (constantes o variables), ordenados en m filas y n columnas. Denotamos dichos elementos por $a(i, j)$ o por a_{ij} con $i = 1, 2, \dots, m, j = 1, 2, \dots, n$. El primer índice denota la fila y el segundo la columna de la matriz \mathbf{A} a la que el elemento a_{ij} pertenece. Las *dimensiones* de una matriz son el par (m, n) que denotan el número de filas y de columnas de la matriz. Una matriz \mathbf{A} es *cuadrada* si $m = n$. En tal caso, diremos que \mathbf{A} es una matriz cuadrada de orden n . Un *vector fila* es una matriz con $m = 1$. Un *vector columna* es una matriz con $n = 1$. Si utilizamos la letra \mathbf{x} para denotar a un vector, escribiremos simplemente \mathbf{x} si dicho vector es un vector columna, y escribiremos \mathbf{x}' si es un vector fila. La dimensión de un vector es una única cifra, que indica el número de componentes del vector.

A partir de una matriz \mathbf{A} , una *submatriz* se obtiene eliminando algunas filas y columnas de \mathbf{A} .

La *diagonal principal* de una matriz cuadrada es la diagonal formada por los elementos $a_{ii}, i = 1, 2, \dots, n$. Una matriz es *diagonal* si todos los elementos fuera de la diagonal principal son cero. Una matriz es *diagonal a bloques* si los únicos elementos no nulos están ordenados como bloques de submatrices cuadradas a lo largo de la diagonal principal. Una matriz es *triangular superior* (*triangular inferior*) si todos los elementos por debajo (encima) de la diagonal principal son cero. Una matriz *escalar* es una matriz diagonal que tiene todos los elementos en la diagonal principal iguales entre sí. Una matriz *identidad* es una matriz escalar con elementos iguales a 1.

La *traspuesta* de una matriz \mathbf{A} $m \times n$ es una matriz \mathbf{B} $n \times m$ que tiene por filas y columnas las columnas y filas de \mathbf{A} . Esto se indica por $\mathbf{A}' = \mathbf{B}$. Si a_{ij} y b_{ij} denotan los elementos en la fila i , columna j de las matrices \mathbf{A} y \mathbf{B} , se tiene que $a_{ij} = b_{ji}$. Una matriz es *simétrica* si es igual a su traspuesta. Nótese que para que una matriz sea simétrica, es necesario que sea cuadrada.

1.1.a. Operaciones con matrices

Sean **A** y **B** dos matrices, ambas de dimensión $m \times n$. La suma de ambas es otra matriz **C**, también de dimensión $m \times n$, y tal que $c_{ij} = a_{ij} + b_{ij}$. La diferencia de ambas matrices es otra matriz **D**, de dimensión $m \times n$, tal que $d_{ij} = a_{ij} - b_{ij}$.

El producto de una matriz **A** $m \times n$ por una constante λ es otra matriz **B**, tal que $b_{ij} = \lambda a_{ij}$. Es decir, para multiplicar una matriz por una constante, se multiplica cada elemento de la matriz por dicha constante.

El producto **AB** de dos matrices **A** ($m \times n$) y **B** ($p \times q$) puede obtenerse sólo cuando $n = p$. En tal caso, el resultado es una matriz **C** ($m \times q$). El elemento (i, j) de la matriz **C** se obtiene multiplicando elemento a elemento la fila i -ésima de **A** por la columna j -ésima de **B**. Esta fila y columna mencionadas tienen n (o lo que es lo mismo, p) elementos cada una. Así, se tiene $c_{ij} = \sum_{s=1}^n a_{is} b_{sj}$.

Nótese que una matriz cuadrada es la única que puede multiplicarse por ella misma. Una matriz cuadrada **A** es *idempotente* si se tiene que $\mathbf{AA} = \mathbf{A}$. Cuando ello ocurre, entonces **A**, multiplicada por ella misma un número cualquiera de veces, es también igual a **A**. No es preciso que una matriz sea simétrica para ser idempotente, pero las matrices idempotentes que veremos en sucesivos capítulos resultan ser además simétricas.

El producto de un vector fila $1 \times n$ por un vector columna $n \times 1$ es un escalar. En particular, dado un vector columna \mathbf{a} ($n \times 1$), entonces el producto $\mathbf{a}'\mathbf{a}$ es un número, igual a la suma de los cuadrados de los n componentes del vector \mathbf{a} : $\mathbf{a}'\mathbf{a} = \sum_1^n a_i^2$. El producto de un vector columna $n \times 1$ por un vector fila $1 \times n$ también es factible, pero el resultado es ahora una matriz cuadrada de orden n : $\mathbf{aa}' = \mathbf{B}$, con $b_{ij} = a_i a_j$. Nótese que este producto es una matriz simétrica.

Es importante recordar que el producto de matrices, así como su suma, son ambas operaciones con propiedad asociativa. La suma tiene además la propiedad conmutativa. El producto tiene la propiedad distributiva con respecto a la suma, es decir: $\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$; $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$; $\mathbf{A}(\mathbf{BC}) = \mathbf{ABC} = (\mathbf{AB})\mathbf{C}$; $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

La traspuesta de una suma de matrices es igual a la suma de las traspuestas de las dos matrices **A** y **B**: $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$. La traspuesta de un producto de matrices es igual al producto de las traspuestas, pero en orden inverso: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. Nótese que si **A** es $m \times n$ y **B** es $n \times p$, éste es el único modo en que las dos traspuestas pueden multiplicarse. Dada una matriz **X** $m \times n$ cualquiera, entonces el producto $\mathbf{X}'\mathbf{X}$ es una matriz $n \times n$ simétrica.

La siguiente interpretación del producto de dos matrices tiene interés: Consideremos las matrices **M** = (m_{ij}) , $p \times n$ y la matriz **X** = (x_{ij}) , $n \times k$. Entonces su producto es otra matriz **R** $p \times k$:

$$\mathbf{MX} = \begin{pmatrix} m_{11}x_{11} + m_{12}x_{21} + \dots + m_{1n}x_{n1}, \dots, m_{11}x_{1k} + m_{12}x_{2k} + \dots + m_{1n}x_{nk} \\ m_{21}x_{11} + m_{22}x_{21} + \dots + m_{2n}x_{n1}, \dots, m_{21}x_{1k} + m_{22}x_{2k} + \dots + m_{2n}x_{nk} \\ \dots \\ m_{p1}x_{11} + m_{p2}x_{21} + \dots + m_{pn}x_{n1}, \dots, m_{p1}x_{1k} + m_{p2}x_{2k} + \dots + m_{pn}x_{nk} \end{pmatrix} = \mathbf{R}$$

que permite interpretar las columnas de \mathbf{R} como combinaciones lineales (k combinaciones) de las n columnas de la matriz \mathbf{M} . Así, la primera columna de \mathbf{R} es una combinación lineal de las columnas de la matriz \mathbf{M} , con coeficientes $x_{11}, x_{21}, \dots, x_{n1}$. La última columna de \mathbf{R} es una combinación lineal de las columnas de la matriz \mathbf{M} con coeficientes $x_{1k}, x_{2k}, \dots, x_{nk}$. Por otra parte, las filas de la matriz \mathbf{R} son p combinaciones lineales de las n filas de la matriz \mathbf{X} .

La traza de una matriz cuadrada es la suma de los elementos de su diagonal principal. Es inmediato ver que la traza de una suma de matrices se puede obtener como suma de las trazas de las matrices $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$. Cuando se multiplican tres o más matrices, entonces la traza del producto no se altera cuando se permutan circularmente los factores de dicho producto. Es decir, supongamos que \mathbf{A} , \mathbf{B} y \mathbf{C} son tres matrices $m \times n$, $n \times p$ y $p \times m$, respectivamente. Entonces se tiene:

$$tr(\mathbf{ABC}) = tr(\mathbf{CAB}) = tr(\mathbf{BCA})$$

Otra propiedad que utilizaremos en lo sucesivo es la siguiente: Supongamos que los elementos (algunos o todos) de una matriz \mathbf{A} $n \times n$ son variables aleatorias. Entonces es fácil ver que es equivalente calcular la esperanza matemática de cada elemento de la matriz y luego sumar los elementos que resultan en la diagonal principal, o sumar dichos elementos en la matriz original y luego calcular su esperanza matemática, es decir: $E[tr(\mathbf{A})] = tr[E(\mathbf{A})]$. Ello se debe a que la traza sería en este caso una suma de variables aleatorias.

Introducimos a continuación la notación con la que designaremos unas matrices que aparecerán habitualmente en el desarrollo de los temas que siguen: Denotamos por \mathbf{I}_n la matriz identidad de orden n . Dada una matriz \mathbf{A} , de dimensión $m \times n$, se tiene $\mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A}$. Denotaremos por $\mathbf{0}_{m \times n}$ una matriz $m \times n$ que tiene todos sus elementos iguales a cero. La notación $\mathbf{1}_n$ denota un vector columna de dimensión n , con todos sus componentes iguales a 1, mientras que $\mathbf{0}_n$ denota un vector columna de dimensión n , con todos sus componentes iguales a cero. Algunas propiedades de fácil comprobación son:

a) Dado un vector \mathbf{y} de dimensión n se tiene $\frac{1}{n} \mathbf{1}'_n \mathbf{y} = \bar{y}$, donde \bar{y} es la media aritmética de los componentes del vector \mathbf{y} .

b) Dada una matriz \mathbf{X} de dimensión $m \times n$ se tiene $\frac{1}{m} \mathbf{1}'_m \mathbf{X} = \bar{\mathbf{x}}'$, donde $\bar{\mathbf{x}}'$ es el vector fila de dimensión n formado por las medias aritméticas de los elementos que integran cada columna de la matriz \mathbf{X} .

Denotaremos por \mathbf{Q} una matriz de dimensión $n \times n$, que nos será de especial utilidad en la discusión de algunas cuestiones, y en particular en la Sección 3.8. Esta matriz viene definida por:

$$\mathbf{Q} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n$$

Es fácil ver (lo dejamos como ejercicio) que esta matriz es simétrica e idempotente. Otras propiedades suyas que el lector puede comprobar sin dificultad son:

$$a) \quad \mathbf{Qy} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \dots \\ y_n - \bar{y} \end{pmatrix}; \quad \mathbf{Q1}_n = \mathbf{0}_n$$

es decir, que la matriz \mathbf{Q} transforma un vector en el vector de diferencias con respecto a la media aritmética de sus componentes. Como consecuencia, si el vector \mathbf{y} tiene sus componentes iguales entre sí, entonces el producto \mathbf{Qy} es igual al vector cero. Otra consecuencia es que si la media aritmética de las componentes del vector \mathbf{y} es cero, entonces $\mathbf{Qy} = \mathbf{y}$.

b) Dada una matriz \mathbf{A} $n \times p$ se tiene:

$$\mathbf{QA} = \begin{pmatrix} a_{11} - \bar{a}_1, & a_{12} - \bar{a}_2, & \dots, & a_{1p} - \bar{a}_p \\ a_{21} - \bar{a}_1, & a_{22} - \bar{a}_2, & \dots, & a_{2p} - \bar{a}_p \\ a_{31} - \bar{a}_1, & a_{32} - \bar{a}_2, & \dots, & a_{3p} - \bar{a}_p \\ \dots & \dots & \dots & \dots \\ a_{n1} - \bar{a}_1, & a_{n2} - \bar{a}_2, & \dots, & a_{np} - \bar{a}_p \end{pmatrix}$$

es decir, que la matriz \mathbf{Q} transforma una matriz \mathbf{A} en la matriz de diferencias con respecto a las medias aritméticas de cada columna de \mathbf{A} . Una forma intuitiva de ver este resultado es considerando la matriz \mathbf{A} como una colección de p vectores columna, cada uno con n componentes, y utilizar la propiedad a) aplicada a cada vector columna.

Otra propiedad que utilizaremos con frecuencia y que el lector puede probar sin dificultad es:

Dada una matriz \mathbf{X} , de dimensión $m \times n$, la matriz $\mathbf{M} = \mathbf{I}_m - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es simétrica, idempotente y cumple $\mathbf{MX} = \mathbf{0}_{m \times n}$.

1.2. DETERMINANTES

El *determinante* es una función real de una matriz cuadrada, es decir, una función que asocia un número real definido de modo unívoco a toda matriz cuadrada. Las matrices cuadradas más sencillas son 1×1 , y su determinante es numéricamente igual al único elemento en dicha matriz. Dada una matriz 2×2 :

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

su determinante es igual a $a_{11}a_{22} - a_{12}a_{21}$. Dada una matriz 3×3 :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

su determinante es igual a $a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{12}a_{31}a_{23} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{21}a_{12}a_{33}$.

Dada una matriz cuadrada $n \times n$, se llama *cofactor* del elemento (i, j) de dicha matriz al determinante de la matriz $(n - 1) \times (n - 1)$ que resulta al suprimir de la matriz original la fila i -ésima y la columna j -ésima, multiplicado por $(-1)^{i+j}$. Puede comprobarse que si en la matriz 3×3 anterior se elige una fila o columna, se multiplica cada elemento de dicha línea por su correspondiente cofactor y se suman los tres resultados así obtenidos, se obtiene precisamente el determinante de la matriz 3×3 . Este resultado no es casual, sino que es válido para toda matriz cuadrada, y así se tiene: El determinante de una matriz cuadrada puede obtenerse desarrollando por los elementos de una línea cualquiera (fila o columna), sin más que multiplicar cada uno de los elementos de dicha línea por su cofactor correspondiente, y sumando los productos obtenidos. Utilizando este resultado puede obtenerse el determinante de una matriz 4×4 , utilizando la expresión que antes vimos para el determinante de una matriz 3×3 . Una vez que sabemos calcular éste, se puede obtener el determinante de matrices 5×5 , y así sucesivamente. Las siguientes son propiedades fundamentales del determinante de una matriz:

- Una matriz cuadrada y su traspuesta tienen el mismo determinante.
- Si se intercambian de orden dos líneas cualesquiera de una matriz (dos filas o dos columnas entre sí), entonces el determinante de la matriz cambia de signo.
- Si una matriz tiene dos filas o columnas iguales entre sí, su determinante es cero.
- Una matriz con una línea de ceros tiene determinante igual a cero.
- Si se multiplican los elementos de una línea por los cofactores de una línea paralela, el resultado es siempre cero.
- Si se suma o resta a una línea una combinación lineal de líneas paralelas, el determinante no varía.
- Una matriz tiene una línea que es combinación lineal de otras si y sólo si su determinante es cero.
- El determinante del producto de dos matrices cuadradas de igual orden es igual al producto de sus respectivos determinantes.
- Si se multiplica una línea de una matriz $n \times n$ por una constante, el determinante queda multiplicado por esa constante. Si se multiplica una matriz cuadrada por una constante, su determinante queda multiplicado por dicha constante elevada a la potencia n -ésima.
- El determinante de una matriz diagonal o de una matriz triangular es igual al producto de los elementos de la diagonal.

Una matriz con determinante igual a cero se dice *singular*.

1.3. MATRIZ INVERSA

Dada una matriz cuadrada \mathbf{A} $n \times n$, si existe otra matriz \mathbf{B} $n \times n$, tal que $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, entonces se dice que la matriz \mathbf{B} es la *inversa* de la matriz \mathbf{A} , y se denota $\mathbf{A}^{-1} = \mathbf{B}$. Una matriz cuya inversa es igual a su traspuesta ($\mathbf{A}^{-1} = \mathbf{A}'$) se llama *ortogonal*. Una matriz cuadrada \mathbf{A} tiene inversa siempre que no es *singular*. Dicha matriz inversa puede calcularse numéricamente de la siguiente forma:

- Dada la matriz \mathbf{A} , no singular, se sustituye cada elemento de \mathbf{A} por su cofactor, teniendo cuidado de no olvidar el signo en el cálculo de los cofactores.
- Se traspone la matriz así obtenida.
- Se divide cada elemento en esta nueva matriz por el determinante de la matriz original \mathbf{A} .

Debe quedar claro que el cálculo de la inversa de una matriz de orden 4 o superior es una tarea laboriosa si se hace a mano, cuya dificultad se incrementa más que proporcionalmente con el orden de la matriz a invertir. Por suerte, hoy en día todos los paquetes de ordenador tienen instrucciones o subrutinas que permiten el cálculo de dichas matrices inversas sin que necesitemos hacerlo manualmente. El aspecto interesante que surge es que teóricamente toda matriz cuyo determinante es numéricamente distinto de cero es invertible, pero, sin embargo, hay muchos casos en que el determinante de una matriz es lo suficientemente próximo a cero (aunque distinto de cero), como para que la subrutina que el ordenador utiliza no sea capaz de obtener su matriz inversa. Otras veces la subrutina funciona, pero como al obtener la matriz inversa dividimos por el determinante de la matriz original, se obtienen matrices inversas con elementos gigantescos. (En este sentido la inversa de una matriz es algo análogo al inverso de un número real.) En estos casos, pequeñas diferencias numéricas en la matriz original pueden producir enormes diferencias en la matriz inversa.

Las siguientes son propiedades fundamentales de la inversa de una matriz:

- Si invertimos dos veces una matriz, se obtiene la matriz original.
- Invertir y transponer matrices son dos operaciones cuyo orden de aplicación puede permutarse sin que se altere el resultado obtenido, es decir: $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$.
- La inversa del producto de dos matrices cuadradas es el producto de las inversas de los factores, pero con el orden invertido: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, siempre y cuando las inversas, tanto de \mathbf{A} como de \mathbf{B} , existan.
- Las inversas de matrices triangulares superiores o inferiores son matrices del mismo tipo.
- La inversa de una matriz diagonal es también diagonal. Sus elementos son los inversos de los elementos de la matriz original. La inversa de una matriz diagonal a bloques es una matriz del mismo tipo. Cada uno de sus bloques es el inverso de los bloques de la matriz original.

- Los determinantes de una matriz y de su inversa son inversos el uno del otro: $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$.
- Un útil resultado se refiere a la inversa de una expresión matricial del tipo $\mathbf{A} + \mathbf{gh}'$, donde \mathbf{A} es una matriz $n \times n$ y \mathbf{g} , \mathbf{h} son vectores $g \times 1$. Se tiene:

$$(\mathbf{A} + \mathbf{gh}')^{-1} = \left(\mathbf{I}_n - \mathbf{A}^{-1} \frac{\mathbf{gh}'}{1 + \mathbf{h}'\mathbf{A}^{-1}\mathbf{g}} \right) \mathbf{A}^{-1}$$

- Dada una matriz \mathbf{A} , $n \times n$, supongamos que se particiona:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

donde \mathbf{A}_{11} , \mathbf{A}_{12} , \mathbf{A}_{21} y \mathbf{A}_{22} son respectivamente $p \times p$, $p \times q$, $q \times p$ y $q \times q$, con $p + q = n$. Sea \mathbf{B} , $n \times n$, la matriz inversa de \mathbf{A} , y particionemos \mathbf{B} de modo análogo a como hicimos con \mathbf{A} :

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

donde cada bloque \mathbf{B}_{ij} tiene las mismas dimensiones que la submatriz \mathbf{A}_{ij} . Las matrices cuadradas \mathbf{B}_{11} y \mathbf{B}_{22} no son las inversas de los bloques \mathbf{A}_{11} y \mathbf{A}_{22} , pero hay una relación entre ellas que resulta ser de gran utilidad para el desarrollo de los capítulos que siguen:

$$\begin{aligned} \mathbf{B}_{11} &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} & \mathbf{B}_{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11} \end{aligned}$$

Aunque estas fórmulas son algo aparatosas, no es difícil recordarlas, pues los productos que en ellas aparecen sólo podrían efectuarse de dicho modo si se quiere que ambos miembros de las igualdades tengan las mismas dimensiones.

Este es sólo un ejemplo de lo sencillo que resulta trabajar con matrices particionadas en bloques cuando las dimensiones de dichos bloques permiten efectuar las operaciones deseadas. En realidad, se maneja cada bloque como si de un elemento se tratase. Por ejemplo, supongamos que se quiere efectuar el producto de \mathbf{A} , una matriz $m \times n$, por la matriz \mathbf{B} $n \times p$, que $m_1 + m_2 = m$, $n_1 + n_2 = n$ y $p_1 + p_2 = p$, y que descomponemos ambas matrices:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

donde \mathbf{A}_{11} es $m_1 \times n_1$; \mathbf{A}_{12} , $m_1 \times n_2$; \mathbf{A}_{21} , $m_2 \times n_1$; \mathbf{A}_{22} , $m_2 \times n_2$; \mathbf{B}_{11} , $n_1 \times p_1$; \mathbf{B}_{12} , $n_1 \times p_2$; \mathbf{B}_{21} , $n_2 \times p_1$, y \mathbf{B}_{22} , $n_2 \times p_2$. Entonces se tiene:

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}$$

El producto de Kronecker de una matriz \mathbf{A} de dimensión $m \times n$ y una matriz \mathbf{B} de dimensión $p \times q$ es otra matriz \mathbf{C} de dimensión $mp \times nq$, definida por:

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & a_{13}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & a_{23}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & a_{m3}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}$$

donde a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$ son los elementos de la matriz \mathbf{A} .

La inversa de la matriz \mathbf{C} es igual a $\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. Nótese que, a diferencia de la inversa del producto habitual de dos matrices, en este caso no se altera el orden de los factores. De modo similar, se tiene $\mathbf{C}' = \mathbf{A}' \otimes \mathbf{B}'$.

En varias ocasiones a lo largo del texto es preciso obtener la derivada de un producto de matrices con respecto a uno o varios parámetros, para lo que será conveniente recordar las siguientes reglas:

a) La derivada de una función $f(\theta_1, \theta_2, \dots, \theta_k)$ con respecto al vector $(\theta_1, \theta_2, \dots, \theta_k)$ es igual a un vector columna, llamado el *gradiente* de la función. El vector tiene k componentes, cada uno de ellos igual a una de las derivadas parciales: $\partial f / \partial \theta_i$.

b) La matriz hessiana de la función $f(\theta_1, \theta_2, \dots, \theta_k)$ es una matriz cuadrada, de dimensión $k \times k$, cada uno de cuyos componentes es igual a $\partial^2 f / \partial \theta_i \partial \theta_j$. En particular, los elementos de la diagonal de dicha matriz son las derivadas segundas con respecto a cada uno de los parámetros θ_i .

c) El gradiente de la función lineal $\mathbf{A}\boldsymbol{\theta}$, donde \mathbf{A} es $k \times k$ y $\boldsymbol{\theta}$ $k \times 1$, con respecto al vector $\boldsymbol{\theta}$, es la matriz \mathbf{A} .

d) La derivada de la forma cuadrática $\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}$ con respecto al vector $\boldsymbol{\theta}$ es igual a $2\mathbf{A}\boldsymbol{\theta}$. La derivada de $\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}$ respecto de \mathbf{A} es igual a $\boldsymbol{\theta}\boldsymbol{\theta}'$.

e) La derivada del determinante $|\mathbf{A}|$ con respecto a la matriz \mathbf{A} es igual a la matriz $|\mathbf{A}|(\mathbf{A}')^{-1}$.

f) La derivada de $\ln|\mathbf{A}|$ con respecto a \mathbf{A} es igual a $(\mathbf{A}')^{-1}$, y en particular la derivada de $\ln|\mathbf{A}|$ con respecto a a_{ij} es a^{ji} .

1.4. RANGO DE UNA MATRIZ

Una matriz \mathbf{A} , de dimensión $m \times n$, puede interpretarse como una colección de m vectores fila de dimensión n , o como una colección de n vectores columna de dimensión m . Ello permite aplicar la teoría algebraica de los

espacios vectoriales a las filas o columnas de una matriz. En particular, puede hablarse de filas linealmente independientes o dependientes, en el sentido en que vectores de dimensión n lo son, y lo mismo con las columnas de la matriz.

Como vamos a ver, el máximo número de filas linealmente independientes en una matriz es igual al máximo número de columnas linealmente independientes. Dicho número se llama *rango* de la matriz. De ello se deduce inmediatamente que una matriz y su traspuesta tienen el mismo rango. También se deduce que, dada una matriz A $m \times n$, su rango es $\leq \min\{m, n\}$.

Lema 1.1. El máximo número de filas linealmente independientes de una matriz cualquiera es igual al máximo número de columnas linealmente independientes.

Demostración. Sea A una matriz $m \times n$ y sea p el máximo número de filas linealmente independientes. Tomemos p filas linealmente independientes y formemos con ellas una submatriz A^* , $p \times n$. Sea q el máximo número de columnas linealmente independientes de la matriz A . Como no puede haber más columnas linealmente independientes en A^* que en A , se tiene que q es también el máximo número de columnas linealmente independientes de A^* . Cada columna de la matriz A^* tiene p elementos, lo que implica que $q \leq p$, puesto que el máximo número de vectores de dimensión p linealmente independientes entre sí es, precisamente, igual a p .

Por otra parte, construyamos ahora A^{**} , la matriz $m \times q$ que se obtiene con q columnas linealmente independientes de la matriz A . Como p es el máximo número de filas linealmente independientes de A , también es el máximo número de filas linealmente independientes de A^{**} , y como cada una de ellas tiene q elementos, entonces $p \leq q$, por lo que, finalmente, $p = q$.

Si una línea (fila o columna) de una matriz es combinación lineal de las demás, su determinante es cero, y la matriz es singular. Por tanto, si una matriz cuadrada de orden n es no singular, su rango es igual a n . Recíprocamente, si una matriz cuadrada es singular, entonces una cualquiera de sus filas puede escribirse como combinación lineal de las restantes filas, y lo mismo ocurre con cada una de sus columnas.

Una propiedad importante del rango de una matriz es que no cambia si se premultiplica o postmultiplica dicha matriz por una matriz cuadrada no singular. Es importante recordar la interpretación del producto de dos matrices AB (donde B es una matriz cuadrada no singular), como otra matriz C cuyas columnas son combinación lineal de las columnas de A . Siempre que a partir de un conjunto de n vectores se genera otro conjunto tomando combinaciones lineales de los primeros, el número de vectores que son linealmente independientes en el segundo conjunto puede ser como mucho igual al número de vectores linealmente independientes en el primer conjunto, pero nunca superior. Puesto que la matriz producto C es un conjunto de n vectores columna que son combinación lineal de los n vectores columna de A , entonces el número de columnas de C linealmente independientes no puede ser superior al de A , aunque sí que podría ser inferior. Este resultado admite

un nivel de generalización adicional: Dadas dos matrices \mathbf{A} $m \times n$ y \mathbf{B} $n \times p$, se tiene que $\text{Rango}(\mathbf{AB}) \leq \min\{\text{Rango}(\mathbf{A}), \text{Rango}(\mathbf{B})\}$, proposición que pasamos a probar tras unos resultados preliminares.

Sea \mathbf{A} una matriz $m \times n$ con $\text{Rango}(\mathbf{A}) = r \leq \min\{m, n\}$ y sea \mathbf{x} un vector de dimensión n . Particionemos \mathbf{A} y \mathbf{x} del siguiente modo:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$$

donde \mathbf{A}_{11} es $r \times r$; \mathbf{A}_{12} , $r \times (n - r)$; \mathbf{A}_{21} , $(m - r) \times r$, y \mathbf{A}_{22} , $(m - r) \times (n - r)$. Por otra parte, \mathbf{x}_1 es de dimensión r , mientras que \mathbf{x}_2 es de dimensión $n - r$. Las filas y columnas de \mathbf{A} se han ordenado de forma que \mathbf{A}_{11} es no singular. Ello es posible ya que $\text{Rango}(\mathbf{A}) = r$.

Consideremos ahora el sistema $\mathbf{Ax} = \mathbf{0}$, cuyo conjunto de soluciones constituye el *subespacio nulo* de la matriz \mathbf{A} , que denotamos por $\text{SN}(\mathbf{A})$. Abandonando las últimas $m - r$ filas de \mathbf{A} se tiene el sistema de r ecuaciones en $n (\geq r)$ incógnitas: $\mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 = \mathbf{0}$. La información contenida en las $m - r$ ecuaciones desechadas está incorporada en el subsistema de r ecuaciones seleccionado, pues los $m - r$ filas de \mathbf{A} que se abandonaron son combinación lineal de las r filas que se han conservado. Si el determinante de \mathbf{A}_{11} no es cero, entonces puede resolverse $\mathbf{x}_1 = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{x}_2$, es decir, que dado \mathbf{x}_2 , \mathbf{x}_1 queda determinado. Por tanto, $n - r$ vectores determinan todas las soluciones al sistema $\mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 = \mathbf{0}$, y se tiene:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{I}_{n-r} \end{pmatrix} \mathbf{x}_2$$

que da las soluciones al sistema de ecuaciones anterior como combinaciones lineales de las columnas de la matriz $\begin{pmatrix} -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{I}_{n-r} \end{pmatrix}$ (con coeficientes igual a las coordenadas del vector \mathbf{x}_2). Estas columnas constituyen un conjunto de $n - r$ vectores de dimensión n , y son linealmente independientes debido a la presencia de la submatriz \mathbf{I}_{n-r} . Como consecuencia, la dimensión del subespacio nulo es igual a $n - r$. Por tanto, *el número de columnas de la matriz \mathbf{A} (n) es igual a su rango (r) más la dimensión del subespacio nulo ($n - r$)*⁽¹⁾.

Ejemplo 1.1. La matriz

$$\begin{pmatrix} -1 & 0 & 2 & 4 \\ 2 & 1 & 0 & 3 \\ 0 & 1 & 4 & 11 \end{pmatrix}$$

⁽¹⁾ El subespacio nulo de una matriz podría también definirse como el conjunto de vectores en R^n , que son solución del sistema $\mathbf{yA} = \mathbf{0}_m$. En tal caso, el número de filas de \mathbf{A} es igual al rango de \mathbf{A} más la dimensión del subespacio nulo.

tiene rango igual a 2. Si eliminamos una fila (la segunda), se tiene, como sistema $A_{11}x_1 + A_{12}x_2 = 0$:

$$\begin{aligned}x_1 &= 2x_3 + 4x_4 \\x_2 &= -11x_4 + 4x_3\end{aligned}$$

por lo que la dimensión del subespacio nulo es 2:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ -11 & -4 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_4 \\ x_3 \end{pmatrix}$$

mientras que el número de columnas de la matriz de partida es 4.

Excluyendo la tercera ecuación se llegaría a:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2 & -3 \\ \frac{1}{2} & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_4 \end{pmatrix}$$

pero estas dos columnas son combinación lineal de las dos columnas de la matriz anterior, y representan por tanto el mismo subespacio nulo.

Teorema 1.1. Dada una matriz \mathbf{X} , $T \times k$, con $\text{Rango}(\mathbf{X}) = p \leq \min\{T, k\}$, se tiene:

$$\text{Rango}(\mathbf{X}'\mathbf{X}) = \text{Rango}(\mathbf{X}\mathbf{X}') = \text{Rango}(\mathbf{X})$$

Demostración. 1. Demostramos, en primer lugar, que el rango de una matriz no cambia al premultiplicarla por su traspuesta. El subespacio nulo de la matriz \mathbf{X} tiene dimensión $k - p$, que es mayor o igual que cero. Sea \mathbf{m} un vector perteneciente a dicho subespacio. Entonces se tiene $\mathbf{X}\mathbf{m} = \mathbf{0}_T$ y por consiguiente $\mathbf{X}'\mathbf{X}\mathbf{m} = \mathbf{0}_k$. Por tanto, $\text{Subespacio Nulo}(\mathbf{X}) \subset \text{Subespacio Nulo}(\mathbf{X}'\mathbf{X})$. Sea ahora un vector \mathbf{s} del subespacio nulo de la matriz $\mathbf{X}'\mathbf{X}$. Entonces se tiene $(\mathbf{X}'\mathbf{X})\mathbf{s} = \mathbf{0}_k$, y por tanto $\mathbf{s}'(\mathbf{X}'\mathbf{X})\mathbf{s} = (\mathbf{X}\mathbf{s})'(\mathbf{X}\mathbf{s}) = 0$, lo que solamente puede ocurrir si $\mathbf{X}\mathbf{s} = \mathbf{0}_T$, es decir, si $\mathbf{s} \in \text{SN}(\mathbf{X})$. En consecuencia, ambas matrices tienen el mismo subespacio nulo, por lo que la dimensión del $\text{SN}(\mathbf{X}'\mathbf{X})$ es también igual a $k - p$ y por tanto $\text{Rango}(\mathbf{X}'\mathbf{X}) = p$.

Nuestro resultado 1 puede escribirse como $\text{Rango}(\mathbf{X}\mathbf{X}') = \text{Rango}(\mathbf{X})$, pero una matriz y su traspuesta tienen igual rango, por lo que, finalmente, $\text{Rango}(\mathbf{X}\mathbf{X}') = \text{Rango}(\mathbf{X}) = \text{Rango}(\mathbf{X}'\mathbf{X})$.

Como aplicación de este resultado, supongamos que $T > k$ y que las k columnas de \mathbf{X} son linealmente independientes. En tal caso, $\text{Rango}(\mathbf{X}) = k = \text{Rango}(\mathbf{X}'\mathbf{X}) = \text{Rango}(\mathbf{X}\mathbf{X}')$. Como $\mathbf{X}'\mathbf{X}$ es una matriz $k \times k$, este resultado

implica que dicha matriz es no singular y por tanto invertible. En cambio, XX' es una matriz $T \times T$, y el resultado implica que es singular.

Teorema 1.2. El rango de una matriz A no cambia al premultiplicar o postmultiplicar dicha matriz por una matriz cuadrada no singular. Es decir, si A es una matriz $m \times n$ y P , Q son matrices $m \times m$ y $n \times n$ no singulares, se tiene: $\text{Rango}(A) = \text{Rango}(PA) = \text{Rango}(AQ) = \text{Rango}(PAQ)$.

Demostración. 1. Si $x \in \text{SN}(A)$, entonces $Ax = 0$ y por tanto $PAx = 0$, por lo que $x \in \text{SN}(PA)$. Por otra parte, si $x \in \text{SN}(PA)$, entonces $PAx = 0$, lo que implica que $P^{-1}(PAx) = Ax = 0$ y $x \in \text{SN}(A)$. En consecuencia, como A y PA tienen igual número de columnas, también tienen el mismo rango.

2. Puesto que el rango de una matriz es igual al rango de su traspuesta, se tiene: $\text{Rango}(AQ) = \text{Rango}(Q'A')$. De acuerdo con 1 se tiene: $\text{Rango}(Q'A') = \text{Rango}(A')$, y es igual a $\text{Rango}(A)$.

3. $\text{Rango}(PAQ) = \text{Rango}(A)$ es ahora obvio, sin más que utilizar 1 y 2 sucesivamente.

Teorema 1.3. $\text{Rango}(AB) \leq \min\{\text{Rango}(A), \text{Rango}(B)\}$.

Demostración. 1. Sea x un vector del $\text{SN}(B)$. Entonces $Bx = 0$ y por tanto $ABx = 0$, por lo que $x \in \text{SN}(AB)$. Por tanto, $\text{SN}(B) \subset \text{SN}(AB)$. Como B y AB tienen el mismo número de columnas, entonces $\text{Rango}(B) \geq \text{Rango}(AB)$.

2. $\text{Rango}(AB) = \text{Rango}(B'A') \leq \text{Rango}(A') = \text{Rango}(A)$, donde la desigualdad proviene del resultado 1.

1.5. VALORES Y VECTORES PROPIOS DE UNA MATRIZ

Dada una matriz cuadrada A $n \times n$, entonces una constante λ y un vector x , $n \times 1$ no nulo que satisfagan el sistema de ecuaciones $Ax = \lambda x$ se llaman, respectivamente, *valor propio* y *vector propio* de la matriz A . Si la matriz $A - \lambda I_n$ no es singular, entonces la única solución a la ecuación anterior es la trivial: $x = 0_n$. Para que haya una solución no nula, debe ocurrir que $|A - \lambda I_n| = 0$. Esta es la *ecuación característica* de la matriz A . Dicha ecuación tiene n soluciones, aunque no necesariamente diferentes entre sí. Por ser las raíces de un polinomio de grado n , los valores propios pueden, en principio, ser números complejos.

Para cada solución (valor propio) existe un vector propio asociado, que se obtiene sustituyendo el valor de λ en la ecuación $Ax = \lambda x$. Cuando un valor propio es una solución múltiple de la ecuación característica, entonces el número de vectores propios linealmente independientes asociados con dicho valor propio es igual a su orden de multiplicidad.

Es importante observar que cada vector propio correspondiente a un valor propio no nulo es la solución a un sistema lineal homogéneo y, por tanto, no está únicamente determinado, sino que es, en realidad, un subespacio de

dimensión uno. Para determinar unívocamente el vector propio se utilizan distintos criterios, como normalizar los vectores propios para que tengan módulo igual a 1, lo que haremos de ahora en adelante.

Como ejemplo puede comprobarse que los valores propios de la matriz $\mathbf{A} = \begin{pmatrix} 5 & 3/2 \\ 3/2 & 1 \end{pmatrix}$ son $\frac{11}{2}$ y $\frac{1}{2}$, y que sus vectores propios, normalizados para que tengan módulo unidad, son $\left(\frac{3}{\sqrt{10}}; \frac{1}{\sqrt{10}}\right)$ y $\left(\frac{1}{\sqrt{10}}; -\frac{3}{\sqrt{10}}\right)$. Además, puede comprobarse que estos dos vectores propios son ortogonales entre sí. Aunque puede hablarse de vectores y valores propios de cualquier matriz cuadrada, en este texto sólo calcularemos valores y vectores propios de matrices simétricas.

Las siguientes son propiedades fundamentales de los valores propios de una matriz:

Proposición 1.1. Los valores propios de una matriz simétrica son reales: En el caso de una matriz $\mathbf{A} = (a_{ij})$, 2×2 , los valores propios vienen dados por $\lambda = \frac{1}{2} [(a_{11} + a_{22}) \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}]$. Si la matriz es simétrica, entonces $\lambda = \frac{1}{2} [(a_{11} + a_{22}) \pm \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}]$ y por tanto son reales, ya que el radicando es positivo.

Proposición 1.2. Los vectores propios correspondientes a distintos valores propios de una matriz simétrica son ortogonales entre sí, es decir, su producto es cero.

Demostración. Sean \mathbf{x}_1 y \mathbf{x}_2 vectores propios de la matriz \mathbf{A} correspondientes a los valores propios λ_1 y λ_2 , respectivamente. Entonces se tiene $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ y $\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_2$. Por tanto, $\mathbf{x}_2'\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_2'\mathbf{x}_1$ y $\mathbf{x}_1'\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_1'\mathbf{x}_2$. Ahora bien, $\mathbf{x}_1'\mathbf{A}\mathbf{x}_2$ y $\mathbf{x}_2'\mathbf{A}\mathbf{x}_1$ son matrices traspuestas una de la otra, y de dimensión 1×1 , es decir, escalares, y por tanto son iguales entre sí. En consecuencia, $(\lambda_1 - \lambda_2)\mathbf{x}_1'\mathbf{x}_2 = 0$. Si $\lambda_1 \neq \lambda_2$, entonces, necesariamente, $\mathbf{x}_1'\mathbf{x}_2 = 0$.

Esta proposición tiene una aplicación de interés: dada una matriz \mathbf{A} , de dimensión $n \times n$, construyamos otra matriz \mathbf{B} que tiene por columnas los vectores propios de la matriz \mathbf{A} . Como acabamos de ver, dichas columnas son vectores ortogonales entre sí. Por tanto, si formamos el producto $\mathbf{C} = \mathbf{B}'\mathbf{B}$, su elemento genérico c_{ij} está formado por el producto de las columnas i y j de \mathbf{B} . Dicho producto será cero, salvo cuando $i = j$, en cuyo caso será igual a 1, de acuerdo con nuestro convenio de normalización. Por tanto, $\mathbf{B}'\mathbf{B} = \mathbf{I}_n$; pero, por esta razón, también se tiene $\mathbf{B}\mathbf{B}' = \mathbf{I}_n$, y \mathbf{B}' es la inversa de \mathbf{B} , por lo que \mathbf{B} es una matriz ortogonal.

Proposición 1.3. (Descomposición canónica de una matriz cuadrada.)

Dada una matriz simétrica \mathbf{A} $n \times n$, sea \mathbf{B} la matriz que tiene por columnas los vectores propios de \mathbf{A} . Entonces, se cumple:

$$\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{\Gamma}$$

donde $\mathbf{\Gamma}$ es una matriz diagonal que tiene por elementos los valores propios de \mathbf{A} .

Demostración. En efecto, si denotamos por \mathbf{C} al producto $\mathbf{B}'\mathbf{A}\mathbf{B}$, el elemento genérico c_{ij} se obtiene multiplicando la fila i -ésima de \mathbf{B}' (columna i -ésima de \mathbf{B}) por \mathbf{A} y por la columna j -ésima de \mathbf{B} : $\mathbf{x}'_i\mathbf{A}\mathbf{x}_j = \mathbf{x}'_i(\lambda_j\mathbf{x}_j) = \lambda_j(\mathbf{x}'_i\mathbf{x}_j)$. Por tanto, dicho producto es cero salvo cuando i es igual a j , en cuyo caso $\mathbf{x}'_i\mathbf{A}\mathbf{x}_i = \lambda_i(\mathbf{x}'_i\mathbf{x}_i) = \lambda_i$.

Proposición 1.4. Dada una matriz \mathbf{A} simétrica, definida positiva, existe una matriz no singular \mathbf{P} tal que $\mathbf{A} = \mathbf{P}\mathbf{P}'$.

Demostración. Por la descomposición anterior $\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{\Gamma}$ y por tanto $\mathbf{B}(\mathbf{B}'\mathbf{A}\mathbf{B})\mathbf{B}' = \mathbf{B}\mathbf{\Gamma}\mathbf{B}'$. Pero \mathbf{B} es una matriz ortogonal, por lo que $\mathbf{B}\mathbf{B}' = \mathbf{B}'\mathbf{B} = \mathbf{I}_n$, y finalmente $\mathbf{A} = \mathbf{B}\mathbf{\Gamma}\mathbf{B}'$. Si denotamos por $\mathbf{\Gamma}^{\frac{1}{2}}$ la matriz diagonal que tiene por elementos las raíces cuadradas de los valores propios de la matriz \mathbf{A} , entonces $\mathbf{A} = \mathbf{P}\mathbf{P}'$, donde $\mathbf{P} = (\mathbf{B}\mathbf{\Gamma}^{\frac{1}{2}})$.

Notemos que \mathbf{A} no necesita ser simétrica. Basta que tenga k vectores propios linealmente independientes, en cuyo caso $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{\Gamma}$, donde las columnas de \mathbf{B} son los vectores propios de \mathbf{A} , y $\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \mathbf{B}^{-1}\mathbf{B}\mathbf{\Gamma} = \mathbf{\Gamma}$, sólo que ahora \mathbf{B} no es necesariamente ortogonal.

Proposición 1.5. La suma de los valores propios de una matriz \mathbf{A} es igual a su *traza*, es decir, a la suma de los elementos de su diagonal principal.

Demostración. Por la descomposición antes vista: $\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{\Gamma}$, donde \mathbf{B} es una matriz ortogonal, y utilizando la propiedad circular de la traza: $\text{traza}(\mathbf{\Gamma}) = \text{traza}(\mathbf{B}'\mathbf{A}\mathbf{B}) = \text{traza}(\mathbf{B}\mathbf{B}'\mathbf{A}) = \text{traza}(\mathbf{A})$, pero $\text{traza}(\mathbf{\Gamma})$ es, simplemente, la suma de los valores propios de la matriz \mathbf{A} .

Proposición 1.6. El producto de los valores propios de una matriz es igual al determinante de dicha matriz. De ello se deduce que una matriz es singular si y sólo si tiene al menos un valor propio igual a cero.

Demostración (para matrices simétricas). Utilizando de nuevo la descomposición $\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{\Gamma}$ se tiene $|\mathbf{B}'||\mathbf{A}||\mathbf{B}| = |\mathbf{\Gamma}| = \prod_{i=1}^n \lambda_i$. Ahora bien, por ser \mathbf{B} una matriz ortogonal se tiene $\mathbf{B}'\mathbf{B} = \mathbf{I}_n$ y por tanto $|\mathbf{B}'\mathbf{B}| = |\mathbf{B}'||\mathbf{B}| = |\mathbf{B}|^2 = 1$. Consecuentemente, $|\mathbf{B}| = +1$ o -1 , lo que llevado a la expresión anterior implica finalmente $|\mathbf{A}| = \prod_{i=1}^n \lambda_i$.

Proposición 1.7. El rango de una matriz es igual al número de valores propios no nulos.

Demostración. Puesto que el rango de una matriz no varía al pre o postmultiplicar dicha matriz por una matriz no singular, entonces por la descomposición canónica anterior se tiene $\text{Rango}(A) = \text{Rango}(\Gamma)$. Ahora bien, las filas de esta última matriz sólo tienen un elemento no nulo, aquél en la diagonal de Γ . Tales filas serán necesariamente independientes entre sí, salvo que alguno de dichos elementos (valores propios de A) sea igual a cero.

Proposición 1.8. Los valores propios de la matriz $A^2 (=AA)$ son igual al cuadrado de los valores propios de la matriz A . Sin embargo, los vectores propios de A^2 son iguales a los de A . (Claramente, este resultado prueba que el determinante de la matriz A^2 es positivo.)

Demostración. Si x es un vector propio asociado al valor propio λ , entonces $Ax = \lambda x$, por lo que también se tiene $A^2x = A(Ax) = \lambda(Ax) = \lambda^2x$, por lo que λ^2 es un valor propio de la matriz A^2 y x un vector asociado a dicho valor propio.

Proposición 1.9. a) Los valores propios de la matriz A^{-1} son los inversos de los valores propios de la matriz A , pero los vectores propios son los mismos que los de A . b) Los valores propios de una matriz idempotente son igual a 0 ó 1.

Demostración. a) Puesto que $Ax = \lambda x$ y premultiplicando por A^{-1} se tiene $x = \lambda A^{-1}x$, por lo que λ^{-1} es un valor propio de A^{-1} y x es el vector propio correspondiente. b) Sea λ un vector propio de A y sea x un vector propio asociado a λ . Entonces, λ^2 es un valor propio de A^2 y x un vector propio asociado a dicho valor propio. Así se tiene $Ax = \lambda x$ y también $A^2x = \lambda^2x$. Ahora bien, como $A^2x = Ax$ (por ser A idempotente), entonces se tiene $\lambda^2x = \lambda x$, por lo que $\lambda(\lambda - 1)x = 0$, es decir, que bien $\lambda = 0$, o $\lambda = 1$.

Proposición 1.10. El rango de una matriz idempotente es igual al número de valores propios iguales a 1, e igual a su traza.

Demostración. El rango de una matriz A es igual al rango de la matriz diagonal Γ formada con sus valores propios. En una matriz idempotente éstos son 0 ó 1, por lo que el rango de A coincide con el número de valores propios iguales a 1. Finalmente, la traza de A es igual a la traza de Γ , es decir, la suma de los valores propios de A , y ésta coincide, en este caso, con el número de valores propios iguales a 1.

En consecuencia, el rango de la matriz Q introducida en la Sección 1.1.a es: $\text{Rango}(Q) = n - 1$.

Proposición 1.11. Dado un valor propio de multiplicidad k , existen k vectores linealmente independientes asociados con dicho valor propio.

Una matriz simétrica A de dimensión $n \times n$ se dice *definida positiva* (*semidefinida positiva*) si para cualquier vector a de dimensión n distinto de 0_n ,

se tiene que $\mathbf{a}'\mathbf{A}\mathbf{a} > 0$ (≥ 0). Una matriz \mathbf{A} simétrica se dice *definida negativa* (*semidefinida negativa*) si $-\mathbf{A}$ es definida positiva (*semidefinida positiva*). Una condición necesaria y suficiente para que una matriz sea definida positiva es que todos sus valores propios sean positivos.

Vamos a probar algunas propiedades que nos serán de utilidad en temas posteriores:

Proposición 1.12. Supongamos que la matriz \mathbf{A} de dimensión $m \times n$ tiene sus columnas linealmente independientes entre sí, lo que exige que $m \geq n$, puesto que las columnas son vectores de dimensión m , y si $m < n$, no podríamos tener n vectores independientes. El producto $\mathbf{A}'\mathbf{A}$ es una matriz definida positiva.

Demostración. Nótese que la matriz $\mathbf{A}'\mathbf{A}$ es $n \times n$. Sea \mathbf{x} un vector columna no nulo de dimensión n ($\mathbf{x} \neq \mathbf{0}_n$) y definamos el vector de dimensión m : $\mathbf{y} = \mathbf{A}\mathbf{x}$. Entonces se tiene $\mathbf{y}'\mathbf{y} = (\mathbf{A}\mathbf{x})'(\mathbf{A}\mathbf{x}) = \mathbf{x}'(\mathbf{A}'\mathbf{A})\mathbf{x}$, que es siempre positivo, puesto que $\mathbf{y}'\mathbf{y}$ es la suma de los cuadrados de los componentes del vector \mathbf{y} . Por tanto, $\mathbf{x}'(\mathbf{A}'\mathbf{A})\mathbf{x} > 0$ para todo vector \mathbf{x} de dimensión n no nulo, y como consecuencia la matriz $\mathbf{A}'\mathbf{A}$ es definida positiva.

El único caso en que el producto $\mathbf{y}'\mathbf{y}$ no sería positivo es que fuese cero, pero recordemos que dicho producto es igual a la suma de cuadrados de las componentes de \mathbf{y} , es decir, una suma de sumandos no negativos. En consecuencia, ello sólo podría ocurrir si el vector \mathbf{y} fuese cero. Ahora bien, como $\mathbf{y} = \mathbf{A}\mathbf{x}$, el vector \mathbf{y} es una combinación lineal de las columnas de la matriz \mathbf{A} , que por hipótesis son linealmente independientes, por lo que ninguna combinación lineal de dichas columnas puede ser igual al vector $\mathbf{0}_m$.

Proposición 1.13. Sea \mathbf{A} una matriz $m \times m$ definida positiva y \mathbf{P} una matriz $m \times n$, de rango m . El producto $\mathbf{P}'\mathbf{A}\mathbf{P}$ es una matriz definida positiva.

Demostración. Dado un vector \mathbf{y} de dimensión n , el producto $\mathbf{P}\mathbf{y}$ es otro vector de dimensión m , por lo que se tiene $(\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y}) > 0$. Pero $(\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y}) = \mathbf{y}'(\mathbf{P}'\mathbf{A}\mathbf{P})\mathbf{y}$, que será por tanto positivo. Como ello ocurre para todo vector \mathbf{y} , se tiene que $\mathbf{P}'\mathbf{A}\mathbf{P}$ es una matriz definida positiva.

Proposición 1.14. Los elementos de la diagonal principal de una matriz definida positiva son estrictamente positivos, mientras que los elementos de la diagonal principal de una matriz semidefinida positiva son no negativos.

Demostración. Para probar cualquiera de estos resultados para el elemento i -ésimo de la diagonal basta utilizar en la definición el vector \mathbf{a}_i que tiene todos sus elementos iguales a cero, excepto el i -ésimo que es igual a 1.

CAPITULO 2

ANALISIS ESTADISTICO

2.1. INTRODUCCION: VARIABLE ALEATORIA, DISTRIBUCIONES DISCRETAS

El objeto de este capítulo consiste en revisar algunos resultados estadísticos que serán de utilidad en el análisis de los métodos y modelos econométricos que vamos a efectuar en este texto. No podemos llevar a cabo un tratamiento exhaustivo de los distintos aspectos tratados, por lo que la lectura de este capítulo no puede suplir al estudio de un buen manual de Estadística. En castellano, los textos de D. Peña (1992) y Arnaiz (1978) son especialmente recomendables.

Interpretamos cada variable económica como teniendo naturaleza aleatoria. Así, cuando decimos que el Producto Interior Bruto (PIB) español en un determinado año ha sido de 68,3 billones de pesetas, entendemos que dicho valor numérico no es sino uno de los (posiblemente muchos) valores que el PIB pudo haber tomado en dicho año. A priori no podemos saber con exactitud el valor futuro del PIB, y esa incertidumbre es, precisamente, el reflejo de su naturaleza aleatoria.

Un *suceso* es cada uno de los resultados que pueden observarse acerca del comportamiento de una variable aleatoria: así que el PIB sea igual a 68,3 billones de pesetas es un suceso, y que sea igual a 71,2 billones es otro suceso. Hay sucesos más complejos: que el PIB exceda de 60 billones es un suceso, de igual modo que el que esté comprendido entre 65 y 70 billones de pesetas es otro suceso.

Una *probabilidad* es una medida positiva, con valor máximo igual a la unidad, y con la propiedad de que la medida de una unión finita o, a lo más, numerable de sucesos disjuntos es igual a la suma de sus medidas individuales. La unión de *todos* los sucesos posibles tiene probabilidad igual a 1. En gran parte, el interés por el trabajo empírico en Economía estriba en asignar probabilidades a los posibles sucesos que pueden acaecer con la variable económica en estudio. La Econometría proporciona herramientas útiles para tal fin.

Si denotamos por X el PIB y por x_0 una determinada cifra, $P(X = x_0)$ denota la probabilidad de que el PIB sea igual a x_0 . Teóricamente puede asociarse a cada valor posible de una variable aleatoria X una determinada *probabilidad*; en la práctica, evaluar tales probabilidades puede resultar muy difícil. El conjunto formado por los valores posibles de una variable, junto con las probabilidades a ellos asociadas, constituyen su *distribución de probabilidad*. En general, aunque no es necesario, ni siempre cierto:

1. Muchos de los hipotéticos valores x_0 de una variable tendrán una probabilidad nula. Más aún, el *soporte* de una variable económica, es decir, el rango máximo de valores que tienen probabilidad no nula, es acotado.
2. La distribución de probabilidad de una variable económica cambia a través del tiempo.
3. Dicha distribución de probabilidad puede venir influida, y quizá determinada, por otras variables económicas, es decir, que existen relaciones de dependencia entre ellas.

Una variable aleatoria se dice *discreta* si distribuye la masa unitaria de probabilidad entre un conjunto discreto de valores (puntos en la recta real). Así, una variable aleatoria discreta puede definirse como un conjunto, quizá infinito, de números reales x_i , cada uno de los cuales tiene asignada una determinada probabilidad p_i , de modo que $0 < p_i \leq 1$ y $\sum p_i = 1$.

Toda variable aleatoria discreta tiene asociada una *función de distribución*:

$$F: \mathbb{R} \longrightarrow [0, 1]$$

definida por:

$$F(x_0) = P(X \leq x_0)$$

de modo que el valor de la función de distribución en un punto x es la medida de probabilidad del intervalo $(-\infty, x_0]$.

De esta definición se deducen las siguientes propiedades de toda función de distribución:

1. $F(-\infty) = 0$, $F(+\infty) = 1$.
2. F es no decreciente.
3. F es continua por la derecha.
4. El conjunto de puntos de discontinuidad de F es a lo más numerable.

Toda función real que satisface las cuatro propiedades anteriores es una función de distribución de una determinada variable aleatoria.

Las funciones de distribución más sencillas son del tipo escalón, como en la Figura 2.1.a, que corresponde a una variable aleatoria cuya masa total de probabilidad está colocada en el punto x_0 . En efecto, de acuerdo con esta función de distribución, sea x un punto a la izquierda de x_0 ; la probabilidad de que la variable aleatoria tome un valor inferior o igual a x , que viene medida por $F(x)$, es igual a cero.

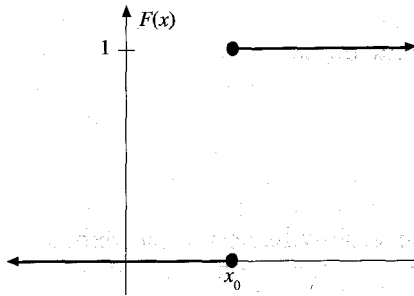


FIGURA 2.1.a.

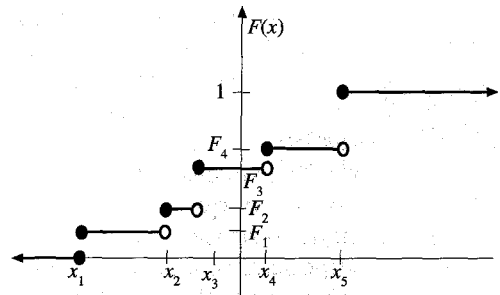


FIGURA 2.1.b.

Consideremos ahora cualquier punto x a la derecha de x_0 . La probabilidad de que la variable aleatoria X tome un valor inferior o igual a x es 1. La misma afirmación puede hacerse acerca de puntos x tan próximos a x_0 por su derecha como se quiera. Pero, lo que es más importante, como $F(x_0) = 1$, se tiene que la probabilidad de que la variable aleatoria X tome un valor superior o igual a x_0 es 1, mientras que es cero si nos movemos tan sólo ligeramente a la izquierda del punto x_0 . En definitiva, la variable aleatoria X toma el valor x_0 con probabilidad 1, y tiene probabilidad igual a 0 de tomar cualquier otro valor distinto de x_0 . Toda *constante* puede interpretarse como una variable aleatoria de este tipo.

Un grado superior de complejidad lo constituyen las funciones de distribución que constan de varios escalones, como en la Figura 2.1.b. Por ser función de distribución, dicha función satisface las condiciones 1 a 4 de la definición anterior. En particular, su rango es acotado superior e inferiormente, ya que $F(-\infty) = 0$ y $F(+\infty) = 1$. En consecuencia, el conjunto de puntos en los que registra un salto (o escalón) es, a lo más, numerable. Dicho conjunto no tiene por qué ser finito: por ejemplo, considérese la variable aleatoria que en cada entero $n = 1, 2, 3, \dots$ tiene una masa de probabilidad igual a $\frac{1}{2^n}$.

Veamos qué tipo de variable aleatoria es aquella que tiene por función de distribución la representada en la Figura 2.1.b. Si tomamos un $\varepsilon_1 > 0$ tal que $x_1 + \varepsilon_1 < x_2$ y consideramos:

$$F(x_1 + \varepsilon_1) - F(x_1 - \varepsilon_1) = P[X \leq x_1 + \varepsilon_1] - P[X \leq x_1 - \varepsilon_1] = \text{haciendo tender } \varepsilon_1 \text{ a cero} = P(X = x_1) = F_1$$

Por tanto, el salto en la función de distribución en el punto x_1 es igual a la probabilidad en dicho punto. Sea ahora ε_2 un número positivo tal que $x_1 + \varepsilon_2 < x_2 < x_3 - \varepsilon_2$. Entonces se tiene:

$$F(x_2 + \varepsilon_2) - F(x_2 - \varepsilon_2) = P[X \leq x_2 + \varepsilon_2] - P[X \leq x_2 - \varepsilon_2] = F_2 - F_1$$

y haciendo tender ε_2 hacia 0 se tiene que $P(X = x_2) = F_2 - F_1$. Repitiendo el razonamiento se tiene que la variable aleatoria representada en la Figu-

ra 2.1.b asigna probabilidad no nula a los puntos de la recta real en que su función de distribución experimenta un salto. Además, la probabilidad que asigna a cada uno de dichos puntos coincide con la magnitud del salto.

Hasta ahora hemos llevado a cabo el ejercicio que consiste en deducir la distribución de probabilidad de una variable aleatoria a partir de su función de distribución. Sin embargo, también puede llevarse a cabo el ejercicio contrario, que consistirá en deducir la función de distribución de una determinada variable aleatoria, a partir de su distribución de probabilidad. En efecto, supongamos que la variable aleatoria X puede tomar como posibles valores los números reales: $x_1 < x_2 < x_3 < \dots < x_n$, con probabilidades $p_1, p_2, p_3, \dots, p_n$, de modo que $p_1 + p_2 + \dots + p_n = 1$.

Consideremos un punto $x < x_1$. En dicho punto se tiene $F(x) = P[X \leq x] = 0$, por la definición de esta variable aleatoria. Si desplazándonos ligeramente hacia la derecha llegamos al punto x_1 , se tiene $F(x_1) = P[X \leq x_1] = P[X = x_1] = p_1$, de nuevo, por la definición de la variable aleatoria. Siguiendo hacia la derecha, consideremos cualquier punto x de la recta real comprendido entre x_1 y x_2 . En dicho punto se tiene $F(x) = P[X \leq x] = P[X = x_1] = p_1$. Lo mismo ocurrirá en todos los puntos que recorremos al desplazarnos hacia la derecha hasta que lleguemos al punto x_2 . En dicho punto: $F(x_2) = P[X \leq x_2] = P[X = x_1] + P[X = x_2] = p_1 + p_2$. Siguiendo con este tipo de razonamiento, el lector puede ver fácilmente que se genera la función de distribución que aparece en la Figura 2.1.b, con $p_i = F_{i+1} - F_i$.

La notación $dF(x)$ mide la variación en la función de distribución $F(x)$ en un entorno infinitesimal a la derecha del punto x , es decir: $dF(x) = F(x + \varepsilon) - F(x)$. Esta notación, relacionada con la medida de Riemann-Stieltjes, permite utilizar cálculo integral con variables aleatorias discretas. Así, con tales variables:

$$\int_a^b dF(x) = F(b) - F(a) = \sum_{a < x_i \leq b} P(x = x_i) = \sum_{a < x_i \leq b} p_i$$

que mide la probabilidad en un intervalo finito de la recta real, mientras que:

$$\int_{-\infty}^b dF(x) = F(b) - F(-\infty) = F(b) = \sum_{x_i \leq b} P(X = x_i);$$

$$\int_a^{\infty} dF(x) = F(\infty) - F(a) = 1 - F(a) = 1 - \sum_{x_i \leq a} P(X = x_i) = 1 - \sum_{x_i \leq a} p_i = \sum_{x_j > a} p_j$$

Ejemplos de distribuciones de probabilidad de tipo discreto

1. Bernoulli: La variable aleatoria puede tomar valores 0 ó 1, con probabilidades p y $1 - p$.
2. Binomial: La variable puede tomar valores 0, 1, 2, ..., n , con probabilidades $p_0, p_1, p_2, p_3, \dots, p_n$, dadas por $p_i = C(n, i)\pi^i(1 - \pi)^{n-i}$, para un

cierto π dado entre 0 y 1 $\left(C(n, i) = \frac{n!}{i!(n-i)!} \right)$. Recordemos que, por la fórmula del binomio: $(a + b)^n = \sum_0^n C(n, i) a^i b^{n-i}$, de modo que $\sum_1^n p_i = 1$.

3. Distribución geométrica: La variable puede tomar los valores $x_0, x_1, x_2, \dots, x_n, \dots$ con probabilidades $p_i = (1 - \pi)$, $0 < \pi < 1$. El lector puede comprobar que, de nuevo, $\sum_i p_i = 1$.

4. Poisson: La variable aleatoria puede tomar valores $x_0, x_1, x_2, \dots, x_n, \dots$ con probabilidades $p_i = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$.

DISTRIBUCIONES CONTINUAS. FUNCION DE DENSIDAD

En otros casos, la masa total de probabilidad está repartida de modo continuo sobre un intervalo de la recta real (quizá sobre toda ella). En tal caso, $F(x)$ es derivable en el interior de tal intervalo, excepto, a lo más, en un conjunto numerable de puntos aislados, y se tiene:

$$dF(x) = f(x)dx, \quad \text{donde } f(x) \text{ denota la derivada } F'(x)$$

La función $f(x)$ es la *función de densidad* de $F(x)$ y está definida únicamente en los puntos en que F es derivable. Toda función de densidad satisface, en los puntos en que está definida: $f(x) \geq 0$. No es preciso definir $dF(x)$ en los puntos en que F es continua, pero no derivable, y puede suponerse, a todos los efectos, que $dF(x)$ es igual a cero en tales puntos.

Con estos criterios, el valor de la función de distribución en un punto x_0 de la recta real viene dado por:

$$F(x_0) = \int_{-\infty}^{x_0} dF(x) = \int_{-\infty}^{x_0} f(x)dx$$

teniéndose además que:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Un caso más general es el de las *mezclas* o *mixturas* de distribuciones discretas y continuas, representadas por una función de densidad con un número finito de puntos de discontinuidad, correspondientes a puntos en que está concentrada una masa positiva de probabilidad. En tales puntos, $dF(x) = p_x$, donde p_x denota la magnitud del salto en la distribución F .

En la Figura 2.2.a aparece una función de densidad que es constante en el intervalo (a, b) . Cuando existe, la función de densidad puede acumularse a lo largo de un intervalo de la recta, para medir la probabilidad contenida en el

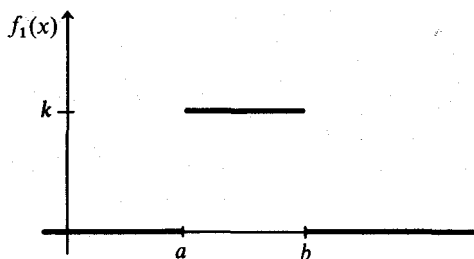


FIGURA 2.2.a.

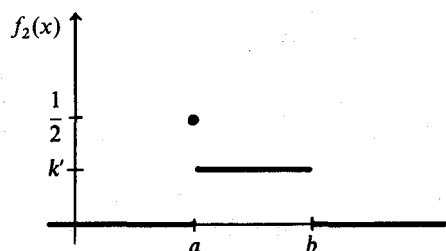


FIGURA 2.2.b.

mismo. Así, en dicha figura, la probabilidad está uniformemente distribuida a lo largo del intervalo (a, b) . De este modo, la probabilidad en un subintervalo es proporcional a la longitud del mismo, por lo que la probabilidad en (a, b) es igual a $k(b - a)$. Puesto que la masa total de probabilidad es igual a 1, en realidad, la función que aparece dibujada en la Figura 2.2.a sólo es una función de densidad si $k = \frac{1}{b - a}$.

En la Figura 2.2.b hay una masa de probabilidad igual a $\frac{1}{2}$ en el punto $x = a$. La probabilidad restante, que es también igual a $\frac{1}{2}$, está uniformemente repartida en el intervalo (a, b) . Por un razonamiento análogo al anterior, la constante k' debe ser igual a $\frac{1}{2(b - a)}$.

Derivemos las funciones de distribución respectivas:

a) En el caso de la Figura 2.2.a se tiene:

$$y < a \quad \Rightarrow F_1(y) = P[X \leq y] = 0$$

$$y = a \quad \Rightarrow F_1(a) = P[X \leq a] = 0$$

$$\begin{aligned} a < y < b &\Rightarrow F_1(y) = \int_{-\infty}^y dF_1(x) = \int_{-\infty}^a dF_1(x) + \int_a^y dF_1(x) = \\ &= 0 + \int_a^y f_1(x) dx = \int_a^y \frac{1}{b - a} dx = \frac{y - a}{b - a} \end{aligned}$$

$$y = b \quad \Rightarrow F_1(b) = 1$$

$$y > b \quad \Rightarrow F_1(y) = 1$$

b) En el caso de la Figura 2.2.b se tiene:

$$y < a \quad \Rightarrow F_2(y) = 0$$

$$y = a \quad \Rightarrow F_2(a) = \frac{1}{2}$$

$$\begin{aligned}
 a < y < b &\Rightarrow F_2(y) = \int_{-\infty}^y dF_2(x) = \int_{-\infty}^a dF_2(x) + \int_a^y dF_2(x) = \\
 &= F_2(a) + \int_a^y dF_2(x) = \frac{1}{2} + \frac{y-a}{2(b-a)},
 \end{aligned}$$

que tiende a $\frac{1}{2}$ según nos acercamos al punto

$y = a$, y tiende a 1 según nos acercamos al punto

$y = b$

$$y \geq b \Rightarrow F_2(y) = 1$$

El lector debe representar gráficamente estas dos funciones de distribución.

Algunas distribuciones continuas

1. La distribución exponencial está definida por la función de densidad $f(x) = \lambda e^{-\lambda x}$ para todo $x \geq 0$.

2. La distribución uniforme en el intervalo $[a, b]$ está definida por la función de densidad $f(x) = \frac{1}{b-a}$.

3. La distribución Normal $N(\mu, \sigma^2)$ está definida por la función de densidad $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ sobre la recta real. En particular, la distribución

$N(0, 1)$ está caracterizada por la función de densidad $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

2.3. MOMENTOS DE UNA DISTRIBUCION

2.3.a. Momentos poblacionales con respecto al origen

Los momentos de orden 1, 2, ... con respecto al origen se definen por:

$$\begin{aligned}
 \mu_r &= E(X^r) = \int_{-\infty}^{\infty} x^r dF(x) = \\
 &= \int_{-\infty}^{\infty} x^r f(x) dx \quad \text{si la distribución es continua, y} \\
 &= \sum_1^n x_i^r p_i \quad \text{si la distribución es de tipo discreto}
 \end{aligned}$$

De éstos, el más importante es el momento de orden 1, que suele denotarse simplemente por μ (sin subíndice) y que recibe el nombre de *esperanza matemática* de la variable aleatoria. Por ser integrales impropias, los momentos de una variable aleatoria pueden no existir. La esperanza de una constante es la misma constante.

Conviene recordar que la esperanza de una variable aleatoria no es, necesariamente, el valor más probable de dicha variable. De hecho, pudiera ocurrir que una variable aleatoria tuviese una probabilidad nula de tomar un valor igual a su esperanza matemática. El valor más probable de una variable

es su *moda*, y puede no ser única. Un número real m que satisface la propiedad $P(X < m) \leq \frac{1}{2}$ y $P(X \leq m) \geq \frac{1}{2}$ es una *mediana* de la variable, que puede presentar algunos problemas de definición en una variable aleatoria discreta.

2.3.b. Momentos poblacionales con respecto a la media

Cuando existe la esperanza matemática de una variable aleatoria, pueden definirse asimismo momentos con respecto a dicha esperanza por medio de:

$$\begin{aligned} m_r &= E[(x - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r dF(x) = \\ &= \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad \text{si la distribución es de tipo continuo, y} \\ &= \sum_1^n (x_i - \mu)^r p_i \quad \text{si la distribución es discreta} \end{aligned}$$

El momento m_1 es igual a cero para toda variable aleatoria, ya que $m_1 = E[(X - \mu)] = EX - \mu = 0$. El momento más importante con respecto a la media es, sin duda, m_2 , la *varianza* de la variable aleatoria, denotada también por $\text{Var}(X)$ o σ^2 . Por otra parte, la *desviación típica* σ se define como la raíz cuadrada, con signo positivo, de la varianza de la variable aleatoria. La varianza de toda constante es igual a cero. Los cocientes $\frac{m_3}{\sigma^3}$ y $\frac{m_4}{\sigma^4}$ se denominan *coeficientes de asimetría* y *curtosis*, respectivamente. Si una distribución es *simétrica*: $f(\mu - x) = f(\mu + x)$ para todo x , su coeficiente de asimetría es cero.

Una variable Bernoulli tiene esperanza p y varianza igual a $p(1 - p)$. Una variable binomial (n, π) tiene esperanza $n\pi$ y varianza $n\pi(1 - \pi)$. Una variable de Poisson tiene esperanza y varianza iguales a λ . Una distribución uniforme tiene esperanza igual a $\frac{a + b}{2}$ y varianza igual a $\frac{(b - a)^2}{12}$. Una distribución exponencial tiene esperanza igual a $\frac{1}{\lambda}$ y varianza igual a $\frac{1}{\lambda^2}$.

Además, los momentos de una variable aleatoria satisfacen las siguientes propiedades:

1. $E(X + a) = \mu + a$.
2. $E(aX) = a\mu$.
3. $E(aX + b) = a\mu + b$.
4. $\text{Var}(X + a) = \text{Var}(X)$.
5. $\text{Var}(aX) = a^2 \text{Var}(X)$.
6. $\text{Var}(X) = E(X^2) - \mu^2$.
7. $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Ejemplo. Si X es una variable $N(\mu, \sigma^2)$, entonces la variable $\frac{X - \mu}{\sigma}$ tiene distribución $N(0, 1)$. Una variable $N(\mu, \sigma^2)$ tiene coeficiente de asimetría cero y coeficiente de curtosis igual a 3.

Igualmente, se tienen las siguientes propiedades:

Proposición 2.1. La esperanza matemática de una variable aleatoria X es la constante para la que es mínima la esperanza $E[(X - a)^2]$.

Demostración. Basta desarrollar la expresión anterior y derivar⁽¹⁾ con respecto al parámetro a .

Esta proposición admite una interpretación interesante: Supongamos que se pretende predecir el valor que tomará la variable aleatoria X , sobre la que no se dispone de otra información que la de su distribución de probabilidad. Evidentemente, X podrá tomar cualquier valor de su soporte; supongamos que el investigador asocia al posible error de predicción que cometa una pérdida cuadrática que es tanto mayor cuanto mayor es dicho error, con independencia de que sea por exceso o por defecto. En tales condiciones, el investigador minimizará el valor esperado de su función de pérdida utilizando como predicción de X su esperanza matemática.

Proposición 2.2. Si la variable aleatoria toma sólo valores positivos y si existe su esperanza $E(X) = \mu$, entonces $E(X) = \int [1 - F(x)] dx$.

Proposición 2.3 (desigualdad de Chebychev). Para cualquier variable aleatoria X , cuya esperanza matemática μ y varianza σ^2 existen, se tiene:

$$P[|X - \mu| \geq \lambda\sigma] \leq \lambda^{-2} \quad \text{para todo } \lambda > 0$$

2.3.c. Momentos muestrales

Al igual que en el caso poblacional pueden definirse momentos con respecto al origen y con respecto a la media en una muestra obtenida a partir de una determinada variable aleatoria. Una vez más, los momentos más importantes son la media muestral \bar{X} y la varianza muestral s^2 , definidos:

$$\bar{X} = \frac{\sum_1^n x_i}{n}$$

$$s^2 = \frac{\sum_1^n (x_i - \bar{X})^2}{n}$$

⁽¹⁾ Supuesto que pueda derivarse bajo el signo integral, es decir, que la derivada de la integral sea igual a la integral de la derivada.

La *desviación típica* es la raíz cuadrada de la varianza, tomada con signo positivo, y puede interpretarse como la distancia promedio entre las observaciones muestrales y la media muestral. Cuando la media muestral es cero, la desviación típica es un indicador del tamaño de la variable aleatoria. La *cuasi-varianza* es igual a la varianza, multiplicada por $\frac{n}{n-1}$. También pueden definirse coeficientes de asimetría y curtosis muestrales.

Los momentos muestrales, por ser función de la muestra recogida, son variables aleatorias y su valor cambia de una muestra a otra. Pues bien, la media muestral (como variable aleatoria que es) tiene una esperanza matemática igual a la de la distribución de la que se obtuvo la muestra $E(\bar{X}) = \mu$, mientras que si las observaciones muestrales son independientes, su varianza es igual a la varianza de la variable aleatoria de la que se obtuvo la muestra, dividida por el tamaño muestral $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. Es decir, dada una población, la media muestral tiene una varianza tanto más pequeña cuanto mayor sea el tamaño de la muestra en la que se calculó.

2.4. DISTRIBUCIONES BIVARIANTES

Aunque hasta aquí hemos examinado tan sólo distribuciones de probabilidad univariantes, es decir, correspondientes a una sola variable aleatoria, existen también distribuciones de probabilidad multivariantes, es decir, concernientes a la distribución de pares (o n -tuplas de variables aleatorias). Una distribución de probabilidad *bivariante* (es decir, correspondiente a dos variables aleatorias) puede venir definida por la tabla de números reales y probabilidades siguiente:

$X_1 \backslash X_2 =$	-2	-1	0	1	2
-1	$\frac{2}{24}$	0	$\frac{2}{24}$	$\frac{4}{24}$	0
0	0	$\frac{1}{24}$	$\frac{2}{24}$	0	$\frac{2}{24}$
2	0	$\frac{3}{24}$	$\frac{2}{24}$	0	$\frac{6}{24}$

Como puede apreciarse, las 15 cifras que aparecen en el cuadro suman exactamente 1, y son las probabilidades asignadas a una determinada variable aleatoria. Dicha variable toma, en realidad, pares de valores (X_1, X_2) : $(-2, -1)$, $(-2, 0)$, $(-2, 2)$, $(-1, -1)$, $(-1, 0)$, $(-1, 2)$, $(0, -1)$, $(0, 0)$, $(0, 2)$, $(1, -1)$, $(1, 0)$, $(1, 2)$, $(2, -1)$, $(2, 0)$, $(2, 2)$; con las probabilidades

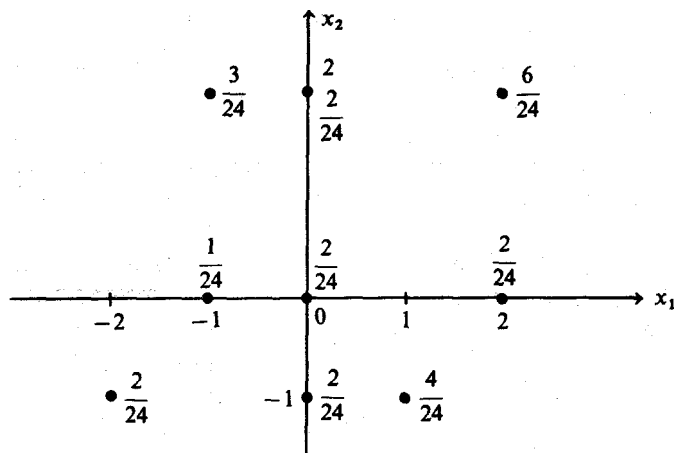


FIGURA 2.3.

reseñadas en la tabla. La información que la tabla proporciona puede, a su vez, organizarse de distintos modos. Por ejemplo, alguien podría cuestionarse acerca de las probabilidades asignadas a los posibles valores de la variable X_1 , con independencia de los valores que tome la variable X_2 . Esta constituye la distribución de probabilidad *marginal* de la variable X_1 , y es:

Valores de X_1	-2	-1	0	1	2
Probabilidades P_1	$\frac{2}{24}$	$\frac{4}{24}$	$\frac{6}{24}$	$\frac{4}{24}$	$\frac{8}{24}$

que implica $E(X_1) = \frac{1}{2}$, $\text{Var}(X_1) = \frac{1}{28}$, mientras que la distribución de probabilidad *marginal* de la variable aleatoria X_2 es:

Valores de X_2	-1	0	2
Probabilidades P_2	$\frac{8}{24}$	$\frac{5}{24}$	$\frac{11}{24}$

con $E(X_2) = \frac{7}{12}$ y $\text{Var}(X_2) = \frac{263}{144}$.

Así, nótese que las probabilidades correspondientes a la distribución *marginal* se obtienen de:

$$P_1[X_1 = -2] = \sum_{x_2} P[X_1 = -2] = P[X_1 = -2/X_2 = -1] + P[X_1 = -2/X_2 = 0] + P[X_1 = -2/X_2 = 2]$$

La distribución de probabilidad de la variable X_1 *condicional* en el valor tomado por la variable X_2 se obtiene dividiendo cada fila de la distribución *bivalente* por la probabilidad *marginal* del valor correspondiente de X_2 :

Valores posibles X_1 :	-2	-1	0	1	2
Probabilidades: Si $X_2 = -1$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$	0
Si $X_2 = 0$	0	$\frac{1}{5}$	$\frac{2}{5}$	0	$\frac{2}{5}$
Si $X_2 = 2$	0	$\frac{3}{11}$	$\frac{2}{11}$	0	$\frac{6}{11}$

Como puede verse, la representación analítica de la distribución de probabilidad condicionada es complicada, pues es, en realidad, una distribución de probabilidad diferente para cada posible valor de la variable que condiciona X_2 . Del mismo modo, la distribución de probabilidad de X_2 *condicionada* en el valor tomado por la variable X_1 vendría dada por las *columnas* de la tabla inicial, divididas por la probabilidad del valor correspondiente de X_1 .

La *esperanza condicional* de X_1 , dada la variable X_2 , es la esperanza matemática de la distribución de probabilidad condicionada. La esperanza condicional es una *variable aleatoria*, pues su valor numérico depende del valor que tome la variable aleatoria X_2 . Así, en nuestro ejemplo se tiene:

$$E(X_1/X_2) = \begin{cases} 0 & \text{con probabilidad } \frac{8}{24} \quad (\text{si } X_2 = -1) \\ \frac{3}{5} & \text{con probabilidad } \frac{5}{24} \quad (\text{si } X_2 = 0) \\ \frac{9}{11} & \text{con probabilidad } \frac{11}{24} \quad (\text{si } X_2 = 2) \end{cases}$$

que es la *distribución de probabilidad* de la *esperanza condicional* de la variable X_1 cuando la condición es el valor tomado por la variable aleatoria X_2 . Las probabilidades $\frac{8}{24}$, $\frac{5}{24}$ y $\frac{11}{24}$ se corresponden con las probabilidades de que X_2 tome los valores -1 , 0 y 2 , respectivamente. Como tal variable aleatoria, tiene (salvo que no exista) esperanza matemática, y además se tiene:

$$E[E(X_1/X_2)] = E(X_1)$$

En este ejemplo:

$$E[E(X_1/X_2)] = 0 \cdot \frac{8}{24} + \frac{3}{5} \cdot \frac{5}{24} + \frac{9}{11} \cdot \frac{11}{24} = \frac{12}{24} = \frac{1}{2}$$

precisamente la esperanza matemática de la variable X_1 , es decir, la esperanza matemática que se obtiene de la distribución de probabilidad marginal de la variable X_1 .

Así pues, la distribución de probabilidad de una variable aleatoria *condicionada* por otra variable aleatoria se obtiene dividiendo las probabilidades de pares de valores de ambas variables por la probabilidad del valor de aquella variable que se utiliza como condicionante:

$$P(X_1 = x_1 / X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$

Dos variables aleatorias X_1 y X_2 son *independientes* si la distribución de probabilidad de cada una de ellas, condicionada por la otra, coincide con su distribución de probabilidad marginal. La intuición es clara pues, si se cumple esta condición, se tiene que el valor observado de una variable, por ejemplo, X_2 , no aporta ninguna información relevante acerca de la distribución de probabilidad de la variable X_1 . Ello se refleja en el hecho de que al obtener la distribución de X_1 condicionada por X_2 se obtiene, precisamente, la distribución marginal de X_1 .

En tal caso se tiene $P(X_1 = x_1 / X_2 = x_2) = P(X_1 = x_1)$ y, en consecuencia, $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$. Si las funciones de distribución son derivables y por tanto existen sus funciones de densidad, entonces se tiene $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. En nuestro ejemplo las variables X_1 y X_2 *no son independientes*, pues, como puede verse, la distribución de probabilidad de la variable X_1 depende del valor que tome la variable X_2 .

Otra forma de ver que ambas variables no son independientes, consiste en tomar un valor de la función de probabilidad conjunta como, por ejemplo, en el punto $(2, 0)$: $P(X_1 = 2, X_2 = 0) = \frac{2}{24}$, y el valor de la distribución de probabilidad marginal de X_2 en el punto 0, $P(X_2 = 0) = \frac{5}{24}$. El cociente entre ambos es $\frac{2}{5}$, que no coincide con $P(X_1 = 2)$, que es $\frac{8}{24}$.

2.5. MOMENTOS EN UNA DISTRIBUCION BIVARIANTE

En el caso de una distribución de probabilidad bivalente de tipo continuo, si denotamos por $f(x, y)$ la función de densidad conjunta de ambas variables debe cumplirse que:

$$\iint f(x, y) dx dy = 1$$

y se tienen las siguientes definiciones:

- Funciones de densidad marginales:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

- Funciones de densidad condicionales:

$$f(x/y) = \frac{f(x, y)}{f_2(y)} \quad \text{siempre que } f_2(y) \neq 0$$

$$f(y/x) = \frac{f(x, y)}{f_1(x)} \quad \text{siempre que } f_1(x) \neq 0$$

- Momentos conjuntos con respecto al origen:

$$\mu_{rs} = \iint_{-\infty}^{\infty} x^r y^s f(x, y) dx dy$$

- Momentos con respecto a las medias:

$$m_{rs} = \iint_{-\infty}^{\infty} (x - \mu_x)^r (y - \mu_y)^s f(x, y) dx dy$$

donde $\mu_x = \mu_{10}$ y $\mu_y = \mu_{01}$ son las esperanzas matemáticas de x e y .

En particular, la *covarianza* de las variables aleatorias X e Y es el momento de orden $(1, 1)$ con respecto a las medias, es decir: $\text{Cov}(x, y) = m_{11}$. El *coeficiente de correlación* entre X e Y es el cociente entre la covarianza y el producto de las desviaciones típicas de X e Y .

Proposición 2.4. Si dos variables aleatorias X y Y son independientes, entonces $\mu_{rs} = E(x^r)E(y^s)$.

Demostración

$$\begin{aligned} \mu_{rs} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f_1(x) f_2(y) dx dy = \\ &= \left(\int_{-\infty}^{\infty} x^r f_1(x) dx \right) \left(\int_{-\infty}^{\infty} y^s f_2(y) dy \right) = E(x^r)E(y^s) \end{aligned}$$

Proposición 2.5. Si dos variables aleatorias X e Y son independientes, entonces su covarianza es igual a cero.

Demostración

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = [E(X - EX)][E(Y - EY)] = 0$$

Proposición 2.6. Los momentos de una distribución bivalente satisfacen las siguientes propiedades:

- $E(aX + bY) = aE(X) + bE(Y)$.
- $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$.
- Si X e Y son independientes, entonces $\text{Var}(X + Y) = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.
- $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$.

Demostración. Se deja como ejercicio para el lector.

Proposición 2.7. Para todo par de variables aleatorias X e Y se tiene:

$$\text{Cov}(X, Y) = E(XY) - \mu_x \mu_y = E[(X - \mu_x)Y] = E[X(Y - \mu_y)]$$

Ejemplo. Dada la distribución de probabilidad conjunta definida por la función de densidad:

$$f(x, y) = kx(x + y) \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

se tiene:

$$a) \quad k = \frac{12}{7} \text{ para que la función anterior sea función de densidad.}$$

$$b) \quad f_1(x) = \left(\frac{12}{7}\right)x\left(x + \frac{1}{2}\right), \quad f_2(y) = \left(\frac{12}{7}\right)\left(\frac{1}{3} + \frac{y}{2}\right)$$

$$\bar{E}(x) = \frac{5}{7}, \quad \text{Var}(x) = \frac{23}{490}, \quad E(y) = \frac{4}{7}, \quad \text{Var}(y) = \frac{23}{294}$$

$$c) \quad f(x/y) = \frac{x(x+y)}{\frac{1}{3} + \frac{y}{2}}, \quad f(y/x) = \frac{x+y}{x + \frac{1}{2}}$$

$$d) \quad E(x/y) = \frac{\frac{1}{4} + \frac{y}{3}}{\frac{1}{3} + \frac{y}{2}}, \quad E(y/x) = \frac{\frac{x}{2} + \frac{1}{3}}{x + \frac{1}{2}}$$

$$e) \quad E[E(x/y)] = \frac{5}{7}, \quad E[E(y/x)] = \frac{4}{7}$$

$$f) \quad E(xy) = \frac{17}{42}$$

$$g) \quad \text{Cov}(x, y) = -\frac{1}{294}$$

La siguiente proposición recoge algunas de las principales propiedades de la esperanza condicional, cuya demostración se pide en uno de los ejercicios al final del capítulo.

Proposición 2.8

- a) $\text{Cov}[X, E(Y/X)] = \text{Cov}(X, Y)$.
- b) $\text{Var}(X) = E_y[\text{Var}(X/Y)] + \text{Var}_y[E(X/Y)]$.
- c) $E[g(X)Y/X] = g(X)E(Y/X)$.
- d) $E[g(X)\{Y - E(Y/X)\}/X] = 0$.

El lector puede comprobar que *a)* y *b)* se satisfacen en los ejemplos anteriores.

2.6. PROPIEDADES DE UN ESTIMADOR

Cuando se extrae una muestra de una población de la que desconocemos alguno de los parámetros que la caracterizan, tiene interés cuestionarse cómo utilizar la información muestral para obtener una idea acerca del valor del parámetro desconocido. La función utilizada para resumir la información muestral en un número, que se asociará al parámetro desconocido, se llama *estimador*, y su valor en una muestra determinada se denomina *estimación*. Un estimador, siendo función de la muestra, es una variable aleatoria y tiene su propia distribución de probabilidad.

Denotemos por θ el parámetro desconocido y por $\hat{\theta}$ el estimador utilizado. El *sesgo* de un estimador $\hat{\theta}$ es la diferencia $E(\hat{\theta}) - \theta$. Dicho estimador se dice *insesgado* si $E(\hat{\theta}) = \theta$, y se dice *eficiente* o *de mínima varianza* si no hay ningún otro estimador insesgado que tenga una varianza menor que $\hat{\theta}$. En general se trata de utilizar estimadores con varianza pequeña, pues de este modo la estimación es más precisa.

Por ejemplo, la media muestral es un estimador insesgado de la esperanza poblacional, ya que $E(\bar{X}) = \mu$. Del mismo modo, la cuasivarianza muestral $(s')^2$ es un estimador insesgado de la varianza poblacional, ya que $E(s')^2 = \sigma^2$.

Si se conoce la distribución de probabilidad de un estimador, pueden obtenerse intervalos de confianza acerca del verdadero valor del parámetro desconocido. Como caso sencillo, supongamos que se trata de estimar la esperanza μ de una variable $N(\mu, 1)$ y que, tras tomar una muestra de 25 observaciones, hemos obtenido $\bar{X} = 10$. En estas condiciones, el estimador $\hat{\theta}$, que en este caso es la media muestral \bar{X} , tiene una distribución $N(\mu, 0,20^2)$, por lo que $\frac{\bar{X} - \mu}{0,20}$ tiene una distribución $N(0, 1)$. Como esta última variable tiene una probabilidad 0,95 de tomar un valor en el intervalo $(-1,96; 1,96)$, se tiene:

$$\begin{aligned} 0,95 &= P[-1,96 \leq 5(\bar{X} - \mu) \leq 1,96] = P[\bar{X} - 0,40 \leq \mu \leq \bar{X} + 0,40] = \\ &= P[9,6 \leq \mu \leq 10,4] \end{aligned}$$

que constituye el *intervalo de confianza* del 95 por 100 del parámetro μ . Puede verse que cuanto menor sea la varianza del estimador (en este caso \bar{X}) menor será la amplitud del intervalo de confianza, por lo que, al nivel de confianza dado, el rango de valores que daremos como posibles para μ será menor. Dicho de otro modo, habremos estimado el parámetro μ con una mayor precisión.

El razonamiento es válido con mayor generalidad: Basta volver a la desigualdad de Chebychev para observar que cuanto menor sea la varianza de un estimador insesgado, menor será el intervalo asociado al parámetro desconocido para una misma cota inferior de probabilidad.

El *error cuadrático medio* de un estimador se define por $ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$ y se tiene:

$$ECM(\hat{\theta}) = E\{[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2\} = E[(\hat{\theta} - E(\hat{\theta}))^2] + E[(E(\hat{\theta}) - \theta)^2]$$

ya que el producto cruzado de ambos términos tiene esperanza cero. Finalmente, se tiene:

$$ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + [\text{Sesgo}(\hat{\theta})]^2$$

por lo que, si un estimador es insesgado, su error cuadrático medio y su varianza coinciden. El ECM es un estadístico interesante, pues conjuga el sesgo de un estimador con su varianza en una única medida. En ocasiones, el investigador puede estar dispuesto a utilizar un estimador sesgado, bien porque no dispone de ningún estimador insesgado, o bien porque existe un estimador que, aun siendo sesgado, tiene una varianza reducida. Para poder decidir si conviene utilizar dicho estimador, se evalúa su ECM y se prefiere sobre un estimador insesgado si su varianza es suficientemente pequeña como para compensar el sesgo.

2.7 CAMBIO DE VARIABLE EN DISTRIBUCIONES DE PROBABILIDAD

Supongamos que tenemos caracterizada la distribución de una variable aleatoria x mediante su función de densidad $f(x)$. Consideremos otra variable aleatoria y de la que sabemos que es función exacta de la variable x , por ejemplo: $y = 5x + 3$. Puesto que una variable es función exacta de la otra, basta conocer la distribución de una de las variables para tener perfectamente determinada la distribución de probabilidad de la otra variable. En principio podría pensarse que para obtener la función de densidad de la variable y basta sustituir x en su función de densidad por su expresión en función de y , $x = \frac{y-3}{5}$. Sin embargo, tal argumento es erróneo. Por ejemplo, supongamos que x es una variable con función de densidad:

$$f(x) = 2x \quad \text{en el intervalo} \quad 0 \leq x \leq 1$$

El razonamiento anterior llevaría a sugerir como función de densidad de y :

$$f(y) = \left(\frac{2}{5}\right)(y-3) \quad \text{en el intervalo} \quad 3 \leq y \leq 8$$

Sin embargo, la integral de $f(y)$ a lo largo de dicho intervalo es igual a 5, por lo que $f(y)$ no puede ser función de densidad de la variable aleatoria y . Bastaría que, al hacer el cambio de variable, hubiésemos multiplicado la supuesta función de densidad de y por el valor de la derivada $\frac{dx}{dy}$, que resulta ser igual a $\frac{1}{5}$, para que la función así obtenida:

$$f(y) = \frac{2}{25}(y-3) \quad \text{para} \quad 3 \leq y \leq 8$$

sea la función de densidad de la variable aleatoria y .

Así, la relación entre ambas funciones de densidad es:

$$f_y(y) = f_x(h^{-1}(y)) \left| \frac{dx}{dy} \right|$$

supuesto que la función $y = h(x)$ es biunívoca y con derivada continua. Este resultado puede extenderse al caso de distribuciones bivariantes o multivariantes, siendo entonces $|dx/dy|$ el jacobiano de la transformación.

2.8. DISTRIBUCIONES DERIVADAS

1. La suma de los cuadrados de n variables aleatorias con distribución $N(0, 1)$, independientes entre sí, es una variable con distribución chi-cuadrado con n grados de libertad. La esperanza de una variable chi-cuadrado es igual a su número de grados de libertad, mientras que su varianza es el doble del número de grados de libertad. La suma de k variables aleatorias independientes con distribución chi-cuadrado, con grados de libertad igual a n_i es una variable con distribución chi-cuadrado con grados de libertad igual a $\sum_i^k n_i$.

Dado un conjunto de n variables $x_i \sim N(\mu, \sigma^2)$ independientes entre sí, entonces $x = \sum_1^n \left(\frac{x_i^2}{\sigma_i^2}\right)$ se distribuye como una *chi-cuadrado no central* con

n grados de libertad y parámetro de no centralidad igual a $\sum_{i=1}^n \frac{\mu_i^2}{\sigma_i^2}$. Compara-

da con una variable chi-cuadrado con igual número de grados de libertad, una chi-cuadrado no central tiene una mayor masa de probabilidad alejada del origen.

2. El cociente

$$Z = \frac{Y}{\sqrt{\frac{X}{k}}}$$

donde Y es una variable con distribución $N(0, 1)$ y X una variable chi-cuadrado con k grados de libertad, siendo Y y X independientes entre sí, es una variable Z con distribución t de Student con k grados de libertad.

Si Y fuese una variable $N(\mu, \sigma^2)$, independiente de X , entonces se tendría que

$$\frac{\frac{Y - \mu}{\sigma}}{\sqrt{\frac{X}{k}}}$$

es una distribución t de Student con k grados de libertad.

3. El cociente de dos variables chi-cuadrado independientes entre sí, divididas por sus grados de libertad m y n , respectivamente, define la variable F de Snedecor, a la que se asocian como grados de libertad el par (m, n) . Una distribución F con grados de libertad $(1, n)$ es equivalente al cuadrado de una distribución t de Student con n grados de libertad.

El cociente de una variable chi-cuadrado con m grados de libertad y parámetro de no centralidad igual a λ , y una variable chi-cuadrado central, independiente de la anterior, con n grados de libertad, es una variable F no central, con (m, n) grados de libertad y parámetro de no centralidad igual a λ . Al igual que ocurría con la distribución chi-cuadrado, una variable F no central tiene una mayor masa de probabilidad alejada del origen, en comparación con una F central con sus mismos grados de libertad.

4. Un vector \mathbf{X} de dimensión k sigue una distribución Normal multivariante si su función de densidad, definida en \mathbf{R}^k , es:

$$f(\mathbf{x}) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})\right]$$

5. Si el vector \mathbf{X} tiene una distribución $N_k(\boldsymbol{\mu}, \Sigma)$ y si \mathbf{T} es una matriz $p \times k$ de rango p , entonces el vector \mathbf{TX} sigue una distribución $N_p(\mathbf{T}\boldsymbol{\mu}, \mathbf{T}\Sigma\mathbf{T}')$.

Debe notarse que la dimensión de la distribución Normal ha cambiado de k a p . En particular, este resultado también garantiza que la distribución de cualquier componente de un vector que sigue una distribución Normal multivariante sigue una distribución del mismo tipo, pues un componente puede generarse a partir de \mathbf{X} con una matriz \mathbf{T} formada por una sola fila: $(0, 0, 0, \dots, 1, \dots, 0, 0)$.

6. Si el vector \mathbf{x} sigue una distribución Normal multivariante $N_k(\mathbf{0}, \mathbf{I}_k)$, y si \mathbf{A} es una matriz simétrica e idempotente, entonces:

- a) $\mathbf{x}'\mathbf{A}\mathbf{x}$ sigue una distribución chi-cuadrado con un número de grados de libertad igual al rango de la matriz \mathbf{A} . En este resultado es crucial que la matriz de covarianzas del vector \mathbf{x} sea la identidad. En realidad, lo que es importante es que, siendo una matriz escalar, las coordenadas del vector \mathbf{x} sean variables aleatorias independientes entre sí. Si su varianza, siendo igual, fuese un parámetro σ^2 , bastaría dividir el vector \mathbf{x} por σ para tener que $\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\sigma^2}$ sigue una distribución chi-cuadrado.
- b) Si \mathbf{B} es otra matriz simétrica e idempotente, las distribuciones chi-cuadrado $\mathbf{x}'\mathbf{A}\mathbf{x}$ y $\mathbf{x}'\mathbf{B}\mathbf{x}$ son independientes entre sí si y sólo si $\mathbf{A}\mathbf{B} = \mathbf{0}$. Dada una matriz \mathbf{M} de dimensión $p \times k$ y de rango p , el vector aleatorio $\mathbf{M}\mathbf{x}$, que tiene distribución Normal $(\mathbf{M}\boldsymbol{\mu}, \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}')$, y la variable chi-cuadrado $\mathbf{x}'\mathbf{A}\mathbf{x}$ son independientes si y sólo si $\mathbf{M}\mathbf{A} = \mathbf{0}_p$.
7. Sea $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ un vector con distribución Normal multivariante, $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

siendo \mathbf{y}_1 e \mathbf{y}_2 , al igual que $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$, vectores de dimensión k_1 y k_2 , y $\boldsymbol{\Sigma}_{11}$ y $\boldsymbol{\Sigma}_{22}$ matrices simétricas, de dimensión $k_1 \times k_1$ y $k_2 \times k_2$, con $k_1 + k_2 = k$, y $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$. Entonces la distribución marginal del subvector \mathbf{y}_1 es $N_{k_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, mientras que la distribución del subvector \mathbf{y}_1 condicionada al subvector \mathbf{y}_2 es:

$$N_{k_1}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

2.9. EL ESTIMADOR DE MAXIMA VEROSIMILITUD

Uno de los procedimientos de estimación más populares es el de máxima verosimilitud, cuya justificación descansa sobre algunos resultados que examinaremos en la Sección 2.11. A diferencia de otros métodos de estimación, el de máxima verosimilitud se basa en un determinado supuesto acerca del tipo de distribución de donde se obtuvo la muestra. La *función de verosimilitud* es la función de probabilidad conjunta de la muestra; cuando ésta es aleatoria simple, las distintas observaciones muestrales son independientes entre sí y la función de verosimilitud es el producto de los valores de la función de densidad para cada una de las observaciones.

Hay que tener presente que el valor de θ que maximiza la función $L(\theta)$ coincide con el que maximiza una función estrictamente monótona (creciente o decreciente) de $L(\theta)$, si bien sus valores máximos serán diferentes. Utilizaremos exhaustivamente este resultado en este texto, aplicándolo a la función de verosimilitud, que denotaremos precisamente por $L(\theta)$, y a su logaritmo neperiano que es, generalmente, más sencillo de maximizar.

En el caso de una muestra de tamaño n extraída de una población $N(\mu, \sigma^2)$ se tiene:

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_1^n (x_i - \mu)^2$$

cuyas derivadas con respecto a los parámetros desconocidos μ y σ^2 , igualadas a cero, conducen a los estimadores de máxima verosimilitud:

$$\hat{\mu} = \bar{X} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_1^n (x_i - \bar{X})^2 = s^2$$

Si denotamos por θ el vector $\theta = (\mu, \sigma^2)$, la matriz de derivadas segundas es:

$$\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} = \begin{pmatrix} -\frac{n}{\sigma^2} & -n \left(\frac{\bar{X} - \mu}{\sigma^4} \right) \\ -n \left(\frac{\bar{X} - \mu}{\sigma^4} \right) & \frac{n}{2\sigma^4} - \sum_1^n \frac{(X_i - \mu)^2}{\sigma^6} \end{pmatrix}$$

que evaluada en los estimadores de máxima verosimilitud se reduce a:

$$\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

que es definida negativa, por lo que la solución de las condiciones necesarias de optimización antes obtenida era realmente la solución a dicho problema.

2.10. TEORIA ASINTOTICA

Comenzamos dando varios criterios de convergencia de una sucesión de variables aleatorias, así como algunas de las propiedades de dichos criterios.

2.10.a. Convergencia en probabilidad

Definición 2.1. Se dice que la sucesión de variables aleatorias X_1, X_2, \dots, X_t , converge en probabilidad a la variable aleatoria X si:

$$\forall \varepsilon > 0, \lim_t P[|X_t - X| < \varepsilon] = 1$$

o, equivalentemente, si:

$$\forall \varepsilon, \eta > 0 \text{ existe un } t_0 \text{ tal que si } t > t_0, \text{ entonces } P[|X_t - X| < \varepsilon] > 1 - \eta$$

La convergencia en probabilidad de la sucesión anterior se indica mediante la notación $\text{plim } X_t = X$. Una sucesión de variables aleatorias interesantes es la formada por los estimadores $\hat{\theta}_t$ de un parámetro desconocido θ que van obteniéndose según aumenta el tamaño muestral. Así, $\hat{\theta}_t$ denota el estimador de θ obtenido con las t primeras observaciones muestrales, $\hat{\theta}_{t+1}$ denota el estimador obtenido con las primeras $t + 1$ observaciones, y así sucesivamente. Si la sucesión $\hat{\theta}_t$ converge en probabilidad al verdadero valor (desconocido) del parámetro θ , entonces se dice que el estimador $\hat{\theta}$ es *consistente*. Es decir:

Definición 2.2. El estimador $\hat{\theta}$ del parámetro θ es *consistente* si al obtener la sucesión $\{\hat{\theta}_t\}_{t=1}^{\infty}$ del modo descrito se tiene:

$$\forall \varepsilon > 0, \lim_t P[|\hat{\theta}_t - \theta| < \varepsilon] = 1$$

Cuando un estimador no es consistente, a la diferencia $\text{plim } E(\hat{\theta}_t) - \theta$ se le denomina *sesgo asintótico*.

Es preciso hacer notar que:

1. La definición de convergencia en probabilidad se generaliza inmediatamente al caso de una sucesión de vectores aleatorios de dimensión k fija, sin más que sustituir el valor absoluto en la definición anterior por la norma euclídea del vector diferencia $\hat{\theta}_t - \theta$. Este es, como veremos en este texto, el caso habitual en Econometría.

2. La variable aleatoria límite X en la definición pudiera tener una distribución de probabilidad degenerada en un punto, es decir, ser una constante. De hecho, en aplicaciones econométricas le preocupa al investigador saber si el estimador que utiliza es consistente, lo que equivale a averiguar si su límite en probabilidad coincide con el vector *constante* de parámetros desconocidos.

Una condición suficiente para la convergencia en probabilidad de una sucesión de variables aleatorias a una constante viene dada por la:

Proposición 2.9. Si $E(X_t) = c$ para todo t y si $\lim_t \text{Var}(X_t) = \lim_t \sigma_t^2 = 0$, entonces $\text{plim } X_t = c$.

Demostración. Es una aplicación del teorema de Chebychev, por el que se tiene:

$$P[|X_t - c| \geq k\sigma_t] \leq \frac{1}{k^2}$$

Si elegimos la constante $k = \frac{\varepsilon}{\sigma_t}$, entonces:

$$P[|X_t - c| \geq \varepsilon] \leq \frac{\sigma_t^2}{\varepsilon^2}$$

que, por hipótesis, converge a cero cuando aumenta el tamaño muestral. En consecuencia, X_t converge en probabilidad a c .

La interpretación de este resultado en términos de un problema de estimación es que si se dispone de un estimador $\hat{\theta}_t$ que es insesgado para cualquier tamaño muestral y si su varianza tiende a cero al aumentar el número de observaciones, entonces dicho estimador es consistente. Un ejemplo lo constituye la media muestral \bar{X} , que tiene las propiedades:

$$E(\bar{X}) = \mu \quad \text{y} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

con independencia del tipo de población de la que se obtuvo la muestra, siempre que tenga esperanza y varianza finitas. En virtud de la Proposición 2.9, la media muestral es un estimador consistente de la esperanza poblacional de cualquier distribución de probabilidad.

Este resultado también es cierto cuando el estimador no es insesgado en muestras finitas, siempre que, tanto su sesgo como su varianza, tiendan a cero cuando crece el tamaño muestral:

Proposición 2.10. Si $\lim_t E(X_t) = c$ y si $\lim_t \text{Var}(X_t) = 0$, entonces: $\text{plim } X_t = c$.

Demostración. Basta descomponer $X_t - c = (X_t - E(X_t)) + (E(X_t) - c)$ y probar que ambos elementos convergen en probabilidad a cero.

Por ejemplo, la cuasivarianza muestral es un estimador insesgado de la varianza poblacional, pero la varianza muestral s^2 es un estimador sesgado de σ^2 . Sin embargo, $Es^2 = \frac{T-1}{T} \sigma^2$, por lo que el sesgo tiende a 0 al aumentar el tamaño muestral T . Pero, además, si la población es $N(0, \sigma^2)$, entonces $\frac{(T-1)s^2}{\sigma^2}$ sigue una distribución chi-cuadrado con $T-1$ grados de libertad.

Por tanto:

$$\text{Var}\left(\frac{(T-1)s^2}{\sigma^2}\right) = 2(T-1)$$

lo que implica:

$$\text{Var}(s^2) = \frac{2\sigma^4}{T-1}$$

que tiende a cero al aumentar el tamaño muestral. Por tanto, la varianza muestral es un estimador sesgado, aunque consistente, de la varianza poblacional en una población Normal $(0, \sigma^2)$.

En realidad, ni siquiera es necesario que la sucesión de variables aleatorias tenga varianza finita:

Proposición 2.11 (Khinchine). Si las variables aleatorias $X_1, X_2, \dots, X_t, \dots$ son independientes e idénticamente distribuidas con esperanza finita μ , y si \bar{X}_t denota a $\frac{1}{t} \sum_i X_i$, entonces la media muestral \bar{X}_t converge en probabilidad a la esperanza poblacional:

$$plim \bar{X}_t = \mu$$

Ejemplo. Consideremos la sucesión de variables aleatorias:

$$X_t = \begin{cases} Z & \text{con probabilidad } \frac{t-1}{t}, \text{ donde } Z \text{ se distribuye } N(0, 1) \\ t & \text{con probabilidad } \frac{1}{t} \end{cases}$$

cuyo límite en probabilidad es la variable $N(0, 1)$.

Ejemplo. Consideremos la sucesión de variables aleatorias:

$$X_t = \begin{cases} 0 & \text{con probabilidad } \frac{t-1}{t} \\ t^2 & \text{con probabilidad } \frac{1}{t} \end{cases}$$

En este caso, $plim X_t = 0$, a pesar de que el valor no nulo de la variable X_t crece rápidamente.

2.10.b. Convergencia en distribución

Definición 2.3. Sea $X_1, X_2, \dots, X_t, \dots$ una sucesión de variables aleatorias con funciones de distribución $F_1, F_2, \dots, F_t, \dots$. Supongamos $F_t(x) \longrightarrow F(x)$ en todos los puntos de continuidad de la función $F(x)$ y que, en ellos, $F(x)$ es una función de distribución. Entonces se dice que la sucesión de variables aleatorias X_t converge en distribución a X , donde X es una variable aleatoria con función de distribución F , y se representa por $X_t \xrightarrow{D} X$.

La definición de convergencia en distribución se extiende, sin ninguna dificultad, al caso de vectores aleatorios.

Un resultado que nos será de gran utilidad en este texto es el que concierne a la convergencia de una función continua de una sucesión de vectores aleatorios que converge en probabilidad o en distribución a un vector X .

Proposición 2.12. Si X_t es una sucesión de vectores aleatorios que converge en probabilidad (o en distribución) al vector aleatorio X y si $g(\cdot)$ es una función continua, entonces la sucesión de vectores aleatorios $g(X_t)$ converge en probabilidad (o en distribución) al vector aleatorio $g(X)$.

Es crucial tener en cuenta que la función g no puede depender del tiempo t . Por ejemplo, $g(X_t) = tX_t^2$ es una función continua de X_t , pero dependiente del tiempo.

El resultado anterior se satisface también para una función g cuyo conjunto de puntos de discontinuidad tenga medida cero bajo la distribución límite $F(\cdot)$. En el caso de distribuciones multivariantes se tiene:

Proposición 2.13. Sea X_t una sucesión de vectores aleatorios de dimensión k que convergen en probabilidad al vector aleatorio X y sea Y_t una sucesión de vectores de dimensión m que convergen asimismo en probabilidad al vector constante c . Si además $g(\cdot)$ es una función continua definida sobre R^{k+m} , entonces la sucesión $g(X_t, Y_t)$ converge en probabilidad a $g(X, c)$.

En este resultado hay que tener en cuenta que los vectores a los que se hace referencia pueden también ser matrices. Las aplicaciones más usuales de este importante resultado son:

Si $\text{plim } X_t = X$, variable aleatoria, y si $\text{plim } Y_t = c$ constante, entonces: $\text{plim}(X_t + Y_t) = X + c$, $\text{plim}(X_t Y_t) = cX$, y, en particular, si $\text{plim } Y_t = 0$, entonces $\text{plim } X_t Y_t = 0$. También se tiene $\text{plim } \frac{X_t}{Y_t} = \frac{X}{c}$, siempre que c sea distinta de cero.

A continuación damos algunos resultados que relacionan las convergencias en probabilidad y en distribución.

Proposición 2.14. a) Si $\text{plim } X_t = X$, entonces la sucesión de variables aleatorias X_t converge en distribución a X .

b) Si la sucesión de variables aleatorias X_t converge en distribución a la constante c , entonces $\text{plim}(X_t) = c$.

Proposición 2.15. Dadas dos sucesiones de variables aleatorias X_t e Y_t , si se tiene que $\text{plim}(X_t - Y_t) = 0$ y también que Y_t converge a Y en distribución, entonces se tiene que $X_t \xrightarrow{D} Y$. En particular, si la diferencia de dos variables aleatorias $X_t - Y_t$ converge en probabilidad a cero y una de ellas (por ejemplo, Y_t) converge en probabilidad a una constante, entonces X_t converge en probabilidad (y en distribución) a dicha constante. Este resultado puede generalizarse al caso vectorial.

Proposición 2.16. Sea Y_t una matriz aleatoria de dimensión fija $p \times q$ y sea X_t un vector aleatorio de dimensión $q \times 1$. Entonces, si

$$\begin{cases} \mathbf{Y}_t \xrightarrow{P} \mathbf{A} \text{ (matriz de constantes) y} \\ \mathbf{X}_t \xrightarrow{D} \mathbf{X} \end{cases}$$

se tiene que $\mathbf{Y}_t \mathbf{X}_t \xrightarrow{D} \mathbf{A} \mathbf{X}$.

El siguiente resultado es una importante caracterización alternativa de la convergencia en distribución, pues es el modo más frecuente de probar que una sucesión de variables aleatorias converge en distribución a una determinada variable aleatoria.

Dada una variable aleatoria X , su función característica se define como $\Psi_x(z) = E(e^{izx})$, siendo i la unidad imaginaria.

Proposición 2.17. Sea $X_1, X_2, \dots, X_t, \dots$ una sucesión de variables aleatorias con funciones características $\Psi_1(z), \Psi_2(z), \dots, \Psi_t(z), \dots$. La sucesión de variables aleatorias X_t converge en distribución a la variable aleatoria X si y sólo si la sucesión de funciones características $\Psi_t(z)$ converge a la función característica $\Psi(z)$ de X en todo punto $z \in \mathbb{R}$, y si $\Psi(z)$ es continua en $z = 0$.

2.11. TEOREMA CENTRAL DEL LIMITE

Los resultados contenidos en esta sección aseguran que una determinada función lineal de un estimador converge en distribución a una variable Normal. Ello justifica la utilización de distribuciones Normales en casos en que el investigador desconoce, o no está dispuesto a suponer, una determinada distribución de probabilidad para las variables aleatorias con que trabaja.

Teorema Central del Límite. Sea $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t, \dots$ una sucesión de vectores aleatorios de dimensión k , independientes entre sí e idénticamente distribuidos, con $E(\mathbf{X}_i) = \boldsymbol{\mu}$ y $\text{Var}(\mathbf{X}_i) = \boldsymbol{\Sigma}$. Entonces se tiene:

$$\sqrt{T} \left[\frac{1}{T} \sum_1^T \mathbf{X}_t - \boldsymbol{\mu} \right] \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{\Sigma})$$

Por ejemplo, una muestra aleatoria simple de una variable con esperanza matemática μ y varianza igual a σ^2 constituye una sucesión como la del teorema (sólo que ahora son variables aleatorias y no vectores). Sin necesidad de suponer que la variable de la que se obtiene la muestra tiene una determinada distribución de probabilidad, el teorema garantiza que $\sqrt{T}[\bar{X} - \mu]$ converge en distribución a una variable $N(0, \sigma^2)$, lo que permite aproximar la distribución de la media aritmética \bar{X} por una $N\left(\mu, \frac{\sigma^2}{T}\right)$, la misma que se hubiese obtenido si hubiésemos supuesto la Normalidad de la variable de la que se obtuvo la muestra.

Por tanto, pueden calcularse intervalos de confianza para μ como si la variable fuese Normal. Estos intervalos serán tanto más correctos cuanto

mayor sea el tamaño muestral, siempre que las observaciones muestrales sean independientes entre sí.

Nótese, por último, que como $\text{Var}(\bar{X}) = \frac{\sigma^2}{T}$, entonces, según crece el tamaño muestral, la distribución límite de la media muestral tiende a ser degenerada en el punto μ . El teorema anterior asegura que si se multiplica \bar{X} por el factor \sqrt{T} , la distribución límite está bien definida y, concretamente, la aproximación de la media muestral por una distribución $N\left(\mu, \frac{\sigma^2}{T}\right)$ es válida.

Teorema de Mann-Wald. Sea \mathbf{X} una matriz $T \times k$ y \mathbf{u} un vector de dimensión T tales que:

- i) $E(\mathbf{u}) = \mathbf{0}$, $E(\mathbf{u}\mathbf{u}') = \sigma_u^2 \mathbf{I}_T$.
- ii) $E(\mathbf{X}_i' \mathbf{u}) = 0$, $i = 1, 2, \dots, k$, donde \mathbf{X}_i es la columna i -ésima de la matriz \mathbf{X} .
- iii) $\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right) = \Sigma_{xx} < \infty$, donde Σ_{xx} es una matriz simétrica, definida positiva.

Entonces se tiene que:

- a) $\text{plim}\left(\frac{\mathbf{X}'\mathbf{u}}{T}\right) = \mathbf{0}_k$.
- b) $\frac{\mathbf{X}'\mathbf{u}}{\sqrt{T}} \xrightarrow{D} N(\mathbf{0}, \sigma_u^2 \Sigma_{xx})$.

Por último, a veces no se está interesado en la distribución de un estimador, sino en la de una función del mismo. En tales casos, el resultado a utilizar es el siguiente:

Proposición 2.18. Sea \mathbf{X}_T un vector en R^k y supongamos que

$$\sqrt{T}(\mathbf{X}_T - \mathbf{a}) \xrightarrow{D} N(\mathbf{0}, \Sigma)$$

Si f es una función diferenciable de R^k en R^m , se tiene:

$$\sqrt{T}(f(\mathbf{X}_T) - f(\mathbf{a})) \xrightarrow{D} N_m(\mathbf{0}, \nabla f(\mathbf{a}) \Sigma \nabla f(\mathbf{a})')$$

donde $\nabla f(\cdot)$ denota la matriz gradiente de la función vectorial $f(\cdot)$.

Volviendo al ejemplo anterior, puesto que $\sqrt{T}(\bar{X} - \mu)$ converge en distribución a una $N(0, \sigma^2)$, se tiene que $\sqrt{T}(\bar{X}^2 - \mu^2)$ converge en distribución

a una variable $N(0, 4\mu^2\sigma^2)$, lo que permite aproximar el cuadrado de la media muestral por medio de una variable $N\left(\mu^2, \frac{4\mu^2\sigma^2}{T}\right)$.

La distribución asintótica del estimador de máxima verosimilitud

Previamente hemos definido un estimador eficiente como aquel de mínima varianza, y hemos hecho referencia al interés de tal propiedad. La cota de Cramer-Rao proporciona la menor varianza que puede tener un estimador insesgado⁽²⁾ $\hat{\theta}$ de un vector de parámetros θ . Dicha cota viene dada por la inversa de la matriz:

$$-E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'}\right] = E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)'\right]$$

llamada *matriz de información*, denotada por $I(\theta)$.

Bajo ciertas condiciones de regularidad (véase Rao, 1973), el estimador de máxima verosimilitud es consistente, tiene distribución asintótica Normal, y la matriz de covarianzas de dicha distribución límite es la menor posible, es decir, el estimador es asintóticamente eficiente. El resultado puede también expresarse como:

$$\sqrt{T}(\hat{\theta}_{MV} - \theta) \xrightarrow{D} N\left(\mathbf{0}_k, \lim\left[\frac{I(\theta)}{T}\right]^{-1}\right)$$

donde $I(\theta)$ es la matriz de información. Para obtener valores numéricos, se evalúa $I(\theta)$ en $\theta = \hat{\theta}_{MV}$.

2.12. CONTRASTES DE HIPOTESIS

Una vez que se dispone de una estimación de un parámetro desconocido, el investigador está generalmente interesado en contrastar una determinada hipótesis nula: $H_0: \theta \in \Theta$ acerca del verdadero valor desconocido del parámetro de interés, frente a una alternativa: $H_1: \theta \in \Omega$. Existen distintas estrategias para llevar a cabo tales contrastes, que vamos a repasar brevemente en esta sección. Cada una de estas estrategias, aplicada a un contraste particular, le confiere

⁽²⁾ La cota de Cramer-Rao se derivó para estimadores cualesquiera, no necesariamente insesgados, en cuyo caso su formulación analítica es ligeramente más compleja. Su utilización en este texto se referirá tan sólo a estos últimos.

En la práctica, la matriz de información se estima por:

$$\sum_{i=1}^T \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta}\right) \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta}\right)'$$

unas determinadas características: el *error del tipo I* de un contraste se produce cuando se rechaza la hipótesis nula siendo cierta, mientras que el *error de tipo II* se produce cuando se mantiene la hipótesis nula, a pesar de ser falsa. El *tamaño* o *nivel de significación* α de un contraste es la probabilidad de cometer un error de tipo I, y su complemento $1 - \alpha$ es el *nivel de confianza* del contraste. La *potencia* del contraste es el complementario de la probabilidad β de cometer un error de tipo II: la probabilidad de rechazar la hipótesis nula cuando ésta es falsa. Un contraste es *insesgado* cuando su potencia es mayor que su tamaño para todos los posibles valores del parámetro desconocido.

Lógicamente, el investigador quiere que un contraste sea potente, y también que las probabilidades de ambos errores sean pequeñas, pero ello no es posible simultáneamente; generalmente sólo puede reducirse el error de tipo I de un contraste a costa de aumentar el error de tipo II, y viceversa.

Un primer procedimiento para llevar a cabo contrastes de hipótesis se basa en el uso de intervalos de confianza. Este método es especialmente adecuado cuando se contrasta una hipótesis nula simple (es decir, cuando Θ contiene un solo valor numérico, θ_0), frente a una alternativa compuesta como $H_1: \theta \neq \theta_0$. Estimado $\hat{\theta}$, se construye un intervalo del nivel de confianza escogido alrededor de $\hat{\theta}$ y se rechaza H_0 si θ_0 cae fuera de dicho intervalo.

Otro procedimiento de contrastación consiste en evaluar la discrepancia con que la estimación de que se dispone deja de satisfacer las restricciones que se especifican en la hipótesis nula H_0 . Si, por ejemplo, H_0 es $\theta = \theta_0$, siendo θ_0 un determinado valor numérico, entonces la discrepancia mencionada es $\hat{\theta} - \theta_0$. Se trata de decidir si ésta es grande o pequeña; lo que sabemos es que se trata de una variable aleatoria, puesto que $\hat{\theta}$ lo es. Para saber si es grande, debe compararse con su desviación típica. El *estadístico de Wald* compara el cuadrado de la discrepancia $(\hat{\theta} - \theta_0)^2$ con la varianza $\text{Var}(\hat{\theta} - \theta_0)$, es decir, evalúa el cociente:

$$\frac{(\hat{\theta} - \theta_0)^2}{\text{Var}(\hat{\theta})}$$

donde se ha utilizado $\text{Var}(\hat{\theta}) = \text{Var}(\hat{\theta} - \theta_0)$. Dicho estadístico se distribuye como una chi-cuadrado con un grado de libertad. Si θ fuese un vector, el estadístico de Wald sería:

$$W = (\hat{\theta} - \theta_0)' [\text{Var}(\hat{\theta})]^{-1} (\hat{\theta} - \theta_0)$$

con distribución χ_k^2 , siendo k la dimensión de θ .

Para un contraste de restricciones lineales más general $H_0: \mathbf{R}\theta = \mathbf{r}$, donde \mathbf{R} es una matriz $q \times k$ y \mathbf{r} un vector $q \times 1$, ambos conocidos, el estadístico de Wald es:

$$(\mathbf{R}\hat{\theta} - \mathbf{r})' [\text{Var}(\mathbf{R}\hat{\theta} - \mathbf{r})]^{-1} (\mathbf{R}\hat{\theta} - \mathbf{r}) = (\mathbf{R}\hat{\theta} - \mathbf{r})' [\mathbf{R} \text{Var}(\hat{\theta}) \mathbf{R}']^{-1} (\mathbf{R}\hat{\theta} - \mathbf{r})$$

con distribución χ_q^2 .

En el Capítulo 11 analizamos la utilización de este estadístico para el contraste de hipótesis no lineales.

La función de verosimilitud puede también utilizarse para llevar a cabo contrastes de hipótesis. Supongamos que, acerca de los parámetros θ de una distribución de probabilidad, se establece la hipótesis nula $H_0: \theta \in \Theta$ frente a la alternativa $H_1: \theta \in \Omega$. Si denotamos por $L(\Theta)$ y por $L(\Omega)$ los valores máximos de dicha función en los subespacios de valores paramétricos Θ y Ω , entonces se tiene que el doble del *estadístico de razón de verosimilitudes*, cambiado de signo:

$$RV = -2[\ln L(\Theta) - \ln L(\Omega)]$$

sigue una distribución chi-cuadrado con k grados de libertad, donde k es el número de restricciones que definen el subespacio Θ . Es muy habitual que el subespacio Θ se reduzca a un punto, de modo que la hipótesis nula postula que los parámetros toman un determinado valor. En tal caso, $L(\Theta)$ es, simplemente, la evaluación de la función de verosimilitud en dicho punto. Simultáneamente, el subespacio Ω suele coincidir con todo el espacio paramétrico, por lo que $L(\Omega)$ es el valor de la función en el estimador de máxima verosimilitud.

La interpretación del contraste es que, si la hipótesis nula es cierta, entonces ambos valores de L serán muy similares, por lo que el estadístico será pequeño, inferior al valor de las tablas de la variable chi-cuadrado, y no se rechazará la hipótesis nula. Lo contrario tenderá a ocurrir si la hipótesis nula es falsa.

Utilizando asimismo la función de verosimilitud, el *estadístico de los multiplicadores de Lagrange* viene dado por:

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)' [\mathbf{I}(\hat{\theta}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)$$

donde $\hat{\theta}_R$ denota el estimador de máxima verosimilitud bajo las restricciones que aparecen en la hipótesis nula H_0 . Cuando H_0 es una hipótesis simple, $\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R}$ es el vector gradiente evaluado en θ_0 , el único punto recogido en H_0 . El estadístico LM se distribuye como una chi-cuadrado con número de grados de libertad igual al número de restricciones. El lector debe notar que para el cálculo de LM sólo se precisa del estimador restringido, de igual modo que para el cálculo de W sólo se precisaba del estimador sin restricciones. Por el contrario, para el cálculo de RV se precisan ambas estimaciones.

2.13. DISTRIBUCIONES TRUNCADAS

La distribución de probabilidad de una variable aleatoria truncada se obtiene normalizando la parte relevante de la función de distribución, de modo que

la probabilidad total sea igual a 1. Por ejemplo, dada una variable aleatoria X con función de densidad $f(x)$ sobre la recta real, supongamos que truncamos su distribución de probabilidad por debajo del punto a , fijo. Tenemos así una distribución de probabilidad diferente de la original, que corresponde a una variable X^* que coincide con X cuando ésta toma un valor superior a a , y no se observa en caso contrario.

La función de densidad de X^* queda definida por:

$$f(x^*) = \frac{f(x)}{1 - F(a)}$$

Proposición 2.19. Si $f(x)$ es una densidad Normal(μ, σ), la variable X^* recibe el nombre de *normal truncada* y definiendo $\alpha = (a - \mu)/\sigma$ se tiene:

$$P(X > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right) = 1 - \Phi(\alpha)$$

$$f(x/x > a) = \frac{f(x)}{P(X > a)} = \frac{\left(\frac{1}{2\pi\sigma^2}\right)^{1/2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{1 - \Phi(\alpha)}$$

donde ϕ y Φ denotan las funciones de densidad y distribución de una variable Normal(0, 1), con momentos:

$$E(X^*) = \mu + \sigma\lambda(\alpha)$$

$$\text{Var}(X^*) = \sigma^2(1 - \delta(\alpha))$$

donde:

$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{si el truncamiento es del tipo } x > a$$

$$\lambda(\alpha) = -\frac{\phi(\alpha)}{\Phi(\alpha)} \quad \text{si el truncamiento es del tipo } x < a$$

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

Además se tiene $0 < \delta(\alpha) < 1$, lo que implica:

1. Si el truncamiento es por debajo de a_0 , la esperanza de la variable truncada es mayor que la de la variable original. Lo contrario ocurriría si el truncamiento fuese por encima de a_0 , es decir, si sólo se observasen los valores de X inferiores a a_0 ,

2. El truncamiento reduce la varianza de una variable.

En términos notacionales es importante observar que:

$$\frac{d\phi(\alpha)}{d\alpha} = -\alpha\phi(\alpha) \quad \text{que implica} \quad \delta(\alpha) = \frac{d\lambda(\alpha)}{d\alpha}$$

La función $\lambda(\alpha)$ se denomina *Ratio de Mills*.

Ejemplo. Consideremos la variable aleatoria X , con densidad Uniforme $[0, 1]$, truncada inferiormente en el punto $x = 1/4$. Su densidad truncada es:

$$f(x^*) = \frac{f(x)}{P(X > 1/4)} = \frac{1}{3/4} = \frac{4}{3} \quad \text{si} \quad \frac{1}{4} < x^* \leq 1$$

$$= 0 \quad \text{si} \quad x^* \leq 1/4$$

cuya masa total de probabilidad es 1, puesto que es una variable $U(1/4, 1)$. Su esperanza matemática es:

$$E(X^*) = \frac{5}{8}$$

que es superior a $E(X)$, y su varianza:

$$\text{Var}(X^*) = \int_{1/4}^1 \frac{4}{3} \left(u - \frac{5}{8}\right)^2 du = \frac{4}{9} \left[\left(u - \frac{5}{8}\right)^3 \right]_{1/4}^1 = \frac{3}{64}$$

que es inferior a la varianza de X .

PROBLEMAS

Problema 2.1. Sea $f(x, y) = kxy$ una función de densidad conjunta definida sobre el rectángulo $0 \leq x \leq 1$, $0 \leq y \leq 1$. Hallar el valor de la constante k . Encontrar las funciones de densidad marginales de x e y , así como sus esperanzas y varianzas. Obtener las funciones de densidad de x condicionada por y , así como la de la variable y condicionada por x . Probar que x e y son independientes.

Problema 2.2. Sea X una variable con distribución uniforme en el intervalo $(1, 2)$ y sea y una variable tal que $y = \ln x$. Probar que

$$E(y) = 2 \ln 2 - 1$$

Problema 2.3. Sea una variable X uniforme en el intervalo $(0, 1)$ y considérese la función $y = \frac{x}{1+x}$. Hallar la función de densidad de y . Probar que su esperanza es igual a $1 + \ln \frac{1}{2}$.

Problema 2.4. Probar que las variables aleatorias, cuya función de densidad conjunta es:

$$f(x, y) = \frac{1}{2} \quad \text{si } x \in (0, 2), \quad y \in (0, 1)$$

son independientes entre sí.

Problema 2.5. Sea X una variable aleatoria con distribución:

Valores posibles de X :	-3	-1	0	1	3
Probabilidades:	$\frac{6}{24}$	$\frac{2}{24}$	$\frac{4}{24}$	$\frac{10}{24}$	$\frac{2}{24}$

Calcular la distribución de probabilidad de la variable aleatoria $Y = X^2$.

Problema 2.6. Supongamos que la variable Y tiene la siguiente distribución de probabilidad:

$$Y = \begin{cases} -1 & \text{con probabilidad } \frac{1}{3} \\ 1 & \text{con probabilidad } \frac{2}{3} \end{cases}$$

y que la variable aleatoria X sigue una distribución uniforme sobre el intervalo $[0, 2]$ si $Y = -1$, y una distribución uniforme sobre el intervalo $(1, 5)$ si $Y = 1$.

- Obtener la distribución de probabilidad de la variable X condicionada por la variable Y .
- Obtener la distribución de probabilidad de la esperanza condicional $E(X/Y)$.
- Obtener la distribución marginal de la variable X . Representar sus funciones de densidad y de distribución.
- Comprobar que la esperanza matemática de esta última distribución coincide con la de la variable aleatoria $E(X/Y)$.

Problema 2.7. Obtenga el estimador de máxima verosimilitud de los parámetros de una distribución $N(\mu, \sigma^2)$. Obtenga su matriz de información y compruebe si las varianzas de los estimadores propuestos alcanzan las cotas mínimas de Cramer-Rao. ¿Puede estar seguro de su eficiencia?

Problema 2.8. Obtener el estimador de máxima verosimilitud de:

- El parámetro λ en una distribución de Poisson.
- El parámetro π en una distribución binomial.
- El parámetro π en una distribución geométrica.
- El parámetro a en una distribución uniforme en $[0, a]$.
- El parámetro λ en una distribución exponencial.

Problema 2.9. Demuestre la Proposición 2.10.

Problema 2.10. Utilizar la Proposición 2.17 para probar que $\bar{X}_T \xrightarrow{D} \mu$, donde \bar{X}_T es la media muestral calculada a partir de una muestra de tamaño T de una población Normal (μ, σ^2) .

[Indicación: La función característica de una variable $N(\mu, \sigma^2)$ es $\Psi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$.]

Problema 2.11. Probar que si la sucesión de variables aleatorias X_t converge en distribución a la constante c , entonces también converge en probabilidad a dicha constante.

Problema 2.12. Sea $\{X_t\}_{t=1}^{\infty}$ una sucesión de variables aleatorias tales que $\lim_t EX_t^2 = 0$. Probar que $\text{plim } X_t = 0$.

Problema 2.13. Obtener el contraste de razón de verosimilitudes correspondientes a la hipótesis nula: $H_0: \mu = \mu_0$, a partir de una muestra obtenida de una población $N(\mu, \sigma^2)$, con $\sigma^2 = 1$. ¿Cómo cambiaría su respuesta si σ^2 fuese desconocido?

Problema 2.14. Dada la variable aleatoria X con función de densidad $f(x) = \frac{3}{8}(x+1)^2$ en el intervalo $(-1, 1)$ y dada la variable $Y = \sqrt{X+1}$, hallar la función de densidad de esta última, así como su esperanza y su varianza.

Problema 2.15. Dada una muestra aleatoria $X_1 X_2 \dots X_T$ de una población cuyos momentos existen, probar que la distribución del estadístico $\frac{1}{T} \sum_1^T (X_t - \bar{X})^2$ puede aproximarse por una $N\left(\sigma^2, \frac{m_4 - \sigma^4}{T}\right)$.

Probar que $(s')^2 = \frac{1}{T-1} \sum_1^T (X_t - \bar{X})^2$ tiene la misma distribución límite que el estadístico anterior y que su varianza común, $\frac{m_4 - \sigma^4}{T}$, se convierte en $\frac{2\sigma^4}{T}$ si la población de la que se extrajo la muestra es Normal.

Por último, probar que $s' = \sqrt{(s')^2}$ puede aproximarse por una distribución $N\left(\sigma, \frac{m_4 - \sigma^4}{4T\sigma^2}\right)$. Simplificar esta varianza en el caso de una población Normal.

Problema 2.16. Demostrar la Proposición 2.6.

Problema 2.17. Demuestre las Proposiciones 2.7 y 2.8.

Problema 2.18. Obtenga los resultados concernientes a la función de densidad bivariente:

$$f(x, y) = kx(x+y), \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

que se muestran en la Sección 2.5.

Problema 2.19. Compare los errores cuadráticos medios de la varianza y cuasivarianza muestral en la estimación de σ^2 en una población $N(\mu, \sigma^2)$.

Problema 2.20. Encontrar la mejor predicción de y dado el valor de x si ambas variables aleatorias tienen como función de densidad conjunta

$$f(x, y) = xy, \quad 0 \leq x \leq 2, \quad 0 \leq y \leq 1$$

y se pretende minimizar el *error cuadrático medio* de la predicción obtenida.

Problema 2.21. Suponga que la distribución conjunta de dos variables x e y viene dada por la función de densidad

$$f(x, y) = \frac{\theta e^{-(\beta + \theta)y} \cdot (\beta y)^x}{x!}, \quad \beta, \theta > 0, \quad y \geq 0, \quad x = 0, 1, 2, \dots$$

1. Encuentre los estimadores de máxima verosimilitud (MV) de β y θ y su distribución asintótica conjunta.

2. Encuentre el estimador MV de $\frac{\theta}{\beta + \theta}$ y su distribución asintótica.

3. Demuestre que la función de densidad marginal de la variable x es de la forma

$$f(x) = \gamma(1 - \gamma)^x, \quad x = 0, 1, 2, \dots$$

y halle el estimador MV de γ y su distribución asintótica.

4. Demuestre que la densidad condicional $f(y/x)$ es de la forma

$$\frac{\lambda e^{-\lambda y} (\lambda y)^x}{x!}$$

y halle el estimador MV de x y su distribución asintótica.

5. Demuestre que la densidad marginal de y es

$$f(y) = \theta e^{-\theta y}, \quad y \geq 0$$

y halle el estimador MV de θ .

6. Demuestre que la densidad condicional $f(x/y)$ es de la forma

$$f(x/y) = \frac{e^{-\beta y} (\beta y)^x}{x!}$$

y halle el estimador MV de β .

CAPITULO 3

EL MODELO LINEAL GENERAL

3.1. INTRODUCCION

El objeto de la Econometría consiste en:

- a) especificar un modelo de relación entre variables económicas,
- b) utilizar información muestral acerca de los valores tomados por dichas variables, con el objeto de cuantificar la magnitud de la dependencia entre ellas,
- c) evaluar críticamente la validez de hipótesis propuestas por la Teoría Económica acerca de las relaciones estimadas y, en algunos casos,
- d) efectuar un ejercicio de seguimiento coyuntural y de previsión de las variables analizadas.

Así, al utilizar los métodos econométricos tratamos de responder a preguntas como:

1. ¿Cuáles son los determinantes de la tasa de inflación?
2. Sobre la base de la información histórica disponible, ¿cuál es la importancia cuantitativa de cada uno de dichos determinantes?
3. ¿Podemos contrastar algunas de las implicaciones de la Teoría Económica acerca del efecto que variables como el crecimiento monetario tienen sobre la tasa de inflación?
4. ¿Qué sugiere nuestro modelo estimado para la tasa de inflación acerca del comportamiento de esta variable durante el próximo año?

En consecuencia, el analista económico debe comenzar especificando muy claramente cuál es el centro de atención de su trabajo empírico; luego, debe tratar de identificar cuáles son los determinantes que explican la evolución de esta variable; debe escoger cuidadosamente la información estadística relevante para cuantificar tal relación, y debe proceder, finalmente, a su cuantificación. Por último, utilizará el *modelo de relación estimado*, ya sea a efectos de contrastación de algún supuesto teórico, o como elemento de análisis y seguimiento de la variable cuyo comportamiento escogió explicar.

El objeto de nuestro estudio a lo largo de este texto es un modelo de relación entre variables económicas, que denotamos genéricamente:

$$y = f(x_1, x_2, x_3, \dots, x_k, u/\beta) \quad [3.1]$$

que trata de explicar el comportamiento de una variable económica y utilizando la información proporcionada por un conjunto de k variables explicativas con un claro significado económico, así como por una variable aleatoria, *no observable* y , por consiguiente, sin significado conceptual económico, que denotaremos por u . Las variables observables constituyen el vector x , de dimensión $k \times 1$, o representado como una fila $x' = (x_1, x_2, x_3, \dots, x_k)$, y la relación de dependencia entre la variable y y el vector x envolverá, generalmente, un vector de parámetros que denotamos por β .

Con objeto de analizar *empíricamente*, es decir, utilizando datos reales, las características de la relación [3.1], recogemos *información muestral*, que consiste en una lista ordenada de valores numéricos de las variables y, x_1, x_2, \dots, x_k . **En una muestra de sección cruzada**, diversos agentes económicos de una naturaleza similar proporcionan la información solicitada *en un mismo instante de tiempo*. Alternativamente, el investigador económico trabaja en ocasiones con datos de **series temporales**, en las que se dispone de información acerca de una unidad económica, como puede ser un país, *a lo largo del tiempo*; estas muestras pueden tener frecuencia diaria, mensual, anual, etc., según la frecuencia de observación de los datos. En ocasiones, series temporales observadas frecuentemente (diariamente, por ejemplo) se transforman en series mensuales o trimestrales antes de su utilización, ya sea mediante agregación o tomando promedios de observaciones sucesivas.

Cuando utilizamos una muestra de sección cruzada, empleamos el subíndice i para denotar los valores de las variables correspondientes a la unidad económica i -ésima; cuando utilizamos datos de series temporales, utilizamos el subíndice t para denotar las observaciones correspondientes a un mismo instante de tiempo. De este modo, disponemos en realidad de una lista de relaciones:

$$y_i = f(x_{1i}, x_{2i}, x_{3i}, \dots, x_{ki}, u_i/\beta), \quad i = 1, 2, \dots, N \quad [3.2]$$

que relacionan los valores correspondientes $y_i, x_{1i}, x_{2i}, \dots, x_{ki}$ que componen cada una de las N observaciones muestrales. El modelo anterior está escrito para el caso de una sección cruzada de datos; en el caso de datos de series temporales tenemos:

$$y_t = f(x_{1t}, x_{2t}, x_{3t}, \dots, x_{kt}, u_t/\beta), \quad t = 1, 2, \dots, T \quad [3.3]$$

Salvo algunas excepciones que resultarán explícitas en su momento, tratamos en este libro con relaciones de dependencia lineal, es decir:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, 2, \dots, N \quad [3.4]$$

que denominamos *modelo de regresión lineal múltiple*, *modelo econométrico* o, más sencillamente, *modelo lineal general*. En él, los componentes del vector β son los *coeficientes* de las variables explicativas en el modelo lineal.

La variable aleatoria u_i , a la que nos referimos en lo sucesivo como *término de error* del modelo, entra aditivamente en el modelo y no precisa ir acompañada de ningún coeficiente, por razones que resultarán claras más adelante. La variable y se denomina *variable endógena*, mientras que las variables explicativas $x_1, x_2, x_3, \dots, x_k$ se denominan *variables explicativas* del modelo. Aunque volveremos a este punto más adelante, debe resultar claro que los coeficientes $\beta_1, \beta_2, \dots, \beta_k$ recogen la magnitud del impacto de cada una de las variables explicativas sobre la variable endógena. Hay otros modelos en que la forma funcional de la dependencia entre la variable endógena y las explicativas es más compleja que la que aparece en [3.4]; en otras ocasiones, el modelo econométrico consta de más de una ecuación: todos estos aspectos serán discutidos en capítulos posteriores.

En muchas ocasiones, el modelo de relación incorpora un término constante:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, 2, \dots, N \quad [3.5]$$

que interpretaremos como acompañando a una primera variable explicativa x_{1i} cuyo valor es siempre igual a 1: $x_{1i} = 1, i = 1, 2, \dots, N$. Si dispusiésemos de una muestra de series temporales, escribiríamos:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t, \quad t = 1, 2, \dots, T \quad [3.6]$$

En este caso, a los demás coeficientes, es decir, a todos menos al término independiente, se les denomina *pendientes* del modelo de regresión.

Ejemplo 3.1. Un investigador se propone medir la propensión marginal al consumo de las familias españolas, para lo que recoge datos acerca de los gastos de consumo, de los ingresos laborales y rentas de capital, y de los impuestos que pagan un conjunto de N familias. Con ellos especifica el modelo:

$$\text{Consumo}_i = \alpha + \beta_1 \text{Ingresos totales}_i + \beta_2 \text{Impuestos}_i + u_i, \quad i = 1, 2, \dots, N$$

en el que ha agregado en la variable «Ingresos Totales» las rentas salariales y de capital. El investigador esperará obtener en su estimación un valor positivo para el coeficiente β_1 y negativo para β_2 , y quizá querrá contrastar la hipótesis nula $H_0: \beta_2 = -\beta_1$, que implicaría que el consumo depende de la Renta Disponible más que de los Ingresos Totales de las familias.

Ejemplo 3.2. Un analista de la economía española pretende estimar un modelo que explique la evolución temporal de la tasa de inflación. Para ello, recoge datos mensuales desde enero de 1976 hasta diciembre de 1991 acerca

del Índice de Precios al Consumo (IPC), así como de los Activos Líquidos en Manos del Público (ALP), y especifica el modelo

$$\text{IPC}_t = \alpha + \beta \text{ALP}_t + u_t, \quad t = 1, 2, \dots, 192$$

aunque también considera la posibilidad de obtener las tasas anuales de crecimiento de cada una de estas variables, denotarlas por π_t —tasa de inflación, es decir, tasa anual de crecimiento del IPC— y m_t —tasa de crecimiento de los ALP— y especificar el modelo

$$\pi_t = \alpha + \beta m_t + v_t, \quad t = 1, 2, \dots, 16$$

para, una vez estimado, contrastar la hipótesis de neutralidad monetaria $H_0: \beta = 1$, que significaría que toda expansión monetaria se traduce en un crecimiento similar de los precios, sin tener efectos reales.

3.1.a. Características del modelo

1. El modelo econométrico es estocástico

La presencia del término de error u hace que la relación entre la variable endógena y las explicativas sea estocástica, como contraposición a la posibilidad de que y hubiese dependido del vector x de modo determinista, es decir, a través de una relación funcional exacta, como en:

$$y_i = 3,7 - 5,1x_{1i} + 2,4x_{2i}, \quad i = 1, 2, \dots, N$$

El investigador considera que la relación es estocástica por las siguientes razones:

- a) Porque el modelo [3.5] es sólo una aproximación al verdadero modelo de relación entre y y el vector x , que es mucho más complejo y de difícil especificación.
- b) Porque gran parte de las variables económicas de interés están sujetas a errores de medida, en muchos casos por inferirse su valor a partir de muestras finitas y , en otros casos, por no ajustarse exactamente al concepto económico que el investigador querría incorporar en su modelo econométrico.
- c) Porque reconocemos la posible existencia de otros factores determinantes del comportamiento de y que no hemos incluido en el modelo, bien por desconocimiento de dichos factores o porque no dispongamos de observaciones numéricas de los mismos.

Suponemos que la *esperanza matemática* del término de error u_i del modelo es cero; si, por el contrario, tuviésemos $E(u_i) = a \neq 0$, éste sería un efecto constante y , por ello, determinista sobre y_i , y debería incluirse como parte de la constante β_1 en [3.5]. Este supuesto propone una esperanza matemática nula para cada una de dichas variables aleatorias, ya correspondan a las

distintas observaciones de sección cruzada o a períodos diferentes de tiempo, de modo que podemos recoger todos ellos formulando $E(\mathbf{u}) = \mathbf{0}_N$, ($E(\mathbf{u}) = \mathbf{0}_T$ en el caso de [3.6]), donde \mathbf{u} denota el vector formado por todos los términos de error del modelo.

Una situación en que este supuesto no se cumpliría es cuando el investigador, por error, omite del modelo una variable explicativa relevante. Así, supongamos que en vez de especificar el modelo

$$y_t = \alpha + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

se especifica el modelo

$$y_t = \alpha + \beta_2 x_{2t} + v_t$$

en cuyo caso el término de error v_t sería igual a $v_t = u_t + \beta_3 x_{3t}$ y su esperanza matemática, si la variable x_{3t} se supone determinista, será:

$$E(v_t) = E(u_t) + \beta_3 E(x_{3t}) = \beta_3 x_{3t}$$

que será distinta de cero en general.

El término de error del modelo [3.5] es un vector de dimensión $N \times 1$ (o $T \times 1$ en el caso de [3.6]). En consecuencia, su matriz de covarianzas $\text{Var}(\mathbf{u})$, que tiene por elemento genérico $\sigma_{ij} = \text{Cov}(u_i, u_j)$, es simétrica, definida positiva, de dimensión $N \times N$ ($T \times T$ en el caso de datos de series temporales):

$$\text{Var}(\mathbf{u}) = \begin{pmatrix} \text{Var } u_1 & \text{Cov}(u_1, u_2) & \cdots & \text{Cov}(u_1, u_{N-1}) & \text{Cov}(u_1, u_N) \\ \text{Cov}(u_2, u_1) & \text{Var } u_2 & \cdots & \text{Cov}(u_2, u_{N-1}) & \text{Cov}(u_2, u_N) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \text{Cov}(u_{N-1}, u_1) & \text{Cov}(u_{N-1}, u_2) & \cdots & \text{Var}(u_{N-1}) & \text{Cov}(u_{N-1}, u_N) \\ \text{Cov}(u_N, u_1) & \text{Cov}(u_N, u_2) & \cdots & \text{Cov}(u_N, u_{N-1}) & \text{Var}(u_N) \end{pmatrix}$$

Suponemos inicialmente que esta matriz de covarianzas tiene estructura:

$$\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_N \quad [3.7]$$

(o $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_T$ en el caso de [3.6]), donde hemos recogido simultáneamente dos hipótesis:

a) Por una parte, estamos suponiendo que los elementos de fuera de la diagonal de la matriz son iguales a cero. Estos elementos son las covarianzas entre los términos de error correspondientes a observaciones distintas; al suponer que sus covarianzas son iguales a cero, estamos afirmando que dichos términos de error están incorrelacionados entre sí.

b) En segundo lugar, decimos que la matriz es escalar, es decir, la diagonal principal es constante. Los elementos de dicha diagonal son las varianzas de los términos de error de las distintas observaciones; en conse-

cuencia, estamos suponiendo que la varianza del término de error del modelo es constante, ya sea a lo largo del tiempo o a través de la sección cruzada.

A lo largo de este texto, dedicamos capítulos específicos al análisis de modelos econométricos en que estos supuestos no se satisfacen. Es importante que pensemos ahora, sin embargo, que una cierta hipótesis acerca de la estructura de la matriz $\text{Var}(\mathbf{u})$ es necesaria, puesto que dicha matriz contiene, en general, $N(N + 1)/2$ [o $T(T + 1)/2$] elementos diferentes; éste es un número que crece muy rápidamente con el número de observaciones disponibles, haciendo imposible la estimación de todos y cada uno de dichos parámetros. El supuesto que aquí hemos hecho es el más sencillo de todos, haciendo que todos los elementos de la matriz sean función de un único parámetro, σ_u^2 .

2. El modelo econométrico es lineal

En la mayor parte de este libro —excepto en un capítulo al efecto— suponemos que el modelo que relaciona variables endógena y explicativas es lineal en los coeficientes β , como ocurre en [3.5] o [3.6]. Evidentemente, no toda relación entre variables es lineal, pero aunque no lo fuese, podría considerarse a [3.5] y [3.6] como una aproximación lineal en serie de Taylor del verdadero modelo no lineal de relación entre y y x que aparece en [3.1].

Debemos observar también que algunas relaciones no lineales entre variables pueden transformarse fácilmente en lineales, como ocurre cuando se toman logaritmos en una función de producción Cobb-Douglas, del modo que vimos en la Introducción. Otras funciones, en cambio, no aceptan una transformación similar, como es el caso de la función de producción con elasticidad constante de sustitución:

$$Q_t = e^\alpha [(1 - \delta)L_t^{-\gamma} + \delta K_t^{-\gamma}]^{-\frac{\beta}{\gamma}} \varepsilon_t$$

donde Q_t denota la cantidad de producto obtenida, K_t y L_t son las cantidades de los factores capital y trabajo utilizadas en la producción de Q_t y ε_t es una perturbación aleatoria en la función de producción. Tomando logaritmos se tiene:

$$\ln Q_t = \alpha - \frac{\beta}{\gamma} \ln [(1 - \delta)L_t^{-\gamma} + \delta K_t^{-\gamma}] + \ln \varepsilon_t$$

que no es lineal en sus parámetros α , β , γ y $\delta^{(1)}$.

⁽¹⁾ Afortunadamente, aun siendo complicada, esta función admite un tratamiento lineal, pues, dados valores numéricos de δ y γ , el modelo es lineal en α y β , parámetros que pueden ser estimados por procedimientos sencillos, *condicionalmente* en los valores supuestos para γ y δ . Si el rango de valores admisibles para estos dos últimos parámetros es acotado, entonces puede establecerse una red de pares de valores para ellos, calcular las estimaciones de α y β para cada par (γ, δ) y elegir aquel conjunto de estimaciones numéricas para los cuatro parámetros que resulte óptimo de acuerdo con alguno de los criterios de estimación que introduciremos más adelante.

3. Los coeficientes del modelo $\beta_1, \beta_2, \dots, \beta_k$ son constantes en el tiempo

Mantenemos este supuesto a lo largo de este libro, por lo que *el problema de estimación* consiste en utilizar la información muestral para asociar valores numéricos a estos k coeficientes. Si, por el contrario, permitiésemos que los coeficientes variasen en el tiempo, el problema de estimación sería más complejo; tendría, sin embargo, la ventaja de que permitiría discutir acerca de la estabilidad temporal del modelo.

4. Existe una relación causal desde las variables explicativas hacia la variable endógena

Cuando especificamos un modelo econométrico como [3.5] o [3.6], entendemos que la Teoría Económica subyacente aporta suficientes elementos como para sugerir que las variables explicativas x influyen sobre la variable y , y no al revés. En tal situación, se denomina a las variables x como **exógenas**, queriendo significar con ello que, desde el punto de vista del modelo que se está analizando, sus valores se toman como dados, ya que no reciben influencia alguna de la variable que se pretende explicar y que, por contraposición, recibe el calificativo de **endógena**. Es importante entender que tal clasificación depende del modelo en que se incluye una determinada variable, y no es una propiedad que una variable económica arrastra consigo por todos los modelos a los que se incorpora.

En nuestro Ejemplo 2, si durante el período muestral se ha seguido una política monetaria consistente en fijar un determinado crecimiento anual para la cantidad de dinero y preocuparse únicamente porque el crecimiento monetario se ajuste lo más posible a dicho objetivo, el crecimiento monetario es una variable exógena en el modelo que pretende explicar la tasa de inflación. Si, por el contrario, se ha seguido una política monetaria con realimentación, es decir, donde el crecimiento monetario se ha decidido en cada período con la atención puesta en las tasas de inflación que hasta entonces se han registrado, entonces no estaría justificado calificar de exógeno al crecimiento monetario.

5. Las variables x no son linealmente dependientes

Esta es una de las características menos estrictas de las que aquí describimos, pues excluye tan sólo la posibilidad de que alguna de las variables explicativas del modelo econométrico pueda escribirse como *combinación lineal exacta* de las demás. En realidad, todas las variables económicas muestran algún grado de correlación entre sí, y ello no nos produce excesivas dificultades, excepto cuando se llega a una situación de dependencia total, que es lo que excluimos al afirmar que las variables explicativas no son linealmente dependientes entre sí.

6. Las variables x son deterministas

Este supuesto, que mantenemos en nuestra discusión inicial del modelo, significa que si tuviésemos la oportunidad de obtener otra muestra, además de la ya disponible, los valores de las variables explicativas serían los mismos que los ya observados en la muestra que en este momento tenemos. Nótese que el supuesto no incluye a la variable endógena y . De hecho, ya hemos visto que y es aleatoria (por ser función del término de error u), y sus valores observados serían diferentes si pudiésemos disponer de una muestra distinta.

Consideremos, en el marco de nuestro Ejemplo 2, el alcance de este supuesto: de acuerdo con el mismo, si pudiésemos volver al año inicial (1976) en las mismas condiciones económicas entonces existentes y recoger otra muestra para el mismo período (enero 1976-diciembre 1991), obtendríamos los mismos valores del crecimiento monetario. Desde este punto de vista, las tasas de crecimiento de la oferta monetaria que se han observado en este período son las únicas que pudieron haber ocurrido, con independencia de la información de que dispuso la autoridad monetaria y de los objetivos de política económica que se trazaron. Nótese que, en esta hipotética situación, las tasas de inflación observadas para el período serían diferentes entre distintas muestras, debido a su componente estocástica u_t .

Enlazando con la discusión que mantuvimos en el punto anterior, podría tener sentido mantener el supuesto de variables explicativas x deterministas bajo una política monetaria de fijación de una tasa de crecimiento constante todos los años; por el contrario, no podría mantenerse el supuesto bajo una política en que el crecimiento de la oferta monetaria se hace depender del «estado» de economía y , en particular, de la evolución de la tasa de inflación. De este modo, la clasificación de las variables explicativas en «exógenas» o «endógenas» está ligada a que podamos mantener o no el supuesto de que dichas variables son de naturaleza determinista.

Cuando se trabaja con datos de series temporales, en ocasiones aparecen valores retardados y_{t-j} de la variable endógena como variables explicativas. Hay muchos argumentos que justifican tal inclusión, pues la presencia de coste de ajuste de uno u otro tipo impide que las variables económicas tomen de inmediato el valor deseado por los agentes evolucionando, por el contrario, con cierta inercia. Sin embargo, si bien es muy aceptable que los retardos de y influyan sobre su valor actual, lo contrario no es aceptable. Cuando tales retardos se incluyen como variables explicativas, decimos que son variables *predeterminadas*; sería una contradicción decir que son exógenas, pero, sin embargo, su valor numérico está determinado *previamente* al de y_t , lo que justifica el apelativo propuesto. Introducimos estos conceptos con mayor rigor en el Capítulo 9.

3.1.b. Descripción de los capítulos posteriores

Nuestra pretensión es desarrollar en los Capítulos 3 y 4 la teoría de estimación, contrastación de hipótesis y predicción en modelos econométricos lineales como el que aquí hemos introducido. En capítulos posteriores iremos

abandonando algunas de las características de este modelo para extender el tratamiento a situaciones más generales, que son de interés para el analista de datos económicos.

1. El supuesto $\text{Var}(\mathbf{u}) = \sigma_u^2 \cdot \mathbf{I}_T$ puede abandonarse en cada una de las dos dimensiones que antes citamos:

- a) Cuando la varianza del término de error es diferente para cada observación muestral, decimos que tenemos una situación de *heteroscedasticidad*, que analizamos en el Capítulo 6.
- b) Cuando, con datos de series temporales, los términos de error correspondientes a períodos diferentes están correlacionados, decimos que existe *autocorrelación*. Esta situación se analiza en el Capítulo 7, mientras que el Capítulo 5 constituye una introducción común a los problemas de heteroscedasticidad y autocorrelación.

2. Si las variables explicativas no son todas independientes entre sí, se producen dificultades cuando se trata de *estimar* los coeficientes β_i del modelo, es decir, de asociar valores numéricos a los mismos, sobre la base de la información muestral disponible, utilizando los procedimientos que introducimos en la próxima sección. Ello se debe a que, al existir correlaciones elevadas entre algunas de las variables explicativas, es difícil desagregar su capacidad explicativa global en las componentes atribuibles a cada una de ellas. Esta situación se denomina como *multicolinealidad* y se estudia en el Capítulo 10.

3. El Capítulo 13 analiza *modelos univariantes de series temporales*, que son aquellos en que las variables explicativas son todas valores retardados de la variable endógena.

4. Las variables explicativas dejan de ser determinista en varios casos:

- a) Cuando, tratando con datos de series temporales, se incluye como explicativa algún retardo de la variable endógena, como y_{t-1} ; al igual que y_t , sus retardos son variables aleatorias y pueden exigir un tratamiento especial. Denominamos *dinámicos* a estos modelos y los analizamos en el Capítulo 9.
- b) En ocasiones hay que aceptar que alguna de las variables explicativas es endógena y tiene una determinación simultánea con la variable que inicialmente pretendíamos explicar. En tales casos hablamos de *causalidad bidireccional*, y tiene sentido especificar ecuaciones adicionales que, junto con la primera, constituyen un *sistema de ecuaciones simultáneas*, que analizamos en los Capítulos 17 y 18.

5. Se abandona el supuesto de linealidad en el Capítulo 11, donde se extiende el tratamiento que proponemos para modelos lineales a formas funcionales más generales. En el Capítulo 12 se introducen elementos de cálculo que resultan necesarios para el análisis de modelos no lineales.

6. En ocasiones, la variable que pretendemos explicar no toma valores continuos, sino tan sólo un rango discreto, como es el caso de quien trata de explicar si el nivel de estudios que escoge un individuo en función de su renta,

edad y medio social en que vive es de: estudios superiores, medios, formación profesional o inferiores. Este tipo de modelos se analiza en el Capítulo 16.

7. Ocurre frecuentemente en Economía que las variables que se pretenden relacionar tienen un comportamiento inestable en el tiempo, en un sentido que definiremos con mayor precisión más adelante. La existencia de tendencias es uno de tales casos, que requieren un tratamiento especial que se presenta en el Capítulo 14.

8. En ocasiones, especialmente con datos microeconómicos, no se dispone de observaciones de variables que pueden ser muy relevantes en la determinación de la variable endógena. El tratamiento adecuado en tales muestras se examina en el Capítulo 15.

2. EL ESTIMADOR DE MINIMOS CUADRADOS ORDINARIOS

En esta primera presentación del modelo econométrico lineal suponemos que el investigador dispone de observaciones de series temporales; el lector puede adaptar la formulación al caso en que se dispone de una muestra de sección cruzada, sin excesiva dificultad.

Ejemplo 3.3. Consideremos cinco observaciones anuales acerca de dos variables que denotamos por x_t e y_t , la última de las cuales queremos explicar por medio de la primera. Los datos se recogen en las primeras columnas de la Tabla 3.1.

Como ya hemos visto en la introducción, el modelo econométrico sugiere que, en cada período muestral, el valor observado de la variable endógena y_t

TABLA 3.1

t	y_t	x_t	x_t^2	$x_t y_t$	$(y_t - \bar{y})^2$	\hat{u}_t	$x_t \hat{u}_t$	$(\hat{u}_t - \bar{\hat{u}})^2$
1	84	32	1024	2688	256	-3,6	-115,2	12,96
2	108	40	1600	4320	64	8,0	320,0	64,0
3	92	36	1296	3312	64	-1,8	-64,8	3,24
4	110	44	1936	4840	100	3,8	167,2	14,44
5	106	48	2304	5088	36	-6,4	-307,2	40,96
Suma	500	200	8160	20248	520	0	0	135,6
Media	100	40	1632	4049,6	104	0	0	27,2

puede explicarse en parte por los valores tomados por las variables explicativas, en este caso únicamente una constante y x_t . Dichas relaciones pueden escribirse:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + \beta_k x_{k1} & + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + \beta_k x_{k2} & + u_2 \\ \dots &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \dots \\ y_{T-1} &= \beta_1 + \beta_2 x_{2T-1} + \beta_3 x_{3T-1} + \cdots + \beta_k x_{kT-1} & + u_{T-1} \\ y_T &= \beta_1 + \beta_2 x_{2T} + \beta_3 x_{3T} + \cdots + \beta_k x_{kT} & + u_T \end{aligned} \quad [3.8]$$

Si ordenamos las observaciones de que disponemos mediante un vector y y de orden $T \times 1$ y una matriz X de orden $T \times k$, tenemos:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{T-1} \\ y_T \end{pmatrix} = \begin{pmatrix} 84 \\ 108 \\ 92 \\ 110 \\ 106 \end{pmatrix} \quad X = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ x_{1T} & x_{2T} & \dots & x_{kT} \end{pmatrix} = \begin{pmatrix} 1 & 32 \\ 1 & 40 \\ 1 & 36 \\ 1 & 44 \\ 1 & 48 \end{pmatrix} \quad [3.9]$$

donde hemos incorporado las observaciones muestrales correspondientes al ejemplo que estamos considerando, que constituyen un vector y de orden 5×1 y una matriz X , 5×2 , puesto que hay que contar la constante del modelo como una variable explicativa más.

Si además consideramos el vector formado por los k coeficientes del modelo $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ y el vector formado por los términos de error de los distintos períodos muestrales $u = (u_1, u_2, \dots, u_T)'$, entonces podemos escribir las T relaciones [3.8] en notación matricial:

$$y = X\beta + u \quad [3.10]$$

En nuestro ejemplo $\beta = (\beta_1, \beta_2)$ y $u = (u_1, u_2, u_3, u_4, u_5)$.

El primer objetivo del análisis econométrico es el de obtener estimaciones de los parámetros desconocidos del modelo. Estos son, por un lado, los coeficientes β de las variables explicativas y, por otro, los que entran en la matriz de covarianzas del término de error. Estimar consiste en utilizar la información muestral para asignar valores numéricos a dichos parámetros. Debe notarse que un estimador es, por tanto, una función del espacio muestral (el conjunto de todas las observaciones posibles que pudieron haberse tenido de las variables exógena y endógena) sobre el espacio paramétrico (el conjunto de todos los valores admisibles de los parámetros). Los distintos estimadores posibles difieren unos de otros en el modo de resumir la evidencia muestral para asociar valores numéricos a los parámetros del modelo.

Una vez *estimados* los coeficientes β , se puede calcular para cada instante t :

$$\hat{y}_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} \quad [3.11]$$

en el que las estimaciones $\hat{\beta}_i$, $1 \leq i \leq k$ han sustituido a los verdaderos valores, desconocidos. La expresión [3.11] representa la estimación, de acuerdo con el modelo econométrico, del valor que debía haber tomado la variable endógena y_t . Habrá siempre una discrepancia entre el valor realmente observado de y_t y la estimación anterior, a la que denominamos *residuo* correspondiente a dicho período $\hat{u}_t = y_t - \hat{y}_t$. De este modo podemos generar una serie de T residuos que, representados en forma matricial, como un vector $T \times 1$, son:

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad [3.12]$$

Parece razonable descomponer el problema de estimación de un modelo econométrico en dos partes: en primer lugar estimamos los coeficientes $\boldsymbol{\beta}$; con ellos obtenemos el vector de residuos $\hat{\mathbf{u}}$ y, a partir de éstos, estimamos los parámetros de la matriz de covarianzas. Hay tantos estimadores como funciones entre espacio muestral y espacio paramétrico, un número sin duda excesivamente amplio, por lo que es preciso introducir algún criterio que permita seleccionar un estimador entre el conjunto de todos los posibles. Debe quedar claro que criterios diferentes conducirán a estimadores diferentes, incluso a partir de una misma muestra; de modo similar, un determinado *estimador* producirá una *estimación* a partir de una muestra dada, pero generaría una *estimación* diferente si se dispusiese de una muestra distinta.

Parece razonable que un posible criterio que defina a un estimador sea la minimización de la magnitud de los residuos que dicho estimador genera. Tal idea es correcta, pero hay varias dificultades para hacerla práctica: en primer lugar, los residuos constituyen un vector $T \times 1$, por lo que no se trata de minimizar un residuo determinado, sino una medida conjunta del tamaño global de todos ellos.

Dado un vector de estimaciones $\hat{\boldsymbol{\beta}}$, podría pensarse en sumar los T residuos por él generados y escoger como estimación aquel vector $\hat{\boldsymbol{\beta}}$ cuya suma de residuos fuese la menor posible. Una dificultad con tal procedimiento es la cancelación de residuos en los períodos en que el vector $\hat{\boldsymbol{\beta}}$ genera una estimación \hat{y}_t por encima del valor observado (residuo negativo) con aquellos en que ocurre lo contrario (residuo positivo). Además, si realmente se pretendiese minimizar la suma de residuos, bastaría generar residuos de tamaño muy grande, pero negativos, lo cual no parece muy adecuado.

El *estimador de mínimos cuadrados* que introducimos en esta sección utiliza como criterio la minimización de la norma euclídea del vector \mathbf{u} , es decir, de la suma $SR = \sum_1^T \hat{u}_t^2$, que denominaremos en lo sucesivo *suma residual*, sin mencionar explícitamente su dependencia de los cuadrados de los residuos. La suma residual puede expresarse en notación matricial como $\hat{\mathbf{u}}'\hat{\mathbf{u}}$, siendo $\hat{\mathbf{u}}$ el vector $T \times 1$ de residuos. La suma residual es una función de: a) las observaciones muestrales y b) las estimaciones $\hat{\boldsymbol{\beta}}$, puesto que:

$$SR(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad [3.13]$$

En principio parece que se trataría de utilizar cada posible vector de estimaciones $\hat{\boldsymbol{\beta}}$ para evaluar la suma residual y elegir, finalmente, aquel vector

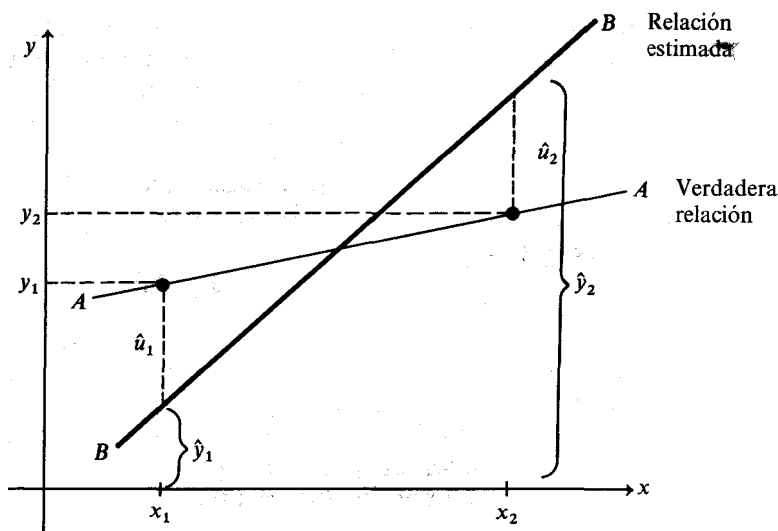


FIGURA 3.1.

que generase una suma residual mínima. Tal procedimiento no es factible, pues el número de vectores numéricos $\hat{\beta}$ a considerar es, en general, infinito. Afortunadamente, la teoría de la optimización matemática nos permite resolver nuestro problema:

$$\min_{\hat{\beta}} SR(\hat{\beta}) = \min_{\hat{\beta}} (y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}) \quad [3.14]$$

Utilizando los resultados acerca de derivadas matriciales que vimos en el Capítulo 1 se tiene, derivando respecto al vector $\hat{\beta}$ en [3.14], que:

$$\frac{\partial SR(\hat{\beta})}{\partial \hat{\beta}} = \frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} \quad [3.15]$$

La solución al problema de minimización de $SR(\hat{\beta})$ requiere, en primer lugar, que este vector gradiente sea igual a cero, es decir, que:

$$(X'X)\hat{\beta} = X'y \quad [3.16]$$

Además debe cumplirse que la matriz de segundas derivadas o matriz hessiana de $SR(\hat{\beta})$ sea definida positiva. Pero dicha matriz es:

$$\frac{\partial^2 SR(\hat{\beta})}{\partial \hat{\beta} \partial \hat{\beta}'} = X'X$$

que, como vimos en la Proposición 1.12, es siempre semidefinida positiva.

Puesto que $X'X$ es una matriz $k \times k$ y $X'y$ un vector $k \times 1$, la ecuación matricial [3.16] es, en realidad, un sistema de k ecuaciones lineales en los k coeficientes desconocidos $\beta_1, \beta_2, \dots, \beta_k$. Este sistema se denomina *sistema de ecuaciones normales* y tiene, generalmente, una única solución. Dicha solución al sistema de ecuaciones normales es el *estimador de mínimos cuadrados ordinarios* del vector β , que denotaremos en lo sucesivo por MCO y que puede escribirse:

$$\hat{\beta} = (X'X)^{-1}X'y \quad [3.17]$$

Es interesante obtener la versión desarrollada del sistema de ecuaciones normales:

$$\begin{pmatrix} \sum_1^T x_{1t}^2 & \sum_1^T x_{1t}x_{2t} & \cdots & \sum_1^T x_{1t}x_{kt} \\ \sum_1^T x_{2t}x_{1t} & \sum_1^T x_{2t}^2 & \cdots & \sum_1^T x_{2t}x_{kt} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_1^T x_{kt}x_{1t} & \sum_1^T x_{kt}x_{2t} & \cdots & \sum_1^T x_{kt}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum_1^T x_{1t}y_t \\ \sum_1^T x_{2t}y_t \\ \cdots \\ \sum_1^T x_{kt}y_t \end{pmatrix} \quad [3.18]$$

Cuando la primera variable explicativa es una constante, es decir, $x_{1t} = 1$ para todo t , la primera fila (y la primera columna) de la matriz $X'X$ se convierten en $\left(T, \sum_1^T x_{2t}, \dots, \sum_1^T x_{kt}\right)$ y la primera coordenada del vector $X'y$ pasa a ser $\sum_1^T y_t$.

La única dificultad para la obtención numérica del estimador MCO aparece cuando la matriz $X'X$ sea singular puesto que, por no ser invertible, no podemos utilizar la expresión [3.17]. En tal caso, no es que el sistema de ecuaciones normales [3.16] no tenga solución, sino que tiene infinitas soluciones, como veremos en el Capítulo 10. Ello ocurre en situaciones de *multicolinealidad*, es decir, cuando se vulnera nuestro supuesto 5 de la Sección 3.1. Si, por ejemplo, la variable x_1 es combinación lineal de las restantes, existen valores paramétricos $\lambda_2, \lambda_3, \dots, \lambda_k$, tales que $x_{1t} = \lambda_2 x_{2t} + \lambda_3 x_{3t} + \dots + \lambda_k x_{kt}$, de modo que los elementos (1, 1) y (1, 2) de la matriz $X'X$ resultan:

$$\begin{cases} \sum_1^T x_{1t}^2 = \lambda_2 \sum_1^T x_{1t}x_{2t} + \lambda_3 \sum_1^T x_{1t}x_{3t} + \cdots + \lambda_k \sum_1^T x_{1t}x_{kt} \\ \sum_1^T x_{1t}x_{2t} = \lambda_2 \sum_1^T x_{2t}^2 + \lambda_3 \sum_1^T x_{2t}x_{3t} + \cdots + \lambda_k \sum_1^T x_{2t}x_{kt} \end{cases}$$

y algo similar ocurre con los demás elementos de la fila 1 de la matriz $X'X$, que puede escribirse como combinación lineal de las demás filas, con coeficien-

tes $\lambda_2, \lambda_3, \dots, \lambda_k$. En consecuencia, el determinante de la matriz $\mathbf{X}'\mathbf{X}$ será igual a cero.

Ejemplo 3.3 (continúa). Utilizando los datos de las columnas 2 a 5 de la Tabla 3.1 tenemos:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 5 & \sum_1^5 x_{2t} \\ \sum_1^5 x_{2t} & \sum_1^5 x_{2t}^2 \end{pmatrix} = \begin{pmatrix} 5 & 200 \\ 200 & 8160 \end{pmatrix} \Rightarrow (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 10,2 & -0,25 \\ -0,25 & 0,00625 \end{pmatrix};$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 500 \\ 20248 \end{pmatrix}$$

por lo que la estimación MCO es:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 10,2 & -0,25 \\ -0,25 & 0,00625 \end{pmatrix} \begin{pmatrix} 500 \\ 20248 \end{pmatrix} = \begin{pmatrix} 38 \\ 1,55 \end{pmatrix}$$

teniéndose el modelo estimado $y_t = 38 + 1,55x_t + \hat{u}_t$, que genera unos residuos $\hat{u}_t = y_t - 38 - 1,55x_t$, cuyos valores numéricos, que el lector debe comprobar, se muestran en la columna 7 de la tabla.

Antes de analizar las propiedades del estimador MCO, conviene insistir en que dicho estimador está definido de modo único siempre y cuando la matriz producto $\mathbf{X}'\mathbf{X}$ sea invertible, como puede verse en [3.17], lo cual ocurre siempre que:

1. Las k columnas de la matriz \mathbf{X} son linealmente independientes, es decir, siempre que las k variables explicativas del modelo no sean linealmente dependientes entre sí.
2. Se disponga de al menos tantas observaciones como variables explicativas, es decir: $T \geq k$.

Esta última propiedad merece cierta discusión: cuando $T = k$, es decir, cuando se dispone exactamente de tantas observaciones muestrales como variables explicativas hay en el modelo (incluyendo la posible constante), se tiene que la matriz \mathbf{X} es de orden $k \times k$; por tanto, salvo multicolinealidad, dicha matriz es invertible, por lo que se tiene $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}$ y por tanto:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{-1}(\mathbf{X}')^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{-1}\mathbf{y}$$

El vector de residuos es $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}^{-1}\mathbf{y}) = \mathbf{y} - \mathbf{y} = \mathbf{0}_T$, por lo que podemos decir que el ajuste es perfecto, ya que *cada uno* de los residuos es igual a cero, y lo mismo ocurre con la suma residual, que toma su mínimo valor posible, es decir, cero.

Sin embargo, ello no debe interpretarse como una característica muy positiva de la estimación que hemos obtenido. Es cierto que el ajuste del modelo de regresión a una muestra con un número de observaciones $T = k$ es *perfecto*, en el sentido de que la suma residual es igual a cero, pero ello ocurre tan sólo porque disponemos de una información muestral muy reducida. De hecho, nuestras estimaciones son muy poco robustas y, por consiguiente, muy imprecisas: si obtenemos una observación muestral adicional y eliminamos una de las actuales, obtendremos de nuevo una suma residual mínima, igual a cero, pero las estimaciones $\hat{\beta}$ pueden diferir arbitrariamente de las precedentes.

En realidad, para lograr precisión en la estimación MCO es preciso disponer de un número de observaciones notablemente superior al de variables explicativas, es decir, $T \gg k$. A la diferencia $T - k$ se le conoce como *número de grados de libertad de la estimación*. Cuando $T = k$, no se dispone de grados de libertad; las k observaciones determinan exactamente el valor numérico del estimador MCO, lo que conduce a una suma residual igual a cero.

De este modo, por un lado es mejor disponer de un número alto de observaciones muestrales para obtener mayor precisión y robustez en las estimaciones; por otro lado, cada observación adicional generará un residuo más y, con ello, la suma residual tenderá generalmente a aumentar. La aspiración del investigador consiste en disponer de un elevado número de observaciones y un modelo econométrico que las represente suficientemente bien, de modo que ningún residuo, por sí solo, tenga una contribución importante a la suma residual.

3.3 PROPIEDADES DEL ESTIMADOR DE MINIMOS CUADRADOS ORDINARIOS

Caracterizamos en esta sección las propiedades del estimador MCO. Hemos de comenzar observando que el estimador MCO es un vector aleatorio, pues, dependiendo del vector de observaciones de la variable endógena y , depende también del vector de término de error u :

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u \quad [3.19]$$

Como tal vector aleatorio, cabe preguntarse acerca de su esperanza matemática y su matriz de covarianzas, interrogante que despejamos en las dos proposiciones siguientes:

Proposición 3.1. Si $E(u) = 0_T$, entonces el estimador MCO es insesgado, es decir, $E(\hat{\beta}) = \beta$.

Demostración. Tomando esperanzas en [3.19] se tiene:

$$E(\hat{\beta}) = E[\beta + (X'X)^{-1}X'u] = \beta + (X'X)^{-1}X'E(u) = \beta$$

donde hemos utilizado el hecho de que el vector β , si bien desconocido, es un vector de constantes, así como que las variables explicativas son deterministas.

Nótese que [3.19] puede utilizarse para definir el *error de estimación*:

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

que utilizamos en la prueba de la siguiente propiedad:

Proposición 3.2. Si $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_T$, la matriz de covarianzas del estimador MCO es igual a $\text{Var}(\hat{\beta}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$.

Demostración:

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma_u^2 \mathbf{I}_T)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

donde, además de [3.19], hemos utilizado el supuesto $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_T$. Entre otras cosas, estas dos propiedades implican que el estimador $\hat{\beta}_i$ de uno cualquiera de los coeficientes β_i tiene como esperanza matemática β_i , el verdadero valor del parámetro que se pretende estimar, y como varianzas $\sigma_u^2 a_{ii}$, donde a_{ii} es el elemento i -ésimo en la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Como debe ocurrir con toda matriz de covarianzas, notemos que $(\mathbf{X}'\mathbf{X})^{-1}$ es una matriz simétrica y definida positiva, puesto que su inversa, $\mathbf{X}'\mathbf{X}$, también lo es.

Ejemplo 3.3 (continúa). De acuerdo con los datos de la Tabla 3.1, hemos obtenido más arriba estimaciones MCO: $\hat{\beta}_1 = 38$; $\hat{\beta}_2 = 15,5$. La Proposición 3.1 nos dice que $E(\hat{\beta}_1) = \beta_1$ y $E(\hat{\beta}_2) = \beta_2$, lo cual, lamentablemente, no nos resulta muy operativo. Bastante más útil es la Proposición 3.2 que nos permite calcular para el estimador MCO del vector (β_1, β_2) la matriz de covarianzas:

$$\text{Var} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2 \begin{pmatrix} 10,2 & -0,25 \\ -0,25 & 0,00625 \end{pmatrix}$$

de modo que $\text{Var}(\hat{\beta}_1) = 10,2\sigma_u^2$; $\text{Var}(\hat{\beta}_2) = 0,00625\sigma_u^2$ y $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -0,25\sigma_u^2$. Sin embargo, no podremos asociar valores numéricos a tales varianzas y covarianzas hasta la Sección 3.4, en que aprenderemos a estimar el parámetro σ_u^2 .

Examinemos el significado de las dos propiedades que acabamos de probar: la primera de ellas garantiza que, bajo nuestros supuestos, el estimador MCO es insesgado. Esta es una propiedad interesante si creemos que la ausencia de sesgo es una condición indispensable. Sin embargo, hay que

matizar que el insesgo del estimador MCO no garantiza nada acerca de lo próximo o lejano que el valor numérico de este estimador se halla con respecto a los verdaderos valores de los coeficientes.

Que el estimador MCO sea insesgado quiere decir lo siguiente: supongamos que se dispone de varias muestras diferentes de datos acerca de las variables y, x_1, x_2, \dots, x_k y que utilizamos cada una de dichas muestras para calcular la estimación MCO correspondiente. Si el número de muestras es elevado, entonces el promedio de las estimaciones obtenidas con cada una de ellas será, aproximadamente, el verdadero valor de los coeficientes β . Sin embargo, en *Economía disponemos siempre de una sola muestra*, lo que implica que tendremos una única realización del vector aleatorio $\hat{\beta}$. Con una sola muestra, difícilmente podemos asegurar nada acerca de la distancia entre esa realización de $\hat{\beta}$ y su esperanza matemática poblacional.

Por esta razón, el conocimiento de la matriz de covarianzas es fundamental. Recordemos (Sección 2.3) que, cuando se trabaja con una variable aleatoria, su desviación típica proporciona, por su definición, un promedio de la distancia entre una observación cualquiera de dicha variable y su esperanza matemática. Cuando se trabaja con un vector aleatorio como $\hat{\beta}$, estas ideas son válidas de un modo similar.

Como ya hemos dicho, contamos con una única realización del vector $\hat{\beta}$ obtenida a partir de la única muestra disponible. En tales condiciones, que $\hat{\beta}$ sea insesgado no es una propiedad excesivamente interesante, y debemos tratar de minimizar la distancia entre la única realización muestral (es decir, el valor tomado por el estimador MCO) y su esperanza matemática que, como ya sabemos, es el vector de verdaderos valores de los parámetros β .

Pero esa distancia viene medida por la matriz de covarianzas de $\hat{\beta}$, que nos interesará que sea tan pequeña como sea posible. En nuestro caso, ello conduce a que tanto el valor del parámetro σ_u^2 (que hemos de estimar) como el tamaño de la matriz $(X'X)^{-1}$ (medido de alguna forma) sean pequeños⁽²⁾.

Por el contrario, cuanto más próxima esté la matriz $X'X$ a ser singular (es decir, más cercano a cero sea su determinante), mayor será el determinante de $(X'X)^{-1}$ y, con ello, mayor será la varianza del estimador MCO o, lo que es lo mismo, menor será su precisión.

Por todo ello, se han propuesto en ocasiones estimadores del modelo lineal que, a diferencia del estimador MCO, son sesgados, pero pudieran tener menor Error Cuadrático Medio (ECM) que éste. Como referencia, notemos que, por ser insesgado, el ECM del estimador MCO coincide con su matriz de covarianzas:

$$ECM(\hat{\beta}) = (\beta - E(\hat{\beta}))(\beta - E(\hat{\beta}))' + \text{Var}(\hat{\beta}) = \text{Var}(\hat{\beta})$$

⁽²⁾ Como vimos en el Capítulo 1, el tamaño de la matriz $X'X$, y, por tanto, de su inversa, está relacionado con la magnitud relativa de sus valores propios. Por ejemplo, el determinante de la matriz es igual al producto de sus valores propios, y una matriz cuadrada es singular si tiene algún valor propio igual a cero.

Proposición 3.3. Cada una de las variables explicativas es ortogonal al vector de residuos mínimo-cuadráticos, es decir:

$$\sum_1^T x_{it} \hat{u}_t = 0, \quad \text{para } i = 1, 2, \dots, k$$

o lo que es lo mismo:

$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}_k$$

Demostración:

$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{y} = \mathbf{0}_k$$

Proposición 3.4. Como consecuencia, si hay término independiente en la regresión, entonces la suma de los residuos mínimo-cuadráticos es cero:

$$x_{1t} = 1 \forall t \Rightarrow \sum_1^T \hat{u}_t = 0$$

Demostración. En efecto, si una de las variables es constante, entonces basta aplicar la propiedad anterior a esa variable en particular para obtener este nuevo resultado.

Observación. Las columnas 7 y 8 de la Tabla 3.1 ilustran las dos proposiciones que acabamos de probar.

El estimador MCO es una función lineal del vector de observaciones de la variable aleatoria \mathbf{y} . Un estimador lineal genérico se escribirá:

$$\hat{\boldsymbol{\beta}}^* = \mathbf{f}(\mathbf{X}) + \mathbf{g}(\mathbf{X})\mathbf{y}$$

El estimador MCO utiliza como funciones \mathbf{f} y \mathbf{g} en la expresión anterior: $\mathbf{f} = \mathbf{0}$ y $\mathbf{g} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Como vamos a ver en las dos proposiciones siguientes, el estimador MCO es el estimador lineal e insesgado del vector de coeficientes $\boldsymbol{\beta}$ con menor varianza, es decir, con mayor precisión.

Cualquier otro estimador de $\boldsymbol{\beta}$ que dependa linealmente del vector \mathbf{y} tendrá matriz de covarianzas mayor que $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, es decir, la diferencia entre ambas, en este orden, será definida positiva.

Comenzamos haciendo notar que para estimar una combinación lineal cualquiera de los coeficientes $\boldsymbol{\beta}$ la combinación lineal análoga de las estimaciones MCO es el estimador de menor varianza.

Proposición 3.5. La combinación lineal $\mathbf{c}'\hat{\boldsymbol{\beta}}$ es el estimador lineal insesgado de mínima varianza de la combinación lineal de parámetros $\mathbf{c}'\boldsymbol{\beta}$.

Dicho de otra forma, si queremos estimar la suma de dos parámetros $\beta_1 + \beta_2$, lo mejor que podemos hacer, si nos queremos restringir a estimadores

lineales, es obtener los estimadores MCO de dichos parámetros y utilizar la suma de sus estimaciones como estimación de $\beta_1 + \beta_2$. Algo análogo se debe hacer con cualquier otra combinación lineal. Por ejemplo, el mejor estimador lineal de $\beta_1 - 5\beta_2$ es precisamente $\hat{\beta}_1 - 5\hat{\beta}_2$.

Esta propiedad tiene una implicación de interés: puesto que un coeficiente β_i es una combinación lineal del tipo $c'\beta$, con un vector c que tiene todas sus componentes cero, excepto un 1 en la posición i -ésima, entonces se tiene como corolario que la estimación lineal insesgada de un coeficiente β_i es precisamente la coordenada correspondiente del estimador MCO.

Esta observación garantiza que los elementos de la diagonal de la matriz de covarianzas $\hat{\beta}$ no serán superiores, uno a uno, a los elementos correspondientes en la matriz de covarianzas de cualquier otro estimador lineal e insesgado. Pero este resultado no implica que la diferencia entre ambas sea semidefinida negativa. De hecho, la diferencia podría no ser ni semidefinida positiva ni negativa.

El resultado que demostramos a continuación, conocido como teorema de Gauss-Markov, es aún más fuerte y afirma la eficiencia del estimador MCO del vector β dentro de los estimadores lineales. Como se puede ver en un problema propuesto al final del capítulo, la Proposición 3.5 no es sino un caso particular del teorema que ahora probamos.

Proposición 3.6. Teorema de Gauss-Markov: El estimador MCO es el estimador lineal insesgado óptimo, en el sentido de que cualquier otro estimador lineal e insesgado tiene una matriz de covarianzas «mayor» que la del estimador MCO.

Demostración. Sea $\tilde{\beta} = \tilde{A}y$ un estimador lineal de β , donde \tilde{A} es una matriz $k \times T$. Denotemos por A (matriz $k \times T$) la diferencia $A = \tilde{A} - (X'X)^{-1}X'$, de modo que:

$$\tilde{\beta} = [A + (X'X)^{-1}X']y = [A + (X'X)^{-1}X'](X\beta + u) = AX\beta + \beta + [A + (X'X)^{-1}X']u$$

y, por tanto, $E(\tilde{\beta}) = AX\beta + \beta$. El estimador $\tilde{\beta}$ será insesgado sólo si la matriz A es tal que $AX = 0_{k \times k}$. Con esta condición, el estimador $\tilde{\beta}$ resulta:

$$\tilde{\beta} = \beta + [A + (X'X)^{-1}X']u$$

y su matriz de covarianzas será:

$$\begin{aligned} \text{Cov}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = E\{([A + (X'X)^{-1}X']u)([A + (X'X)^{-1}X']u)'\} = \\ &= \sigma_u^2 AA' + \sigma_u^2 (X'X)^{-1} \end{aligned}$$

donde se ha utilizado la condición de ausencia de sesgo $AX = 0_{k \times k}$. Como la matriz AA' es semidefinida positiva, se concluye que la diferencia entre las matrices de covarianzas de $\tilde{\beta}$ y $\hat{\beta}$ es una matriz semidefinida positiva, por lo que la primera es igual, si no mayor, que la segunda.

Proposición 3.7. Una expresión que resulta útil para la suma residual (suma de los cuadrados de los residuos mínimo-cuadráticos) es:

$$SR = \hat{u}'\hat{u} = y'y - \hat{\beta}'X'y \quad [3.20]$$

Demostración:

$$\begin{aligned} \hat{u}'\hat{u} &= (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X(X'X)^{-1}X'y = y'y - \hat{\beta}'X'y \end{aligned}$$

Proposición 3.8. Otra expresión alternativa para la suma residual es:

$$SR = \hat{u}'\hat{u} = y'y - \hat{y}'\hat{y} = \sum_1^T y_t^2 - \sum_1^T \hat{y}_t^2 \quad [3.21]$$

es decir, la suma residual es la diferencia entre la suma de cuadrados de las observaciones y la suma de cuadrados de los valores de y_t implicados por el modelo, \hat{y}_t .

Demostración. Basta notar que $\hat{y}'\hat{y} = (\hat{\beta}'X')(X\hat{\beta}) = \hat{\beta}'X'X(X'X)^{-1}X'y = \hat{\beta}'X'y$ y sustituir en la expresión [3.20].

Ejemplo 3.3 (continúa). A partir de las estimaciones obtenidas con los datos de la Tabla 3.1 tenemos:

TABLA 3.2

t	y_t	$y_t - \bar{y}$	$(y_t - \bar{y})^2$	\hat{y}_t	y_t^2	\hat{y}_t^2	\hat{u}_t	\hat{u}_t^2
1	84	-16	256	87,6	7056	7673,76	-3,6	12,96
2	108	8	64	100,0	11664	10000,0	8,0	64,0
3	92	-8	64	93,8	8464	8798,44	-1,8	3,24
4	110	10	100	106,2	12100	11278,44	3,8	14,44
5	106	6	36	112,4	11236	12633,76	-6,4	40,96
Suma	500	0	520	500	50520	50384,4	0	135,6

Las dos últimas columnas de esta tabla muestran los residuos del modelo, así como la suma residual: $SR = 135,6$. La suma de las columnas 4, 6 y 7 son, respectivamente, $ST = 520$, $y'y = 50520$ e $\hat{y}'\hat{y} = 50384,4$, por lo que se satisface la Proposición 3.8.

También tenemos:

$$\hat{\beta}'\mathbf{X}'\mathbf{y} = (38, 1,55) \cdot \begin{pmatrix} 500 \\ 20248 \end{pmatrix} = 50384,4$$

de modo que se satisface la Proposición 3.7.

Proposición 3.9. El vector de residuos mínimo-cuadráticos es una transformación lineal del vector término de error.

$$\text{En efecto: } \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}$$

donde \mathbf{M} es la matriz que introducimos en el Capítulo 1: $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, singular, simétrica e idempotente, y donde la última igualdad viene de $\mathbf{M}\mathbf{X} = \mathbf{0}_{T \times k}$.

De esta propiedad se deduce inmediatamente otra expresión para la suma residual:

$$\text{SR} = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{M}\mathbf{y} = (\mathbf{X}\beta + \mathbf{u})'\mathbf{M}(\mathbf{X}\beta + \mathbf{u}) = \mathbf{u}'\mathbf{M}\mathbf{u}$$

ya que $\mathbf{M}\mathbf{X} = \mathbf{0}_{T \times k}$.

Proposición 3.10. El vector de residuos mínimo-cuadráticos tiene esperanza cero y matriz de covarianzas $\sigma_u^2\mathbf{M}$.

Demostración. Es consecuencia inmediata de la Proposición 3.9.

Este último resultado es interesante porque, generalmente, la matriz \mathbf{M} no estará sistemáticamente rellena de ceros. Es decir, que aunque el vector \mathbf{u} tenga matriz de covarianzas escalar, la matriz de covarianzas del vector $\hat{\mathbf{u}}$ es mucho más compleja.

Introduzcamos ahora las siguientes expresiones:

$$\text{Suma total} = \text{ST} = \sum_1^T (y_t - \bar{y})^2$$

$$\text{Suma explicada} = \text{SE} = \sum_1^T (\hat{y}_t - \bar{y})^2$$

$$\text{Suma residual} = \text{SR} = \sum_1^T (y_t - \hat{y}_t)^2$$

La *suma total* no es sino la varianza muestral de la variable endógena multiplicada por T , por lo que es una medida del tamaño de las fluctuaciones experimentadas por dicha variable alrededor de su valor medio. El objeto fundamental de todo modelo econométrico es tratar de explicar dichas fluctuaciones y, por consiguiente, la suma total es la cantidad que se pretende explicar. La *suma explicada* es el grado de fluctuación de la variable \hat{y}_t alrededor del promedio de y , donde \hat{y}_t es la variable que el modelo genera como y_t . Por tanto, la suma explicada es el nivel de fluctuación de la variable y_t que *el modelo es capaz de explicar*. La *suma residual*, introducida previa-

mente, es un indicador del nivel de error del modelo en su intento de explicar la evolución temporal de la variable y_t .

Proposición 3.11. Si entre las variables explicativas hay un término constante, entonces se tiene:

$$\text{Suma total} = \text{Suma explicada} + \text{Suma residual}$$

Esta proposición implica que la diferencia entre la cantidad a explicar (suma total) y la cantidad explicada (suma explicada) es precisamente la suma residual, siempre y cuando el modelo contenga un término constante. El resultado es interesante puesto que es conforme a nuestra intuición acerca de la utilización de un modelo econométrico: se pretende explicar la variable y_t ; el modelo es capaz de generar la variable \hat{y}_t , que dictamina la parte de y_t que el modelo explica, mientras que el vector de error \hat{u}_t determina el tamaño de la componente de y_t no explicada por el modelo. Pues bien, cuando hay un término independiente en el modelo econométrico, entonces las sumas explicada y no explicada (o residual) se agregan hasta ser exactamente igual a la suma a explicar (suma total).

Demostración. Restando a ambos miembros de [3.21] la cantidad $T\bar{y}^2$ y reagrupando términos se tiene:

$$\mathbf{y}'\mathbf{y} - T\bar{y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - T\bar{y}^2 + \hat{\mathbf{u}}'\hat{\mathbf{u}} \quad [3.22]$$

En primer lugar notemos que:

$$ST = \sum_1^T (y_t - \bar{y})^2 = \sum_1^T y_t^2 - 2\bar{y}\sum_1^T y_t + T\bar{y}^2 = \mathbf{y}'\mathbf{y} - 2\bar{y}(T\bar{y}) + T\bar{y}^2 = \mathbf{y}'\mathbf{y} - T\bar{y}^2$$

de modo que el miembro de la izquierda en [3.22] es precisamente la suma total. Ahora bien, la suma explicada puede escribirse como:

$$SE = \sum_1^T (\hat{y}_t - \bar{y})^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - 2\bar{y}\sum_1^T \hat{y}_t + T\bar{y}^2 \quad [3.23]$$

Por otra parte, puesto que $y_t = \hat{y}_t + \hat{u}_t$, se tiene:

$$\bar{y}\sum_1^T \hat{y}_t = \bar{y}\sum_1^T y_t - \bar{y}\sum_1^T \hat{u}_t = \bar{y}\sum_1^T y_t = T\bar{y}^2$$

donde ya se ha utilizado el hecho de que cuando uno de los regresores es constante, entonces $\sum_1^T \hat{u}_t = 0$. Sustituyendo en [3.23] se tiene $SE = \hat{\mathbf{y}}'\hat{\mathbf{y}} - T\bar{y}^2$ y sustituyendo esta expresión en [3.22] se tiene finalmente el resultado: $ST = SE + SR$.

Proposición 3.12. Cuando existe un término independiente, la suma explicada puede calcularse mediante las expresiones:

$$SE = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - T\bar{y}^2 = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - T\bar{y}^2$$

Demostración:

$$\begin{aligned}
 SR &= \hat{u}'\hat{u} = \mathbf{u}'\mathbf{M}\mathbf{u} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \\
 &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - T\bar{y}^2 - (\hat{\beta}'\mathbf{X}'\mathbf{y} - T\bar{y}^2) = \\
 &= ST - (\hat{\beta}'\mathbf{X}'\mathbf{y} - T\bar{y}^2)
 \end{aligned}$$

de donde se deduce la primera expresión anunciada para la suma explicada. La segunda surge de:

$$\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} = \hat{\beta}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\beta}'\mathbf{X}'\mathbf{y}$$

que es válida incluso cuando no hay término independiente en el modelo.

Ejemplo 3.3 (continúa). Volviendo a las Tablas 3.1 y 3.2 hemos observado que $ST = 520$; $SE = 384,4$; $SR = 135,6$, por lo que se satisface la Proposición 3.11. También vimos que $\hat{\beta}'\mathbf{X}'\mathbf{y} = 50384,4$, y como $T = 5$ e $\bar{y} = 100$, se satisface la Proposición 3.12. También se cumple la segunda parte de dicha proposición, puesto que:

$$\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - T\bar{y}^2 = (38; 1,55) \begin{pmatrix} 5 & 200 \\ 200 & 8160 \end{pmatrix} \begin{pmatrix} 38 \\ 1,55 \end{pmatrix} - 50.000 = 384,4$$

En un modelo lineal se define el *coeficiente de determinación* como la cantidad:

$$R^2 = 1 - \frac{SR}{ST}$$

mientras que su raíz cuadrada positiva, cuando existe, se denomina *coeficiente de correlación* entre y e \hat{y} .

Proposición 3.13

- El coeficiente de determinación es siempre menor o igual que 1.
- Si una de las variables explicativas es constante, entonces se tiene $R^2 \geq 0$.

Demostración

a) Puesto que las dos sumas SR y ST son, por definición, dos números positivos, entonces su cociente también lo es y, por tanto, $R^2 \leq 1$.

b) En estas circunstancias hemos probado que $ST = SE + SR$, y siendo todas estas cantidades positivas, se tiene que, necesariamente, $SR \leq ST$, por lo que $\frac{SR}{ST} \leq 1$, con lo que se tiene $R^2 \geq 0$.

Proposición 3.14. *En un modelo con término independiente, el coeficiente de determinación puede calcularse mediante las expresiones:*

$$R^2 = \frac{\hat{\beta}' X' y - T\bar{y}^2}{y'y - T\bar{y}^2} = \frac{\hat{\beta}' X' X \hat{\beta} - T\bar{y}^2}{y'y - T\bar{y}^2}$$

Demostración. Es consecuencia inmediata de las Propiedades 3.11 y 3.12, ya que en el caso en que exista un término independiente, una definición alternativa del estadístico R^2 es $R^2 = \frac{SE}{ST}$.

Estas expresiones no son, sin embargo, válidas en general, por lo que no es correcto utilizar el cociente anterior: $R^2 = SE/ST$, como definición del coeficiente de determinación. Del mismo modo, tanto la no negatividad de R^2 como la descomposición de la suma total sólo son válidas en tanto en cuanto haya un término independiente en el modelo, y no en otro caso. En consecuencia, cuando no hay un término independiente en el modelo, entonces el coeficiente de determinación puede tomar cualquier valor, siempre menor o igual a 1, pero incluyendo en su rango todos los reales negativos (a pesar del exponente 2 con el que se denota). Dicho exponente es puramente una convención notacional, y, como acabamos de ver, no indica en absoluto que el valor del estadístico R^2 haya de ser necesariamente positivo. Por supuesto que cuando el coeficiente de determinación es negativo, entonces no cabe hablar del coeficiente de correlación lineal, definido como $\sqrt{R^2}$.

Por otra parte, incluso en casos en que R^2 resulte negativo, aún será posible utilizarlo para comparar el grado de poder explicativo de dos modelos. Supongamos que tenemos dos modelos distintos, ambos sin término independiente, los dos con el mismo número de variables explicativas. Ya que ambos modelos tratan de explicar la misma variable endógena, comparten la misma suma total. La única diferencia entre ellos con respecto a la definición del estadístico R^2 es la SR, y claramente hemos de preferir aquel modelo que tenga la menor suma residual o, lo que es lo mismo, aquel con mayor R^2 , no importa que este estadístico sea positivo o negativo. Por tanto, pueden hacerse criterios comparativos con el R^2 , incluso cuando no es positivo.

Un problema más delicado aparece cuando, además, los dos modelos a comparar tienen un número distinto de variables explicativas. Puede probarse que cuando se añade una variable a un modelo, entonces la suma residual siempre disminuye. Por tanto, si uno de los dos modelos contiene las mismas variables que el otro y alguna más (en cuyo caso los dos modelos se dicen *anidados*), entonces este modelo amplio sería siempre el preferido de acuerdo con el criterio del mayor R^2 . Se hace preciso en esta situación utilizar un concepto similar al del coeficiente de determinación (R^2), pero teniendo en cuenta el número de variables explicativas que un modelo utiliza.

Cuando el número de variables es diferente y además los modelos no son anidados, entonces la elección de un modelo «preferido» es aún más delicada. En general, para comparar dos modelos no anidados, incluso si ambos tienen

el mismo número de variables explicativas, conviene tener cierta precaución y es preciso utilizar criterios cuya discusión está fuera del ámbito de este texto. Una situación imposible, en términos del uso del R^2 , se presenta cuando de los dos modelos a comparar uno tiene término independiente, pero no el otro, puesto que en tal caso los coeficientes de determinación respectivos están calculados sobre escalas diferentes, $[0, 1]$ en un caso, y $[-\infty, 1]$ en el otro.

En la literatura econométrica se ha sugerido tradicionalmente la utilización del estadístico conocido como R^2 corregido (o R -cuadrado corregido), que se define como:

$$\bar{R}^2 = 1 - \left(\frac{T-1}{T-k} \right) (1 - R^2) \quad [3.24]$$

El interés de este estadístico reside en que, cuando el número de variables explicativas k aumenta, la fracción $\frac{T-1}{T-k}$ también aumenta, mientras que $1 - R^2$ disminuye, ya que el coeficiente de determinación R^2 aumenta. Como en la definición del R -cuadrado corregido aparece el producto de estos dos factores, la idea es que ambos efectos, el creciente y el decreciente, se compensen aproximadamente, por lo que este estadístico sea una medida de la bondad de ajuste de un modelo econométrico con la propiedad de ser neutral frente a la introducción de variables adicionales. Este estadístico está concebido, por tanto, para la comparación de modelos anidados.

También es importante observar que, en principio, no hay razón por la que haya de usarse el cociente $\frac{T-1}{T-k}$ en la definición del R -cuadrado corregido, puesto que cualquier otra función creciente del número de variables k sería igualmente útil. Es cierto que el cociente $\frac{T-1}{T-k}$ tiene la propiedad de estar acotado inferiormente por 1, lo que hace que el rango de valores de \bar{R}^2 sea $[-\infty, 1]$, igual al rango de R^2 , pero existen otras funciones crecientes con esta propiedad.

Proposición 3.15. Si se dispone de menos observaciones que parámetros se quiere estimar (es decir, si $T < k$), entonces el estimador MCO no está unívocamente definido.

Demostración. Cuando $T < k$, entonces el rango de la matriz \mathbf{X} será necesariamente menor que k , y lo mismo ocurrirá a la matriz $\mathbf{X}'\mathbf{X}$ (véase Teorema 1.1), que será, por tanto, singular. Como consecuencia, el sistema de ecuaciones normales tiene infinitas soluciones.

Es importante resaltar que, para ninguna de las propiedades que venimos discutiendo, es preciso hacer ningún supuesto acerca del carácter de la distribución del término de error u_t . Podría seguir una distribución Normal o binomial, o incluso diferentes distribuciones en distintos instantes de tiempo, y los resultados previos aún serían válidos, siempre que $E(\mathbf{u}) = \mathbf{0}_T$ y $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_T$.

Además de las propiedades anteriores, se puede asegurar que cuando el término de error sigue una distribución Normal, entonces el estimador MCO sigue también una distribución Normal:

Proposición 3.16. Si $\mathbf{u} \sim N_T(\mathbf{0}_T, \sigma_u^2 \mathbf{I}_T)$, entonces $\hat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1})$.

Demostración. El estimador $\hat{\boldsymbol{\beta}}$ es una transformación lineal del vector aleatorio \mathbf{u} , ya que $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}\mathbf{u}$, donde $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, por lo que si \mathbf{u} sigue una distribución Normal, entonces $\hat{\boldsymbol{\beta}}$ también seguirá una distribución Normal, cuya esperanza y varianza se han obtenido en proposiciones anteriores.

Nótese, sin embargo, la diferente dimensionalidad de las dos distribuciones normales. Por otra parte, como sabemos por la Proposición 3.9 que el vector de residuos MCO es una transformación lineal del vector de error \mathbf{u} , también se tiene:

$$\text{Si } \mathbf{u} \sim N_T(\mathbf{0}_T, \sigma_u^2 \mathbf{I}_T) \quad \text{entonces} \quad \hat{\mathbf{u}} \sim N_T(\mathbf{0}_T, \sigma_u^2 \mathbf{M})$$

3.4. ESTIMACION DE σ_u^2

La matriz de covarianzas del estimador MCO es importante por dos razones:

1. Por un lado ya hemos visto que cuando sólo disponemos de una muestra para llevar a cabo la estimación del vector $\boldsymbol{\beta}$ entonces es crucial conocer la matriz de covarianzas del estimador para poder juzgar la exactitud con que la estimación obtenida se aproxima a su esperanza matemática $\boldsymbol{\beta}$.

2. Por otra parte, la matriz de covarianzas es precisa para conocer las varianzas y covarianzas de cada elemento del vector $\hat{\boldsymbol{\beta}}$, de modo que podamos hacer contrastes de hipótesis acerca de valores individuales, o de varios coeficientes, o de combinaciones lineales de ellos, del modo que veremos en el capítulo siguiente.

Pero al ser el parámetro σ_u^2 desconocido, la matriz de covarianzas de $\hat{\boldsymbol{\beta}}$ también es desconocida. Sólo podemos aspirar a estimar dicha matriz, estimando previamente el parámetro σ_u^2 y multiplicando tal estimación por la matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

En primer lugar, recordemos que el vector de residuos $\hat{\mathbf{u}}$ puede escribirse como una transformación lineal del vector de errores \mathbf{u} :

$$\hat{\mathbf{u}} = \mathbf{M}\mathbf{u}$$

donde $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, matriz $T \times T$, es idempotente y satisface $\mathbf{M}\mathbf{X} = \mathbf{0}_{T \times k}$. Por tanto, podemos calcular el valor esperado de la suma residual $E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'\mathbf{M}\mathbf{u}) =$ ya que $\mathbf{u}'\mathbf{M}\mathbf{u}$ es un escalar (matriz 1×1) $= E[\text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})] =$ (por las distintas propiedades del operador traza vistas en el Capítulo 1) $= E[\text{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')] = \text{tr}[E(\mathbf{M}\mathbf{u}\mathbf{u}')] = \text{tr}[\mathbf{M}E(\mathbf{u}\mathbf{u}')] = \text{tr}(\mathbf{M}\sigma_u^2 \mathbf{I}_T) = \sigma_u^2 \text{tr } \mathbf{M} =$ $= \sigma_u^2 [\text{tr } \mathbf{I}_T - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')] = \sigma_u^2 (T - k)$.

La última igualdad viene de $\text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \text{tr}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \text{tr} \mathbf{I}_k = k$. De este resultado se deduce que el cociente $\hat{\sigma}_u^2 = \frac{\text{SR}}{T-k} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}$ es un estimador insesgado del parámetro σ_u^2 , puesto que se tiene:

$$E(\hat{\sigma}_u^2) = E\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}\right) = \left(\frac{1}{T-k}\right) E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = \left(\frac{1}{T-k}\right) (T-k)\sigma_u^2 = \sigma_u^2$$

Este cociente se conoce como *estimador de mínimos cuadrados de la varianza del término de error* σ_u^2 . Nótese, sin embargo, que el estimador no proviene de ningún criterio de minimización sino sólo del interés de obtener un estimador de σ_u^2 que sea insesgado, juntamente con la utilización de los residuos mínimos-cuadráticos.

Por último, bajo el supuesto de Normalidad del término de error $\mathbf{u} \sim N(\mathbf{0}_T, \sigma_u^2 \mathbf{I}_T)$, luego $\mathbf{u}/\sigma_u \sim N(\mathbf{0}_T, \mathbf{I}_T)$, y los resultados del epígrafe 2.8.6 garantizan que $\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2} = \frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{\sigma_u^2}$ se distribuye como una chi-cuadrado con $T-k$ grados de libertad, pues ese es el rango de la matriz \mathbf{M} , que coincide con su traza, por ser \mathbf{M} idempotente.

Ello significa que $(T-k) \frac{\hat{\sigma}_u^2}{\sigma_u^2}$ se distribuye del mismo modo, lo que puede utilizarse para obtener intervalos de confianza para el parámetro σ_u^2 .

Ejemplo 3.3 (continúa). La última columna de la Tabla 3.1 muestra que la suma residual del modelo es 135,6. Como el número de grados de libertad es $T-k = 5-2 = 3$, la estimación de la varianza de la perturbación es $\hat{\sigma}_u^2 = 45,2$. Con ello podemos finalmente evaluar las varianzas y covarianzas de las estimaciones MCO:

$$\text{Var} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} = (45,2) \begin{pmatrix} 10,2 & -0,25 \\ -0,25 & 0,00625 \end{pmatrix}$$

de modo que $\widehat{\text{Var}}(\hat{\beta}_1) = 461,0$; $\widehat{\text{Var}}(\hat{\beta}_2) = 0,283$ y $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -11,3$.

Escribiendo las desviaciones típicas estimadas debajo de cada coeficiente tenemos el modelo estimado:

$$y_t = \underset{(21,5)}{38} + \underset{(0,53)}{1,55} x_t + \hat{u}_t$$

El coeficiente de determinación del modelo es:

$$R^2 = 1 - \frac{\text{SR}}{\text{ST}} = 1 - \frac{135,6}{520} = 0,74$$

mientras que el coeficiente de determinación corregido es:

$$\bar{R}^2 = 1 - \frac{T-1}{T-k} (1 - R^2) = 1 - \frac{4}{3} (0,26) = 0,65$$

Ejemplo 3.4. Con objeto de estimar una función de producción agregada para la economía española, utilizamos datos anuales de stock de capital, empleo y PIB, este último en pesetas constantes de 1980, tomados de Molinas, Sebastián y Zabalza (1991) para el período 1964-1988, como aparecen en la Tabla 3.3. Transformando las variables en logaritmos estimamos el modelo⁽³⁾:

$$\ln Y_t = 0,329 \cdot 10^{-2} + 0,322 \ln L_t + 0,630 \ln K_t$$

(1,053) (0,106) (0,013)

$$T = 25; \quad \bar{R}^2 = 0,991; \quad SR = 0,0136635; \quad \hat{\sigma}_u = 0,025;$$

$$DW = 0,365; \quad Q(12) = 36,7$$

que sugiere la posible existencia de rendimientos constantes a escala ya que si interpretamos el modelo estimado como el logaritmo de una función de producción Cobb-Douglas: $Y_t = AK_t^\alpha L_t^{1-\alpha}$, la suma de las elasticidades estimadas en ambos inputs es próxima a la unidad, si bien con una participación mayor del capital (próxima a 2/3) que del factor trabajo (en torno a 1/3).

La matriz de covarianzas de los coeficientes estimados fue:

$$\Sigma = \begin{pmatrix} 1,1095 & -0,1108 & -0,0067 \\ -0,1108 & 0,0112 & 0,0005 \\ -0,0067 & 0,0005 & 0,0002 \end{pmatrix}$$

y las raíces cuadradas de los elementos de la diagonal, es decir, las desviaciones típicas de los coeficientes estimados, se muestran entre paréntesis debajo de dichos coeficientes.

3.5. CONTRASTE DE NORMALIDAD

Puesto que algunas de las propiedades del estimador MCO que hemos visto dependen del supuesto de Normalidad de término de error del modelo, tiene interés disponer de un contraste de tal hipótesis. El contraste más habitual, propuesto por Bera y Jarque (1981), se basa en los coeficientes de asimetría

⁽³⁾ El significado de los estadísticos denotados por DW y Q será analizado en capítulos posteriores.

TABLA 3.3

Año	Stock de capital	Empleo	PIB (pesetas constantes) de 1980
64: 1	11.302,5	11.703,0	7.527,4
65: 1	12.578,4	11.990,9	8.004,2
66: 1	14.024,7	12.123,4	8.568,9
67: 1	15.471,3	12.198,2	8.939,1
68: 1	17.036,7	12.256,5	9.544,7
69: 1	18.814,3	12.333,6	10.398,0
70: 1	20.536,9	12.330,7	10.822,3
71: 1	22.061,6	12.427,1	11.318,0
72: 1	23.888,2	12.650,4	12.227,1
73: 1	26.010,6	12.875,4	13.166,9
74: 1	28.214,8	13.041,7	13.866,5
75: 1	30.094,6	12.823,0	13.940,9
76: 1	31.819,3	12.587,4	14.397,2
77: 1	33.396,0	12.581,8	14.829,2
78: 1	34.774,5	12.373,4	15.044,0
79: 1	35.909,7	12.173,9	15.023,1
80: 1	36.991,9	11.811,5	15.209,1
81: 1	37.892,5	11.471,8	15.171,3
82: 1	38.755,6	11.358,3	15.355,9
83: 1	39.488,6	11.279,3	15.633,2
84: 1	39.966,1	10.954,9	15.914,5
85: 1	40.607,3	10.854,9	16.282,8
86: 1	41.504,4	11.111,2	16.816,4
87: 1	42.851,5	11.452,2	17.748,7
88: 1	44.656,5	11.780,6	18.676,6

y curtosis que introdujimos en el Capítulo 2, aplicados a los residuos del modelo. Estos autores probaron que la distribución del estadístico

$$BJ = T \cdot \left(\frac{(\text{asimetría})^2}{6} + \frac{(\text{curtosis} - 3)^2}{24} \right)$$

es, para muestras grandes, una chi-cuadrado con 2 grados de libertad.

La distribución Normal, por ser simétrica, tiene coeficiente de asimetría igual a cero; además, su coeficiente de curtosis es igual a 3, por lo que el valor del estadístico BJ sería cero. Si el término de error del modelo es Normal, entonces, en la medida en que los residuos pueden interpretarse como una estimación del término de error, tendrán un valor pequeño del estadístico BJ.

Si el valor del estadístico BJ calculado para los residuos de un modelo estimado exceden del valor de las tablas de una distribución χ_2^2 , debe rechazarse la hipótesis nula de Normalidad, manteniéndose si el valor del estadístico BJ es inferior al de las tablas.

3.6. EL ESTIMADOR DE MÁXIMA VEROSIMILITUD

Hasta ahora hemos adoptado el criterio de estimación, consistente en escoger los valores de los parámetros β y σ_u^2 , de modo que se obtenga la menor suma de cuadrados de residuos posible. Sin embargo, ya consideramos en el Capítulo 2 otro procedimiento de estimación, el de máxima verosimilitud. A diferencia del de mínimos cuadrados, el método de máxima verosimilitud descansa sobre un determinado supuesto acerca del tipo de distribución seguido por el término de error del modelo econométrico. Si el supuesto que se hace acerca de dicha distribución es aproximadamente correcto, se ganará eficiencia utilizando tal información adicional. Distintos supuestos acerca de dicha distribución generan estimadores de máxima verosimilitud diferentes, aunque casi con exclusividad se utiliza la hipótesis de Normalidad.

Dado el modelo econométrico $y = X\beta + u$, supongamos que el vector u sigue una distribución Normal y mantengamos el supuesto de que su esperanza es 0_T y su varianza la matriz $\sigma_u^2 I_T$. Bajo estas hipótesis, la función de densidad del vector u es:

$$\begin{aligned} f(u) &= \frac{1}{(2\pi)^{T/2}} \frac{1}{|\sigma_u^2 I_T|^{1/2}} \exp\left(-\frac{1}{2} u'(\sigma_u^2 I_T)^{-1} u\right) = \\ &= \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma_u^2)^{T/2}} \exp\left(-\frac{1}{2\sigma_u^2} u'u\right) \end{aligned} \quad [3.25]$$

La función de densidad anterior puede transformarse en la función de verosimilitud muestral si expresamos el vector u como función de las matrices y y X . Para ello habremos de tener en cuenta el resultado de la Sección 2.7 acerca del cambio de variable en funciones de densidad. El jacobiano de la transformación que convierte el vector aleatorio u en el vector aleatorio y es la matriz:

$$\left(\frac{\partial u}{\partial y}\right) = \begin{pmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \dots & \frac{\partial u_1}{\partial y_T} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \dots & \frac{\partial u_2}{\partial y_T} \\ \dots & \dots & \dots & \dots \\ \frac{\partial u_T}{\partial y_1} & \frac{\partial u_T}{\partial y_2} & \dots & \frac{\partial u_T}{\partial y_T} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

que tiene por determinante la unidad. En consecuencia, la función de verosimilitud se obtiene sustituyendo en [3.25] el vector u como función de y , para obtener:

$$L(y, X; \beta, \sigma_u^2) = \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma_u^2)^{T/2}} \exp\left(-\frac{1}{2\sigma_u^2} (y - X\beta)'(y - X\beta)\right) \quad [3.26]$$

El estimador de máxima verosimilitud de β y σ_u^2 está formado por aquellos valores de estos parámetros que maximizan la función de verosimilitud [3.26]. Por tanto, se trata de maximizar [3.26] con respecto a sus argumentos β y σ_u^2 , para una muestra dada. En primer lugar cabe recordar el resultado mencionado en el Capítulo 2 acerca de que es equivalente maximizar el logaritmo neperiano de la función de verosimilitud:

$$\ln L(\mathbf{y}, \mathbf{X}; \beta, \sigma_u^2) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)$$

e igualando a cero las derivadas con respecto a sus argumentos se tiene:

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} &= -\frac{1}{2\sigma_u^2} [-2\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)] = \mathbf{0}_k \\ \frac{\partial \ln L}{\partial \sigma_u^2} &= -\frac{T}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) = 0 \end{aligned}$$

o, lo que es lo mismo:

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{y} \quad [3.27]$$

$$\hat{\sigma}_u^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta})}{T} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T} \quad [3.28]$$

La ecuación [3.27] revela que, bajo el supuesto de Normalidad del término de error, el estimador de máxima verosimilitud del vector β coincide con el estimador MCO. En particular, esta equivalencia nos permite concluir, sin necesidad de demostrar: $E(\hat{\beta}_{MV}) = \beta$ y $\text{Var}(\hat{\beta}_{MV}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$.

Por el contrario, el estimador MV del parámetro σ_u^2 difiere del estimador MCO que introdujimos en la Sección 3.2. Además, el estimador MV de σ_u^2 es sesgado:

$$E(\hat{\sigma}_{MV}^2) = E\left(\frac{T-k}{T} \hat{\sigma}_{MCO}^2\right) = \frac{T-k}{T} \sigma_u^2 \quad [3.29]$$

si bien es cierto que, al aumentar el tamaño muestral, el estimador MV de σ_u^2 tiende al estimador MCO de dicho parámetro. En consecuencia, como muestra [3.29], el sesgo de $\hat{\sigma}_{MV}^2$ tiende a cero al aumentar el tamaño muestral.

Si obtenemos las derivadas segundas de la función $\ln L(\mathbf{y}, \mathbf{X}; \beta, \sigma_u^2)$:

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} &= -\frac{1}{\sigma_u^2} \mathbf{X}'\mathbf{X} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_u^2} &= -\frac{1}{\sigma_u^4} \mathbf{X}'(\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

$$\frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2} = \frac{T}{2\sigma_u^4} - \frac{1}{\sigma_u^6} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

y tenemos en cuenta que:

- a) $E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = E(\mathbf{u}'\mathbf{u}) = E(\sum_1^T u_i^2) = T\sigma_u^2$
 b) $E[\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = E[\mathbf{X}'\mathbf{u}] = \mathbf{X}'E(\mathbf{u}) = \mathbf{0}_k$

se tiene la matriz de información:

$$\mathbf{I}(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \begin{pmatrix} \frac{1}{\sigma_u^2} \mathbf{X}'\mathbf{X} & \mathbf{0}_k \\ \mathbf{0}_k & \frac{T}{2\sigma_u^4} \end{pmatrix}$$

cuya inversa es la cota de Cramer-Rao (véase Sección 2.11), es decir, la menor varianza que pueda tener un estimador insesgado.

Por comparación con los resultados anteriores puede verse que mientras que el estimador $\hat{\boldsymbol{\beta}}_{\text{MV}}$ alcanza la cota mínima de Cramer-Rao para la varianza de un estimador insesgado $(\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1})$ no ocurre así para el estimador $\hat{\sigma}_{\text{MV}}^2$. En efecto:

$$\text{Var}(\hat{\sigma}_{\text{MV}}^2) = \left(\frac{T-k}{T} \right)^2 \text{Var}(\hat{\sigma}_{\text{MCO}}^2) \quad [3.30]$$

Ahora bien, recordemos que $\frac{(T-k)\hat{\sigma}_{\text{MCO}}^2}{\sigma_u^2}$ sigue una distribución chi-cuadrado con $T-k$ grados de libertad y, en consecuencia, $\text{Var}(\hat{\sigma}_{\text{MCO}}^2) = \frac{2\sigma_u^4}{T-k}$. Esta varianza excede a la cota de Cramer-Rao para la varianza de un estimador insesgado de σ_u^2 , que como acabamos de ver es $\frac{2\sigma_u^4}{T}$; sin embargo, no existe un estimador insesgado de σ_u^2 que alcance dicha cota, por lo que el estimador MCO de σ_u^2 es eficiente.

Por otra parte, si llevamos esta última expresión a [3.30], se tiene:

$$\text{Var}(\hat{\sigma}_{\text{MV}}^2) = \left(\frac{(T-k)^2}{T^2} \right) \frac{2\sigma_u^4}{T-k} = 2(T-k) \frac{\sigma_u^4}{T^2} \quad [3.31]$$

que es inferior a la cota de Cramer-Rao, aunque hemos de recordar que $\hat{\sigma}_{\text{MV}}^2$ es un estimador sesgado de σ_u^2 . En el Problema 3.27 se prosigue con esta comparación.

7. REGRESION PARTICIONADA

Supongamos que se pretende estimar el modelo:

$$y = X\beta + Z\delta + u$$

Las ecuaciones normales son $\begin{bmatrix} (X'X) & (X'Z) \\ (Z'X) & (Z'Z) \end{bmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$, es decir:

$$\begin{cases} (X'X)\hat{\beta} + (X'Z)\hat{\delta} = X'y & [3.32] \\ (Z'X)\hat{\beta} + (Z'Z)\hat{\delta} = Z'y & [3.33] \end{cases}$$

lo que implica, de [3.33]:

$$\hat{\delta} = (Z'Z)^{-1}Z'(y - X\hat{\beta})$$

y también, sustituyendo en [3.32]:

$$X'y = (X'X)\hat{\beta} + (X'Z)(Z'Z)^{-1}Z'(y - X\hat{\beta})$$

por lo que:

$$X'M_z y = (X'M_z X)\hat{\beta} \quad [3.34]$$

donde la matriz M_z viene definida por:

$$M_z = I - Z(Z'Z)^{-1}Z'$$

y es simétrica e idempotente.

Supongamos, alternativamente, que para estimar el modelo se pretende seguir la estrategia de «descontar» de las variables y y X el efecto común que puedan tener las variables Z , para después estimar un modelo de la variable y «neta de X » sobre las variables X «netas de Z ».

Al estimar las regresiones de y y X sobre Z , los residuos son, de acuerdo con la Proposición 3.9:

$$u_y = M_z y$$

$$u_x = M_z X$$

y el estimador MCO de los coeficientes θ en la regresión de u_y sobre u_x : $u_y = \theta u_x + v$:

$$\hat{\theta} = (u_x' u_x)^{-1} u_x' u_y = (X' M_z X)^{-1} X' M_z y$$

es igual al coeficiente estimado $\hat{\beta}_{MCO}$ para el modelo original en [3.34].

Este resultado sugiere que la estimación MCO de un modelo de regresión puede efectuarse en dos etapas, estimando primero regresiones auxiliares sobre las variables «menos importantes» del modelo, y utilizando los residuos así generados para estimar los coeficientes de las variables más relevantes.

Una aplicación habitual de este resultado se discute en el Problema 3.8.

3.8. EL MODELO LINEAL EN DESVIACIONES CON RESPECTO A LA MEDIA

Cuando se especifica un modelo econométrico con término independiente:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t \quad [3.35]$$

para explicar el comportamiento de la variable endógena y_t a través del tiempo, conviene distinguir sus tres componentes: a) el término constante, b) el conjunto de las variables explicativas $\beta_2 x_{2t} + \cdots + \beta_k x_{kt}$, c) el término de error, u_t . Supongamos que el modelo ya ha sido estimado, obteniéndose:

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \cdots + \hat{\beta}_k x_{kt} + \hat{u}_t \quad [3.36]$$

Si sumamos esta expresión para valores del índice temporal t entre 1 y T y dividimos por el tamaño muestral T , tenemos:

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x}_2 + \hat{\beta}_3 \bar{x}_3 + \cdots + \hat{\beta}_k \bar{x}_k \quad [3.37]$$

ya que por haber un término constante en el modelo la media muestral de los residuos mínimo-cuadráticos es igual a cero. Por tanto, se tiene:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}_2 - \hat{\beta}_3 \bar{x}_3 - \cdots - \hat{\beta}_k \bar{x}_k \quad [3.38]$$

lo que indica que, en realidad, todo lo que debe preocuparnos es la estimación de los coeficientes β_2, \dots, β_k , puesto que la estimación del término independiente β_1 puede obtenerse posteriormente sin más que utilizar la expresión [3.38].

Pero hay algo más: si restamos de [3.36] la expresión [3.37], tenemos:

$$y_t - \bar{y} = \hat{\beta}_2 (x_{2t} - \bar{x}_2) + \hat{\beta}_3 (x_{3t} - \bar{x}_3) + \cdots + \hat{\beta}_k (x_{kt} - \bar{x}_k) + \hat{u}_t \quad [3.39]$$

que es una expresión similar a [3.36], aunque con dos diferencias importantes

- El nuevo modelo [3.39] no tiene término independiente.
- Todas las variables de [3.39] están medidas en diferencias con respecto a sus promedios muestrales.

También hay notables similitudes entre ambos modelos:

- Los coeficientes estimados son exactamente los mismos en ambos modelos, excepto el término independiente que no aparece en [3.39]
- Los residuos de ambos modelos son los mismos.

La mayoría de las veces, el interés del analista económico se centra en cuantificar la magnitud del impacto que las variables explicativas tienen sobre la variable endógena. Desde este punto de vista, el término constante no es sino una *corrección necesaria* que, como perfectamente ilustra [3.38], garantiza que los promedios muestrales de ambos miembros del modelo econométrico coincidan. En definitiva, el término constante juega un papel secundario en el modelo, y el modelo en desviaciones respecto a la media [3.39] es todo lo que el analista necesita. Una vez estimado éste, la estimación del término constante se recupera utilizando [3.38].

Tratemos de formalizar analíticamente este argumento: Consideremos el modelo en notación matricial $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, en el que suponemos que hay un término independiente, es decir, $x_{1t} = 1$ para todo t , y consideremos asimismo su versión estimada $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{u}$. Descomponemos la matriz \mathbf{X} en dos bloques: $\mathbf{X} = (\mathbf{1}_T; \mathbf{X}_2)$, donde \mathbf{X}_2 es la submatriz de orden $T \times (k - 1)$ formada por las columnas de observaciones de las variables explicativas, exceptuando al término constante. Descomponemos de modo similar el vector de estimaciones $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1; \hat{\boldsymbol{\beta}}_2)$, donde $\hat{\beta}_1$ es la estimación del término independiente y $\hat{\boldsymbol{\beta}}_2$ el subvector de estimaciones mínimo-cuadráticas del resto de los coeficientes.

Utilizamos a continuación la matriz idempotente $\mathbf{Q} = \mathbf{I}_T - (1/T)\mathbf{1}_T\mathbf{1}'_T$, de dimensión $T \times T$ que introdujimos en la Sección 1.1, por lo que el lector debe repasar sus propiedades. Premultiplicando el modelo por dicha matriz obtenemos:

$$\mathbf{Qy} = \mathbf{QX}\hat{\boldsymbol{\beta}} + \mathbf{Q}\hat{\mathbf{u}} = \mathbf{Q}\mathbf{1}_T\hat{\beta}_1 + \mathbf{QX}_2\hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}} = \mathbf{QX}_2\hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}} \quad [3.40]$$

donde se ha utilizado el hecho de que el producto $\mathbf{Q}\mathbf{1}_T$ es un vector columna de ceros y también que la suma de los residuos es cero, por lo que $\mathbf{Q}\hat{\mathbf{u}} = \hat{\mathbf{u}}$. El modelo [3.40] es, en notación matricial, precisamente el modelo [3.39]. Si premultiplicamos la expresión [3.40] ahora por la traspuesta de \mathbf{X}_2 , de dimensión $(k - 1) \times T$, se tiene $\mathbf{X}'_2\mathbf{Qy} = \mathbf{X}'_2\mathbf{QX}_2\hat{\boldsymbol{\beta}}_2$ (ya que $\mathbf{X}'_2\hat{\mathbf{u}} = \mathbf{0}_{k-1}$), y por ser la matriz \mathbf{Q} idempotente finalmente se tiene $\mathbf{X}'_2\mathbf{Qy} = \mathbf{X}'_2\mathbf{Q}\mathbf{Qy} = \mathbf{X}'_2\mathbf{Q}\mathbf{QX}_2\hat{\boldsymbol{\beta}}_2$, que puede ordenarse como:

$$(\mathbf{QX}_2)'(\mathbf{Qy}) = (\mathbf{QX}_2)'(\mathbf{QX}_2)\hat{\boldsymbol{\beta}}_2 \quad [3.41]$$

Denotemos por $\tilde{\mathbf{X}}_2 = \mathbf{QX}_2$ la matriz de dimensión $T \times (k - 1)$ cuyas columnas son las observaciones de las variables explicativas (excepto la constante) en desviaciones respecto a sus medias muestrales; análogamente, sea $\tilde{\mathbf{y}} = \mathbf{Qy}$ el vector de desviaciones de la variable endógena con respecto a su media muestral. El sistema [3.41] puede escribirse:

$$(\tilde{\mathbf{X}}'_2\tilde{\mathbf{X}}_2)\hat{\boldsymbol{\beta}}_2 = \tilde{\mathbf{X}}'_2\tilde{\mathbf{y}} \quad [3.42]$$

Puesto que la matriz \mathbf{QX}_2 es $T \times (k - 1)$ y el producto \mathbf{Qy} es un vector $T \times 1$, la igualdad [3.41] es realmente un sistema de $k - 1$ ecuaciones. Por comparación con [3.16] puede verse que [3.42] no es sino el sistema de ecuaciones normales [3.16], donde las distintas matrices y vectores han sido

sustituidos por su producto por \mathbf{Q} , mientras que el vector $\boldsymbol{\beta}$ ha sido sustituido por el subvector $\boldsymbol{\beta}_2$.

Dicho de otra forma: Si transformamos todas las variables del modelo en desviaciones con respecto a sus medias muestrales, las $k - 1$ ecuaciones normales que pueden formarse con estas variables tienen *exactamente* las mismas soluciones que el sistema [3.16] excepto, claro está, el término independiente β_1 , que no aparece en [3.42]. En consecuencia, *las estimaciones mínimo-cuadráticas de los coeficientes β_2, \dots, β_k en ambos modelos son las mismas.*

Por otra parte, si premultiplicamos el modelo estimado $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ por $\begin{pmatrix} 1 \\ \mathbf{1}'_T \end{pmatrix}$, donde $\mathbf{1}_T$ es un vector columna de unos, se tiene:

$$\begin{pmatrix} 1 \\ \mathbf{1}'_T \end{pmatrix} \mathbf{y} = \begin{pmatrix} 1 \\ \mathbf{1}'_T \end{pmatrix} (\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}})$$

es decir, $\bar{y} = \hat{\beta}_1 + \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}_2$, donde $\bar{\mathbf{x}}$ es el vector de medias muestrales de las variables explicativas, exceptuando el término independiente. Esta es exactamente nuestra anterior ecuación [3.37], y nos permite obtener la estimación mínimo-cuadrática del término independiente una vez que hemos resuelto el sistema de ecuaciones normales [3.42].

Continuemos ahora premultiplicando la ecuación [3.40] por \mathbf{y}' para obtener:

$$\begin{aligned} \mathbf{y}'\mathbf{Q}\mathbf{y} &= (\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}})'(\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}}) = \hat{\boldsymbol{\beta}}_2'\mathbf{X}'\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\beta}}_2'\mathbf{X}'\hat{\mathbf{u}} + \hat{\mathbf{u}}'\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \\ &+ \hat{\mathbf{u}}'\hat{\mathbf{u}} = \hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}}'\hat{\mathbf{u}} = \hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{y} + \hat{\mathbf{u}}'\hat{\mathbf{u}} \end{aligned} \quad [3.43]$$

donde hemos utilizado que $\mathbf{Q}'\hat{\mathbf{u}} = \mathbf{0}_T$ junto con la Proposición 3.3 y también que $\mathbf{Q}\mathbf{X} = (\mathbf{0}_T; \mathbf{Q}\mathbf{X}_2)$, por lo que $\mathbf{Q}\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{0}_T; \mathbf{Q}\mathbf{X}_2)\hat{\boldsymbol{\beta}} = \mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$.

Examinemos la ecuación [3.43]: La expresión $\mathbf{y}'\mathbf{Q}\mathbf{y}$ es la suma de cuadrados de las diferencias entre los valores muestrales de la variable y_i y su valor medio, es decir, la suma total de y_i . Por otra parte, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ es la suma residual del modelo original. Como dicho modelo incluía un término independiente (por hipótesis), entonces necesariamente la expresión $\hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$ o su equivalente $\hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{y}$ no son sino la suma explicada (ya que, en este caso, $\text{ST} = \text{SE} + \text{SR}$). Notemos, por último, que el producto $\mathbf{X}'_2\mathbf{Q}\mathbf{X}_2$ es la matriz de productos cruzados de las variables x_2, \dots, x_k , todas ellas en desviaciones con respecto a sus medidas muestrales.

Por tanto, en el caso en que el modelo contenga un término independiente, entonces el coeficiente de determinación (o R -cuadrado) puede también calcularse mediante la expresión:

$$R^2 = 1 - \left(\frac{\text{SR}}{\text{ST}} \right) = \frac{\hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{X}_2\hat{\boldsymbol{\beta}}_2}{\mathbf{y}'\mathbf{Q}\mathbf{y}} = \frac{\hat{\boldsymbol{\beta}}_2'\mathbf{X}'_2\mathbf{Q}\mathbf{y}}{\mathbf{y}'\mathbf{Q}\mathbf{y}}$$

donde hay que observar con cierto cuidado que todos los cálculos están efectuados con las variables medidas en diferencias con respecto a sus medias, que es lo que la presencia de la matriz \mathbf{Q} indica. El lector debe comprobar la veracidad de este resultado en el Ejemplo 3.3.

Estas expresiones son totalmente similares, aunque no idénticas, a las obtenidas en la Proposición 3.14 para el caso de un modelo econométrico con término independiente. Las diferencias son que en el modelo en desviaciones:

- a) Se utilizan la submatriz \mathbf{X}_2 y el subvector $\hat{\beta}_2$ en vez de la matriz \mathbf{X} y el vector $\hat{\beta}$.
- b) No aparece la corrección mediante el término $T\bar{y}_2$ que aparece en la Proposición 3.14.

Estos cambios son muy naturales, puesto que, como hemos visto, en el modelo en desviaciones desaparece el coeficiente β_1 , y además la media muestral de la variable dependiente es cero.

Por tanto, en un modelo $y_t = x_t'\beta + u_t$, donde las variables se hallen en desviaciones respecto a la media, el coeficiente de determinación puede escribirse:

$$R^2 = \frac{\hat{\beta}' \tilde{\mathbf{X}}_2' \tilde{\mathbf{y}}}{\tilde{\mathbf{y}}' \tilde{\mathbf{y}}} = \frac{\mathbf{y}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}}{\mathbf{y}' \mathbf{y}} = \frac{\left(\sum_1^T y_t x_t' \right) \left(\sum_1^T x_t x_t' \right)^{-1} \left(\sum_1^T x_t y_t \right)}{\sum_1^T y_t^2}$$

Volviendo a la estimación del término independiente hemos visto únicamente cómo recuperar el valor numérico de dicho estimador, pero no hemos hablado de su varianza y covarianzas con los demás estimadores. En el Problema 3.21 se demuestra que la media muestral \bar{y} es independiente del estimador MCO del vector β_2 formado por los $k - 1$ coeficientes del modelo que excluyen al término independiente. Como consecuencia, se tiene:

$$\text{Var}(\hat{\beta}_1) = \text{Var}(\bar{y} - \bar{x}_2' \hat{\beta}_2) = \text{Var}(\bar{y}) + \bar{x}_2' \text{Var}(\hat{\beta}_2) \bar{x}_2 = \frac{\sigma_u^2}{T} + \sigma_u^2 \bar{x}_2' (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1} \bar{x}_2$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{Cov}(\bar{y} - \bar{x}_2' \hat{\beta}_2, \hat{\beta}_2) = -\bar{x}_2' \text{Var}(\hat{\beta}_2) = -\sigma_u^2 \bar{x}_2' (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1}$$

Ejemplo 3.5. Suponemos en este ejemplo que disponemos de cinco observaciones temporales acerca de una variable endógena y_t y dos variables explicativas x_{2t} y x_{3t} , además de un término constante:

t	y	x_1	x_2	x_3	\tilde{y}	\tilde{x}_2	\tilde{x}_3
1	3	1	3	5	-1	0	0
2	1	1	1	5	-3	-2	-1
3	8	1	5	6	4	2	1
4	3	1	2	4	-1	-1	-1
5	5	1	4	6	1	1	1

Las medidas muestrales son: $\bar{y} = 4$, $\bar{x}_2 = 3$, $\bar{x}_3 = 5$, y transformando las variables en desviaciones con respecto a la media se tienen las matrices:

$$\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2 = \begin{pmatrix} 10 & 6 \\ 6 & 4 \end{pmatrix}; \quad \tilde{\mathbf{X}}_2' \tilde{\mathbf{y}} = \begin{pmatrix} 16 \\ 9 \end{pmatrix}; \quad (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1} = \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2} \end{pmatrix}; \quad \tilde{\mathbf{y}}' \tilde{\mathbf{y}} = 28$$

que generan las estimaciones $\hat{\beta}_2 = \frac{5}{2}$, $\hat{\beta}_3 = -\frac{3}{2}$, con matriz de covarianzas:

$$\text{Var} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \sigma_u^2 (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1} = \sigma_u^2 \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

A partir de estas estimaciones recuperamos $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}_2 - \hat{\beta}_3 \bar{x}_3 = 4$, con varianza:

$$\text{Var}(\hat{\beta}_1) = \sigma_u^2 \left(\frac{1}{T} + \bar{\mathbf{x}}_2' (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1} \bar{\mathbf{x}}_2 \right) = \sigma_u^2 \left[\frac{1}{5} + (3; 5) (\tilde{\mathbf{X}}_2' \tilde{\mathbf{X}}_2)^{-1} \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right] = 26,7 \sigma_u^2$$

que, como el lector puede comprobar, coincide con el elemento (1, 1) de la matriz $\sigma_u^2 (\mathbf{X}' \mathbf{X})^{-1}$ calculada a partir de las variables originales. De modo similar:

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{Cov} \left[\hat{\beta}_1, \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} \right] = -(3; 5) \cdot \text{Var}(\hat{\beta}_2) = \sigma_u^2 \cdot \begin{pmatrix} 9/2 \\ -8 \end{pmatrix}$$

La estimación del parámetro σ_u^2 es:

$$\hat{\sigma}_u^2 = \frac{\tilde{\mathbf{y}}' \tilde{\mathbf{y}} - \hat{\beta}' \tilde{\mathbf{X}}_2' \tilde{\mathbf{y}}}{T - k} = \frac{3/2}{5 - 3} = \frac{3/2}{2} = \frac{3}{4}$$

En consecuencia, se tiene el modelo estimado:

$$y_t = 4 + 2,5x_{2t} - 1,5x_{3t} + \hat{u}_t$$

(4,47) (0,87) (1,37)

mientras que la matriz

$$\Sigma = \begin{pmatrix} 20,03 & 3,38 & -6,0 \\ 0,87 & 0,75 & -1,13 \\ -0,98 & -0,95 & 1,88 \end{pmatrix}$$

tiene:

- a) En su diagonal principal las varianzas de los coeficientes estimados.
- b) En su triángulo superior, las covarianzas entre ellas.

- c) En su triángulo inferior, los coeficientes de correlación entre coeficientes estimados que, como puede observarse, son todos muy elevados.

A modo de resumen, citemos las condiciones bajo las que son válidas algunas de las propiedades que hemos discutido en secciones anteriores:

<i>Propiedad</i>	<i>Es cierta</i>
1. $SE = \hat{\beta}'X'y - T\bar{y}^2$	Si el modelo tiene término independiente.
2. $SE = \hat{\beta}'X'y$	Si las variables del modelo están en desviaciones con respecto a sus medias.
3. $ST = SE + SR$	Si el modelo tiene término independiente.
4. $R^2 = 1 - SR/ST$	Siempre
5. $R^2 = \frac{SE}{ST}$	Si el modelo tiene término independiente.
6. $R^2 \leq 1$	Siempre.
7. $R^2 \geq 0$	Si el modelo tiene término independiente.
8. $SR = y'y - \hat{\beta}'X'y$	Siempre.
9. $\sum_1^T \hat{u}_t = 0$	Si el modelo tiene término independiente.
10. $X'\hat{u} = 0_k$	Siempre

3.9. ALGUNOS MODELOS LINEALES SENCILLOS

Por su sencillez, así como por la frecuencia con que aparecen en la práctica, es interesante particularizar los resultados de las secciones previas a algunos casos en que el modelo econométrico que se pretende estimar incluye un número reducido de variables. Un interés adicional de este análisis es que nos permite conseguir una interpretación más intuitiva de algunas de las propiedades del estimador MCO ya vistas.

Caso 1. Modelo que sólo incluye como variable explicativa una constante.

Comencemos analizando el modelo:

$$y_t = \alpha + u_t, \quad t = 1, 2, \dots, T \quad [3.44]$$

donde se pretende explicar la variable endógena únicamente mediante una constante. En este modelo, la matriz \mathbf{X} se reduce al vector columna $\mathbf{1}_T$, por lo que $\mathbf{X}'\mathbf{X} = T$, $\mathbf{X}'\mathbf{y} = \sum_1^T y_t$ y en consecuencia $\hat{\alpha} = \bar{y}$. Es decir, la mejor constante (en el sentido de mínimos cuadrados) que puede utilizarse para explicar el comportamiento de una variable aleatoria y_t es su media muestral.

La varianza de dicho estimador es $\text{Var}(\hat{\alpha}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma_u^2}{T}$, lo cual era de esperar ya que $\hat{\alpha} = \bar{y}$, y de acuerdo con [3.44]: $\text{Var } y_t = \text{Var } u_t = \sigma_u^2$.

Los residuos son $\hat{u} = y_t - \hat{\alpha}$ y, por tanto, $\text{SR} = \sum_1^T (y_t - \bar{y})^2 = \text{ST}$. Como el modelo tiene término independiente, se tiene $\text{SE} = \text{ST} - \text{SR} = 0$. Es decir, explicando de modo óptimo el comportamiento de y_t mediante una constante se obtiene una suma explicada igual a cero. Esto es consistente con el argumento utilizado en secciones anteriores acerca de que lo que realmente importa es explicar las fluctuaciones de y_t alrededor de su media, no tanto el valor de ésta. Con este modelo no se explican las fluctuaciones de y_t en absoluto. De hecho, la predicción generada por el modelo y_t es constante e igual a \bar{y} para todo t . Por último, la estimación de σ_u^2 es $\hat{\sigma}_u^2 = \frac{\text{ST}}{T-1}$, que no es sino la estimación insesgada de la varianza de y_t .

Caso 2. Modelo con una variable explicativa y una constante.

Consideremos ahora el llamado *modelo lineal simple*:

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T$$

cuyas ecuaciones normales son:

$$\sum_1^T y_t = T\hat{\alpha} + \hat{\beta} \sum_1^T x_t$$

$$\sum_1^T x_t y_t = \hat{\alpha} \sum_1^T x_t + \hat{\beta} \sum_1^T x_t^2$$

De la primera ecuación se tiene, dividiendo por T : $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$, y sustituyendo esta expresión en la segunda ecuación se tiene:

$$\hat{\beta} = \frac{\sum_1^T x_t y_t - (1/T) \sum_1^T x_t \sum_1^T y_t}{\sum_1^T x_t^2 - (1/T) (\sum_1^T x_t)^2} = \frac{\sum_1^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_1^T (x_t - \bar{x})^2} = \frac{S_{xy}}{S_x^2} \quad [3.45]$$

donde S_{xy} , S_x^2 , S_y^2 denotan, respectivamente, la covarianza y varianzas muestrales de x e y , siendo S_x , S_y la raíz cuadrada de S_x^2 , S_y^2 . Finalmente, una vez calculado el valor de $\hat{\beta}$ se obtiene $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$.

Invirtiendo la matriz $\mathbf{X}'\mathbf{X} = \begin{pmatrix} T & \sum_1^T x_t \\ \sum_1^T x_t & \sum_1^T x_t^2 \end{pmatrix}$ se tiene que las varianzas de estos estimadores son:

$$\begin{aligned}\text{Var}(\hat{\beta}) &= \frac{\sigma_u^2}{\sum_1^T (x_t - \bar{x})^2} = \frac{\sigma_u^2}{TS_x^2} \\ \text{Var}(\hat{\alpha}) &= \frac{\sigma_u^2 \sum_1^T x_t^2}{T \sum_1^T (x_t - \bar{x})^2} = \frac{\sigma_u^2 \sum_1^T x_t^2}{T^2 S_x^2} \\ \text{Cov}(\hat{\alpha}, \hat{\beta}) &= \frac{-\bar{x} \sigma_u^2}{\sum_1^T (x_t - \bar{x})^2} = \frac{-\bar{x} \sigma_u^2}{TS_x^2}\end{aligned}\quad [3.46]$$

Es inmediato comprobar que la suma residual de este modelo es igual a $T \left(S_y^2 - \frac{S_{xy}^2}{S_x^2} \right)$, mientras que $T \frac{S_{xy}^2}{S_x^2}$ es la suma explicada del modelo. La varianza σ_u^2 se estima mediante el cociente $\frac{\text{SR}}{T-2}$. El coeficiente de determinación viene dado por:

$$R^2 = 1 - \frac{\text{SR}}{\text{ST}} = 1 - \frac{T \left(S_y^2 - \frac{S_{xy}^2}{S_x^2} \right)}{TS_y^2} = \frac{S_{xy}^2}{S_x^2 S_y^2}$$

mientras que el coeficiente de correlación lineal muestral entre estas dos variables se define como la raíz cuadrada del anterior, adoptando el signo de la covarianza entre x e y :

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

y, por tanto, se tiene:

$$r_{xy} = \hat{\beta} \frac{S_x}{S_y}$$

y, en este sentido, la estimación del modelo lineal simple nos da una idea de la correlación lineal existente entre las variables x e y . En efecto, el estimador MCO del coeficiente de la variable x_t es $\hat{\beta} = \frac{r_{xy} S_y}{S_x}$. Así pues, si bien es cierto que la estimación del parámetro β no coincide con el coeficiente de correlación lineal entre x_t e y_t , es proporcional al mismo.

La relación que acabamos de obtener admite además una interpretación bastante intuitiva. Supongamos que ambas variables están positiva y perfectamente correlacionadas, es decir, $r_{xy} = 1$, y que sus varianzas muestrales son $S_y^2 = 100$ y $S_x^2 = 25$. Supongamos asimismo que a partir de sus valores medios se pretende conocer el incremento que es preciso conseguir en la variable explicativa para que la variable endógena se desvíe diez unidades por encima

de su media. En estas condiciones, $\hat{\beta} = \frac{S_y}{S_x} = 2$, y para conseguir el objetivo deseado es preciso que la variable explicativa se desvíe en cinco unidades por encima de su media muestral. Dado que el tamaño de las fluctuaciones que experimenta la variable x alrededor de su media en la muestra es notablemente inferior al de las fluctuaciones de la variable y , no es preciso que la desviación de x con respecto a su media sea tan grande como la que se pretende conseguir en la variable y .

Si la correlación no fuese perfecta, sino que $r_{xy} = \frac{1}{2}$, por ejemplo, entonces sería preciso que x tomase un valor de diez unidades por encima de su media. Si el coeficiente de correlación tuviese signo negativo, entonces las desviaciones x deberían ser por debajo y no por encima de su media muestral.

Conviene tener presente que el *coeficiente de correlación lineal* r_{xy} mide el grado de asociación que existe entre ambas variables cuando se ajusta a su nube de puntos una *línea recta*, pero no mide el grado de ajuste de una curva a la nube de puntos. Podría darse el caso de que la relación entre dos variables fuese muy estrecha, sólo que distribuida a lo largo de una curva, en cuyo caso, al ajustar una recta se podría tener un coeficiente de correlación lineal (así como un coeficiente de determinación) bajo. Un caso extremo es el de la nube de puntos:

x	-3	-2	-1	0	1	2	3
y	9	4	1	0	1	4	9

que representan la relación funcional exacta $y = x^2$, de modo que están alineados a lo largo de dicha parábola, y, sin embargo, *el coeficiente de correlación lineal entre ambas variables es cero*.

Caso 3. Modelo con dos variables explicativas y una constante.

Consideremos el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t, \quad t = 1, 2, \dots, T$$

La solución al sistema de ecuaciones normales produce los estimadores (veáse Problema 3.19):

$$\hat{\beta}_2 = \frac{\sum_1^T \tilde{x}_{2t} \tilde{y}_t \sum_1^T \tilde{x}_{3t}^2 - \sum_1^T \tilde{x}_{3t} \tilde{y}_t \sum_1^T \tilde{x}_{2t} \tilde{x}_{3t}}{\sum_1^T \tilde{x}_{2t}^2 \sum_1^T \tilde{x}_{3t}^2 - (\sum_1^T \tilde{x}_{2t} \tilde{x}_{3t})^2} \tag{3.47}$$

$$\hat{\beta}_3 = \frac{\sum_1^T \tilde{x}_{3t} \tilde{y}_t \sum_1^T \tilde{x}_{2t}^2 - \sum_1^T \tilde{x}_{2t} \tilde{y}_t \sum_1^T \tilde{x}_{2t} \tilde{x}_{3t}}{\sum_1^T \tilde{x}_{2t}^2 \sum_1^T \tilde{x}_{3t}^2 - (\sum_1^T \tilde{x}_{2t} \tilde{x}_{3t})^2}$$

donde $\tilde{y}_t, \tilde{x}_{2t}, \tilde{x}_{3t}$ denotan las variables en diferencias con respecto a la media. Los estimadores MCO en [3.47] pueden escribirse en función de los estimados

res que se obtendrían en los modelos simples. Así, si denotamos por $\hat{\beta}_{13}$ al estimador MCO de β en la regresión $y_t = \alpha + \beta x_{3t} + u_t$ y definimos análogamente $\hat{\beta}_{12}$, $\hat{\beta}_{23}$ y $\hat{\beta}_{32}$, entonces, dividiendo en [3.47] por $\Sigma \tilde{x}_{2t}^2 \Sigma \tilde{x}_{3t}^2$, se tiene:

$$\hat{\beta}_2 = \frac{\hat{\beta}_{12} - \hat{\beta}_{13}\hat{\beta}_{32}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}} \quad [3.48]$$

$$\hat{\beta}_3 = \frac{\hat{\beta}_{13} - \hat{\beta}_{12}\hat{\beta}_{23}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}}$$

Es importante observar que si las variables x_2 y x_3 fuesen independientes en la muestra, es decir, si $S_{x_2 x_3} = 0$, entonces se tendría:

$$\hat{\beta}_2 = \hat{\beta}_{12} \quad \text{y} \quad \hat{\beta}_3 = \hat{\beta}_{13}$$

por lo que si x_2 y x_3 son ortogonales, entonces las estimaciones MCO del modelo que incluye a ambas simultáneamente como variables explicativas coinciden con las estimaciones que se obtendrían de modelos que las incluyesen por separado como explicativas. Aunque no se ha probado en este capítulo, este resultado puede generalizarse al siguiente (véase Problema 3.26):

Teorema 3.1. Supongamos que en el modelo lineal general: $y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$, el vector de las variables explicativas puede dividirse en dos subvectores \mathbf{x}_{1t} y \mathbf{x}_{2t} , ortogonales entre sí, al menos en la muestra, es decir, $S_{x_i x_j} = 0$ para toda variable x_i en el subvector \mathbf{x}_{1t} y toda variable x_j en el subvector \mathbf{x}_{2t} . Entonces las estimaciones MCO de los subvectores de coeficientes $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$ en el modelo:

$$y_t = \mathbf{x}'_{1t} \boldsymbol{\beta}_1 + \mathbf{x}'_{2t} \boldsymbol{\beta}_2 + u_t$$

coincidirían con las estimaciones MCO que se obtendrían en los modelos:

$$y_t = \mathbf{x}'_{1t} \boldsymbol{\beta}_1 + u_{1t}$$

$$y_t = \mathbf{x}'_{2t} \boldsymbol{\beta}_2 + u_{2t}$$

Una vez se han expresado los estimadores en [3.47] como función de los estimadores MCO de los modelos lineales simples, es sencillo expresarlos como función de los coeficientes de correlación simples:

$$\hat{\beta}_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \cdot \frac{S_y}{S_{x_2}} \quad [3.49]$$

$$\hat{\beta}_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \cdot \frac{S_y}{S_{x_3}}$$

donde los subíndices tienen la misma interpretación que en [3.48].

Definición. El coeficiente de correlación parcial muestral entre las variables y_t y x_{2t} del modelo de tres variables que estamos considerando se define como el coeficiente de correlación simple entre estas variables, una vez que se han corregido de la influencia de la variable x_{3t} .

En la definición se habla de descontar de y_t el efecto de x_{3t} , antes de correlacionar la variable y_t corregida con la variable x_{2t} , corregida de modo similar. ¿Cómo se mide el efecto de x_{3t} sobre y_t ? Lógicamente, estimando la regresión $\tilde{y}_t = \beta_{13}\tilde{x}_{3t} + u_{3t}$. La diferencia $\tilde{y}_t - \hat{\beta}_{13}\tilde{x}_{3t}$ puede, efectivamente, interpretarse como el resultado de «descontar» de y_t el efecto de la variable x_{3t} . De modo análogo podría construirse la variable «corregida» $\tilde{x}_{2t} - \hat{\beta}_{23}\tilde{x}_{3t}$. Pues bien, al coeficiente de correlación lineal simple entre estas variables:

$$r_{12,3} = \frac{\sum (\tilde{y}_t - \hat{\beta}_{13}\tilde{x}_{3t}) (\tilde{x}_{2t} - \hat{\beta}_{23}\tilde{x}_{3t})}{\sqrt{\sum (\tilde{y}_t - \hat{\beta}_{13}\tilde{x}_{3t})^2} \sqrt{\sum (\tilde{x}_{2t} - \hat{\beta}_{23}\tilde{x}_{3t})^2}}$$

se le denomina *coeficiente de correlación parcial entre y_t y x_{2t}* . La calificación de «correlación parcial» proviene del hecho de que $r_{12,3}$ mide la correlación entre ambas variables que no proviene del hecho de que ambas tienen una causa común, x_{3t} . Obsérvese que de modo análogo podría definirse $r_{13,2}$.

El concepto de correlación parcial puede asimismo introducirse en un modelo con k variables. En tal caso, en una primera etapa se estimaría una regresión de y sobre x_3, x_4, \dots, x_k , así como una regresión de x_2 sobre las mismas variables. En una segunda etapa se calcularía el coeficiente de correlación simple de los residuos de ambas regresiones. Nótese que éste es precisamente el procedimiento seguido en el caso del modelo de tres variables.

3.10. CAMBIOS DE ESCALA Y DE ORIGEN

En ocasiones, al estimar un modelo econométrico, interesa cambiar de unidades una variable para hacer sus valores numéricos más comparables con los de las demás variables. Para ello se multiplican o dividen todas sus observaciones por una misma constante. Otras veces interesa «cambiar de origen» o «trasladar» una variable, lo que en términos de la serie equivale a sumar o restar una misma constante a todas sus observaciones.

Es importante saber en estos casos cómo se verán afectados por estas transformaciones en los datos los estadísticos relevantes del modelo, a saber:

- a) Los estimadores MCO de los parámetros.
- b) El vector de residuos y , por consiguiente, la suma residual.
- c) El coeficiente de determinación.

Distintas variedades de problemas pueden plantearse, según que las transformaciones sean tan sólo un cambio de escala o, alternativamente, un cambio de origen, que afecten a todas las variables explicativas o sólo a algunas de ellas, y que se lleven a cabo únicamente en las variables explicativas, o en

ellas y también en la variable endógena. Vamos a discutir en esta sección los casos que consideramos, por su frecuencia, más importantes, dejando los demás para el lector, aunque la discusión que aquí presentamos es suficiente para sugerir el tipo de efectos que se derivará en las otras situaciones posibles.

Caso 1. Cambios de escala en todas las variables del modelo (incluida la variable endógena).

Supongamos que se multiplican todas las observaciones de cada una de las variables explicativas x_{it} por una constante λ_i , $\lambda_i > 0$, $i = 1, 2, \dots, k$. Estas constantes pueden ser de la forma 10^{-1} ó 10^{-3} , en cuyo caso estaríamos dividiendo todas las observaciones de la variable x_{it} por 10 o por 1.000, respectivamente. Un valor $\lambda = 1$ indica que la variable explicativa correspondiente no cambia. Denotamos por λ la constante por la que se transforma la variable endógena y . Matricialmente, esta transformación de variables puede expresarse:

$$\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ x_{13} & x_{23} & \cdots & x_{k3} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda_k \end{pmatrix} = \begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \cdots & \lambda_k x_{k1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \cdots & \lambda_k x_{k2} \\ \lambda_1 x_{13} & \lambda_2 x_{23} & \cdots & \lambda_k x_{k3} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1 x_{1T} & \lambda_2 x_{2T} & \cdots & \lambda_k x_{kT} \end{pmatrix}$$

o $\mathbf{X}\mathbf{\Lambda} = \mathbf{X}^*$, donde $\mathbf{\Lambda}$ es la matriz diagonal $k \times k$ que contiene los factores de cambio de escala y \mathbf{X}^* la matriz $T \times k$ de las observaciones de las variables x_1, x_2, \dots, x_k , transformadas. El vector \mathbf{y}^* de observaciones de la variable y transformada se obtiene por $\mathbf{y}^* = \lambda \mathbf{y}$. Denotemos el modelo en las variables transformadas:

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{u}^*$$

donde, como se ve, admitimos la posibilidad de que tanto el vector de coeficientes como el término de error sean ahora diferentes de los del modelo original. El estimador MCO del vector $\boldsymbol{\beta}$ viene dado por:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{y}^* = (\boldsymbol{\Lambda}' \mathbf{X}' \mathbf{X} \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}' \mathbf{X}' \lambda \mathbf{y} = \\ &= \lambda \boldsymbol{\Lambda}^{-1} (\mathbf{X}' \mathbf{X})^{-1} (\boldsymbol{\Lambda}')^{-1} \boldsymbol{\Lambda}' \mathbf{X}' \mathbf{y} = \lambda \boldsymbol{\Lambda}^{-1} \hat{\boldsymbol{\beta}} \end{aligned}$$

donde hemos utilizado el hecho de que la matriz $\boldsymbol{\Lambda}$ es invertible (salvo en el caso en que $\lambda_i = 0$ para algún i). Este resultado implica que hay una relación sencilla entre los estimadores MCO de ambos modelos. En efecto, para cada coeficiente se tiene:

$$\hat{\beta}_i^* = \left(\frac{\lambda}{\lambda_i} \right) \hat{\beta}_i, \quad i = 1, 2, \dots, k$$

una relación que depende de los factores de cambio de escala de la variable endógena y la correspondiente a ese parámetro. En particular, si la variable endógena no se modifica ($\lambda = 1$), se tiene el intuitivo resultado de que *al dividir los valores de una variable explicativa por una constante el estimador MCO de su coeficiente quedará multiplicado por dicha constante*. Lo contrario ocurre si la variable explicativa se multiplica por una constante. Sin embargo, cuando también se modifica la variable endógena, entonces el efecto dependerá del valor relativo de los dos factores de escala.

En cuanto al vector de residuos MCO:

$$\hat{\mathbf{u}}^* = \mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}^* = \lambda \mathbf{y} - \mathbf{X} \lambda \lambda^{-1} \hat{\boldsymbol{\beta}} = \lambda \hat{\mathbf{u}}$$

es decir, que el vector de residuos sólo se ve afectado cuando hay una transformación en la variable endógena. Este resultado es bastante intuitivo, pues \hat{u}_i es el error cometido al estimar y_i debido a su naturaleza aleatoria, y cabe esperar que al multiplicar y_i por una constante el error correspondiente sufra la misma modificación. Como consecuencia de lo anterior, es fácil obtener las siguientes relaciones entre las sumas total y residual de ambos modelos:

$$ST^* = \sum_1^T (y_i^* - \bar{y}^*)^2 = \sum_1^T [\lambda(y_i - \bar{y})]^2 = \lambda^2 \sum_1^T (y_i - \bar{y})^2 = \lambda^2 ST$$

$$SR^* = \hat{\mathbf{u}}^{*'} \hat{\mathbf{u}}^* = (\lambda \hat{\mathbf{u}})' (\lambda \hat{\mathbf{u}}) = \lambda^2 \hat{\mathbf{u}}' \hat{\mathbf{u}} = \lambda^2 SR$$

donde hemos utilizado las relaciones anteriores junto con la propiedad $\bar{y}^* = \lambda \bar{y}$. Sin más que recordar la definición del coeficiente de determinación puede verse inmediatamente que:

$$R^{*2} = R^2$$

y, como consecuencia, en tanto en cuanto se acepte el R -cuadrado como una medida de la bondad de ajuste del modelo, ambos modelos muestran el mismo grado de ajuste. En cuanto a la estimación de σ_u^2 :

$$\hat{\sigma}_u^{*2} = \frac{\hat{\mathbf{u}}^{*'} \hat{\mathbf{u}}^*}{T - k} = \frac{\lambda^2 \hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k} = \lambda^2 \hat{\sigma}_u^2$$

Caso 2. Traslación en las variables explicativas.

Consideremos el modelo transformado:

$$\mathbf{y} = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{u}^* = 1_T \beta_1^* + \mathbf{X}_2^* \boldsymbol{\beta}_2^* + \mathbf{u}^*$$

donde β_1^* es el término independiente y $\boldsymbol{\beta}_2^*$ el subvector de $\boldsymbol{\beta}$ que contiene a los restantes coeficientes. Supongamos que las transformaciones son del tipo traslación, es decir, cada variable explicativa tiene todas sus observaciones aumentadas o disminuidas por un valor constante (posiblemente cero). Sea \mathbf{B} la matriz $T \times (k - 1)$ que refleja esos cambios:

$$\mathbf{B} = \begin{pmatrix} b_2 & b_3 & b_4 & \cdots & b_k \\ b_2 & b_3 & b_4 & \cdots & b_k \\ b_2 & b_3 & b_4 & \cdots & b_k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_2 & b_3 & b_4 & \cdots & b_k \end{pmatrix} = \mathbf{1}_T \mathbf{b}'$$

donde \mathbf{b}' es el $(k-1)$ vector fila (b_2, b_3, \dots, b_k) , de modo que las nuevas variables son:

$$\mathbf{X}_2^* = \mathbf{X}_2 + \mathbf{B} = \mathbf{X}_2 + \mathbf{1}_T \mathbf{b}'$$

Claramente bastaría que una de las constantes b_i fuese cero para que la variable correspondiente no sufriese alteración. El modelo en las variables transformadas puede escribirse:

$$\begin{aligned} \mathbf{y} &= \beta_1^* \mathbf{1}_T + \mathbf{X}_2^* \beta_2^* + \mathbf{u}^* = \mathbf{1}_T \beta_1^* + [\mathbf{X}_2 + (\mathbf{1}_T \mathbf{b}')] \beta_2^* + \mathbf{u}^* = \\ &= \mathbf{1}_T (\beta_1^* + \mathbf{b}' \beta_2^*) + \mathbf{X}_2 \beta_2^* + \mathbf{u}^* \end{aligned}$$

que es un modelo que tiene todas las variables iguales a las del modelo original (no hay cambios de escala en ellas). Como consecuencia, las estimaciones MCO de los coeficientes β_2 serán iguales en ambos modelos. Como también la variable endógena es la misma (pues suponemos que los cambios ocurren únicamente en las variables explicativas), entonces el valor numérico del estimador MCO del término independiente será también el mismo (para ver por qué puede pensarse en obtener, en cada caso, el modelo en diferencias con respecto a la media). De modo que se tiene:

$$\begin{aligned} \hat{\beta}_2^* &= \hat{\beta}_2 \\ \hat{\beta}_1^* + \mathbf{b}' \hat{\beta}_2^* &= \hat{\beta}_1 \Rightarrow \hat{\beta}_1^* = \hat{\beta}_1 - \mathbf{b}' \hat{\beta}_2 \end{aligned}$$

El único efecto estriba, como puede verse, en la estimación del parámetro β_1 .

Caso 3. Cambio de escala en todas las variables (incluida la variable endógena) y traslación en las variables explicativas, simultáneamente.

Las variables transformadas son ahora:

$$\mathbf{y}^* = \lambda \mathbf{y}; \quad \mathbf{X}_2^* = \mathbf{X}_2 \Lambda + (\mathbf{1}_T \mathbf{b}')$$

por lo que, con estas variables, el modelo puede escribirse:

$$\mathbf{y}^* = \lambda \mathbf{y} = \mathbf{1}_T (\beta_1^* + \mathbf{b}' \beta_2^*) + (\mathbf{X}_2 \Lambda) \beta_2^* + \mathbf{u}^*$$

que puede interpretarse como si se hubiese efectuado un cambio de escala en todas las variables (Caso 1), seguido de una traslación de las variables expli-

cativas (Caso 2). Aplicando primero el resultado obtenido en el Caso 1, se tiene la siguiente relación entre los estimadores del subvector $\hat{\beta}_2$:

$$\hat{\beta}_2^* = \lambda \Lambda^{-1} \hat{\beta}_2$$

Por otra parte, utilizando el resultado obtenido en el Caso 2, se tiene:

$$\hat{\beta}_1^* + \mathbf{b}' \hat{\beta}_2^* = \lambda \hat{\beta}_1$$

o, lo que es lo mismo:

$$\hat{\beta}_1^* = \lambda \hat{\beta}_1 - \lambda \mathbf{b}' \Lambda^{-1} \hat{\beta}_2$$

es decir, todos los parámetros estimados son ahora diferentes de los del modelo original.

3.11. ERRORES DE ESPECIFICACION

La lista de variables que el investigador decide incluir como explicativas puede ser errónea por distintas razones: en algún caso puede omitirse una variable explicativa relevante por carecer de datos fiables acerca de la misma o por ignorar su relevancia; otras veces se incluyen, equivocadamente, como explicativas variables que no lo son.

En general, supongamos que el verdadero modelo que describe el comportamiento de la variable y_t es $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, con $E(\mathbf{u}) = \mathbf{0}_T$, $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_T$, pero que el investigador, sin embargo, se propone estimar el modelo $\mathbf{y} = \mathbf{X}_0 \boldsymbol{\beta}_0 + \mathbf{v}$, donde \mathbf{X}_0 es una matriz $T \times p$. La estimación MCO de los coeficientes de este último modelo serán:

$$\begin{aligned} \hat{\beta}_0 &= (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{y} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \\ &= (\mathbf{X}_0' \mathbf{X})^{-1} \mathbf{X}_0' \mathbf{X}\boldsymbol{\beta} + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{u} \end{aligned} \quad [3.50]$$

por lo que se tiene $E(\hat{\beta}_0) = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}\boldsymbol{\beta}$, que será, en general, distinto del vector $\boldsymbol{\beta}$, por lo que el error cometido al escribir la lista de variables explicativas puede conducir a sesgos en el estimador MCO. Más adelante veremos, sin embargo, que en algunos casos dicho estimador continúa siendo insesgado.

Por otra parte, si denotamos por $\hat{\mathbf{v}}$ los residuos del modelo estimado, se tiene:

$$\hat{\mathbf{v}} \hat{\mathbf{v}}' = \mathbf{y}' \mathbf{M}_0 \mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{u})' \mathbf{M}_0 (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \mathbf{u}' \mathbf{M}_0 \mathbf{u} + \boldsymbol{\beta}' \mathbf{X}' \mathbf{M}_0 \mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}' \mathbf{X}' \mathbf{M}_0 \mathbf{u}$$

donde $\mathbf{M}_0 = \mathbf{I}_T - \mathbf{X}_0 (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0'$. Tomando la esperanza matemática de ambos miembros de esta expresión (y utilizando un razonamiento análogo al de la Sección 3.4) se tiene:

$$E(\hat{\mathbf{v}} \hat{\mathbf{v}}') = \sigma_u^2 \text{tr}(\mathbf{M}_0) + \boldsymbol{\beta}' \mathbf{X}' \mathbf{M}_0 \mathbf{X}\boldsymbol{\beta} \quad [3.51]$$

Si no se hubiese cometido ningún error de especificación, la matriz \mathbf{M}_0 sería simplemente $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $E(\hat{v}'\hat{v}) = \sigma_u^2 \text{tr}(\mathbf{M}) = (T - k)\sigma_u^2$ y $\hat{v}'\hat{v}/(T - k)$ sería un estimador insesgado de σ_u^2 . En todos los demás casos la forma cuadrática $\beta' \mathbf{X}' \mathbf{M}_0 \mathbf{X} \beta$ es positiva (véase Problema 3.16), por lo que $E(\hat{v}'\hat{v})$ excederá a $\sigma_u^2 \text{tr}(\mathbf{M}_0)$, siendo $\hat{v}'\hat{v}/(T - p)$ un estimador de σ_u^2 con sesgo positivo.

Es arriesgado comparar distintas especificaciones entre sí, pues la matriz \mathbf{M}_0 cambia de unas a otras, pero el argumento anterior ha conducido en ocasiones a la recomendación de escoger, entre varios modelos posibles, aquel que produzca una suma residual menor. Si la variable endógena es la misma en todos ellos, esto equivale a elegir el modelo con un mayor coeficiente de determinación corregido.

Consideremos ahora algunos casos particulares:

Caso 1. Omisión de variables relevantes.

Supongamos que la matriz \mathbf{X} está formada por $\mathbf{X} = (\mathbf{X}_0; \mathbf{Z})$, donde \mathbf{X}_0 es $T \times r$ y \mathbf{Z} es $T \times (k - r)$, es decir, que el modelo que se estima excluye alguna de las variables explicativas del modelo verdadero. En tal caso:

$$\begin{aligned} (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{X} &= (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 (\mathbf{X}_0; \mathbf{Z}) = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} (\mathbf{X}'_0 \mathbf{X}_0; \mathbf{X}'_0 \mathbf{Z}) = \\ &= [\mathbf{I}_r; (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{Z}] \end{aligned}$$

y, en consecuencia, de [3.50]:

$$E(\hat{\beta}_0) = [\mathbf{I}_r; (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{Z}] \beta = \beta_0 + (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{Z} \beta_z$$

vector de dimensión $r \times 1$, donde β_0 y β_z denotan, respectivamente, los verdaderos vectores de coeficientes de las variables en \mathbf{X}_0 y \mathbf{Z} .

Como consecuencia de la omisión de las variables explicativas, las estimaciones MCO de los coeficientes de las variables incluidas en \mathbf{X}_0 son sesgadas.

Como caso especial, si bien improbable, supongamos que las variables \mathbf{Z} y \mathbf{X}_0 son ortogonales, es decir, $\mathbf{Z}'\mathbf{X}_0 = \mathbf{0}_{(k-r) \times r}$. En tal caso es inmediato ver que la exclusión de las variables que componen \mathbf{X} no genera ningún sesgo en el estimador MCO, como ya vimos en el Teorema 3.1. En un modelo en desviaciones con respecto a la media, la condición anterior equivaldría al supuesto de que el coeficiente de correlación entre cualquier variable en \mathbf{X}_0 y cualquier variable en \mathbf{X} es cero.

En general, el sesgo producido por omisión de variables viene dado por el producto de dos factores:

- a) Una matriz que tiene por columnas las estimaciones de mínimos cuadrados de los coeficientes de las regresiones de las variables explicativas excluidas de la regresión (es decir, aquellas en \mathbf{Z}) sobre las incluidas (es decir, aquellas en \mathbf{X}_0).
- b) El vector de coeficientes que, en el modelo verdadero, tienen las variables excluidas \mathbf{Z} .

Por ejemplo, supongamos que en vez del modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

el investigador especifica $y_t = \beta_1 + \beta_2 x_{2t} + u_t$ omitiendo, por error, la variable x_{3t} . Si denotamos por $\hat{\beta}_1$ y $\hat{\beta}_2$ las estimaciones MCO de este último modelo, se tendrá:

$$E \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \beta_1 + a\beta_3 \\ \beta_2 + c\beta_3 \end{pmatrix} \quad \text{donde} \quad \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} T & \Sigma_1^T x_{2t} \\ \Sigma_1^T x_{2t} & \Sigma_1^T x_{2t}^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \Sigma_1^T x_{3t} \\ \Sigma_1^T x_{2t} x_{3t} \end{pmatrix}$$

La segunda cuestión importante a considerar es el efecto que sobre la estimación de σ_u^2 tendrá el error de especificación que consideramos. En este caso se tiene $\text{tr}(\mathbf{M}_0) = T - r$, de modo que:

$$E \left(\frac{\hat{v}'\hat{v}}{T-r} \right) = \sigma_u^2 + \frac{\beta' \mathbf{X}' \mathbf{M}_0 \mathbf{X} \beta}{T-r} = \sigma_u^2 + \frac{\beta_z' \mathbf{Z}' \mathbf{M}_0 \mathbf{Z} \beta_z}{T-r} \quad [3.52]$$

por lo que la estimación de σ_u^2 excederá a su verdadero valor ya que $\beta' \mathbf{X}' \mathbf{M}_0 \mathbf{X} \beta \geq 0$ (véase Problema 3.16). Nótese que el sesgo en la estimación del parámetro σ_u^2 persistirá incluso si las variables explicativas excluidas fuesen ortogonales a las incluidas.

Caso 2. Inclusión de variables irrelevantes.

Como caso dual del anterior, supongamos que el investigador comete el error de incluir como explicativas variables que no pertenecen al modelo. Es decir, en este caso, $\mathbf{X}_0 = (\mathbf{X}; \mathbf{Z})$, donde \mathbf{X} es la matriz habitual $T \times k$ de observaciones de las variables explicativas, mientras que \mathbf{Z} es $T \times s$. En este caso, en vez de especificar el modelo correcto $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ se ha especificado el modelo:

$$\mathbf{y} = \mathbf{X}\beta_1 + \mathbf{Z}\beta_2 + \mathbf{v}$$

Como $(\mathbf{X}'_0 \mathbf{X}_0)^{-1} (\mathbf{X}'_0 \mathbf{X}_0) = \mathbf{I}_{k+s} = (\mathbf{X}'_0 \mathbf{X}_0)^{-1} (\mathbf{X}'_0 \mathbf{X}; \mathbf{X}'_0 \mathbf{Z})$ se tiene:

$$(\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mathbf{X}'_0 \mathbf{X} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{s \times k} \end{pmatrix}$$

y, a partir de [3.50]:

$$E \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{s \times k} \end{pmatrix} \beta = \begin{pmatrix} \beta \\ \mathbf{0}_s \end{pmatrix}$$

es decir, que el estimador MCO de los coeficientes de las variables \mathbf{X} es insesgado, mientras que el estimador de los coeficientes de las variables \mathbf{Z} (incorrectamente incluidas) tiene esperanza cero. Por tanto, debería esperarse que a.

estimar el modelo, los coeficientes de estas últimas variables resultasen no significativos.

En segundo lugar, una vez estimado el modelo $y = \mathbf{X}\beta_1 + \mathbf{Z}\beta_2 + v$, el investigador procedería a estimar el parámetro σ_u^2 dividiendo la suma residual por el número de grados de libertad del modelo estimado, es decir, mediante la expresión:

$$\hat{\sigma}_u^2 = \frac{\hat{v}'\hat{v}}{T - k - s}$$

Ahora bien, por las propiedades de la matriz \mathbf{M}_0 (véase Capítulo 1) se tiene $\mathbf{M}_0\mathbf{X} = \mathbf{M}_0\mathbf{Z} = \mathbf{0}_T$, por lo que a partir de [3.51] se tiene inmediatamente que:

$$E(\hat{v}'\hat{v}) = \sigma_u^2 \text{tr}(\mathbf{M}_0) = (T - k - s)\sigma_u^2$$

por lo que el estimador de σ_u^2 es insesgado en este caso.

Puesto que la inclusión de variables irrelevantes no sesga el estimador MCO, a diferencia de lo que ocurre si omitimos variables relevantes, podría considerarse razonable una estrategia de introducir un alto número de variables explicativas en el modelo de regresión. Sin embargo, por razones que quedarán claras en el Capítulo 9, tal estrategia conducirá, de modo inevitable, a aumentar la varianza con que se estiman los coeficientes de las variables explicativas verdaderamente relevantes, sobre los que perdemos, por consiguiente, precisión.

Además, como veremos en el capítulo siguiente, al aumentar la varianza de sus coeficientes podríamos creer equivocadamente que dichas variables no son importantes para explicar la evolución de y , cuando lo que ocurre es que la incorporación de variables irrelevantes nos hace perder precisión en la estimación de los coeficientes de todas ellas.

En definitiva, el investigador debe ser muy cuidadoso al incluir o excluir variables explicativas de un modelo econométrico. Para ayudar en tales decisiones se utilizan procedimientos de inferencia estadística, que describimos en el próximo capítulo.

PROBLEMAS

Problema 3.1. Dadas las observaciones muestrales:

y	x_2	x_3
1	1	2
1	0	-1
-1	1	0
0	0	-1
-1	1	2

- a) Obtener los estimadores MCO de los coeficientes del modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

- b) Estimar σ_u^2 y la matriz $\text{Var}(\hat{\beta})$.
 c) Calcular el coeficiente de determinación.
 d) Comprobar que la suma de los residuos es cero y también que $X'\hat{u} = \mathbf{0}_3$.
 e) Obtener el intervalo de confianza del 95 por 100 para σ_u^2 .

Problema 3.2. Probar que en un modelo econométrico con término independiente la suma explicada puede obtenerse, indistintamente, mediante $\hat{\beta}'_2 X'_2 Q X_2 \hat{\beta}_2$ o $\hat{\beta}' X' X \hat{\beta} - T\bar{y}^2$, donde Q y $\hat{\beta}_2$ responden a la notación introducida en la Sección 3.7.

Problema 3.3. Encontrar las ecuaciones normales, así como los estimadores MCO de los parámetros del modelo:

$$y_t = \alpha + \beta \frac{1}{x_t} + u_t$$

Probar que las expresiones obtenidas coinciden con las que se habrían calculado tras hacer el cambio de variable, $z_t = \frac{1}{x_t}$, y utilizar MCO en el modelo:

$$y_t = \alpha + \beta z_t + u_t$$

Problema 3.4. Obtener el estimador de máxima verosimilitud del modelo $y_t = \alpha + \beta x_t + u_t$, $t = 1, 2, \dots, T$. En particular, probar que de las condiciones de optimización de la función de verosimilitud se tiene:

$$\begin{aligned}\hat{\beta}_{MV} &= \frac{S_{xy}}{S_x^2} \\ \hat{\alpha}_{MV} &= \bar{y} - \hat{\beta}_{MV} \bar{x} \\ \hat{\sigma}_{MV}^2 &= \frac{\hat{u}_{MV}' \hat{u}_{MV}}{T}\end{aligned}$$

donde \bar{x} e \bar{y} denotan las medias muestrales de x_t e y_t , S_{xy} es la covarianza muestral de x e y , mientras que S_x^2 es la varianza muestral de x .

Problema 3.5. Para estudiar la relación entre las variables x e y se han estimado los modelos:

- a) $y_i = \alpha + \beta x_i + u_i$.
 b) $\ln y_i = \alpha + \beta x_i + u_i$.
 c) $y_i = \alpha + \beta \ln x_i + u_i$.
 d) $\ln y_i = \alpha + \beta \ln x_i + u_i$.

Discutir la interpretación que tendría, en cada caso, el valor estimado del coeficiente β .

Problema 3.6. Probar que la estimación MCO del coeficiente β en el modelo $y_i = \alpha + \beta x_i + u_i$ es el inverso del estimador MCO del coeficiente δ en el modelo $x_i = \gamma + \delta y_i + v_i$ si y sólo si el coeficiente de determinación del primer modelo (y del segundo) es igual a 1.

Problema 3.7. Probar que en la regresión particionada de la Sección 3.7 no es preciso descontar de y el efecto de Z , puesto que la regresión de y sobre el vector u_x genera, en realidad, el mismo estimador θ que se obtuvo en dicha sección.

Problema 3.8. Comentar acerca de la aplicación del resultado de la regresión particionada a la práctica siguiente: para estimar el modelo

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \delta t + u_i$$

donde t denota una tendencia lineal se elimina la tendencia de las variables y_i , x_i mediante regresiones:

$$\begin{cases} y_i = at + \varepsilon_{y_i} \\ x_{it} = b_i t + \varepsilon_{x_{it}}, \quad i = 1, 2, \dots, k \end{cases}$$

y a continuación se estima el modelo

$$\varepsilon_{y_i} = \boldsymbol{\varepsilon}'_i \boldsymbol{\beta} + u_i$$

donde $\boldsymbol{\varepsilon}'_i = (\varepsilon_{x_{i1}}, \dots, \varepsilon_{x_{ik}})$.

¿Qué procedimiento alternativo sugiere el resultado de la Sección 3.7?

Problema 3.9. Para estimar el modelo $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ se dispuso de diez observaciones, a partir de las cuales se calcularon las matrices de momentos muestrales respecto a la media:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1,0 & -1,5 \\ -1,5 & 2,5 \end{pmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 16 \\ 9 \end{pmatrix}$$

obteniendo una suma residual de 5,2. Las medias muestrales fueron $\bar{y} = 4$, $\bar{x}_2 = 3$, $\bar{x}_3 = 5$. Recuperar las estimaciones MCO de los coeficientes β_1 , β_2 y β_3 , así como su matriz de varianzas y covarianzas.

Problema 3.10. En un estudio de los determinantes de la inversión se utilizaron veinte observaciones anuales. Las variables utilizadas fueron:

$$y_i = \alpha + \beta x_i + \delta z_i + u_i$$

donde:

y_i : Inversión (en billones de pesetas) en el año i .

x_i : Tipo de interés (en porcentajes) en el año i .

z_i : Variación anual en el Producto Interior Bruto en el año i (en billones de pesetas).

A partir de la muestra utilizada se obtuvieron los siguientes momentos muestrales:

$$\begin{array}{lll} \Sigma x_i = 100 & \Sigma (x_i - \bar{x})^2 = 9 & \Sigma (x_i - \bar{x})(y_i - \bar{y}) = -14 \\ \Sigma y_i = 5 & \Sigma (y_i - \bar{y})^2 = 60 & \Sigma (z_i - \bar{z})(y_i - \bar{y}) = 7 \\ \Sigma z_i = 24 & \Sigma (z_i - \bar{z})^2 = 1 & \Sigma (x_i - \bar{x})(z_i - \bar{z}) = -1 \end{array}$$

- Estimar la regresión de y_i sobre x_i y z_i .
- ¿Qué porcentaje de la evolución temporal de la inversión puede explicarse por la influencia lineal de los tipos de interés y las variaciones en el producto?
- ¿Puede decirse sin ninguna ambigüedad qué tipos de interés elevados conducen a un nivel de inversión bajo?
- A tipos del 10 por 100 y con una variación anual de 2 billones de pesetas en el Producto Interior Bruto, ¿qué puede esperarse del nivel anual de la inversión?

Problema 3.11. A partir de una muestra con veinte observaciones de cuatro variables (y, x_1, x_2, x_3) se calculó la siguiente matriz de covarianzas muestrales:

	y	x_1	x_2	x_3
y	30	12	15	27
x_1	12	14	10	0
x_2	15	10	16	0
x_3	27	0	0	50

1. Considerando la regresión de y sobre x_1, x_2, x_3 , predecir el efecto que sobre y tendrán las siguientes variaciones:

- $\Delta x_1 = 1; \quad \Delta x_2 = -1; \quad \Delta x_3 = 0.$
- $\Delta x_1 = 1; \quad \Delta x_2 = -1; \quad \Delta x_3 = 2.$

2. Utilizando la regresión y sobre x_1 y x_2 , predecir el efecto que sobre la variable y tendrán variaciones como las propuestas en 1.a). Comparar los resultados con los allí obtenidos.

3. Utilizando la regresión de y sobre x_3 , predecir el efecto que sobre y tendrán los incrementos sugeridos en 1.b).

4. Establecer la bondad del ajuste en los casos 1, 2 y 3 anteriores.

Problema 3.12. Obtener las expresiones [3.46] para las varianzas y covarianzas de los estimadores MCO de los parámetros del modelo $y_t = \alpha + \beta x_t + u_t$, de dos modos alternativos:

- Mediante la expresión matricial $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$.
- Obteniendo $\text{Var}(\hat{\beta})$ del modelo en diferencias y $\text{Var}(\hat{\alpha})$ a partir de la expresión: $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$.
- Probar que $\text{SR} = T \left(S_y^2 - \frac{S_{xy}^2}{S_x^2} \right)$ y $\text{SE} = T \frac{S_{xy}^2}{S_x^2}$.

Problema 3.13. Encontrar las expresiones [3.49] que relacionan los estimadores del modelo de 3 variables con los coeficientes de correlación simples entre las variables de dicho modelo.

Problema 3.14. Probar que el estimador MCO del único parámetro en la regresión de la variable $\bar{y} - \hat{\beta}_{13}\bar{x}_3$ sobre la variable $\bar{x}_2 - \hat{\beta}_{23}\bar{x}_3$ es igual a $\hat{\beta}_2$, el estimador MCO del coeficiente de x_2 en la regresión de y sobre x_2 y x_3 , donde $\bar{y}, \bar{x}_2, \bar{x}_3, \hat{\beta}_{13}, \hat{\beta}_{23}$ se corresponden con la notación de la Sección 3.9.

Obtenga esta misma conclusión a partir del resultado de la regresión particionada.

Problema 3.15. Probar que el coeficiente de determinación del modelo de regresión con tres variables (una de ellas constante) pueden escribirse:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}$$

[Sugerencia: Partir de la expresión $R^2 = \frac{\hat{\beta}'X'y}{ST}$ para el modelo en diferencias y utilizar [3.49]. El cociente anterior es válido para el cálculo del R^2 , ya que hay un término independiente en el modelo.]

Problema 3.16. Para probar que el producto $\beta'X'M_0X\beta$ que aparece en la Sección 3.11 es no negativo, basta probar que la matriz $X'M_0X$ es semidefinida positiva. Lleve a cabo tal demostración comprobando que $X'M_0X$ es la matriz de covarianzas de los vectores de residuos del conjunto de regresiones que tienen a las variables X por variables endógenas y a las variables X_0 por explicativas.

Problema 3.17. Demostrar que si en el modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t$$

se omite la variable x_{kt} y esta variable es ortogonal a las variables $x_{2t}, x_{3t}, \dots, x_{k-1t}$, el estimador MCO del término independiente será sesgado y su sesgo igual al producto $\bar{x}_k \beta_k$.

Problema 3.18. Demostrar que si en vez del verdadero modelo

$$y_t = X\beta + \delta z_t + u_t$$

se especifica el modelo

$$y_t = X\beta + \gamma s_t + v_t$$

se tiene $E(\hat{\beta}) = \beta + a\delta$ y $E(\hat{\gamma}) = b\delta$, donde a (vector con $k-1$ componentes) y b (escalar) contienen los coeficientes estimados en una regresión de z_t sobre las variables en X junto con la variable s_t .

Problema 3.19. a) Obtenga las expresiones analíticas de los estimadores MCO de los parámetros β_2 y β_3 en el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$. Si denotamos por $\hat{\beta}_{12}$, $\hat{\beta}_{13}$, $\hat{\beta}_{23}$ y $\hat{\beta}_{32}$ las estimaciones MCO de los coeficientes de las regresiones:

$$\begin{aligned} y_t &= \alpha_1 + \beta_{12} x_{2t} + v_{1t} \\ y_t &= \alpha_2 + \beta_{13} x_{3t} + v_{2t} \\ x_{2t} &= \alpha_3 + \beta_{23} x_{3t} + v_{3t} \\ x_{3t} &= \alpha_4 + \beta_{32} x_{2t} + v_{4t} \end{aligned}$$

b) Pruebe que

$$\hat{\beta}_2 = \frac{\hat{\beta}_{12} - \hat{\beta}_{13}\hat{\beta}_{32}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}}$$

y obtenga una expresión similar para $\hat{\beta}_3$.

c) Se tiene la matriz de momentos muestrales con respecto a la media:

	y_t	x_{2t}	x_{3t}
y_t	10	6	8
x_{2t}	6	5	2
x_{3t}	8	2	15

donde, por ejemplo, $T^{-1} \sum_1^T (x_{2t} - \bar{x}_2)(x_{3t} - \bar{x}_3) = 2$.

Utilizando esta información, calcule $\hat{\beta}_2$ y $\hat{\beta}_3$ de acuerdo con las expresiones en a) y b) y compruebe su equivalencia.

d) ¿Bajo qué condiciones será $\hat{\beta}_2 = \hat{\beta}_{12}$? Interprete su afirmación.

e) Calcule los coeficientes de correlación parcial entre y_t y x_{2t} , así como entre y_t y x_{3t} , con los datos de la tabla anterior, y compare con los valores de los correspondientes coeficientes de correlación simples.

Problema 3.20. Probar que en el modelo $y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$, donde las variables están en diferencias con respecto a la media, la varianza de $\hat{\beta}_2$ puede escribirse:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum_1^T x_{2t}^2 (1 - r_{23}^2)}$$

y también que:

$$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = - \frac{\sigma_u^2 r_{23}^2}{(1 - r_{23}^2) \sum_1^T x_{2t} x_{3t}}$$

donde r_{23} es el coeficiente de correlación entre x_{2t} y x_{3t} , y discutir qué implicaciones tendría el que $r_{23} = 1$.

Problema 3.21. Demostrar que la media muestral de la variable endógena, \bar{y} , es independiente del estimador MCO del subvector β_2 que excluye el término independiente.

Problema 3.22. a) Explique la influencia del error de especificación sobre el sesgo del estimador de mínimos cuadrados ordinarios.

b) Suponga que en vez del *modelo correcto*:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_3 + \beta_4 X_{4t} + u_t$$

el investigador, *erróneamente*, especifica:

$$Y_t = \alpha_1 + \alpha_2 X_{3t} + v_t$$

Calcule el estimador MCO de α_1 , α_2 y σ_v^2 . Evalúe el sesgo de este estimador si se dispone de la matriz de productos cruzados y promedios muestrales:

	y_t	x_{2t}	x_{3t}	x_{4t}
y_t	6000	100	-200	300
x_{2t}		500	0	100
x_{3t}			100	-400
x_{4t}				200

$$\begin{cases} \bar{X}_2 = 4 \\ \bar{X}_3 = -2 \\ \bar{X}_4 = 1 \\ \bar{Y} = 1 \end{cases}$$

donde $x_{it} = X_{it} - \bar{X}_i$, $i = 2, 3, 4$ e $y_t = Y_t - \bar{Y}$, calculadas a partir de una muestra de 20 observaciones. (Por ejemplo, $\sum_1^{20} x_{3t}y_t = -200$; $\sum_1^{20} x_{4t}^2 = 200$.)

c) Evalúe el sesgo en el estimador del parámetro σ_u^2 .

Problema 3.23. Compruebe que las expresiones [3.48] y [3.49] se satisfacen en el Ejemplo 3.5.

Problema 3.24. Repita el Problema 3.1 transformando las variables en diferencias respecto a la media muestral.

Problema 3.25. Reproduzca un razonamiento similar al utilizado en la demostración del teorema de Gauss-Markov para probar el siguiente resultado: «La combinación lineal $c'\hat{\beta}$, donde $\hat{\beta}$ denota al estimador MCO del vector β , es el estimador lineal insesgado de mínima varianza de la combinación lineal $c'\beta$ ».

Problema 3.26. Suponga que las variables explicativas del modelo lineal general $y_t = x_t'\beta + u_t$ pueden descomponerse en dos subvectores x_1 , x_2 , con la propiedad de que la covarianza $S_{x_1, x_2} = 0$ para todo par de variables x_i de x_1 y x_j de x_2 .

Demuestre que las estimaciones mínimo-cuadráticas de los subvectores de coeficientes β_1 y β_2 en el modelo anterior coinciden numéricamente con las que se obtendrían de los modelos parciales:

$$\begin{cases} y_t = x_{1t}'\beta_1 + u_{1t} \\ y_t = x_{2t}'\beta_2 + u_{2t} \end{cases}$$

Problema 3.27. Halle las condiciones para que el estimador MV de σ_u^2 tenga menor Error Cuadrático Medio que su estimador MCO.

Problema 3.28. ¿Cuál de las siguientes matrices no pueden ser consideradas como matrices producto $X'X$ del modelo $y_t = x_t'\beta + u_t$?

$$X'X = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}; \quad X'X = \begin{pmatrix} 4 & 8 \\ 8 & 2 \end{pmatrix}; \quad X'X = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix}; \quad X'X = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}; \quad X'X = \begin{pmatrix} -4 & 0 \\ 0 & 2 \end{pmatrix}$$

Problema 3.29. a) Demuestre que si el verdadero modelo econométrico que relaciona una variable endógena con los subvectores de variables exógenas x_1 y x_2 es $y = X_1\beta_1 + X_2\beta_2 + u$, donde X_1 y X_2 son matrices $T \times k_1$, $T \times k_2$, y se estima por error el modelo $y = X_1\beta_1 + v$, el estimador MCO de β_1 es sesgado.

b) Demuestre que, sin embargo, el estimador MCO del vector β_1 en el modelo $y = X_1^*\beta_1 + \omega$, donde X_1^* son los residuos de las regresiones que tienen al vector X_1 por variables a explicar y a X_2 como variables explicativas, es insesgado.

Problema 3.30. Para estimar el modelo $y_i = \beta x_i + u_i$ se propone el estimador:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\frac{\sigma^2}{\beta^2} + \sum_{i=1}^n x_i^2}$$

- a) Pruebe que dicho estimador está sesgado hacia cero.
b) Pruebe que

$$E(\hat{\beta} - \beta)^2 = \frac{\sigma^2}{\frac{\sigma^2}{\beta^2} + \sum_{i=1}^n x_i^2}$$

- c) Pruebe que su varianza es inferior a la del estimador MCO.

Problema 3.31 (Johnston 2.2). Suponga que para estimar el modelo $y_t = \alpha + \beta x_t + u_t$ se dispone de las observaciones:

$t =$	1	2	3	4	5	6
x_t	1	2	3	4	5	6

y se propone utilizar el estimador $\hat{\beta} = \frac{1}{8} (y_6 + y_5 - y_2 - y_1)$.

Calcule su varianza muestral y compárela con la del estimador MCO.

Problema 3.32. Disponiendo de las dos observaciones:

$i =$	1	2
y_i	4	4
x_i	1	-2

obtenga las suma residual y total y el coeficiente de determinación de los modelos

- a) $y_i = \beta x_i + u_i$
b) $y_i = \beta_0 + \beta_1 x_i + v_i$

e interprete los resultados obtenidos.

Problema 3.33. Para estimar el modelo lineal simple $y_i = \alpha + \beta x_i + u_i$, $i = 1, 2, \dots, N$, un investigador dispone de una muestra clasificada en dos submuestras A y B . En la submuestra B , el investigador tiene las observaciones de y_i , pero no las de x_i . Para poder utilizar toda la muestra en la estimación del modelo, el investigador decide sustituir los valores ausentes de la submuestra B por su promedio en A , \bar{x}_A .

a) Pruebe que el estimador MCO que obtendrá de esta manera coincide con el que obtendría utilizando tan sólo la submuestra A , a pesar de lo cual:

b) El coeficiente de determinación R^2 será inferior al que habría obtenido utilizando tan sólo la submuestra A .

Problema 3.34. Suponga que para estimar el modelo lineal simple $y_i = x_i\beta + u_i$ se dispone de una muestra de tamaño N . En la primera parte de la muestra, de tamaño N_A , se observan ambas variables, mientras que en la segunda submuestra, de tamaño $N_B = N - N_A$, sólo se observa y_i . Suponga que se decide sustituir los valores no observados de x_i por ceros, a la vez que añadir al modelo una variable ficticia que toma el valor 0 cuando x_i se observa y el valor 1 cuando x_i no se observa.

Caracterice al estimador MCO, comparándolo con el resultante de la propuesta alternativa del problema anterior.

Problema 3.35. Con objeto de estimar el modelo de regresión lineal simple $Y_t = \alpha + \beta X_t + u_t$, se han propuesto los siguientes estimadores de β :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum Y_t}{\sum X_t} \\ \hat{\beta}_2 &= \frac{1}{T} \cdot \sum \frac{Y_t}{X_t} \\ \hat{\beta}_3 &= \frac{\sum X_t Y_t}{\sum X_t^2} \\ \hat{\beta}_4 &= \frac{\sum y_t}{\sum x_t} \\ \hat{\beta}_5 &= \frac{1}{T} \cdot \sum \left(\frac{y_t}{x_t} \right) \\ \hat{\beta}_6 &= \frac{\sum x_t y_t}{\sum x_t^2}\end{aligned}$$

donde letras minúsculas indican diferencias entre los valores representados por las mayúsculas y sus respectivos promedios muestrales. Todas las sumas anteriores son desde $t = 1$ hasta $t = T$, donde T es el tamaño muestral. Calcular la esperanza y la varianza de cada estimador y sugerir cuál de ellos debería utilizarse.

Problema 3.36. a) Demuestre que la varianza del estimador de mínimos cuadrados ordinarios $\hat{\beta}_2$ del coeficiente β_2 en el modelo

$$y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad [1]$$

donde las variables están en desviaciones con respecto a la media, puede expresarse:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum_1^T x_{2t}^2 (1 - r_{23}^2)}$$

donde r_{23} es el coeficiente de correlación entre x_{2t} y x_{3t} .

b) Obtenga la esperanza y varianza del estimador MCO del coeficiente δ_2 en el modelo *incorrectamente especificado*:

$$y_t = \delta_2 x_{2t} + v_t \quad [2]$$

- c) ¿Bajo qué condiciones será insesgado este estimador de δ_2 ?
- d) Obtenga los errores cuadráticos medios de los estimadores de β_2 y δ_2 descritos en a) y b).
- e) Pruebe que la relación entre dichos errores cuadráticos medios puede escribirse:

$$\frac{\text{ECM}(\hat{\delta}_2)}{\text{ECM}(\hat{\beta}_2)} = 1 + r_{23}^2(t^2 - 1)$$

donde t es el estadístico de Student para la significación del estimador MCO de β_3 en el modelo [1].

Problema 3.37. Considere una regresión en que las variables explicativas son deterministas y linealmente independientes, y el término de error u_t se distribuye $N(0, \sigma_u^2)$.

Suponga que una de las variables explicativas es $x_{it} = \lambda^i$, con $|\lambda| < 1$. Muestre que el estimador MCO del coeficiente de dicha variable es inconsistente, aunque los coeficientes de las restantes variables se estiman consistentemente.

[Ayuda: Resuelva inicialmente el problema para el caso de un término constante y λ^1 .]

CAPITULO 4

INFERENCIA EN EL MODELO LINEAL

1. INTRODUCCION

Hemos visto en capítulos anteriores que una de las tareas fundamentales del trabajo econométrico es la de aportar un conocimiento descriptivo de una economía real. Uno de los métodos por los que se consigue dicho conocimiento es mediante la contrastación de supuestos alternativos que la Teoría Económica efectúa acerca de la estructura de relación entre variables que estamos analizando empíricamente. Por ejemplo, supongamos que para representar la tecnología existente en un determinado sector industrial se especifica una función de producción Cobb-Douglas en los distintos inputs que se utilizan en dicho sector: trabajo, L_t ; capital, K_t , y energía, E_t . Como ya explicamos en la introducción a este texto, con un término de error multiplicativo dicha función puede transformarse en lineal sin más que tomar logaritmos:

$$\ln Q_t = a + \alpha \ln K_t + \beta \ln L_t + \gamma \ln E_t + u_t$$

donde Q_t es el nivel de producto y el término independiente a es el logaritmo de la constante tecnológica que aparece en la función de producción Cobb-Douglas. Los coeficientes α , β y γ son entonces las elasticidades de los distintos inputs. La Teoría Económica considera tres situaciones posibles con una tecnología de este tipo, según que haya rendimientos a escala crecientes, constantes o decrecientes. Estas tres situaciones vienen caracterizadas por la condición de que la suma de las tres elasticidades sea mayor, igual o menor que 1.

Conocer en cuál de las situaciones opera la industria es importante porque cada una de ellas tiene una implicación diferente acerca del incremento en producción que se obtendrá si se multiplican por alguna constante las cantidades de los inputs que se vienen utilizando. Por tanto, una vez estimados los parámetros α , β y γ , tiene interés contrastar ciertas hipótesis acerca del valor

de la suma de dichas elasticidades. Obtener los métodos necesarios para llevar a cabo estos contrastes es parte del objetivo de este capítulo.

Una diferencia importante con respecto a lo desarrollado hasta ahora es que vamos a mantener a lo largo del capítulo el supuesto de que el término de error sigue en cada período (o, en general, para cada observación) una distribución Normal. Ya hemos explicado en el Capítulo 3 que debido a que el estimador MCO es una transformación lineal del vector \mathbf{u} , entonces si el vector \mathbf{u} se distribuye $N_T(\mathbf{0}_T; \sigma_u^2 \mathbf{I}_T)$, se tiene que $\hat{\boldsymbol{\beta}}$ sigue una distribución $N_k(\boldsymbol{\beta}; \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1})$.

Recordemos asimismo (Proposición 3.9) que:

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{M}\mathbf{u}$$

donde $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y, por consiguiente:

$$\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\mathbf{y} - \mathbf{M}\mathbf{u}) - \mathbf{X}\boldsymbol{\beta} = \mathbf{u} - \mathbf{M}\mathbf{u} = \mathbf{N}\mathbf{u}$$

donde $\mathbf{N} = \mathbf{I}_T - \mathbf{M}$, al igual que \mathbf{M} , es una matriz simétrica e idempotente, de dimensión $T \times T$. Por ser idempotente, el rango de \mathbf{N} es igual a su traza y se tiene:

$$\text{Rango}(\mathbf{N}) = \text{traza}(\mathbf{N}) = \text{tr}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \text{tr}(\mathbf{I}_k) = k$$

donde se ha utilizado la propiedad circular de la traza de matrices (Sección 1.1). Es claro entonces que $\text{Rango}(\mathbf{M}) = T - k$, por lo que, de acuerdo con los resultados de la Sección 2.8, se tiene que la forma cuadrática

$$\frac{1}{\sigma_u^2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sigma_u^2} \mathbf{u}'\mathbf{N}\mathbf{u}$$

sigue una distribución chi-cuadrado con k grados de libertad, mientras que

$$(T - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} = \frac{1}{\sigma_u^2} \hat{\mathbf{u}}'\hat{\mathbf{u}} = \frac{1}{\sigma_u^2} \mathbf{u}'\mathbf{M}\mathbf{u}$$

sigue una distribución chi-cuadrado con $T - k$ grados de libertad.

Además, las matrices \mathbf{M} y \mathbf{N} son ortogonales, es decir, $\mathbf{M}\mathbf{N} = \mathbf{0}_{T \times T}$, por lo que las dos formas cuadráticas anteriores son independientes entre sí. Como consecuencia, se tiene el siguiente resultado:

Proposición 4.1. Los estimadores MCO del vector $\boldsymbol{\beta}$ y del parámetro σ_u^2 son independientes entre sí.

Demostración. Notemos que en la primera de las formas cuadráticas anteriores el único elemento aleatorio que aparece es el vector de estimadores $\hat{\boldsymbol{\beta}}$, mientras que la segunda es proporcional a $\frac{\hat{\sigma}_u^2}{\sigma_u^2}$, donde σ_u^2 es una constante

desconocida. Como acabamos de ver, ambas formas cuadráticas son independientes, pero ello sólo puede ocurrir si el vector aleatorio $\hat{\beta}$ es independiente de la variable aleatoria $\hat{\sigma}_u^2$.

Nótese que esto ocurre a pesar de que los vectores $\hat{\beta}$ y \hat{u} no son independientes ($\hat{u} = \mathbf{y} - \mathbf{X}\hat{\beta}$). Aun siendo una función del vector de residuos \hat{u} , el estimador $\hat{\sigma}_u^2$ resume la información estadística contenida en dicho vector de tal manera que resulta independiente del vector $\hat{\beta}$.

4.2 CONTRASTE DE HIPOTESIS: UN TRATAMIENTO INTRODUCTORIO

De esta discusión se obtiene una conclusión adicional importante, y es que de acuerdo con la Sección 2.8 se tiene que el cociente

$$\frac{\frac{\mathbf{u}'\mathbf{Nu}}{k}}{\frac{\mathbf{u}'\mathbf{Mu}}{T-k}} = \frac{(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)/k}{\frac{\hat{u}'\hat{u}}{T-k}} = (\hat{\beta} - \beta)' [\hat{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1}]^{-1} (\hat{\beta} - \beta)/k \quad [4.1]$$

sigue una distribución $F_{k, T-k}$, por ser el cociente de dos variables chi-cuadrado independientes.

Este resultado ya nos permite contrastar algunas hipótesis sencillas acerca de los valores de los coeficientes del modelo. Por ejemplo, supongamos que se hace la hipótesis de que el vector de parámetros β toma unos valores numéricos concretos, β^0 , es decir, se establece la hipótesis nula $H_0: \beta = \beta^0$. Supongamos ahora que estimamos por MCO dicho vector de parámetros y se obtiene la serie de residuos correspondiente y la estimación de σ_u^2 . Si además sustituimos el vector β en [4.1] por el vector β^0 de la hipótesis nula, todos los componentes de dicho estadístico son conocidos, por lo que podría evaluarse numéricamente y comparar con el valor tabulado de la distribución $F_{k, T-k}$.

4.2.a. Interpretación del estadístico F

Veamos cuál es la interpretación del procedimiento que acabamos de describir para llevar a cabo el contraste de la hipótesis $H_0: \beta = \beta^0$. Si la hipótesis nula fuese correcta, entonces el vector de estimaciones $\hat{\beta}$ no debería ser «muy diferente» del vector β^0 especificado en dicha hipótesis. Es decir, esperaríamos que el vector diferencia $(\hat{\beta} - \beta^0)$ fuese «pequeño». La cuestión por dilucidar es lo «pequeña» que puede ser esta diferencia o, en otras palabras, qué distancia puede aceptarse entre el vector β^0 que se ha especificado en la hipótesis nula y el vector estimado $\hat{\beta}$.

Como vimos en la Sección 2.3, cuando una variable aleatoria tiene esperanza cero, tiene sentido utilizar su desviación típica como un indicador del «tamaño» de dicha variable⁽¹⁾. Comprobemos que esto es lo que se hace en el estadístico [4.1].

En primer lugar, observamos que el estadístico [4.1] está formado por la norma del vector diferencia⁽²⁾ $\hat{\beta} - \beta^0$, donde las coordenadas están ponderadas de acuerdo con la inversa de la matriz de covarianzas estimada del vector $\hat{\beta}$. Para comprender mejor el significado de esta operación, supongamos por un momento que la matriz de covarianzas fuese diagonal, es decir, que las variables explicativas fuesen ortogonales entre sí ($\sum_1^T x_{it}x_{jt} = 0, \forall i \neq j, i, j = 1, \dots, k$). En tal caso, el estadístico F anterior constaría de la suma de cuadrados de las diferencias entre los parámetros estimados ($\hat{\beta}_i$) y sus valores en la hipótesis nula (β_i^0), cada una de ellas dividida por la varianza de $\hat{\beta}_i$:

$$F = \frac{1}{k} \sum_{i=1}^k \frac{(\hat{\beta}_i - \beta_i^0)^2}{\text{Var}(\hat{\beta}_i)}$$

De este modo, una diferencia $\hat{\beta}_i - \beta_i^0$ es tanto más importante en el estadístico F cuanto mayor sea la precisión con que se ha estimado el parámetro β_i , es decir, cuanto menor sea su varianza. Si un parámetro se estima con poca precisión, entonces valores relativamente apreciables de la diferencia $\hat{\beta}_i - \beta_i^0$ recibirán una menor ponderación en el cálculo del estadístico F .

Ahora bien, bajo la hipótesis nula, β^0 es el verdadero valor del vector β y, como tal, es constante, por lo que la matriz de covarianzas de la diferencia $\hat{\beta} - \beta^0$ coincide con la matriz de covarianzas de $\hat{\beta}$, cuya inversa aparece en [4.1]. Así, la expresión [4.1] consiste en dividir la norma del vector diferencia por un indicador de su tamaño, para decidir si dicha diferencia es grande o no.

Consecuentemente, si el valor de nuestro estadístico es menor que el valor de las tablas de la distribución $F_{k, T-k}$, entonces diremos que la diferencia entre $\hat{\beta}$ y β^0 es pequeña, por lo que no rechazaremos la hipótesis establecida H_0 , mientras que la rechazaremos en caso contrario.

¿Qué ocurriría si en la hipótesis nula se hace un supuesto acerca del valor de un solo coeficiente del modelo?⁽³⁾ Una consecuencia de que $\hat{\beta}$ se distribuya $N_k(\beta; \sigma_u^2(X'X)^{-1})$ es que, en particular, $\hat{\beta}_i$ se distribuye $N(\beta_i; \sigma_u^2 a_{ii})$, donde a_{ii} es el elemento i -ésimo en la diagonal principal de la matriz $(X'X)^{-1}$. Por tanto,

(1) Puesto que es una medida promedio de las desviaciones con respecto a la esperanza matemática, cuando ésta es cero, la desviación típica puede interpretarse como el tamaño promedio de la variable aleatoria.

(2) Recordemos que la *norma euclídea* de un vector es la raíz cuadrada de la suma de los cuadrados de sus coordenadas.

(3) Por ejemplo, éste sería el caso si se quisiese contrastar que una de las variables inicialmente incluidas en el modelo es realmente significativa, en cuyo caso estableceríamos la hipótesis $H_0: \beta_i = 0$, donde x_i es la variable de cuya significación o relevancia dudamos.

$$\frac{\hat{\beta}_i - \beta_i}{\sigma_u \sqrt{a_{ii}}} \quad [4.2]$$

se distribuye como una $N(0, 1)$.

El problema, como es habitual, es que el parámetro σ_u^2 es desconocido. Sin embargo, ya hemos visto al comienzo del capítulo que $\frac{(T-k)\hat{\sigma}_u^2}{\sigma_u^2} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2}$ se distribuye como una chi-cuadrado con $T-k$ grados de libertad e independientemente de cada uno de los parámetros estimados $\hat{\beta}_i$. Por tanto (Sección 2.8), el cociente de la variable $N(0, 1)$ dada en [4.2] y la raíz cuadrada de la variable chi-cuadrado dividida por sus grados de libertad

$$\frac{\frac{\hat{\beta}_i - \beta_i}{\sigma_u \sqrt{a_{ii}}}}{\sqrt{\frac{(T-k)\hat{\sigma}_u^2}{\sigma_u^2}} / (T-k)} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}}$$

se distribuye como una t de Student con $T-k$ grados de libertad. Es importante observar que la única diferencia entre esta expresión y [4.2] es la sustitución del parámetro desconocido σ_u por su valor estimado $\hat{\sigma}_u$. Esta sustitución implica un cambio en la distribución de probabilidad del cociente que consideramos, que hemos pasado de comparar con una distribución $N(0, 1)$ a compararlo con una distribución t_{T-k} . Recuérdese también que cuando el número de grados de libertad $T-k$ tiende a infinito ambas distribuciones coinciden.

El estadístico F en [4.1] podría adaptarse al contraste de otras restricciones diferentes de las que aquí hemos analizado, pero ello requiere un esfuerzo individualizado de adaptación del estadístico a cada contraste específico. Para evitar esta situación, sería conveniente disponer de un estadístico que permitiera efectuar, sin necesidad de adaptación previa, el contraste de cualquier conjunto de hipótesis lineales. Esto es lo que hacemos en la sección siguiente.

4.3. CONTRASTE DE HIPOTESIS: TRATAMIENTO GENERAL

4.3.a. La formulación del problema

Es el momento de desarrollar un método general de contrastación de hipótesis lineales, extendiendo los ejemplos que acabamos de discutir en dos direcciones. Por una parte, el procedimiento que vamos a obtener nos permitirá contrastar hipótesis que afirman que una determinada combinación lineal de

los parámetros del modelo es igual a una cierta constante, por ejemplo: $H_0: 3\beta_2 - 2\beta_5 = 4$. El análisis que vamos a llevar a cabo permite además contrastar que varias de estas hipótesis son *simultáneamente* ciertas. Las hipótesis que hemos mencionado en la sección anterior son de este tipo, sólo que con coeficientes muy sencillos, y se obtienen como casos particulares de este contraste más general.

Si, por ejemplo, se quisiese proceder al contraste conjunto de las hipótesis:

$$H_0: \begin{cases} \beta_2 = 0 \\ \beta_4 = -1 \\ \beta_5 = 3,10 \end{cases}$$

en el modelo $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + u_t$, definiendo la matriz \mathbf{R} y el vector \mathbf{r} por:

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} 0 \\ -1 \\ 3,10 \end{pmatrix}$$

se tiene la siguiente expresión matricial para el conjunto de restricciones anterior: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

Si se tuviese en cambio el conjunto de restricciones:

$$H_0: \begin{cases} 7\beta_3 - \beta_5 = 12 \\ \beta_4 + \beta_1 = 0 \\ 3\beta_5 - \beta_2 = 6 \end{cases}$$

entonces las matrices relevantes son:

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & 7 & 0 & -1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 3 \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} 12 \\ 0 \\ 6 \end{pmatrix}$$

y las restricciones pueden denotarse por $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. Nótese que cada fila de \mathbf{R} está formada por los coeficientes de los parámetros $\boldsymbol{\beta}$ en cada una de las restricciones.

El lector puede considerar cualquier otro conjunto de hipótesis, quizá más complejas, y tratar de encontrar las matrices \mathbf{R} y \mathbf{r} que permitan traducir su conjunto de hipótesis a la formulación general $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. Las matrices \mathbf{R} y \mathbf{r} siempre existen, supuesto que todas las restricciones sean lineales.

Consideremos ahora un conjunto de q restricciones lineales genérico: $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, donde \mathbf{R} es una matriz $q \times k$, \mathbf{r} un vector columna de dimensión q , y ambos \mathbf{R} y \mathbf{r} son conocidos. Supongamos además que $q \leq k$ y también que $\text{Rango}(\mathbf{R}) = q$.

Conviene comentar brevemente acerca de estos supuestos. Hemos supuesto que el número de hipótesis que se contrastan, q , es inferior o igual a k , el número de variables del modelo, por lo que el rango de la matriz \mathbf{R} , de dimensión $q \times k$, será $\leq q$. Al suponer que dicho rango es exactamente igual a q y no inferior, estamos suponiendo que las restricciones son linealmente independientes, es decir, no hay en el conjunto dos restricciones como $\beta_2 - 2\beta_3 = 4$ y $6\beta_3 - 3\beta_2 + 12 = 0$.

4.3.b. El estadístico F para el contraste de cualquier conjunto de hipótesis lineales

Comencemos ahora el desarrollo analítico de la metodología general para el contraste de hipótesis lineales, obteniendo un resultado similar al utilizado en la sección anterior. Para ello utilizamos los resultados 5 y 6 de la Sección 2.8.

Puesto que $\hat{\boldsymbol{\beta}}$ se distribuye $N_k(\boldsymbol{\beta}; \sigma_u^2(\mathbf{X}'\mathbf{X}^{-1}))$, entonces, por tener la matriz \mathbf{R} rango q , se tiene que el vector $\mathbf{R}\hat{\boldsymbol{\beta}}$, de dimensión q , se distribuye $N_q(\mathbf{R}\boldsymbol{\beta}; \sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')$ y, como consecuencia, $\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ se distribuye $N_q(\mathbf{0}; \sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')$.

Ahora bien, si la hipótesis H_0 es cierta, entonces $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, y por tanto se tiene que $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ se distribuye $N_q(\mathbf{0}; \sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')$. Finalmente se tiene que

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \quad [4.3]$$

se distribuye como una chi-cuadrado con q grados de libertad, es decir, el número de restricciones. El problema habitual de que el parámetro σ_u^2 sea desconocido se resuelve utilizando la propiedad de que $\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{\sigma_u^2}$ se distribuye como una chi-cuadrado independiente de la forma cuadrática anterior, que depende únicamente del vector aleatorio $\hat{\boldsymbol{\beta}}$. Consecuentemente se tiene que el cociente

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k}} \quad [4.4]$$

se distribuye como una $F_{q, T-k}$. Dos observaciones importantes acerca de esta expresión [4.4]:

- a) El parámetro desconocido σ_u^2 ya no aparece en el estadístico F .
- b) Puesto que el denominador del estadístico [4.4] no es sino la estimación MCO del parámetro σ_u^2 , en realidad el resultado que acabamos de obtener puede expresarse afirmando que el estadístico

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\hat{\sigma}_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q$$

sigue una distribución $F_{q, T-k}$, lo que se presta a una interpretación paralela a la que se hizo en la Sección 4.2. Así, ya que queremos medir la proximidad entre el vector $\mathbf{R}\hat{\boldsymbol{\beta}}$ y el vector \mathbf{r} , entonces la hipótesis H_0 se debe rechazar si el valor numérico del estadístico [4.4] calculado a partir de las estimaciones MCO supera el valor dado por las tablas para la distribución F con grados de libertad $(q, T - k)$.

Conviene observar que el denominador del estadístico [4.4] depende del vector de estimaciones $\hat{\boldsymbol{\beta}}$, pero no del verdadero valor de los parámetros, y siempre tiene una distribución chi-cuadrado. Sin embargo, como hemos visto, el numerador sólo sigue una distribución chi-cuadrado si las restricciones $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ son ciertas, es decir, sólo si la hipótesis nula es cierta. En todos los otros casos tendrá también una distribución chi-cuadrado, pero no central. Por mantenerse la independencia de ambas distribuciones chi-cuadrado, el cociente seguiría teniendo una distribución F , pero en este caso no centrada.

Esto sugiere otra forma de interpretar la contrastación de hipótesis: Si la hipótesis nula es cierta, la distribución del estadístico [4.4] es una F centrada. Por tanto, la mayor parte de su probabilidad estará concentrada cerca del origen. Si obtenemos en nuestra aplicación un valor del estadístico alejado de cero y, por tanto, poco probable, entonces pensaremos que lo que ocurre es que en nuestra aplicación el estadístico [4.4] no tiene una distribución centrada, lo que equivale a decir que la hipótesis nula no es cierta.

Ejemplo 4.1. Para estimar el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$ se dispone de las matrices:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

que han sido calculadas a partir de las 90 observaciones muestrales. También se obtuvo $\mathbf{y}'\mathbf{y} = \sum_1^{90} y_t^2 = 80$. La estimación MCO de los coeficientes del modelo es:

$$y_t = 11 - 7x_{2t} + 12x_{3t} + 3x_{4t} + \hat{u}_t$$

(1,6) (1,7) (1,4) (1,4)

donde los valores en paréntesis son las desviaciones típicas estimadas de los coeficientes, utilizando la estimación de σ_u^2 :

$$\hat{\sigma}_u^2 = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{T - k} = \frac{80 - 37}{86} = \frac{1}{2}$$

Supongamos que se quiere contrastar la hipótesis $H_0: \beta_3 + 2\beta_2 = 3$. Para esta restricción se tiene $\mathbf{R} = (0, 2, 1, 0)$ y $\mathbf{r} = 3$, y por tanto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = 20$

y $R\hat{\beta} - r = -5$, y el estadístico F toma el valor
$$\frac{(-5) \frac{1}{20} (-5)}{\frac{1}{2}} = 2,5.$$
 El valor

de las tablas de la variable $F_{1,86}$ al 95 por 100 de confianza es 3,95, por lo que no rechazamos la hipótesis de que los coeficientes β_2 y β_3 satisfacen la restricción anterior. El valor de las tablas al 99 por 100 de confianza es 6,92, por lo que tampoco se rechaza la hipótesis a este nivel de confianza.

Imaginemos ahora que los momentos muestrales anteriores hubieran sido obtenidos a partir de una muestra de 47 observaciones. En tal caso, el número de grados de libertad del modelo sería 43, y la estimación de σ_u^2 sería $\hat{\sigma}_u^2 = 1,0$.

El estadístico F tomaría el valor
$$\frac{(-5) \frac{1}{20} (-5)}{1} = 1,25,$$
 por lo que la evidencia muestral sería aún más clara en favor de no rechazar la hipótesis nula. Ello se debe a que, con menos información, las estimaciones son menos precisas (las varianzas estimadas de los coeficientes serían ahora mayores, ya que $\hat{\sigma}_u^2$ ha pasado de $\frac{1}{2}$ a 1). En consecuencia, los intervalos de confianza son mayores, y es más difícil rechazar una hipótesis, ya que el rango de valores admisibles ha aumentado.

Por el contrario, si los momentos muestrales se hubiesen calculado a partir de 176 observaciones, entonces el número de grados de libertad sería de 172,

$\hat{\sigma}_u^2 = \frac{1}{4}$, y el estadístico F sería
$$\frac{(-5) \frac{1}{20} (-5)}{\frac{1}{4}} = 5,$$
 por lo que se rechazaría

la hipótesis nula al 95 por 100 de confianza, aunque no se rechazaría al 99 por 100 de confianza.

4.3.c. Un procedimiento alternativo

De acuerdo con la interpretación que dimos anteriormente para el estadístico F , el contraste de hipótesis puede también llevarse a cabo examinando la «holgura» de la restricción, es decir, el margen por el que las estimaciones de los coeficientes dejan de satisfacer las restricciones.

Ejemplo 4.2. En nuestro ejemplo $\hat{\beta}_3 + 2\hat{\beta}_2 - 3 = -5$, y se trata de decidir si esta holgura ($= -5$) es «grande» o «pequeña» en el caso $T = 90$. Para ello es preciso compararla con su varianza, es decir, con:

$$\text{Var}(\hat{\beta}_3 + 2\hat{\beta}_2 - 3) = \text{Var}(\hat{\beta}_3) + 4 \text{Var}(\hat{\beta}_2) + 4 \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{1}{2} (4 + 24 - 8) = 10$$

donde hemos extraído algunos elementos de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. El cociente entre la norma de la holgura (es decir, su cuadrado): 5^2 y su varianza, 10, sigue una distribución $F_{1,86}$. Pero dicho cociente, 2,5, coincide con el valor del estadístico F que obtuvimos antes.

Supongamos ahora que se quisieran contrastar simultáneamente las dos hipótesis: $H_0: \beta_3 + 2\beta_2 = 3$ y $\beta_4 = 6$. Las matrices \mathbf{R} y \mathbf{r} son:

$$\mathbf{R} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

y se tiene:

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = \begin{pmatrix} 20 & -5 \\ -5 & 4 \end{pmatrix} \quad \mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \begin{pmatrix} -5 \\ -3 \end{pmatrix}$$

y el estadístico F toma el valor:

$$\frac{(-5, -3) \frac{1}{55} \begin{pmatrix} 4 & 5 \\ 5 & 20 \end{pmatrix} \begin{pmatrix} -5 \\ -3 \end{pmatrix}}{2 \cdot \frac{1}{2}} = 7,818 \quad [4.5]$$

Los valores tabulados de la variable $F_{2,86}$ al 95 y 99 por 100 de confianza son, respectivamente, 3,10 y 4,86, por lo que la hipótesis nula se rechaza claramente, tanto al 95 como al 99 por 100 de confianza.

Este contraste de hipótesis puede también llevarse a cabo comparando el vector de «holguras» de ambas restricciones, con su matriz de covarianzas. Para ello debemos obtener la varianza de la segunda restricción: $\text{Var}(\hat{\beta}_4 - 6) = \text{Var}(\hat{\beta}_4) = \frac{1}{2} \cdot 4 = 2$, así como la covarianza entre ambas restricciones:

$$\text{Cov}(\hat{\beta}_3 + 2\hat{\beta}_2 - 1, \hat{\beta}_4 - 6) = \text{Cov}(\hat{\beta}_3, \hat{\beta}_4) + 2 \text{Cov}(\hat{\beta}_2, \hat{\beta}_4) = \frac{1}{2} (3 - 8) = -\frac{5}{2}$$

y como la forma cuadrática

$$\left(\begin{array}{c} \text{Vector de} \\ \text{holguras} \end{array} \right)' \left(\begin{array}{c} \text{Matriz de covarianzas} \\ \text{de las restricciones} \end{array} \right)^{-1} \left(\begin{array}{c} \text{Vector de} \\ \text{holguras} \end{array} \right)$$

sigue una distribución $F_{q, T-k}$, se tiene en nuestro caso el estadístico:

$$(-5, -3) \left[\frac{1}{2} \begin{pmatrix} 20 & -5 \\ -5 & 4 \end{pmatrix} \right]^{-1} \begin{pmatrix} -5 \\ -3 \end{pmatrix}$$

que coincide con [4.5].

4.4. APLICACION A ALGUNOS CASOS PARTICULARES

Habiendo obtenido en la sección anterior un estadístico F para el contraste de cualquier conjunto de restricciones lineales, vamos a derivar en esta sección su expresión particular cuando se trata de contrastar algunas de las hipótesis más habituales en el trabajo aplicado.

4.4.a. Contraste de hipótesis acerca de un coeficiente del modelo

Si la hipótesis nula contiene únicamente una hipótesis acerca del valor de un solo coeficiente, como en $H_0: \beta_i = \beta_i^0$, entonces la matriz \mathbf{R} se reduce al vector fila $[0, 0, \dots, 1, 0, \dots, 0]$, donde el número 1 ocupa la posición i -ésima. En este caso, $\mathbf{r} = \beta_i^0$ y $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ es simplemente igual a la diferencia $\hat{\beta}_i - \beta_i^0$, mientras que el producto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ se reduce a a_{ii} , el elemento i -ésimo en la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Es decir, un vector \mathbf{R} como el antes especificado, al multiplicar por la izquierda y la derecha una matriz cuadrada, simplemente «selecciona» el elemento i -ésimo en su diagonal principal. Así, en este caso particular, el estadístico [4.4] se convierte en

$$\frac{(\hat{\beta}_i - \beta_i^0)^2}{\hat{\sigma}_u^2 a_{ii}}$$

con una distribución $F_{1, T-k}$, o lo que es lo mismo, la raíz cuadrada de esta variable aleatoria

$$\frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_u \sqrt{a_{ii}}}$$

sigue una distribución t_{T-k} , resultado que coincide con el que obtuvimos en la Sección 4.2.a para el contraste de esta hipótesis. Por supuesto que si en este caso el ahora escalar \mathbf{r} es cero, entonces el estadístico es:

$$\frac{\hat{\beta}_i}{\hat{\sigma}_u \sqrt{a_{ii}}}$$

que seguirá una distribución t_{T-k} .

Este cociente, que es el apropiado para contrastar la significación de la variable explicativa x_i en el modelo econométrico, se suele conocer como *estadístico t del coeficiente estimado $\hat{\beta}_i$* .

4.4.b. Contraste de hipótesis acerca de todos los coeficientes del modelo (excepto el término independiente)

Supongamos que se quiere contrastar la hipótesis nula $H_0: \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$, donde $\boldsymbol{\beta}_2^0$ es el vector de dimensión $k - 1$ formado por todos los coeficientes, excepto

el término independiente. En este contexto, las matrices \mathbf{R} y \mathbf{r} tienen la estructura $\mathbf{R} = [\mathbf{0}_{k-1} \mathbf{I}_{k-1}]$, $\mathbf{r} = \beta_2^0$, donde β_2^0 , así como $\mathbf{0}_{k-1}$, son vectores columna de dimensión $k-1$ e \mathbf{I}_{k-1} es la matriz identidad:

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} = [\mathbf{0}_{k-1} \mathbf{I}_{k-1}]$$

Por consiguiente, el vector $\mathbf{R}\hat{\beta} - \mathbf{r}$ se reduce a la diferencia entre el subvector columna de estimaciones $(\hat{\beta}_2, \dots, \hat{\beta}_k)$ y el vector β_2^0 . Por otra parte, la matriz \mathbf{X} puede escribirse como $\mathbf{X} = [\mathbf{1}_T \mathbf{X}_2]$, donde $\mathbf{1}_T$ es un vector columna de unos y \mathbf{X}_2 denota la matriz de observaciones de todas las variables explicativas, excepto el término independiente, por lo que se tiene:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} T & \mathbf{1}'_T \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{1}_T & \mathbf{X}'_2 \mathbf{X}_2 \end{pmatrix}$$

En este caso, el producto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ es igual a la matriz que resulta de eliminar la primera fila y la primera columna de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Esta submatriz, de orden $(k-1) \times (k-1)$, es igual a (de acuerdo con los resultados de la Sección 1.3 acerca de la inversión de matrices particionadas) $(\mathbf{X}'_2 \mathbf{X}_2 - \mathbf{X}'_2 \mathbf{1}_T \frac{1}{T} \mathbf{1}'_T \mathbf{X}_2)^{-1}$, es decir, igual a $(\mathbf{X}'_2 \mathbf{Q} \mathbf{X}_2)^{-1}$, donde $\mathbf{Q} = \mathbf{I}_T - \frac{1}{T} \mathbf{1}_T \mathbf{1}'_T$ es la matriz que introdujimos en el Capítulo 1. Con todo ello, el estadístico F resulta ser:

$$\frac{(\hat{\beta}_2 - \beta_2^0)' (\mathbf{X}'_2 \mathbf{Q} \mathbf{X}_2) (\hat{\beta}_2 - \beta_2^0) / (k-1)}{\frac{\hat{u}'\hat{u}}{T-k}} \quad [4.6]$$

que sigue una distribución $F_{k-1, T-k}$. Si el valor numérico del cociente [4.6] supera al de las tablas $F_{k-1, T-k}$, concluimos que la estimación $\hat{\beta}_2$ difiere significativamente del hipotético β_2^0 y rechazamos la hipótesis nula, manteniéndola en caso contrario.

4.4.c. El contraste de significación global del modelo econométrico

Si la hipótesis que se contrasta es la significación conjunta de los coeficientes β_2 , es decir, $H_0: \beta_2 = \mathbf{0}_{k-1}$, entonces el estadístico [4.6] se convierte en:

$$\frac{\beta_2' (\mathbf{X}'_2 \mathbf{Q} \mathbf{X}_2) \hat{\beta}_2 / (k-1)}{\frac{\hat{u}'\hat{u}}{T-k}}$$

con distribución $F_{k-1, T-k}$.

Recordemos que el término independiente relaciona las medias muestrales de la variable endógena y las variables explicativas, estas últimas multiplicadas por sus coeficientes estimados. La significación global de un modelo no debe depender de la significación del término independiente. En efecto, incluso si todos los coeficientes (excepto el término independiente) resultasen no significativos, entonces el término independiente sería, aproximadamente, la media de la variable endógena, y podría perfectamente resultar significativo. Por tanto, el contraste que acabamos de analizar es el de la significación global del modelo.

En este caso, utilizando la expresión derivada en la Sección 3.8 para la suma explicada, el estadístico F admite una representación alternativa:

$$\frac{\hat{\beta}'_2(\mathbf{X}'_2\mathbf{Q}\mathbf{X}_2)\hat{\beta}_2/(k-1)}{\frac{\hat{u}'\hat{u}}{T-k}} = \frac{\frac{\text{Suma explicada}}{k-1}}{\frac{\text{Suma residual}}{T-k}} = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$$

con distribución $F_{k-1, T-k}$. Nótese que estas igualdades son válidas únicamente porque se supone que el modelo contiene un término independiente. Es importante observar que, por tanto, la práctica habitual de contrastar la significación global del modelo mediante su coeficiente de determinación R^2 sólo está justificada cuando el modelo incluye un término independiente.

Si el valor numérico del estadístico supera al de las tablas $F_{k-1, T-k}$, concluimos que el vector $\hat{\beta}_2$ es significativamente distinto de $\mathbf{0}_{k-1}$, luego el modelo es globalmente válido. Nótese que ello es equivalente a un R^2 alto, por ser la función $R^2/(1-R^2)$ creciente en R^2 .

4.4.d. Contraste acerca de un subvector de s variables ($1 \leq s \leq k-1$)

Supongamos ahora que la hipótesis que se contrasta es la significación conjunta de un grupo de s variables explicativas que, sin pérdida de generalidad, supondremos que son las últimas variables explicativas del modelo. Para ello, elijamos una matriz $\mathbf{R} = [\mathbf{0}_{s, k-s}; \mathbf{I}_s]$ y un vector $\mathbf{r} = \mathbf{0}_s$ y consideremos el contraste del conjunto de restricciones lineales $\mathbf{R}\beta = \mathbf{r}$.

Comencemos particionando la matriz $\mathbf{X} = [\mathbf{X}_1; \mathbf{X}_2]$, donde \mathbf{X}_1 es ahora de dimensión $T \times (k-s)$ y \mathbf{X}_2 es $T \times s$. Concretamente, las columnas de \mathbf{X}_2 son las observaciones muestrales de las variables cuya significación contrastamos. Particionemos también el vector de parámetros $\beta = (\beta_1, \beta_2)$, donde β_1 y β_2 son los coeficientes en el modelo de las variables que están incluidas en las submatrices \mathbf{X}_1 y \mathbf{X}_2 , respectivamente. Entonces el modelo econométrico puede escribirse:

$$\mathbf{y} = (\mathbf{X}_1; \mathbf{X}_2) \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \hat{u} = \mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2 + \hat{u} \quad [4.7]$$

Dada la estructura de la matriz \mathbf{R} y del vector \mathbf{r} en este caso se tiene:

$$\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \hat{\boldsymbol{\beta}}_2$$

Además, el producto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ selecciona la submatriz (s, s) en la esquina inferior derecha de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Recordemos que para obtener tal submatriz no es correcto hallar la inversa de la correspondiente submatriz de $\mathbf{X}'\mathbf{X}$, sino que deben utilizarse las expresiones para los bloques de las inversas de matrices particionadas. De acuerdo con aquellos resultados, si denotamos por \mathbf{B} a la submatriz $s \times s$ que buscamos se tiene:

$$\mathbf{B} = (\mathbf{X}'_2\mathbf{X}_2 - \mathbf{X}'_2\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2)^{-1} = (\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}$$

donde $\mathbf{M}_1 = \mathbf{I}_T - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$ y, por tanto, el estadístico F se transforma en este caso en

$$\frac{\hat{\boldsymbol{\beta}}'_2(\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)\hat{\boldsymbol{\beta}}_2/s}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}} \quad [4.8]$$

con distribución $F_{s, T-k}$, y el contraste de la hipótesis nula se resuelve como en los casos anteriores.

Nada en este razonamiento hace que sea preciso que $\hat{\boldsymbol{\beta}}_2^0 = \mathbf{0}_{k-1}$. Si no fuese cero, entonces $\mathbf{R}\hat{\boldsymbol{\beta}}_2 - \mathbf{r} = \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0$ y el estadístico

$$\frac{(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)'\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)/s}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}} \quad [4.9]$$

tiene una distribución $F_{s, T-k}$.

Caso particular. $s = k - 1$.

Es claro que el número de variables cuya significación contrastamos puede tomar cualquier valor entre 1 y $k - 1$. Supongamos que $s = k - 1$. Entonces $\mathbf{X}_1 \equiv \mathbf{1}_T$ y la matriz \mathbf{M}_1 se convierte en:

$$\mathbf{M}_1 = \mathbf{I}_T - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1 = \mathbf{I}_T - \mathbf{1}_T(\mathbf{1}'_T\mathbf{1}_T)^{-1}\mathbf{1}'_T = \mathbf{I}_T - \mathbf{1}_T\frac{1}{T}\mathbf{1}'_T = \mathbf{Q}$$

por lo que los estadísticos [4.8] y [4.9] coinciden en este caso con los que obtuvimos en la Sección 4.4.b.

Caso especial. $s = 1$.

Supongamos ahora que se quiere contrastar la significación de un solo parámetro, es decir, $s = 1$. Ordenemos las variables explicativas de forma que

la variable cuya significación contrastamos sea la k -ésima, es decir, la última variable del modelo. Entonces $\mathbf{X}'_2 \mathbf{M}_1 \mathbf{X}_2$ es de dimensión 1×1 y es el inverso del elemento (k, k) en la matriz $(\mathbf{X}'\mathbf{X})^{-1}$, que venimos denotando por a_{kk} , por lo que el estadístico [4.8] se reduce en este caso a:

$$\frac{\hat{\beta}_k^2 \left(\frac{1}{a_{kk}} \right)}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k}} = \frac{\hat{\beta}_k^2}{\hat{\sigma}_u^2 a_{kk}}$$

con distribución $F_{1, T-k}$, que es precisamente el mismo estadístico que ya obtuvimos en la Sección 4.4.a para este contraste de hipótesis, puesto que $(t_{T-k})^2 = F_{1, T-k}$.

4.5. CONTRASTES DE SIGNIFICACION MEDIANTE SUMAS RESIDUALES

Hemos visto en estos ejemplos anteriores la particularización del estadístico F introducido en [4.4] para el contraste de un conjunto de hipótesis lineales. Todos estos contrastes utilizan el vector de coeficientes estimados, así como el vector de residuos mínimo-cuadráticos. Vamos a probar a continuación que existe un procedimiento alternativo para llevar a cabo los contrastes de hipótesis lineales, que no utiliza los coeficientes estimados, sino tan sólo la suma residual del modelo estimado dos veces: en un caso, sin imponer las restricciones que aparecen en la hipótesis nula y, en el otro, restringido por las mismas.

En este momento vamos a llevar a cabo tal discusión únicamente para el caso en que el contraste de hipótesis es un *contraste de significación conjunta* de un subvector de coeficientes del modelo, es decir, el caso tratado en la Sección 4.4.d. En secciones posteriores extenderemos el resultado al contraste de cualquier conjunto de hipótesis lineales.

Proposición 4.2. Sea SRS (suma residual sin restringir) la suma residual del modelo en consideración estimado por MCO, es decir, $\text{SRS} = \hat{\mathbf{u}}' \hat{\mathbf{u}}$. Supongamos que se quiere contrastar la hipótesis nula $H_0: \boldsymbol{\beta}_2 = \mathbf{0}_s$, donde $\boldsymbol{\beta}_2$ es un subvector de s variables. Sea SRR (suma residual restringida) la suma residual del modelo estimado excluyendo las variables que figuran en el subvector $\boldsymbol{\beta}_2$. El estadístico F dado por [4.8] para el contraste de esta hipótesis nula puede escribirse:

$$\frac{(\text{SRR} - \text{SRS})/s}{\frac{\text{SRS}}{T - k}} \quad [4.10]$$

con distribución $F_{s, T-k}$.

Demostración. De la Proposición 3.9 sabemos que el vector de residuos de la regresión restringida puede expresarse: $\hat{\mathbf{u}}_1 = \mathbf{M}_1 \mathbf{y}$, donde \mathbf{M}_1 es la matriz que definimos en las secciones anteriores. Premultiplicando la expresión [4.7] por \mathbf{M}_1 se tiene:

$$\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \mathbf{M}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \mathbf{M}_1 \hat{\mathbf{u}} = \mathbf{M}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}} \quad [4.11]$$

donde hemos utilizado la propiedad $\mathbf{M}_1 \mathbf{X}_1 = \mathbf{0}$ y también que, por la ortogonalidad entre residuos y variables explicativas, se tiene:

$$\mathbf{X}' \hat{\mathbf{u}} = (\mathbf{X}_1'; \mathbf{X}_2') \hat{\mathbf{u}} = (\mathbf{X}_1' \hat{\mathbf{u}}; \mathbf{X}_2' \hat{\mathbf{u}}) = (\mathbf{0}_{k-s}; \mathbf{0}_s)$$

lo que implica:

$$\mathbf{M}_1 \hat{\mathbf{u}} = [\mathbf{I}_T - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'] \hat{\mathbf{u}} = \hat{\mathbf{u}}$$

Si trasponemos la expresión [4.11] y la multiplicamos por sí misma, se tiene:

$$\mathbf{y}' \mathbf{M}_1 \mathbf{y} = \hat{\boldsymbol{\beta}}_2' \mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}}' \hat{\mathbf{u}} \quad (\text{puesto que } \hat{\mathbf{u}} = \mathbf{M}_1 \hat{\mathbf{u}} \text{ y } \mathbf{X}_2' \hat{\mathbf{u}} = \mathbf{0}_s)$$

Pero, por otra parte, como vimos en la Proposición 3.9, $\mathbf{y}' \mathbf{M}_1 \mathbf{y}$ es la suma residual del modelo que excluye las s variables últimas, es decir, $\mathbf{y}' \mathbf{M}_1 \mathbf{y} = \text{SRR} = \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1$, por lo que el estadístico F dado por [4.8] puede finalmente escribirse:

$$\frac{(\text{SRR} - \text{SRS})/s}{\frac{\text{SRS}}{T-k}} \quad [4.12]$$

con distribución $F_{s, T-k}$.

La contrastación de hipótesis se lleva a cabo del modo habitual: Si la hipótesis nula fuese cierta, ello equivaldría a decir que las últimas s variables no son relevantes para explicar el comportamiento de la variable y_t . Por consiguiente, sus coeficientes estimados no serían significativos; cuando se estimase una regresión excluyendo estas variables, los coeficientes estimados para las primeras $k - s$ variables no diferirían de los que se obtuvieron con la inclusión de todas las variables.

Como consecuencia, los vectores de residuos de ambas regresiones serían muy similares, y así lo serían también las correspondientes sumas residuales. Por tanto, debe esperarse que el estadístico [4.12] fuese pequeño cuando la hipótesis nula es cierta. Así, si el valor numérico del estadístico F de nuestra aplicación fuese menor que el de las tablas, entonces pensaríamos que ambas sumas residuales son suficientemente similares, es decir, que las últimas s variables no son realmente significativas, por lo que no rechazaríamos la hipótesis nula. Si, por el contrario, nuestro estadístico es mayor que el valor

obtenido en las tablas, entonces pensaremos que las dos sumas residuales son suficientemente diferentes, por lo que debemos rechazar la hipótesis nula de no significatividad de las últimas s variables.

Aunque el resultado anterior se ha demostrado en un caso general, tiene interés analizar su aplicación en algún caso particular:

Caso particular. $s = k - 1$.

Consideremos el caso en que $s = k - 1$, es decir, cuando se contrasta la significación global de un modelo que incluye un término independiente. El modelo restringido tiene entonces como única variable explicativa la variable constante, y como único coeficiente, el término independiente. El estimador MCO de dicho parámetro es la media muestral de la variable y (Sección 3.9, Caso 1), por lo que los residuos son iguales a $y_i - \bar{y}$, las desviaciones de y_i con respecto a su media muestral. Como consecuencia, las sumas residual (SRR) y total (ST) coinciden en este caso. Como, por otra parte, el modelo tiene término independiente, entonces la suma explicada es igual a $ST - SRS$, por lo que la expresión [4.10]

$$\frac{(SRR - SRS)/s}{SRS/(T - k)}$$

se reduce en este caso a

$$\frac{(ST - SRS)/(k - 1)}{SRS/(T - k)} = \frac{SE/(k - 1)}{SRS/(T - k)}$$

con distribución $F_{k-1, T-k}$. Pero éste es precisamente el resultado que obtuvimos en la Sección 4.4.c.

Caso particular. $s = 1$.

El otro caso extremo, cuando $s = 1$, es decir, cuando se quiere contrastar la significación de un solo coeficiente del modelo, se deja como ejercicio al lector.

De acuerdo con los resultados que hemos visto, el contraste de *hipótesis de significación* puede realizarse de dos formas distintas, aunque equivalentes: por una parte, puede estimarse el modelo una sola vez, sin imponer las restricciones, y contrastar mediante el estadístico [4.4] que el conjunto de restricciones propuesto es satisfecho por los coeficientes estimados.

Alternativamente, puede estimarse el modelo dos veces, sin imponer las restricciones (modelo sin restringir: MSR), y también imponiéndolas, es decir, eliminando los coeficientes que aparecen en H_0 (modelo restringido: MR), y efectuar el contraste de dicha hipótesis de significación mediante el estadís-

tico F escrito como función de las sumas residuales de estos dos modelos como se muestra en [4.12]. Este resultado acerca de la equivalencia entre las dos formas de llevar a cabo el contraste de hipótesis, que facilita enormemente el trabajo empírico, puede generalizarse al contraste de cualquier hipótesis del tipo $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, como demostraremos en la Sección 4.8.

4.6. INTERVALOS Y REGIONES DE CONFIANZA

Proporcionar un estimador puntual de un coeficiente desconocido no es una información de excesiva utilidad a no ser que se acompañe de un intervalo de confianza para dicha estimación. Para un nivel de confianza dado es importante conocer el rango de valores admisibles del coeficiente que se estima y no tan sólo qué valor consideramos como más probable. Distinguiremos el caso en que se estima un coeficiente del caso en que se estiman varios.

4.6.a. Intervalo de confianza para un solo coeficiente

Hemos visto en las secciones previas que el estadístico

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}}$$

sigue una distribución t_{T-k} , por lo que si se quiere un intervalo de confianza del $(1 - \alpha)\%$ para el parámetro β_i (es decir, de significación del $\alpha\%$), basta obtener de las tablas de esta distribución el valor λ_α correspondiente, y entonces:

$$1 - \alpha = P \left[-\lambda_\alpha \leq \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \leq \lambda_\alpha \right] = P [\hat{\beta}_i - \lambda_\alpha \hat{\sigma}_u \sqrt{a_{ii}} \leq \beta_i \leq \hat{\beta}_i + \lambda_\alpha \hat{\sigma}_u \sqrt{a_{ii}}]$$

y puesto que $\hat{\beta}_i$, α , λ_α , $\hat{\sigma}_u$ y a_{ii} son todos valores numéricos conocidos, esta expresión nos da el intervalo de confianza que buscamos. Una vez que se ha construido el intervalo, se puede contrastar una hipótesis $H_0: \beta_i = \beta_i^0$ al nivel de significación α sin más que ver si β_i^0 está dentro del intervalo (en cuyo caso no rechazamos la hipótesis nula), o fuera de él (en cuyo caso rechazamos la hipótesis nula).

Por ejemplo, en el modelo estimado en la Sección 4.3.b, los intervalos del 95 por 100 de confianza para cada coeficiente son: (7,84, 14,16), (-10,54, -3,46), (9,17, 14,83), (0,17, 5,83), por lo que todos ellos son significativamente diferentes de cero.

4.6.b. Regiones de confianza para varios coeficientes

Cuando se busca un rango de valores para varios coeficientes simultáneamente, no se tiene un intervalo, sino una región de confianza. El razonamiento es análogo al del caso anterior, pero en vez de utilizar la distribución t_{T-k} como hicimos para un solo coeficiente, utilizaremos el estadístico

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k}} \tag{4.13}$$

que se distribuye como una variable $F_{q, T-k}$.

Como se trata de obtener regiones de confianza acerca de coeficientes y no de restricciones lineales más generales, la matriz \mathbf{R} es la que selecciona dichos coeficientes de entre todo el vector $\boldsymbol{\beta}$, de modo que el estadístico [4.13] se convierte en:

$$\frac{(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)' [(\mathbf{X}'\mathbf{X})_2^{-1}]^{-1} (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)/s}{\hat{\sigma}_u^2}$$

con distribución $F_{s, T-k}$, donde $\boldsymbol{\beta}_2$ denota el subvector $s \times 1$, $s \leq k$ cuya región de confianza se quiere construir y $(\mathbf{X}'\mathbf{X})_2^{-1}$ denota la submatriz $s \times s$ correspondiente, extraída de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Si el subvector $\boldsymbol{\beta}_2$ se coloca al final del modelo, entonces ya vimos en la Sección 4.4.d que $(\mathbf{X}'\mathbf{X})_2^{-1} = (\mathbf{X}'_2 \mathbf{M}_1 \mathbf{X}_2)^{-1}$.

En definitiva, la región de confianza se obtendría mediante:

$$1 - \alpha = P[(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)' (\mathbf{X}'_2 \mathbf{M}_1 \mathbf{X}_2) (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \leq \lambda_\alpha s \hat{\sigma}_u^2] \tag{4.14}$$

donde λ_α es el valor numérico que aparece en las tablas de la distribución $F_{s, T-k}$ al nivel de significación del α %.

En la obtención práctica de tales regiones es muy útil recordar que $\mathbf{X}'_2 \mathbf{M}_1 \mathbf{X}_2$ no es sino la submatriz de $\mathbf{X}'\mathbf{X}$ correspondiente a los coeficientes incluidos en $\boldsymbol{\beta}_2$.

Ejemplo 4.3. Así, si disponemos de un modelo como el estimado en la Sección 4.3.b, la región de confianza conjunta de los coeficientes β_2 y β_3 resulta ser:

$$-0,2043(-7 - \beta_2)^2 + 2(0,4731)(-7 - \beta_2)(12 - \beta_3) - 0,0430(12 - \beta_3)^2 \leq 2\lambda_\alpha \sigma_u^2$$

es decir:

$$-95,6835 - 0,2043\beta_2^2 - 0,0430\beta_3^2 - 14,2146\beta_2 + 7,6554\beta_3 + 0,9462\beta_2\beta_3 \leq \lambda_\alpha$$

El miembro izquierdo de la desigualdad que aparece en [4.14] es un polinomio en el que aparecen tan sólo los coeficientes cuya región de

confianza se construye. La dificultad estriba en que el grado del polinomio es igual al número de tales coeficientes, por lo que la utilidad de estas regiones de confianza se restringe al caso de dos coeficientes. Entonces la región de confianza es una elipse, centrada en el estimador MCO. Al igual que ocurre con los intervalos de confianza para un solo coeficiente, las dimensiones de la elipse dependen directamente de las varianzas con que se han estimado los coeficientes: cuanto mayor sea la precisión de estas estimaciones, menores serán las varianzas, y más pequeña será la elipse de confianza del $(1 - \alpha) \%$. Por último, la inclinación de la elipse depende de la covarianza entre las estimaciones de los dos coeficientes para los que se ha construido: la elipse asciende de izquierda a derecha si tal covarianza es positiva, descendiendo en caso contrario.

Al igual que en el caso de un solo coeficiente, se puede contrastar la hipótesis conjunta $H_0: (\beta_1, \beta_2) = (\beta_1^0, \beta_2^0)$ sin más que examinar si el punto (β_1^0, β_2^0) está dentro o fuera de la región de confianza construida. Se mantiene la hipótesis nula en el primer caso, rechazándola si se da la segunda situación.

4.7. ESTIMACION BAJO RESTRICCIONES

Un aspecto básico de la inferencia estadística que se lleva a cabo en Economía es que el investigador sólo contrasta hipótesis en cuya validez está dispuesto a creer a priori, de modo que si su contraste no las rechaza, entonces pasa a imponerlas en la representación estructural que está considerando.

Supongamos por un momento que se contrastara la hipótesis de que el dinero no tiene ningún efecto sobre el nivel de precios, hipótesis en la que el investigador no cree, y el resultado de dicho contraste es que la hipótesis nula no se rechaza. El investigador se encontraría así en una situación difícil de explicar. Generalmente, cuando los resultados empíricos están en total desacuerdo con las sugerencias de la Teoría Económica, es más conveniente interpretarlos como una «casualidad estadística» que como un nuevo resultado que revolucionará la Ciencia Económica, por estar en contradicción directa con los resultados hasta entonces admitidos.

En definitiva, las características de *potencia* de un contraste hacen que sea siempre posible mantener una hipótesis nula incluso si es falsa; por ello es que sólo tiene sentido contrastar hipótesis nulas *aceptables* desde el punto de vista conceptual.

De esta discusión se deduce que si la hipótesis nula no se rechaza, entonces sería muy interesante disponer de un procedimiento para estimar de nuevo el modelo, pero esta vez imponiendo ese conjunto de hipótesis que hemos contrastado y no rechazado. La idea de eficiencia está ligada a la utilización óptima de toda la información disponible. Si se cree que los coeficientes del modelo satisfacen ciertas restricciones, entonces se ganaría eficiencia introduciendo dichas restricciones en el proceso de estimación.

El objetivo de esta sección es describir el modo en que este proceso puede llevarse a cabo. Vamos a encontrar el estimador que minimice la suma de cuadrados de los residuos, de igual modo que hicimos en la Sección 3.2, pero

esta vez imponiendo las restricciones $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, es decir, que, a diferencia de lo que allí se hizo, se trata esta vez de resolver un problema de optimización sujeto a restricciones lineales. El lagrangiano de tal problema será:

$$L = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2\boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\beta} - \mathbf{r})$$

donde $\boldsymbol{\lambda}$ es un vector de dimensión $q \times 1$ de multiplicadores de Lagrange (tantos como restricciones). Tomando derivadas parciales de este lagrangiano con respecto a $\boldsymbol{\beta}$, así como con respecto a $\boldsymbol{\lambda}$, se tiene:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{R}'\boldsymbol{\lambda} \quad (k \text{ derivadas})$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = -2(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}) \quad (q \text{ derivadas})$$

Igualando a cero estas derivadas parciales y resolviendo el sistema de $k + q$ ecuaciones que así se obtienen resulta:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}_R - \mathbf{X}'\mathbf{y} - \mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}_k \quad [4.15.a]$$

$$\mathbf{R}\hat{\boldsymbol{\beta}}_R - \mathbf{r} = \mathbf{0}_q \quad [4.15.b]$$

y las soluciones a este sistema de ecuaciones, $\hat{\boldsymbol{\beta}}_R$, son el estimador de mínimos cuadrados restringidos (MCR), y el vector de precios sombra (multiplicadores de Lagrange) de las q restricciones. Premultiplicando [4.15.a] por $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$ se tiene:

$$\mathbf{R}\hat{\boldsymbol{\beta}}_R - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}_q \quad [4.16]$$

y finalmente, utilizando la ecuación [4.15.b] (es decir, bajo la hipótesis nula $\mathbf{R}\hat{\boldsymbol{\beta}}_R = \mathbf{r}$) se tiene:

$$\boldsymbol{\lambda} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}}) \quad [4.17]$$

donde $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ es el estimador MCO habitual, es decir, sin imponer ninguna de las restricciones en H_0 . Sustituyendo en [4.15.a] se tiene:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}_R - \mathbf{X}'\mathbf{y} - \mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}}) = \mathbf{0}_k$$

y, finalmente:

$$\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}}) \quad [4.18]$$

donde $\hat{\boldsymbol{\beta}}_R$, la solución (junto con [4.17]) del sistema [4.15], es el estimador de mínimos cuadrados restringidos (MCR) del modelo.

La interpretación de esta expresión es que el estimador MCR es una corrección del estimador sin restringir. El tamaño de dicha corrección viene dado por el segundo sumando en la expresión [4.18]. En él se ve que la corrección será tanto mayor cuanto más lejos esté el estimador MCO no restringido $\hat{\beta}$ de satisfacer las restricciones.

Es importante hacer las siguientes observaciones acerca del estimador MCR:

1. El estimador MCR es insesgado sólo si las restricciones $R\beta = r$ bajo las que se ha obtenido son ciertas (basta tomar esperanzas en [4.18]).

2. El estimador MCR, $\hat{\beta}_R$, satisface las restricciones $R\beta = r$. En efecto, es fácil ver que $R\hat{\beta}_R - r = 0_q$.

3. El estimador MCR difiere del estimador MCO sólo si este último no satisface las restricciones en H_0 (lo que en general ocurrirá). Pero si resulta que el estimador MCO satisface exactamente las restricciones cuya validez se contrasta, entonces el estimador MCR coincidiría con el estimador MCO, como puede verse en [4.18].

4. La matriz de covarianzas del estimador MCR (véase Problema 4.1) es siempre inferior a la matriz de covarianzas del estimador MCO, incluso si las restricciones no son ciertas. Aun pareciendo paradójico, este resultado tiene sentido, pues al imponer las restricciones limitamos la región del espacio paramétrico en la que buscamos el estimador mínimo-cuadrático, por lo que podremos estimarlo con una mayor precisión, aunque quizá con sesgo.

En efecto, la matriz de covarianzas del estimador restringido es:

$$\text{Var}(\hat{\beta}_R) = \sigma_u^2[(X'X)^{-1} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}]$$

y, por tanto:

$$\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_R) = \sigma_u^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

que, como se pide probar en el Problema 4.1, es una matriz semidefinida positiva; como consecuencia, los elementos de la diagonal de la matriz de covarianzas de $\hat{\beta}_R$ son inferiores a los elementos correspondientes de la matriz de covarianzas de $\hat{\beta}$ para una estimación dada del parámetro σ_u^2 . El único problema es que, según se utilicen los residuos calculados con el estimador restringido o sin restringir, la estimación de dicho parámetro será diferente.

Puntualicemos, finalmente, que la imposición de restricciones falsas introduce sesgos en el estimador restringido pero, como acabamos de ver, produce una matriz de covarianzas menor que la del estimador sin dichas restricciones. En términos del Error Cuadrático Medio, la preferencia por uno u otro estimador depende de las situaciones.

Ejemplo 4.4. En el caso del modelo discutido en la Sección 4.3.b, el estimador MCO restringido por $\beta_3 + 2\beta_2 = 3$ y $\beta_4 = 6$, las hipótesis consideradas en la Sección 4.3.c, se obtiene:

$$\hat{\beta}_R = \begin{pmatrix} 11 \\ -7 \\ 12 \\ 3 \end{pmatrix} + \begin{pmatrix} -4 & 0 \\ 10 & -4 \\ 0 & 3 \\ -5 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 55 \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 5 & 20 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} -17/11 \\ -75/11 \\ 183/11 \\ 6 \end{pmatrix}$$

que, como puede verse, satisface las dos restricciones. La estimación del parámetro σ_u^2 es $\hat{\sigma}_u^2 = 0,81$, y la matriz de covarianzas del estimador es:

$$\text{Var}(\hat{\beta}_R) = \frac{0,81}{55} \begin{pmatrix} 211 & -85 & 170 & 0 \\ -85 & 10 & -20 & 0 \\ 50 & 20 & 40 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

cuya última fila y columna son cero porque una de las restricciones fija el valor del parámetro β_4 .

CONTRASTE DE HIPÓTESIS LINEALES MEDIANTE SUMAS RESIDUALES

Ahora estamos en condiciones de generalizar el resultado que obtuvimos en la Sección 4.5 acerca de la posibilidad de llevar a cabo los contrastes de hipótesis lineales de dos formas alternativas, pero equivalentes. Si denotamos por \hat{u}_R el vector de residuos generado por el estimador MCR, se tiene:

$$\begin{aligned} \hat{u}_R &= \mathbf{y} - \mathbf{X}\hat{\beta}_R = \mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{X}(\hat{\beta}_R - \hat{\beta}) = \hat{u} - \mathbf{X}(\hat{\beta}_R - \hat{\beta}) = \\ &= \hat{u} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\beta}) \end{aligned} \quad [4.19]$$

Premultiplicando la expresión [4.19] por su traspuesta y utilizando la propiedad $\mathbf{X}'\hat{u} = \mathbf{0}_k$ se tiene:

$$\text{SRR} = \hat{u}'_R \hat{u}_R = \hat{u}'\hat{u} + (\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})$$

y utilizando esta última igualdad y teniendo en cuenta que $\hat{u}'\hat{u} = \text{SRS}$ el estadístico [4.4] puede escribirse:

$$\frac{\text{SRR} - \text{SRS}}{\frac{q}{\text{SRS}}} \quad [4.20]$$

$$\frac{\text{SRR} - \text{SRS}}{T - k}$$

con distribución $F_{q, T-k}$, por lo que el estadístico F para el contraste de hipótesis lineales *siempre* puede construirse sin más que utilizar las sumas residuales restringida y sin restringir.

La demostración que acabamos de presentar no supone ninguna estructura particular para la matriz \mathbf{R} o para el vector \mathbf{r} , y sirve por tanto para el caso del contraste de significación $\mathbf{r} = \mathbf{0}_q$, así como para casos en que $\mathbf{r} = \boldsymbol{\beta}^0$, y también para una o más restricciones lineales, sean de significación o no, entre los coeficientes del modelo econométrico. Esta demostración utiliza la expresión [4.18] del estimador MCR que, por ser entonces desconocida, no pudimos utilizar en la Proposición 4.2. Sin embargo, debe quedar claro que aquella proposición no es sino un caso particular de la que acabamos de probar, si bien tiene interés didáctico en sí misma.

Ejemplo 4.5. Si tomamos de nuevo la estimación de la función de producción agregada de la economía española, del Ejemplo 3.4: $Y_t = \beta_1 + \beta_2 \ln L_t + \beta_3 \ln K_t + u_t$, podemos contrastar la significación estadística de ambos inputs en la función de producción. Para el contraste de la hipótesis nula: $H_0: \beta_2 = 0$, el estadístico

$$\frac{\hat{\beta}_2}{\sqrt{\text{Var}(\hat{\beta}_2)}}$$

se distribuye como una t de Student con $T - k = 25 - 3 = 22$ grados de libertad. El valor del estadístico anterior en la estimación fue de $0,322/0,106 = 3,04$, que excede de los valores críticos de las tablas de la mencionada distribución t , que son de 1,321 al 90 por 100 de confianza, de 1,717 al 95 por 100 y de 2,508 al 99 por 100 de confianza, por lo que se rechaza la hipótesis nula de no significación estadística del empleo en la producción. Un contraste similar para la significación del stock de capital, que puede llevarse a cabo con los datos que aparecen en el Ejemplo 3.4, conduce a un resultado análogo, de modo aún más claro.

Aceptando la significación estadística de ambos inputs, tiene sentido plantearse el contraste de la hipótesis de rendimientos constantes a escala. Esta es una hipótesis acerca de los dos coeficientes conjuntamente: $H_0: \beta_2 + \beta_3 = 1$, por lo que puede formularse como un contraste F con $\mathbf{R} = (0, 1, 1)$ y $\mathbf{r} = 1$. El estadístico es:

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(T - k)} = \frac{\left[\begin{array}{c} (0, 1, 1) \begin{pmatrix} 0,330 \cdot 10^{-2} \\ 0,322 \\ 0,630 \end{pmatrix} - 1 \end{array} \right]^2}{\left[\begin{array}{c} (0, 1, 1) \begin{pmatrix} 1,1095 \cdot 10^{-4} & -0,1108 & -0,0067 \\ -0,1108 & 0,0112 & 0,0005 \\ -0,0067 & 0,0005 & 0,0002 \end{pmatrix} \begin{pmatrix} 0,1108 \\ 0,0005 \\ 0,0002 \end{pmatrix} \end{array} \right]} \cdot 22$$

cuyo valor numérico, 0,1806, es inferior a los valores críticos de la $F_{1,22}$, que son de: 4,30 y 7,95 al 95 y 99 por 100 de confianza respectivamente, por lo que no se rechaza la hipótesis de existencia de rendimientos constantes a escala.

Hemos discutido la posibilidad de llevar a cabo el contraste de hipótesis lineales mediante comparación de las sumas residuales de los modelos restringido y sin restringir. Si imponemos en el modelo de producción la restricción de rendimientos constantes a escala, tenemos:

$$\ln Y_t = \beta_1 + \beta_2 \ln L_t + (1 - \beta_2) \ln K_t + u_t$$

o, lo que es lo mismo:

$$\ln Y_t - \ln K_t = \beta_1 + \beta_2 (\ln L_t - \ln K_t) + u_t$$

que, estimado, resultó en:

$$\ln Y_t - \ln K_t = -0,444 + 0,367 (\ln L_t - \ln K_t)$$

(0,011) (0,011)

$$T = 25; \quad \bar{R}^2 = 0,978; \quad SR = 0,0137756; \quad \hat{\sigma}_u = 0,0245;$$

$$DW = 0,36; \quad Q(12) = 33,95$$

Comparando este resultado con la estimación del modelo sin restringir, el lector puede comprobar sin dificultad que el estadístico F , calculado a partir de ambas sumas residuales, alcanza el mismo valor numérico que antes proporcionamos, 0,1806.

4.9. CONTRASTE DE HIPOTESIS MEDIANTE SUSTITUCION DE LAS MISMAS

Hemos visto ya que los contrastes de hipótesis lineales pueden llevarse a cabo de dos formas, bien sea mediante el estadístico [4.4], para lo que es preciso únicamente el estimador sin restringir, o mediante las sumas residuales, para lo que es preciso estimar el modelo dos veces, imponiendo y sin imponer las restricciones. Si se escoge la segunda opción, cabe observar que el estimador restringido puede, a su vez, obtenerse por dos procedimientos diferentes. Uno de ellos sería mediante la utilización de la expresión [4.18] que para dicho estimador obtuvimos. Otra posibilidad consiste en utilizar las restricciones para eliminar algunos de los coeficientes del modelo y estimar los restantes.

Ejemplo 4.6. Así, consideremos el modelo

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

donde las variables se hallan en diferencias con respecto a la media, y los productos cruzados calculados a partir de 103 observaciones muestrales:

	y	x_1	x_2	x_3
y	10	3	-3	-2
x_1	3	10	4	-2
x_2	-3	4	8	0
x_3	-2	-2	0	6

que genera las estimaciones MCO: $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \left(\frac{23}{44}, -\frac{28}{44}, -\frac{7}{44}\right)$. La suma residual sin restringir es $SRS = 10 - \left(\frac{23}{44} \cdot 3 + \frac{28}{44} \cdot 3 + \frac{7}{44} \cdot 2\right) = 6,20$.

Consideremos ahora la restricción $\beta_1 + \beta_2 = 0$. La utilización de la expresión [4.18] para el estimador MCO restringido produce como resultado $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)_R = \left(\frac{4}{7}, -\frac{4}{7}, -\frac{1}{7}\right)$. La suma residual restringida es:

$$SRR = \mathbf{y}'\mathbf{y} - (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)_R \mathbf{X}'\mathbf{y} = 10 - \left(\frac{4}{7}, -\frac{4}{7}, -\frac{1}{7}\right) \begin{pmatrix} 3 \\ -3 \\ -2 \end{pmatrix} = \frac{44}{7} = 6,28$$

El valor numérico del estadístico [4.20] resulta:

$$\frac{\frac{SRR - SRS}{q}}{\frac{SRS}{T - k}} = \frac{\frac{6,28 - 6,20}{1}}{\frac{6,20}{100}} = 1,29$$

que es inferior a los valores críticos proporcionados por las tablas de distribución $F_{1,100}$, tanto al 95 como al 99 por 100 de confianza. En consecuencia, no rechazamos la hipótesis nula $H_0: \beta_1 + \beta_2 = 0$.

Como alternativa, podría sustituirse la restricción en el modelo para obtener el modelo transformado:

$$y_t = \beta_1(x_{1t} - x_{2t}) + \beta_3 x_{3t} + u_t = \beta_1 x_{1t}^* + \beta_3 x_{3t} + u_t$$

donde $x_{1t}^* = x_{1t} - x_{2t}$. Las matrices de productos correspondientes (calculadas a partir de la tabla anterior) son:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 10 & -2 \\ -2 & 6 \end{pmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 6 \\ -2 \end{pmatrix}$$

y el estimador MCO de este modelo transformado es $(\hat{\beta}_1, \hat{\beta}_3) = \left(\frac{4}{7}, -\frac{1}{7}\right)$.

que coincide con las estimaciones de estos coeficientes que obtuvimos antes. Evidentemente, $\hat{\beta}_2 = -\hat{\beta}_1 = -\frac{4}{7}$. La suma residual restringida se obtendría:

$$\text{SRR} = 10 - \begin{pmatrix} 4 & -1 \\ 7 & 7 \end{pmatrix} \begin{pmatrix} 6 \\ -2 \end{pmatrix} = \frac{44}{7}$$

Por tanto, la estimación restringida mediante sustitución de las hipótesis equivale a efectuar un cambio de variables en el modelo. La suma residual del modelo en el que se han sustituido las restricciones es la *suma residual restringida*. Esta suma se compara con la del modelo original para obtener el valor del estadístico F que hemos venido proponiendo a lo largo de este capítulo.

4.10. CONTRASTE DE CAMBIO ESTRUCTURAL. TEST DE CHOW

4.10.a. Test de Chow

Un contraste que tiene especial importancia por su interés, así como por la frecuencia con que aparece en aplicaciones empíricas, es el de la hipótesis nula de que dos submuestras han sido generadas por una misma estructura económica. Se contrasta la hipótesis nula de *ausencia de cambio estructural*, y el contraste suele denominarse como test de Chow. Usualmente se produce cuando se tiene cierta información acerca de una variación estructural que ocurrió en algún momento durante el período muestral, y se pretende contrastar si dicha variación fue suficientemente importante como para generar cambios en los coeficientes del modelo. El *modelo restringido* (MR) es:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T_1, T_1 + 1, \dots, T \quad [4.21]$$

mientras que el *modelo sin restringir* (MSR) es:

$$\begin{aligned} y_t &= \mathbf{x}'_t \boldsymbol{\beta}_1 + u_t, & t &= 1, 2, \dots, T_1 \\ y_t &= \mathbf{x}'_t \boldsymbol{\beta}_2 + u_t, & t &= T_1 + 1, \dots, T \end{aligned} \quad [4.22]$$

que consta de una regresión diferente para cada una de las dos submuestras. Como puede verse, ambos modelos en [4.22] tienen las mismas variables explicativas, y la hipótesis nula consiste en la igualdad de cada uno de los coeficientes β_i en las dos submuestras, lo que constituye un conjunto de k hipótesis lineales.

La suma residual restringida es, en este caso, la que proviene de la estimación del modelo restringido [4.21] denotada por SRR, mientras que la suma residual sin restringir es el agregado de las sumas residuales de cada

una de las regresiones en [4.22], que denotamos por SR_1 y SR_2 , respectivamente. El estadístico F para el contraste de la hipótesis nula de ausencia de cambio estructural es, por tanto:

$$\frac{\frac{SRR - (SR_1 + SR_2)}{k}}{\frac{SR_1 + SR_2}{T - 2k}} \quad [4.23]$$

con distribución $F_{k, T-2k}$, donde $T - 2k$ es el número de grados de libertad del modelo sin restringir, ya que se estiman k parámetros con las primeras T_1 observaciones muestrales y otros k parámetros con las restantes T_2 observaciones. En total, T observaciones y $2k$ parámetros estimados.

En muchos casos, este contraste está motivado por la necesidad de comprobar si las últimas observaciones recibidas suponen un cambio significativo con respecto al resto de la muestra. En tal caso, T_2 puede ser muy pequeño (quizá igual a 1 ó 2), y *no es posible estimar el segundo de los modelos de [4.22] por falta de grados de libertad.*

La dificultad estriba en que, como vimos en la Proposición 3.15, cuando se dispone de menos observaciones que coeficientes a estimar, existen infinitas soluciones al sistema de ecuaciones normales. En este sentido, el estimador MCO está mal definido, pero, sin embargo, cualquiera de tales soluciones genera una suma residual igual a cero. En realidad, no sólo es la suma residual igual a cero, sino que cada uno de los residuos es también igual a cero.

Supongamos que estamos en el caso límite en que T_2 es precisamente igual al número de coeficientes del modelo, k . En tal caso, $SR_2 = 0$, y el número de restricciones q es igual a k e igual a T_2 , por lo que el estadístico F de [4.23] se reduce a:

$$\frac{(SR - SR_1)/T_2}{SR_1/(T - 2k)} = \frac{(SR - SR_1)/T_2}{SR_1/(T_1 - k)} \quad [4.24]$$

con distribución F_{T_2, T_1-k} . Cuando descendemos a situaciones con $T_2 < k$, seguimos teniendo $SR_2 = 0$, aunque no hay razón para mantener T_2 y $T_1 - k$ como divisores en ambos términos del cociente en [4.24]. Sin embargo, puede demostrarse que exactamente la misma expresión [4.24] continúa teniendo distribución F_{T_2, T_1-k} en estos casos en que $T_2 < k$, pudiéndose llevar a cabo el contraste de cambio estructural del modo habitual.

4.10.b. Contrastes de estabilidad

Cuando se dispone de datos de series temporales, es siempre interesante contrastar la hipótesis nula de homogeneidad temporal del modelo, frente a

la alternativa de que se ha producido algún cambio a lo largo del período muestral. Este contraste tiene especial interés si el modelo se ha elaborado con fines predictivos.

Además del test de Chow, existen otros contrastes basados en los denominados «residuos recursivos». El residuo recursivo correspondiente a la observación t se define como el error de predicción de y_t , utilizando el estimador MCO obtenido con las observaciones 1, 2, ..., $t - 1$:

$$\hat{e}_t = y_t - \mathbf{x}'_t \hat{\boldsymbol{\beta}}_{t-1}$$

donde el subíndice temporal del estimador hace referencia a la muestra utilizada en su cálculo.

La varianza de dicho error de predicción es:

$$\text{Var}(\hat{e}_t) = \sigma_u^2 [1 + \mathbf{x}'_t (\mathbf{X}'_{t-1} \mathbf{X}_{t-1})^{-1} \mathbf{x}_t]$$

donde \mathbf{X}_{t-1} denota la matriz $(t - 1) \times k$ formada por las $t - 1$ primeras observaciones de las variables explicativas. Si definimos el *residuo recursivo normalizado*:

$$\tilde{e}_t = \frac{\hat{e}_t}{\sqrt{1 + \mathbf{x}'_t (\mathbf{X}'_{t-1} \mathbf{X}_{t-1})^{-1} \mathbf{x}_t}}$$

bajo la hipótesis nula de estabilidad y el supuesto de Normalidad, se tiene que \tilde{e}_t se distribuye $N(0, \sigma_u^2)$, y es independiente de \tilde{e}_s para todo $s \neq t$.

Si los valores de \tilde{e}_t cambian en el tiempo de manera sistemática, se tomará como evidencia de inestabilidad del modelo. Hay dos estadísticos que permiten contrastar tal hipótesis: El estadístico CUSUM se basa en la suma acumulada de los residuos normalizados:

$$W_t = \sum_{r=k+1}^t \frac{\tilde{e}_r}{\hat{\sigma}}$$

donde

$$\hat{\sigma} = \frac{1}{T-k} \cdot \sum_{r=k+1}^T (\tilde{e}_r - \bar{e})^2$$

y

$$\bar{e} = \frac{1}{T-k} \cdot \sum_{r=k+1}^T e_r$$

Bajo la hipótesis nula de estabilidad, W_t tiene esperanza cero y varianza aproximadamente igual al número de residuos acumulados, de modo que el contraste se lleva a cabo con el gráfico de W_t a través del tiempo. Se

construyen bandas de confianza para la serie de W_t mediante las rectas que unen los puntos $(k, \pm a\sqrt{T-k})$ y $(T, \pm 3a\sqrt{T-k})$. Al 95 por 100 de confianza, $a = 0,948$, mientras que al 99 por 100, $a = 1,143$. Se rechaza la hipótesis de estabilidad si W_t traspasa dichas bandas.

El segundo estadístico, CUSUMSQ, utiliza los cuadrados de los residuos normalizados:

$$S_t = \frac{\sum_{r=k}^t \tilde{e}_r^2}{\sum_{r=k}^T \tilde{e}_r^2}$$

Como los residuos recursivos son independientes, cada término de esta suma es una chi-cuadrado con un grado de libertad, por lo que $E(S_t)$ es aproximadamente igual a $(t-k)/(T-k)$. El contraste consiste en dibujar la serie temporal de S_t , así como las líneas que limitan la banda de confianza: $E(S_t) \pm C_0$, donde el valor crítico de C_0 se muestra en la Tabla A.10. Nuevamente, si S_t se sale de la banda, se rechaza la hipótesis nula de homogeneidad del modelo.

Un último contraste se basa en el hecho de que, bajo la hipótesis nula, la media muestral de los residuos recursivos, \bar{e} , se distribuye normalmente, con esperanza cero y varianza $\sigma_u^2/(T-k)$, por lo que puede construirse un contraste «tipo t »:

$$t = \frac{\sqrt{T-k}}{\hat{\sigma}} \cdot \bar{e}$$

donde

$$\hat{\sigma} = \frac{1}{T-k-1} \cdot \sum_{r=k+1}^T (e_r - \bar{e})^2$$

y se compara con las tablas de la t de Student para $T-k-1$ grados de libertad.

4.11. VARIABLES FICTICIAS

Para llevar a cabo algunos de los contrastes de homogeneidad analizados en la sección anterior, ha sido tradicional en Econometría la consideración de las llamadas variables ficticias. Estas son variables que toman el valor 1 en una submuestra y 0 en el resto de la muestra. Si el número de submuestras es mayor que dos, se define una variable ficticia para cada una de las submuestras, tomando el valor 1 en dicha submuestra y el valor 0 en el resto de las observaciones muestrales.

Por ejemplo, si se quiere contrastar la homogeneidad del término independiente en las dos submuestras en que se ha dividido una muestra, podría especificarse el modelo

$$y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T_1, T_1 + 1, \dots, T \quad [4.25]$$

donde $D_{1t} = 1$ y $D_{2t} = 0$ si $0 \leq t \leq T_1$, mientras que $D_{1t} = 0$ y $D_{2t} = 1$ si $T_1 < t \leq T$. Con estas definiciones, la matriz $T \times (k + 1)$ de observaciones de las variables explicativas sería:

$$\begin{pmatrix} \mathbf{1}_{T_1} & \mathbf{0}_{T_1} & \mathbf{X}_1 \\ \mathbf{0}_{T_2} & \mathbf{1}_{T_2} & \mathbf{X}_2 \end{pmatrix}$$

El contraste de igualdad de los términos independientes entre ambas submuestras sería equivalente a un contraste de igualdad de los coeficientes α_1 y α_2 en el modelo único [4.25].

Obsérvese que la especificación

$$y_t = \mu + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T_1, T_1 + 1, \dots, T$$

hubiese sido incorrecta, puesto que se cumple $D_{1t} + D_{2t} = 1$ para todo t , por lo que existe dependencia lineal exacta entre las tres primeras variables del modelo. Esto contradice uno de los supuestos que discutimos en la Sección 3.1, produciendo *multicolinealidad exacta*, cuyos nocivos efectos analizaremos en el Capítulo 10. Sería factible, en cambio, especificar el modelo

$$y_t = \mu + \alpha D_{2t} + \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T_1, T_1 + 1, \dots, T \quad [4.26]$$

en cuyo caso, el término independiente de la primera submuestra vendrá estimado por $\hat{\mu}$, mientras que el de la segunda submuestra se estima por $\hat{\mu} + \hat{\alpha}$. Con esta especificación, el contraste de homogeneidad de los términos independientes se formularía como un contraste de significación del coeficiente α . Si dicho parámetro resulta no significativo, entonces podría mantenerse la hipótesis nula de que los términos independientes son los mismos en las dos submuestras. La matriz $T \times (k + 1)$ de observaciones de las variables explicativas es en [4.26]:

$$\begin{pmatrix} \mathbf{1}_{T_1} & \mathbf{0}_{T_1} & \mathbf{X}_1 \\ \mathbf{1}_{T_2} & \mathbf{1}_{T_2} & \mathbf{X}_2 \end{pmatrix}$$

La heterogeneidad de pendientes entre submuestras (el conjunto de coeficientes exceptuando el término independiente) puede formularse con variables ficticias que multiplican a las variables explicativas, como por ejemplo:

$$y_t = \alpha + \mathbf{x}'_{1t} \boldsymbol{\beta} + D_{1t} \mathbf{x}'_{1t} \boldsymbol{\delta} + u_t$$

donde D_{1t} sería la variable ficticia antes definida. El producto $D_{1t} \mathbf{x}'_{1t}$ es un vector de variables ficticias que coincide con \mathbf{x}'_{1t} en la primera parte de la

muestra, y es igual al vector $\mathbf{0}_k$ en su segunda parte. Esta especificación tiene como matriz $T \times (2k - 1)$ de observaciones para las variables explicativas:

$$\begin{pmatrix} \mathbf{1}_{T_1} & \mathbf{X}_1 & \mathbf{X}_1 \\ \mathbf{1}_{T_2} & \mathbf{X}_2 & \mathbf{0}_{T_2 \times (k-1)} \end{pmatrix}$$

y el contraste de significación del coeficiente δ sería un contraste de homogeneidad de pendientes en las dos submuestras, suponiendo que los términos independientes son iguales. Como vemos, *un contraste de homogeneidad de coeficientes entre submuestras puede formularse como un contraste de significación de coeficientes de variables ficticias*.

Si se quiere contrastar la homogeneidad de términos independientes en las tres submuestras en que se ha dividido una muestra, se especificaría el modelo:

$$y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

donde $D_{it} = 1$ en la i -ésima submuestra, y $D_{it} = 0$ en el resto de la muestra, y se contrastaría la igualdad de los tres coeficientes α : $H_0: \alpha_1 = \alpha_2 = \alpha_3$ (nótese que esto son dos restricciones, no tres). Una formulación equivalente consistiría en especificar el modelo:

$$y_t = \mu + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

en el que contrastar la significación conjunta de los coeficientes α_1 y α_2 : $H_0: \alpha_1 = \alpha_2 = 0$, equivale a contrastar la homogeneidad de términos independientes. Si los coeficientes α_1 y α_2 resultan significativos, sus valores numéricos indicarían las diferencias entre los términos independientes de la primera y segunda submuestras, ambas con respecto a la tercera submuestra, cuyo término independiente es μ .

En resumen. *Especificar modelos con variables ficticias, estimar sus coeficientes y llevar a cabo los contrastes de significación citados es equivalente a estimar los modelos restringido y sin restringir que introducimos en la Sección 4.5 y llevar a cabo el contraste de homogeneidad correspondiente.*

En ocasiones se dispone únicamente de información cualitativa acerca de un conjunto de individuos, lo que puede representarse por variables ficticias. Supongamos que para analizar un modelo de determinación de salarios se ha recogido información en una muestra de 2.000 trabajadores acerca de su nivel de estudios (si es inferior a estudios primarios, primarios, medios y superiores), su sexo y el salario que reciben, y se especifica el modelo:

$$S_{it} = \alpha_1 + \alpha_2 E_{2i} + \alpha_3 E_{3i} + \alpha_4 E_{4i} + \beta \text{Sexo}_i + u_i \quad [4.27]$$

donde la variable E_2 toma el valor 1 para los trabajadores con estudios primarios y cero para el resto, la variable E_3 toma el valor 1 para los trabajadores con estudios medios y cero en el resto, y la variable E_4 toma e.

valor 1 para los trabajadores con estudios superiores. La variable Sexo toma el valor 1 para las mujeres, y el valor cero para los hombres.

En este contexto, un contraste de diferencias salariales en base al sexo del trabajador, a igualdad de educación, equivaldría a un contraste de significación del parámetro β : $H_0: \beta = 0$. Un contraste de diferencias salariales debidas al nivel de educación sería un contraste de significación conjunta de los coeficientes α_2 , α_3 y α_4 : $H_0: \alpha_2 = \alpha_3 = \alpha_4 = 0$. Las predicciones salariales de este modelo son:

Salarios	<E. primarios	E. primarios	E. medios	E. superiores
Mujeres	$\alpha_1 + \beta$	$\alpha_1 + \alpha_2 + \beta$	$\alpha_1 + \alpha_3 + \beta$	$\alpha_1 + \alpha_4 + \beta$
Hombres	α_1	$\alpha_1 + \alpha_2$	$\alpha_1 + \alpha_3$	$\alpha_1 + \alpha_4$

Una peculiaridad de este modelo es que, como puede verse en la tabla, sus predicciones implican que la diferencia entre el salario de los hombres y mujeres que tienen el mismo nivel de educación es igual a $-\beta$, con independencia de cual sea su nivel de estudios.

Una estimación negativa de β significaría que, a igualdad de nivel educativo, el salario percibido por un hombre es superior al percibido por una mujer. Ello podría interpretarse como un efecto de discriminación en contra de la mujer trabajadora, salvo que se pensara que el resultado puede deberse a la necesidad de incorporar otros factores de determinación salarial ausentes de [4.27].

Otra característica de este modelo es que la diferencia entre los salarios percibidos por trabajadores con diferente nivel de estudios es independiente de si éstos son hombres o mujeres.

Esta posible limitación desaparece si se especifica el modelo

$$S_i = \alpha_1 + \sum_{j=2}^4 \alpha_j E_{ji} + \beta \text{Sexo}_i + \sum_{j=2}^4 \delta_j (E_{ji} \cdot \text{Sexo}_i) + u_i$$

cuyas predicciones son:

Salarios	<E. primarios	E. primarios	E. medios	E. superiores
Mujeres	$\alpha_1 + \beta$	$\alpha_1 + \alpha_2 + \beta + \delta_2$	$\alpha_1 + \alpha_3 + \beta + \delta_3$	$\alpha_1 + \alpha_4 + \beta + \delta_4$
Hombres	α_1	$\alpha_1 + \alpha_2$	$\alpha_1 + \alpha_3$	$\alpha_1 + \alpha_4$

donde puede verse que este modelo recoge las interacciones que puedan existir entre el nivel educativo de un trabajador y su sexo en cuanto al efecto que tienen sobre el salario recibido.

Es precisamente en estos casos en que se pretende recoger interacciones entre variables explicativas, cuando más interés tiene la utilización de variables ficticias.

4.12. PREDICCIÓN EN EL MODELO LINEAL

Supongamos que habiendo estimado el modelo $y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$ con datos de series temporales se quiere utilizar la versión estimada de dicho modelo para hacer predicciones acerca de los valores futuros de la variable endógena y_t . El modelo lineal representa el mecanismo de relación entre el vector \mathbf{x}_t y la variable y_t , y tiene sentido utilizarlo para obtener predicciones de y_t salvo en casos de gran inestabilidad de dicha relación.

4.12.a. Cálculo de las predicciones

Una vez conocida la *mejor* relación lineal entre y_t y el vector \mathbf{x}_t a lo largo del intervalo muestral (la estimación MCO), *supongamos* que esa relación estimada es también la mejor relación entre y y las variables x_1, x_2, \dots, x_k en el período $T + 1$.

Bajo este supuesto de estabilidad, si denotamos por E_T el valor esperado en base a la información disponible hasta el instante T , la función a utilizar para predecir y_{T+1} es:

$$E_T y_{T+1} = E_T(\mathbf{x}'_{T+1} \boldsymbol{\beta} + u_{T+1}) = E_T(\mathbf{x}'_{T+1} \boldsymbol{\beta}) + E_T u_{T+1} = (\mathbf{E}_T \mathbf{x}_{T+1})' \hat{\boldsymbol{\beta}}_T + E_T u_{T+1}$$

donde⁽⁴⁾:

- a) Para predecir el producto $\mathbf{x}'_{T+1} \boldsymbol{\beta}$ multiplicamos el vector de estimaciones MCO por los valores previstos para las variables explicativas el próximo período.
- b) Representamos el estimador por $\hat{\boldsymbol{\beta}}_T$ para mostrar de manera explícita que las estimaciones dependen del tamaño muestral utilizado. Hemos supuesto que la dependencia de y_t respecto del vector \mathbf{x}_t es suficientemente estable como para poder utilizar $\hat{\boldsymbol{\beta}}_T$ para predecir y_{T+1} .
- c) Necesitamos predecir el vector de variables explicativas \mathbf{x}_{T+1} .
- d) Necesitamos predecir el término de error u_{T+1} .

Dediquemos ahora algo de atención a las dos últimas observaciones: con respecto a las variables \mathbf{x}_{T+1} , puede haber algunos casos en que sus valores futuros se conozcan de antemano y no haya necesidad de predecirlas. Por ejemplo, si el modelo se ha especificado para explicar la cifra de ventas de una determinada empresa como función del precio de su producto, la empresa fija la variable explicativa (precio) para el siguiente período y utiliza ese precio conocido para predecir el volumen de ventas.

A veces se citan las variables monetarias como otro ejemplo. El banco emisor puede tener un control de los agregados monetarios que permita la

⁽⁴⁾ La cadena de igualdades anterior es válida únicamente bajo supuestos matemáticos en cuya descripción no entramos. Nos limitamos aquí a justificar heurísticamente su validez.

utilización de cifras futuras del crecimiento de la oferta monetaria para predecir efectos inflacionistas, por ejemplo. Sin embargo, el control de los agregados monetarios es menos que perfecto, y la sustitución de $E_T x_{T+1}$ por un valor numérico x_{T+1} no deja de ser sino una aproximación. Los salarios reales, si sus valores para los dos años próximos han sido fijados en virtud de un convenio, serían otro ejemplo de una variable explicativa (en una ecuación de empleo, por ejemplo) cuyos valores futuros son conocidos de antemano.

En el resto de los casos los valores futuros de las variables explicativas se desconocen, por lo que habría que estimarlos previamente. Estas estimaciones pueden estar basadas simplemente en nuestra intuición o ser algo más rigurosas. En el Capítulo 13 estudiaremos métodos para desarrollar modelos que permiten predecir una variable utilizando únicamente la información contenida en los valores pasados de esa misma variable. Indudablemente, dichos métodos serían de gran interés en esta situación.

Por último, hemos de determinar la predicción $E_T u_{T+1}$, es decir, hemos de responder a la pregunta: ¿cuál es la mejor predicción, basándonos en la información hoy disponible, del valor del término de error el próximo período? Como vimos en la Proposición 2.1, si se carece de información muestral, la mejor predicción de una variable aleatoria es su esperanza matemática. Esto es debido a que la esperanza matemática es la constante con respecto a la cual se minimizan las fluctuaciones de la variable aleatoria. Hemos supuesto que el término de error es una sucesión de variables aleatorias independientes entre sí, lo que implica que u_{T+1} es independiente de la información de que hoy disponemos. En estas circunstancias, la muestra no nos proporciona ninguna información útil que ayude a predecir el valor futuro del término de error y , en consecuencia, predecimos u_{T+1} por medio de su esperanza matemática, es decir, $E_T u_{T+1} = E u_{T+1} = 0$.

De la discusión que hemos hecho se deducen una serie de condiciones necesarias para que la predicción así obtenida sea fiable:

1. Que la relación lineal estimada entre y_t y las variables x_{1t}, \dots, x_{kt} se mantenga en el futuro.
2. Que los coeficientes sean suficientemente estables como para que sus estimaciones obtenidas con la muestra actual sean una buena aproximación a los valores que se obtendrían al incorporar observaciones futuras.
3. Que se conozcan los valores futuros de las variables x , o que los modelos de predicción que utilicemos para dichas variables sean suficientemente fiables.
4. Que el modelo lineal que hemos estimado esté correctamente especificado.
5. Que el horizonte de predicción no sea muy lejano.

Ningún problema de estimación es completo si la estimación que se obtiene no viene acompañada de un intervalo de confianza, y la predicción no es diferente en este respecto. Vamos a suponer (como aproximación) que

los valores futuros \mathbf{x}_{T+1} son conocidos. En estas condiciones, si consideramos el modelo

$$y_{T+1} = \mathbf{x}'_{T+1}\boldsymbol{\beta} + u_{T+1}$$

entonces la predicción mínimo-cuadrática es, en resumen:

$$E_T y_{T+1} = (E_T \mathbf{x}_{T+1})' \hat{\boldsymbol{\beta}}_T = \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T$$

y pasamos a calcular su desviación típica e intervalos de confianza.

4.12.b. El error de predicción y su varianza

El error de predicción un período hacia el futuro se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida:

$$e_T(1) = y_{T+1} - E_T y_{T+1} = \mathbf{x}'_{T+1}\boldsymbol{\beta} - \mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_T + u_{T+1} = \mathbf{x}'_{T+1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) + u_{T+1}$$

que es, en el instante T , una variable aleatoria con valor desconocido, puesto que su realización ocurrirá en el período $T+1$.

Puede verse en esta expresión cuáles son las fuentes de error de predicción:

1. El error en la estimación del vector $\boldsymbol{\beta}$.
2. El error en la predicción del vector \mathbf{x}_{T+1} .
3. El error estocástico inherente al modelo, u_{T+1} .

Como el estimador MCO es insesgado, entonces se tiene que, bajo nuestros supuestos, el error de predicción tiene esperanza cero: $E(e_T(1)) = 0^{(5)}$. En este sentido se dice que cuando las variables \mathbf{x}_{T+1} son conocidas de antemano, entonces la predicción obtenida a partir del estimador MCO del modo que aquí hemos descrito es insesgada.

Calculemos ahora la varianza del error de predicción, $\sigma_e^2 = \text{Var}(e_T(1))$:

$$\begin{aligned} \sigma_e^2 &= E\{\mathbf{x}'_{T+1}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})'\mathbf{x}_{T+1} + 2\mathbf{x}'_{T+1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)u_{T+1} + u_{T+1}^2\} = \\ &= \mathbf{x}'_{T+1}E[(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_T - \hat{\boldsymbol{\beta}}_T)']\mathbf{x}_{T+1} + E u_{T+1}^2 = \sigma_u^2 \mathbf{x}'_{T+1}(\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{x}_{T+1} + \sigma_u^2 \end{aligned} \quad [4.28]$$

donde hemos utilizado el hecho de que, debido a que el término de error es independiente en el tiempo, se tiene $E[\boldsymbol{\beta}_T u_{T+1}] = 0$ (ya que $\boldsymbol{\beta}_T$ depende de los términos de error sólo hasta el instante T , inclusive).

⁽⁵⁾ Indudablemente, esta es una propiedad mínima que toda predicción debe tener. En caso contrario, si la esperanza del error de predicción fuese una constante no nula, ello querría decir que se comete un error sistemático en la predicción que, corregido, mejoraría las predicciones. Que el error de predicción tenga esperanza nula garantiza que es cero en promedio, es decir, que siendo positivos y negativos en distintos periodos, la suma de los errores de predicción tiende a

En la expresión [4.28] hemos denotado por \mathbf{X}_T la matriz $T \times k$ de observaciones de las variables x_1, x_2, \dots, x_k , a la que hemos añadido el subíndice T para indicar mejor la dependencia temporal. Como puede verse, la varianza del error de predicción es función únicamente de:

- La matriz $T \times k$ de observaciones \mathbf{X}_T .
- El vector formado por las k observaciones \mathbf{x}_{T+1} , que venimos suponiendo conocido.
- El parámetro σ_u^2 desconocido, que sustituiremos por una estimación suya.

4.12.c. Intervalo de confianza para la predicción

La estimación resultante de $\hat{\sigma}_e$ puede utilizarse en la construcción de un intervalo de confianza para $E_T y_{T+1}$, como sigue:

Bajo el supuesto de Normalidad del término de error, el error de predicción $e_T(1)$ es combinación lineal de dos variables normales:

$$e_T(1) = -\mathbf{x}'_{T+1}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) + u_{T+1} = -\mathbf{x}_{T+1}(\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T u_T + u_{T+1}$$

y tiene, por consiguiente, distribución Normal: $e_T(1) \sim N(0, \sigma_e^2)$, donde σ_e^2 se calculó en [4.28]. Por un razonamiento análogo al de secciones previas se tiene que

$$\frac{e_T(1)}{\hat{\sigma}_e} = \frac{y_{T+1} - E_T y_{T+1}}{\hat{\sigma}_e}$$

sigue una distribución t_{T-k} .

Por consiguiente, dada una predicción puntual $E_T y_{T+1}$ y una estimación de la desviación típica del error de predicción $\hat{\sigma}_e$, podemos utilizar esta expresión para obtener un intervalo de confianza para el valor futuro y_{T+1} .

Es importante puntualizar que los resultados que acabamos de presentar dejarán de ser rigurosamente válidos si dejan de ser ciertas las hipótesis bajo las cuales los obtuvimos y, en particular, si las variables \mathbf{x}_{T+1} no son conocidas con certeza. En tal caso, las expresiones anteriores no son sino *cotas inferiores para la varianza del error de predicción*.

En tales situaciones puede aplicarse la desigualdad de Chebyshev, que asegura:

$$P[|E_T y_{T+1} - y_{T+1}| \geq \lambda \sigma_e] \leq \frac{1}{\lambda^2}$$

y elegir un valor de λ apropiado (por ejemplo, de modo que $\frac{1}{\lambda^2} = 0,05$). Se

puede confiar en que sustituir σ_e por una estimación $\hat{\sigma}_e$ no supondrá un gran error y así utilizaríamos como intervalo de confianza:

$$P[E_T y_{T+1} - \lambda \hat{\sigma}_e \leq y_{T+1} \leq E_T y_{T+1} + \lambda \hat{\sigma}_e] \geq 1 - \frac{1}{\lambda^2}$$

con la salvedad de que ahora la probabilidad dentro de este intervalo no es exacta, sino superior a una cierta cota inferior que hemos establecido con la elección del valor del parámetro λ . Así, si se quiere hacer una afirmación a un nivel de confianza aproximado del 95 por 100, se tomaría $\lambda = 4,472$ y se diría que la probabilidad de que la realización futura de la variable aleatoria y_{T+1} se halle entre $E_T y_{T+1} - 4,472 \hat{\sigma}_e$ y $E_T y_{T+1} + 4,472 \hat{\sigma}_e$ es de al menos 0,95. Por supuesto, una cota inferior diferente para el nivel de significación implicaría un número diferente de 4,472 en la construcción del intervalo de confianza.

4.12.d. Predicción de un vector de valores futuros de la variable endógena

Supongamos ahora que queremos predecir la evolución de la variable endógena a lo largo de un número n de periodos futuros; denotemos por $y_F = (y_{T+1}, y_{T+2}, \dots, y_{T+n})$ el vector formado por dichos valores futuros. De acuerdo con el modelo econométrico se tiene:

$$y_F = X_F \beta + u_F$$

donde X_F es la matriz $n \times k$ de valores de las variables explicativas sobre el horizonte de predicción y u_F el vector $n \times 1$ de los términos de error correspondientes. En esta representación hemos incorporado el supuesto de que el vector de coeficiente β es estable en el tiempo.

Siguiendo la lógica de la discusión anterior, es fácil ver que la predicción debe obtenerse mediante:

$$E_T y_F = X_F \hat{\beta}_T = X_F (X_T' X_T)^{-1} X_T' y_T$$

donde X_T es la matriz $T \times k$ e y_T el vector $T \times 1$ de observaciones muestrales.

El error de predicción (vector de dimensión n) es igual a:

$$\begin{aligned} e_T(n) &= y_F - E_T y_F = X_F \beta + u_F - E_T y_F = \\ &= X_F \beta + u_F - X_F (X_T' X_T)^{-1} X_T' (X_T \beta + u_T) = u_F - X_F (X_T' X_T)^{-1} X_T' u_T \end{aligned}$$

siendo u_T el vector $T \times 1$ formado por los términos de error a lo largo del intervalo muestral. Es claro que

$$E(e_T(n)) = \mathbf{0}_n$$

por lo que la matriz de covarianzas del vector de errores de predicción es:

$$\text{Var}(\mathbf{e}_T(n)) = E[\mathbf{e}_T(n)\mathbf{e}_T(n)'] = \text{Var}(\mathbf{u}_F) + \mathbf{X}_F(\mathbf{X}'_T\mathbf{X}_T)^{-1}\mathbf{X}'_T \text{Var}(\mathbf{u})\mathbf{X}_T(\mathbf{X}'_T\mathbf{X}_T)^{-1}\mathbf{X}'_F - 2 \text{Cov}[\mathbf{X}_F(\mathbf{X}'_T\mathbf{X}_T)^{-1}\mathbf{X}'_T\mathbf{u}, \mathbf{u}_F]$$

Si suponemos que el término de error continúa siendo independiente sobre el horizonte de predicción, se tiene:

$$\text{Cov}(\mathbf{u}, \mathbf{u}_F) = \mathbf{0}_{T \times n}$$

y, por consiguiente:

$$\text{Var}(\mathbf{e}_T(n)) = \sigma_u^2[\mathbf{I}_n + \mathbf{X}_F(\mathbf{X}'_T\mathbf{X}_T)^{-1}\mathbf{X}'_F] = \sigma_u^2\mathbf{V} \quad [4.29]$$

donde \mathbf{V} denota la matriz que aparece dentro del corchete, de dimensión $n \times n$. Si el término de error del modelo econométrico tiene una distribución Normal, entonces el vector de errores de predicción también seguirá una distribución Normal ($\mathbf{0}_n, \sigma_u^2\mathbf{V}$).

El razonamiento anterior se basa en el supuesto de estabilidad de los coeficientes del modelo econométrico. Bajo dicho supuesto, el vector aleatorio $\mathbf{e}_T(n)$ formado por los errores de predicción no debería alejarse excesivamente de su valor esperado, $\mathbf{0}_n$. Una vez transcurrido el horizonte temporal para el que se obtuvieron las previsiones, puede calcularse el vector $\mathbf{e}_T(n)$. Una comparación entre su norma y su varianza [dada por la forma cuadrática $\mathbf{e}_T(n)'(\sigma_u^2\mathbf{V})^{-1}\mathbf{e}_T(n)$] puede utilizarse para obtener un *contraste de estabilidad de los coeficientes del modelo* a lo largo del tiempo.

El desconocimiento del parámetro σ_u^2 se resuelve por el procedimiento habitual de considerar el cociente de formas cuadráticas:

$$\frac{\mathbf{e}_T(n)'(\sigma_u^2\mathbf{V})^{-1}\mathbf{e}_T(n)/n}{(T-k)\frac{\hat{\sigma}_u^2}{\sigma_u^2}} = \frac{\mathbf{e}_T(n)'\mathbf{V}^{-1}\mathbf{e}_T(n)}{n\hat{\sigma}_u^2} \quad [4.30]$$

que sigue ahora una distribución $F_{n, T-k}$, donde $\hat{\sigma}_u^2$ denota la estimación de dicho parámetro obtenida con las T observaciones muestrales. Como se pide probar en el Problema 4.6, este estadístico coincide con el estadístico de Chow que vimos en la Sección 4.10 para contrastar la hipótesis de estabilidad temporal de los coeficientes del modelo econométrico.

4.12.e. Evaluación de la bondad predictiva del modelo

En el período $T+n$ habrá transcurrido el horizonte para el que se obtuvieron las predicciones y el investigador dispondrá de las observaciones $y_{T+1}, y_{T+2}, \dots, y_{T+n}$, por lo que podrá evaluarse la bondad de las predicciones que se hicieron. Entre otras cosas, si había varios modelos de predicción

alternativos, la comparación de sus medidas correspondientes permitirá comprobar cuál de ellos fue superior.

Denotando por $y_{T+1}^p, y_{T+2}^p, \dots, y_{T+n}^p$ las predicciones, la raíz del error cuadrático medio de la predicción t en términos porcentuales se define:

$$\text{RECM} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_{t+i}^p - y_{t+i}}{y_{t+i}} \right)^2} \quad [4.31]$$

Otra medida es el estadístico U propuesto por Theil (1961), muy utilizado para la evaluación de predicciones:

$$U = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (y_{t+i}^p - y_{t+i})^2}{\frac{1}{n} \sum_{i=1}^n y_{t+i}^2}} \quad [4.32]$$

En ocasiones se evalúa esta medida para las fluctuaciones en y_t :

$$U_{\Delta} = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (\Delta y_{t+i}^p - \Delta y_{t+i})^2}{\frac{1}{n} \sum_{i=1}^n (\Delta y_{t+i})^2}} \quad [4.33]$$

donde Δy_t puede denotar bien la variación absoluta en y_t : $\Delta y_t = y_t - y_{t-1}$, o su tasa de variación porcentual: $\Delta y_t = \frac{y_t - y_{t-1}}{y_{t-1}}$. El estadístico U_{Δ} evalúa la bondad del modelo predictivo para anticipar puntos de giro en y_t .

4.12.f. Predicción de la «media» de la variable y_{T+1}

En ocasiones, el investigador no está tan interesado en predecir el valor futuro de la variable endógena y_{T+1} , sino tan sólo su valor esperado Ey_{T+1} que es igual a $Ey_{T+1} = \mathbf{x}'_{T+1} \boldsymbol{\beta}$. Si conocemos los valores futuros de las variables explicativas \mathbf{x}_{T+1} , la predicción sería $\mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T$, es decir, la misma que utilizamos por predecir y_{T+1} . La diferencia es que el error de predicción de Ey_{T+1} es:

$$\tilde{e}_T(1) = Ey_{T+1} - \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T = \mathbf{x}'_{T+1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) = \mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T u_T$$

de modo que su matriz de covarianzas es:

$$\begin{aligned} \text{Var}(\tilde{e}_T(1)) &= E[\tilde{e}_T(1) \tilde{e}_T(1)'] = E[\mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T u_T u_T' \mathbf{X}_T (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{x}_{T+1}] = \\ &= \mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T (\sigma_u^2 \mathbf{I}_T) \mathbf{X}_T (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{x}_{T+1} = \sigma_u^2 [\mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{x}_{T+1}] \end{aligned}$$

que es inferior a $\text{Var}(e_T(1))$, la matriz de covarianzas del error de predicción del dato de y_{T+1} . Ello se debe a que, al predecir y_{T+1} , pretendemos prever no sólo el componente explicado por \mathbf{x}_{T+1} , sino también la componente puramente aleatoria, u_{T+1} . Por el contrario, esta componente no forma parte del objetivo de predicción cuando se prevé el valor esperado Ey_{T+1} .

PROBLEMAS

Problema 4.1. Probar que la matriz de covarianzas del estimador de mínimos cuadrados restringidos es

$$\text{Var}(\hat{\beta}_R) = \sigma_u^2[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}]^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}]$$

y demostrar que la diferencia entre la matriz de covarianzas del estimador MCO sin restringir, $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, y la matriz anterior es una matriz semidefinida positiva.

Problema 4.2. Demostrar que, cuando la hipótesis nula está constituida por k restricciones lineales independientes entre sí, el estimador de mínimos cuadrados restringidos coincide con la solución al sistema de ecuaciones formado por dichas restricciones.

Problema 4.3. Probar que las expresiones [4.4] y [4.6] para el estadístico F para el contraste de una hipótesis del tipo $H_0: \beta = \beta^0$ coinciden.

Problema 4.4. Escribiendo las sumas residuales del modelo restringido y sin restringir como $\mathbf{y}'\mathbf{M}_1\mathbf{y}$ y $\mathbf{y}'\mathbf{M}\mathbf{y}$, con la notación de la Sección 4.5, probar que $\mathbf{y}'(\mathbf{M}_1 - \mathbf{M})\mathbf{y} = \text{SRR} - \text{SRS}$ directamente (sin partir de [4.7]) y, en consecuencia, que el estadístico [4.12] sigue una distribución $F_{s, T-k}$.

Problema 4.5. Probar que el predictor $E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta}$ es el predictor lineal insesgado de mínima varianza, de entre los que satisfacen:

$$E(\mathbf{x}'_{T+1} \hat{\beta}) = \mathbf{x}'_{T+1} E(\hat{\beta}) = \mathbf{x}'_{T+1} \beta = E(y_{T+1})$$

Indicación: Escribir cualquier otro predictor como:

$$[\mathbf{x}'_{T+1}(\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T + \mathbf{B}]\mathbf{y}$$

Problema 4.6. Demostrar que, cuando $n < k$, el estadístico [4.23] para el contraste de homogeneidad de coeficientes en el modelo econométrico que vimos en la Sección 4.10 coincide con el test de Chow.

Problema 4.7. a) Efectuar el contraste de la hipótesis $H_0: 3\beta_1 - \beta_2 = 7$ en el modelo

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

si se ha obtenido $\hat{\beta}_1 = 4,5$, $\hat{\beta}_2 = 5,0$, a partir de:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 50 & 13 & 21 & -120 \\ & 4 & 2 & -20 \\ & & 6 & -10 \\ & & & 10 \end{pmatrix}$$

$$\text{Suma residual} = 40,0, \quad T = 25$$

b) ¿Cómo se contrastaría la hipótesis anterior junto con la hipótesis $\beta_3 = 0$ si se ha estimado $\hat{\beta}_3 = -1,5$?

Problema 4.8. Probar que en el modelo $y_t = \alpha + \beta x_t + u_t$ la varianza del error de predicción $e_T(1)$ es igual a

$$\sigma_e^2 = \sigma_u^2 \left[1 + \frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_1^T (x_i - \bar{x})^2} \right]$$

e interpretar la contribución de cada uno de los tres términos que aparecen en el corchete a dicha varianza.

Problema 4.9. Dada la matriz de productos cruzados de las variables:

	y	x_1	x_2	x_3
y	10	3	-3	-2
x_1	3	10	4	-2
x_2	-3	4	8	0
x_3	-2	-2	0	6

- Obtener las estimaciones MCO.
- Obtener el estimador restringido por la relación $\beta_2 + 2\beta_3 = -1$.
- Obtener el mismo estimador introduciendo la restricción en el modelo, del modo que se ha descrito en la Sección 4.9.

Problema 4.10. Dada la matriz de productos cruzados:

	y	x_1	x_2	x_3	x_4
y	10	3	-3	-2	3
x_1	3	10	4	-2	0
x_2	-3	4	8	0	0
x_3	-2	-2	0	6	0
x_4	3	0	0	0	8

a) Obtener las estimaciones MCO del modelo:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

b) Obtener las estimaciones MCO bajo las restricciones:

$$\begin{aligned} \beta_1 + \beta_2 &= 0 \\ -4\beta_3 + 2\beta_4 &= 3 \end{aligned}$$

así como la matriz de covarianzas de dicho estimador.

Problema 4.11. Probar que el estadístico [4.1] para el contraste de la hipótesis nula $H_0: \beta = \beta^0$ acerca del vector de coeficientes del modelo econométrico, supuesto que σ_u^2 es conocido, coincide con el resultado que daría el contraste de razón de verosimilitudes (véase Sección 2.12). ¿Puede generalizarse este resultado al caso en que σ_u^2 es desconocido?

Problema 4.12. Demostrar que la distribución chi-cuadrado con $T - 1$ grados de libertad para el estadístico $(T - 1) \frac{\hat{\sigma}_u^2}{\sigma_u^2}$ puede obtenerse como un contraste de razón de verosimilitudes.

Problema 4.13. Halle la matriz de covarianzas entre el estimador MCG y la diferencia entre éste y el estimador MCO:

$$\text{COV}[\hat{\beta}_{\text{MCG}}, (\hat{\beta}_{\text{MCG}} - \hat{\beta}_{\text{MCO}})]$$

Problema 4.14. Obtenga el estadístico de razón de verosimilitudes para el contraste de la significación conjunta de un subvector de coeficientes del modelo de regresión lineal múltiple.

Problema 4.15. Probar el caso particular de $s = 1$ de la Sección 4.4.d.

Problema 4.16. Probar que la expresión [4.23] para el estadístico F en el que se basa el test de Chow para el contraste de cambio estructural es el mismo al que se llegaría utilizando la expresión general [4.4].

Problema 4.17. Dados los siguientes datos:

x_{i2}	y_i
169,6	71,20
166,8	58,20
157,1	56
181,1	64,50
158,4	53
165,6	52,4
166,7	56,8
157,5	49,2
168,1	55,6
165,3	77,8

Se pide:

- Efectuar la regresión $y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$.
- Hallar la matriz de covarianzas de β .
- Contrastar $H_0: \beta_1 = 0$.
- Contrastar $H_0: \beta_2 = 0$.
- Obtener los residuos. ¿Existen valores atípicos? ¿Existe alguna pauta definida claramente en los residuos?

Problema 4.18. El estimador de máxima verosimilitud restringido $\hat{\beta}_{MVR}$ por el conjunto de restricciones lineales $H_0: \mathbf{R}\beta = \mathbf{r}$ se obtiene maximizando la función de verosimilitud, sujeta a dichas restricciones. Demuestre que dicho estimador coincide con el estimador de mínimos cuadrados restringido $\hat{\beta}_{MCR}$.

Problema 4.19. Discuta las implicaciones que para el contraste de hipótesis lineales pueden tener los errores de especificación, según se produzcan por omitir variables relevantes o por incluir variables irrelevantes.

Problema 4.20. Discuta las implicaciones que los cambios de escala y de origen en las variables del modelo lineal tienen sobre la contrastación de hipótesis.

Problema 4.21 (Johnston). Considere la regresión en desviaciones con respecto a la media:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

donde se sabe que:

$$T = 100; \quad \sum_1^T y_t^2 = \frac{493}{3}; \quad \sum_1^T x_{1t}^2 = 30; \quad \sum_1^T x_{2t}^2 = 3; \quad \sum_1^T x_{2t} y_t = 20;$$

$$\sum_1^T x_{1t} y_t = 30; \quad \sum_1^T x_{1t} x_{2t} = 0$$

- Obtenga las estimaciones MCO.
- Contraste la hipótesis nula $H_0: \beta_2 = 7$.
- Contraste la hipótesis nula $H_0: \beta_1 = \beta_2 = 0$.

Problema 4.22 (Johnston). a) Estime por MCO el modelo $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$ disponiendo de la información muestral:

$$y = \begin{pmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{pmatrix}; \quad x = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix}$$

- b) Contraste la hipótesis nula $H_0: \beta_2 = \beta_3 = 0$.
- c) Contraste la hipótesis nula $H_0: \beta_3 = 0$.
- d) Contraste la hipótesis nula $H_0: \beta_2 + \beta_3 = 0$.
- e) Construya un intervalo de confianza del 95 por 100 para $\hat{\beta}_2$.
- f) Halle una región de confianza conjunta para (β_2, β_3) .

Problema 4.23. Supongamos que se quieren relacionar las horas de lectura de libros de ficción con el nivel de estudios de las personas en una muestra, para lo que se agrupan los individuos encuestados en tres clases:

- Clase 1: Si la persona tiene estudios superiores.
- Clase 2: Si la persona tiene estudios medios (BUP).
- Clase 3: Si la persona tiene estudios inferiores a éstos.

Posteriormente se definen tres variables: E_1, E_2, E_3 , que toman valores: $E_i = 1$ si la persona pertenece a la clase i -ésima, y $E_i = 0$ en caso contrario, y se estima la regresión:

$$y_j = \alpha_1 E_{1j} + \alpha_2 E_{2j} + \alpha_3 E_{3j} + u_j, \quad j = 1, 2, \dots, n$$

- a) Probar que $\hat{\alpha}_i^{MCO} = \bar{y}_i$, $i = 1, 2, 3$, donde \bar{y}_i denota la media muestral de y_j dentro de la clase i -ésima.
- b) Si, por el contrario, se especifica el modelo $y_j = \beta_1 + \beta_2 E_{2j} + \beta_3 E_{3j} + v_j$, pruebe que $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)_{MCO} = (\bar{y}_1, \bar{y}_2 - \bar{y}_1, \bar{y}_3 - \bar{y}_1)$.
- c) Por último, se considera el factor sexo, para lo que se define la variable $S_j = 1$ si la persona encuestada es una mujer, e igual a cero si es un hombre. Disponiendo de datos recogidos de diez personas que han manifestado realizar las horas de lectura semanales:

	E_1	E_2	E_3
Mujeres	8, 10, 12	12, 14	20
Hombres	5, 6	10	20

estime el modelo $y_j = \alpha_1 + \alpha_2 E_{2j} + \alpha_3 E_{3j} + \alpha_4 S_j$ y calcule las horas estimadas de lectura para cada combinación (E_i, S_j) , $i = 1, 2, 3$.

Problema 4.24. Para estimar el modelo $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ se dispone de las observaciones:

t	1	2	3	4	5	6	7	8	9	10
y	3	6	10	5	10	12	5	10	10	8
x_1	1	1	1	1	1	1	1	1	1	1
x_2	1	0	-1	1	0	-1	1	0	-1	0
x_3	1	1	1	0	0	0	-1	-1	-1	0

a) Estime $\beta_1, \beta_2, \beta_3, \sigma_u^2$ y obtenga la matriz de varianzas y covarianzas del vector β .

b) Considere la hipótesis conjunta $H_0: \beta_1 = 7, \beta_2 = 2\beta_3$. Escríbala en la forma $R\beta = r$ y calcule el valor del estadístico F para su contraste. Muestre que no puede rechazarse la hipótesis al 95 por 100 de confianza.

c) Si la hipótesis fuese cierta, la regresión podría escribirse: $y_t - 7x_{1t} = \beta_3(2x_{2t} + x_{3t}) + u_t$ o, de forma similar: $y_t^* = \beta_3 x_t^* + u_t$. Efectúe la regresión mínimo-cuadrática en este nuevo modelo, calcule el valor del estadístico

$$\frac{(\text{SRR} - \text{SRS})/q}{\text{SRS}/(T - k)}$$

y utilícelo para contrastar dicha hipótesis.

Problema 4.25. Con datos de una cierta economía se ha estimado con datos anuales para el período 1961-1974 la relación entre Consumo, Exportaciones y la Oferta Monetaria:

$$C_t = 655,8 + 3,359 Ex_t + 0,0556 OM_t; \quad R^2 = 0,969$$

(11,39) (4,282) (0,312) SR = 53938,8

donde las cifras entre paréntesis son los valores de los estadísticos t de Student. A continuación se lleva a cabo la regresión del consumo sobre la oferta monetaria, obteniendo:

$$C_t = 851,3 + 0,7945 OM_t; \quad R^2 = 0,9182$$

(15,52) (11,60) SR = 143868,69

Se sabe además que la inversa de la matriz $X'X$ es:

$$(X'X)^{-1} = \begin{pmatrix} 0,6758016 & -0,007303 & 0,001341 \\ -0,007303 & 0,000125 & -0,0000276 \\ 0,001341 & -0,0000276 & 0,0000064 \end{pmatrix}$$

Lleve a cabo un contraste de la significatividad de la variable Exportaciones en el primer modelo por tres procedimientos diferentes y compruebe que son equivalentes.

Problema 4.26 (Johnston). Se ha estimado la función de producción:

$$\ln Q_t = 1,37 + 0,632 \ln K_t + 0,452 \ln L_t$$

(0,257) (0,219)

$$R^2 = 0,98; \quad \text{Cov}(\hat{\beta}_k, \hat{\beta}_l) = 0,055$$

donde los valores entre paréntesis son desviaciones típicas. Contraste:

a) La hipótesis de que las elasticidades del producto con respecto al capital y al trabajo son iguales.

b) La existencia de rendimientos constantes a escala.

Problema 4.27. Para estimar el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ se dispone de dos muestras de tamaño 100 y 300, en las que los productos cruzados de las variables son, respectivamente:

	y_t	x_{1t}	x_{2t}	x_{3t}		y_t	x_{1t}	x_{2t}	x_{3t}
y_t	200	-100	-200	400	y_t	400	100	-300	800
x_{1t}		100	0	200	x_{1t}		300	100	400
x_{2t}			200	0	x_{2t}			200	0
x_{3t}				300	x_{3t}				600

- a) Obtenga el estimador MCO de $(\beta_2, \beta_3, \beta_4)$ y σ_u^2 con cada una de las dos submuestras.
- b) Contraste la hipótesis nula de que el vector $(\beta_2, \beta_3, \beta_4)$ es igual en ambas muestras.
- c) Contraste la hipótesis de que la varianza σ_u^2 es igual en ambas muestras.

Problema 4.28. Estime el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ sujeto a la restricción $\beta_2 + \beta_3 = 1$ disponiendo únicamente de la información:

$$\sum_1^T (y_t - x_{3t})^2 = 8250; \quad \sum_1^T x_{2t} = 200; \quad \sum_1^T (x_{2t} - x_{3t})^2 = 600; \quad T = 100$$

$$\sum_1^T x_{3t} = 400; \quad \sum_1^T y_t = 600; \quad \sum_1^T (y_t - x_{3t})(x_{2t} - x_{3t}) = 200$$

y obtenga la matriz de covarianzas de sus estimaciones.

Problema 4.29. a) Utilice la expresión [4.18] del estimador MCR para probar que el estimador de los coeficientes del modelo:

$$y = X_1 \beta_1 + X_2 \beta_2 + u$$

donde X_1 es $T \times k_1$ y X_2 $T \times k_2$, sujeto a las restricciones $H_0: \beta_2 = 0$, es $\hat{\beta} = (\hat{\beta}_1, 0)$, donde $\hat{\beta}_1$ denota el estimador MCO en una regresión de y sobre X_1 .

b) Demuestre asimismo que el estimador de mínimos cuadrados, restringido por $H_0: \beta_2 = \beta_2^0$, siendo β_2^0 un vector de constantes conocidas, puede escribirse:

$$\hat{\beta}_1^* = (X_1' X_1)^{-1} X_1' (y - X_2 \beta_2^0)$$

c) ¿Qué le sugieren estos resultados acerca de modos alternativos de llevar a cabo la contrastación de hipótesis lineales?

Problema 4.30. Demuestre que la expresión [4.18] del estimador MCR del modelo $y = X\beta + u$ sujeto a las restricciones lineales $R\beta = r$ puede escribirse en la forma:

$$\hat{\beta}_{MCR} = [I - CR]\hat{\beta} + w$$

donde el vector w no depende de $\hat{\beta}$. ¿Cuál es la matriz C en dicha representación?

Problema 4.31. Derive la expresión analítica del contraste de razón de verosimilitudes de la hipótesis nula $H_0: \beta = \beta^0$ en el modelo lineal $y_t = x_t'\beta + u_t$.

Problema 4.32. Para estimar el modelo $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ se dispone de los productos matriciales:

$$(X'X) = \begin{pmatrix} 20 & -20 & -18 & 3 \\ -20 & 25 & 20 & -2 \\ -18 & 20 & 4 & 0 \\ 3 & -2 & 0 & 1 \end{pmatrix} \quad X'y = \begin{pmatrix} 10 \\ 10 \\ -4 \\ -10 \end{pmatrix}; \quad y'y = 454$$

obtenidos a partir de las 104 observaciones de que consta la muestra.

- Obtener el estimador MCO.
- Contrastar, al 95 por 100 de confianza, la hipótesis $H_0: \beta_3 = 0$.
- Contrastar, al 95 por 100 de confianza, la hipótesis $H_0: 2\beta_3 - \beta_1 = -2$.
- Construir un intervalo del 95 por 100 de confianza para σ_u^2 .
- Contrastar las dos hipótesis del apartado b), simultáneamente, con $\beta_3 + \beta_1 = 5$.

Problema 4.33. Obtenga la expresión analítica del contraste de multiplicadores de Lagrange para el contraste de la significación conjunta de un subvector de coeficientes del modelo de regresión lineal múltiple.

Problema 4.34. Derive el test de Razón de Verosimilitudes para el contraste de la significación conjunta de un subvector de variables en un modelo de regresión lineal.

Problema 4.35. Derive el test de Razón de Verosimilitudes para el contraste de la hipótesis de cambio estructural.

Problema 4.36. Probar que en el modelo $y_t = \alpha + \beta x_t + u_t$ el contraste de la hipótesis nula $H_0: \beta = \beta^0$ puede llevarse a cabo de modo equivalente mediante el estadístico F , o sustituyendo la restricción directamente en el modelo, y estimando el modelo resultante.

Problema 4.37. Demuestre que la inclusión de una variable explicativa adicional en un modelo econométrico hace aumentar el coeficiente de determinación ajustado \bar{R}^2 si y sólo si el estadístico t de dicha variable es superior a la unidad.

Problema 4.38. Demuestre que la inclusión de r variables explicativas adicionales reduce la estimación de la varianza residual si y sólo si el estadístico F conjunto de dichas variables excede a la unidad.

CAPITULO 5

MATRICES DE COVARIANZAS NO ESCALARES

5.1. INTRODUCCION

Hemos analizado en los Capítulos 3 y 4 las propiedades del estimador MCO de los coeficientes de un modelo lineal cuyo término de error tiene una matriz de covarianzas escalar: todos sus elementos son cero, excepto los de la diagonal principal, y éstos son todos iguales a σ_u^2 . Existen, sin embargo, dos situaciones en que dicha matriz de covarianzas tiene una estructura más compleja, en cuyo caso las propiedades analizadas en los capítulos previos podrían dejar de ser válidas.

Una de tales situaciones se produce cuando la varianza del término de error es distinta de unas observaciones a otras. Con datos de series temporales, ello ocurre cuando $\text{Var}(u_t) = \sigma_t^2$, con $\sigma_t^2 \neq \sigma_s^2$ para $t \neq s$. En el caso de datos de sección cruzada, esta situación se produce cuando $\text{Var}(u_i) = \sigma_i^2$, con $\sigma_i^2 \neq \sigma_j^2$ para $i \neq j$. En ambos casos, los elementos de la diagonal principal de la matriz de covarianzas no serán iguales entre sí. A esta situación se le denomina como de *heteroscedasticidad*, a diferencia del caso en que la varianza de u_t es constante, que se conoce como *homoscedasticidad*.

Una segunda situación ocurre cuando, disponiendo de datos de series temporales, los términos de error correspondientes a distintos períodos no son independientes entre sí, es decir, $\text{Cov}(u_t, u_{t-k}) = E(u_t \cdot u_{t-k}) \neq 0$ para algún $k > 0$. Ello hace que los elementos de fuera de la diagonal en la matriz de covarianzas no sean *todos* nulos, por lo que ésta no será diagonal. A esta situación se la denomina como de *autocorrelación*, para reflejar el hecho de que el término de error está correlacionado consigo mismo a través del tiempo.

Aunque esta propiedad está naturalmente asociada a los datos de series temporales, puede presentarse en muestras de sección cruzada una situación similar: si la ordenación de las observaciones en la muestra sigue unos criterios «geográficos», como pueden ser las distintas áreas de un país, o las familias de un vecindario, entonces los términos de error correspondientes a observaciones sucesivas pueden estar correlacionados entre sí; esta situación se deno-

mina como *correlación espacial*, para la que se han desarrollado algunos procedimientos específicos que no tratamos en este texto.

En bastantes ocasiones no se presta en el trabajo económico aplicado suficiente atención a la existencia de estas dificultades, salvo que aparezca evidencia clara en sentido contrario. Sin embargo, tan plausibles son ambas situaciones que un investigador riguroso debe comenzar su trabajo aplicado admitiendo su posible presencia —que deberá contrastar por los procedimientos adecuados—, según el tipo de datos que utilice. Especialmente, cuando se trabaja con datos temporales, cabe considerar la presencia tanto de heteroscedasticidad como de autocorrelación. En consecuencia, el investigador debe utilizar los procedimientos de estimación e inferencia que desarrollaremos en este y sucesivos capítulos, salvo que pueda justificar que la presencia de ambas situaciones no se produce en una magnitud suficiente como para que sea precisa la utilización de procedimientos más complejos que los de MCO.

La primera cuestión que debemos plantear se refiere a las propiedades del estimador MCO en presencia de heteroscedasticidad y/o autocorrelación. Ya vimos en el Capítulo 3 que, cuando la matriz de covarianzas del término de error es escalar, dicho estimador es óptimo. Así, las preguntas que vamos a responder en este capítulo son:

1. ¿Es el estimador MCO insesgado cuando el término de error tiene una matriz de covarianzas que no es escalar?

2. Si es insesgado, ¿es óptimo (es decir, el de mínima varianza entre los estimadores lineales e insesgados), al igual que ocurría con una matriz de covarianzas escalar?

3. Si no es óptimo, ¿existe algún otro estimador con menor varianza, es decir, más eficiente?

4. Más generalmente, ¿en qué medida dependen las propiedades del estimador MCO de la estructura de la matriz de covarianzas del vector \mathbf{u} ? Por ejemplo, ¿es el estimador de mínimos cuadrados ordinarios aun de máxima verosimilitud?

5.2. PROPIEDADES DEL ESTIMADOR DE MINIMOS CUADRADOS ORDINARIOS

La definición del estimador MCO que hicimos en la Sección 3.2 no depende de la estructura de la matriz de covarianzas del término de error del modelo econométrico. Por consiguiente, podemos establecer la siguiente definición, que es más general, por ser aplicable a casos en que la matriz de covarianzas no es escalar.

Definición 5.1. Dado el modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde \mathbf{y} , \mathbf{u} son $T \times 1$ y \mathbf{X} la matriz $T \times k$ de observaciones de las variables explicativas, con $E(\mathbf{u}) = \mathbf{0}$ y $\text{Var}(\mathbf{u}) = \sigma_u^2 \boldsymbol{\Sigma}$, el estimador MCO del vector de parámetros $\boldsymbol{\beta}$ es una solución al sistema de ecuaciones normales:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Cuando la matriz $\mathbf{X}'\mathbf{X}$ es invertible, entonces dicho sistema tiene una solución única, dada por $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, y al igual que probamos al demostrar la Proposición 3.1, se tiene:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \quad [5.1]$$

Cuando $\boldsymbol{\Sigma} = \mathbf{I}_T$, la matriz de covarianzas es escalar y los resultados que veremos en este capítulo acerca del estimador MCO se reducen a los que vimos en los Capítulos 3 y 4. Así, por ejemplo, la definición del estimador de mínimos cuadrados que acabamos de proponer incluye claramente a la definición de la Sección 3.2 en el caso en que $\text{Var}(\mathbf{u}) = \sigma_u^2\mathbf{I}_T$.

5.2.a. El estimador MCO es insesgado

Una implicación de la expresión [5.1] es que las condiciones para que el estimador de mínimos cuadrados ordinarios sea insesgado cuando $\text{Var}(\mathbf{u}) = \sigma_u^2\boldsymbol{\Sigma}$ son las mismas que eran precisas para que dicho estimador fuese insesgado cuando $\text{Var}(\mathbf{u}) = \sigma^2\mathbf{I}_T$, es decir:

$$E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \mathbf{0}_k \quad [5.2]$$

Proposición 5.1. Si las variables explicativas del modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ son deterministas y $\text{Var}(\mathbf{u}) = \sigma_u^2\boldsymbol{\Sigma}$, entonces el estimador de mínimos cuadrados ordinarios es insesgado.

Demostración. Si las variables \mathbf{X} son deterministas, entonces claramente se cumple la condición [5.2], siempre que $E(\mathbf{u}) = \mathbf{0}_T$, puesto que:

$$E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0}_T = \mathbf{0}_k$$

5.2.b. Matriz de covarianzas del estimador MCO

Proposición 5.2. Si $\text{Var}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma_u^2\boldsymbol{\Sigma}$ y las variables explicativas son deterministas, entonces la matriz de covarianzas del estimador MCO es:

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{MCO}}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Demostración:

$$\begin{aligned} \text{Var } \hat{\boldsymbol{\beta}}_{\text{MCO}} &= E[(\hat{\boldsymbol{\beta}}_{\text{MCO}} - E\hat{\boldsymbol{\beta}}_{\text{MCO}})(\hat{\boldsymbol{\beta}}_{\text{MCO}} - E\hat{\boldsymbol{\beta}}_{\text{MCO}})'] = \\ &= \text{según [5.1]} = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad [5.3]$$

Una vez obtenidas las primeras propiedades del estimador MCO cuando la matriz de covarianzas del término de error no es escalar, conviene hacer algunas observaciones importantes:

1. La expresión [5.3] es muy distinta en apariencia de la matriz de covarianzas $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ del estimador MCO cuando $\text{Var}(\mathbf{u}) = \sigma_u^2\mathbf{I}_T$. Sin embargo, es fácil ver que cuando $\Sigma = \mathbf{I}_T$, la matriz de covarianzas [5.3] se reduce a la matriz $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$.

En consecuencia, no hay ninguna razón por la que deba imponerse como matriz de covarianzas del estimador MCO una expresión restringida como es $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, sino que más bien debería utilizarse [5.3] y permitir así que en las circunstancias apropiadas (cuando no hay heteroscedasticidad ni autocorrelación) dicha matriz se reduzca numéricamente a $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$.

2. Por tanto, la expresión [5.3] es *realmente* la matriz de covarianzas del estimador MCO. Como ya hemos mencionado, no se trata de discutir si en un modelo de regresión existe heteroscedasticidad y/o autocorrelación, sino que más bien, dando como un hecho el que ambos problemas están siempre presentes en todo trabajo econométrico, entonces la cuestión reside en analizar en qué medida aparecen en cada aplicación empírica. Suponer ausencia de autocorrelación o de heteroscedasticidad es siempre tan sólo una aproximación, especialmente si el supuesto se adopta sin contrastación.

Utilizar $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ como matriz de covarianzas del estimador MCO cuando puede existir autocorrelación y/o heteroscedasticidad es simplemente *incorrecto* y puede conducir a errores de inferencia arbitrarios. Además, al no cumplirse las hipótesis básicas, los estadísticos que introdujimos en el Capítulo 4 no tienen las distribuciones t y F .

3. Como muestra [5.1], el estimador MCO está definido en este caso más general por la misma expresión que cuando $\Sigma = \mathbf{I}_T$ y es, por tanto, una función lineal del término de error \mathbf{u} . Por ello, sigue siendo cierto que si las variables \mathbf{X} son deterministas y el término de error tiene una distribución Normal, es decir, $\mathbf{u} \sim N_T(\mathbf{0}_T, \sigma_u^2\Sigma)$, entonces el estimador MCO también se distribuye normalmente, con el vector de esperanzas y matriz de covarianzas que hemos calculado en las Proposiciones 5.1 y 5.2:

$$\hat{\beta}_{\text{MCO}} \sim N_k(\beta; \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})$$

5.3. EL ESTIMADOR DE MINIMOS CUADRADOS GENERALIZADOS

El supuesto de que la matriz de covarianzas de \mathbf{u} es $\sigma_u^2\Sigma$ con $\Sigma \neq \mathbf{I}_T$ implica, como acabamos de ver, que la matriz de covarianzas del estimador MCO ya no tiene la expresión tan sencilla que obtuvimos en la Proposición 3.2. Este puede ser un cambio importante, porque allí probamos que, cuando $\text{Var}(\mathbf{u}) = \sigma_u^2\mathbf{I}_T$, el estimador MCO era de mínima varianza entre los estimadores lineales e insesgados; ahora bien, si bajo las nuevas propiedades del término de error la matriz de covarianzas de $\hat{\beta}_{\text{MCO}}$ es diferente, podría ocurrir

que $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ ya no fuese la menor matriz de covarianzas entre todos los estimadores de la clase mencionada.

5.3.a. Estimación del vector de coeficientes β . El estimador de mínimos cuadrados generalizados

En estas circunstancias, sería interesante poder transformar el modelo económico en otro cuyos coeficientes fuesen los mismos que los del modelo original, pero cuyo término de error tuviese una matriz de covarianzas escalar. Si ello fuese posible, entonces podríamos utilizar el estimador MCO en este modelo transformado. Por los resultados del Capítulo 3, dicho estimador sería eficiente, es decir, el de mínima varianza entre los estimadores que son lineales e insesgados.

Para ello premultipliquemos el modelo por una matriz \mathbf{P} de dimensiones $T \times T$, para obtener:

$$\mathbf{P}\mathbf{y} = (\mathbf{P}\mathbf{X})\boldsymbol{\beta} + \mathbf{P}\mathbf{u} \quad [5.4]$$

y denotemos $\mathbf{y}^* = \mathbf{P}\mathbf{y}$, $\mathbf{X}^* = \mathbf{P}\mathbf{X}$, $\mathbf{u}^* = \mathbf{P}\mathbf{u}$. Notemos que \mathbf{y}^* y \mathbf{u}^* son todavía vectores $T \times 1$ y \mathbf{X}^* una matriz $T \times k$. La transformación efectuada es, en realidad, un cambio de variables. Cada observación de la nueva variable \mathbf{y}^* es una combinación lineal de todas las observaciones contenidas en el vector original \mathbf{y} ; los coeficientes en dichas combinaciones lineales son las filas de la matriz \mathbf{P} .

Algo similar ocurre con las variables explicativas. Ahora tenemos una matriz \mathbf{X}^* de T observaciones para cada una de las k nuevas variables. La columna i -ésima contiene T observaciones de la variable x_i^* , cada una de las cuales es una combinación lineal de *todas* las observaciones de la variable x_i del modelo original, aunque no depende de ninguna otra variable explicativa. En general, ninguna de estas nuevas variables, y^* , x_1^* , ..., x_k^* , tiene un significado económico claro.

Por otra parte, la linealidad del modelo permite que los coeficientes $\boldsymbol{\beta}$ del modelo transformado sean precisamente los mismos que los del modelo original. La matriz de covarianzas del nuevo término de error es:

$$\text{Var}(\mathbf{u}^*) = \text{Var}(\mathbf{P}\mathbf{u}) = \sigma_u^2 \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}' \quad [5.5]$$

La pregunta que nos hacemos es: ¿Existe una matriz \mathbf{P} tal que el término de error del modelo transformado tenga como matriz de covarianzas $\text{Var}(\mathbf{u}^*) = \sigma_u^2 \mathbf{I}_T$? La matriz $\boldsymbol{\Sigma}$ es simétrica, definida positiva, y hemos visto en el Capítulo 1. que en tal caso siempre existe una matriz cuadrada, no singular, \mathbf{V} de modo que $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}'$, o, equivalentemente:

$$\mathbf{V}^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{-1})' = \mathbf{I}_T \quad \text{y por tanto,} \quad \boldsymbol{\Sigma}^{-1} = (\mathbf{V}^{-1})'\mathbf{V}^{-1} \quad [5.6]$$

Basta comparar las expresiones [5.5] y [5.6] para notar que si se utiliza precisamente la matriz V^{-1} como matriz de transformación P , entonces el término de error resultante tiene matriz de covarianzas escalar. En efecto, una vez que se ha obtenido el modelo transformado:

$$y^* = X^* \beta + u^* \quad [5.7]$$

con $y^* = V^{-1}y$, $X^* = V^{-1}X$, $u^* = V^{-1}u$, se tiene $E(u^*) = E(V^{-1}u) = \mathbf{0}_T$, $\text{Var}(u^*) = \sigma_u^2 V^{-1} \Sigma (V^{-1})' = \sigma_u^2 I_T$, por lo que el modelo [5.7] satisface las condiciones bajo las que obtuvimos los resultados del Capítulo 3. En particular, el estimador de MCO de los parámetros del modelo [5.7] es

$$\hat{\beta}_{\text{MCG}} = (X^{*'} X^*)^{-1} X^{*'} y^*$$

y recibe el nombre de *estimador de mínimos cuadrados generalizados* de los coeficientes β del modelo original (que denotamos en lo sucesivo por MCG).

En función de las variables originales se tiene que:

$$\hat{\beta}_{\text{MCG}} = (X'(V^{-1})'V^{-1}X)^{-1} X'(V^{-1})'V^{-1}y = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y \quad [5.8]$$

Es claro que este estimador no es equivalente a aplicar MCO directamente al modelo original $y = X\beta + u$, puesto que las variables que en él aparecen son distintas de las del modelo [5.7] que se ha estimado.

5.3.b. Propiedades del estimador MCG

Por obtenerse el estimador MCG mediante la aplicación del procedimiento MCO a un modelo transformado, podemos asegurar inmediatamente algunas de sus propiedades, como:

a) El estimador MCG es *insesgado* puesto que $\hat{\beta}_{\text{MCG}} = \beta + (X^{*'} X^*)^{-1} X^{*'} u^*$, y como $E(u^*) = \mathbf{0}_T$, entonces $E(\hat{\beta}_{\text{MCG}}) = \beta$.

b) **Matriz de covarianzas** del estimador MCG:

$$\text{Var}(\hat{\beta}_{\text{MCG}}) = \sigma_u^2 (X^{*'} X^*)^{-1} = \sigma_u^2 (X'(V^{-1})'V^{-1}X)^{-1} = \sigma_u^2 (X'\Sigma^{-1}X)^{-1} \quad [5.9]$$

c) **Ecuaciones normales:** al igual que el estimador MCO, el estimador MCG también satisface un sistema de ecuaciones normales. En efecto, puesto que este estimador no es sino el estimador MCO del modelo transformado, se tiene $(X^{*'} X^*) \hat{\beta}_{\text{MCG}} = X^{*'} y^*$ o, lo que es lo mismo:

$$(X'\Sigma^{-1}X) \hat{\beta}_{\text{MCG}} = X'\Sigma^{-1}y$$

que es el sistema de ecuaciones normales del estimador MCG.

d) **Obtención** del estimador MCG: de la discusión que acabamos de realizar se deduce que hay dos formas equivalentes de obtener el estimador de MCG de un modelo econométrico:

- Descomponiendo la matriz Σ como el producto VV' , transformando las matrices de datos del modelo original mediante premultiplicación por la inversa de la matriz V y estimando por MCO con dichas variables transformadas.
- Utilizando la expresión matricial que aparece en [5.8].

La segunda alternativa exige invertir una matriz de orden $T \times T$, así como efectuar unos productos matriciales que pueden resultar bastante exhaustivos computacionalmente. Por eso, como veremos, generalmente se adopta la primera alternativa, tras haber hecho una determinada hipótesis acerca de la estructura de la matriz Σ . Ello nos permitirá establecer fácilmente la forma de la matriz V^{-1} y así llevar a cabo la transformación del modelo original. Todo ello requerirá invertir tan sólo la matriz X^*X^* , cuya dimensión, $k \times k$, no crece con el tamaño muestral.

e) **Eficiencia** del estimador MCG:

Proposición 5.3. El estimador MCG es el estimador lineal insesgado de mínima varianza del vector de parámetros β .

Demostración. Como el modelo transformado [5.7] satisface las condiciones supuestas en el Capítulo 3, entonces, por el teorema de Gauss-Markov, (Proposición 3.6), el estimador MCO de [5.7] es el estimador lineal, insesgado de mínima varianza del vector β . Pero ese estimador es *precisamente* el estimador MCG del vector β en el modelo original, que es el mismo vector de coeficientes que el del modelo transformado y, por consiguiente, tiene también esta propiedad de optimalidad.

En particular, el estimador MCG es más eficiente que el estimador MCO, por lo que la diferencia

$$(X'\Sigma^{-1}X)^{-1} - (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

es semidefinida negativa. Sin embargo, el estimador MCG no es siempre estrictamente más eficiente que el estimador MCO. Rao (1968) probó el siguiente lema (que no demostramos):

Lema 5.1. Sea Z una matriz $T \times r$ tal que $X'Z = 0$. Los estimadores MCO y MCG, así como sus matrices de covarianzas, son idénticos si y sólo si la matriz Σ tiene la forma:

$$\Sigma = XCX' + ZDZ' + \sigma^2I_T$$

donde C y D son matrices definidas no negativas, pero arbitrarias.

La condición de equivalencia de los estimadores MCO y MCG dada en este lema es analíticamente compleja; sin embargo, un caso sencillo en que dicha condición se satisface es cuando:

$$\Sigma = E(\mathbf{uu}') = \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \dots & \dots & \dots & \dots & \dots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

puesto que entonces:

$$\Sigma = \sigma_u^2(1 - \rho)\mathbf{I}_T + \sigma_u^2\rho\mathbf{1}_T\mathbf{1}_T' = \sigma^2\mathbf{I}_T + \mathbf{X}\mathbf{C}\mathbf{X}'$$

donde:

$$\mathbf{C} = \begin{pmatrix} \rho\sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \quad \text{y} \quad \sigma^2 = \sigma_u^2(1 - \rho)$$

5.3.c. Estimación del parámetro σ_u^2

El problema fundamental en la obtención del estimador MCG es que tanto σ_u^2 como Σ son desconocidos, y se requiere una estimación previa de los mismos. Sin embargo, como puede verse en [5.8], al igual que ocurría con el estimador MCO, no es preciso conocer el parámetro σ_u^2 para obtener el estimador $\hat{\beta}_{\text{MCG}}$, aunque sí para calcular su matriz de covarianzas, de acuerdo con [5.9].

La estimación de mínimos cuadrados de $\hat{\sigma}_{\text{MCG}}^2$ se obtiene a partir del modelo [5.7], del modo que se vio en la Sección 3.4, es decir:

$$\hat{\sigma}_{\text{MCG}}^2 = \frac{\hat{\mathbf{u}}^{*\prime}\hat{\mathbf{u}}^*}{T - k}$$

o, en función de las variables del modelo original:

$$\hat{\sigma}_{\text{MCG}}^2 = \frac{\hat{\mathbf{u}}_{\text{MCG}}'(\mathbf{V}^{-1})'\mathbf{V}^{-1}\hat{\mathbf{u}}_{\text{MCG}}}{T - k} = \frac{\hat{\mathbf{u}}_{\text{MCG}}'\Sigma^{-1}\hat{\mathbf{u}}_{\text{MCG}}}{T - k}$$

donde $\hat{\mathbf{u}}_{\text{MCG}} = \mathbf{y} - \mathbf{X}\hat{\beta}_{\text{MCG}}$.

Proposición 5.4. El estimador $\hat{\sigma}_{\text{MCG}}^2$ es insesgado.

Demostración. Puesto que $\text{Var}(\mathbf{u}^*) = \sigma_u^2\mathbf{I}_T$, puede utilizarse en el modelo transformado [5.7] el análisis que acerca de las propiedades del estimador

de σ_u^2 se llevó a cabo en la Sección 3.4. En particular, allí se probó que tras estimar por MCO, los residuos obtenidos satisfacen $E(\hat{\mathbf{u}}^* \hat{\mathbf{u}}^*) = (T - k)\sigma_u^2$, por lo que $\hat{\sigma}_{\text{MCG}}^2$ es un estimador insesgado de σ_u^2 .

La suma de cuadrados de residuos del modelo transformado no es igual a la suma de cuadrados de los residuos del modelo original, sino a la forma cuadrática $\hat{\mathbf{u}}'_{\text{MCG}} \hat{\Sigma}^{-1} \hat{\mathbf{u}}_{\text{MCG}}$ que con ellos se obtiene. Ello sugiere que, para mantener una nomenclatura análoga a la hasta ahora utilizada, definamos la suma residual de un modo más preciso.

Definición 5.2. La suma residual del modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, con $\text{Var}(\mathbf{u}) = \sigma_u^2 \boldsymbol{\Sigma}$, es $\text{SR} = \hat{\mathbf{u}}'_{\text{MCG}} \boldsymbol{\Sigma}^{-1} \hat{\mathbf{u}}_{\text{MCG}}$.

Es decir, que la definición de suma residual que vimos en la Sección 3.3 como la suma de los cuadrados de los residuos no tiene carácter general y sólo es válida en un modelo cuyo término de error tiene matriz de covarianzas escalar. Nótese, sin embargo, que una vez más, si la matriz $\boldsymbol{\Sigma}$ resultase ser escalar, ambas definiciones coincidirían.

Es fácil comprobar que una expresión alternativa para el estimador MCG de σ_u^2 es:

$$\hat{\sigma}_{\text{MCG}}^2 = \frac{\mathbf{y}' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \hat{\boldsymbol{\beta}}'_{\text{MCG}} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y}}{T - k}$$

puesto que el numerador de este cociente es una expresión alternativa para la suma residual que acabamos de definir.

5.3.d. El coeficiente de determinación

Una dificultad que aparece en este contexto es el no poder utilizar el estadístico R^2 como medida de ajuste del modelo. En primer lugar, el modelo transformado puede no tener término constante —especialmente en situaciones de heteroscedasticidad, como veremos—, por lo que el R^2 calculado en este modelo no está acotado entre 0 y 1. En cualquier caso, con él mediríamos la capacidad de explicar \mathbf{y}^* , que no es la variable de interés. También podríamos calcular el estadístico R^2 con las variables del modelo original, pero nuevamente no podemos garantizar que su valor numérico esté dentro del intervalo unidad.

Como resumen de las ideas introducidas en esta sección, que serán de gran utilidad a lo largo de los próximos capítulos, digamos que el investigador económico debe reconocer que las situaciones de autocorrelación y/o heteroscedasticidad son la regla y no la excepción, y que debe hacerse siempre un análisis tan exhaustivo como sea posible acerca del grado en que estos problemas están presentes en una aplicación empírica.

En consecuencia, si quiere utilizarse el procedimiento de mínimos cuadrados para estimar los coeficientes de un modelo de regresión lineal, hay que

recordar que el *estimador eficiente* es el de mínimos cuadrados generalizados [5.7] cuya matriz de covarianzas es $\text{Var}(\hat{\beta}_{\text{MCG}}) = \sigma_u^2 \mathbf{X}'\Sigma^{-1}\mathbf{X}$. Si, por alguna razón, se estima por mínimos cuadrados ordinarios, es importante recordar que su matriz de covarianzas es $\text{Var}(\hat{\beta}_{\text{MCO}}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\Sigma^{-1}\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1}$ y que utilizar $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$ como matriz de covarianzas $\hat{\beta}_{\text{MCO}}$ es inapropiado, por constituir un estimador sesgado de $\text{Var}(\hat{\beta}_{\text{MCO}})$.

Como veremos en algunos ejercicios de simulación, dicho sesgo puede ser además de gran magnitud.

Por otra parte, si se dan las condiciones bajo las cuales el estimador MCO tiene buenas propiedades estadísticas (es decir, si $\Sigma = \mathbf{I}_T$), entonces la utilización de las expresiones del estimador MCG, tanto en lo que respecta al vector β como a su matriz de varianzas, conducirá al estimador MCO.

5.4. INTRODUCCION AL PROBLEMA DE LA HETEROSCEDASTICIDAD

Como ejemplo de la situación de estimación que estamos discutiendo en este capítulo, supongamos que el término de error del modelo lineal

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t$$

tiene matriz de covarianzas:

$$\text{Var}(\mathbf{u}) = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \sigma_T^2 \end{pmatrix} \quad [5.10]$$

es decir, que la varianza σ_t^2 del término de error u_t varía a lo largo del tiempo. Sin embargo, seguimos suponiendo que $E(u_t u_s) = 0$ para todo $t \neq s$, es decir, que los términos de error correspondientes a dos instantes de tiempo distintos son independientes entre sí. Ello hace que los elementos de fuera de la diagonal principal de la matriz $\text{Var}(\mathbf{u})$ sean cero.

Supongamos que los parámetros $\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2$ son conocidos. Entonces la matriz de covarianzas se puede descomponer:

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \sigma_T^2 \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_T \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_T \end{pmatrix} = \mathbf{V}\mathbf{V}'$$

y se tiene:

$$\mathbf{V}^{-1} = \begin{pmatrix} 1/\sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1/\sigma_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1/\sigma_T \end{pmatrix}$$

Si utilizamos esta matriz para transformar las variables $y_t, x_{1t}, x_{2t}, \dots, x_{kt}$, se tiene:

$$\mathbf{y}^* = \mathbf{V}^{-1}\mathbf{y} = \begin{pmatrix} y_1/\sigma_1 \\ y_2/\sigma_2 \\ y_3/\sigma_3 \\ \cdots \\ y_T/\sigma_T \end{pmatrix};$$

[5.11]

$$\mathbf{X}^* = \mathbf{V}^{-1}\mathbf{X} = \begin{pmatrix} x_{11}/\sigma_1 & x_{21}/\sigma_1 & x_{31}/\sigma_1 & \cdots & x_{k1}/\sigma_1 \\ x_{12}/\sigma_2 & x_{22}/\sigma_2 & x_{32}/\sigma_2 & \cdots & x_{k2}/\sigma_2 \\ x_{13}/\sigma_3 & x_{23}/\sigma_3 & x_{33}/\sigma_3 & \cdots & x_{k3}/\sigma_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1T}/\sigma_T & x_{2T}/\sigma_T & x_{3T}/\sigma_T & \cdots & x_{kT}/\sigma_T \end{pmatrix}$$

y el estimador MCO en la regresión de \mathbf{y}^* sobre \mathbf{X}^* es el estimador MCG del modelo original:

$$\hat{\beta}_{\text{MCG}} = \begin{pmatrix} \sum_1^T(x_{1t}^2/\sigma_t^2) & \sum_1^T(x_{1t}x_{2t}/\sigma_t^2) & \cdots & \sum_1^T(x_{1t}x_{kt}/\sigma_t^2) \\ \cdots & \sum_1^T(x_{2t}^2/\sigma_t^2) & \cdots & \sum_1^T(x_{2t}x_{kt}/\sigma_t^2) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \sum_1^T(x_{kt}^2/\sigma_t^2) \end{pmatrix}^{-1} \begin{pmatrix} \sum_1^T(x_{1t}y_t/\sigma_t^2) \\ \sum_1^T(x_{2t}y_t/\sigma_t^2) \\ \cdots \\ \sum_1^T(x_{kt}y_t/\sigma_t^2) \end{pmatrix}$$

[5.12]

Consideremos ahora el modelo más sencillo:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad \text{Var}(u_t) = \sigma_t^2$$

en el que se tiene, utilizando [5.12] que:

$$\hat{\beta}_{\text{MCG}} = \begin{pmatrix} \sum_1^T (1/\sigma_t^2) & \sum_1^T (x_{2t}/\sigma_t^2) \\ \sum_1^T (x_{2t}/\sigma_t^2) & \sum_1^T (x_{2t}^2/\sigma_t^2) \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_1^T (y_t/\sigma_t^2) \\ \sum_1^T (y_t x_{2t}/\sigma_t^2) \end{pmatrix} \quad [5.13]$$

Ahora, si en vez de transformar las variables del modelo original mediante la matriz V^{-1} utilizásemos la estructura [5.10] de la matriz de covarianzas del vector u en la expresión matricial [5.8], se tendría:

$$\hat{\beta}_{\text{MCG}} = \begin{bmatrix} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{21} & x_{22} & x_{23} & \dots & x_{2T} \end{pmatrix} \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & 1/\sigma_3^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/\sigma_T^2 \end{pmatrix} \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ \dots & \dots \\ 1 & x_{2T} \end{pmatrix} \\ \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{21} & x_{22} & x_{23} & \dots & x_{2T} \end{pmatrix} \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & 1/\sigma_3^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/\sigma_T^2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_T \end{pmatrix} \end{bmatrix}^{-1} \quad [5.14]$$

de donde se obtiene exactamente la expresión [5.12], lo cual no puede resultar sorprendente, ya que no hemos hecho sino utilizar dos formas equivalentes de obtener el estimador MCG.

Con la estructura de la matriz Σ que hemos supuesto, las expresiones [5.11] para y^* , X^* muestran que el método de MCG puede interpretarse como el resultado de dividir las observaciones de todas las variables en cada instante t por la desviación típica del término de error, σ_t , para luego aplicar MCO.

Lo que se hace así es ponderar cada período con un peso de $\frac{1}{\sigma_t}$, menor cuanto mayor sea la varianza del término de error de dicho período. Parece razonable que así se haga: una mayor varianza σ_t^2 implica un mayor error muestral en la observación de la variable endógena y_t ; por consiguiente, debería asignarse una menor importancia a dicha observación en el proceso de estimación del modelo. Por esta razón, la aplicación de mínimos cuadrados generalizados con una matriz Σ del tipo aquí considerado se conoce también como *mínimos cuadrados ponderados*. Por el contrario, el estimador MCO pondera a todas las observaciones muestrales de igual forma.

Un problema fundamental en la práctica es que se desconocen los valores de los parámetros $\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2$, por lo que hay que proceder a su estimación antes de poder calcular el estimador MCG, ya sea por medio de la expresión

matricial [5.8] o mediante la estimación del modelo transformado [5.7]. Sin embargo, incluso cuando los elementos de fuera de la diagonal principal de la matriz Σ son cero (es decir, incluso si sólo existe heteroscedasticidad en el término de error), la matriz de covarianzas contendría T parámetros diferentes a lo largo de la diagonal, que habría que estimar. Lamentablemente, carecemos de suficiente información para estimar el modelo, pues con cada nueva observación aparece un nuevo parámetro por estimar, σ_t^2 . Por tanto, se hace preciso parametrizar la estructura de varianzas del término de error de modo que la sucesión $\sigma_1^2, \sigma_2^2, \dots, \sigma_T^2$, y, por tanto, la matriz Σ dependa tan sólo de un número reducido de parámetros.

Por ejemplo, supongamos que el investigador está dispuesto a admitir a priori que $\sigma_t^2 = \sigma_u^2 x_{2t}$, es decir, que la varianza de u_t es en cada periodo proporcional al valor que toma una de las variables explicativas del modelo. Bajo este supuesto se tiene:

$$\Sigma = \sigma_u^2 \begin{pmatrix} x_{21} & 0 & 0 & 0 \\ 0 & x_{22} & 0 & 0 \\ 0 & 0 & x_{23} & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & x_{2T} \end{pmatrix}$$

por lo que la matriz Σ se hace depender de tan sólo un parámetro desconocido σ_u^2 , que habría que estimar.

La hipótesis que acabamos de hacer acerca de la estructura de la matriz Σ es bastante verosímil en modelos econométricos estimados con datos de sección cruzada. Supongamos que se estima una función de consumo en la que los gastos de consumo de cada familia se hacen depender de su nivel de ingresos. Las familias de renta elevada tienen una amplia gama de posibilidades de uso de sus ingresos o excedentes, una vez satisfechas sus necesidades mínimas; por tanto, la variabilidad de la componente de los gastos de consumo de la unidad familiar que no viene explicada por su nivel de ingresos será mayor en estas familias. Ello puede formalizarse analíticamente haciendo que la desviación típica del término de error correspondiente a cada observación crezca con la variable renta, exactamente como en el ejemplo anterior.

5.5. INTRODUCCION AL PROBLEMA DE AUTOCORRELACION

El problema de estimar la matriz Σ es aún más importante cuando, además de heteroscedasticidad, existe autocorrelación, pues en tal caso los elementos de fuera de la diagonal de la matriz Σ son, en principio, diferentes de cero,

por lo que dicha matriz contiene potencialmente $\frac{T(T+1)}{2}$ parámetros distintos, que deben ser estimados para obtener el estimador de mínimos cuadrados. Si, por ejemplo, la muestra consta de 100 observaciones, entonces dicho número de parámetros desconocidos sería de 5050, y su estimación se convierte en una tarea estadística sin solución.

Supongamos ahora que la matriz Σ tiene los elementos de la diagonal iguales a 1, es decir, no hay heteroscedasticidad. En tal caso, aún quedan los $\frac{T(T-1)}{2}$ elementos de fuera de dicha diagonal por estimar. Este es un número muy elevado en cuanto T sea algo grande, y además crece más rápido que el tamaño muestral, por lo que aparece de nuevo la necesidad de representar Σ como función de un número reducido de parámetros.

Una posible parametrización podría obtenerse mediante una hipótesis sencilla acerca del comportamiento intertemporal del término de error del modelo, como ocurre, por ejemplo, en el modelo de *autocorrelación autorregresiva de primer orden*:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

donde $|\rho| < 1$ es un parámetro desconocido y ε_t , en cada instante, una variable aleatoria con $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ y $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$. En el Capítulo 7 veremos que el supuesto $|\rho| < 1$ implica que u_t depende sólo de ε_t y de valores pasados de ε_t , pero no depende de valores futuros de ε_t , propiedad que tomamos aquí como probada. El siguiente lema, que demostramos en un capítulo posterior, proporciona la forma de dicha dependencia:

Lema 5.2. Si en el modelo $u_t = \rho u_{t-1} + \varepsilon_t$ se tiene $|\rho| < 1$, entonces u_t admite la representación:

$$u_t = \sum_0^\infty \rho^i \varepsilon_{t-i}$$

Utilizando este lema, junto con los supuestos que hicimos en el párrafo anterior acerca de las propiedades de ε_t , se tiene:

- a) $\sigma_u^2 = \text{Var}(u_t) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$, constante para todo t ,
- b) $\sigma_{12} = \text{Cov}(u_1, u_2) = E(u_1 u_2) = E[u_1(\rho u_1 + \varepsilon_2)] = E(\rho u_1^2 + u_1 \varepsilon_2) = \rho E u_1^2 + E(u_1 \varepsilon_2) = \rho \sigma_u^2$

donde hemos utilizado el hecho $E(u_1^2) = \text{Var}(u_1) = \sigma_u^2$ y también que $E(u_1 \varepsilon_2) = 0$. Ello se debe a que u_1 depende tan sólo de ε_1 y anteriores, junto con el supuesto de que ε_t no tiene autocorrelación intertemporal.

El mismo razonamiento puede utilizarse para probar que $E(u_{t-1}u_t) = \rho\sigma_u^2$ para todo t . Por otra parte:

$$c) \quad E(u_{t-2}u_t) = E[u_{t-2}(\rho u_{t-1} + \varepsilon_t)] = \rho E(u_{t-1}u_{t-2}) + E(u_{t-2}\varepsilon_t) = \rho^2\sigma_u^2 + 0 = \rho^2\sigma_u^2$$

donde se ha utilizado b), así como el hecho de que u_{t-2} dependa de ε_{t-2} y valores previos, pero de ninguno posterior a ε_{t-2} .

En consecuencia, la matriz de covarianzas del término de error resulta ser:

$$\text{Cov}(\mathbf{u}) = \sigma_u^2 \Sigma = \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{T-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & 1 \end{pmatrix} \quad [5.15]$$

que, como puede verse, depende únicamente de dos parámetros desconocidos, ρ y σ_u^2 .

Como ocurría en la situación de heteroscedasticidad, el problema importante en la práctica es que, tras conseguir alguna de estas parametrizaciones simplificadas, los valores de los parámetros (ρ y σ_u^2 en este caso) son desconocidos, por lo que los métodos que antes hemos visto para obtener el estimador de mínimos cuadrados no son operativos. Sin embargo, es posible obtener una estimación $\hat{\Sigma}$ de la matriz Σ y a continuación calcular el estimador de mínimos cuadrados generalizados, bien sea mediante la descomposición de dicha matriz $\hat{\Sigma}$ y posterior transformación de las variables del modelo o utilizando directamente la expresión matricial:

$$\hat{\beta}_{\text{MCG}} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{y}$$

Conviene aclarar que el argumento que hicimos en la Proposición 5.3 para probar que el estimador de MCG es de mínima varianza de entre los estimadores lineales e insesgados de β deja de ser válido si se sustituye la matriz Σ por una estimación suya. Las propiedades del estimador de MCG obtenido mediante la previa estimación de la matriz Σ dependerá de la calidad de dicha estimación $\hat{\Sigma}$. Esta es una importante cuestión que no analizamos en este texto, pero es conveniente saber que los métodos que aquí sugerimos para resolver el problema que surge por desconocimiento de Σ tienen una justificación matemática rigurosa cuando el tamaño muestral tiende a infinito. El lector interesado puede consultar Theil (1971).

En nuestro ejemplo, el modelo podría estimarse inicialmente por MCO, para obtener los residuos \hat{u}_t correspondientes y estimar el parámetro ρ , así

como σ_ε^2 mediante una regresión de \hat{u}_t sobre \hat{u}_{t-1} , de nuevo por MCO. Debe hacerse hincapié en que éstos no serían los residuos MCG de los que hablamos en secciones previas, sino los procedentes de la estimación inicial (ineficiente) mediante MCO.

Una vez obtenido $\hat{\rho}$ y $\hat{\sigma}_\varepsilon^2$, se estima $\hat{\sigma}_u^2$ mediante la relación $\hat{\sigma}_u^2 = \frac{\hat{\sigma}_\varepsilon^2}{1 - \hat{\rho}^2}$.

Así, se puede obtener la estimación de la matriz de covarianzas $\sigma_u^2 \hat{\Sigma}$ que se mostró en [5.15] que se utilizaría, finalmente, para obtener el estimador de MCG. Como veremos en los capítulos siguientes, a pesar de no ser un estimador eficiente, el estimador MCO del modelo original es muy útil en una primera etapa del proceso de estimación por MCG.

5.5.a. Estimación mediante transformación de variables

Veamos en qué consiste la transformación de variables que permitiría obtener el estimador MCG en este modelo con autocorrelación autorregresiva de primer orden. Consideremos el modelo formado por las ecuaciones:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t \quad [5.16]$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \quad [5.17]$$

y llevemos a cabo el siguiente procedimiento:

a) Escribamos [5.16] para el instante $t - 1$:

$$y_{t-1} = \mathbf{x}'_{t-1} \boldsymbol{\beta} + u_{t-1} \quad [5.18]$$

b) Restemos de [5.16] el producto del modelo [5.18] por la constante ρ , para obtener:

$$y_t - \rho y_{t-1} = \mathbf{x}'_t \boldsymbol{\beta} - \rho (\mathbf{x}'_{t-1} \boldsymbol{\beta}) + u_t - \rho u_{t-1} = (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \boldsymbol{\beta} + \varepsilon_t \quad [5.19]$$

donde se ha utilizado el modelo de autocorrelación del término de error u_t para obtener un nuevo modelo [5.19], cuyo término de error ε_t no tiene autocorrelación.

c) Definamos ahora nuevas variables:

$$\begin{aligned} y_t^* &= y_t - \rho y_{t-1}, & t &= 2, 3, \dots, T \\ \mathbf{x}_{it}^* &= \mathbf{x}_{it} - \rho \mathbf{x}_{it-1}, & t &= 2, 3, \dots, T, \quad i = 1, 2, \dots, k \end{aligned} \quad [5.20]$$

con lo que el modelo [5.19] se convierte en:

$$y_t^* = \mathbf{x}_t^{*'} \boldsymbol{\beta} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [5.21]$$

en el que el estimador de mínimos cuadrados ordinarios del vector $\boldsymbol{\beta}$ es de mínima varianza, ya que coincide con el estimador de mínimos cuadrados

generalizados. La única salvedad es que la transformación que acabamos de proponer ignora la primera observación muestral, lo que conduce a una ligera pérdida de eficiencia. Este problema es, sin embargo, prácticamente irrelevante en cuanto el tamaño muestral sea siquiera moderadamente grande.

Como hemos mencionado antes, el parámetro desconocido ρ puede estimarse comenzando de una estimación previa del modelo [5.16] por MCO, es decir, ignorando la autocorrelación de u_t , y a continuación estimando ρ mediante una regresión de \hat{u}_t , sobre \hat{u}_{t-1} , de acuerdo con [5.17]. Con el valor así estimado de ρ se efectúa la transformación de variables [5.20] y se estima β a partir del modelo [5.21]. Es muy frecuente en las aplicaciones prácticas de este método que cuando se estima [5.17] y se obtienen los residuos, éstos presenten todavía alguna autocorrelación. Ello se debe a que provienen de la utilización de una estimación (ineficiente) $\hat{\rho}$ y no del verdadero valor del parámetro desconocido ρ .

Otra razón por la que es posible que no se resuelva en la práctica el problema de la autocorrelación en una primera aplicación del método antes sugerido es porque las estructuras de autocorrelación sencillas del tipo [5.17] se imponen más bien como una aproximación que por el convencimiento del investigador de que dicha estructura puede realmente resumir toda la estructura de autocorrelación del término de error.

Cuando esto ocurre, entonces se itera el procedimiento, es decir, se postula un modelo de comportamiento para ε_t similar a [5.17], que pretenda explicar su estructura de autocorrelación, y se transforman las variables en el modelo [5.21] para obtener un nuevo modelo:

$$y_t^{**} = x_t^{**} \beta + v_t \quad [5.22]$$

donde v_t es supuestamente un proceso sin autocorrelación e y_t^{**} , x_t^{**} se han obtenido de un modo similar a [5.20].

Cuando se llevan a cabo varias de estas sucesivas transformaciones, la interpretación de las variables resultantes es cada vez más difícil, por lo que al investigador le puede quedar la duda de cuál es el significado de los coeficientes β finalmente estimados, así como de cuáles son los verdaderos residuos del modelo, si \hat{u}_t , $\hat{\varepsilon}_t$ o \hat{v}_t . Para ello, conviene tener presente que las sucesivas transformaciones que se efectúen sobre el modelo original se hacen únicamente con el objeto de obtener unas estimaciones de β tan eficientes como sea posible, y dichas estimaciones se asocian a los coeficientes del modelo de partida [5.16], por lo que ni mucho menos es preciso hacer ninguna interpretación ni análisis en el contexto del modelo transformado final.

Consecuentemente, a todos los efectos y, en particular, para llevar a cabo el análisis de inferencia, como veremos más adelante, los residuos del modelo son los que se derivan del modelo original [5.16], en el que se imponen los coeficientes β finalmente obtenidos.

Veremos en el Capítulo 7 que el método de estimación que aquí hemos sugerido para el caso de un modelo en que el término de error tiene estructura autorregresiva de orden 1 (es decir, [5.17]) se puede generalizar a estructuras de autocorrelación más complicadas.

5.5.b. Obtención del estimador MCG mediante productos matriciales

En esta subsección obtenemos el estimador MCG de los modelos [5.16] y [5.17] utilizando directamente la expresión [5.8]. Simultáneamente, probamos que la transformación de variables propuesta en [5.20] coincide aproximadamente con el estimador MCG.

Comencemos invirtiendo la matriz de covarianzas de este modelo, que está dada por [5.15], teniéndose:

$$\Sigma^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}$$

que se puede descomponer como $\Sigma^{-1} = \frac{1}{1-\rho^2} (\mathbf{V}^{-1})' \mathbf{V}^{-1}$, donde

$$\mathbf{V}^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

Como ya vimos anteriormente en un contexto genérico, si se utiliza esta matriz \mathbf{V}^{-1} para transformar las variables del modelo original, se obtiene el estimador de mínimos cuadrados generalizados. Pero dicho producto genera precisamente la misma transformación que hemos sugerido en la sección anterior (5.5.a), excepto la primera observación, que se transforma en:

$$\sqrt{1-\rho^2} y_1 = (\sqrt{1-\rho^2}) \mathbf{x}'_1 \boldsymbol{\beta} + \sqrt{1-\rho^2} u_1$$

El que esta observación merezca un trato especial se debe a que para ella no existe «observación anterior» con la que pudiera llevarse a cabo una transformación como [5.20]. Por otra parte, tiene sentido que la primera observación se transforme del modo señalado, puesto que así su nuevo término de error, $\sqrt{1-\rho^2} u_1$, tiene varianza igual a $(1-\rho^2) \sigma_u^2$, precisamente igual a la varianza $\sigma_\varepsilon^2 = (1-\rho^2) \sigma_u^2$, del término de error transformado de las siguientes observaciones, que es $\varepsilon_t = u_t - \rho u_{t-1}$.

Si ignoramos la primera fila de la matriz V^{-1} , se obtiene la matriz⁽¹⁾ V^{*-1} de dimensión $(T-1) \times T$:

$$V^{*-1} = \begin{pmatrix} -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & -\rho & 1 \end{pmatrix}$$

que, claramente, transforma las variables y_t , x_t y u_t del mismo modo que hemos sugerido en [5.20], excepto la primera observación, que ignora. En efecto, al ser V^* de dimensión $(T-1) \times T$, las matrices transformadas $V^{*-1}y$ y $V^{*-1}X$ son de dimensión $(T-1) \times 1$ y $(T-1) \times k$, respectivamente. Por tanto, el procedimiento sugerido en la subsección 5.5.a genera (aproximadamente, porque se ignora la primera observación) el estimador de mínimos cuadrados generalizados.

La matriz de covarianzas del término de error del modelo transformado es $\sigma_\varepsilon^2 I_T$, o lo que es lo mismo: $(1-\rho^2)\sigma_u^2 I_T$. Abandonar la primera observación no tendrá un gran impacto, a no ser que se trabaje con una muestra pequeña, mientras que evita tener que prestar atención a una observación de modo diferenciado al resto de ellas.

Conviene resaltar, por último, que con estas transformaciones no es exactamente cierto que los coeficientes del modelo queden inalterados. En efecto, puesto que todas las variables del modelo sufren una transformación del tipo [5.20], el término independiente del modelo transformado es igual a $(1-\rho)\beta_1$, donde β_1 es el término independiente del modelo original. En consecuencia, el parámetro β_1 del modelo original se recupera a partir de la estimación del término independiente del modelo transformado y de la estimación previa del parámetro ρ .

5.6. EL ESTIMADOR DE MÁXIMA VEROSIMILITUD

En esta sección vamos a obtener el estimador de máxima verosimilitud de un modelo lineal cuyo término de error se supone que tiene una distribución Normal, aunque con matriz de covarianzas no escalar. El principal resultado que veremos es que, al igual que probamos en la Sección 3.6 en un contexto más simplificado, este estimador de máxima verosimilitud coincide con el estimador de mínimos cuadrados generalizados.

En efecto, si suponemos que el vector u , formado por los términos de error del modelo $y_t = x_t'\beta + u_t$, sigue una distribución normal $u \sim N_T(0, \sigma_u^2 \Sigma)$, entonces:

⁽¹⁾ El símbolo «-1» es simplemente una cuestión de notación, pues, lógicamente, no existe ninguna matriz cuya inversa sea igual a V^{*-1} .

$$L(\mathbf{u}/\sigma_u^2, \Sigma) = \frac{1}{(2\pi)^T} \cdot \frac{1}{(\sigma_u^2)^T} \cdot \frac{1}{|\Sigma|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2\sigma_u^2} \mathbf{u}'\Sigma^{-1}\mathbf{u}\right)$$

y teniendo en cuenta que el jacobiano de la transformación es:

$$|\mathbf{J}| = \left| \frac{\partial \mathbf{u}_t}{\partial y_t} \right| = \begin{vmatrix} \partial u_1 / \partial y_1 & \cdots & \partial u_T / \partial y_1 \\ \cdots & \cdots & \cdots \\ \partial u_1 / \partial y_T & \cdots & \partial u_T / \partial y_T \end{vmatrix} = |\mathbf{I}_T| = 1$$

se tiene la función de verosimilitud:

$$L(\mathbf{y}, \mathbf{X}/\beta, \sigma_u^2, \Sigma) = \frac{1}{(2\pi)^T} \cdot \frac{1}{(\sigma_u^2)^T} \cdot \frac{1}{|\Sigma|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2\sigma_u^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta)\right)$$

y, si tomamos logaritmos:

$$\begin{aligned} \ln L(\mathbf{y}, \mathbf{X}/\beta, \sigma_u^2, \Sigma) &= \\ &= -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2} \ln |\Sigma| - \frac{1}{2\sigma_u^2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

y al tomar derivadas con respecto al vector β se obtiene el sistema de k ecuaciones:

$$-\left(\frac{1}{\hat{\sigma}_{\text{MV}}^2}\right) \mathbf{X}' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_{\text{MV}}) = \mathbf{0}$$

o, lo que es lo mismo:

$$\mathbf{X}' \Sigma^{-1} \mathbf{y} = (\mathbf{X}' \Sigma^{-1} \mathbf{X}) \hat{\beta}_{\text{MV}} \quad [5.23]$$

que es precisamente el sistema de ecuaciones normales del estimador de mínimos cuadrados generalizados, por lo que *el estimador de máxima verosimilitud del vector β y su estimador de mínimos cuadrados generalizados coinciden*. Tomando ahora derivadas con respecto a σ_u^2 se tiene:

$$-\frac{T}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \hat{\mathbf{u}}' \Sigma^{-1} \hat{\mathbf{u}} = \mathbf{0}, \quad \text{es decir,} \quad \hat{\sigma}_{\text{MV}}^2 = \frac{\hat{\mathbf{u}}'_{\text{MCG}} \Sigma^{-1} \hat{\mathbf{u}}_{\text{MCG}}}{T}$$

por lo que los estimadores MCG y MV de este parámetro difieren tan sólo en el denominador de la suma residual, al igual que ocurría en el caso en que la matriz de covarianzas del vector \mathbf{u} era escalar; puesto que el estimador MCG de σ_u^2 es insesgado, entonces su estimador de máxima verosimilitud será sesgado, pues ambos difieren en un factor $\frac{T}{T-k}$:

$$\hat{\sigma}_{\text{MV}}^2 = \frac{T-k}{T} \hat{\sigma}_{\text{MCG}}^2 \quad [5.24]$$

Es fácil ver que el sesgo del estimador $\hat{\sigma}_{\text{MV}}^2$ disminuye al aumentar el tamaño muestral, puesto que:

$$E(\hat{\sigma}_{\text{MV}}^2) = E\left(\frac{T-k}{T} \sigma_{\text{MCG}}^2\right) = \frac{T-k}{T} E(\hat{\sigma}_{\text{MCG}}^2) = \frac{T-k}{T} \sigma_u^2$$

y el cociente $\frac{T-k}{T}$ tiende a 1 cuando T tiende a infinito, por lo que $E(\hat{\sigma}_{\text{MV}}^2)$ tiende a σ_u^2 .

5.7. INFERENCIA ESTADÍSTICA CON MATRICES DE COVARIANZAS NO ESCALARES

Las cuestiones concernientes a la contrastación de hipótesis en el modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{u}) = \mathbf{0}, \quad \text{Var}(\mathbf{u}) = \sigma_u^2 \boldsymbol{\Sigma}, \quad \boldsymbol{\Sigma} \neq \mathbf{I}_T \quad [5.25]$$

pueden resolverse de nuevo mediante la consideración del modelo transformado:

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{u}^*, \quad E(\mathbf{u}^*) = \mathbf{0}, \quad \text{Var}(\mathbf{u}^*) = \sigma_u^2 \mathbf{I}_T \quad [5.26]$$

donde:

$$\mathbf{y}^* = \mathbf{V}^{-1} \mathbf{y}, \quad \mathbf{X}^* = \mathbf{V}^{-1} \mathbf{X}, \quad \mathbf{u}^* = \mathbf{V}^{-1} \mathbf{u}, \quad \mathbf{V}\mathbf{V}' = \boldsymbol{\Sigma}$$

Debido a la linealidad del modelo, los coeficientes $\boldsymbol{\beta}$ en los modelos [5.25] y [5.26] son los mismos y, por tanto, el contraste de un conjunto de q hipótesis lineales, $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, sobre los componentes del vector $\boldsymbol{\beta}$ puede hacerse indistintamente sobre cualquiera de ambos modelos. La ventaja del modelo [5.26] es que satisface las condiciones para las que desarrollamos el análisis de inferencia en el Capítulo 4, por lo que podemos utilizar todos los resultados allí obtenidos. Así, si suponemos que el término de error sigue una distribución Normal: $\mathbf{u} \sim N_T(\mathbf{0}_T; \sigma_u^2 \boldsymbol{\Sigma})$, entonces:

$$\mathbf{u}^* = \mathbf{V}^{-1} \mathbf{u} \sim N_T(\mathbf{0}; \sigma_u^2 \mathbf{V}^{-1} \boldsymbol{\Sigma} (\mathbf{V}^{-1})') = N_T(\mathbf{0}; \sigma_u^2 \mathbf{I}_T)$$

y, como consecuencia, utilizando los resultados del Capítulo 4, se concluye que el estadístico

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}}_{\text{MCG}} - \mathbf{r})[\mathbf{R}(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_{\text{MCG}} - \mathbf{r})/q}{\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*/(T-k)}$$

tiene una distribución $F_{q, T-k}$ si el vector $\hat{\beta}$ que aparece en este estadístico es el estimador de MCO del modelo [5.26] o, lo que es lo mismo, el estimador MCG del modelo [5.25]. En función de las variables originales, se tiene que

$$\frac{(\mathbf{R}\hat{\beta}_{\text{MCG}} - \mathbf{r})[\mathbf{R}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta}_{\text{MCG}} - \mathbf{r})/q}{\hat{\mathbf{u}}'_{\text{MCG}}\Sigma^{-1}\hat{\mathbf{u}}_{\text{MCG}}/(T-k)} \quad [5.27]$$

tiene una distribución $F_{q, T-k}$. Notemos que en el denominador de este estadístico aparece la suma residual que, insistimos una vez más, no es la suma de cuadrados de los residuos del modelo original, sino de los residuos del modelo transformado.

La dificultad práctica estriba, de nuevo, en que la matriz Σ es desconocida y ha de sustituirse por una estimación $\hat{\Sigma}$, lo que hace que el cociente [5.27] ya no tenga una distribución F , sino sólo aproximadamente. De nuevo el grado de aproximación depende de la bondad de la estimación de $\hat{\Sigma}$, que, a su vez, dependerá del tamaño muestral, pero, de nuevo, ésta es una cuestión que no analizamos en este texto y que puede consultarse en Theil (1971).

Cuando un conjunto de hipótesis lineales $H_0: \mathbf{R}\beta = \mathbf{r}$ no queda rechazado tras su contraste, entonces se debe estimar sujeto a estas restricciones. Si se plantea un problema de optimización análogo al de [4.7], es decir, encontrar el vector β que minimiza la «Suma ponderada de los cuadrados de los residuos» $\hat{\mathbf{u}}'\Sigma^{-1}\hat{\mathbf{u}}$ o, lo que es lo mismo, $\hat{\mathbf{u}}^*\hat{\mathbf{u}}^*$, sujeta a las restricciones $H_0: \mathbf{R}\beta = \mathbf{r}$, entonces el desarrollo del Capítulo 4, junto con razonamientos similares a los que venimos haciendo en este capítulo, conducen a la expresión:

$$\hat{\beta}_{\text{MCGR}} = \hat{\beta}_{\text{MCG}} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}\Sigma^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta}_{\text{MCG}} - \mathbf{r}) \quad [5.28]$$

donde $\hat{\beta}_{\text{MCG}}$ y $\hat{\beta}_{\text{MCGR}}$ son los estimadores de MCG sin restricciones, y sujeto a las restricciones, respectivamente. Otra forma de probar que éste es el estimador de mínimos cuadrados restringidos en esta situación es obteniendo el estimador restringido del vector β en el modelo transformado [5.7] y sustituyendo las variables \mathbf{X}^* por las variables originales \mathbf{X} . Para ello es importante recordar de nuevo que los coeficientes β en ambos modelos son los mismos, aunque las variables de ambos modelos sean diferentes.

Por las razones expuestas en el Capítulo 4, el estadístico [5.27] puede también escribirse:

$$\frac{\frac{\hat{\mathbf{u}}'_R \Sigma^{-1} \hat{\mathbf{u}}_R - \hat{\mathbf{u}}'_{\text{MCG}} \Sigma^{-1} \hat{\mathbf{u}}_{\text{MCG}}}{q}}{\frac{\hat{\mathbf{u}}'_{\text{MCG}} \Sigma^{-1} \hat{\mathbf{u}}_{\text{MCG}}}{T-k}} = \frac{\text{SRR} - \text{SRS}}{\text{SRS}} \quad [5.29]$$

donde $\hat{\mathbf{u}}_{\text{MCG}}$ y $\hat{\mathbf{u}}_R$ denotan los vectores de residuos obtenidos del modelo original, sustituyendo el vector β de parámetros desconocidos por su estimador MCG sin restringir y restringido, respectivamente.

La matriz de covarianzas Σ que aparece en [5.27] o [5.29] debe sustituirse por una estimación MCO previa, y ésta podría ser restringida o sin restringir. Sobre la base de análisis de simulación, es aceptado que la matriz $\hat{\Sigma}$ sea la calculada a partir de los residuos MCO en el modelo *con las restricciones*.

5.8. PREDICCIÓN EN UN MODELO CON MATRIZ DE COVARIANZAS GENERICA

Una matriz de covarianzas no escalar para el término de error del modelo econométrico puede afectar al modo en que se obtienen predicciones a partir de dicho modelo de dos formas diferentes: alterando el modo de generar las predicciones numéricas o alterando la forma en que se construyen sus intervalos de confianza.

Como primera aproximación al problema recordemos que las predicciones que discutimos en el Capítulo 4 se basaban en el supuesto $E_T u_{T+1} = 0$. Este supuesto es válido incluso si el término de error tiene heteroscedasticidad, pero deja de serlo cuando existe autocorrelación. En este último caso la muestra aporta información relevante acerca de la realización futura de la variable u_{T+1} . Esto se debe a que u_T está correlacionada con u_{T+1} , y tenemos una idea aproximada del valor de u_T , a través del residuo \hat{u}_T , que podemos calcular numéricamente. La consecuencia es que $E_T u_{T+1}$ será una función no nula de la muestra.

Por otra parte, la heteroscedasticidad hace que no tenga sentido hablar del parámetro σ_u^2 como una constante, sino como una función del tiempo, lo que habrá de ser tenido en cuenta al calcular la amplitud de los intervalos de confianza para la predicción. En resumen, la heteroscedasticidad parece afectar al modo en que se obtienen los intervalos de confianza de la predicción, aunque no a ésta, mientras que la autocorrelación tiene el efecto contrario.

5.8.a. Predicción bajo heteroscedasticidad

Las ecuaciones relevantes en la obtención de predicciones e intervalos de confianza a partir del modelo econométrico son:

$$E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta} + E_T u_{T+1}$$

$$e_T(1) = y_{T+1} - E_T y_{T+1} = \mathbf{x}'_{T+1} (\beta - \hat{\beta}) + u_{T+1}$$

$$\text{Var}(e_T(1)) = \text{Var} u_{T+1} + \mathbf{x}'_{T+1} \text{Var}(\hat{\beta}) \mathbf{x}_{T+1} + 2(\beta - \hat{\beta})' \text{Cov}(\mathbf{x}'_{T+1} (\beta - \hat{\beta}), u_{T+1})$$

donde el estimador $\hat{\beta}$ se ha obtenido con las T primeras observaciones, y donde se supone, entre otras cosas, que los valores futuros de las variables explicativas son conocidos de antemano. Si existe heteroscedasticidad en el modelo, pero no autocorrelación, entonces, tras estimar el modelo por MCG,

la predicción de y_{T+1} es $E_t y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta}$ y la varianza del error de predicción se convierte en:

$$\text{Var}(e_T(1)) = \text{Var } u_{T+1} + \sigma_u^2 [\mathbf{x}'_{T+1} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{x}_{T+1}] \quad [5.30]$$

Por ejemplo, consideremos el modelo $\tilde{y} = \beta \tilde{x}_t + u_t$, ambas variables medidas en diferencias con respecto a la media muestral. Supongamos que $\sigma_t^2 = \text{Var}(u_t) = \sigma_u^2 \tilde{x}_t^2$, con lo que la expresión anterior se transforma en:

$$\begin{aligned} \text{Var}(e_T(1)) &= \sigma_u^2 \tilde{x}_{T+1}^2 + \left[\tilde{x}_{T+1}^2 \left(\Sigma_1^T \frac{\tilde{x}_t^2}{\sigma_t^2} \right)^{-1} \right] = \\ &= \sigma_u^2 \tilde{x}_{T+1}^2 + \tilde{x}_{T+1}^2 \left(\Sigma_1^T \frac{1}{\sigma_u^2} \right)^{-1} = \sigma_u^2 \tilde{x}_{T+1}^2 \left(1 + \frac{1}{T} \right) \end{aligned}$$

5.8.b. Predicción bajo autocorrelación

Si el término de error del modelo econométrico presenta autocorrelación, entonces los vectores \mathbf{u} (términos de error durante el intervalo muestral) y \mathbf{u}_F (términos de error de los n períodos futuros) no son independientes, teniendo una matriz de covarianzas:

$$\text{Var} \begin{pmatrix} \mathbf{u} \\ \mathbf{u}_F \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}$$

donde la estructura de las matrices Σ_{ij} , $ij = 1, 2$ dependerá de la estructura de autocorrelación del término de error del modelo. Puede probarse (Goldberger, 1962) que la predicción óptima es, en estas condiciones:

$$E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta}_{\text{MCG}} + \Sigma'_{12} (\Sigma_{11})^{-1} \hat{\mathbf{u}} \quad [5.31]$$

donde $\hat{\mathbf{u}}$ denota el vector (de dimensión T) formado por los residuos MCG durante el intervalo muestral. Aun sin analizar la demostración de tal propiedad, es fácil ver su interpretación: bajo el supuesto de Normalidad:

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{u}_F \end{pmatrix} \sim \text{Normal} \left[\begin{pmatrix} \mathbf{0}_T \\ \mathbf{0}_n \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \right]$$

y la esperanza condicional de \mathbf{u}_F , dado el vector \mathbf{u} , es (véase Capítulo 2) $E(\mathbf{u}_F/\mathbf{u}) = \Sigma'_{12} (\Sigma_{11})^{-1} \mathbf{u}$, por lo que [5.31] puede interpretarse como:

$$E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta}_{\text{MCG}} + E(\mathbf{u}_F/\mathbf{u})$$

Es fácil ver que en el caso en que la estructura de autocorrelación es

$$u_t = \rho u_{t-1} + \varepsilon_t$$

y se predice para un solo período futuro, la expresión [5.31] se reduce a $E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_{\text{MCG}} + \rho \hat{u}_T$, resultado al que asimismo se llegaría utilizando la transformación MCG: $y_{T+1} - \rho y_T = (\mathbf{x}_{T+1} - \mathbf{x}_T)' \boldsymbol{\beta} + \varepsilon_{T+1}$ y tomando esperanzas condicionales en el período T .

Para n períodos hacia el futuro, tendríamos:

$$E_T y_{T+n} = \mathbf{x}'_{T+n} \hat{\boldsymbol{\beta}}_{\text{MCG}} + \rho^n \hat{u}_T$$

5.9. EJERCICIOS PRACTICOS

Ejercicio 5.1. Con objeto de analizar los efectos de la heteroscedasticidad en el modelo lineal, se estudió el modelo econométrico:

$$y_t = 50,0 + x_{2t} - 3,0x_{3t} + u_t \quad [5.32]$$

donde u_t se distribuye $N(0, \sigma_t^2)$, $\sigma_t^2 = 25x_{3t}$.

Generación de datos. La muestra utilizada consta de 120 observaciones. Los 120 valores de las variables explicativas x_{2t} y x_{3t} fueron generados del siguiente modo:

$$\begin{aligned} x_{21} &= 100,0 \\ x_{2t} &= 25,0 + 0,75x_{2t-1} + \varepsilon_{1t} & t = 2, 3, \dots, 120 \\ x_{3t} &= 25,0 + 2t + \varepsilon_{2t} & t = 1, 2, \dots, 120 \end{aligned} \quad [5.33]$$

donde tanto ε_{1t} como ε_{2t} tienen distribución Normal $(0; 25)$, pero son independientes entre sí, así como independientes de u_t . El ejercicio de estimación comienza generando series aleatorias de 120 observaciones para ε_{1t} y ε_{2t} , a partir de las distribuciones citadas. A continuación se utilizan las ecuaciones [5.33] para generar las series temporales de x_{2t} y x_{3t} .

La siguiente etapa consiste en obtener una realización del término de error u_t , teniendo en cuenta que su distribución es diferente en cada período, puesto que su varianza evoluciona de acuerdo con $\sigma_t^2 = 25x_{3t}$. En consecuencia, se trata de considerar en cada período t , $1 \leq t \leq 120$, una distribución $N(0, 25x_{3t})$, donde x_{3t} es el valor que en ese período toma la variable x_3 en la serie temporal que se generó para dicha variable. De esta distribución Normal se toma una muestra de tamaño 1, que es el valor de la variable aleatoria u_t para el período en cuestión.

Estimación del modelo. Por último, se utiliza el modelo [5.32], junto con la realización de u_t , para obtener la serie de datos correspondiente a la variable endógena y_t . Procediendo como si se desconociesen los valores numéricos de los coeficientes de [5.32], se estima el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

con las series generadas para y_t , x_{2t} , x_{3t} .

Este procedimiento se llevó a cabo cien veces, manteniendo en cada caso las realizaciones de las variables explicativas, pero cambiando la realización del término de error y, consiguientemente, la realización de la variable endógena y_t . De este modo se obtuvieron 100 estimaciones para cada uno de los tres parámetros del modelo [5.32]. Se calcularon tanto las estimaciones MCO como las estimaciones MCG, utilizando el hecho de que la estructura de σ_t^2 es conocida. Es importante hacer notar que en la práctica no sería muy probable que σ_t^2 tuviese un comportamiento tan claramente definido como función de una variable explicativa, y menos aún que tuviésemos un conocimiento del mismo. La variable x_3 , es, en su parte determinista, una tendencia que crece desde 27,0 hasta 265,0, por lo que la desviación típica de u_t crece a lo largo de la muestra, aproximadamente, desde 25,0 hasta 81,5.

Significación estadística de las estimaciones. Las 100 realizaciones de cada uno de los parámetros estimados generan una distribución empírica, que puede representarse por los histogramas de frecuencias que aparecen en las Figuras 5.1, 5.2 y 5.3. Utilizando los estadísticos t de Student para cada uno de los tres coeficientes del modelo, la estimación del parámetro β_3 resultó siempre significativa, tanto por MCO como por MCG. La estimación MCO de β_2 fue significativa en 8 de las 110 estimaciones, mientras que la estimación MCG resultó significativa un 26 por 100 de las veces. La estimación MCO de β_1 fue significativa en 8 de las 100 estimaciones, mientras que su estimación MCG era significativa en 14 ocasiones.

No es casual que las estimaciones MCG resulten significativas más a menudo que las estimaciones MCO, puesto que son más precisas, es decir, tienen una menor varianza, por lo que sus intervalos de confianza, siendo más estrechos, tienen menor probabilidad de incluir al origen. Por otra parte, el que β_3 resulte significativo mucho más a menudo que los otros coeficientes del modelo se debe, en parte, a ser el de mayor valor absoluto. Sin embargo, este hecho también está relacionado con la estructura de las tres variables explicativas, aunque ésta es una cuestión que sobrepasa el ámbito de nuestra discusión.

Precisión de las estimaciones. Los intervalos de 95 por 100 de confianza para el contraste de la hipótesis nula de que el valor de cada uno de los tres coeficientes es su verdadero valor, es decir, 50, 1 y -3 , son respectivamente iguales a: $(-142, 242)$, $(-0,806, 2,806)$ y $(-2,825, -3,175)$. Con estos intervalos se rechazó cada una de las hipótesis nulas en cuatro, cinco y seis ocasiones respectivamente, muy en consonancia con los resultados que teóricamente cabría esperar. En efecto, ya que la distribución de probabilidad del término de error es Normal (nosotros la hemos elegido así), un contraste al 95 por 100 de confianza que incluyera en la hipótesis nula el verdadero valor de uno de los parámetros (que en este caso conocemos, pues los hemos elegido) debería aceptar dicha hipótesis, en promedio, un 95 por 100 de las veces.

Los intervalos anteriores muestran asimismo que, mientras β_3 es estimado con mucha precisión, β_2 se estima con cierta precisión, pero, sin embargo, la

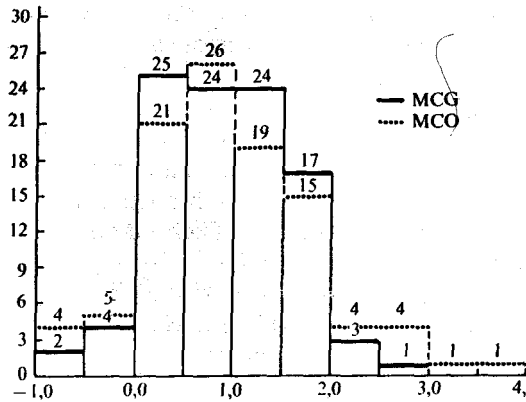


FIGURA 5.1. Distributions empíricas de β_2 .

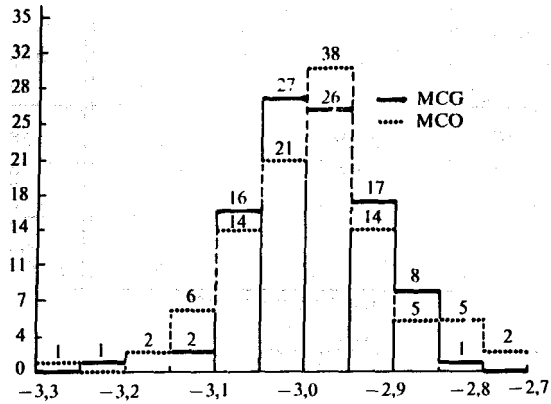


FIGURA 5.2. Distributions empíricas de β_3 .

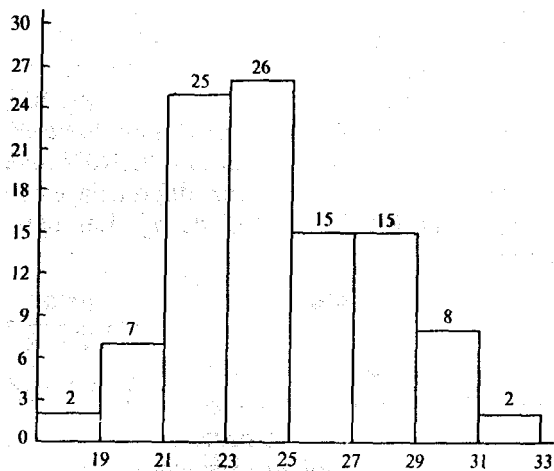


FIGURA 5.3. Distribución empírica de σ_u^2 .

precisión en la estimación del término independiente es muy pequeña. Tiene interés hacer algunas consideraciones acerca de la estimación de β_2 . Hemos visto en el párrafo anterior que si estuviésemos interesados en contrastar la hipótesis nula de que dicho parámetro toma el valor 1, en la mayoría de los casos dicha hipótesis no sería rechazada. Sin embargo, el rango de valores en el intervalo del 95 por 100 es suficientemente amplio para que, incluso si el valor de la elasticidad en la hipótesis nula fuese muy distinta de 1, no rechazásemos la hipótesis nula. Dicho de otra forma, la potencia del contraste no sería muy alta.

Distribución empírica de las estimaciones. Las medias y desviaciones típicas obtenidas a partir del conjunto de 100 estimaciones de cada uno de los parámetros del modelo fueron:

	Estimaciones MCO		Estimaciones MCG	
	Media	Desviación típica	Media	Desviación típica
$\hat{\beta}_1$	52,9	(91,4)	58,15	(70,2)
$\hat{\beta}_2$	0,96	(0,86)	0,92	(0,67)
$\hat{\beta}_3$	-2,99	(0,085)	-2,99	(0,071)
$\hat{\sigma}_u^2$	1319,6	(243,6)	24,6	(3,2)

Estos resultados sugieren varias observaciones:

a) A pesar de la imprecisión con que se han estimado β_1 y β_2 , sin embargo los promedios de las 100 estimaciones están muy próximos a los verdaderos valores de los parámetros. Este es precisamente el significado de que tanto el estimador MCO como el estimador MCG sean insesgados. Como dijimos en el Capítulo 3, el que un estimador sea insesgado no garantiza que su valor, obtenido a partir de una única muestra, esté cercano al verdadero valor del parámetro que estima. Pero sí que asegura que si, al contrario de lo que ocurre en Economía, se dispusiera de varias muestras, entonces el valor medio de las estimaciones de un parámetro convergería a su verdadero valor.

b) Puede observarse la gran diferencia existente entre los promedios de las estimaciones MCO y MCG de σ_u^2 . Las expresiones para ambos estimadores son:

$$\hat{\sigma}_{\text{MCO}}^2 = \frac{\hat{u}'\hat{u}}{T-k}; \quad \hat{\sigma}_{\text{MCG}}^2 = \frac{\hat{u}'^*\hat{u}^*}{T-k} = \frac{\hat{u}'_{\text{MCG}}\Sigma^{-1}\hat{u}_{\text{MCG}}}{T-k}$$

donde \hat{u} y \hat{u}_{MCG} denotan los residuos obtenidos con los estimadores correspondientes. Sin embargo, la principal diferencia entre ambas es que la suma de cuadrados de residuos *del modelo original* está ponderada en el estimador MCG. Con heteroscedasticidad, Σ es una matriz diagonal, cuyos elementos

son los inversos de la sucesión σ_t^2 . En consecuencia, ponderar por Σ^{-1} equivale a dividir cada término u_t^2 de la suma residual por el correspondiente σ_t^2 . En este modelo, descontando el factor 25,0, la varianza σ_t^2 oscila, aproximadamente, desde 27,0 hasta 265,0 con un promedio de 146,0. La relación entre las estimaciones MCO y MCG de σ_u^2 es aproximadamente igual a este valor.

Matrices de covarianzas de las estimaciones. Como las series x_{2t} y x_{3t} se mantienen constantes para las 100 estimaciones, las matrices de covarianzas de los estimadores de β_2 y β_3 que sólo dependen de las variables explicativas y de la estructura de σ_t^2 (que a su vez sólo depende de x_{3t}) son las mismas en todas las simulaciones. La siguiente tabla presenta las estimaciones de las desviaciones típicas de los coeficientes (las raíces cuadradas de los elementos en la diagonal de ambas matrices de covarianzas).

	Desviaciones típicas MCO	Desviaciones típicas MCG
$\hat{\beta}_1$	96,00	72,90
$\hat{\beta}_2$	0,90	0,70
$\hat{\beta}_3$	0,085	0,075

Como se explicó en el Capítulo 3, éstas son las desviaciones típicas con que se obtendrían las estimaciones de cada uno de los parámetros del modelo si se dispusiese de diferentes muestras. Aunque, como allí se dijo, esto nunca ocurre en Economía, en cambio sí que ocurre en este ejercicio «experimental», en el que hemos contado con 100 muestras.

Así, puede verse que las desviaciones típicas que acabamos de presentar son aproximadamente iguales a las desviaciones típicas de la distribución empírica a partir de las 100 estimaciones, que se presentaron en la tabla anterior.

Análogos ejercicios podrían plantearse mediante simulaciones de modelos con autocorrelación, para lo cual basta disponer de un microordenador y alguno de los muchos paquetes estadísticos existentes. El lector experimentado, así como el profesor de la asignatura, deben insistir en este tipo de discusiones.

PROBLEMAS

Problema 5.1. Probar que $\hat{\sigma}_{MCG}^2 = \frac{\mathbf{y}'\Sigma^{-1}\mathbf{y} - \hat{\beta}'\mathbf{X}'\Sigma^{-1}\mathbf{y}}{T - k}$.

Problema 5.2. Probar que [5.13] y [5.14] son expresiones equivalentes, incluso en el caso general, con k variables explicativas.

Problema 5.3. Probar, mediante un argumento análogo al del Capítulo 4, que la solución al problema de optimización

$$\text{mín } \mathbf{u}'\Sigma^{-1}\mathbf{u}$$

sujeto a las restricciones

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

es precisamente el estimador MCG restringido de la expresión [5.28].

Problema 5.4. En el modelo $y_t = \alpha + \beta x_t + u_t$, con $E(u_t) = 0$, $E(u_t^2) = \sigma_t^2$, $E(u_t u_s) = 0$, obtener la expresión analítica de los estimadores $\hat{\alpha}_{\text{MCG}}$ y $\hat{\beta}_{\text{MCG}}$, y particularizarlas a los casos:

- a) $\sigma_t^2 = \sigma^2$ para todo t , b) $\sigma_t^2 = kx_t$, k dado.

Problema 5.5. Considere el modelo $y_t = \alpha + \beta x_t + u_t$, $E(u_t) = 0$, $\text{Var}(u_t) = \sigma_t^2 = \beta^2 x_t^2$, del que se tienen las siguientes observaciones:

t	y_t	x_t
1	4	2
2	2	1
3	1	4
4	5	2
5	3	1

- a) Obtenga el estimador MCO de α y β , así como su matriz de covarianzas. ¿Existe algún problema en la definición de dicho estimador en este modelo?
- b) Obtenga el estimador MCG y su matriz de covarianzas. ¿Existe algún problema en su definición? Compare su matriz de covarianzas con la del estimador MCO.
- c) Obtenga las dos ecuaciones que proporcionan el estimador de máxima verosimilitud bajo el supuesto $u_t \sim N(0, \sigma_t^2)$, así como la expresión analítica de su matriz de covarianzas.
- d) Calcule el valor numérico del estimador de máxima verosimilitud. ¿Está unívocamente definido en este modelo? ¿Por qué?
- e) Obtenga los valores numéricos de la matriz de covarianzas del estimador de máxima verosimilitud y compare dicha matriz con la matriz de covarianzas del estimador MCG.

Problema 5.6. Probar que [5.27] y [5.29] son expresiones equivalentes.

Problema 5.7. Demuestre que si los valores futuros de las variables explicativas \mathbf{x}_{T+1} son conocidos en un modelo de regresión con heteroscedasticidad, pero no autocorrelación, entonces la varianza del error de predicción de y_{T+1} viene dada por [5.30].

Problema 5.8. Para estimar el modelo de regresión lineal:

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i; \quad i = 1, 2, \dots, N; \quad u_i \sim N(0, \sigma_u^2)$$

el investigador podría ordenar la muestra según los valores muestrales de una de las variables explicativas, clasificarlos en intervalos, calcular las medias muestrales de x e y para cada intervalo y estimar un modelo con dichos promedios.

a) Explique cuál sería el estimador *eficiente* en dicho modelo y demuestre que es insesgado.

b) Demuestre que la razón para no recomendar este estimador, que es sencillo por utilizar un número reducido de observaciones, es su ineficiencia con respecto al estimador MCO. Para ello, demuestre que la diferencia entre las matrices de covarianzas de ambos estimadores tiene un signo bien definido.

Problema 5.9. Demostrar que la matriz de covarianzas del estimador MV del modelo

$$y_t = x_t \beta + u_t \\ E u_t = 0, \quad \text{Var } u_t = \sigma^2 z_t^\alpha, \quad \alpha > 0, \quad u_t \sim \text{Normal}$$

es

$$\Sigma = \begin{pmatrix} \frac{1}{\sigma^2} \sum_1^T \frac{x_t x_t'}{z_t^\alpha} & 0 & 0 \\ 0 & \frac{T}{2\sigma^4} & \frac{1}{2\sigma^2} \sum_1^T \ln z_t \\ 0 & \frac{1}{2\sigma^2} \sum_1^T \ln z_t & \frac{1}{2} \sum_1^T (\ln z_t)^2 \end{pmatrix}$$

Problema 5.10. Un empresario intenta cuantificar la relación existente entre el volumen de ventas (y) y los gastos de publicidad (x). A priori se sabe que la relación entre estas dos variables es del tipo:

$$y_t = \beta_1 + \beta_2 x_t + u_t, \quad E u_t = 0, \quad E(u_t \cdot u_{t'}) = 0 \quad \text{si } t \neq t', \quad E(u_t^2) = \sigma_u^2 x_t$$

y dispone de los datos muestrales:

Volumen de ventas	Gastos
120	4
130	25
150	30
160	36
180	40
200	49

a) Obtener las estimaciones MCO y MCG de β_1 y β_2 y sus respectivas matrices de varianzas y covarianzas.

b) Con los estimadores más eficientes contrastar $H_0: \beta_2 = 0$ frente a $H_1: \beta_2 \neq 0$.

c) Obtener las predicciones MCO y MCG y sus respectivas varianzas para un valor de x igual a 60. ¿Cuál de las dos predicciones es más eficiente?

Problema 5.11. Dado el modelo de regresión:

$$y_t = \beta_1 x_t + \beta_2 z_t + u_t \quad t = 1, \dots, n$$

tal que:

$$\text{Var}(u_t) = \sigma^2 x_t^2$$

hallar la expresión del test F para contrastar la hipótesis $H_0: \mathbf{A}\boldsymbol{\beta} = \mathbf{c}$.

CAPITULO 6

HETEROSCEDASTICIDAD

6.1. INTRODUCCION

En el Capítulo 5 hemos considerado distintas situaciones en que la matriz de covarianzas del término de error del modelo econométrico dejaba de tener una estructura escalar. Un caso en que esto ocurre es cuando, conservando su estructural diagonal, los elementos de la diagonal no son iguales entre sí. Decimos entonces que el término de error tiene *heteroscedasticidad*, siendo su varianza diferente para las distintas observaciones que integran la muestra.

Estudiamos entonces el modo de estimar en tal situación, en que el estimador lineal eficiente, mínimos cuadrados generalizados (MCG), adopta el nombre especial de mínimos cuadrados ponderados, por cuanto consiste en conceder ponderaciones diferentes a las distintas observaciones muestrales al evaluar la suma residual que se pretende minimizar. Vamos a ver en este capítulo que esta idea de ponderar las observaciones muestrales al utilizar el método de mínimos cuadrados es más general de lo que pudo parecer en una primera instancia.

6.2. POSIBLES CAUSAS DE HETEROSCEDASTICIDAD

6.2.a. La presencia de heteroscedasticidad en un modelo de consumo estimado con datos de ingresos y gasto en consumo procedentes de N familias:

$$\text{Gasto en consumo}_i = \beta_1 + \beta_2 \text{ Ingresos}_i + u_i$$

puede surgir porque, una vez satisfechas sus necesidades primordiales, una familia de mayores ingresos dispone de un mayor excedente de renta, del que debe decidir qué proporción ahorrar y cuál consumir. Distintas familias, dentro de un mismo nivel de ingresos, tomarán distintas decisiones respecto a su exceso de renta, por lo que cabe esperar que las cifras de gasto en bienes

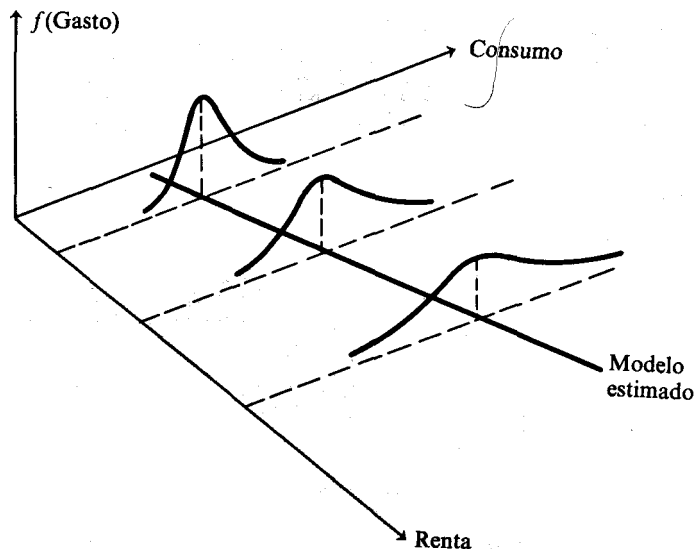


FIGURA 6.1.

de consumo tengan una mayor varianza en las familias de mayores ingresos que en las de ingresos inferiores. La Figura 6.1 ilustra esta situación, donde la dispersión de las cifras de gastos de consumo para un nivel de ingresos dado es superior cuanto mayores son los ingresos.

Algo similar ocurre con las empresas que deben decidir qué porcentaje de sus beneficios repartir como dividendos. De nuevo, las empresas con mayores beneficios tienen un margen de discreción muy superior al fijar su política de dividendos, por lo que si se pretende estimar el modelo:

$$\text{Dividendos}_i = \beta_1 + \beta_2 \text{Beneficios}_i + u_i$$

cabría esperar que la varianza de u_i , σ_i^2 dependa del nivel de beneficios de la empresa i -ésima.

6.2.b. Uno de los ejercicios propuestos al final del capítulo trata con el caso en que los datos de que dispone el investigador se han obtenido agregando los datos inicialmente recogidos a los distintos grupos en que se han clasificado los individuos encuestados. Algo similar ocurre si los datos son promedios de los proporcionados por tales grupos. Por ejemplo, si se dispone de datos de consumo y renta per cápita promedios de diversas regiones, sus varianzas serían inversamente proporcionales al número de individuos que se han promediado en cada colectivo.

6.2.c. Si se omite del modelo una variable relevante, se produce, aparte de los sesgos ya analizados en la Sección 3.11, una situación de heteroscedastici-

dad, por una razón similar a la del caso anterior. Ahora, en vez de estimar el modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

estaríamos estimando el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + v_t$$

donde $v_t = u_t + \beta_3 x_{3t}$, con lo que el residuo en cada período tendrá una magnitud creciente con la de x_{3t} y, en términos de momentos muestrales, el investigador podría llegar a creer que:

$$\text{Var}(v_t) = \sigma_u^2 + \beta_3^2 x_{3t}^2$$

cambiante en el tiempo de acuerdo con los valores de la variable x_{3t} . Debido a esta relación es que, como insistiremos en este capítulo, cuando se detecta la dependencia de σ_t^2 de una variable no incluida del modelo, debe considerarse su inclusión en una reformulación del mismo.

Cuando un modelo presenta una situación de heteroscedasticidad, hay varias cuestiones de importancia:

1. ¿Cuáles son las consecuencias de la heteroscedasticidad sobre el estimador de mínimos cuadrados ordinarios y su matriz de covarianzas? ¿Siguen siendo el estimador de mínimos cuadrados ordinarios óptimo entre los que son lineales e insesgados?
2. ¿Cómo puede detectarse la presencia de heteroscedasticidad?
3. ¿Cómo debe estimarse un modelo que presenta heteroscedasticidad?
4. ¿Cuál es la forma correcta de hacer contrastes de hipótesis lineales en un modelo con heteroscedasticidad?
5. ¿Cómo se elaboran las predicciones del modelo econométrico en tal situación?

Los puntos 1, 3 y 5 han sido ya discutidos en el Capítulo 5 en un caso sencillo. Como recordatorio, baste decir que en estas condiciones, el estimador de mínimos cuadrados ordinarios (MCO) continúa siendo lineal e insesgado, aunque ya no es el estimador de mínima varianza en dicha clase. Esa propiedad corresponde al estimador de mínimos cuadrados generalizados (MCG) que, supuesto que la matriz Σ fuese conocida, se define como la solución al sistema de ecuaciones normales:

$$(\mathbf{X}'\Sigma^{-1}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\Sigma^{-1}\mathbf{y} \quad [6.1]$$

6.3. ESTIMACION MINIMO-CUADRATICA EN PRESENCIA EN HETEROSCEDASTICIDAD

Como se dijo en el Capítulo 4, un problema fundamental que surge con la heteroscedasticidad es que, si se permite que la varianza σ_i^2 del término de error sea diferente en cada período, entonces el número de parámetros a estimar crecería con el número de observaciones ya que con cada observación aparece un nuevo parámetro, σ_i^2 . Es preciso, por tanto, establecer algún supuesto acerca del modo en que σ_i^2 varía a lo largo del tiempo (o a través de las distintas familias, o individuos, o países que integran la muestra).

Esta es una restricción importante porque, tanto la detección de la heteroscedasticidad como la estimación del modelo en presencia de ésta, se ven *condicionados* por el supuesto específico que se haya establecido acerca del modo en que la varianza del término de error varía entre observaciones muestrales. Así, los contrastes que veremos en la próxima sección están diseñados para detectar la presencia de heteroscedasticidad, supuesto que ésta adopte una determinada estructura; si no se detecta heteroscedasticidad, puede ser que no exista tal problema, pero también podría ocurrir que fuese de tipo diferente a aquel en que se basa el contraste. De modo análogo, el procedimiento de estimación MCG es eficiente siempre y cuando la heteroscedasticidad sea de la forma que se ha supuesto al diseñar tal estimador, pero no podemos garantizar sus propiedades y, en particular, su eficiencia, si la heteroscedasticidad es de otro tipo.

Además, debe pensarse que cualquier supuesto que se adopte acerca de la estructura de heteroscedasticidad del modelo no se efectúa sino como una aproximación a la «verdadera», pero, por supuesto, desconocida, estructura. Entonces ¿tiene sentido utilizar el estimador MCO incluso si se sospecha la existencia de heteroscedasticidad? La respuesta puede ser afirmativa si no se confía en poder hacer una especificación suficientemente aproximada de la estructura que adopta la sucesión $\{\sigma_i^2\}$. Un error en dicha especificación hará que el estimador MCG no sea totalmente eficiente, y su ganancia con respecto al estimador MCO quede en entredicho. Por otra parte, una especificación incorrecta conduce a una estimación sesgada de la matriz de covarianzas Σ .

Por consiguiente, es siempre una buena idea utilizar el estimador MCO como referencia con el que comparar el estimador MCG que posteriormente pueda obtenerse bajo una determinada estructura de heteroscedasticidad. En todo caso, hay que recordar que la utilización del estimador MCO no evita tener que «aproximar» Σ , puesto que forma parte de la matriz de covarianzas de dicho estimador.

Una nota de atención adicional se refiere a la estimación del parámetro σ_u^2 que acompaña a la matriz Σ en presencia de heteroscedasticidad. Para ello, *no puede utilizarse la suma residual de la estimación MCO*. La razón es que, como vimos en el ejercicio práctico al final del Capítulo 4, dicha suma residual sufre un problema de escala con respecto a la suma residual que se obtendría de los residuos MCG, que es la que se debe utilizar.

White (1980) ha propuesto una aproximación a la matriz de covarianzas del estimador MCO que no precisa de una representación específica de la

forma funcional que adopta la heteroscedasticidad, por lo que no genera los sesgos que pudieran generarse por adoptar una representación incorrecta. De hecho, se ha comprobado que la sugerencia de White:

$$\text{Var}(\hat{\beta}_{\text{MCO}}) = T\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} \left(\frac{1}{T} \mathbf{X}'\Sigma\mathbf{X} \right) (\mathbf{X}'\mathbf{X})^{-1}$$

donde $\sigma_u^2 \frac{1}{T} (\mathbf{X}'\Sigma\mathbf{X})$ se estima por:

$$\widehat{\sigma_u^2 \frac{1}{T} (\mathbf{X}'\Sigma\mathbf{X})} = \frac{1}{T} \sum_i^T \hat{u}_i^2(\mathbf{x}_i, \mathbf{x}_i')$$

tiene buenas propiedades, por lo que es siempre recomendable su uso. Su comparación con la expresión particular $\text{Var}(\hat{\beta}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ puede dar una idea del grado de heteroscedasticidad presente en el modelo.

El estimador MCG tiene matriz de covarianzas $\sigma_u^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$. Por otra parte, la matriz de covarianzas del estimador MCO es $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Siendo el estimador MCG lineal, insesgado y de mínima varianza, entonces la diferencia entre su matriz de covarianzas y la del estimador MCO es siempre una matriz semidefinida negativa⁽¹⁾. En particular, esto implica que los elementos de la diagonal de la primera (las varianzas de las estimaciones MCG de los coeficientes del modelo) son menores que los elementos correspondientes de la segunda matriz (las varianzas de las estimaciones MCO).

Otra propiedad de interés es que si el término de error del modelo econométrico sigue una distribución Normal, entonces el estimador MCG coincide con el estimador de máxima verosimilitud, con lo que pasa a ser el estimador insesgado óptimo (es decir, el estimador insesgado de mínima varianza).

El procedimiento de estimación por MCG bajo heteroscedasticidad es siempre el mismo, y sigue las líneas apuntadas en el Capítulo 5:

1. Se estima el modelo por MCO, ignorando la heteroscedasticidad del término de error.
2. Se establece un supuesto acerca de la estructura de la sucesión σ_i^2 .
3. Se utilizan los residuos MCO para estimar la forma funcional supuesta para σ_i^2 en el apartado anterior.
4. Se divide cada observación por la estimación $\hat{\sigma}_i$ (raíz cuadrada de $\hat{\sigma}_i^2$).
5. Se vuelve a estimar el modelo original con las variables transformadas en 4.

De estas normas puede verse que la idea de ponderar las observaciones es de aplicación generalizada al estimar por MCG un modelo con heteroscedasti-

⁽¹⁾ Por otra parte, si se calcula la matriz $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, entonces los elementos de su diagonal pueden ser mayores o menores que los elementos correspondientes de $\sigma_u^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$. Esto es, sin embargo, totalmente irrelevante, puesto que en presencia de heteroscedasticidad, la primera de estas dos matrices no es la matriz de covarianzas de ningún estimador lineal.

cidad. Dicho procedimiento equivale a dar a cada observación una importancia inversamente relacionada con la varianza del término de error en ese período. Este es un comportamiento bastante natural. Si en un cierto período el error tiene una varianza muy grande, entonces la observación recogida para la variable endógena y_t en ese período estará sujeta a un componente puramente aleatorio muy importante.

En consecuencia, dicha observación no es muy informativa en cuanto a caracterizar la relación entre y_t y las variables explicativas del modelo. A diferencia del estimador de MCG, el estimador MCO pondera a todas las observaciones de igual modo. Si las varianzas de los términos de error de cada período son diferentes, entonces asignar la misma ponderación a cada observación no puede ser un procedimiento eficiente, como ya vimos de modo formal en el Capítulo 5.

Ejemplo 6.1. Consideremos el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$ y supongamos que la varianza del término de error evoluciona en el tiempo de acuerdo con el esquema $\sigma_t^2 = \sigma_u^2 x_{3t}$, es decir, que en cada período la varianza de dicho término es proporcional al valor que toma la variable x_3 . En esta situación, particularizando [6.1] se tiene que el estimador de MCG viene dado por la solución al sistema de ecuaciones normales:

$$\begin{pmatrix} \Sigma_1^T(1/x_{3t}) & \Sigma_1^T(x_{2t}/x_{3t}) & T \\ & \Sigma_1^T(x_{2t}^2/x_{3t}) & \Sigma_1^T x_{2t} \\ & & \Sigma_1^T x_{3t} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \Sigma_1^T(y_t/x_{3t}) \\ \Sigma_1^T(x_{2t}y_t/x_{3t}) \\ \Sigma_1^T y_t \end{pmatrix} \quad [6.2]$$

que equivale a estimar el modelo con la variable endógena y variables explicativas transformadas: $\frac{y_t}{\sqrt{x_{3t}}}$, $\frac{1}{\sqrt{x_{3t}}}$, $\frac{x_{2t}}{\sqrt{x_{3t}}}$, $\sqrt{x_{3t}}$, respectivamente. De nuevo aparece aquí la idea de *ponderar* las observaciones muestrales.

La matriz de covarianzas del estimador resultante se estima multiplicando una estimación insesgada del parámetro σ_u^2 por la inversa de la matriz de la izquierda en la igualdad anterior. Una estimación insesgada de σ_u^2 viene dada por $\frac{\mathbf{y}'\Sigma^{-1}\mathbf{y} - \hat{\beta}'_{\text{MCG}}\mathbf{X}'\Sigma^{-1}\mathbf{y}}{T-k}$. En este ejemplo $\mathbf{y}'\Sigma^{-1}\mathbf{y}$ viene dado por $\frac{\Sigma_1^T y_t^2}{x_{3t}}$, mientras que $\mathbf{X}'\Sigma^{-1}\mathbf{y}$ es precisamente el vector a la derecha de [6.2]. Como puede verse, no es preciso conocer el parámetro σ_u^2 para estimar el vector β .

La matriz de covarianzas del estimador MCO es:

$$\sigma_u^2 \begin{pmatrix} T & \Sigma x_{2t} & \Sigma x_{3t} \\ \Sigma x_{2t}^2 & \Sigma(x_{2t}x_{3t}) & \Sigma x_{3t}^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma x_{3t} & \Sigma x_{2t}x_{3t} & \Sigma x_{3t}^2 \\ \Sigma(x_{2t}^2x_{3t}) & \Sigma(x_{2t}x_{3t}^2) & \Sigma x_{3t}^3 \end{pmatrix} \begin{pmatrix} T & \Sigma x_{2t} & \Sigma x_{3t} \\ \Sigma x_{2t}^2 & \Sigma(x_{2t}x_{3t}) & \Sigma x_{3t}^2 \end{pmatrix}^{-1}$$

donde el parámetro σ_u^2 debería estimarse mediante la misma expresión anterior.

La ponderación de las observaciones introduce una dificultad, y es que la correlación entre variables transformadas como $\frac{y}{\sqrt{x_{3t}}}$ y $\frac{x_{2t}}{\sqrt{x_{3t}}}$ va a ser, en general, superior a la correlación entre las variables originales y_t y x_{2t} , debido a la presencia común del divisor $\sqrt{x_{3t}}$. Este problema es conocido en Econometría como *correlación espúrea*, que, evidentemente, conviene distinguir de la correlación existente entre las variables originales, que es, en definitiva, la que se pretendía estimar.

Ejemplo 6.2. Supongamos que en el modelo $y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$ se tiene $E(u_t) = 0$ para todo t , pero $\text{Var}(u_t) = \sigma_t^2 = \sigma_u^2 (E y_t)^2$, es decir, que la varianza del término de error crece con el cuadrado de la esperanza de la variable endógena. Dicho valor esperado es igual a $\mathbf{x}'_t \boldsymbol{\beta}$, por lo que varía a través del tiempo, según van cambiando los valores de las variables exógenas.

La ponderación que hay que introducir en este caso para hacer que el término de error resultante tuviese matriz de covarianzas escalar es igual a $\mathbf{x}'_t \boldsymbol{\beta}$. Estas ponderaciones son desconocidas, puesto que dependen del vector $\boldsymbol{\beta}$; por tanto, habría que estimar el modelo inicialmente por MCO y dividir los valores observados de las variables endógena y exógena por $\mathbf{x}'_t \hat{\boldsymbol{\beta}}$ en cada período. En consecuencia, el estimador MCG resulta ser:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = \left(\sum_1^T \frac{\mathbf{x}_t \mathbf{x}'_t}{(\mathbf{x}'_t \hat{\boldsymbol{\beta}})^2} \right)^{-1} \left(\sum_1^T \frac{\mathbf{x}_t y_t}{(\mathbf{x}'_t \hat{\boldsymbol{\beta}})^2} \right)$$

con matriz de covarianzas:

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{MCG}}) = \sigma_u^2 \left(\sum_1^T \frac{\mathbf{x}_t \mathbf{x}'_t}{(\mathbf{x}'_t \hat{\boldsymbol{\beta}})^2} \right)^{-1}$$

y el parámetro σ_u^2 se estimaría por:

$$\hat{\sigma}_u^2 = \left(\frac{1}{T - k} \right) \sum_1^T \left[\frac{(y_t - \mathbf{x}'_t \hat{\boldsymbol{\beta}}_{\text{MCO}})^2}{(\mathbf{x}'_t \hat{\boldsymbol{\beta}})^2} \right]$$

6.4. CONTRASTES DE HETEROSCEDASTICIDAD

Vamos a pasar revista en esta sección a los principales contrastes que han sido propuestos para detectar la presencia de heteroscedasticidad en un modelo econométrico. Todos ellos contrastan la hipótesis nula de *ausencia de heteroscedasticidad*. Algunos de ellos sugieren la forma funcional de la heteroscedasticidad cuando se rechaza la hipótesis nula, por lo que la transformación de variables necesaria para estimar por MCG es inmediata. Otros, en cambio, no proporcionan tal información.

6.4.a. El contraste de Goldfeld y Quandt

Este contraste, propuesto en Goldfeld y Quandt (1965), parte del supuesto de que la magnitud de σ_t^2 depende de una variable z_t . Esta variable es, generalmente, una de las variables explicativas, aunque no es preciso que lo sea para llevar a cabo el contraste de hipótesis. En todo caso, necesitamos disponer de información muestral acerca de dicha variable.

Supongamos que dicha dependencia es positiva, es decir, que valores mayores de σ_t^2 ocurren en periodos en los que z_t es grande. Entonces el contraste consiste en:

a) Ordenar las observaciones por valores de la variable z_t , de menor a mayor.

b) Omitir p observaciones en mitad de la muestra.

c) Estimar dos veces el modelo original, una con las $\frac{T-p}{2}$ primeras

observaciones muestrales y otra con las $\frac{T-p}{2}$ últimas observaciones en la muestra. Nótese que el número de observaciones p omitidas en b) ha de ser suficientemente pequeño de modo que $\frac{T-p}{2}$ sea sustancialmente mayor que el número de parámetros en el modelo.

d) Sean SR_1 y SR_2 las sumas residuales de ambas regresiones. Entonces, bajo el supuesto de homoscedasticidad y Normalidad del término de error, el cociente

$$\lambda = \frac{SR_2}{SR_1} = \frac{\sigma_2^2}{\sigma_1^2}$$

sigue una distribución $F_{m,m}$, donde $m = \left(\frac{T-p}{2}\right) - k$.

La idea de este contraste es la siguiente: Si existe heteroscedasticidad del tipo que se ha supuesto desde un principio, entonces, con la ordenación de la muestra que se ha hecho, la varianza del término de error será mayor hacia el final de la muestra (es decir, cuando la variable z_t toma sus valores más grandes) que al principio. Como el cuadrado de los residuos está asociado con la varianza de u_t , entonces SR_2 debería ser sensiblemente mayor que SR_1 . Por tanto, si el valor del estadístico λ excede el valor que las tablas de la distribución F proporcionen, se rechazará la hipótesis nula de ausencia de heteroscedasticidad.

Algunas observaciones sobre este contraste:

1. Si se sospecha que la varianza del término de error depende *inversamente* de los valores que toma una variable z_t , entonces se debería ordenar la muestra de acuerdo con los valores *decrecientes* de dicha variable (es decir, de valores mayores a valores menores) y proceder exactamente del modo

descrito en las etapas a)-c) anteriores, sin ninguna variación, ni siquiera en el modo de comparar con las tablas de la distribución F .

2. ¿Cómo se elige el valor de p , el número de observaciones a omitir en la parte central de la muestra? Excluir «demasiadas» observaciones tiene dos efectos de signo contrario: Por una parte se pierden muchos grados de libertad para las dos regresiones que después se estiman, con lo que su estimación es poco precisa y el contraste global tiende a perder potencia. Por otro lado, la potencia tiende a aumentar porque, si existe heteroscedasticidad, al excluir muchas observaciones centrales, las submuestras resultantes se hacen más diferentes entre sí y será más fácil detectar el problema.

Si excluimos pocas observaciones, las observaciones del final de la primera parte de la muestra serán «demasiado» parecidas a las del comienzo de la última parte de la muestra, con lo que se premia el supuesto de homoscedasticidad, frente a la posibilidad de detectar la posible heteroscedasticidad que pudiera existir. Por consiguiente, se disminuye la potencia del contraste⁽²⁾. Harvey y Phillips (1974) sugieren no eliminar más de la tercera parte de las observaciones.

3. Si este contraste llevase a la conclusión de que el término de error del modelo no presenta heteroscedasticidad, podría deberse a que hemos comenzado de una mala especificación del parámetro σ_t^2 , que quizá depende de una variable diferente a la que hemos supuesto. Por ello, el contraste debería llevarse a cabo repetidamente con variables de las que podamos sospechar a priori que puede depender la varianza del término de error.

6.4.b. El contraste de Breusch y Pagan

Supongamos que la varianza del término de error en cada período (o para cada familia, etc.) depende de un vector de variables \mathbf{z}_t de dimensión p , es decir:

$$\sigma_t^2 = h(\mathbf{z}_t' \boldsymbol{\alpha}) = h(\alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \dots + \alpha_p z_{pt})$$

Nótese que si todos los coeficientes de la combinación lineal $\mathbf{z}_t' \boldsymbol{\alpha}$, excepto el término independiente α_0 , fuesen cero, entonces se tendría una situación de ausencia de heteroscedasticidad. En efecto, en tal caso: $\sigma_t^2 = h(\alpha_0)$, que es una constante y, por tanto, independiente de los valores de las variables \mathbf{z}_t . Por consiguiente, si pudiésemos estimar los coeficientes $\alpha_1, \alpha_2, \dots, \alpha_p$, un contraste de la hipótesis nula de homoscedasticidad vendría dado por la contrastación conjunta de las p restricciones lineales:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_p = 0$$

⁽²⁾ Recordemos que potencia de un contraste es la probabilidad de rechazar la hipótesis nula cuando es falsa, y también que la hipótesis nula en este caso es la de homoscedasticidad.

Este contraste se puede efectuar como sigue:

a) Se estima por MCO el modelo econométrico original y se obtienen los residuos correspondientes.

b) Se obtiene la serie (o sección cruzada, en su caso) de residuos normalizados al cuadrado:

$$\hat{\varepsilon}_t^2 = \frac{\hat{u}_t^2}{\hat{\sigma}_u^2}, \quad t = 1, 2, \dots, T$$

donde $\hat{\sigma}_u^2$ es la estimación de MV de la varianza del término de error bajo la hipótesis nula (homoscedasticidad), es decir: $\hat{\sigma}_u^2 = \frac{SR}{T}$, donde SR es la suma residual de la regresión en a).

c) Se estima una regresión de $\hat{\varepsilon}_t^2$ sobre una constante y las variables z_{1t} , z_{2t} , ..., z_{pt} , y se obtiene la suma explicada en dicha regresión. Notemos que como hay término independiente en esta regresión auxiliar, entonces la descomposición $ST = SE + SR$ es válida.

d) Bajo la hipótesis nula de homoscedasticidad, y supuesta una distribución Normal para el término de error, el cociente $\frac{SE}{2}$ calculado para la regresión en c) se distribuye, según crece el tamaño muestral, como una variable chi-cuadrado con p grados de libertad.

La interpretación de este contraste reside en que si los residuos fuesen homoscedásticos, entonces las variables z_t no deberían tener ningún poder explicativo acerca de los residuos transformados $\hat{\varepsilon}_t$ y, por consiguiente, SE debería ser pequeña. Si $\frac{SE}{2}$ es mayor que el valor de la chi-cuadrado en las tablas al nivel de significación escogido, entonces se considera suficientemente grande y se rechaza la hipótesis nula de homoscedasticidad.

En tal caso, se podría dividir cada observación por $\sqrt{z_t' \hat{\alpha}}$ como una aproximación a la desviación típica de cada periodo. La estimación MCO de este modelo transformado sería la estimación MCG del modelo original. En la práctica, este procedimiento de estimación no es muy recomendable porque aunque la aproximación lineal de la función sea suficientemente buena para la detección de heteroscedasticidad cuando ésta existe, sin embargo puede no serlo para la estimación de MCG, por lo que el modelo transformado puede no estar libre de heteroscedasticidad.

Es interesante hacer algunos comentarios con respecto a este contraste de heteroscedasticidad:

1. Si la heteroscedasticidad es tal que la varianza del término de error crece con el tiempo, los valores de u_t podrán ser positivos o negativos, pero tenderán a crecer en el tiempo en valor absoluto. Por tanto, un contraste de heteroscedasticidad debe basarse en la magnitud de los residuos, más que su

valores con signo. Por esto es que la variable a explicar en la regresión estimada en la etapa c) son los cuadrados de los residuos MCO normalizados. En seguida veremos que, como alternativa, otros contrastes utilizan $|\hat{u}_i|$ como variable a explicar.

2. Este contraste permite cierta flexibilidad en el sentido de que no precisa que el investigador especifique completamente la función h que explica la dependencia de la magnitud de los residuos con respecto a las variables z_1, \dots, z_p . La idea es que no importa cuál sea la forma funcional de dicha dependencia (dada por la función h), una representación lineal de la relación existente entre σ_i^2 y estas variables, entendida como una aproximación lineal de Taylor a la verdadera forma funcional h , debería tener un poder explicativo significativo.

3. La lista de variables z_i debería tener dos propiedades: a) ser corta, b) incluir pocas variables que no estén ya incluidas como variables explicativas en el modelo original. En efecto, mientras que es fácil pensar en situaciones en que la varianza del término de error del modelo cambia con los valores que toma alguna de las variables explicativas, no es obvio cómo justificar el que dicha varianza dependa de variables que no están en el modelo. Ello podría estar indicando una mala especificación del modelo lineal por omisión de la variable que explica a $\hat{\sigma}_i^2$. De este modo, el resultado de este contraste bien podría conducir a una nueva especificación del modelo econométrico.

En presencia de estacionalidad, las variables z_{it} deben incluir variables ficticias estacionales. Los cuadrados de las variables explicativas x_{it} son otros candidatos razonables a formar parte del vector z .

6.4.c. El contraste de Glesjer

Este contraste es más ambicioso que el anterior, pues trata de estimar la verdadera estructura de heteroscedasticidad, no limitándose al análisis de estructuras lineales. Sin embargo, una limitación del contraste de Glesjer es que sólo resulta útil cuando se cree que dicha estructura puede explicarse con tan sólo una variable, quizá junto con un término constante.

Las etapas en que se desarrolla el contraste son:

1. Estimar el modelo por MCO y obtener los residuos correspondientes.
2. Estimar una regresión del *valor absoluto* de \hat{u}_i , o de su cuadrado \hat{u}_i^2 , sobre una potencia de la variable z_i , es decir:

$$|\hat{u}_i| = \delta_0 + \delta_1 z_i^h + v_i$$

para distintos valores del exponente h : $h = \left\{ -1, 1, \frac{1}{2}, -\frac{1}{2} \right\}$. Escoger el valor de h que proporcione una «mejor» regresión. Los criterios de acuerdo con los que se escogen la «mejor» regresión deben tener en cuenta un coeficiente δ_1

significativo y una SR pequeña (quizá la menor de todas las regresiones que se hayan intentado)⁽³⁾.

3. Una vez seleccionado el valor del parámetro h , entonces se divide el vector de dimensión $k + 1$ formado por las observaciones (y_t, x_t) de cada período por $\hat{\delta}_0 + \hat{\delta}_1 z_t^h$ si se estimó una regresión de $|\hat{u}_t|$, o por $\sqrt{\hat{\delta}_0 + \hat{\delta}_1 z_t}$ si se estimó una regresión para \hat{u}_t^2 , y se estima el modelo, de nuevo por MCO, pero ahora con las variables transformadas, para obtener el estimador MCG del modelo original.

Algunas observaciones de interés:

a) Si ningún valor del parámetro h produce una regresión aceptable, entonces quizá el supuesto inicial de que la variable z_t explique la estructura de heteroscedasticidad sea inapropiado y se debiera repetir el intento con otra variable. De igual modo, la obtención de una regresión aceptable no excluye la posibilidad de que se pudiese encontrar otra regresión que utilizase una variable explicativa distinta de z_t , que pudiera ser al menos tan buena como la que ya hemos obtenido. Por ello, procede intentar todas las alternativas que el investigador estuviera a priori dispuesto a considerar.

b) Recordemos que en una situación de heteroscedasticidad, el método de mínimos cuadrados ponderados (es decir, MCG) equivale a repetir la aplicación de MCO tras llevar a cabo una ponderación de las observaciones. Dicha ponderación debe consistir *siempre* en dividir las variables de un período por una estimación de la *desviación típica del término de error* para ese período.

6.4.d. El contraste de Harvey

En algunos casos, una especificación del tipo $\sigma_t^2 = e^{z_t \alpha}$ es razonable. Por ejemplo, dicha función incluye como caso particular la expresión $\sigma_t^2 = \sigma_u^2 x_t^s$ cuando $z_t = (\ln \sigma_u, \ln x_t)$ y $\alpha = (2, s)$. Esta estructura de heteroscedasticidad se conoce como *multiplicativa*. El contraste consta de las siguientes etapas:

a) Estimar el modelo por MCO ignorando la posible heteroscedasticidad y obtener los residuos \hat{u}_t .

b) Estimar por MCO la regresión:

$$\ln \hat{u}_t^2 = z_t' \alpha + \varepsilon_t = \alpha_1 + \alpha_2 z_{2t} + \dots + \alpha_p z_{pt} + \varepsilon_t \quad [6.3]$$

c) El estadístico F para el contraste de significación global de esta regresión sigue una distribución chi-cuadrado, con $p - 1$ grados de libertad, y puede interpretarse como el contraste de la hipótesis nula de homoscedasticidad, por las mismas razones que dimos en el contraste de Breusch-Pagan.

⁽³⁾ Sin embargo, Goldfeld y Quandt [*Nonlinear Methods in Econometrics*, North Holland (1972)] han demostrado que el término de error de estas regresiones tiene esperanza no nula, heteroscedasticidad y autocorrelación, por lo que los contrastes de hipótesis contruidos sobre una estimación MCO no son válidos. La elección del valor «óptimo» del parámetro h se convierte así en un delicado problema que limita el interés de este contraste.

Estos resultados fueron obtenidos por Harvey (1976), quien además probó que las estimaciones de los coeficientes α del modelo en la regresión [6.3] son consistentes (es decir, su posible sesgo desaparece cuando el tamaño muestral tiende a infinito), excepto para el término independiente α_1 , que es sesgado con independencia del tamaño muestral.

La matriz de covarianzas asintótica del vector $(\alpha_2, \alpha_3, \dots, \alpha_p)$ es la matriz D que resulta al excluir la primera fila y columna de $(Z'Z)^{-1}$, multiplicada por 4,935. En consecuencia, el contraste de la hipótesis nula se lleva a cabo comparando el valor de $\frac{\hat{\alpha}^* D^{-1} \hat{\alpha}^*}{4,935}$ con los valores tabulados de la variable chi-cuadrado con $p - 1$ grados de libertad. El vector $\hat{\alpha}^*$ es el vector de coeficientes en [6.3], excluyendo el término independiente α_1 .

Es importante hacer notar que el sesgo del coeficiente α_1 no tiene ninguna influencia sobre este contraste, pues los únicos parámetros relevantes son los restantes parámetros $\alpha_i, i = 2, \dots, p$. El único problema con este contraste es el habitual de que utiliza una matriz de covarianzas estimada bajo un determinado supuesto de heteroscedasticidad, y cabría considerar distintas alternativas al mismo.

Por otra parte, dicho sesgo tampoco tiene ninguna influencia sobre la estimación de los parámetros β_i , puesto que la varianza del término de error viene dada por $\sigma_t^2 = e^{\alpha_1} e^{\alpha_2 z_{2t} + \dots + \alpha_p z_{pt}}$, y el término e^{α_1} puede sacarse factor común, pasando a jugar el papel del parámetro σ_u^2 que, como sabemos, no interviene en la estimación de mínimos cuadrados generalizados. De hecho, si se rechaza la hipótesis de homoscedasticidad, entonces dividiríamos las observaciones del período t por $\sqrt{e^{\alpha_2 z_{2t} + \dots + \alpha_p z_{pt}}}$, ignorando la estimación de α_1 . Una vez obtenido el estimador del vector β , su matriz de covarianzas se obtiene multiplicando la matriz $(X^* X^*)^{-1}$ de las variables transformadas por la estimación de σ_u^2 , que se obtendrá de la expresión:

$$\hat{\sigma}_u^2 = \frac{\hat{u}' \Sigma^{-1} \hat{u}}{T - k}$$

6.4.e. El contraste de White

White (1980) ha propuesto un contraste general, que no precisa especificar la forma que puede adoptar la heteroscedasticidad:

- a) Estimar el modelo original por MCO, ignorando la posible heteroscedasticidad.
- b) Estimar una regresión del cuadrado de los residuos mínimo-cuadráticos anteriores, sobre una constante, los regresores del modelo original, sus cuadrados y productos cruzados de segundo orden⁽⁴⁾.

⁽⁴⁾ Es decir, que si hay varios regresores en el modelo original, en esta regresión auxiliar se incluirán x_{2t}^2 , así como $x_{2t}x_{3t}$, pero no $x_{2t}x_{3t}x_{4t}$.

c) Al aumentar el tamaño muestral, el producto TR^2 , donde T es el tamaño muestral y R^2 el coeficiente de determinación de la última regresión, sigue una distribución chi-cuadrado con $p - 1$ grados de libertad, donde p es el número de regresores del modelo estimado en b).

Mientras que el tamaño muestral crece con el número de observaciones, el coeficiente de determinación tenderá a cero bajo la hipótesis nula de homoscedasticidad. Sólo cuando la varianza del término de error depende (aunque sea de forma no lineal) de las variables explicativas del modelo, el coeficiente R^2 no tenderá a cero. En tal caso, el producto TR^2 permanecerá a un cierto nivel, lejos de cero, y es de esperar que supere al valor de tablas de la distribución chi-cuadrado. Como puede verse, el contraste no sólo requiere que los regresores no tengan un alto poder explicativo de σ_t^2 , sino que el coeficiente de determinación de dicha regresión tienda hacia cero rápidamente.

6.4.f. El contraste de Spearman

Este contraste se basa en la intuición de que si la varianza del término de error σ_t^2 depende directamente de los valores de la variable x_t , entonces el tamaño de los residuos debería estar relacionado con el tamaño de dicha variable. Así, tras estimar el modelo por MCO, se ordenan en sentido creciente tanto el valor absoluto de los residuos obtenidos, $|\hat{u}_t|$, como los valores de x_t , y se calcula el *coeficiente de correlación de rangos*:

$$r_s = 1 - \frac{6\sum_1^T d^2}{T(T^2 - 1)}$$

donde d es la diferencia en el puesto que ocupan en dichas clasificaciones el valor x_t y el valor absoluto $|\hat{u}_t|$ correspondiente a un mismo período.

Si el tamaño muestral es grande, entonces la expresión

$$\frac{r_s \sqrt{T-2}}{\sqrt{1-r_s^2}}$$

se distribuye, aproximadamente, como una t de Student con $T - 2$ grados de libertad. Esta expresión podría calcularse para *cada una* de las variables explicativas del modelo, con el objeto de tratar de hallar un modelo de dependencia para σ_t^2 .

6.5. CONTRASTE DE IGUALDAD DE VARIANZA ENTRE SUBMUESTRAS

Un contraste especial de heteroscedasticidad aparece con mucha frecuencia en aplicaciones empíricas. Se trata del caso en que disponiéndose de muestras recogidas de diferentes poblaciones, y bajo el supuesto de que

todas ellas tienen matriz de covarianzas escalar, se pretende contrastar si sus varianzas son iguales. Este tipo de contraste será de gran utilidad en capítulos posteriores.

Bajo el supuesto de que existen m variables aleatorias u_i , distribuidas $N(0, \sigma_i^2)$, se trata de contrastar la hipótesis nula:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$$

El contraste se lleva a cabo del siguiente modo: Se obtiene la varianza muestral de la variable endógena dentro de cada grupo, $s_1^2, s_2^2, \dots, s_m^2$, así como la varianza calculada con *todas* las observaciones muestrales, s^2 . Al aumentar el tamaño muestral, el estadístico

$$Q = (T - m) \ln s^2 - \sum_1^m (T_i - 1) \ln s_i^2$$

sigue, aproximadamente, una distribución chi-cuadrado con $m - 1$ grados de libertad. Basta comparar el valor del estadístico Q con el valor correspondiente en las tablas de dicha distribución. Cabe observar que este resultado es más aproximado si se divide el valor del estadístico Q por:

$$1 + \frac{1}{3(m-1)} \left(\sum_1^m \frac{1}{T_i} - \frac{1}{T} \right)$$

aunque, como puede verse, cuando el número de observaciones de las distintas submuestras crece, esta corrección deja de ser importante.

6.6. TRANSFORMACION DE BOX Y COX

Consideremos el modelo no lineal:

$$y_t = e^{\mathbf{x}_t' \boldsymbol{\beta}} \cdot e^{u_t}, \quad t = 1, 2, \dots, T \quad [6.4]$$

donde u_t es un ruido blanco con distribución Normal $(0, \sigma_u^2)$. En este modelo, y_t tiene una distribución log-Normal, es decir, su logaritmo es Normal, ya que:

$$\ln y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t \quad [6.5]$$

Por otra parte:

$$\text{Var } y_t = (\text{Var } e^{u_t}) \cdot (e^{\mathbf{x}_t' \boldsymbol{\beta}})^2 = (e^{\mathbf{x}_t' \boldsymbol{\beta}})^2 \cdot e^{\sigma_u^2} (e^{\sigma_u^2} - 1) \quad [6.6]$$

luego y_t tiene heteroscedasticidad. Sin embargo, es claro de [6.5] que su logaritmo, $\ln y_t$, es homoscedástico, puesto que:

$$\text{Var}(\ln y_t) = \text{Var } u_t = \sigma_u^2$$

por lo que el estimador MCO aplicado a [6.5] es eficiente.

La transformación logarítmica es un caso particular de la familia de transformaciones propuestas por Box y Cox (1964), consistentes en suponer que existe un valor de λ tal que:

$$g(y_t, \lambda) = \frac{y_t^\lambda - 1}{\lambda} = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

donde u_t es ruido blanco con distribución Normal. Si $\lambda = 0$, $g(y_t, 0) = \ln y_t$, como puede verse pasando al límite, mientras que si $\lambda = 1$ tenemos el modelo lineal.

Como muestra [6.6], en el modelo [6.4] la desviación típica de la variable endógena es proporcional a su valor esperado, $e^{\mathbf{x}_t' \boldsymbol{\beta}}$, y lo que consigue la transformación logarítmica es transformar el modelo en homoscedástico.

6.7. HETEROSCEDASTICIDAD CONDICIONAL AUTORREGRESIVA (ARCH)

En algunas variables económicas, y especialmente en el comportamiento de las tasas de rentabilidad de activos financieros, se observan períodos de relativa estabilidad, seguidos de intervalos de alta volatilidad, antes de volver a la estabilidad.

Parece adecuado en tales casos especificar un modelo de heteroscedasticidad del tipo:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

donde u_t es $N(0, \sigma_t^2)$ con:

$$\sigma_t^2 = \delta_0 + \delta_1 u_{t-1}^2$$

que refleja que la varianza del término de error, aproximada por u_t^2 , evoluciona en el tiempo con cierta suavidad, alternando intervalos de valores pequeños de σ_t^2 (estabilidad) con otros de valores elevados (volatilidad).

El modelo anterior es estacionario si $|\delta_1| < 1$. En realidad, no sólo es lógico esperar que δ_1 sea positivo, sino que, si no lo fuera, podríamos tener $\sigma_t^2 < 0$ para algún t . En consecuencia, $0 < \delta_1 < 1$, junto con $\delta_0 > 0$, son desigualdades que las estimaciones del modelo deben satisfacer.

Bajo Normalidad de u_t , el logaritmo de la verosimilitud es:

$$\ln L = -\frac{1}{2} \sum_2^T \ln(\delta_0 + \delta_1 u_{t-1}^2) - \frac{1}{2} \sum_2^T \frac{u_t^2}{\delta_0 + \delta_1 u_{t-1}^2}$$

donde $u_t = y_t - \mathbf{x}_t' \boldsymbol{\beta}$.

Puede probarse (Gracia-Díez y Novales, 1993) que la estimación de máxima verosimilitud puede obtenerse por el siguiente procedimiento:

1. Estimar β por MCO y obtener los residuos \hat{u}_t .
2. A partir de una estimación inicial $(\hat{\delta}_0, \hat{\delta}_1)$, definir las variables:

$$h_t^2 = \hat{\delta}_0 + \hat{\delta}_1 \hat{u}_{t-1}^2 \quad t = 2, \dots, T$$

$$z_t' = \left(\frac{1}{h_t^2}, \frac{\hat{u}_{t-1}^2}{h_t^2} \right) \quad t = 2, \dots, T$$

$$w_t = \frac{\hat{u}_t^2}{h_t^2} - 1 \quad t = 2, \dots, T$$

y estimar una regresión de w_t sobre el vector z_t . Los coeficientes estimados son las *correcciones a añadir* a la estimación inicial $(\hat{\delta}_0, \hat{\delta}_1)$:

$$\begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix}_i = \begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix}_{i-1} + (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{W}$$

Este es un procedimiento iterativo que debe utilizarse hasta lograr la convergencia. Es razonable comenzar de $\hat{\delta}_0 = \hat{\sigma}_u^2$, la estimación MCO, y $\hat{\delta}_1 = 0$ o quizá $\frac{1}{2}$. La matriz de covarianzas del estimador resultante es

$$\text{Var}(\hat{\delta}_0, \hat{\delta}_1)' = 2(\mathbf{Z}'\mathbf{Z})^{-1}$$

3. Una vez que se dispone de $(\hat{\delta}_0, \hat{\delta}_1)$, se calcula:

$$r_t^2 = \frac{1}{h_t^2} + 2 \frac{(\hat{\delta}_1 \hat{u}_t)^2}{(h_{t+1}^2)^2}$$

$$c_t = \frac{1}{h_t^2} - \frac{\hat{\delta}_1}{h_{t+1}^2} \left(\frac{\hat{u}_{t+1}^2}{h_{t+1}^2} - 1 \right)$$

y se estima una regresión de $\tilde{u}_t = \frac{\hat{u}_t c_t}{r_t}$ sobre el vector $\tilde{\mathbf{x}}_t = \mathbf{x}_t r_t$ para obtener la corrección a introducir en el vector β :

$$\hat{\beta}_i = \hat{\beta}_{i-1} + (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{u}}$$

Cuando el algoritmo converge, la matriz de covarianzas del estimador resultante es $\text{Var}(\hat{\beta}) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$. Una vez que se dispone de este estimador de β , podría pensarse en iterar desde el principio, pero las ganancias que así se obtienen son mínimas.

El modelo anterior se conoce como modelo ARCH(1) de regresión. El modelo ARCH(p) tiene como ecuación de σ_t^2 :

$$\sigma_t^2 = \delta_0 + \delta_1 u_{t-1}^2 + \dots + \delta_p u_{t-p}^2$$

El vector \mathbf{z}_t es ahora $\mathbf{z}'_t = \left(\frac{1}{h_t^2}, \frac{\hat{u}_{t-1}^2}{h_t^2}, \dots, \frac{\hat{u}_{t-p}^2}{h_t^2} \right)$, mientras que

$$r_t^2 = \frac{1}{h_t^2} + 2\hat{u}_t^2 \sum_1^p \frac{\delta_i^2}{h_{t+i}^4}$$

$$c_t = \frac{1}{h_t^2} - \sum_1^p \frac{\delta_i}{h_{t+i}^2} \left(\frac{\hat{u}_{t+i}^2}{h_{t+i}^2} - 1 \right)$$

Para disminuir el número de parámetros, el investigador representa a veces el modelo ARCH(p) como:

$$h_t^2 = \delta_0 + \delta_1 S_t$$

donde:

$$S_t = \sum_1^p b_i \hat{u}_{t-i}^2$$

tomando los coeficientes b_i valores fijados de antemano (por ejemplo, decreciendo linealmente). En tal caso, la iteración sobre el vector δ se consigue mediante una regresión de $w_t = \frac{\hat{u}_t^2}{h_t^2} - 1$ sobre el vector $\left(\frac{1}{h_t^2}, S_t \right)$, de modo que equivale a tratar a S_t^2 como el retardo \hat{u}_{t-1}^2 en un ARCH(1). La matriz de covarianzas es igual a dos veces la inversa de la matriz de momentos cruzados de las variables explicativas de esta regresión.

Para iterar sobre el vector β se define

$$r_t^2 = \frac{1}{h_t^2} + 2(\delta_1^2 \hat{u}_t^2) \sum_1^p \frac{b_i^2}{h_{t+i}^4}$$

$$c_t = \frac{1}{h_t^2} - \delta_1 \sum_1^p \frac{b_i}{h_{t+i}^2} \cdot (\hat{u}_{t+i}^2 - h_{t+i}^2)$$

con matriz de covarianzas como en el modelo ARCH habitual.

Para contrastar la existencia de una estructura ARCH(p) se utiliza el producto TR^2 , donde R^2 es el coeficiente de determinación de una regresión de \mathbf{W}_0 sobre \mathbf{Z} , donde \mathbf{W}_0 es el vector \mathbf{W} evaluado bajo H_0 . Si la hipótesis nula es $H_0: \delta = \mathbf{0}$, entonces \mathbf{W}_0 son los residuos MCO al cuadrado, normalizados por $\hat{\sigma}_u^2$. Equivalentemente, R^2 es el coeficiente de determinación de la regresión de \hat{u}_t^2 sobre una constante y p retardos suyos. Un procedimiento similar puede utilizarse para contrastar $H_0: \text{ARCH}(p-1)$ frente a $H_1: \text{ARCH}(p)$.

Otros modelos más generales son el ARCH en media:

$$y_t = \mathbf{x}'_t \beta + \rho h_t^2 + u_t$$

$$h_t^2 = \delta_0 + \delta_1 u_{t-1}^2$$

Con ellos, estimamos un modelo inicial.

$$\text{PIB}_i = -187,7 + 3,76 \text{ OCU}_i, \quad i = 1, 2, \dots, 18$$

$$(182,9) \quad (0,19) \quad [6.7]$$

$$\bar{R}^2 = 0,96; \quad \hat{\sigma}_u = 518,8$$

en la que los estadísticos t de ambos coeficientes son $-1,0$ y $20,3$, respectivamente. Es decir, aunque la constante se estima con poca precisión (alta varianza), y por ello resulta no significativa, la relación que aparece entre producción y empleo en las distintas comunidades, supuesto igual para todas ellas, se estima con bastante precisión, evaluándose en un aumento de 3.760 millones de pesetas por cada 1.000 empleados, con datos de 1989.

Existe la posibilidad de que, por razones similares a las expuestas en la Sección 6.2, la varianza de la componente del PIB no explicada por el empleo aumente con éste, es decir, que haya heteroscedasticidad con σ_i^2 dependiendo directamente de OCU_i^2 , es decir, con la desviación típica σ_i dependiendo de OCU_i .

El contraste de Breusch-Pagan consiste en estimar una regresión de los residuos MCO normalizados sobre una constante y OCU_i^2 , obteniendo en ella un valor de $\text{SE}/2$ igual a 8,97, que se distribuye como una chi-cuadrado con un grado de libertad. En nuestro caso, 8,97 excede claramente al valor crítico de las tablas de χ_i^2 al 95 por 100, que es 3,84, por lo que rechazamos la hipótesis nula de ausencia de heteroscedasticidad.

Para el contraste de Goldfeld y Quandt ordenamos las comunidades crecientemente de acuerdo con su nivel de empleo, omitimos las seis comunidades que ocupan el lugar central en esta ordenación, y estimamos la relación entre PIB y empleo con cada uno de los otros dos grupos de seis comunidades. El cociente de sumas residuales entre ambas regresiones fue $\text{SR}_2/\text{SR}_1 = 65,2$, que excede claramente al valor crítico de una variable $F_{4,4}$ al 95 por 100, que es de 6,39, por lo que se rechaza nuevamente la hipótesis nula de ausencia de heteroscedasticidad.

Para llevar a cabo el contraste de correlación de rangos ordenamos las comunidades, de acuerdo con los valores crecientes de su nivel de empleo y el valor absoluto de sus residuos MCO, del modo que aparece en las últimas columnas de la Tabla 6.1, con $\sum_1^{18} d_i^2 = 1.004$. El coeficiente de correlación de rangos resulta, por consiguiente, negativo, por lo que no puede llevarse a cabo este contraste. En todo caso, no proporciona evidencia en contra de la hipótesis nula de ausencia de heteroscedasticidad.

El contraste de Glesjer consiste en estimar una serie de regresiones a partir de los residuos MCO, con los que obtuvimos:

$$1. \quad \hat{u}_i^2 = -68.600,1 + 419,7 \text{ OCU}_i; \quad \bar{R}^2 = 0,38; \quad \hat{\sigma}_u = 345.319,9$$

$$(123,4)$$

$$2. \quad \hat{u}_i^2 = 77.998,8 + 0,16 \text{ OCU}_i^2; \quad \bar{R}^2 = 0,30; \quad \hat{\sigma}_u = 369.124,7$$

$$(0,06)$$

$$3. \hat{u}_i^2 = -330.144,3 + 23.333,0 \sqrt{OCU_i}; \quad \bar{R}^2 = 0,37; \quad \hat{\sigma}_u = 347.652,4 \\ (6.973,7)$$

$$4. \hat{u}_i^2 = 496.995,6 - 4.711.435,0 \frac{1}{\sqrt{OCU_i}}; \quad \bar{R}^2 = 0,11; \quad \hat{\sigma}_u = 415.665,4 \\ (2.709.583)$$

que sugieren que lo más adecuado es una representación de σ_i^2 como función creciente de OCU_i o de OCU_i^2 . A continuación podríamos dividir los datos de PIB y empleo de cada comunidad por el modelo estimado 1 y volver a estimar por MCO obteniendo esta vez el estimador MCG.

Generalizando algo más el modelo, introducimos una segunda variable explicativa que discriminase entre las comunidades que tienen un PIB per cápita superior a la media nacional⁽⁵⁾ de aquellas que lo tienen inferior, obteniendo:

$$PIB_i = -372,1 + 3,66 OCU_i + 581,1 ALTO_i \quad [6.8] \\ (168,8) \quad (0,16) \quad (213,8)$$

$$\bar{R}^2 = 0,97; \quad \hat{\sigma}_u = 438,6$$

en la que las tres variables explicativas tienen estadísticos t superior a 2. Este modelo supone una ganancia relevante con respecto al inicial, como se aprecia en una reducción de $\hat{\sigma}_u$ superior al 15 por 100.

El contraste de heteroscedasticidad de White consiste en estimar una regresión de los residuos del modelo anterior, al cuadrado, sobre las variables explicativas, sus cuadrados y productos cruzados, en la que obtuvimos:

$$\hat{u}_i^2 = -32.374,0 + 384,9 OCU_i - 12.369,7 ALTO_i - \\ - 0,047 OCU_i^2 - 95,6 (OCU_i * ALTO_i)$$

y en la que $TR^2 = 7,77$ se distribuye como una χ_5^2 , cuyos valores críticos al 90 y 95 por 100 son, respectivamente, 9,24, 11,1, por lo que, de acuerdo con este contraste, no rechazaríamos la hipótesis nula de ausencia de heteroscedasticidad en [6.8]. No hay que olvidar, sin embargo, que tanto este contraste como los anteriores sólo tienen validez asintótica, y los estamos utilizando en este ejemplo con un número reducido de observaciones.

La detección de heteroscedasticidad puede sugerir en algún caso una especificación incompleta del modelo. La regresión estimada con los residuos de [6.8]:

$$\hat{u}_i^2 = 81.140,9 + 0,081 OCU_i^2; \quad R^2 = 0,29 \\ (56.142,2) \quad (0,031)$$

⁽⁵⁾ Así, construimos una variable ficticia, $ALTO_i$, que toma el valor 1 en las comunidades con PIB per cápita alto: Aragón, Baleares, Cataluña, Comunidad Valenciana, Madrid, Navarra, País Vasco y La Rioja, y 0 en el resto.

sugiere posible heteroscedasticidad, con σ_i^2 creciendo con OCU_i^2 . Si incluimos esta variable en el modelo, tenemos:

$$PIB_i = -28,4 + 2,56 OCU_i + 0,00051 OCU_i^2 + 508,5 ALTO_i \quad [6.9]$$

(227,7) (0,56) (0,00025)

$$\bar{R}^2 = 0,976; \quad \hat{\sigma}_u = 398,5$$

en la que hemos reducido $\hat{\sigma}_u$ en casi un 10 por 100 adicional, lo que sugiere que la relación ente PIB y empleo no es lineal⁽⁶⁾. El análisis posterior de los residuos de [6.9] no indicó evidencia adicional de heteroscedasticidad.

Por otra parte, una vez detectada la dependencia de σ_i^2 respecto de OCU_i^2 en el modelo [6.8], podemos pasar a la estimación MCG sin más que dividir todas las variables por OCU_i y estimar el modelo:

$$\frac{PIB_i}{OCU_i} = -6,43 \cdot \frac{1}{OCU_i} + 3,36 + 61,3 \cdot \frac{ALTO_i}{OCU_i} \quad [6.10]$$

(14,9) (0,13) (40,4)

en el que hemos mantenido la constante en el centro del modelo porque proviene de dividir OCU_i por si misma. Dicho de otro modo, para obtener la relación entre las estimaciones de los coeficientes en [6.8] y [6.10], debemos escribir este último:

$$PIB_i = -6,43 + 3,36 OCU_i + 61,3 ALTO_i \quad [6.11]$$

(14,9) (0,13) (40,4)

que proporciona una estimación más precisa de la relación entre PIB y empleo, a la vez que sugiere que la diferencia entre la parte del PIB que no queda explicada por el empleo, en las comunidades de mayor y menor producto per cápita, es menos importante de lo que inicialmente habíamos estimado (ineficientemente). Como puede verse, las varianzas de los tres coeficientes estimados se han reducido notablemente con respecto a las estimaciones MCO de [6.8].

6.9. EJERCICIOS PRACTICOS

Ejercicio 6.1. Consideremos de nuevo el ejercicio práctico 5.1, en el que se estimaba el modelo:

$$y_t = 50 + x_{2t} - 3x_{3t} + u_t, \quad t = 1, 2, \dots, 120$$

utilizando 100 muestras diferentes. La varianza del término de error en cada período era $\sigma_t^2 = 25x_{3t}$, es decir, que la desviación típica de u_t era proporcio-

⁽⁶⁾ La interpretación económica del modelo [6.9] es algo más delicada, pues sugiere la existencia de rendimientos crecientes en el factor trabajo, pero es demasiado sencillo como para interpretarlo rigurosamente en ese sentido.

nal a la raíz cuadrada del valor de la variable x_3 en cada período. En el Ejercicio 5.1 se establecían comparaciones entre las distribuciones teóricas y empíricas de los estimadores MCO y de MCG, así como entre sus matrices de covarianzas.

A diferencia de las matrices de covarianzas de estos dos estimadores, la corrección de White para la matriz de covarianzas del estimador MCO depende de los residuos MCO y, por tanto, de la realización de la variable endógena y_t . Como consecuencia, dicha matriz de covarianzas cambia de una muestra a otra. Puesto que conocemos la verdadera estructura de covarianzas del término de error, podemos comparar las varianzas estimadas del estimador MCO, $\hat{\beta}_i$, $i = 1, 2, 3$, tanto cuando se utiliza en la expresión $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, la verdadera estructura de σ_i^2 , como cuando se utiliza la corrección de White. De las 100 estimaciones las varianzas obtenidas por el primer método superaron a las del segundo en 71 ocasiones para $\hat{\beta}_1$, y en 72 ocasiones tanto para $\hat{\beta}_2$ como para $\hat{\beta}_3$. La sugerencia de White parece subestimar las varianzas de los estimadores MCO, al menos en este ejemplo.

Se efectuaron asimismo los contrastes de heteroscedasticidad que se han sugerido en este capítulo:

1. El contraste de Goldfeld y Quandt se llevó a cabo estimando el modelo, por una parte, con las primeras 40 observaciones, y por otro lado, con las observaciones 81 a 120. Contrastando al 95 por 100 de confianza, se detectó heteroscedasticidad en 93 ocasiones, y en 99 ocasiones cuando se contrastó al 99 por 100 de confianza.

Las 40 observaciones omitidas en la parte central de la muestra son seguramente un número excesivo aunque, a la vista de los resultados, el contraste no perdió mucha potencia.

2. El contraste de Breusch y Pagan consiste, en este ejemplo, en estimar una regresión del cuadrado de los residuos MCO sobre una constante y la variable x_{3t} . La mitad de la suma explicada en dicha regresión sigue una distribución chi-cuadrado con un grado de libertad. En este caso particular, ésta es precisamente la regresión que describe la estructura de heteroscedasticidad, por lo que es de esperar que el contraste dé buenos resultados. Contrastando al 95 por 100 de confianza, se detectó heteroscedasticidad en 91 ocasiones, y contrastando al 99 por 100 de confianza se detectó heteroscedasticidad precisamente en 99 ocasiones, lo que es bastante consistente con el concepto estadístico de nivel de confianza.

3. El contraste de Glesjer se llevó a cabo mediante regresiones de $|\hat{u}_t|$, tanto sobre $\sqrt{x_{3t}}$ como sobre x_{3t} y x_{3t}^2 , además de un término constante en cada caso. Dada la estructura de heteroscedasticidad que hemos impuesto, la primera de estas regresiones debería ser la que diera mejor resultado. Los coeficientes de la variable explicativa resultaron siempre significativos. Por comparación de las sumas residuales, la primera regresión resultó ser «la mejor» en 72 ocasiones, la segunda fue «la mejor» en 17 ocasiones y la tercera en 11 ocasiones.

Ejercicio 6.2. Se estimó el mismo modelo del Ejercicio 6.1, es decir:

$$y_t = 50 + x_{2t} - 3x_{3t} + u_t$$

donde las variables x_2 y x_3 fueron generadas de acuerdo con:

$$x_{2t} = 25 + 0,75x_{2t-1} + \varepsilon_{2t}$$

$$x_{3t} = 25 + 2t + \varepsilon_{3t}$$

con $\varepsilon_{2t} \sim N(0,25)$ y $\varepsilon_{3t} \sim \text{Normal}(0,100)$, independientes entre sí. La estructura de heteroscedasticidad fue cambiada a:

$$\text{Var}(u_t) = \sigma_t^2 = 400x_{2t}$$

es decir, que la varianza de u_t depende ahora de la segunda variable explicativa, no de la tercera como en el Ejercicio 6.1. Se efectuaron 50 estimaciones. con los siguientes resultados:

	Estimación MCO			Estimación de mínimos cuadrados generalizados		
	Media	Desviaciones típicas Empírica	Teórica	Media	Desviaciones típicas Empírica	Teórica
$\hat{\beta}_1$	114,6	(304,5)	(317,0)	116,1	(305,6)	(315,6)
$\hat{\beta}_2$	0,36	(2,91)	(3,0)	0,34	(2,92)	(3,0)
$\hat{\beta}_3$	-3,00	(0,28)	(0,28)	-3,01	(0,28)	(0,28)
$\hat{\sigma}_u^2$				411,15	(57,9)	(52,3)

Como puede verse, el estimador MCG es apenas más eficiente que el estimador MCO en este modelo. Curiosamente, las desviaciones empíricas MCO son ligeramente inferiores a las de MCG. El contraste de White, efectuado al 95 por 100 de confianza, no detectó heteroscedasticidad en ninguna de la 100 estimaciones. El contraste de Breusch y Pagan detectó heteroscedasticidad en sólo ocho ocasiones al 95 por 100 de confianza, y en una sola trabajando al 99 por 100 de confianza. Además, se efectuó el contraste de Glesjer, llevado a cabo mediante regresiones de $|\hat{u}_t|$ sobre $\sqrt{x_{2t}}$, x_{2t} y x_{2t}^2 . La primera de estas regresiones describe la verdadera estructura de heteroscedasticidad. De las 100 estimaciones se eligió la primera regresión como la mejor tan sólo 56 veces, la segunda en dos ocasiones y la tercera en 42.

Al contrastar una a una las hipótesis nulas, de que cada parámetro es igual a su verdadero valor, 50, 1 y -3, respectivamente, éstas se mantuvieron en 98, 98 y 94 ocasiones, respectivamente, tanto cuando se utilizaron las estimaciones MCO como cuando se usaron las estimaciones MCG.

La media de las estimaciones de σ_u^2 es bastante próxima al verdadero valor de 400,0. Teóricamente, $117 \frac{\hat{\sigma}_u^2}{\sigma_u^2}$ o, lo que es lo mismo, $0,2925 \hat{\sigma}_u^2$, sigue una distribución chi-cuadrado con 117 grados de libertad. Ello implica para $\hat{\sigma}_u^2$ una esperanza de $\frac{117}{0,2925} = 400$ y una varianza igual a $\frac{234}{0,2925^2} = 2735$, ambas muy similares a los momentos de la distribución empírica del estimador $\hat{\sigma}_u^2$ que hemos reflejado en la tabla anterior.

La diferencia fundamental entre este ejercicio y el previo es que en éste la estructura de σ_i^2 cambia con x_{2i} , que es una variable que fluctúa alrededor de su esperanza, que es 100, pero que no tiene una tendencia determinista. Por ello, en un sentido estricto, la heteroscedasticidad en este caso existe, puesto que la varianza del término de error cambia en el tiempo. Sin embargo, no tiene una estructura que afecte en gran medida a la precisión de las estimaciones, como hemos visto. Las diferencias entre las estimaciones MCO y de MCG existen cuando la varianza sufre variaciones importantes que hacen ineficiente ponderar por igual a todas las observaciones en la estimación del modelo. El caso más claro es cuando, como en el Ejercicio 6.1, la heteroscedasticidad tiene una tendencia importante, ya sea creciente o decreciente.

De hecho, ya hemos visto que en la situación de este ejercicio incluso es difícil detectar la existencia de heteroscedasticidad por medio de los contrastes habituales.

PROBLEMAS

Problema 6.1. (Heteroscedasticidad por grupos.) Se pretende estimar la ecuación de consumo:

$$c_i = a + by_i^d + u_i \quad [1]$$

donde c_i denota el consumo familiar y y_i^d la renta disponible. Para ello se recoge información de N familias, estructuradas en cinco subgrupos de tamaños N_i , $i = 1, 2, \dots, 5$. De cada uno de ellos se obtiene el consumo y la renta disponible agregados, $C_j = \sum_1^{N_j} c_i$, $Y_j^d = \sum_1^{N_j} y_i^d$, con lo que se especifica el modelo:

$$C_j = \alpha + \beta Y_j^d + v_j, \quad j = 1, 2, \dots, 5 \quad [2]$$

donde C_j e Y_j denotan los agregados para cada subgrupo poblacional, y $\sum_1^5 N_j = N$.

a) Determinar la relación entre los coeficientes α , β del modelo [2] y los coeficientes a , b , del modelo [1], así como la relación entre los términos de error v_j y u_i . En particular, si suponemos que u_i es independiente en el tiempo con $E(u_i) = 0$, $\text{Var}(u_i) = \sigma_u^2$ para todo i , ¿cuáles serán las propiedades de v_j ?

b) Encontrar el sistema de ecuaciones normales para la estimación del modelo [2].

c) Probar que si la información recogida consistiese en los promedios del consumo y la renta disponible en vez de sus agregados, entonces el sistema de ecuaciones normales (y, por tanto, las estimaciones de mínimos cuadrados) serían exactamente iguales a las obtenidas en b).

d) Estimar dicho modelo y obtener la matriz de covarianzas de las estimaciones de los parámetros, teniendo en cuenta la siguiente información muestral:

N_j	Y_j^d	C_j
20	2	6
5	5	4
30	2	4
25	0	6
30	1	2

Problema 6.2. Se dispone de 20 observaciones de cada uno de dos países diferentes, y se pretende estimar, para cada uno de ellos, el modelo:

$$y_i = \alpha + \beta x_i + u_i$$

El término de error es en cada país una variable aleatoria con distribución Normal e independiente en el tiempo. La varianza de u_i es, cada período, igual a 1,0 en el primer país, e igual a 2,0 en el segundo. Además se sabe que, debido a shocks comunes a ambos países, la covarianza entre los términos de error de uno y otro país es, en cada período, igual a 1,0. Los momentos muestrales de segundo orden, así como las medias muestrales de las variables, vienen resumidos en las tablas:

	y_1	y_2	x_1	x_2		<i>Medias</i>
y_1	10	-1	1	-1	y_1	2
y_2	-1	15	5	1	y_2	3
x_1	1	5	11	1	x_1	0
x_2	-1	1	1	1	x_2	5

- Estimar el modelo para cada país por MCO.
- Justificar la razón por la que dicha estimación no es eficiente de entre las que son lineales e insesgadas.
- Obtener una estimación que tenga dicha propiedad.
- Contrastar la hipótesis de que los coeficientes β son los mismos para los dos países.
- Contrastar la homogeneidad global del modelo para los dos países.

Problema 6.3. (Justificación del contraste de igualdad de varianzas submuestrales.)

- Supuesto que se disponga de muestras de tamaños T_1, T_2, \dots, T_m , escribir la función de verosimilitud, tanto en el caso restringido en que se supone que las varianzas de cada submuestra son iguales, como en el caso sin restringir en que se supone que son diferentes.
- Obtener las expresiones para los estimadores de máxima verosimilitud, restringido y sin restringir, tanto para el vector β como para los parámetros σ_i^2 . Interpretarlos.
- Encontrar la expresión del estadístico utilizado en el contraste de razón de verosimilitudes.

d) Probar que dicho estadístico coincide, para muestras grandes, con el estadístico Q de la Sección 6.5. Tener presente que las varianzas que aparecen en el estadístico Q son varianzas de la variable endógena y_t , mientras que las del contraste de razón de verosimilitudes son varianzas de los términos de error.

Problema 6.4. Obtenga expresiones matriciales para el estimador MCG, su matriz de covarianzas, el estimador MCO, su matriz de covarianzas, así como el estimador de σ_u^2 , para el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

donde $E(u_t) = 0$ para todo t y $E(u_t^2) = t\sigma_u^2$, $E(u_t u_s) = 0 \quad \forall t \neq s$.

Problema 6.5. Considere el modelo $y_t = \beta x_t + u_t$, con $\sigma_t^2 = k(\beta x_t)^2$, donde las variables se hallan en diferencias respecto a sus medias muestrales.

a) Pruebe que el estimador de MCG de β es igual al promedio muestral del cociente $\frac{y_t}{x_t}$. Halle su varianza.

b) ¿Qué tipo de problemas surgirían en esta estimación si $x_t = 0$ para algún t ? ¿Qué inferencia obtendríamos de tal resultado?

Problema 6.6. Supongamos que para estimar el modelo

$$y_t = \beta x_t + u_t, \quad E(u_t) = 0, \quad \text{Var}(u_t) = k(\beta x_t)^2$$

donde las variables están en desviaciones respecto a sus medias muestrales, se dispone de las observaciones:

x_t	2	0	-1	1	-2
y_t	-5	4	1	-2	2

- Obtenga el estimador MCO de β y de su varianza.
- Obtenga el estimador de MCG de β y de su varianza. Dé una medida de la eficiencia relativa del estimador MCG frente al estimador MCO.
- Calcule el coeficiente de determinación.
- Estime el parámetro k .

Problema 6.7. Encuentre una expresión analítica general (como función de determinados momentos muestrales) para la matriz de covarianzas de los estimadores MCO y de MCG en el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad E(u_t) = 0, \quad \text{Var}(u_t) = kx_t$$

¿Qué peculiaridad presentarían ambas matrices en el caso en que se tuviese $\sum_1^T x_t = 0$? Interprete esta situación.

Problema 6.8. Interprete que si el término de error del modelo discutido en el Ejemplo 2 de la Sección 6.3 sigue una distribución Normal, el estimador de máxima verosimilitud del vector β y del parámetro σ^2 difiere del estimador allí sugerido. ¿A la

vista de los resultados de la Sección 6.6, cómo puede justificarse tal diferencia? Obtenga la matriz de covarianzas del estimador de máxima verosimilitud.

Problema 6.9. Suponga que el modelo de regresión lineal simple:

$$y_{ig} = \beta x_{ig} + u_{ig}, \quad E(u_{ig}) = 0, \quad \text{Var}(u_{ig}) = \sigma_u^2$$

es válido para cada observación i de cada uno de los cinco subgrupos en que se ha dividido la muestra $g = 1, 2, 3, 4, 5$, con $n_1 + n_2 + n_3 + n_4 + n_5 = 100$.

Suponga, asimismo, que se observan tan sólo las medias muestrales de cada grupo:

Grupo	1	2	3	4	5
n_g	10	20	15	30	25
Media: \bar{y}_g	8	16	12	4	20
Media: \bar{x}_g	2	8	4	2	12

Se proponen tres estimadores de β :

$$\hat{\beta}_1 = \frac{\bar{y}}{\bar{x}} = \frac{\sum n_g \bar{y}_g / \sum n_g}{\sum n_g \bar{x}_g / \sum n_g}$$

$$\hat{\beta}_2 = \frac{\sum \bar{y}_g \bar{x}_g}{\sum \bar{x}_g^2}$$

$$\hat{\beta}_3 = \frac{\sum n_g \bar{x}_g \bar{y}_g}{\sum n_g \bar{x}_g^2}$$

- Pruebe que los tres estimadores son insesgados.
- Obtenga la varianza de cada estimador.
- Calcule numéricamente cada estimador y su varianza, suponiendo que $\sigma_u^2 = 1$. ¿Cuáles son los estimadores más y menos eficientes?
- Demuestre que la varianza del tercer estimador es siempre menor que la del segundo.

Problema 6.10. Proponga un estadístico para el contraste de la hipótesis nula $H_0: \beta_2 = 0$ en el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad \text{Var}(u_t) = \sigma_t^2 = \sigma_u^2 x_{2t}^2$$

Problema 6.11. Suponga que se dispone de un modelo:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \quad i = 1, 2, \dots, N$$

donde $N = N_1 + N_2 + \dots + N_m$, en el que las N_1 primeras observaciones provienen de una submuestra con $\text{Var}(u_1) = \sigma_1^2$, las N_2 observaciones siguientes de una submuestra con $\text{Var}(u_1) = \sigma_2^2$, y así sucesivamente. El vector $\boldsymbol{\beta}$ es común a todas las submuestras. Obtenga el estadístico de razón de verosimilitudes para el contraste de la hipótesis nula $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$.

Problema 6.12. (Heteroscedasticidad multiplicativa.) Derive el estadístico de los multiplicadores de Lagrange para el contraste de la hipótesis nula de ausencia de heteroscedasticidad en el modelo:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T$$

$$E(u_t) = 0, \quad \text{Var}(u_t) = \sigma_u^2 e^{\omega_t \delta}, \quad \omega_t \delta = \delta_0 + \omega_{1t} \delta_1 + \dots + \omega_{pt} \delta_p$$

Problema 6.13. Pruebe que el contraste de los multiplicadores de Lagrange para el contraste de la hipótesis nula de ausencia de heteroscedasticidad en el modelo de heteroscedasticidad multiplicativa se reduce al contraste de Breusch-Pagan.

Problema 6.14. (Formulación genérica del problema de heteroscedasticidad.) Encuentre las condiciones necesarias para la obtención del estimador de máxima verosimilitud de $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ y σ^2 en el modelo de regresión:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

$$E(u_t) = 0, \quad \text{Var}(u_t) = \sigma_u^2 f(\mathbf{z}_t, \boldsymbol{\theta})$$

donde se supone que los vectores $\boldsymbol{\beta}$ y $\boldsymbol{\theta}$ no tienen ningún elemento en común.

Problema 6.15. Si en el modelo de regresión lineal general se hace el supuesto de Prais y Houttaker acerca de la estructura de heteroscedasticidad:

$$\text{Var}(y_t) = \text{Var}(u_t) = \sigma_t^2 = \sigma_u^2 (\mathbf{x}'_t \boldsymbol{\beta})^2 = \sigma_u^2 (E y_t)^2$$

- a) Obtenga la expresión analítica del estimador MCG, y de su varianza.
- b) Suponga que u_t se distribuye $N(0, \sigma_t^2)$, y obtenga el estimador de máxima verosimilitud.

Problema 6.16. Dado el modelo de regresión $y_t = \mu + \varepsilon_t$, donde $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 x_t^2$, con $x_t > 0$:

- a) ¿Cuál es el estimador lineal más eficiente del parámetro μ ? ¿Cuál es su varianza?
- b) ¿Cuál es el estimador MCO de μ y cuál es su varianza?
- c) Demuestre que el estimador en a) es al menos tan eficiente como el estimador MCO.
- d) Suponga ahora que ε_t se distribuye $N(0, \sigma^2(1 + \gamma^2 x_t^2))$. Pruebe que γ^2 y σ^2 pueden estimarse consistentemente mediante una regresión de mínimos cuadrados sobre una constante y x_t^2 . ¿Es eficiente este estimador?

Problema 6.17. Pruebe que el contraste de razón de verosimilitudes de la hipótesis nula de homoscedasticidad, bajo la hipótesis alternativa: $\sigma_t^2 = \exp(\mathbf{z}'_t \boldsymbol{\alpha})$, consiste en comparar el estadístico:

$$\lambda = T \log \hat{\sigma}_u^2 - \sum_1^T \mathbf{z}'_t \boldsymbol{\alpha}$$

donde $\hat{\sigma}_u^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{T}$, con una variable chi-cuadrado con $p - 1$ grados de libertad.

Problema 6.18. Obtenga los estimadores de máxima verosimilitud y la matriz de información de los parámetros β y σ_u^2 del modelo:

$$y_t = \mathbf{x}'_t \beta + u_t, \quad \text{con } E(u_t) = 0, \quad \text{Var}(u_t) = \sigma_t^2 = \sigma_u^2 (\mathbf{x}'_t \beta)^2$$

Problema 6.19. Demuestre que en el modelo $y_t = \beta x_t + u_t$, con $E(u_t^2) = \sigma_t^2$, se tiene que

$$\text{Var}(\hat{\beta}_{\text{MCO}}) - \text{Var}(\hat{\beta}_{\text{MCG}}) = \frac{\sum_1^T x_t^2 \sigma_t^2}{(\sum_1^T x_t^2)^2} - \frac{1}{\sum_1^T \left(\frac{x_t^2}{\sigma_t^2} \right)}$$

y razone que ambos estimadores son igualmente eficientes si σ_t^2 es, en realidad, constante. Por otra parte, si $\sigma_t^2 = \sigma_u^2 x_t^2$, entonces la ganancia en eficiencia del estimador MCG crece con el grado de volatilidad de x_t^2 .

Obtenga la expresión (incorrecta) para la matriz de covarianzas del estimador MCO ($\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$), y la correcta, y demuestre que la diferencia entre ambas crece con la correlación entre σ_t^2 y x_t^2 .

Problema 6.20. Obtenga la función de verosimilitud, las condiciones necesarias de optimización y la matriz de información correspondientes al modelo:

$$y_t = \mathbf{x}'_t \beta + u_t, \quad u_t \sim N(0, \sigma^2 (\gamma' z_t)^2)$$

Problema 6.21. Dado el modelo lineal sin término constante y un solo regresor:

$$y_t = \beta x_t + u_t$$

donde $E(u_t) = 0$, $E(u_t^2) = \sigma_t^2$, y suponiendo que las varianzas cambian de acuerdo con el esquema $\sigma_t^2 = \sigma_u^2 z_t$, donde z_t es una variable conocida, obtenga la expresión analítica del estimador MCG, así como de su varianza. Utilice la desigualdad de Cauchy-Schwartz para comparar dicha varianza con la del estimador MCO. ¿Qué ocurriría si a pesar de la heteroscedasticidad se utilizase $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$ como matriz de covarianzas del estimador MCO?

Nota: La desigualdad de Cauchy-Schwartz garantiza que para dos variables cualesquiera x_t e y_t se tiene:

$$(\sum_1^T x_t y_t)^2 \leq (\sum_1^T x_t^2) (\sum_1^T y_t^2)$$

Problema 6.22. Suponga que en el modelo lineal general $y_t = \mathbf{x}'_t \beta + u_t$ se tiene $E u_t = 0$, $\text{Var}(u_t) = \sigma_t^2 = \sigma_u^2 x_{it}^2$ para un cierto $i = 1, 2, \dots, k$. Encuentre las expresiones para el cálculo del estimador de máxima verosimilitud de β , α y σ_u^2 , así como su matriz de covarianzas.

Problema 6.23. Para analizar la dependencia entre los gastos en alquiler y los ingresos de las familias no propietarias de viviendas se dividió a una muestra de 20 familias de acuerdo con sus niveles de ingresos, que se redondearon a 2, 4, 6 y 8 millones de pesetas anuales. En la muestra había cinco familias de cada uno de estos grupos, que generaron las siguientes observaciones:

Grupo	Gastos de alquiler					Ingresos
I	4,5	4	3,5	4,5	3,5	2
II	6	7	7,5	8	6,5	4
III	8,5	7	9,5	10	10	6
IV	12	14	13	10	11	8

a) Estime el modelo: $\text{Alquiler}_i = \alpha + \beta \cdot \text{Ingresos}_i + u_i$, obteniendo las varianzas de los parámetros estimados.

b) Este modelo no debe estimarse por MCO para cada uno de los subgrupos. ¿Por qué? ¿Cuál sería un estimador razonable de β en los subgrupos? Comparar los valores estimados de σ_u^2 en cada uno de ellos.

c) Suponiendo que dichas varianzas pudiesen ser significativamente diferentes, encuentre, analíticamente, la expresión del estimador eficiente del parámetro β (note que no se pide nada acerca del parámetro α).

d) Utilizando la expresión que se ha encontrado en c), estime eficientemente β , utilizando toda la muestra, y obtenga su varianza. Comparar con la encontrada en a).

e) A la vista de los resultados, ¿qué puede decirse acerca de la importancia de la heteroscedasticidad presente en el modelo estimado en a)?

CAPITULO 7

AUTOCORRELACION

7.1. INTRODUCCION

En el Capítulo 5 examinamos dos situaciones diferentes en que el término de error de un modelo econométrico tenía una matriz de covarianzas con estructura no escalar. En particular, en el modelo examinado en la Sección 5.5, la matriz de covarianzas no es escalar debido a que algunos elementos fuera de la diagonal principal son distintos de cero. Estadísticamente, esta situación proviene del hecho de que el término de error del modelo tiene correlación consigo mismo a través del tiempo. El ejemplo allí discutido fue:

$$\begin{aligned}y_t &= \mathbf{x}'_t \boldsymbol{\beta} + u_t, & t = 1, 2, \dots, T \\u_t &= \rho u_t u_{t-1} + \varepsilon_t\end{aligned}\tag{7.1}$$

Aun siendo sencilla esta especificación para la autocorrelación del término de error, tiene especial interés, puesto que siendo una primera aproximación a situaciones de autocorrelación más complicadas se presenta muy frecuentemente en el análisis práctico. Los objetivos de este capítulo son:

- a) generalizar la discusión de la Sección 5.5 a otros modelos de autocorrelación,
- b) explicar cómo contrastar la hipótesis nula de ausencia de autocorrelación en el término de error de un modelo econométrico, y en caso de rechazar dicha hipótesis:
- c) estimar el modelo econométrico bajo una determinada estructura de autocorrelación.

7.2. NATURALEZA Y CAUSAS DE LA AUTOCORRELACION

Existe autocorrelación cuando el término de error de un modelo econométrico está correlacionado consigo mismo a través del tiempo, es decir: $E(u_t u_s) \neq 0$. Por supuesto, no es preciso que u_t esté correlacionado consigo mismo en cada dos instantes de tiempo diferentes, sino que basta que la correlación se extienda a algunos periodos. Por ejemplo, podría ser que $E(u_t u_{t-1}) \neq 0$ para todo t , pero que $E(u_t u_{t-s}) = 0$ para todo $s \geq 2$.

El lector no debe dejarse llevar por la impresión de que un modelo como [7.1] genera tan sólo correlación entre dos valores *consecutivos* de u_t . Al contrario, como vimos en la matriz de covarianzas [5.15], la especificación [7.1] genera correlación entre dos valores cualesquiera de u_t , con independencia del número de periodos intercalados entre ambos. En el Capítulo 13 veremos modelos analíticos de autocorrelación que, por el contrario, generan correlación durante un corto número de periodos.

La autocorrelación puede venir producida por diversas causas:

Existencia de ciclos y tendencias. Si la autocorrelación es positiva (en el caso del modelo [7.1], si el coeficiente ρ es positivo), entonces un valor alto de u_t , que genera un valor de y_t por encima de su promedio, tendrá una probabilidad elevada de ir seguido de un valor alto de u_{t+1} , y por ello, de un valor de y_{t+1} por encima de su promedio; lo mismo ocurriría para valores de y_t por debajo de su promedio. En el caso de la autocorrelación negativa, valores de y_t por encima de su promedio irán seguidos, con mucha probabilidad, de valores de y_t por debajo de su promedio. Por tanto, la autocorrelación positiva está asociada a la existencia de rachas de valores altos y bajos de y_t . Si, como ocurre debido a la inercia de la mayoría de las variables macroeconómicas, la variable endógena del modelo econométrico presenta ciclos y éstos no son bien explicados por las variables exógenas del modelo, el término de error tendrá autocorrelación (Fig. 7.1.a).

Igualmente, es cierto que la mayoría de las variables económicas (y especialmente las variables medidas en términos nominales) tienen una tendencia, generalmente creciente. Si el conjunto de variables explicativas del modelo no explican adecuadamente dicho comportamiento, entonces el término de error incorporará dicha tendencia, y ello conduce a la existencia de autocorrelación positiva: Una primera racha de residuos negativos seguida por otra racha de residuos positivos (Fig. 7.1.b).

Variables omitidas. Si el verdadero modelo que explica el comportamiento de la variable endógena es

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t, \quad t = 1, 2, \dots, T$$

pero, en cambio, se especifica el modelo $y_t = \beta_1 + \beta_2 x_{2t} + v_t$, entonces el término de error v_t es igual a $u_t + \beta_3 x_{3t}$. Si la variable x_{3t} , por las razones antes expuestas, presenta autocorrelación, entonces el término de error v_t también estará autocorrelacionado, incluso si u_t estaba libre de tal problema.

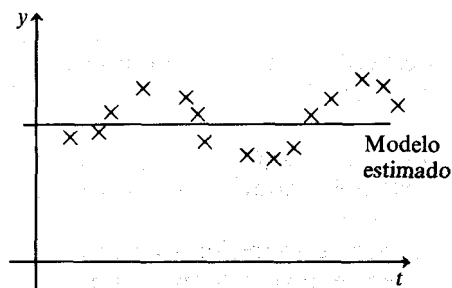


FIGURA 7.1.a. Autocorrelación producida por un ciclo.

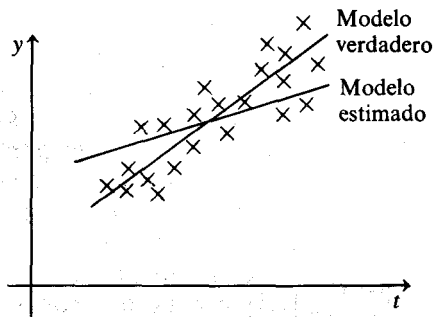


FIGURA 7.1.b. Autocorrelación producida por una tendencia.

Si el investigador sospechara que la presencia de autocorrelación en el modelo econométrico se debe a la existencia de variables relevantes omitidas, debería hacer un esfuerzo por identificarlas, pues la estimación por mínimos cuadrados de un modelo en tales condiciones está sujeta a dificultades adicionales a la mera existencia de autocorrelación (como se vio en la Sección 3.11).

Relaciones no lineales. Supongamos que la verdadera relación entre los tipos de interés y el stock de Deuda Pública D_t , es:

$$r_t = \beta_1 + \beta_2 D_t + \beta_3 D_t^2 + u_t, \quad \beta_2 > 0, \quad \beta_3 < 0, \quad t = 1, 2, \dots, T$$

Este modelo implica que los tipos de interés aumentan al crecer el stock de Deuda, aunque menos que proporcionalmente, puesto que se tiene:

$$\frac{\partial r_t}{\partial D_t} = \beta_2 + 2\beta_3 D_t < \beta_2$$

tanto menor cuanto mayor es D_t . Sin embargo, si se especifica el modelo lineal

$$r_t = \delta_1 + \delta_2 D_t + v_t, \quad t = 1, 2, \dots, T$$

entonces la Figura 7.2 muestra que se tendrá una racha de residuos negativos, seguida de otra racha de residuos positivos, para terminar con otra racha negativa. Ello generará autocorrelación del término de error.

Relaciones dinámicas. La mayoría de las relaciones entre variables económicas se extienden a más de un período, especialmente si el período de observación de los datos es mensual o trimestral. En tal caso, se tienen relaciones entre inflación y crecimiento de la oferta monetaria, del tipo:

$$\pi_t = \beta_1 + \beta_2 m_t + \beta_3 \pi_{t-1} + u_t$$

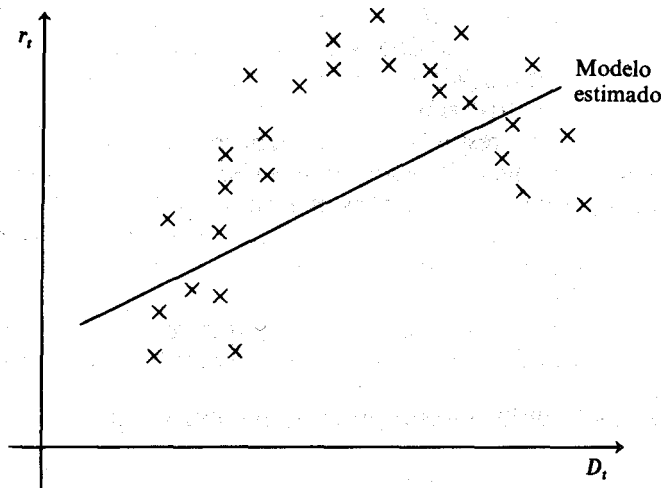


FIGURA 7.2. Autocorrelación producida por una relación no lineal.

La omisión del retardo de la variable endógena haría que el término de error del modelo incorporara dicha variable, mostrando ciertas propiedades sistemáticas, que no serían sino reflejo de la influencia de π_{t-1} sobre π_t .

7.3. CONSECUENCIAS DE LA AUTOCORRELACION

Debemos comenzar recordando del Capítulo 5 que el estimador de mínimos cuadrados ordinarios, aun siendo insesgado en un modelo con autocorrelación en el que las variables explicativas sean deterministas, ya no es el estimador lineal insesgado de mínima varianza, sino que dicha propiedad corresponde a otro estimador que es también lineal e insesgado, el estimador de mínimos cuadrados generalizados.

Este último puede obtenerse de dos maneras alternativas, aunque equivalentes: Una de las alternativas consiste en llevar a cabo una transformación del modelo, premultiplicando las matrices de observaciones de las variables (tanto endógena como exógena). El factor por el que se premultiplican es la matriz V^{-1} , donde V descompone la matriz de covarianzas del modelo original $\Sigma = VV'$. La segunda alternativa consiste en resolver el correspondiente sistema de ecuaciones normales o, lo que es lo mismo, efectuar el producto matricial:

$$\hat{\beta}_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

Si, a pesar de ello, se calcula el estimador de mínimos cuadrados ordinarios, entonces es muy importante recordar que su matriz de covarianzas no es $\sigma_u^2(X'X)^{-1}$, sino $\sigma_u^2(X'X)^{-1}X'\Sigma X(X'X)^{-1}$, y también que la primera es una estimación *sesgada* de la segunda, sin que exista necesariamente ninguna

ordenación entre ambas en términos de sus tamaños. Es decir, la varianza de cada parámetro podría ser inferior o superior, si se utilizase la expresión (incorrecta) $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ para la matriz de covarianzas de $\hat{\beta}_{\text{MCO}}$ que si se utilizase la expresión correcta de dicha matriz de covarianzas. Incluso si se utiliza la expresión correcta, debe recordarse que el estimador de mínimos cuadrados generalizados es más eficiente (tiene una matriz de covarianzas menor) que el de mínimos cuadrados ordinarios.

En ocasiones, el investigador no tiene confianza en el modelo de autocorrelación que siguen los residuos y teme que una mala especificación produzca resultados sesgados. Por ello, puede tener interés en obtener el estimador MCO, al menos a efectos comparativos. Recientemente, en la línea de la sugerencia de White para el caso de heteroscedasticidad, Newey y West (1987)

han propuesto utilizar como estimador de $\frac{1}{T} \mathbf{X}'\Sigma\mathbf{X}$ la matriz:

$$\mathbf{S} = \frac{1}{T} \sum_{j=0}^L \sum_{t=j+1}^T w_j \hat{u}_t \hat{u}_{t-j} [\mathbf{x}_t \mathbf{x}'_{t-j} + \mathbf{x}_{t-j} \mathbf{x}'_t]$$

donde $w_j = 1 - \frac{j}{L+1}$, siendo L el orden máximo de la autocorrelación del término de error, que no siempre será fácil de determinar.

En el primer ejercicio de simulación de este capítulo (en la Sección 7.7) se hace referencia a la eficiencia relativa del estimador MCG frente al estimador de mínimos cuadrados ordinarios. Como allí puede verse, incluso en modelos econométricos sencillos, si existe autocorrelación de primer orden con un coeficiente elevado ($\rho = 0,90$), la varianza del estimador de mínimos cuadrados generalizados puede llegar a ser tan sólo un 10 por 100 de la varianza del estimador MCO. Se muestra asimismo cómo la utilización errónea de la matriz $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ como matriz de covarianzas del estimador de mínimos cuadrados ordinarios puede generar sesgos realmente importantes.

7.4. CONTRASTES DE AUTOCORRELACION

7.4.a. El contraste de Durbin-Watson

Si se sospecha que el término de error del modelo econométrico tiene autocorrelación de primer orden: $u_t = \rho u_{t-1} + \varepsilon_t$, donde ε_t no tiene autocorrelación, entonces el estadístico de Durbin-Watson permite contrastar la hipótesis nula de ausencia de autocorrelación, frente a la hipótesis alternativa que hemos mencionado. Dicho estadístico viene definido por la expresión:

$$d = \frac{\sum_2^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_2^T \hat{u}_t^2}$$

La interpretación del estadístico es la siguiente: La autocorrelación de primer orden con coeficiente ρ positivo hace que valores positivos del término

de error u_t tiendan a venir seguidos de valores positivos, y valores negativos tiendan asimismo a venir seguidos de valores negativos. La razón es que, excepto por el componente puramente aleatorio ε_t , el término de error u_t es igual a ρu_{t-1} . Aunque el término de error del modelo es una variable aleatoria no observable, podemos disponer de los residuos generados por una estimación MCO, y si $\hat{\beta}_{\text{MCO}}$ es insesgado, \hat{u}_t será un estimador insesgado (aunque ineficiente) de u_t .

Por todo ello, con autocorrelación de primer orden con coeficiente positivo, se observarán rachas de valores negativos y positivos de los residuos. En tales condiciones, la diferencia $\hat{u}_t - \hat{u}_{t-1}$ será generalmente menor, en valor absoluto, que el valor de u_t . Como consecuencia $(\hat{u}_t - \hat{u}_{t-1})^2 < \hat{u}_t^2$ y, en consecuencia, el numerador de la fracción d será «pequeño» en relación con el denominador. Por ejemplo, si ρ es aproximadamente igual a 1, entonces \hat{u}_t será aproximadamente igual a \hat{u}_{t-1} y, por tanto, el numerador en d tenderá a ser muy próximo a cero.

Si, por el contrario, el coeficiente de autocorrelación fuese negativo, entonces el efecto sería el de tener valores positivos de \hat{u}_t seguidos de valores negativos, y viceversa. La implicación numérica es que el valor absoluto de las diferencias $\hat{u}_t - \hat{u}_{t-1}$ tenderá a ser mayor que \hat{u}_t . Lo mismo tenderá a ocurrir con sus cuadrados y, como consecuencia, el estadístico d tenderá a tomar valores «grandes». De hecho, el estadístico d está más relacionado con el valor del parámetro ρ de lo que hasta ahora hemos mostrado. Desarrollando el numerador del estadístico d de Durbin-Watson:

$$d = \frac{\sum_2^T \hat{u}_t^2 + \sum_2^T \hat{u}_{t-1}^2 - 2\sum_2^T \hat{u}_t \hat{u}_{t-1}}{\sum_2^T \hat{u}_t^2}$$

Si el número de observaciones es suficientemente grande, entonces $\sum_2^T \hat{u}_t^2 \approx \sum_2^T \hat{u}_{t-1}^2$, y el estadístico d puede aproximarse por $d = 2(1 - \hat{\rho})$, donde $\hat{\rho} = \frac{\sum_2^T \hat{u}_t \hat{u}_{t-1}}{\sum_2^T \hat{u}_t^2}$. Dado que, como veremos en el Capítulo 13, los valores admisibles del parámetro ρ están entre $-1,0$ y $1,0$, entonces el estadístico d oscilará entre 0 y 4, con valores próximos a 0 cuando exista autocorrelación positiva de primer orden, y valores cercanos a 4 cuando exista autocorrelación de primer orden con coeficiente negativo. Cuando el término de error es independiente a lo largo del tiempo, $\hat{\rho}$ será casi nulo, y el valor del estadístico d estará próximo a 2. Sobre el estadístico de Durbin-Watson cabe hacer las siguientes observaciones:

1. El vector de residuos $\hat{\mathbf{u}}$ es igual al producto $\mathbf{M}\mathbf{u}$ (Proposición 3.9), por lo que su distribución (en particular, su matriz de covarianzas) depende de la matriz \mathbf{M} , quien a su vez depende de la matriz de observaciones \mathbf{X} . Como consecuencia, la distribución de d también cambia con cada matriz \mathbf{X} y, por ello, con cada aplicación empírica, lo que hace imposible la construcción de tablas de valores numéricos para dicho estadístico.

Durbin y Watson obtuvieron cotas superior e inferior para los niveles de significación del estadístico d , sobre el conjunto de todas sus posibles distribu-

ciones de probabilidad. Dichos valores críticos se presentan en las Tablas A.7 y A.8, al final de este texto. Tales cotas funcionan del siguiente modo. Supongamos que se pretende contrastar la hipótesis nula de ausencia de autocorrelación, frente a la alternativa de *autocorrelación positiva de primer orden*, y obtenemos un valor del estadístico d por debajo de la cota inferior d_i . Entonces podemos asegurar que, con independencia de los valores de la matriz X , puede rechazarse la hipótesis nula de no autocorrelación, en favor de la alternativa de autocorrelación de primer orden con coeficiente positivo. Además, la estimación del coeficiente del modelo de autocorrelación $\hat{\rho}$ podría obtenerse de la aproximación $d = 2(1 - \hat{\rho})$.

Si, por el contrario, el valor del estadístico d en nuestra aplicación supera la cota superior d_s , entonces concluiremos que con independencia de la distribución de los residuos MCO no puede rechazarse la hipótesis nula de ausencia de autocorrelación de primer orden. Finalmente, si el valor del estadístico d está entre d_i y d_s , no se toma ninguna decisión respecto al contraste de la hipótesis de ausencia de autocorrelación.

2. Si la alternativa fuese de autocorrelación de primer orden pero *coeficiente negativo*, entonces se lleva a cabo el proceso que hemos descrito pero para el estadístico $4 - d$. Así, valores de $4 - d$ próximos a cero (por debajo de d_i) implicarían valores de d próximos a 4 y, por tanto, autocorrelación negativa de primer orden con coeficiente ρ cercano a -1 .

3. Un problema importante en la puesta en práctica de este contraste se deduce inmediatamente de los comentarios anteriores. Estamos sugiriendo utilizar en el contraste cotas superiores e inferiores, obtenidas de entre *todas las distribuciones posibles* del vector de residuos MCO. Exigir que el valor del estadístico d esté por debajo de la cota inferior, o supere la cota superior, es, en general, un requisito demasiado estricto. Dicho de otra forma, lo que ocurre es que existe un gran número de casos prácticos en los que la situación es de indefinición, por encontrarse el valor calculado para el estadístico d entre los valores tabulados d_i y d_s . En tales casos no puede obtenerse ninguna conclusión con respecto al contraste que se pretendía llevar a cabo⁽¹⁾.

4. Las cotas d_i y d_s obtenidas por Durbin y Watson suponen que hay un término constante incluido en la regresión, y no son estrictamente válidas en otro caso. Lo que es más importante, dichas cotas están obtenidas sobre el supuesto de que todas las variables incluidas como explicativas son deterministas. Un caso frecuente en que este supuesto no puede mantenerse, es aquel en que se utilizan como explicativas algunos valores retardados de la variable endógena y_t , como en los modelos que veremos en el Capítulo 13.

En tal situación, el procedimiento que hemos sugerido produce estimacio-

⁽¹⁾ Lamentablemente, se observa con excesiva frecuencia que los investigadores utilizan el gran tamaño de esta región intermedia de indefinición para afirmar que en su aplicación particular «no existe evidencia clara en contra de la hipótesis nula de ausencia de autocorrelación», y en consecuencia, mantener dicha hipótesis nula. Debe quedar claro de nuestra discusión que si los numerosos casos de indefinición del contraste pudieran discernirse, podrían decantarse tanto en un sentido como en otro, por lo que la interpretación más correcta de dicho fenómeno es que «no hay evidencia clara para mantener la hipótesis nula» de ausencia de autocorrelación, lo que debe crear al menos cierta preocupación acerca del mantenimiento de dicha hipótesis.

nes sesgadas del verdadero valor del parámetro ρ . Lo que es aún peor, la estimación de $\hat{\rho}$ estará sesgada en el sentido de disminuir su valor absoluto, lo que facilitará la aceptación de la hipótesis nula de no autocorrelación, incluso en casos en que una estimación insesgada de ρ hubiese conducido al rechazo de dicha hipótesis.

5. Para aminorar este problema en el caso en que las *únicas variables estocásticas* incluidas como explicativas son los retardos de la variable endógena (posiblemente varios), Durbin sugirió utilizar el siguiente estadístico:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T\text{Var}(\hat{\beta}_2)}}$$

donde $\text{Var}(\hat{\beta}_2)$ denota la varianza de la estimación MCO del coeficiente del primer retardo de y_t y $\hat{\rho}$ es la estimación habitual del parámetro ρ , mediante los residuos de dicha regresión. Cabe observar que el *único coeficiente* que entra en este contraste es el del primer retardo de la variable endógena, con independencia de que hubiera retardos adicionales incluidos como variables explicativas.

Si no hay autocorrelación, el estadístico h tiene una distribución que se aproxima a $N(0, 1)$ cuando el tamaño muestral tiende a infinito. Como la hipótesis alternativa es, o bien que exista autocorrelación positiva, o bien que exista autocorrelación negativa de primer orden, el contraste de la hipótesis nula debe ser, lógicamente, contraste de una sola cola, por lo que debe compararse el valor empírico del estadístico h con 1,645, el valor de la $N(0, 1)$ al 95 por 100 para tal contraste.

Todavía podría aparecer un problema adicional si se tuviese que la varianza estimada $\text{Var}(\hat{\beta}_2)$ fuese superior a T^{-1} , pues el radicando sería en tal caso negativo. Durbin probó que cuando el tamaño muestral tiende a infinito, un modo equivalente de llevar a cabo este contraste es estimando una regresión de los residuos MCO de la regresión inicial sobre un retardo de los mismos, todos los retardos de la variable endógena incluidos en el modelo, y las demás variables explicativas, y rechazar la hipótesis nula si el coeficiente de \hat{u}_{t-1} resulta significativamente diferente de cero (de acuerdo con su t de Student).

6. No debe olvidarse que los contrastes que hasta ahora hemos discutido tienen como hipótesis alternativa la de autocorrelación de primer orden, y no otra, por lo que en el caso de que la verdadera estructura de autocorrelación fuese diferente, entonces el resultado del test podría ser arbitrario. Sin embargo, cabe hacer las siguientes observaciones: Aunque la autocorrelación fuese de orden 2 o superior, también la habrá, generalmente, de orden 1, y el contraste detectará la presencia de autocorrelación, y tenderemos a rechazar la hipótesis nula de no autocorrelación. Sin embargo, pudiera ser que la autocorrelación fuese de orden 4, por ejemplo, sin que exista correlación de orden 1, 2 ó 3. Ello podría deberse a la estacionalidad, trabajando con datos trimestrales. En tal caso, los contrastes anteriores bien podrían sugerir que no existe autocorrelación.

Para el caso concreto de autocorrelación de orden 4 (únicamente), Wallis sugirió utilizar un estadístico d_4 construido de manera análoga al estadístico de Durbin-Watson con la sola diferencia de utilizar u_{t-4} en vez de u_{t-1} en dicha definición. Wallis computó los valores críticos precisos para llevar a cabo este contraste, dependiendo de si el modelo incluye o no un término constante. Dichos valores aparecen en la Tabla A.9, al final de este texto.

7.4.b. El contraste de Breusch y Godfrey

Otra alternativa es llevar a cabo contrastes de autocorrelación en los que la hipótesis alternativa incluya especificaciones más generales que la del modelo autorregresivo de primer orden considerado en [7.1]. Generalizando nuestra exposición anterior, consideremos, para distintos valores de k , el conjunto de estadísticos:

$$r_k = \frac{\sum_1^T u_t u_{t-k}}{\sum_1^T u_t^2}$$

El primero de ellos, r_1 , que sería la estimación MCO del parámetro ρ en [7.1], ha sido ya utilizado en la construcción del estadístico Durbin-Watson. Los demás estadísticos r_k serán o no significativos, dependiendo de la estructura de autocorrelación de u_t . Dicha estructura pudiera ser más compleja que la mostrada en [7.1], como es el caso de un modelo autorregresivo de orden p :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$$

u otras que analizaremos en el Capítulo 13. En definitiva, es importante disponer de contrastes de autocorrelación que consideren este tipo de alternativas. El más conocido entre éstos es el contraste del tipo de multiplicadores de Lagrange introducido simultáneamente por Breusch (1978) y Godfrey (1978), quienes sugieren:

a) Estimar el modelo de regresión por MCO y obtener los residuos \hat{u}_t . El modelo puede incluir como explicativas algunos retardos de la variable endógena, sin que por ello cambien las propiedades de este contraste.

b) Estimar una regresión de los residuos \hat{u}_t sobre p retardos: $\hat{u}_{t-1}, \dots, \hat{u}_{t-p}$, así como sobre las variables explicativas x_t del modelo original. El número de retardos de \hat{u}_t en esta regresión auxiliar debe coincidir con el número de estadísticos r_k cuya significación conjunta se pretende contrastar. Obtener el valor del R^2 de esta regresión. (Nótese que habrá que excluir p observaciones de esta regresión.)

c) Comparar TR^2 con las tablas de una distribución chi-cuadrado con p grados de libertad y rechazar la hipótesis nula de no autocorrelación si TR^2 es superior al valor de las tablas.

Si la hipótesis nula de ausencia de autocorrelación es cierta, entonces el coeficiente de determinación de la regresión auxiliar que hemos descrito debe-

ría tender a cero. En dicha regresión hay dos clases de variables explicativas: los retardos de los residuos MCO y las variables explicativas de la regresión original, mientras que la variable dependiente son los residuos MCO. Sin autocorrelación, estos residuos no vendrán explicados por sus retardos. Además, ya sabemos que dichos residuos son ortogonales a las variables explicativas de la regresión en que se obtienen. Como consecuencia, el coeficiente de determinación de dicha regresión debe ser muy pequeño.

El contraste se basa sobre el hecho de que, según aumenta el tamaño muestral, dicho coeficiente debería acercarse a cero incluso más rápidamente que el tamaño muestral tiende a infinito. Una vez más, la validez estricta de este contraste se reduce al caso en que el tamaño muestral es infinito.

Cuando el vector de variables explicativas contiene únicamente variables exógenas, este contraste es equivalente a la comparación del estadístico $Q = T \sum_1^p r_i^2$ con una variable chi-cuadrado con p grados de libertad. Este contraste fue propuesto por Box y Pierce. Un refinamiento de éste, propuesto por Ljung y Box (1979), consiste en utilizar $Q' = T(T+2) \sum_{i=1}^p \frac{r_i^2}{T-i}$ con la misma distribución que Q , si bien la dificultad con estos contrastes estriba en la elección del orden p .

7.4.c. Contrastes gráficos

Ningún contraste de autocorrelación debe excluir un examen riguroso de los residuos generados en la estimación del modelo. Dicho examen debe incluir el gráfico de la serie de residuos, en el que habrá que buscar rachas de residuos de igual signo, que serían un claro indicio de autocorrelación.

Otro procedimiento que resulta muy ilustrativo, es el examen del gráfico de \hat{u}_t contra \hat{u}_{t-1} . Si la mayoría de los puntos en dicho gráfico se hallan en el primer y tercer cuadrantes, ello es un indicio de autocorrelación positiva, mientras que si se hallan en el segundo y cuarto cuadrantes, será un indicio de autocorrelación negativa. La limitación de este procedimiento es que sólo es realmente útil si la autocorrelación puede aproximarse por un modelo autorregresivo de orden 1.

Como hemos discutido antes, un contraste de autocorrelación puede basarse en las rachas de residuos de igual signo, puesto que si hay pocas rachas, ello quiere decir que los residuos apenas cambian de signo, lo que puede indicar correlación positiva. Si hay muchas rachas, se tiene evidencia de autocorrelación negativa. Si denotamos por T el número total de observaciones, por T_1 y T_2 el número de residuos positivos y negativos respectivamente y por n el número de rachas de residuos de igual signo, se tiene:

$$E(n) = 1 + \frac{2T_1 T_2}{T_1 + T_2}$$

$$\text{Var}(n) = \frac{2T_1 T_2 (2T_1 T_2 - T_1 - T_2)}{(T_1 + T_2)^2 (T_1 + T_2 - 1)}$$

Si el tamaño muestral es suficientemente grande, entonces la variable n puede aproximarse por una distribución Normal, por lo que con un 95 por 100 de confianza, sus valores deberían estar en el intervalo $E(n) \pm 1,96\sqrt{\text{Var}(n)}$. En caso contrario, debería rechazarse la hipótesis nula de ausencia de autocorrelación, siendo ésta positiva si n está a la izquierda del intervalo y negativa si está a la derecha.

Por último, puede diseñarse un contraste de autocorrelación mediante una tabla de contingencia, como las usualmente utilizadas en Estadística para contrastar la independencia entre sucesos. Bajo la hipótesis nula de ausencia de autocorrelación de primer orden, los sucesos independientes serían los signos (positivo o negativo) de los residuos del modelo en dos instantes consecutivos de tiempo, $t - 1$ y t . Así, puede diseñarse la tabla:

	Número de residuos positivos en t	Número de residuos negativos en t	
Número de residuos positivos en $t - 1$	T_{11}	T_{12}	T_1
Número de residuos negativos en $t - 1$	T_{21}	T_{22}	T_2
	T_1	T_2	$T - 1$

donde $T - 1$ denota el tamaño muestral útil, mientras que T_{21} , por ejemplo, denota el número de residuos negativos que están seguidos de residuos positivos en el siguiente período. Si el término de error del modelo estuviera libre de autocorrelación, entonces debería ser igualmente probable encontrar residuos positivos y negativos un período después de haber observado un residuo positivo (o después de un residuo negativo). Dicho de otro modo, se debería tener las siguientes igualdades entre los tamaños submuestraes:

$$T_{11} = \frac{T_1^2}{T - 1}; \quad T_{12} = \frac{T_1 T_2}{T - 1}; \quad T_{21} = \frac{T_2 T_1}{T - 1}; \quad T_{22} = \frac{T_2^2}{T - 1}$$

Denotemos ahora por N_{ij} , $i, j = 1, 2$ los valores numéricos de los cuatro cocientes que acabamos de describir, calculados únicamente a partir de T_1 , T_2 y T . Los valores N_{ij} son las frecuencias con las que *teóricamente* debiera haberse observado pares de residuos consecutivos con una determinada combinación de signos. Si el tamaño muestral es suficientemente grande, entonces el estadístico

$$\frac{(T_{11} - N_{11})^2}{N_{11}} + \frac{(T_{12} - N_{12})^2}{N_{12}} + \frac{(T_{21} - N_{21})^2}{N_{21}} + \frac{(T_{22} - N_{22})^2}{N_{22}}$$

sigue una distribución chi-cuadrado con un grado de libertad.

7.4.d. Las funciones de autocorrelación

Quizá las herramientas más útiles para el contraste de ausencia de autocorrelación son las funciones de autocorrelación simple y parcial de los residuos, que introduciremos en el Capítulo 13. El estadístico de Ljung y Box que antes vimos no es sino una síntesis de dichas funciones.

El examen gráfico de la serie de residuos, junto con el contraste de rachas de la sección anterior, constituyen un excelente complemento de las funciones de autocorrelación residual para la contrastación de esta hipótesis.

7.5. ESTIMACION DE MODELOS CON AUTOCORRELACION

7.5.a. Estimación MCG (Procedimiento de Cochrane-Orcutt)

Como sabemos, un modelo en el que se acepta que el término de error presenta autocorrelación, debe estimarse por MCG si se pretende obtener su estimador lineal e insesgado de mínima varianza. En la Sección 5.5 discutimos en detalle la aplicación de tal método de estimación a un modelo bajo el supuesto de que la estructura de autocorrelación del término de error es de primer orden. Como es habitual con el método MCG, había en aquel caso dos formas de llevarlo a la práctica: O bien se efectúa la operación matricial habitual, que resulta algo compleja puesto que envuelve la inversión de una matriz $T \times T$, o bien se descompone dicha matriz de covarianzas, $\Sigma = \mathbf{V}\mathbf{V}'$, se utiliza el factor \mathbf{V} para transformar las matrices de observaciones de las variables del modelo y se estima el modelo transformado mediante MCO.

Vimos en dicha sección que, en el modelo con autocorrelación autorregresiva de primer orden, el procedimiento consistía en tomar «cuasidiferencias» de las variables endógenas y exógenas $y_t^* = y_t - \hat{\rho}y_{t-1}$, $x_t^* = x_t - \hat{\rho}x_{t-1}$, utilizando el parámetro $\hat{\rho}$, estimado por medio de los residuos de una regresión inicial MCO. La primera observación requería un tratamiento diferente, aunque bien es cierto que si el tamaño muestral es suficientemente grande, entonces dicha observación puede olvidarse sin una gran pérdida de eficiencia o precisión en las estimaciones de los parámetros del modelo.

Veámos también que el procedimiento de transformación que acabamos de mencionar tiene una interpretación bastante intuitiva, puesto que si en el modelo

$$\begin{aligned} y_t &= \alpha + \mathbf{x}_t' \boldsymbol{\beta} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

se toman estas cuasidiferencias, se obtiene el modelo:

$$y_t - \rho y_{t-1} = \alpha(1 - \rho) + \mathbf{x}_t' \boldsymbol{\beta} - \rho \mathbf{x}_{t-1}' \boldsymbol{\beta} + (u_t - \rho u_{t-1}) \quad [7.2]$$

o, lo que es lo mismo:

$$y_t^* = \alpha(1 - \rho) + \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$$

en el que el término de error es independiente en el tiempo, por lo que puede estimarse por MCO sin mayor problema. El modelo transformado puede también escribirse:

$$\begin{aligned} y_t &= \alpha(1 - \rho) + \rho y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta} - \rho \mathbf{x}_{t-1}' \boldsymbol{\beta} + \varepsilon_t = \\ &= \delta_1 + \delta_2 y_{t-1} + \mathbf{x}_t' \boldsymbol{\delta}_3 + \mathbf{x}_{t-1}' \boldsymbol{\delta}_4 + \varepsilon_t \end{aligned} \quad [7.3]$$

por lo que un procedimiento alternativo consistiría en estimar directamente esta última ecuación. El problema que esta última sugerencia presenta, sin embargo, es que en ella aparecen menos coeficientes a estimar que parámetros, por lo que habría todo un continuo de posibilidades de recuperación de los parámetros iniciales. Ello se debe a que los coeficientes δ_1 , δ_2 y los vectores $\boldsymbol{\delta}_3$, $\boldsymbol{\delta}_4$ del modelo estimado no son independientes, sino que, por el contrario, satisfacen una restricción no lineal: $\boldsymbol{\delta}_4 = -\delta_2 \boldsymbol{\delta}_3$, y no imponerla genera las infinitas posibilidades a que hemos hecho referencia. Si, por el contrario, tomamos en cuenta dicha restricción, entonces podemos seleccionar de entre el continuo de posibilidades aquella que satisface la restricción mencionada.

En cualquier caso, la ecuación transformada puede escribirse:

$$y_t - \rho y_{t-1} = \alpha(1 - \rho) + (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \boldsymbol{\beta} + \varepsilon_t$$

o bien:

$$(y_t - \alpha - \mathbf{x}_t' \boldsymbol{\beta}) = \rho(y_{t-1} - \alpha - \mathbf{x}_{t-1}' \boldsymbol{\beta}) + \varepsilon_t$$

lo que sugiere llevar a cabo el siguiente *procedimiento iterativo*:

1. Comenzar con la estimación MCO del modelo de regresión, ignorando la presencia (conocida) de autocorrelación de primer orden en el término de error.

2. Utilizar los residuos MCO para estimar el parámetro ρ . Esto puede hacerse mediante una regresión de \hat{u}_t sobre \hat{u}_{t-1} , o a partir del estadístico Durbin-Watson de la estimación MCO inicial.

3. Utilizar dicha estimación para obtener las variables cuasidiferenciadas y_t^* y \mathbf{x}_t^* .

4. Estimar por MCO un modelo con las variables transformadas, para obtener estimaciones del vector de coeficientes $\boldsymbol{\beta}$.

5. Utilizar la estimación del vector $\boldsymbol{\beta}$ para generar una nueva serie de residuos. Utilizar estos residuos para estimar de nuevo el parámetro ρ .

Estas iteraciones deben continuar hasta alcanzar un grado de convergencia fijado de antemano.

Obsérvese que el término independiente es ahora $\alpha(1 - \rho)$, no α , y que habrá que ajustar la estimación numérica obtenida para recuperar de ella la estimación de α .

Este procedimiento de estimación puede generalizarse sin ninguna dificultad al caso de más de una variable explicativa y de autocorrelación de orden superior a uno. Si, por ejemplo, se tiene el modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

entonces las dos ecuaciones a utilizar serían

$$y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} = \beta_1(1 - \rho_1 - \rho_2) + \beta_2(x_{2t} - \rho_1 x_{2,t-1} - \rho_2 x_{2,t-2}) +$$

$$+ \dots + \beta_k(x_{kt} - \rho_1 x_{k,t-1} - \rho_2 x_{k,t-2}) + \varepsilon_t$$

y

$$\hat{u}_t = \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t$$

En la literatura se han propuesto diversos modos de comenzar este algoritmo, alternativos a la estimación MCO. Por ejemplo, Durbin sugirió comenzar de la regresión [7.3] olvidando la restricción no lineal, y utilizando como valor inicial de ρ el coeficiente estimado de y_{t-1} . Todos estos procedimientos son equivalentes cuando el tamaño muestral es suficientemente grande.

La suma de los cuadrados de los residuos puede ser una función complicada de los parámetros del modelo, por lo que incluso una vez que se ha obtenido la convergencia del algoritmo iterativo cabe todavía la duda de si la convergencia se habrá producido a un mínimo local o global de dicha función objetivo. En el modelo autorregresivo de primer orden conviene ayudarse de una «red de búsqueda», que consiste en construir una red o partición del espacio paramétrico $(-1, 1)$, y evaluar la función objetivo (suma de cuadrados de los residuos) en los nudos de dicha red, para finalmente escoger como estimación aquel valor de ρ para el que el valor de la función objetivo es menor.

Este procedimiento puede también utilizarse como un método de estimación en sí mismo. Así, una vez seleccionado el valor de ρ que genera la menor suma residual, se podría construir a su alrededor otra red de búsqueda, para elegir ahora el valor de ρ que genere la menor suma residual de entre el nuevo conjunto de puntos, y así hasta conseguir la precisión deseada. Hay que tener en cuenta que cuanto más fina sea la red construida, más probable será que no se confunda un mínimo local con uno global. También puede utilizarse la primera evaluación de este método de búsqueda como una primera etapa del estimador MCG, que hemos descrito anteriormente, en vez de empezar de la estimación de ρ que generan los residuos mínimo-cuadráticos.

7.5.b. Estimación de máxima verosimilitud del modelo autorregresivo de primer orden

Supongamos que se pretende estimar el modelo de regresión con autocorrelación de primer orden y que se está dispuesto a suponer que el término de

error fundamental, es decir, aquel que genera u_t y, por tanto, el único ruido blanco del modelo, tiene distribución Normal: $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. La función de verosimilitud es:

$$L = \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^T \exp \left[\frac{-\sum_1^T \varepsilon_t^2}{2\sigma_\varepsilon^2} \right]$$

Si se hace el cambio de variable ε_t a la variable u_t , hay que tener en cuenta que la matriz de dicha transformación es la matriz \mathbf{V}^{-1} de la Sección 5.5, cuyo determinante es $\sqrt{1 - \rho^2}$, por lo que se tiene:

$$\begin{aligned} L &= \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^T \sqrt{1 - \rho^2} \exp \left[\frac{-(1 - \rho^2)u_1^2 - \sum_2^T (u_t - \rho u_{t-1})^2}{2\sigma_\varepsilon^2} \right] = \\ &= \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^T \sqrt{1 - \rho^2} \exp \left[\frac{\mathbf{u}' \boldsymbol{\Sigma}^{-1} \mathbf{u}}{2\sigma_\varepsilon^2} \right] \end{aligned}$$

y su logaritmo:

$$\log L = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_\varepsilon^2 + \frac{1}{2} \ln(1 - \rho^2) - \frac{1}{2\sigma_\varepsilon^2} \mathbf{u}' \boldsymbol{\Sigma}^{-1} \mathbf{u}$$

Si se deriva esta función con respecto a σ_ε^2 , se obtiene como condición necesaria de optimalidad:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\mathbf{u}}' \boldsymbol{\Sigma}^{-1} \hat{\mathbf{u}}}{T} = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{T}$$

que proporciona la estimación de σ_ε^2 una vez que se hayan obtenido los residuos $\hat{\varepsilon}_t$, es decir, una vez obtenidas las estimaciones de los coeficientes del modelo. La sustitución de este estimador en vez del parámetro σ_ε^2 en la función de verosimilitud genera la función que se conoce como *verosimilitud concentrada*:

$$\begin{aligned} \ln L &= -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \hat{\mathbf{u}}' \boldsymbol{\Sigma}^{-1} \hat{\mathbf{u}} + \frac{1}{2} \ln(1 - \rho^2) - \frac{T}{2} + \frac{T}{2} \ln T = \\ &= \text{Constante} - \frac{T}{2} \ln[(\hat{\mathbf{u}}' \boldsymbol{\Sigma}^{-1} \hat{\mathbf{u}})(1 - \rho^2)^{-1/T}] \end{aligned}$$

por lo que maximizar la verosimilitud es equivalente a minimizar la expresión dentro del corchete. Dicha expresión es muy parecida a la suma residual del modelo, excepto por el factor $(1 - \rho^2)^{-1/T}$. Este factor tiende a uno cuando el tamaño muestral T tiende a infinito, por lo que las diferencias entre ambos métodos de estimación disminuyen según crece el tamaño muestral. El factor también desaparecería si ignorásemos la primera observación muestral, es

decir, si estimásemos *condicionalmente* en la primera observación muestral. Una diferencia es que el estimador MCG que ignora la primera observación puede conducir a estimaciones numéricas de ρ fuera del intervalo $(-1, 1)$, lo que no ocurre con la estimación de máxima verosimilitud.

Si con el objeto de obtener el estimador de máxima verosimilitud se efectuase una búsqueda de red en el intervalo $(-1, 1)$ para el parámetro ρ , entonces el criterio a minimizar debería ser:

$$\frac{T}{2} \ln(\text{Suma residual}) - \frac{1}{2} \ln(1 - \rho^2) \quad [7.4]$$

Si el tamaño muestral es grande, entonces el segundo término es relativamente despreciable y, de nuevo, el procedimiento de búsqueda con el criterio de minimización de la suma residual tenderá a generar la estimación de máxima verosimilitud.

Como *método aproximado*, puede tratarse de maximizar la *función de verosimilitud condicional*, que trata la primera observación como una constante:

$$\begin{aligned} \ln L &= -\frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \Sigma_2^T \varepsilon_t^2 = \\ &= \frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \mathbf{u}' \Sigma^{-1} \mathbf{u} \end{aligned}$$

y si se elimina el término de error ε_t (o u_t) como función de las variables observables, se tiene:

$$\ln L = -\frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \Sigma_2^T [(y_t - \alpha - \mathbf{x}_t' \boldsymbol{\beta}) - \rho(y_{t-1} - \alpha - \mathbf{x}_{t-1}' \boldsymbol{\beta})]^2$$

El cálculo de las derivadas parciales de primer orden genera el sistema de condiciones necesarias para la maximización de la función de verosimilitud condicionada. Dicho sistema es:

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha} &= (1 - \hat{\rho}) \frac{\Sigma_2^T \hat{\varepsilon}_t}{\hat{\sigma}_\varepsilon^2} = 0 && (1 \text{ ecuación}) \\ \frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \frac{1}{\hat{\sigma}_\varepsilon^2} \Sigma_2^T \mathbf{x}_t^* \hat{\varepsilon}_t = 0 && (k - 1 \text{ ecuaciones}) \\ \frac{\partial \ln L}{\partial \rho} &= \frac{1}{\hat{\sigma}_\varepsilon^2} \Sigma_2^T \hat{u}_{t-1} (\hat{u}_t - \hat{\rho} \hat{u}_{t-1}) = 0 && (1 \text{ ecuación}) \\ \frac{\partial \ln L}{\partial \sigma_\varepsilon^2} &= -\frac{T-1}{2\hat{\sigma}_\varepsilon^2} + \frac{\Sigma_2^T \hat{\varepsilon}_t^2}{2\sigma_\varepsilon^4} = 0 && (1 \text{ ecuación}) \end{aligned} \quad [7.5]$$

La última de estas ecuaciones proporciona la estimación de máxima verosimilitud de σ_ε^2 como el cociente entre la suma residual y el tamaño muestral, que en nuestro caso es $T - 1$. Recordemos que la suma residual es siempre la suma de los cuadrados de los residuos obtenidos para el único proceso del modelo que es ruido blanco, y que en nuestra notación representamos por ε_t , y que $\varepsilon'\varepsilon = \sum_2^T \varepsilon_t^2 = \mathbf{u}'\Sigma^{-1}\mathbf{u}$.

La tercera ecuación del sistema anterior puede escribirse:

$$\hat{\rho} = \frac{\sum_2^T \hat{u}_t \hat{u}_{t-1}}{\sum_2^T \hat{u}_{t-1}^2}$$

y claramente muestra que la estimación máximo verosímil de ρ se obtiene mediante una regresión de \hat{u}_t sobre \hat{u}_{t-1} . Este resultado es importante, porque la estimación MCG del vector (α, β) depende de ρ , que por ser desconocido es preciso estimar previamente. Al discutir el estimador MCG en la Sección 7.5.a, sugerimos que se obtuviera la estimación del parámetro ρ de este modo, pero en aquel momento no teníamos una justificación rigurosa a este respecto. Ahora acabamos de probar que dicha estimación de ρ es de máxima verosimilitud condicionada.

Finalmente, las dos primeras ecuaciones del sistema [7.5] pueden interpretarse como las ecuaciones normales de una regresión que tiene a ε_t como término de error, y a las variables transformadas \mathbf{x}_t^* como explicativas; la primera ecuación muestra que la suma de los residuos de dicha regresión son cero, mientras que el segundo bloque ($k - 1$ ecuaciones) es el conjunto de ecuaciones normales correspondientes a los coeficientes β . Ello implica que las estimaciones MCG coinciden con las de máxima verosimilitud. Este es un resultado que ya conocíamos, pues probamos en la Sección 5.6 que, en un modelo con matriz de covarianzas no escalar, éste es un resultado general.

Vimos en la Sección 2.11 que en muestras grandes, el estimador de máxima verosimilitud puede aproximarse por una distribución Normal, cuya matriz de covarianzas es la inversa de la *matriz de información*:

$$I(\theta) = -E \left(\frac{\partial^2 \log^2 L(\theta)}{\partial^2 \theta} \Big|_{\theta = \theta_0} \right)$$

donde θ es el vector $(\alpha, \beta, \rho, \sigma_\varepsilon^2)$, y θ_0 denota el vector de verdaderos valores (desconocidos) de los parámetros, y la barra vertical indica que las derivadas segundas deben evaluarse en dichos valores antes de tomar su esperanza matemática. Dejamos al lector (Problema 7.4) la comprobación de que dicha matriz de información resulta ser:

$$I \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\rho} \\ \hat{\sigma}_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \frac{(T-1)(1-\rho)^2}{\sigma_\varepsilon^2} & (1-\rho) \frac{\sum_2^T \mathbf{x}_t^*}{\sigma_\varepsilon^2} & 0 & 0 \\ (1-\rho) \frac{\sum_2^T \mathbf{x}_t^*}{\sigma_\varepsilon^2} & \frac{\sum_2^T \mathbf{x}_t^* \mathbf{x}_t^{*'}}{\sigma_\varepsilon^2} & 0 & 0 \\ 0 & 0 & \frac{T-1}{1-\rho^2} & 0 \\ 0 & 0 & 0 & \frac{T-1}{2\sigma_\varepsilon^4} \end{pmatrix} \quad [7.6]$$

que muestra los siguientes resultados:

1. Por su estructura de bloques, las estimaciones máximo-verosimil de ρ y σ_ε^2 son independientes entre sí, así como independientes de la estimación del vector (α, β) .

2. La varianza del estimador de ρ es igual a $\frac{1-\rho^2}{T-1}$, lo que puede utilizarse para construir intervalos de confianza de dicho parámetro, sustituyendo ρ por su estimación en la expresión de dicha varianza y teniendo en cuenta su distribución aproximada Normal en muestras grandes.

3. La varianza de σ_ε^2 es $\frac{2\sigma_\varepsilon^4}{T-1}$, lo que puede asimismo utilizarse para construir intervalos de confianza para σ_ε^2 .

4. El bloque superior 2×2 coincide con la inversa de la matriz de covarianzas del estimador MCG de $((1-\rho)\alpha; \beta)$, que es $\sigma_\varepsilon^2 (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$. No podía ser de otra forma, puesto que ya sabemos que si el término de error del modelo sigue una distribución Normal, entonces dicho estimador es eficiente. Tan sólo hay que tener cuidado en advertir que el término independiente del modelo transformado no es α , el término independiente del modelo original, sino $\alpha(1-\rho)$.

5. La independencia entre los parámetros hace que sea posible estimar eficientemente (α, β) por una parte, ρ después mediante los residuos de la etapa anterior, y σ_ε^2 finalmente, del modo que se describió en la Sección 7.5.a. Si no se hubiesen obtenido los bloques de ceros que aparecen en la matriz de información, entonces la estimación por separado de estos parámetros no habría sido eficiente.

7.6. ANALISIS DE DOS CASOS PRACTICOS: LAS FUNCIONES DE CONSUMO E INVERSION

7.6.a. Una función de consumo

Como ejercicio aplicado, estimamos una sencilla función de consumo para la economía española, en la que el consumo agregado de los sectores público y

privado se hace depender del PIB, valorado a precios de mercado. Los datos anuales, que aparecen en el Tabla 7.1, están tomados de *La economía Española: Una perspectiva macroeconómica*, Molinas, Sebastián, Zabalza (1991), y corresponden al período 1954-1988, estando en pesetas constantes de 1980.

TABLA 7.1

Año	Consumo público	Consumo privado	PIB	Consumo total
1954	618,9	2.917,9	4.322,2	3.536,9
1955	639,9	3.071,5	4.546,6	3.711,5
1956	683,9	3.293,0	4.872,5	3.976,9
1957	720,9	3.399,0	5.080,8	4.120,0
1958	701,1	3.561,4	5.309,9	4.262,6
1959	725,6	3.607,5	5.209,2	4.333,1
1960	751,2	3.489,2	5.331,7	4.240,5
1961	791,1	3.867,5	5.963,0	4.658,7
1962	842,3	4.194,7	6.518,2	5.037,1
1963	917,5	4.638,1	7.089,0	5.555,7
1964	934,2	4.865,6	7.527,4	5.799,9
1965	965,7	5.186,9	8.004,1	6.152,6
1966	982,3	5.545,9	8.568,8	6.528,2
1967	1.006,2	5.883,4	8.939,1	6.889,7
1968	1.027,4	6.248,8	9.544,6	7.276,3
1969	1.069,1	6.674,6	10.397,9	7.743,7
1970	1.129,8	6.980,5	10.822,3	8.110,3
1971	1.178,9	7.333,3	11.318,0	8.512,2
1972	1.239,9	7.941,1	12.227,1	9.181,0
1973	1.319,7	8.557,7	13.166,9	9.877,4
1974	1.442,6	8.990,7	13.866,5	10.433,3
1975	1.517,1	9.152,2	13.940,9	10.669,3
1976	1.622,2	9.660,2	14.397,2	11.282,4
1977	1.686,2	9.805,8	14.829,2	11.492,0
1978	1.777,7	9.898,4	15.044,0	11.676,1
1979	1.851,9	10.023,1	15.023,1	11.875,0
1980	1.929,3	10.080,4	15.209,1	12.009,7
1981	1.965,4	10.020,0	15.171,3	11.985,5
1982	2.061,0	10.038,4	15.355,9	12.099,5
1983	2.141,7	10.072,8	15.633,1	12.214,5
1984	2.203,3	10.034,0	15.914,4	12.237,4
1985	2.305,4	10.273,3	16.282,8	12.578,8
1986	2.437,5	10.644,0	16.816,3	13.081,6
1987	2.649,3	11.224,4	17.748,6	13.873,8
1988	2.788,1	11.752,0	18.676,5	14.540,1

Las Figuras 7.3 y 7.4 muestran ambas series, así como sus variaciones anuales.

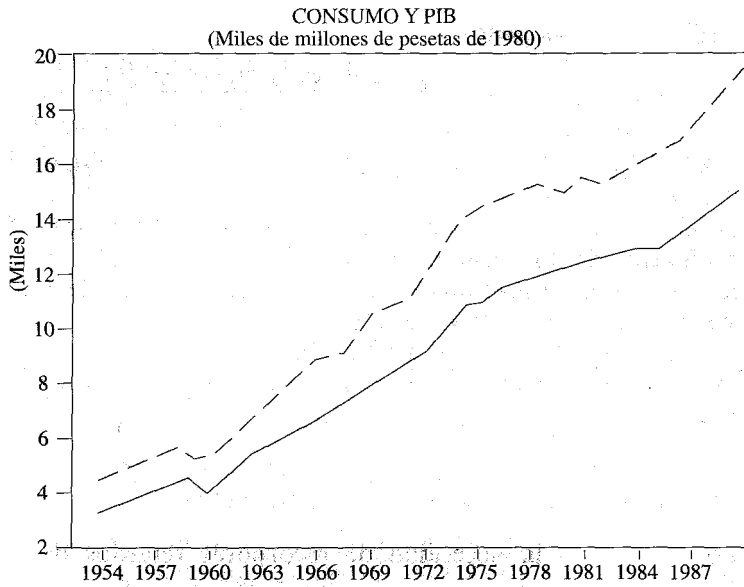


FIGURA 7.3

FIGURA 7.3.

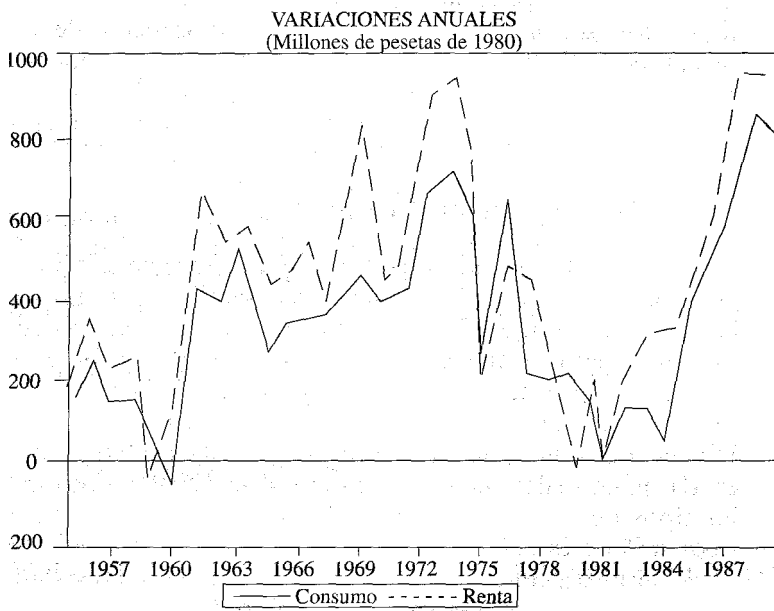


FIGURA 7.4

FIGURA 7.4.

Con ellas se estimó el modelo:

$$C_t = 76,56 + 0,77 Y_t + \hat{u}_t \quad [7.7]$$

(81,9) (0,07)

$$R^2 = 0,997; \quad SR = 1.079.364,7; \quad \hat{\sigma}_u = 180,9; \quad DW = 0,34$$

en el que se tiene una propensión marginal al consumo de 0,77. El elevado coeficiente de correlación es frecuente en datos macroeconómicos que, como los que aquí utilizamos, contienen una clara tendencia temporal, lo que volveremos a estudiar en el Capítulo 14.

El bajo valor del estadístico Durbin-Watson sugiere la existencia de una importante autocorrelación en los residuos. Si se estima un modelo autorregresivo para los residuos, se tiene:

$$\hat{u}_t = 0,825 \hat{u}_{t-1} + \hat{\varepsilon}_t \quad [7.8]$$

(0,097)

no detectando la necesidad de incorporar retardos de \hat{u}_t adicionales. Utilizando el modelo [7.8] para filtrar tanto C_t como Y_t y denotando por $C_t^* = C_t - 0,825 C_{t-1}$ e $Y_t^* = Y_t - 0,825 Y_{t-1}$ se obtuvo el estimador MCG:

$$C_t^* = 3,54 + 0,77 Y_t^* + \hat{u}_t \quad [7.9]$$

(51,6) (0,02)

$$R^2 = 0,977; \quad \hat{\sigma}_u = 102,0; \quad DW = 1,81$$

con idéntica propensión marginal al consumo que en [7.7], pero estimada ahora con mucha mayor precisión y sin indicios de autocorrelación. La constante de [7.9] no es equiparable a la de [7.7], sino que si las denotamos por $\hat{\beta}_1^*$ y $\hat{\beta}_1$, respectivamente, tenemos $\hat{\beta}_1^* = 3,54 = \hat{\beta}_1(1 - 0,825)$, es decir: $\hat{\beta}_1 = 20,23$.

7.6.b. Una función de inversión

Con objeto de estimar una función de inversión, tomamos datos para el período 1964-1988, de la misma fuente anterior, para las variables: Formación bruta de capital fijo y PIB a precios de mercado, ambas en pesetas corrientes, así como para el deflactor implícito del PIB con base en 1980, y tipos de interés, tanto a corto como a largo plazo. Transformando las variables en logaritmos, la primera diferencia del deflactor nos da una medida de inflación. Dividiendo la inversión y el PIB por el deflactor expresamos ambas variables en términos reales (pesetas constantes de 1980), mientras que si restamos de los tipos de interés nominales la inflación, tenemos una aproximación a los tipos de interés reales.

La primera columna de cada panel muestra los resultados de la estimación por MCO, en el primer caso con el tipo de interés real a corto plazo, siendo a largo plazo en el segundo modelo. Ambos presentan lo que parece una elasticidad significativa de la inversión respecto al PIB, evaluada en torno

al 16 por 100, así como una semielasticidad negativa, como cabría esperar, respecto a los tipos de interés reales.

Los coeficientes de determinación son elevados, pero tanto los estadísticos Durbin-Watson como Box-Pierce muestran evidencia clara de autocorrelación residual. Ello hace que la matriz de covarianzas habitual del estimador MCO, $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, de la que se obtuvieron las desviaciones típicas, no sea válida. La segunda columna de cada panel presenta las mismas estimaciones, pero esta vez con la matriz de covarianzas estimada por el procedimiento de Newey-West.

TABLA 7.2

<i>Constante</i>	8.344 (2.755)	8.344 (2.952)	4.410 (2.017)	10.546 (2.726)	10.546 (3.187)	4.499 (1.946)
<i>PIB</i>	0,159 (0,019)	0,159 (0,022)	0,157 (0,030)	0,155 (0,020)	0,155 (0,23)	0,157 (0,031)
<i>Tipo de interés a corto</i>	-0,311 (0,114)	-0,311 (0,155)	-0,120 (0,112)	—	—	—
<i>Tipo de interés a largo</i>	—	—	—	-0,303 (0,131)	-0,303 (0,162)	-0,107 (0,112)
\bar{R}^2	0,76		0,59	0,76		0,53
<i>DW</i>	0,61		1,64	0,58		1,65
$\hat{\sigma}$	2,69		1,66	2,80		1,65
<i>Q(12)</i>	35,3 (0,0)		8,34 (0,69)	37,7 (0,0)		8,02 (0,71)

Ante la evidencia de autocorrelación, procedimos a estimar ambos modelos por MCG, utilizando el procedimiento Cochrane-Orcutt mediante un modelo AR(2) estimado para los residuos de las estimaciones de la primera columna. Con dichos modelos AR(2) se filtraron las tres variables del modelo, volviendo a estimar por MCO con las variables filtradas, obteniendo los resultados de la tercera columna. Los coeficientes de los tipos de interés descienden en magnitud respecto a las estimaciones MCO, y resultan no significativos.

Volveremos a considerar estos dos ejemplos en el Capítulo 9.

7.7. EJERCICIOS DE SIMULACION

Johnston comenta en su texto (1984) la relación que existe entre la varianza de los estimadores MCG y MCO de β en el modelo de regresión simple:

$$\begin{aligned}
 y_t &= \beta x_t + u_t \\
 u_t &= \rho u_{t-1} + \varepsilon_t
 \end{aligned}
 \tag{7.10}$$

Dicha relación depende a su vez del comportamiento de la variable explicativa x_t . Johnston analiza el caso en que la variable x_t sigue un proceso autorregresivo de orden 1 con parámetro λ . Sólo si $\lambda = 0$ será válido el supuesto que mantenemos a lo largo de ese texto acerca de que las variables explicativas del modelo son deterministas, pero tiene interés discutir este aspecto en un contexto algo más general. El cociente entre las varianzas de los estimadores MCG y MCO del único parámetro β es una medida de la eficiencia relativa del estimador MCG. La relación es, bajo estos supuestos:

$$\frac{\text{Var}(\hat{\beta}_{\text{MCG}})}{\text{Var}(\hat{\beta}_{\text{MCO}})} = \frac{(1 - \rho^2)(1 - \rho\lambda)}{(1 + \rho^2 - 2\rho\lambda)(1 + \rho\lambda)}$$

con los siguientes valores:

$\lambda \backslash \rho$	-0,90	-0,50	0,50	0,90
0,0	0,11	0,60	0,60	0,11
0,50	0,19	0,71	0,60	0,08
0,90	0,53	0,92	0,81	0,11

donde puede verse que la ganancia de eficiencia del estimador MCG es mayor cuanto menor sea la autocorrelación de la variable x_t . Por otra parte, valores positivos de ambos parámetros generan una mayor ganancia en eficiencia por parte del estimador MCG.

Otro problema con la utilización del estimador MCO puede aparecer si se calcula como su matriz de covarianzas la tradicional matriz $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$. En el caso particular del modelo [7.10], la relación entre la expresión verdadera para la varianza del estimador MCO de β y su estimación sesgada $\frac{\sigma_u^2}{\sum_1^T x_t^2}$ puede aproximarse por:

$$\tau = \frac{\text{Var}(\hat{\beta}_{\text{MCO}})}{\sigma_u^2 / \sum_1^T x_t^2} = \frac{1 - \rho\lambda}{1 + \rho\lambda}$$

que implica un sesgo en la medición de dicha varianza, igual a $-\frac{2\rho\lambda}{1 + \rho\lambda}$. En la tabla siguiente puede comprobarse el valor de dicho sesgo dependiendo de los valores de los parámetros ρ y λ .

ρ	λ	Sesgo
-0,90	0,50	16,3 por 100 del valor de la varianza
0,90	0,50	-62,1 por 100 del valor de la varianza
0,90	0,90	-89,5 por 100 del valor de la varianza
0,80	0,80	-77,0 por 100 del valor de la varianza

Como puede verse en la tabla, cuando ambos parámetros son positivos, la expresión usual (aunque incorrecta) $\text{Var}(\hat{\beta}_{\text{MCO}}) = \sigma_u^2 / \sum_1^T x_i^2$ será una *subestimación* del verdadero valor de dicha varianza. Por ejemplo, si ambos parámetros valen 0,90, entonces la varianza estimada incorrectamente es tan sólo un 10 por 100 de su verdadero valor. Ello implica unos intervalos de confianza excesivamente pequeños y, como consecuencia, un estadístico t excesivamente grande, por lo que se tenderá a aceptar la significación de β mucho más a menudo de lo que se debiera. Por otra parte, puede verse en el texto de Johnston que si mantenemos el supuesto de que la única variable explicativa sigue un modelo autorregresivo de primer orden se tiene:

$$E(\mathbf{u}'\mathbf{u}) = \sigma_u^2 \left(T - \frac{1 + \rho\lambda}{1 - \rho\lambda} \right)$$

donde si $|\rho| < 1$ y $|\lambda| < 1$, entonces el parámetro resulta también inferior a 1 y de nuevo, si los parámetros ρ y λ tienen el mismo signo, entonces la expresión usual para $\hat{\sigma}_u^2$ será una *subestimación* de la verdadera estimación. Este efecto se acumula al anterior (si ρ y λ tienen el mismo signo), aunque no es excesivamente importante. Por ejemplo, si ambos parámetros tienen un valor de 0,90, entonces $E(\hat{\sigma}_u^2) = 0,915\sigma_u^2$.

PROBLEMAS

Problema 7.1. Para comprender la relación entre tendencia y autocorrelación, considérese una variable determinista de *tendencia* que recoge tan sólo el paso del tiempo, es decir: $y_t = t$, $t = 1, 2, \dots, T$. Calcúlese el estimador mínimo cuadrático del modelo:

$$y_t = \rho y_{t-1} + u_t$$

(a pesar de no existir componente aleatorio en y_t).

Problema 7.2. La variable Y de la tabla adjunta se ha construido añadiendo una tendencia igual a $0,5t$ a la variable X . Calcular el coeficiente de autocorrelación

X	Y
0,0	0,5
-0,1	0,9
0,1	1,6
0,2	2,2
-0,1	2,4
0,2	3,2
0,2	3,7
-0,1	3,9
0,4	4,4
0,3	5,3

$$r_1 = \frac{\sum x_t x_{t-1}}{\sum x_t^2}$$

para ambas variables e interpretar los resultados obtenidos.

Problema 7.3. El contraste de Godfrey se especifica en ocasiones a través de una regresión de los residuos MCO sobre p retardos suyos junto con las variables del modelo original. A continuación, el estadístico

$$\frac{\hat{u}'\mathbf{E}[\mathbf{E}'\mathbf{E} - \mathbf{E}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}]\mathbf{E}'\hat{u}}{\sigma_u^2}$$

se compara con una chi-cuadrado con p grados de libertad. La matriz \mathbf{E} contiene las «observaciones» de los residuos retardados, mientras que \mathbf{X} contiene las observaciones de las variables explicativas del modelo original. Probar la equivalencia de este estadístico con el sugerido en la Sección 7.4.

Problema 7.4. Demostrar que [7.6] es la matriz de información del estimador de máxima verosimilitud que ignora la primera observación muestral.

Problema 7.5. Dado el modelo

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

donde ε_t es *i. i. d.*, $N(0, \sigma_\varepsilon^2)$, y disponiendo de las observaciones numéricas:

t	1	2	3	4	5	6	7	8
y_t	22	26	32	34	40	46	46	50
x_t	4	6	10	12	14	16	20	22

- Obtenga una estimación eficiente de los coeficientes β_0 y β_1 , así como de su matriz de covarianzas, sabiendo que $\rho = 0,50$.
- Contraste la hipótesis $H_0: \beta_1 = 1$.
- Indique cómo se estimaría eficientemente el modelo si no se conociese el valor numérico del parámetro ρ .

Problema 7.6. Considere el modelo lineal $y_t = \alpha + u_t$, donde $E u_t = 0$, $\text{Var } u_t = \sigma_u^2$, $E(u_t \cdot u_{t-s}) = \rho^s \cdot \sigma_u^2 \quad \forall s \neq 0$.
¿Es el estimador MCO de α insesgado? ¿Es consistente?

Problema 7.7. Demuestre que el contraste de la h de Durbin, propuesto para el caso en que aparecen retardos de la variable endógena como variables explicativas, es un test de multiplicadores de Lagrange.

Problema 7.8. Demuestre que el contraste de autocorrelación propuesto por Breusch y Godfrey es un contraste del tipo de multiplicadores de Lagrange.

Problema 7.9. Reducción de autocorrelación mediante primeras diferencias: Considere los modelos

$$\begin{cases} y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{cases} \quad \text{y} \quad \begin{cases} y_t - y_{t-1} = (\mathbf{x}_t' - \mathbf{x}_{t-1}') \boldsymbol{\beta} + v_t \\ v_t = u_t - u_{t-1} \end{cases}$$

y compare la autocorrelación de u_t con la de v_t . ¿Cómo cambiaría su respuesta si la relación entre u_t y ε_t fuese $u_t = \varepsilon_t - \lambda \varepsilon_{t-1}$?

Problema 7.10. Al estimar por MCO un modelo lineal, utilizando 21 observaciones, se obtuvo:

$$y_t = 1,3 + 0,97 y_{t-1} + 2,31 x_t; \quad DW = 1,21$$

$$(0,3) \quad (0,18) \quad (0,41)$$

donde las cifras entre paréntesis son las desviaciones típicas asintóticas. Contraste la presencia de autocorrelación en el término de error.

Problema 7.11. ¿Cómo obtendría predicciones numéricas a partir del modelo

$$\begin{cases} y_t = x_t' \beta + u_t \\ u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t, \end{cases} \quad t = 3, 4, \dots, T$$

para diferentes horizontes predictivos, siendo $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, independiente en el tiempo? ¿Cómo cambiaría su respuesta si la estructura de autocorrelación de u_t fuese $u_t = \varepsilon_t - \lambda \varepsilon_{t-1}$? ¿Cuáles serían las varianzas de las predicciones un periodo hacia el futuro?

Problema 7.12. Se quiere predecir la inflación π_t para 1986 utilizando el modelo:

$$\pi_t = 0,012 + 0,74 \dot{m}_t + u_t, \quad R^2 = 0,75$$

$$(0,85) \quad (5,40) \quad [1]$$

estimado con datos anuales 1964-1985, donde m_t denota el crecimiento de la oferta monetaria y los números en paréntesis son los valores de los estadísticos t de Student. La autoridad monetaria anuncia que la oferta monetaria crecerá al 10 por 100 durante 1986.

- ¿Cuál sería la predicción de la inflación?
- ¿Cómo cambiaría la respuesta anterior si no se tuviese información acerca de la evolución de la oferta monetaria?
- Cuál sería la respuesta si, conociendo el anuncio de la autoridad monetaria y la dificultad de controlar el agregado monetario, se cree que su crecimiento vendrá dado por:

$$\dot{m}_t = 0,10 + \varepsilon_t \quad [2]$$

donde ε_t es un término de error sin autocorrelación, $N(0, \sigma_\varepsilon^2)$.

- Cuál sería la respuesta si se cree que \dot{m}_t será exactamente 0,10 pero el término de error del modelo [1] tiene autocorrelación de primer orden:

$$u_t = \rho u_{t-1} + \xi_t, \quad \text{con } E(\xi_t \cdot \xi_s) = 0, \quad t \neq s, \quad N(0, \sigma_\xi^2)$$

- Comente la especificación [1] desde los puntos de vista siguientes:
 - Valor de R^2 .
 - Si hay alguna razón para suponer que la relación entre π_t y \dot{m}_t es de π_t hacia \dot{m}_t en vez de \dot{m}_t hacia π_t , ¿tendría el tipo de datos utilizados alguna influencia en la respuesta?
 - Mencione posibles variables alternativas a incluir en [1] y [2].

CAPITULO 8

ECUACIONES SIMULTANEAS CON VARIABLES EXPLICATIVAS EXOGENAS

Consideramos en este capítulo, por primera vez, un modelo econométrico constituido por un sistema de ecuaciones de regresión lineal. El capítulo está dividido en dos partes: en la primera examinamos la situación que surge cuando se dispone de varias muestras procedentes de unidades económicas diferentes, aunque comparables, y se pretende contrastar la homogeneidad de todas ellas. En la segunda parte discutimos una situación algo diferente, en que las ecuaciones que constituyen el sistema no son necesariamente similares entre sí, pero existen razones para pensar que sus términos de error están correlacionados en cada período; se trata de estimar eficientemente, sacando provecho de tales correlaciones. Los ejemplos al comienzo de cada una de las dos partes ilustran diversas situaciones de cada tipo.

Parte I

UNA SECCION CRUZADA DE SERIES TEMPORALES

8.1. INTRODUCCION

En ocasiones, el investigador dispone de varias muestras análogas que podría utilizar indistintamente para contrastar las hipótesis de su interés. Un ejemplo típico es la situación que se le presenta al analista que quiere caracterizar la estructura de costes de la industria automovilística y dispone para ello de datos de utilización de los distintos inputs, así como del producto que generan distintos fabricantes.

Una situación similar aparece cuando se pretende contrastar la proposición de neutralidad monetaria a largo plazo: éste es un estudio que podría

llevarse a cabo con datos de España, pero también podría realizarse con datos de Estados Unidos o Alemania, por ejemplo. Si el investigador no tiene un interés especial en un determinado país, sino que tiene más bien un interés general en la proposición de neutralidad, sería interesante realizar el estudio para un conjunto de países. Así, con el ánimo de establecer comparaciones internacionales de los resultados obtenidos, el investigador podría estimar un modelo similar para cada país.

Existen, sin embargo, situaciones en las que sería preferible la estimación simultánea del conjunto de ecuaciones correspondientes a cada uno de los países de los que se dispone información muestral:

1. Supongamos que se piensa que los diversos países han experimentado, durante el período muestral, perturbaciones exógenas comunes —como serían las subidas inesperadas en los precios internacionales del petróleo—, además de sus perturbaciones específicas. En tal caso, los términos de error de las ecuaciones de los distintos países tendrán un componente común, lo que se traducirá en una correlación contemporánea no despreciable. Una estimación simultánea del conjunto de ecuaciones podría utilizar la matriz de correlaciones de los términos de error para ganar eficiencia (tal estrategia de estimación es la que estudiamos en la segunda parte del capítulo).

2. En la gran mayoría de los casos, las comparaciones internacionales que quieran efectuarse van a traducirse en la contrastación de que determinados efectos estimados son iguales entre un conjunto de países. Para llevar a cabo de forma rigurosa tal contraste de hipótesis, es preciso estimar simultáneamente los modelos de dichos países.

El objeto de la primera parte de este capítulo es precisamente el de contrastar la homogeneidad de un conjunto de ecuaciones como el descrito; se trata de modelos de regresión que son conceptualmente idénticos, cada uno de los cuales es estimado con datos de una unidad económica diferente, de cuya homogeneidad se duda. Para contrastar la homogeneidad de las distintas ecuaciones se hace preciso considerar su estimación simultánea. Sin embargo, en esta primera parte del capítulo no aprovechamos las posibles correlaciones entre los términos de error de las distintas ecuaciones; éste será el aspecto analizado en la segunda parte. Mantenemos en todo momento el supuesto de que las variables explicativas de cada ecuación pueden suponerse deterministas.

Puesto que tratamos ahora con una muestra compuesta de submuestras similares recogidas en diversas unidades económicas, necesitamos introducir una notación algo más compleja de lo habitual. El primer subíndice en las variables y matrices que aparecen en este capítulo denota la unidad económica de la que proceden: así, y_{it} y x_{ikt} denotan la observación del período t en la submuestra i -ésima para las variables y y x_k ; y_i y X_i denotan el vector $T \times 1$ y la matriz $T \times (k - 1)$ de observaciones de las variables explicativas de dicha submuestra; cuando queramos considerar todas las variables del modelo simultáneamente, utilizaremos la matriz $W_i = (\mathbf{1}_T; X_i)$, de dimensión $T \times k$, en la que $\mathbf{1}_T$ denota un vector columna de T unos.

Denotamos por α_i el término independiente y por β_i los restantes $k - 1$ coeficientes correspondientes a la unidad i -ésima. Como se aprecia con esta notación, vamos a dar un trato diferenciado al término independiente y a los demás coeficientes del modelo a los que, por analogía con el modelo de regresión lineal simple, nos referiremos como *pendientes* del modelo.

De este modo, tenemos el sistema de ecuaciones:

$$\begin{aligned} y_1 &= \alpha_1 + X_1 \beta_1 + u_1 \\ y_2 &= \alpha_2 + X_2 \beta_2 + u_2 \\ &\dots\dots\dots \\ y_m &= \alpha_m + X_m \beta_m + u_m \end{aligned}$$

donde los vectores u_i , $i = 1, 2, \dots, m$ tienen dimensión $T \times 1$.

Supondremos en el capítulo que se dispone de un número igual de observaciones T para cada una de las unidades económicas. Este supuesto simplifica la notación, pero no es restrictivo, y los resultados que vamos a obtener son generalmente válidos cuando el número de observaciones varía de unas unidades a otras.

Igualmente, podría especificarse un número distinto k_i de variables explicativas para cada unidad muestral. Sin embargo, esta generalización contradice en parte la idea que mantenemos en este capítulo de contrastar la homogeneidad de los coeficientes estimados para modelos similares, por lo que consideramos en lo que sigue que $k_1 = k_2 = \dots = k_m$.

8.2. DIVERSAS ESPECIFICACIONES DE INTERES

Los métodos de inferencia desarrollados en el Capítulo 4 fueron propuestos para contrastar hipótesis acerca de los coeficientes de *un* modelo econométrico. La situación que ahora contemplamos considera, por el contrario, la igualdad de coeficientes de *distintos* modelos. Aparentemente, aquellos procedimientos no serían válidos y deberíamos diseñar nuevos métodos; afortunadamente, la discusión que allí efectuamos es suficientemente general como para que pueda utilizarse en este nuevo contexto sin necesidad de modificaciones.

Para entender este aspecto, recordemos que los procedimientos de contrastación que ya conocemos consisten en formular dos modelos, uno de los cuales es una versión restringida del otro; en tal situación, la comparación de las sumas residuales de ambos modelos genera el estadístico F que permite llevar a cabo el contraste de hipótesis.

Comencemos examinando las especificaciones que resultan de interés cuando se dispone de datos procedentes de distintas unidades económicas. En primer lugar, el modelo M1

Modelos análogos pueden formularse cuando se dispone de un número reducido de observaciones temporales relativas a un número grande m de unidades muestrales. Por ejemplo, la especificación M4 sería en tal caso:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ \dots \\ y_T \end{pmatrix} = \begin{pmatrix} 1_m & X_1 \\ 1_m & X_2 \\ 1_m & X_3 \\ \dots & \dots \\ \dots & \dots \\ 1_m & X_T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \dots \\ \dots \\ \beta_T \end{pmatrix} + u \quad [8.5]$$

en el que se estima un vector de coeficientes para cada período (un número reducido de ellos) utilizando en cada instante t el elevado número de observaciones de la sección cruzada. No existe ninguna diferencia conceptual importante entre ambos casos; por el contrario, son situaciones totalmente análogas, y el lector debe saber formular los modelos adecuados en cada caso. Al igual que con los modelos M1 a M3, es preciso que el número de grados de libertad sea relativamente alto para otorgar una mínima fiabilidad a las estimaciones que se obtengan. En este contexto, el problema aparecería si el número de períodos T para los que se pretende estimar un vector de coeficientes fuese elevado en relación al número de unidades muestrales de que se dispone.

Es importante interpretar cada uno de estos modelos: M1 es una simple agrupación matricial de todas las observaciones muestrales, en la que se ha respetado la posible heterogeneidad de todos los coeficientes a través de las unidades submuestrales; la situación polar de ésta viene representada por M2, que es una formulación severamente restringida por el supuesto de que tanto los términos independientes como los vectores de pendientes son *comunes* a las distintas submuestras.

El modelo M3 supone que los efectos de las variables explicativas sobre la variable endógena son comunes a todas las unidades muestrales, de modo que la heterogeneidad en la muestra se produce únicamente a través de los términos independientes: ello significa que, incluso a igualdad de niveles medios de las variables explicativas en las submuestras, los *niveles medios* de la variable endógena serían diferentes, tanto mayores cuanto mayor sea el valor numérico del término independiente correspondiente.

8.3. CONTRASTES DE HOMOGENEIDAD ENTRE UNIDADES MUESTRALES

Dos aspectos importantes merecen destacarse en las especificaciones anteriores:

1. En primer lugar, todas ellas permiten tratar el conjunto de ecuaciones correspondiente a las distintas unidades muestrales como si de un único modelo econométrico se tratase.

2. En segundo lugar, vamos a ver en esta sección que las hipótesis de homogeneidad más interesantes pueden representarse como restricciones sobre los coeficientes de este único modelo global.

Ambas condiciones juntas hacen que los procedimientos de contrastación de hipótesis estudiados en el Capítulo 4 utilizando el estadístico F sean aplicables. Para ello es preciso que hagamos el supuesto que allí mantuvimos, acerca de que el término de error del modelo global $\mathbf{u}' = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_m)$ tenga una distribución de probabilidad Normal con matriz de covarianzas escalar. Tal supuesto es aquí ciertamente restrictivo, pues exige no sólo que cada unidad submuestral tenga distribución $N(\mathbf{0}_T, \Sigma_i)$ con matriz de covarianzas $\Sigma_i = \sigma_i^2 \mathbf{I}_T$ escalar, sino que los parámetros σ_i^2 sean iguales para todas ellas. En cualquier caso, en tanto en cuanto se admita que las matrices de covarianzas son escalares, si los parámetros σ_i^2 fuesen diferentes podría estimarse en cada unidad submuestral por MCO para obtener con los residuos una estimación de σ_i^2 , y transformar las observaciones de la unidad i -ésima por el cociente $\hat{\sigma}_1/\hat{\sigma}_i$; de este modo, todas las submuestras, ahora con variables transformadas, tendrían la misma matriz de covarianzas⁽¹⁾: $\Sigma_i = \sigma_1^2 \mathbf{I}_T$.

Basta recordar ahora que el estadístico F se construye como cociente de dos términos: el numerador es la diferencia entre suma residual restringida y suma residual sin restringir, dividida por el número de restricciones que se contrastan; el denominador es la suma residual sin restringir, dividida por el número de grados de libertad de su modelo. El número de restricciones es siempre igual a la diferencia entre los grados de libertad del modelo sin restringir y el modelo restringido. Se trata, en cada caso, de estimar ambos modelos y obtener el valor numérico del estadístico F del modo antes indicado para llevar a cabo el contraste de hipótesis que se pretende.

Los contrastes más interesantes son:

Contraste 1. Homogeneidad global del modelo:

		<i>Grados de libertad</i>	
Modelo restringido:	Modelo M2	Numerador	Denominador
Modelo sin restringir:	Modelo M1	$k(m - 1)$	$Tm - km$

Contraste 2. Contraste de Igualdad de Pendientes:

		<i>Grados de libertad</i>	
Modelo restringido:	Modelo M3	Numerador	Denominador
Modelo sin restringir:	Modelo M1	$(m - 1)(k - 1)$	$Tm - km$

⁽¹⁾ Podría haberse escogido otro σ_1 con el que normalizar. También podría dividirse cada submuestra por σ_i , para normalizar todas las desviaciones típicas a 1.

Contraste 3. Contraste de Igualdad de Términos Independientes, supuesto que las pendientes de ambos modelos son iguales:

		Grados de libertad	
Modelo restringido:	Modelo M2	Numerador	Denominador
Modelo sin restringir:	Modelo M3	$(m - 1)$	$Tm - k - m + 1$

El primer contraste compara una situación en que tanto el término independiente α como los coeficientes β son específicos de cada unidad muestral, con otra en que todos ellos son comunes a todas las submuestras, por lo que lo calificamos de *contraste de homogeneidad global*. El segundo contraste examina si la heterogeneidad entre submuestras se debe a las pendientes β , comparando dos modelos, ambos de los cuales incorporan términos independientes específicos de las unidades muestrales. El tercero de los contrastes que presentamos es dual del anterior: manteniendo la igualdad de pendientes entre submuestras, se contrasta si la heterogeneidad proviene tan sólo de los términos independientes.

Hay que hacer notar que este último contraste de igualdad de términos independientes se condiciona en el supuesto mantenido (y no contrastado) de que las pendientes del modelo son iguales. Como hemos mencionado en capítulos anteriores, el economista está más interesado, en general, en estimar los valores numéricos de las pendientes del modelo que en el término independiente. Este funciona como un término de ajuste relacionado con los niveles de las variables endógena y exógena en cada submuestra.

Lógicamente, se debería llevar a cabo en primer lugar un contraste global de homogeneidad (Contraste 1). Si se rechaza la homogeneidad de las submuestras, entonces debería pasarse al Contraste 2, para discutir si la heterogeneidad proviene de diferencias en las pendientes y, en caso contrario, pasar al Contraste 3, para ver si la heterogeneidad se debe a diferencias en los términos independientes, mientras que las pendientes son iguales.

El Contraste 3, que trata de descubrir en el término independiente del modelo lineal un efecto diferenciador producido por la heterogeneidad de las distintas submuestras, se conoce como Análisis de la Covarianza.

Llevar a cabo cada uno de los contrastes e hipótesis descritas se reduce, por tanto, a estimar los dos modelos que en él aparecen. Estos son siempre los que hemos denominado M1 a M3. Por consiguiente, las secciones que restan en este capítulo las dedicamos a discutir la estimación de tales modelos.

8.4. ESTIMACIONES DE MODELOS EN EL CASO DE IGUALDAD DE MATRICES DE COVARIANZAS

Para discutir la estimación de los tres modelos que hemos presentado en la sección anterior, vamos a suponer inicialmente que no hay autocorrelación ni heteroscedasticidad en ninguna de las submuestras y que la varianza del término de error σ_u^2 es igual en todas ellas. Asimismo, suponemos inicialmente

que los términos de error de las distintas ecuaciones son independientes entre sí. Más adelante generalizaremos la discusión a casos en que los supuestos anteriores no se cumplen.

Como la exposición de los procedimientos de estimación que vamos a efectuar es bastante exigente desde el punto de vista analítico, conviene que examinemos de manera resumida los resultados que vamos a obtener:

1. Para estimar M1 podríamos transformar las variables en diferencias respecto a sus promedios, obtenidos dentro de cada submuestra. Como no hay ningún coeficiente común a las distintas submuestras, parece (y así lo demostramos) como si fuese eficiente estimar cada $\hat{\beta}_i$, $i = 1, 2, \dots, m$ con el bloque de variables en diferencias respecto a las medias de su submuestra, para luego recuperar los términos independientes mediante $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i' \hat{\beta}_i$.

2. Para estimar M2 podríamos transformar todas las variables en diferencias respecto a las *medias globales*, es decir, calculadas con las Tm observaciones, para luego recuperar el coeficiente $\hat{\alpha}$ a partir de las estimaciones $\hat{\beta}$ y las medias \bar{y} , \bar{x} .

3. Puesto que el modelo M3 considera heterogeneidad tan sólo en los términos independientes, si transformamos las variables en diferencias respecto a las medias dentro de cada submuestra, tendríamos un modelo con Tm observaciones y tan sólo $k - 1$ coeficientes. Tras estimar este vector común β con todas las observaciones recuperaríamos cada término independiente mediante $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i' \hat{\beta}$.

Para ello es interesante que el lector entienda y se familiarice previamente con las siguientes igualdades:

$$\begin{aligned} \mathbf{1}'_T \mathbf{X}_i &= \sum_1^T (x_{i2t}, x_{i3t}, \dots, x_{ikt}) = T \bar{x}'_i \\ \sum_1^m \mathbf{1}'_T \mathbf{X}_i &= \sum_1^m \sum_1^T (x_{i2t}, x_{i3t}, \dots, x_{ikt}) = T m \bar{x}' \\ \mathbf{1}'_T \mathbf{y}_i &= T \bar{y}_i; \quad \sum_1^m \mathbf{1}'_T \mathbf{y}_i = T m \bar{y} \end{aligned} \quad [8.6]$$

donde hemos denotado por \bar{x}_i el vector columna $k \times 1$ de medias aritméticas de observaciones de las variables explicativas, calculadas sobre la submuestra i -ésima, mientras que \bar{x} denota el vector de medias calculadas con todas las observaciones muestrales. Por \bar{y}_i e \bar{y} denotamos momentos análogos de la variable endógena, ambos escalares. Si transformamos las variables del modelo en diferencias con respecto a la media, las matrices de observaciones correspondientes a la submuestra i -ésima pueden escribirse:

$$\mathbf{X}_i - \mathbf{1}_T \bar{x}'_i \quad \text{e} \quad \mathbf{y}_i - \mathbf{1}_T \bar{y}_i$$

o bien:

$$\mathbf{X}_i - \mathbf{1}_T \bar{x}' \quad \text{e} \quad \mathbf{y}_i - \mathbf{1}_T \bar{y}$$

dependiendo de que las diferencias se tomen con respecto a los promedios muestrales, o a los calculados tan sólo con las observaciones de la submuestra correspondiente.

Estimación del modelo M1. En este modelo, la matriz \mathbf{W} de orden $Tm \times km$ tiene la forma:

$$\mathbf{W} = \begin{pmatrix} \mathbf{1}_T & & & \mathbf{X}_1 & & & \\ & \mathbf{1}_T & & & \mathbf{X}_2 & & \\ & & \dots & & & \dots & \\ & & & \mathbf{1}_T & & & \mathbf{X}_m \end{pmatrix}$$

por lo que la matriz $\mathbf{W}'\mathbf{W}$, $km \times km$, y el producto $\mathbf{W}'\mathbf{y}$, $km \times 1$:

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} T & & & & & & \\ & T & & & & & \\ & & \dots & & & & \\ & & & T & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ T\bar{\mathbf{x}}_1 & & & & & & \\ & T\bar{\mathbf{x}}_2 & & & & & \\ & & \dots & & & & \\ & & & T\bar{\mathbf{x}}_m & & & \\ & & & & \mathbf{X}'_1\mathbf{X}_1 & & \\ & & & & & \mathbf{X}'_2\mathbf{X}_2 & \\ & & & & & & \dots \\ & & & & & & \mathbf{X}'_m\mathbf{X}_m \end{pmatrix} \quad \mathbf{W}'\mathbf{y} = \begin{pmatrix} T\bar{y}_1 \\ T\bar{y}_2 \\ \dots \\ T\bar{y}_m \\ \dots \\ \mathbf{X}'_1\mathbf{y}_1 \\ \mathbf{X}'_2\mathbf{y}_2 \\ \dots \\ \mathbf{X}'_m\mathbf{y}_m \end{pmatrix}$$

y el primer bloque de ecuaciones normales es:

$$\begin{cases} T\hat{\alpha}_1 + T\bar{\mathbf{x}}'_1\hat{\beta}_1 = T\bar{y}_1 \\ \dots \dots \dots \Rightarrow \hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}'_i\hat{\beta}_i, \quad i = 1, 2, \dots, m \\ T\hat{\alpha}_m + T\bar{\mathbf{x}}'_m\hat{\beta}_m = T\bar{y}_m \end{cases} \quad [8.7]$$

y sustituyendo en el segundo bloque de ecuaciones normales:

$$\begin{cases} T\bar{\mathbf{x}}_1(\bar{y}_1 - \bar{\mathbf{x}}'_1\hat{\beta}_1) + (\mathbf{X}'_1\mathbf{X}_1)\hat{\beta}_1 = \mathbf{X}'_1\mathbf{y}_1 \\ \dots \dots \dots \\ T\bar{\mathbf{x}}_m(\bar{y}_m - \bar{\mathbf{x}}'_m\hat{\beta}_m) + (\mathbf{X}'_m\mathbf{X}_m)\hat{\beta}_m = \mathbf{X}'_m\mathbf{y}_m \end{cases}$$

es decir:

$$(\mathbf{X}'_i\mathbf{X}_i - T\bar{\mathbf{x}}_i\bar{\mathbf{x}}'_i)\hat{\beta}_i = \mathbf{X}'_i\mathbf{y}_i - T\bar{\mathbf{x}}'_i\bar{y}_i, \quad i = 1, 2, \dots, m$$

$$[(\mathbf{X}_i - \mathbf{1}_T\bar{\mathbf{x}}_i)'(\mathbf{X}_i - \mathbf{1}_T\bar{\mathbf{x}}_i)]\hat{\beta}_i = (\mathbf{X}_i - \mathbf{1}_T\bar{\mathbf{x}}_i)'(\mathbf{y}_i - \mathbf{1}_T\bar{y}_i)$$

donde hemos utilizado la igualdad $\mathbf{1}'_T\mathbf{X}_i = T\bar{\mathbf{x}}'_i$ para obtener el sistema de ecuaciones normales del modelo M1 tras haber transformado las variables en diferencias con respecto a los promedios de la submuestra correspondiente, y

sugiere que *debe estimarse un modelo diferente para cada submuestra*. Ello es bastante natural, puesto que: a) *todos los coeficientes varían de unas submuestras a otras*, y b) *suponemos ausencia de correlación contemporánea*. Una vez obtenida la estimación de las pendientes β_i , entonces las estimaciones de los distintos términos independientes se obtendría de acuerdo con [8.7]. Debe hacerse hincapié en que éste es un número elevado de coeficientes y sólo si T es suficientemente mayor que k será relevante la estimación de este modelo.

Por tanto:

$$\hat{\beta}_i = [(\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)' (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)]^{-1} [(\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)' (\mathbf{y}_i - \mathbf{1}_T \bar{y}_i)]$$

con matriz de covarianzas igual a:

$$\text{Var}(\hat{\beta}_i) = \sigma_u^2 [(\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)' (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)]^{-1}$$

mientras que la estimación del parámetro σ_u^2 se obtendría mediante la expresión:

$$\hat{\sigma}_u^2 = \sum_1^m \left(\frac{(\mathbf{y}_i - \mathbf{1}_T \bar{y}_i)' (\mathbf{y}_i - \mathbf{1}_T \bar{y}_i) - \hat{\beta}_i' (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)' (\mathbf{y}_i - \mathbf{1}_T \bar{y}_i)}{(T - k)m} \right)$$

y

$$\text{Var}(\hat{\alpha}_i) = \frac{\sigma_u^2}{T} + \bar{x}_i' \text{Var}(\hat{\beta}_i) \bar{x}_i$$

Como puede verse, el modelo M1 se estima con cada submuestra por separado, sin utilizar en ningún momento datos agregados.

Estimación del modelo M2. La estimación del modelo M2 es asimismo bastante sencilla, puesto que al suponer que todos los parámetros son iguales para todas las submuestras, puede utilizarse MCO una sola vez con todas las observaciones para estimar los k parámetros del modelo. Como puede verse en la descripción de los contrastes que propusimos en la sección anterior, este modelo aparece siempre como el modelo restringido para contrastar determinadas hipótesis. La matriz \mathbf{W} de observaciones de las variables explicativas es, en este caso, de dimensión $Tm \times k$, mientras que el vector \mathbf{y} de observaciones de la variable endógena es $Tm \times 1$:

$$\mathbf{W} = \begin{pmatrix} \mathbf{1}_T & \mathbf{X}_1 \\ \mathbf{1}_T & \mathbf{X}_2 \\ \mathbf{1}_T & \mathbf{X}_3 \\ \dots & \dots \\ \mathbf{1}_T & \mathbf{X}_m \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{W}_3 \\ \dots \\ \mathbf{W}_m \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_m \end{pmatrix}$$

Por consiguiente, $W'W$ es $k \times k$, mientras que $W'y$ es $k \times 1$:

$$W'W = \begin{pmatrix} Tm & \Sigma_1^m \mathbf{1}_T' \mathbf{X}_i \\ \Sigma_1^m \mathbf{X}_i' \mathbf{1}_T & \Sigma_1^m \mathbf{X}_i' \mathbf{X}_i \end{pmatrix} = \begin{pmatrix} Tm & Tm\bar{\mathbf{x}}' \\ Tm\bar{\mathbf{x}} & \Sigma_1^m \mathbf{X}_i' \mathbf{X}_i \end{pmatrix} \quad W'y = \begin{pmatrix} Tm\bar{y} \\ \Sigma_1^m \mathbf{X}_i' y_i \end{pmatrix}$$

donde hemos utilizado las igualdades [8.6].

Por tanto, las ecuaciones normales del modelo M2 resultan:

$$\begin{cases} Tm\hat{\alpha} + Tm\bar{\mathbf{x}}' \hat{\boldsymbol{\beta}} = Tm\bar{y} \\ Tm\bar{\mathbf{x}} \hat{\alpha} + (\Sigma_1^m \mathbf{X}_i' \mathbf{X}_i) \hat{\boldsymbol{\beta}} = \Sigma_1^m \mathbf{X}_i' y_i \end{cases}$$

De la primera de ellas se tiene:

$$\hat{\alpha} = \bar{y} - \bar{\mathbf{x}}' \hat{\boldsymbol{\beta}} \quad [8.8]$$

que muestra cómo el estimador MCO del término independiente se obtiene utilizando las medias muestrales globales y el vector $\hat{\boldsymbol{\beta}}$, común a todas las submuestras.

Sustituyendo [8.8] en el segundo bloque de ecuaciones normales:

$$Tm\bar{\mathbf{x}}\bar{y} - Tm\bar{\mathbf{x}}\bar{\mathbf{x}}' \hat{\boldsymbol{\beta}} + (\Sigma_1^m \mathbf{X}_i' \mathbf{X}_i) \hat{\boldsymbol{\beta}} = \Sigma_1^m \mathbf{X}_i' y_i$$

es decir:

$$[\Sigma_1^m (\mathbf{X}_i' \mathbf{X}_i - T\bar{\mathbf{x}}\bar{\mathbf{x}}')] \hat{\boldsymbol{\beta}} = \Sigma_1^m (\mathbf{X}_i' y_i - T\bar{\mathbf{x}}\bar{y})$$

o, lo que es lo mismo, utilizando las expresiones [8.6]⁽²⁾:

$$[\Sigma_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})] \hat{\boldsymbol{\beta}} = \Sigma_1^m [(\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (y_i - \mathbf{1}_T \bar{y})]$$

que son las $k - 1$ ecuaciones normales para la estimación del vector $\boldsymbol{\beta}$ con las variables transformadas en diferencias con respecto a las medias muestrales. Por tanto, la estimación del modelo se reduce a:

1. Calcular dichos promedios.
2. Llevar a cabo las transformaciones mencionadas, con lo cual desaparece el término independiente del modelo.
3. Estimar las pendientes $\boldsymbol{\beta}$ y, finalmente:
4. Recuperar el término independiente como se indica en [8.8].

Así pues, la estimación MCO del vector $\boldsymbol{\beta}$ es:

$$\hat{\boldsymbol{\beta}} = [\Sigma_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})]^{-1} [\Sigma_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (y_i - \mathbf{1}_T \bar{y})]$$

⁽²⁾ Basta notar que, por ejemplo:

$$\begin{aligned} \Sigma_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (y_i - \mathbf{1}_T \bar{y}) &= \Sigma_1^m (\mathbf{X}_i' y_i - \bar{\mathbf{x}}_i \mathbf{1}_T' y_i - \mathbf{X}_i' \mathbf{1}_T \bar{y}_i + \bar{\mathbf{x}}_i \mathbf{1}_T' \mathbf{1}_T \bar{y}_i) = \\ &= \Sigma_1^m (\mathbf{X}_i' y_i - T\bar{\mathbf{x}}_i \bar{y}_i - T\bar{\mathbf{x}}_i \bar{y}_i + T\bar{\mathbf{x}}_i \bar{y}_i) = \Sigma_1^m (\mathbf{X}_i' y_i - T\bar{\mathbf{x}}_i \bar{y}_i) \end{aligned}$$

y su matriz de covarianzas es:

$$\text{Var}(\hat{\beta}) = \sigma_u^2 [\sum_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})]^{-1}$$

con:

$$\text{Var}(\hat{\alpha}) = \frac{\sigma_u^2}{Tm} + \bar{\mathbf{x}}' \text{Var}(\hat{\beta}) \bar{\mathbf{x}}$$

donde el parámetro σ_u^2 se estima a partir de la suma residual obtenida con todas las observaciones muestrales y el vector $\hat{\beta}$. El cociente de dicha suma residual por el número de grados de libertad del modelo, $Tm - k$, es la estimación insesgada de dicho parámetro⁽³⁾:

$$\hat{\sigma}_u^2 = \frac{(\mathbf{y} - \mathbf{1}_{Tm} \bar{y})' (\mathbf{y} - \mathbf{1}_{Tm} \bar{y}) - \hat{\beta}' (\sum_1^m (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}})' (\mathbf{y}_i - \mathbf{1}_T \bar{y}))}{Tm - k}$$

Estimación del modelo M3. Este modelo puede escribirse:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

donde la matriz \mathbf{Z} , de dimensión $Tm \times m$, consta de columnas de unos y ceros que acompañan al vector $m \times 1$ formado por los términos independientes $\alpha_1, \alpha_2, \dots, \alpha_m$:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_T & \mathbf{0}_T & \mathbf{0}_T & \cdots & \mathbf{0}_T \\ \mathbf{0}_T & \mathbf{1}_T & \mathbf{0}_T & \cdots & \mathbf{0}_T \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_T & \mathbf{0}_T & \mathbf{0}_T & \cdots & \mathbf{1}_T \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdots \\ \mathbf{X}_m \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \cdots \\ \mathbf{y}_m \end{pmatrix}$$

mientras que \mathbf{X} denota la matriz $Tm \times (k - 1)$ de observaciones del resto de las variables explicativas, ordenadas como en el caso anterior, ya que, al igual que en el modelo M1, las pendientes del modelo se suponen iguales para las diferentes submuestras.

Las ecuaciones normales del modelo son, por tanto:

$$\begin{cases} (\mathbf{Z}'\mathbf{Z})\hat{\boldsymbol{\alpha}} + (\mathbf{Z}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{y} \\ (\mathbf{X}'\mathbf{Z})\hat{\boldsymbol{\alpha}} + (\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \end{cases}$$

⁽³⁾ Con independencia de la descripción que hemos hecho, interesante desde el punto de vista metodológico, el lector debe observar que el estimador MCO podría también obtenerse si transformamos las variables, mediante:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (\sum_1^m \mathbf{W}_i' \mathbf{W}_i)^{-1} (\sum_1^m \mathbf{W}_i' \mathbf{y}_i)$$

donde \mathbf{W}_i denota la matriz $(\mathbf{1}_T; \mathbf{X}_i)$, de dimensión $T \times k$.

Con esta notación se tiene que $(\mathbf{Z}'\mathbf{Z})^{-1}$ es una matriz diagonal $m \times m$, con elementos $\frac{1}{T}, \frac{1}{T}, \dots, \frac{1}{T}$, y los productos:

$$\mathbf{Z}'\mathbf{X} = \begin{pmatrix} \sum_1^T x_{11t} & \sum_1^T x_{12t} & \dots & \sum_1^T x_{1kt} \\ \sum_1^T x_{21t} & \sum_1^T x_{22t} & \dots & \sum_1^T x_{2kt} \\ \dots & \dots & \dots & \dots \\ \sum_1^T x_{m1t} & \sum_1^T x_{m2t} & \dots & \sum_1^T x_{mkt} \end{pmatrix} = \begin{pmatrix} T\bar{x}'_1 \\ T\bar{x}'_2 \\ \dots \\ T\bar{x}'_m \end{pmatrix} \quad \mathbf{Z}'\mathbf{y} = \begin{pmatrix} T\bar{y}_1 \\ T\bar{y}_2 \\ \dots \\ T\bar{y}_m \end{pmatrix}$$

Con ello, las ecuaciones normales pueden escribirse:

$$\begin{aligned} \hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \dots \\ \hat{\alpha}_m \end{pmatrix} &= (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\hat{\beta}) = \left[\text{diag} \left(\frac{1}{T}, \frac{1}{T}, \dots, \frac{1}{T} \right) \right] \cdot \left[\begin{pmatrix} T\bar{y}_1 \\ T\bar{y}_2 \\ \dots \\ T\bar{y}_m \end{pmatrix} - \begin{pmatrix} T\bar{x}'_1\hat{\beta} \\ T\bar{x}'_2\hat{\beta} \\ \dots \\ T\bar{x}'_m\hat{\beta} \end{pmatrix} \right] = \\ &= \begin{pmatrix} (T\bar{y}_1 - T\bar{x}'_1\hat{\beta})/T \\ (T\bar{y}_2 - T\bar{x}'_2\hat{\beta})/T \\ \dots \\ (T\bar{y}_m - T\bar{x}'_m\hat{\beta})/T \end{pmatrix} = \begin{pmatrix} \bar{y}_1 - \bar{x}'_1\hat{\beta} \\ \bar{y}_2 - \bar{x}'_2\hat{\beta} \\ \dots \\ \bar{y}_m - \bar{x}'_m\hat{\beta} \end{pmatrix} \end{aligned} \quad [8.9]$$

de donde es fácil ver que:

$$\text{Var } \hat{\alpha}_i = \frac{\sigma_u^2}{T} + \bar{x}'_i(\text{Var } \hat{\beta})\bar{x}_i$$

de modo que cada término independiente α_i se obtiene a partir de las estimaciones del mismo vector β , aunque, a diferencia de [8.9], en este caso se utilizan para cada submuestra sus correspondientes momentos muestrales, \bar{x}_i e \bar{y}_i .

Por otra parte, el segundo bloque de ecuaciones normales es:

$$(T\bar{x}'_1; T\bar{x}'_2; \dots; T\bar{x}'_m) \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \dots \\ \hat{\alpha}_m \end{pmatrix} + \mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

y, utilizando [8.6]:

$$\sum_1^m T\bar{x}_i\bar{y}_i - \sum_1^m T(\bar{x}_i\bar{x}'_i)\hat{\beta} + (\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{y}$$

o, lo que es lo mismo:

$$\Sigma_1^m [X_i' X_i - T(\bar{x}_i \bar{x}_i')] \hat{\beta} = \Sigma_1^m (X_i' y_i - T \bar{x}_i \bar{y}_i)$$

o, finalmente:

$$\Sigma_1^m [(X_i - \mathbf{1}_T \bar{x}_i)' (X_i - \mathbf{1}_T \bar{x}_i)] \hat{\beta} = \Sigma_1^m (X_i - \mathbf{1}_T \bar{x}_i)' (y_i - \mathbf{1}_T \bar{y}_i)$$

que equivale a transformar las variables *en diferencias con respecto a sus promedios dentro de cada submuestra*, antes de estimar el vector β , común a todas las submuestras, por mínimos cuadrados utilizando *todas* las observaciones muestrales. Los términos independientes de cada submuestra se obtienen posteriormente, del modo señalado en [8.9] donde ya se hacen intervenir medias sub muestrales, lo que es consistente con la heterogeneidad de los términos independientes. La matriz de covarianzas del vector β es:

$$\text{Var}(\hat{\beta}) = \sigma_u^2 (\Sigma_1^m (X_i - \mathbf{1}_T \bar{x}_i)' (X_i - \mathbf{1}_T \bar{x}_i))^{-1}$$

y la estimación del parámetro σ_u^2 se obtendría dividiendo la suma residual por el número de grados de libertad $(T - 1)m - (k - 1)$:

$$\hat{\sigma}_u^2 = \frac{(\Sigma_1^m (y_i - \mathbf{1}_T \bar{y}_i)' (y_i - \mathbf{1}_T \bar{y}_i) - \Sigma_1^m \hat{\beta}' (X_i - \mathbf{1}_T \bar{x}_i)' (y_i - \mathbf{1}_T \bar{y}_i))}{(T - 1)m - (k - 1)}$$

donde debe apreciarse la utilización de medias sub muestrales.

Por supuesto que hay una alternativa a este procedimiento, que consiste en estimar por mínimos cuadrados el modelo original directamente, estimando los términos independientes a la vez que las «pendientes». Bajo el supuesto de que no hay heteroscedasticidad ni autocorrelación, el modelo se estimaría por MCO, para lo que las matrices relevantes son:

$$X'X = \begin{pmatrix} T & & & & \Sigma_1^T x_{12t} & \dots & \Sigma_1^T x_{1kt} \\ & T & & & \Sigma_1^T x_{22t} & \dots & \Sigma_1^T x_{2kt} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & T & \Sigma_1^T x_{m2t} & \dots & \Sigma_1^T x_{mkt} \\ \Sigma_1^T x_{12t} & \Sigma_1^T x_{22t} & \dots & \Sigma_1^T x_{m2t} & & & \\ \dots & \dots & \dots & \dots & & & \Sigma_1^m X_i' X_i \\ \Sigma_1^T x_{1kt} & \Sigma_1^T x_{2kt} & \dots & \Sigma_1^T x_{mkt} & & & \end{pmatrix}$$

$$X'y = \begin{pmatrix} \Sigma_1^T y_{1t} \\ \Sigma_1^T y_{2t} \\ \dots \\ \Sigma_1^T y_{mt} \\ \Sigma_1^m X_i' y_i \end{pmatrix}$$

Nótese que la matriz $\mathbf{X}'\mathbf{X}$ es $(m + k - 1) \times (m + k - 1)$, lo que indica el número de coeficientes que podremos estimar mediante el método de mínimos cuadrados, que serían m términos independientes (diferentes para las distintas submuestras) y $k - 1$ pendientes (comunes a todas las submuestras). Sin embargo, este número de parámetros pudiera llegar a ser grande si hay un número alto de submuestras; por eso es interesante la transformación de variables que hemos presentado, ya que en ella sólo se estiman $k - 1$ parámetros, para recuperar después las estimaciones de los términos independientes.

El lector debe asegurarse de que entiende las similitudes y diferencias entre los modelos M2 y M3.

8.5. ESTIMACION CON MATRICES DE COVARIANZAS DIFERENTES

Manteniendo el supuesto de que no hay correlación contemporánea entre los términos de error de ecuación diferentes, pero admitiendo que las matrices de covarianzas de los términos de error de cada ecuación puedan ser diferentes entre sí, vamos a discutir en esta sección los cambios que deberían introducirse en la estimación de los modelos M1 a M3 con respecto a la presentación efectuada en la sección previa. Conviene adelantar que el grado de complejidad que introduce esta generalización es suficientemente grande como para que no podamos obtener expresiones analíticas sencillas para los estimadores en todos los casos.

Para simplificar la discusión vamos a distinguir entre el caso en que las matrices de covarianzas de los vectores \mathbf{u}_i son todas del tipo $\sigma_i^2 \mathbf{I}_T$, aunque con parámetros σ_i^2 diferentes, del caso más general en que $\text{Var}(\mathbf{u}_i) = \sigma_i^2 \Sigma_i$, con $\Sigma_i \neq \mathbf{I}_T$. Establecemos esta distinción sólo por interés pedagógico, pues claramente, la primera de estas situaciones es un caso particular de la segunda cuando las matrices Σ_i son todas matrices identidad.

Caso 1. $\text{Var}(\mathbf{u}_i) = \sigma_i^2 \mathbf{I}_T, i = 1, 2, \dots, m.$

El tratamiento del modelo M1 es sencillo, pues, como vimos en la sección anterior, basta estimar con los datos correspondientes a cada submuestra por separado. Notemos además que, al estimar cada submuestra por separado, la heterogeneidad de los parámetros σ_i^2 no presenta ninguna dificultad, puesto que sólo utilizamos observaciones de una submuestra y todas ellas tienen el mismo valor del parámetro σ_u^2 . En consecuencia, no es preciso utilizar MCG, sino que basta con estimar por MCO cada una de ellas. La única diferencia respecto a la Sección 8.4 es que ahora cada parámetro σ_i^2 se estima con los datos de su propia submuestra.

Al ser α y β comunes a las distintas submuestras en el modelo M2, hay que estimar éste simultáneamente, al igual que en la sección anterior. Sin más que utilizar la expresión matricial del estimador MCG en esta situación, puede obtenerse fácilmente la siguiente expresión:

$$\begin{pmatrix} \hat{\alpha}_{\text{MCG}} \\ \hat{\beta}_{\text{MCG}} \end{pmatrix} = \left(\sum_1^m \frac{\mathbf{W}'_i \mathbf{W}_i}{\sigma_i^2} \right)^{-1} \left(\sum_1^m \frac{\mathbf{W}'_i \mathbf{y}_i}{\sigma_i^2} \right)$$

y también que:

$$\text{Var} \begin{pmatrix} \hat{\alpha}_{\text{MCG}} \\ \hat{\beta}_{\text{MCG}} \end{pmatrix} = \left(\sum_1^m \frac{\mathbf{W}'_i \mathbf{W}_i}{\sigma_i^2} \right)^{-1}$$

Los parámetros σ_i^2 , desconocidos, deberían estimarse previamente dividiendo la suma residual procedente de una estimación inicial, por MCO, en la submuestra correspondiente, para obtener:

$$\hat{\sigma}_i^2 = \frac{\mathbf{y}'_i \mathbf{y}_i - \hat{\beta}'_i \mathbf{W}'_i \mathbf{y}_i}{T - k}$$

Si se dividen las observaciones procedentes de cada submuestra por el valor correspondiente de $\hat{\sigma}_i$ y se utiliza MCO con las matrices \mathbf{W}_i e \mathbf{y}_i así transformadas, se obtiene el estimador MCG que aparece arriba.

El modelo M3 se trataría del mismo modo, con la diferencia de que en este último caso la matriz a invertir sería $(m + k - 1) \times (m + k - 1)$, en vez de $k \times k$ como ocurre en el modelo M2.

Caso 2. Si ahora permitimos la existencia de heteroscedasticidad y autocorrelación en cada ecuación, pero no correlación contemporánea, la matriz de covarianzas del vector $\mathbf{u}' = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_m)$ sería diagonal a bloques, con bloques dados por $\sigma_1^2 \Sigma_1, \sigma_2^2 \Sigma_2, \dots, \sigma_m^2 \Sigma_m$.

La estimación del modelo M1 es, de nuevo, sencilla en esta situación, pudiendo obtenerse a partir de la expresión matricial:

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\beta}_1 \\ \dots \\ \hat{\alpha}_m \\ \hat{\beta}_m \end{pmatrix} = \begin{pmatrix} \mathbf{W}'_1 (\Sigma_1^{-1} / \sigma_1^2) \mathbf{W}_1 & & & & \\ & \mathbf{W}'_2 (\Sigma_2^{-1} / \sigma_2^2) \mathbf{W}_2 & & & \\ & & \dots & & \\ & & & \dots & \\ & & & & \mathbf{W}'_m (\Sigma_m^{-1} / \sigma_m^2) \mathbf{W}_m \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{W}'_1 (\Sigma_1^{-1} / \sigma_1^2) \mathbf{y}_1 \\ \mathbf{W}'_2 (\Sigma_2^{-1} / \sigma_2^2) \mathbf{y}_2 \\ \dots \\ \dots \\ \mathbf{W}'_m (\Sigma_m^{-1} / \sigma_m^2) \mathbf{y}_m \end{pmatrix} =$$

$$= \begin{pmatrix} (\mathbf{W}'_1 \Sigma_1^{-1} \mathbf{W}_1)^{-1} \mathbf{W}'_1 \Sigma_1^{-1} \mathbf{y}_1 \\ (\mathbf{W}'_2 \Sigma_2^{-1} \mathbf{W}_2)^{-1} \mathbf{W}'_2 \Sigma_2^{-1} \mathbf{y}_2 \\ \dots \\ \dots \\ (\mathbf{W}'_m \Sigma_m^{-1} \mathbf{W}_m)^{-1} \mathbf{W}'_m \Sigma_m^{-1} \mathbf{y}_m \end{pmatrix}$$

es decir, que equivale a la estimación MCG dentro de cada submuestra por separado. La matriz de covarianzas del par $(\hat{\alpha}_i, \hat{\beta}_i)$ es $[\mathbf{W}'_i (\Sigma_i^{-1} / \sigma_i^2) \mathbf{W}_i]^{-1}$.

Es interesante observar que, al igual que en el caso de los estimadores MCO, las matrices de covarianzas de los estimadores MCG de los vectores β_i en el modelo M1 dependen solamente de los valores de las variables X en su submuestra correspondiente, a diferencia de lo que ocurre en los modelos M2 y M3.

En el caso del modelo M2, el estimador MCG del vector (α, β) , común para todas las submuestras, se obtiene de la expresión:

$$\begin{pmatrix} \hat{\alpha}_{\text{MCG}} \\ \hat{\beta}_{\text{MCG}} \end{pmatrix} = \left[\mathbf{W}'_1 \left(\frac{\Sigma_1^{-1}}{\sigma_1^2} \right) \mathbf{W}_1 + \dots + \mathbf{W}'_m \left(\frac{\Sigma_m^{-1}}{\sigma_m^2} \right) \mathbf{W}_m \right]^{-1} \cdot \left[\mathbf{W}'_1 \left(\frac{\Sigma_1^{-1}}{\sigma_1^2} \right) \mathbf{y}_1 + \dots + \mathbf{W}'_m \left(\frac{\Sigma_m^{-1}}{\sigma_m^2} \right) \mathbf{y}_m \right]$$

Su matriz de covarianzas es:

$$\text{Var}(\hat{\alpha}_{\text{MCG}}; \hat{\beta}_{\text{MCG}}) = \left[\mathbf{W}'_1 \left(\frac{\Sigma_1^{-1}}{\sigma_1^2} \right) \mathbf{W}_1 + \dots + \mathbf{W}'_m \left(\frac{\Sigma_m^{-1}}{\sigma_m^2} \right) \mathbf{W}_m \right]^{-1}$$

Al tratar de obtener estas estimaciones en la práctica, de nuevo nos encontramos con que las matrices de covarianzas $\Sigma_1, \Sigma_2, \dots, \Sigma_m$, así como los parámetros $\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$, son desconocidos, por lo que habría que estimar inicialmente por MCO en cada submuestra. Con los residuos así generados, se estimarían todos ellos, utilizando sus valores estimados para la obtención del estimador MCG, bien mediante su expresión matricial, o bien mediante una apropiada transformación del modelo y aplicación posterior de MCO. En cualquier caso, como es habitual, éste debería ser un procedimiento iterativo hasta conseguir la convergencia del mismo.

El lector debe comprobar que, efectivamente, los procedimientos de estimación sugeridos en el Caso 1 son caso particular de los sugeridos en el Caso 2, cuando $\Sigma_i = \mathbf{I}_T, i = 1, 2, \dots, m$.

8.6. ESTIMACION CUANDO LOS TERMINOS DE ERROR DE LAS DISTINTAS ECUACIONES ESTAN RELACIONADOS

El tratamiento analítico de esta situación es bastante más complejo que el de las tratadas en la sección anterior. A modo de ejemplo, consideremos una situación con dos submuestras cuyos términos de error tienen correlación contemporánea. Este sería el caso en que se pretende estimar un modelo de determinación de la tasa de inflación, disponiendo de observaciones procedentes de dos países, siendo de esperar que hayan existido perturbaciones en los precios que hayan afectado simultáneamente a ambos países, así como perturbaciones específicas de cada país.

En tal caso, la matriz de covarianzas del término de error, de dimensión $2T \times 2T$, es:

$$\Sigma = \begin{pmatrix} \sigma_1^2 \mathbf{I}_T & \sigma_{12} \mathbf{I}_T \\ \sigma_{12} \mathbf{I}_T & \sigma_2^2 \mathbf{I}_T \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \otimes \mathbf{I}_T$$

por lo que su inversa es la matriz:

$$\Sigma^{-1} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}^{-1} \otimes \mathbf{I}_T = k \begin{pmatrix} \sigma_2^2 \mathbf{I}_T & -\sigma_{12} \mathbf{I}_T \\ -\sigma_{12} \mathbf{I}_T & \sigma_1^2 \mathbf{I}_T \end{pmatrix}$$

donde la constante k es igual a $(\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)^{-1}$. Utilizando esta expresión, sería ahora sencillo obtener el estimador mínimo-cuadrático

$$\hat{\beta} = (\mathbf{W}' \Sigma^{-1} \mathbf{W})^{-1} \mathbf{W}' \Sigma^{-1} \mathbf{y}$$

del modelo M2, por ejemplo, en cuyo caso:

$$\mathbf{W} = \begin{pmatrix} \mathbf{1}_T & \mathbf{X}_1 \\ \mathbf{1}_T & \mathbf{X}_2 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$$

A pesar de la sencillez teórica de este modelo, es de gran utilidad práctica, por lo que se incluye al final del capítulo un amplio grupo de ejercicios similares al que aquí hemos analizado.

8.7. COEFICIENTES VARIANDO EN EL TIEMPO

Supongamos ahora que se dispone de una muestra con una dimensión temporal pequeña y un número grande de observaciones dentro de cada sección cruzada. En tal situación, interesaría hacer el supuesto dual del que hasta ahora hemos venido considerando: los coeficientes del modelo de regresión son los mismos para todas las unidades de observación (países, individuos, etcétera) en un período dado, pero son distintos para periodos diferentes. En tal caso, formularíamos modelos del tipo:

$$y_{it} = \alpha_t + \mathbf{x}'_{it} \beta_t + u_{it}, \quad i = 1, 2, \dots, m, \quad t = 1, 2, \dots, T \quad [8.10]$$

donde, a su vez, podemos suponer como caso particular que, bien el término independiente del modelo α_t , o los coeficientes de las variables explicativas β_t , son constantes a lo largo del tiempo. Ello daría lugar a distintas versiones restringidas del modelo general anterior. Dichas versiones se corresponderían con los modelos M1, M2, M3 de la Sección 8.2. En este caso, tiene sentido suponer que se dispone de un número igual de observaciones en cada una de las submuestras, y de ahí el valor común T del tamaño de cada submuestra en [8.10].

Así, podríamos considerar tres contrastes análogos a los analizados en la Sección 8.3. Por ejemplo, el modelo M2' sería aquel que incorporase el

supuesto de homogeneidad global, imponiendo la restricción de que tanto el término independiente α como los coeficientes β son constantes en el tiempo. El modelo M3' sería aquel que mantiene el supuesto de que las «pendientes» son constantes en el tiempo, pero deja variar al término independiente; finalmente, el modelo M1' permitiría que tanto las pendientes como el término independiente variasen en el tiempo, coincidiendo con la especificación general [8.10].

Los métodos a utilizar en la estimación de estos modelos son asimismo totalmente análogos a los empleados en el caso en que los coeficientes dependían de las unidades de observación muestrales. Así, el modelo M2' se estimaría de una vez, con todas las observaciones muestrales simultáneamente, el modelo M3' se estimaría con las variables en diferencias con respecto a la media, para más tarde recuperar los términos independientes del modo habitual.

La diferencia importante sería que, en este caso, las medias muestrales con respecto a las que se toman las diferencias son los promedios de todas las observaciones de un período dado, por lo que la expresión para recuperar los términos independientes en el modelo M3' sería:

$$\hat{\alpha}_t = y_t - \bar{x}_{1t}\hat{\beta}_{1t} - \dots - \bar{x}_{kt}\hat{\beta}_{kt}$$

donde el subíndice temporal de las estimaciones de los coeficientes β denotan el hecho de que, en este caso, dichas estimaciones cambian en el tiempo y

$$\bar{x}_{kt} = \left(\frac{1}{m}\right) \sum_{j=1}^m x_{jkt}$$

Finalmente, el modelo M1' también se estimaría tantas veces como períodos muestrales hay, utilizando cada vez sólo las observaciones de un período determinado.

Los contrastes de hipótesis pueden llevarse a cabo mediante comparaciones entre las sumas residuales restringida y sin restringir, del modo que ya hemos visto en repetidas ocasiones. Es importante tener en cuenta en tal caso el número de grados de libertad de cada una de dichas sumas residuales, que para los tres modelos que hemos mencionado son $(m-k)T$, $mT-k$, $(m-1)T-(k-1)$.

Ejemplo. Para completar la discusión de la Sección 6.7, añadimos a los datos de 1989 del PIB y el Empleo de las Comunidades Autónomas españolas, la información similar correspondiente a 1985. Las cifras de PIB de 1989 se transformaron en pesetas de 1985. La estimación del modelo en ambos períodos mediante MCO es:

Año 1989

$$(a) \quad \text{PIB}_t = -187,68 + 3,759 \cdot \text{Empleo}_t$$

(182,92) (0,185)

$$\bar{R}^2 = 0,96; \quad \text{SR} = 4.307.096; \quad \hat{\sigma} = 518,84; \quad N = 18$$

Año 1985

$$(b) \quad \text{PIB}_i = -99,90 + 1,929 \cdot \text{Empleo}_i$$

$$(95,60) \quad (0,110)$$

$$\bar{R}^2 = 0,95; \quad \text{SR} = 1.129.030; \quad \hat{\sigma} = 265,64; \quad N = 18$$

que ilustran claramente cómo la contribución del Empleo al PIB, en el promedio de las Comunidades Autónomas, fue muy superior en 1985 que en 1989.

El investigador puede tener interés en contrastar la hipótesis de que la relación PIB/Empleo no varió significativamente de un año a otro, es decir: $H_0: \beta_{1985} = \beta_{1989}, \alpha_{1985} = \alpha_{1989}$. Bajo tal supuesto, podrían utilizarse los datos de ambos años para estimar un único modelo, que sería el modelo restringido (MR).

Si se procede a la estimación del modelo utilizando todas las observaciones simultáneamente, se tiene:

$$(c) \quad \text{PIB}_i = -241,52 + 3,034 \cdot \text{Empleo}_i$$

$$(232,15) \quad (0,249)$$

$$\bar{R}^2 = 0,81; \quad \text{SR} = 29.071.732; \quad \hat{\sigma} = 924,69; \quad N = 36$$

El modelo sin restringir (MSR) consiste en las estimaciones (a) y (b), por separado para cada submuestra, y la suma residual sin restringir (SRS) es el agregado de las sumas residuales de ambas regresiones. Así, tenemos el estadístico F :

$$\frac{(\text{SRR} - \text{SRS})/q}{\text{SRS}/\text{gdl}} = \frac{[29.071.732 - (1.129.030 + 4.307.096)]/2}{(1.129.030 + 4.307.096)/32} \sim F_{2,32}$$

donde gdl denota el número de grados de libertad del MSR, que es $(T_1 + T_2) - 2k = 18 + 18 - 4 = 32$.

En definitiva, el valor numérico del estadístico es 69,57, que excede claramente los valores críticos de la distribución $F_{2,32}$ tanto al 95 como al 99 por 100, que son 3,30 y 5,35 respectivamente, por lo que rechazamos la hipótesis nula de igualdad temporal de los coeficientes del modelo.

Sin embargo, el investigador debe reconocer que la constante del modelo puede diferir de un año a otro solamente porque las medias muestrales de ambas variables difieran de 1985 a 1989. En definitiva, la hipótesis que es verdaderamente relevante es $H_0: \beta_{1985} = \beta_{1989}$, permitiendo que el valor de α cambie de un año a otro.

Para llevar a cabo este contraste de hipótesis, definimos una variable ficticia D_{it} que tome el valor 1 en las 18 observaciones correspondientes a 1989 y estimamos el modelo:

$$(d) \quad \text{PIB}_{it} = 387,10 - 1.176,26D_{it} + 2,976 \cdot \text{Empleo}_{it}$$

$$(218,90) \quad (237,33) \quad (0,192)$$

$$\bar{R}^2 = 0,89; \quad \text{SR} = 16.666.016; \quad \hat{\sigma} = 710,66; \quad N = 36$$

que conduce al estadístico F para el contraste de la hipótesis nula:
 $H_0: \beta_{1985} = \beta_{1989}$:

$$\frac{(\text{SRR} - \text{SRS})/q}{\text{SRS/gdl}} = \frac{[16.666.016 - (1.129.030 + 4.307.096)]/1}{(1.129.030 + 4.307.096)/32} = 66,11$$

que conduce nuevamente al rechazo de H_0 .

El que las constantes de los modelos (a) y (c) sean no significativas hace que los valores de los estadísticos F para estos dos contrastes sean muy similares.

Sin embargo, los valores estimados de $\hat{\sigma}$ en los modelos (a) y (b) son suficientemente diferentes como para creer que existe heteroscedasticidad temporal que puede estar sesgando el contraste. Para evitar tal problema, multiplicamos las observaciones de PIB y Empleo de 1985 por el factor $\hat{\sigma}_{1985}/\hat{\sigma}_{1989}$ y volvimos a estimar el modelo para 1985, obteniendo:

$$(a) \quad \text{PIB}_i = -96,09 + 3,759 \cdot \text{Empleo}_i$$

(93,66) (0,185)

$$\bar{R}^2 = 0,96; \quad \text{SR} = 1.129.030; \quad \hat{\sigma} = 265,64; \quad N = 18$$

donde debe observarse la natural coincidencia de valores de SR y $\hat{\sigma}$ con la regresión (b).

La estimación restringida con estos datos transformados resultó ser:

$$(c) \quad \text{PIB}_i = 126,07 + 2,160 \cdot \text{Empleo}_i$$

(149,73) (0,210)

$$\bar{R}^2 = 0,75; \quad \text{SR} = 12.976.630; \quad \hat{\sigma} = 617,80; \quad N = 36$$

El estadístico F para el contraste de $H_0: \alpha_{1985} = \alpha_{1989}; \beta_{1985} = \beta_{1989}$ es ahora:

$$\frac{(\text{SRR} - \text{SRS})/q}{\text{SRS/gdl}} = \frac{[12.976.630 - 2(1.129.030)]/2}{2(1.129.030)/32} = 75,95$$

que conduce nuevamente a rechazar H_0 .

Si nos limitamos a contrastar la hipótesis $H_0: \beta_{1985} = \beta_{1989}$, tenemos el modelo restringido:

$$(e) \quad \text{PIB}_{it} = 413,13 - 825,68 D_{it} + 2,404 \cdot \text{Empleo}_{it}$$

(127,87) (164,32) (0,168)

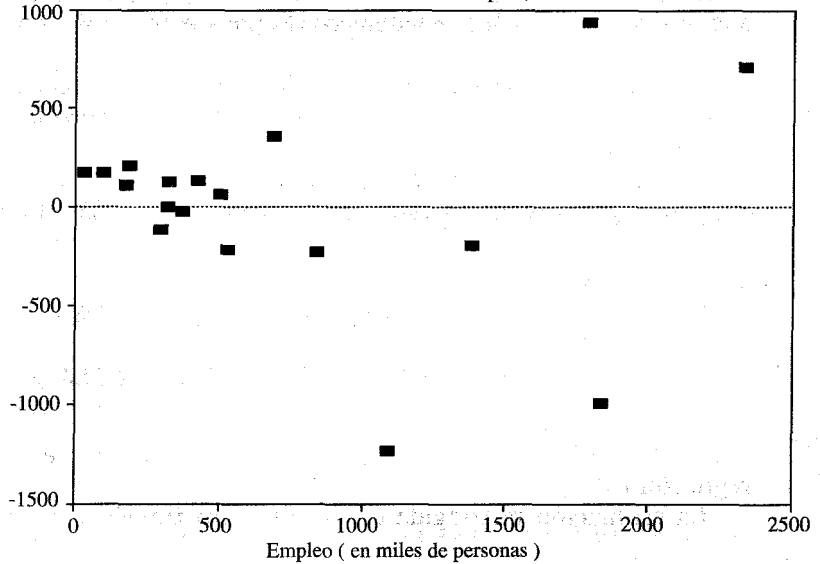
$$\bar{R}^2 = 0,85; \quad \text{SR} = 7.351.839; \quad \hat{\sigma} = 472,0; \quad N = 36$$

con un estadístico F :

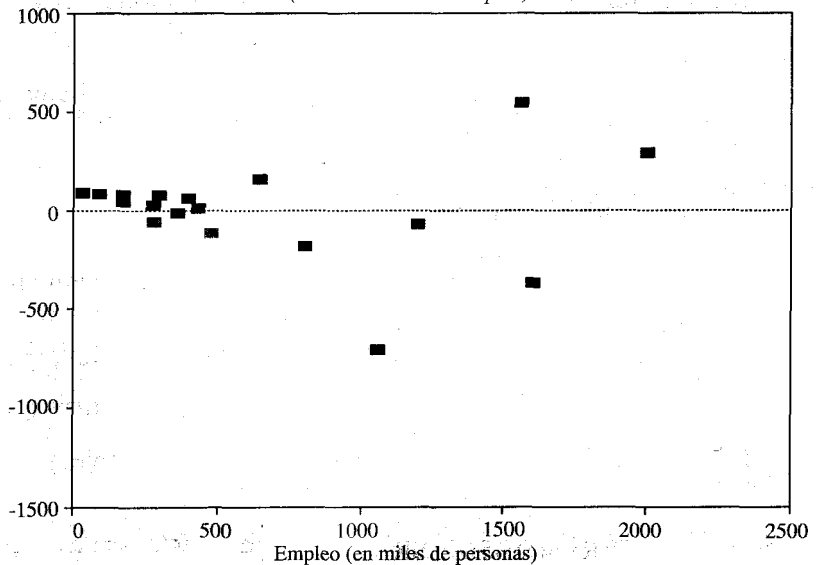
$$\frac{(SRR - SRS)/q}{SRS/gdl} = \frac{[7.351.839 - 2(1.129.030)]/1}{2(1.129.030)/32} = 72,19$$

que implica el rechazo de la hipótesis de homogeneidad temporal de β . El investigador se queda finalmente con las estimaciones (a)-(b).

RESIDUOS PIB-EMPLEO: 1985
(Como función del Empleo)



RESIDUOS PIB-EMPLEO: 1989
(Como función del Empleo)



Parte II

REGRESIONES APARENTEMENTE NO RELACIONADAS

8.8. DESCRIPCIÓN DEL MODELO

A semejanza del modelo que acabamos de analizar, tratamos ahora el caso de un sistema de ecuaciones en el que todas las variables explicativas son deterministas. A diferencia del caso anterior, se trata de ecuaciones que, generalmente, explican variables endógenas diferentes, y se utiliza una única muestra. Algunos ejemplos son:

1. Se quiere estimar la demanda de un conjunto de bienes duraderos, para lo que se dispone de datos de ventas y precios de mercado de cada uno de ellos, así como también de algunas variables de coyuntura: tipos de interés, índices de producción y empleo, tasas de inflación.

En este ejemplo podría pensarse que las elasticidades precio y renta de algunos bienes podrían ser iguales. Ello sugeriría la estimación conjunta de sus funciones de demanda, a efectos de imponer las restricciones de igualdad entre dichos coeficientes. Pero, incluso si se está convencido a priori de que los coeficientes de las funciones de demanda de cada producto son diferentes, tiene sentido pensar que existen factores estructurales (estocásticos) que en un período dado pueden afectar a los términos de error de las distintas ecuaciones de demanda de forma similar, por lo que dichos términos estarán correlacionados contemporáneamente.

2. Considerando la tasa de crecimiento de la oferta monetaria ΔM_t y del stock de Deuda Pública ΔS_t como variables exógenas, se quiere estudiar su efecto sobre la determinación de la tasa de inflación π_t y los tipos de interés r_t , para lo que se especifican ecuaciones:

$$\begin{aligned}\pi_t &= \beta_1 + \beta_2 \pi_{t-1} + \beta_3 \Delta M_t + \beta_4 \Delta S_t + u_t \\ r_t &= \delta_1 + \delta_2 \Delta S_t + \delta_3 \Delta M_t + v_t\end{aligned}$$

Se piensa que hay otros determinantes estructurales, además de las políticas fiscal y monetaria, que pueden tener incidencia simultáneamente sobre la determinación del índice de precios y los tipos de interés, pero no se dispone de suficiente información como para hacer una especificación explícita de dichos factores que permitiera incluirlos como variables explicativas.

3. Para estimar el efecto inflacionario de los shocks producidos en los precios del petróleo en los últimos quince años, se dispone de series temporales de producción, índices de precios, tipos de interés y precios energéticos para distintos países a lo largo del período 1960-1987. Se piensa que los efectos sufridos por los diversos países pueden ser diferentes, pero querríamos contrastar tal hipótesis. Por otra parte, aun si dichos efectos fuesen realmente distintos, se cree que hay efectos sobre la variación de precios a nivel mundial

que no pueden recogerse simplemente con la evolución de los precios energéticos, por lo que la estimación simultánea de las ecuaciones de inflación para este conjunto de países tiene interés.

8.9. ESTIMACION DE UN CONJUNTO DE REGRESIONES APARENTEMENTE NO RELACIONADAS

Por constar de un sistema de ecuaciones relacionadas a través de las correlaciones entre sus términos de error, sin que haya otra conexión explícita, este modelo se denomina como de *regresiones aparentemente no relacionadas*.

En este modelo es fundamental la estructura de correlaciones entre los vectores de error de las distintas ecuaciones. Así, una parte importante del estudio sería la modelización explícita de las correlaciones en los términos de error de las ecuaciones de inflación de los distintos países en un mismo año (e incluso, posiblemente, correlaciones retardadas).

Por tanto, la especificación que se haga de la matriz de covarianzas del vector ampliado que se forma con los términos de error de todas las ecuaciones tendrá gran importancia para la estimación del modelo. Un supuesto sencillo acerca de la estructura de dicha matriz de covarianzas es el formado por las hipótesis:

a) En cada ecuación, el término de error no presenta autocorrelación ni heteroscedasticidad, es decir:

$$E(\mathbf{u}_i \mathbf{u}_i') = \sigma_i^2 \mathbf{I}_T, \quad i = 1, 2, \dots, m$$

donde m es el número de ecuaciones del modelo. Puede apreciarse que admitimos valores diferentes para la varianza del término de error de las distintas ecuaciones. Suponemos que se dispone de igual número de observaciones, T , para todas las variables endógenas del modelo.

b) Con respecto a las covarianzas entre los errores de las distintas ecuaciones, suponemos que toda la correlación es contemporánea, lo que se puede expresar:

$$E(\mathbf{u}_i \mathbf{u}_j') = \sigma_{ij} \mathbf{I}_T, \quad i \neq j, \quad i, j = 1, 2, \dots, m$$

lo que recoge un supuesto doble. Por una parte, se está suponiendo que dicha covarianza es independiente del instante de tiempo considerado $E(u_{it} u_{jt}) = \sigma_{ij}$, $\forall t = 1, 2, \dots, T$. Por otro lado, se está admitiendo que $E(u_{it} u_{js}) = 0$ para todo $t \neq s$, $t, s = 1, 2, \dots, T$.

El modelo se puede representar: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde \mathbf{u} e \mathbf{y} son vectores de dimensión $Tm \times 1$ y \mathbf{X} una matriz $Tm \times \sum_1^m k_i$. La matriz \mathbf{X} es diagonal a bloques, donde cada bloque es la matriz \mathbf{X}_i correspondiente a cada ecuación. de modo que se tiene:

todas ellas de orden $T \times T$, mientras que el bloque (1, 2), por ejemplo, muestra la covarianza entre los vectores \mathbf{u}_1 y \mathbf{u}_2 .

La ausencia de autocorrelación dentro de cada ecuación implica que cada uno de los bloques a lo largo de la diagonal principal es una matriz diagonal. A su vez, el supuesto de homoscedasticidad en cada ecuación implica el que cada uno de esos bloques es realmente una matriz escalar. Cualquiera de estos supuestos podría abandonarse en un posterior proceso de generalización, con lo que se generarían a lo largo de la diagonal bloques con estructura más compleja. Por otra parte, el supuesto de que toda la correlación entre distintas ecuaciones es contemporánea hace que los bloques de fuera de la diagonal sean a su vez matrices escalares.

El estimador de mínimos cuadrados ordinarios del sistema de ecuaciones es:

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_m \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \\ &= \left[\begin{pmatrix} \mathbf{X}'_1 & & & \\ & \mathbf{X}'_2 & & \\ & & \dots & \\ & & & \mathbf{X}'_m \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & & & \\ & \mathbf{X}_2 & & \\ & & \dots & \\ & & & \mathbf{X}_m \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{X}'_1 & & & \\ & \mathbf{X}'_2 & & \\ & & \dots & \\ & & & \mathbf{X}'_m \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \dots \\ \mathbf{y}_m \end{pmatrix} = \\ &= \begin{pmatrix} \mathbf{X}'_1\mathbf{X}_1 & & & \\ & \mathbf{X}'_2\mathbf{X}_2 & & \\ & & \dots & \\ & & & \mathbf{X}'_m\mathbf{X}_m \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_1\mathbf{y}_1 \\ \mathbf{X}'_2\mathbf{y}_2 \\ \dots \\ \mathbf{X}'_m\mathbf{y}_m \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}_1 \\ (\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{y}_2 \\ \dots \\ (\mathbf{X}'_m\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{y}_m \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1^{\text{MCO}} \\ \hat{\beta}_2^{\text{MCO}} \\ \dots \\ \hat{\beta}_m^{\text{MCO}} \end{pmatrix} \end{aligned}$$

que coincide, por tanto, con el estimador de mínimos cuadrados ecuación por ecuación.

La matriz $\mathbf{\Omega}$ no es escalar y, por tanto, el estimador MCO no será eficiente. Es interesante notar que, incluso si se supiese que las varianzas (constantes) del término de error de todas las ecuaciones fuesen iguales, la matriz de covarianzas de \mathbf{u} no sería diagonal debido a que las correlaciones entre los términos de error generan bloques no nulos fuera de la diagonal principal. En consecuencia, el estimador MCO seguiría sin ser eficiente. Son precisamente las covarianzas entre los términos de error de las distintas ecuaciones las que hacen ineficiente a dicho estimador.

El estimador MCO ignora la información contenida en estas covarianzas entre observaciones contemporáneas de distintas ecuaciones. Este comportamiento sólo sería eficiente en el caso de que dichas covarianzas fuesen cero. Por ello, y por ganar generalidad, es que el caso de una matriz de covarianzas diagonal no se debe imponer como especificación inicial en un análisis de este

tipo, puesto que, siendo un caso particular del que aquí presentamos, siempre puede contrastarse después de una estimación inicial del modelo aquí propuesto.

La matriz de covarianzas anterior puede escribirse utilizando el producto de Kronecker:

$$\text{Var}(\mathbf{u}) = \mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T \tag{8.11}$$

donde $\mathbf{\Sigma}$ es la matriz $m \times m$:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1m} \\ & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2m} \\ & & \dots & \dots & \dots \\ & & & \dots & \dots \\ & & & & \sigma_m^2 \end{pmatrix}$$

En las condiciones que hemos supuesto sabemos que el estimador MCG es eficiente, y se tiene:

$$\begin{aligned} \hat{\beta}_{\text{MCG}} &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}) = [\mathbf{X}'(\mathbf{\Sigma} \otimes \mathbf{I}_T)^{-1}\mathbf{X}]^{-1}[\mathbf{X}'(\mathbf{\Sigma} \otimes \mathbf{I}_T)^{-1}\mathbf{y}] = \\ &= [\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1}[\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y}] \end{aligned}$$

con matriz de covarianzas:

$$\text{Var}(\hat{\beta}_{\text{MCG}}) = [\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1}$$

Utilizando las expresiones:

$$\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T) = \begin{pmatrix} \mathbf{X}'_1 & & & & \\ & \mathbf{X}'_2 & & & \\ & & \dots & & \\ & & & \dots & \\ & & & & \mathbf{X}'_m \end{pmatrix} \begin{pmatrix} \sigma^{11}\mathbf{I}_T & \sigma^{12}\mathbf{I}_T & \dots & \sigma^{1m}\mathbf{I}_T \\ & \sigma^{22}\mathbf{I}_T & \dots & \sigma^{2m}\mathbf{I}_T \\ & & \dots & \dots \\ & & & \dots \\ & & & & \sigma^{mm}\mathbf{I}_T \end{pmatrix}$$

donde el primer factor es una matriz $\sum_1^m k_i \times (Tm)$ y el segundo es una matriz $Tm \times Tm$. En ellas, σ^{ij} denota el elemento genérico (i, j) de la matriz $\mathbf{\Sigma}^{-1}$, que, así como $\mathbf{\Sigma}$, también es simétrica. Así, se tiene:

$$\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T) = \begin{pmatrix} \sigma^{11}\mathbf{X}'_1 & \sigma^{12}\mathbf{X}'_1 & \dots & \sigma^{1m}\mathbf{X}'_1 \\ \sigma^{21}\mathbf{X}'_2 & \sigma^{22}\mathbf{X}'_2 & \dots & \sigma^{2m}\mathbf{X}'_2 \\ \dots & \dots & \dots & \dots \\ \sigma^{m1}\mathbf{X}'_m & \sigma^{m2}\mathbf{X}'_m & \dots & \sigma^{mm}\mathbf{X}'_m \end{pmatrix}$$

$$\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X} = \begin{pmatrix} \sigma^{11} \mathbf{X}'_1 \mathbf{X}_1 & \sigma^{12} \mathbf{X}'_1 \mathbf{X}_2 & \dots & \sigma^{1m} \mathbf{X}'_1 \mathbf{X}_m \\ & \sigma^{22} \mathbf{X}'_2 \mathbf{X}_2 & \dots & \sigma^{2m} \mathbf{X}'_m \mathbf{X}_m \\ \dots & \dots & \dots & \dots \\ & & & \sigma^{mm} \mathbf{X}'_m \mathbf{X}_m \end{pmatrix} \quad [8.12]$$

y también:

$$\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y} = \begin{pmatrix} \sigma^{11} \mathbf{X}'_1 \mathbf{y}_1 + \sigma^{12} \mathbf{X}'_1 \mathbf{y}_2 + \dots + \sigma^{1m} \mathbf{X}'_1 \mathbf{y}_m \\ \sigma^{21} \mathbf{X}'_2 \mathbf{y}_1 + \sigma^{22} \mathbf{X}'_2 \mathbf{y}_2 + \dots + \sigma^{2m} \mathbf{X}'_2 \mathbf{y}_m \\ \dots \\ \sigma^{m1} \mathbf{X}'_m \mathbf{y}_1 + \sigma^{m2} \mathbf{X}'_m \mathbf{y}_2 + \dots + \sigma^{mm} \mathbf{X}'_m \mathbf{y}_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m \sigma^{1j} \mathbf{X}'_1 \mathbf{y}_j \\ \sum_{j=1}^m \sigma^{2j} \mathbf{X}'_2 \mathbf{y}_j \\ \dots \\ \sum_{j=1}^m \sigma^{mj} \mathbf{X}'_m \mathbf{y}_j \end{pmatrix}$$

de modo que:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = \begin{pmatrix} \sigma^{11} \mathbf{X}'_1 \mathbf{X}_1 & \sigma^{12} \mathbf{X}'_1 \mathbf{X}_2 & \dots & \sigma^{1m} \mathbf{X}'_1 \mathbf{X}_m \\ & \sigma^{22} \mathbf{X}'_2 \mathbf{X}_2 & \dots & \sigma^{2m} \mathbf{X}'_2 \mathbf{X}_m \\ \dots & \dots & \dots & \dots \\ & & & \sigma^{mm} \mathbf{X}'_m \mathbf{X}_m \end{pmatrix}^{-1} \begin{pmatrix} \sum_1 \sigma^{1j} \mathbf{X}'_1 \mathbf{y}_j \\ \sum_1 \sigma^{2j} \mathbf{X}'_2 \mathbf{y}_j \\ \dots \\ \sum_1 \sigma^{mj} \mathbf{X}'_m \mathbf{y}_j \end{pmatrix} \quad [8.13]$$

Este estimador no puede descomponerse por ecuaciones, como ocurría antes con el estimador de mínimos cuadrados ordinarios. La única excepción a esta afirmación se produce si $\sigma_{ij} = 0, \forall i \neq j$, pero en tal caso, el estimador de mínimos cuadrados generalizados se reduce, como veremos en el corolario a la Proposición 8.1, al estimador de mínimos cuadrados ordinarios y, como ya hemos visto, éste sí se descompone por ecuaciones.

Aquí surge de nuevo el problema habitual de desconocimiento de la matriz $\boldsymbol{\Sigma}$, que habrá que estimar previamente. Esta dificultad se resuelve estimando por MCO ecuación por ecuación, para después estimar los elementos de la matriz de covarianzas anterior por:

$$\hat{\sigma}_{ij} = \frac{\hat{\mathbf{u}}_i \hat{\mathbf{u}}_j}{(T - k_i)^{1/2} (T - k_j)^{1/2}}$$

donde $\hat{\mathbf{u}}_i$ es el T -vector de residuos de la i -ésima ecuación. Es importante observar que en este modelo se estiman tantos coeficientes como columnas tiene la matriz \mathbf{X} , que en este caso en que está organizada por bloques son $\sum_1^m k_i$.

La matriz de covarianzas estimada puede utilizarse para obtener los estimadores MCG de los coeficientes del modelo. A partir de ellos, se podría volver a generar una serie de residuos para cada una de las ecuaciones, y de ellos volver a estimar la matriz de covarianzas, para entrar así en un procedimiento iterativo, que se seguiría hasta alcanzar el criterio de convergencia

deseado. La utilización de MCO y no MCG en esta primera etapa es apropiada porque la única razón para necesitar MCG en la estimación del sistema era la covarianza entre los términos de error de las diferentes ecuaciones, mientras que hemos mantenido el supuesto de que la matriz de covarianzas de cada ecuación era escalar.

Hay que tener presente, sin embargo, que aunque no es el caso habitual, si la dimensión $m \times m$ de la matriz de covarianzas Σ es grande con respecto al tamaño muestral, entonces este procedimiento, siendo aún posible, deja de tener una interpretación estadística razonable, pues estaríamos estimando demasiados parámetros dada la información muestral de que disponemos. También cabe mencionar que, con independencia de que estimemos conjuntamente las ecuaciones del modelo, la estimación individual de las ecuaciones es siempre un ejercicio de interés, puesto que permite establecer comparaciones entre los valores de los coeficientes estimados bajo ambos procedimientos.

En cuanto a una generalización de la estructura de la matriz de covarianzas de los términos de error, es sencillo ver qué ocurre cuando se admite la presencia de autocorrelación y/o heteroscedasticidad dentro de cada muestra si se supone que no existe correlación entre distintas muestras, ni siquiera contemporáneamente. Entonces se tiene el siguiente resultado:

Proposición 8.1. Sea $\sigma_i^2 \Sigma_i$ la matriz de covarianzas del vector \mathbf{u}_i , $i = 1, 2, \dots, m$, no necesariamente diagonales. Si $\sigma_{ij} = 0$ para todo $i \neq j$, entonces el estimador MCG que se obtiene estimando simultáneamente todas las ecuaciones coincide con la utilización del estimador MCO para cada ecuación por separado.

Demostración. En tal caso, la matriz Ω es diagonal a bloques:

$$\Omega = \text{Var}(\mathbf{u}) = \text{Var} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \dots \\ \mathbf{u}_m \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \Sigma_1 & & & \\ & \sigma_2^2 \Sigma_2 & & \\ & & \dots & \\ & & & \sigma_m^2 \Sigma_m \end{pmatrix}$$

y puesto que ahora tanto la matriz \mathbf{X} como la matriz Ω son diagonal a bloques, se tiene:

$$\hat{\beta}_{\text{MCG}} = [\mathbf{X}' \Omega^{-1} \mathbf{X}]^{-1} [\mathbf{X}' \Omega^{-1} \mathbf{y}] =$$

$$= \left[\begin{pmatrix} \mathbf{X}'_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{X}'_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{X}'_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{X}'_m \end{pmatrix} \begin{pmatrix} (1/\sigma_1^2) \Sigma_1^{-1} & & & & \\ & (1/\sigma_2^2) \Sigma_2^{-1} & & & \\ & & (1/\sigma_3^2) \Sigma_3^{-1} & & \\ & & & \dots & \\ & & & & (1/\sigma_m^2) \Sigma_m^{-1} \end{pmatrix} \right]$$

$$\begin{aligned} & \left(\begin{array}{cccccc} \mathbf{X}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{X}_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{X}_m \end{array} \right)^{-1} \\ & \left[\left(\begin{array}{cccc} \mathbf{X}'_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{X}'_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{X}'_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{X}'_m \end{array} \right) \left(\begin{array}{cccc} (1/\sigma_1^2)\Sigma_1^{-1} & 0 & 0 & 0 \\ 0 & (1/\sigma_2^2)\Sigma_2^{-1} & 0 & 0 \\ 0 & 0 & (1/\sigma_3^2)\Sigma_3^{-1} & 0 \\ \dots & \dots & \dots & \dots \\ & & & & (1/\sigma_m^2)\Sigma_m^{-1} \end{array} \right) \left(\begin{array}{c} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \dots \\ \mathbf{y}_m \end{array} \right) \right] = \\ & = \left(\begin{array}{cccc} (1/\sigma_1^2)\mathbf{X}'_1\Sigma_1^{-1}\mathbf{X}_1 & 0 & 0 & 0 \\ 0 & (1/\sigma_2^2)\mathbf{X}'_2\Sigma_2^{-1}\mathbf{X}_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & (1/\sigma_m^2)\mathbf{X}'_m\Sigma_m^{-1}\mathbf{X}_m & 0 \end{array} \right)^{-1} \cdot \left(\begin{array}{c} (1/\sigma_1^2)\mathbf{X}'_1\Sigma_1^{-1}\mathbf{y}_1 \\ (1/\sigma_2^2)\mathbf{X}'_2\Sigma_2^{-1}\mathbf{y}_2 \\ \dots \\ (1/\sigma_m^2)\mathbf{X}'_m\Sigma_m^{-1}\mathbf{y}_m \end{array} \right) = \\ & = \left(\begin{array}{c} (\mathbf{X}'_1\Sigma_1^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma_1^{-1}\mathbf{y}_1 \\ (\mathbf{X}'_2\Sigma_2^{-1}\mathbf{X}_2)^{-1}\mathbf{X}'_2\Sigma_2^{-1}\mathbf{y}_2 \\ \dots \\ (\mathbf{X}'_m\Sigma_m^{-1}\mathbf{X}_m)^{-1}\mathbf{X}'_m\Sigma_m^{-1}\mathbf{y}_m \end{array} \right) = \left(\begin{array}{c} \hat{\beta}_1^{\text{MCG}} \\ \hat{\beta}_2^{\text{MCG}} \\ \dots \\ \hat{\beta}_m^{\text{MCG}} \end{array} \right) \end{aligned}$$

En el ejemplo de los efectos inflacionarios de los shocks registrados en los precios de los productos energéticos, las condiciones de esta proposición se dan si los shocks para dos países diferentes no tienen correlación ni en un mismo instante de tiempo ni tampoco con algún desfase temporal. Algo similar podría decirse en el caso de las demandas de un grupo de bienes duraderos.

Este es el resultado que veníamos anunciando, y reafirma la lógica con que iniciamos la discusión de este tema. Cuando no hay correlaciones entre los términos de error de las distintas ecuaciones, entonces la estimación simultánea, por mínimos cuadrados, del conjunto de ecuaciones coincide con la estimación de mínimos cuadrados ecuación por ecuación, por lo que no se obtiene ningún incremento en eficiencia como consecuencia de la consideración conjunta de las ecuaciones. Cuando los términos de error están correlacionados, entonces las estimaciones mínimo-cuadráticas simultáneamente o ecuación por ecuación serán en general diferentes.

Corolario. Si la matriz de covarianzas de cada vector \mathbf{u}_i es diagonal: $\text{Var } \mathbf{u}_i = \sigma_i^2 \mathbf{I}_T$, y si $\sigma_{ij} = 0$ para todo $i \neq j$, entonces el estimador de MCG

del sistema de ecuaciones es equivalente a estimar por MCO cada ecuación por separado.

Demostración. Basta hacer $\Sigma_i = \mathbf{I}_T$ para $i = 1, 2, \dots, m$ en el teorema anterior.

Utilizando la expresión [8.13] que obtuvimos para el estimador MCG del sistema, es sencillo probar directamente el corolario a la Proposición 8.1. Si suponemos en [8.13] que $\sigma_{ij} = 0, \forall i \neq j$, entonces la matriz Σ^{-1} será también diagonal, y se obtiene el resultado del corolario ya que, en tales condiciones:

$$\begin{aligned} \hat{\beta}_{\text{MCG}} &= [\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1} \mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I}_T)\mathbf{y} = \\ &= \begin{pmatrix} (1/\sigma_1^2)\mathbf{X}'_1\mathbf{X}_1 & 0 & 0 \\ & (1/\sigma_2^2)\mathbf{X}'_2\mathbf{X}_2 & 0 \\ \dots & \dots & \dots \\ & & (1/\sigma_m^2)\mathbf{X}'_m\mathbf{X}_m \end{pmatrix}^{-1} \begin{pmatrix} (1/\sigma_1^2)\mathbf{X}'_1\mathbf{y}_1 \\ (1/\sigma_2^2)\mathbf{X}'_2\mathbf{y}_2 \\ \dots \\ (1/\sigma_m^2)\mathbf{X}'_m\mathbf{y}_m \end{pmatrix} = \\ &= \begin{pmatrix} (\mathbf{X}'_1\mathbf{X}_1)\mathbf{X}'_1\mathbf{y}_1 \\ (\mathbf{X}'_2\mathbf{X}_2)\mathbf{X}'_2\mathbf{y}_2 \\ \dots \\ (\mathbf{X}'_m\mathbf{X}_m)\mathbf{X}'_m\mathbf{y}_m \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1^{\text{MCO}} \\ \hat{\beta}_2^{\text{MCO}} \\ \dots \\ \hat{\beta}_m^{\text{MCO}} \end{pmatrix} \end{aligned}$$

Puede probarse (véase Problema 8.22) que otra situación en que no se obtiene ninguna ganancia de eficiencia con la estimación conjunta es cuando todas las variables explicativas de las distintas ecuaciones, son las mismas. Es preciso puntualizar, sin embargo, que no se trata del caso en que se estiman ecuaciones para explicar la tasa de inflación de un conjunto de países utilizando el crecimiento de la cantidad de dinero como única variable explicativa. Es cierto que se está utilizando «la misma» variable conceptual como explicativa, pero las matrices de observaciones disponibles para ella diferirán de unos países a otros.

Como puede verse, difícilmente se producirá la condición mencionada en el trabajo aplicado. Tan sólo en una situación como la del Ejemplo 1 anterior, si se pretende explicar la demanda de un conjunto de bienes duraderos utilizando los mismos indicadores macroeconómicos para cada uno de ellos (tipos de interés, renta, inflación), y sin utilizar indicadores específicos de cada producto (ni siquiera sus respectivos precios), estaríamos en las condiciones referidas.

En el contexto del modelo que estamos analizando, en ocasiones ocurre que las variables explicativas de una de las ecuaciones aparecen asimismo como explicativas en otra ecuación, quizá junto con otras que no aparecían en la primera, por lo que es importante la siguiente proposición:

Proposición 8.2. Dado el sistema de ecuaciones en bloques:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$$

si \mathbf{X}_2 es un subconjunto de \mathbf{X}_1 , entonces $\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{y}_2$ es precisamente igual al segundo subvector del estimador:

$$\hat{\boldsymbol{\beta}} = [\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{Z}]^{-1} \mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y}$$

donde $\mathbf{Z} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}$ y $\boldsymbol{\beta} = (\boldsymbol{\beta}_1; \boldsymbol{\beta}_2)$, por lo que no se gana eficiencia en la estimación de $\boldsymbol{\beta}_2$ al considerar conjuntamente el primero y segundo bloques. Sin embargo, esto no será cierto en general para el subvector $\boldsymbol{\beta}_1$, que es estimado más eficientemente si se estiman simultáneamente ambas ecuaciones.

Como resultado general, conviene recordar que:

1. Cuanto mayores sean las correlaciones entre las perturbaciones de las distintas ecuaciones, mayor será la ganancia en eficiencia de la estimación MCG simultánea, frente a la estimación de cada ecuación por separado.

2. Cuanto más similares sean las matrices \mathbf{X} de observaciones de las variables explicativas, menor será la ganancia en eficiencia que se obtiene con la estimación simultánea por MCG.

8.10. CONTRASTE DE HIPOTESIS

Por estar formado por variables explicativas exógenas, este modelo no presenta importantes variaciones con respecto al modelo econométrico uniecuacional que hemos discutido en capítulos anteriores. En particular, los resultados que permitieron desarrollar los procedimientos de inferencia estadística del Capítulo 4 siguen siendo válidos. La diferencia estriba en que ahora hemos de tratar con varios vectores de coeficientes $\boldsymbol{\beta}_i$, $i = 1, 2, \dots, m$.

Para el contraste del conjunto de hipótesis lineales $(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})$ el estadístico

$$\frac{(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})' \{ \mathbf{R}[\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1} \mathbf{R}' \}^{-1} (\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})}{\frac{q}{\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{m(T-k)}}$$

sigue, asintóticamente, una distribución $F_{q, m(T-k)}$, donde, como es habitual, q denota el número de restricciones que se pretende contrastar.

De igual modo, el estimador que minimiza la suma de cuadrados restringidos por el conjunto de restricciones lineales es:

$$\hat{\boldsymbol{\beta}}_{\mathbf{R}} = \hat{\boldsymbol{\beta}} + [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1} \mathbf{R}' \{ \mathbf{R}[\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{X}]^{-1} \mathbf{R}' \}^{-1} (\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})$$

Tanto para obtener el estimador restringido como para llevar a cabo el contraste de hipótesis mediante comparación de las sumas residuales, existe la posibilidad ya mencionada en el Capítulo 4 de incorporar al modelo dichas restricciones y estimar el modelo resultante, que sería el modelo restringido, en la denominación que allí utilizamos. Por ejemplo, supongamos que todos los modelos contienen las mismas variables explicativas y que se pretende contrastar la igualdad de los vectores $\beta_1, \beta_2, \dots, \beta_m$. En tal caso, se formaría la matriz X , que tendría ahora ordenadas en columna las matrices X_i , siendo, por tanto, de dimensión $Tm \times k$, así como el vector y , de dimensión $Tm \times 1$, para obtener el estimador restringido de dimensión k :

$$\hat{\beta}^* = [X'(\Sigma^{-1} \otimes I_T)X]^{-1} X'(\Sigma^{-1} \otimes I_T)y$$

La suma residual de este modelo debería compararse con la del modelo sin restringir (aquel que estima vectores diferentes $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$) para contrastar la validez de las restricciones.

PROBLEMAS

Problema 8.1. Una muestra de datos sobre renta (Y) y gasto de consumo (C) de unidades familiares se divide en tres partes, dependiendo de que el cabeza de familia sea trabajador por cuenta ajena, trabajador por cuenta propia, o esté en paro. Se estima una regresión de $\log C$ sobre $\log Y$ para cada submuestra, ambas en desviaciones respecto a la media de cada submuestra, así como para toda la muestra, obteniéndose:

	$\hat{\beta}$	$\hat{\sigma}$	n
Trabajadores por cuenta ajena	1,02 (0,06)	0,24	102
Trabajadores por cuenta propia	0,91 (0,10)	0,46	104
Trabajadores en paro	0,76 (0,08)	0,30	26
Toda la muestra	0,86 (0,05)	0,39	232

donde $\hat{\sigma}$ es la suma de los cuadros de los residuos, n el número de observaciones y $\hat{\beta}$ la pendiente de la recta de regresión.

Contrastar las hipótesis:

- La elasticidad renta del gasto familiar es la misma para cada una de las tres clases consideradas.
- La elasticidad renta del gasto familiar es 1,0 en cada clase.
- Dicha elasticidad es 1,0 para la muestra agregada.

Nota: Las cifras en paréntesis son desviaciones típicas estimadas.

Problema 8.2. Para llevar a cabo un análisis de la demanda de microordenadores en España, se ordenaron sus 54 provincias en orden decreciente de acuerdo con su renta per cápita y a continuación se especificó el modelo:

$$\ln N_i = \alpha + \beta \cdot \ln E_i + u_i$$

donde N_i es el número de microordenadores vendidos en la provincia i -ésima y E_i el número de licenciados superiores de los últimos tres años.

Dicho modelo se estimó tres veces: para las provincias de renta alta, media y baja, con las variables en desviaciones respecto a los promedios de las correspondientes submuestras, teniéndose:

	β	$\hat{\sigma}^2$	T
Renta alta	0,80	1,5	18
Renta media	0,60	1	18
Renta baja	0,50	0,5	18
Todas	0,65	1,2	54

Contraste la hipótesis de que el coeficiente β es igual para los tres grupos de provincias.

Problema 8.3. Se dispone de los siguientes datos anuales del volumen de ventas (Y , en millones de pesetas) y gastos de promoción (X , en miles de pesetas) de dos empresas de un mismo sector:

	<i>Empresa 1</i>		<i>Empresa 2</i>	
	<i>Ventas</i>	<i>Gastos de promoción</i>	<i>Ventas</i>	<i>Gastos de promoción</i>
1978	360	80	540	110
1979	380	90	560	120
1980	340	80	550	130
1981	400	90	590	160
1982	400	90	590	150
1983	440	90	600	160
1984	410	110	620	170
1985	390	120	640	190
1986	420	120	650	200
1987	460	130	660	210

a) Obtener las estimaciones de los parámetros del modelo

$$Y_{it} = \alpha_i + \beta_i X_{it} + u_{it}, \quad i = 1, 2$$

para cada una de las empresas por separado.

b) Contrastar la hipótesis nula:

$$H_0: \alpha_1 = \alpha_2$$

$$\beta_1 = \beta_2$$

c) Estimar el modelo bajo el supuesto: $\beta_1 = \beta_2$.

Problema 8.4. Supongamos que se desea estimar una ecuación de demanda de automóviles (y) y como función de la renta (x) utilizando datos microeconómicos sobre familias. Se postula que la demanda para familias que viven en áreas urbanas es:

$$y_i = \alpha_1 + \beta_1 x_i + u_i, \quad i = 1, \dots, N_1 \quad [1]$$

con $E u_i = 0$, $E u_i^2 = \sigma_1^2$, mientras que dicha demanda para familias que viven en medio rural es:

$$y_j = \alpha_2 + \beta_2 x_j + u_j, \quad j = 1, \dots, N_2 \quad [2]$$

con $E u_j = 0$, $E u_j^2 = \sigma_2^2$. Explicar cómo se estimaría eficientemente esta función de demanda si:

- No se tiene ninguna información acerca de α_1 , α_2 , β_1 , β_2 , σ_1^2 y σ_2^2 .
- Se cree que $\beta_1 = \beta_2$.
- Se cree que $\beta_1 = \beta_2$ y $\sigma_1^2 = \sigma_2^2$.
- Se cree que $\beta_1 = \beta_2$, $\alpha_1 = \alpha_2$ y $\sigma_1^2 = \sigma_2^2$.
- Se cree que $\beta_1 = \beta_2$ y $\alpha_1 = \alpha_2$.

Problema 8.5. Dada la especificación correcta

$$y_t = \alpha + \delta x_t + \varepsilon_t, \quad E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma^2$$

se ignora si los coeficientes α y δ son los mismos para dos muestras de las que se dispone.

A partir de los resultados:

Primera submuestra

$$n = 102, \quad R_1^2 = 0,50, \quad y_1' y_1 = 416, \quad x_1' y_1 = \begin{pmatrix} 40 \\ 45 \end{pmatrix}$$

Segunda submuestra

$$n = 52, \quad R_2^2 = 0,80, \quad y_2' y_2 = 1.950, \quad x_2' y_2 = \begin{pmatrix} 100 \\ 116 \end{pmatrix}$$

Muestra completa

$$R^2 = 0,72$$

Se pide:

- Obtener los estimadores MCO para cada submuestra.

- b) Contrastar la hipótesis de cambio estructural mediante el estadístico F .
 c) Contrastar la hipótesis anterior utilizando la suma de los cuadrados de los residuos.

Problema 8.6. Para estimar el multiplicador del Gasto Público se dispuso de datos de series temporales formadas por T observaciones anuales para Alemania y el Reino Unido. En ambos casos se especificó el modelo:

$$\Delta X_{it} = \beta_1 \Delta r_{it} + \beta_2 \Delta G_{it} + \beta_3 \Delta I_{it-1} + u_{it}, \quad i = 1, 2; \quad t = 1, 2, \dots, T$$

donde X_t es el PNB en términos reales, r_t los tipos de interés nominales a corto plazo, I_{t-1} la inversión y G_t el Gasto Público. Todas las variables son logaritmos de las variables originales.

Se supone que los términos de error de ambos países están correlacionados contemporáneamente, aunque no en distintos períodos, es decir: $E(u_{1t} \cdot u_{2s}) = \sigma_{12}$ si $t = s$ e igual a 0 si $t \neq s$. Además, las varianzas σ_1^2 y σ_2^2 de los términos de error son diferentes para ambos países.

Un análisis preliminar detectó la existencia de autocorrelación de primer orden en ambos países, aunque con parámetros sensiblemente diferentes. Explicar en detalle cómo debería procederse para obtener *estimaciones eficientes* de los parámetros β_i , $i = 1, 2, 3$, bajo el supuesto de que son iguales para ambos países. ¿Cómo podría llevarse a cabo el contraste de igualdad de los coeficientes β_i , $i = 1, 2, 3$? Supuesto que no se rechace dicha hipótesis nula, ¿cómo podría contrastarse la hipótesis de que los coeficientes de autocorrelación del término de error son iguales para ambos países?

Problema 8.7. Se pretende efectuar un estudio sobre las rentas pagadas por familias que habitan viviendas de alquiler. Para ello, se recoge una muestra de 525 familias. Tras especificar el modelo

$$\text{Alquiler}_i = \alpha + \beta \text{Ingresos}_i + u_i, \quad i = 1, 2, \dots, 525$$

se sospecha que dicho modelo puede presentar heteroscedasticidad, por lo que se divide la muestra en dos partes, de acuerdo con el valor mediana de los ingresos. Así, una submuestra consta de las familias con ingresos superiores a la mediana, y la otra submuestra, de las familias con ingresos inferiores a la mediana. Se estima el modelo para cada subgrupo, con los siguientes resultados:

- a) Familias de ingresos superiores:

$$\text{Alquiler}_i = 1.500 + 0,20 \cdot \text{Ingresos}_i, \quad R^2 = 0,80$$

(1,5) (3,0)

$$\text{Suma residual} = \sum \hat{u}_i^2 = 30.000$$

- b) Familias de ingresos inferiores:

$$\text{Alquiler}_i = 600 + 0,30 \cdot \text{Ingresos}_i, \quad R^2 = 0,60,$$

(3,0) (1,8)

$$\text{Suma residual} = \sum \hat{u}_i^2 = 15.000$$

Con esta información, contrastar la hipótesis nula de ausencia de heteroscedasticidad. ¿Podría diseñarse un contraste de mayor potencia? Si se rechazase la hipótesis

nula, ¿cómo debería procederse para obtener estimaciones eficientes de los parámetros α y β ?

Problema 8.8. Consideremos los dos modelos de regresión:

$$\begin{aligned} Y_a &= X_a \beta_a + u_a, & u_a &\sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_5) \\ Y_b &= X_b \beta_b + u_b, & u_b &\sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_5) \end{aligned}$$

sobre los que se tienen los datos:

$$X_a = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & -2 & -1 \\ 1 & 2 & 1 \end{pmatrix} \quad Y_a = \begin{pmatrix} -1 \\ 5 \\ 3 \\ 6 \\ 12 \end{pmatrix} \quad X_b = X_a \quad Y_b = \begin{pmatrix} 4 \\ 10 \\ 8 \\ 10 \\ 13 \end{pmatrix}$$

- Utilizando los datos del grupo *a*), construir un intervalo de confianza del 95 por 100 para la combinación lineal de parámetros: $\theta = \beta_1 - 3\beta_2 + 4\beta_3$.
- Utilizando los datos del grupo *b*, verificar la hipótesis $H_0: \beta_1 + \beta_2 = 0$ al 95 por 100 de confianza.
- Calcular el estimador de máxima verosimilitud de β_b dado que $\beta_1 + \beta_2 = 0$.
- Contrastar la homogeneidad global de los modelos al 95 por 100 de confianza.
- Suponiendo que $E(u_a u_b) = 4\sigma^2 \mathbf{I}_5$, contrastar la hipótesis de homogeneidad global al 95 por 100 de confianza.

Problema 8.9. Se sabe que las variables X e Y obedecen al modelo:

$$y_i = \alpha + \beta x_i + u_i, \quad E(u_i) = 0, \quad E(u_i^2) = \sigma^2, \quad \text{para todo } i$$

Obtenidas diez observaciones de ambas variables, se sospecha que las cinco primeras parecen corresponder al modelo especificado pero con parámetros diferentes de los correspondientes a las cinco últimas observaciones.

Estimando el modelo por MCO con las cinco primeras observaciones se obtuvo:

$$R_1^2 = 0,42 \quad \hat{\sigma}_1^2 (X_1' X_1)^{-1} = \begin{pmatrix} 12,26 & -5,0 \\ -5,0 & 2,27 \end{pmatrix}; \quad y_1' y_1 = 337; \quad X_1' y_1 = \begin{pmatrix} 39 \\ 92 \end{pmatrix}$$

mientras que con las cinco últimas:

$$R_2^2 = 0,80 \quad \hat{\sigma}_2^2 (X_2' X_2)^{-1} = \begin{pmatrix} 45,18 & -6,5 \\ -6,5 & 1,0 \end{pmatrix}; \quad y_2' y_2 = 2.000; \quad X_2' y_2 = \begin{pmatrix} 96 \\ 672 \end{pmatrix}$$

Con estos datos se desea:

- Obtener las estimaciones MCO de β para cada uno de los grupos de cinco observaciones.
- Contrastar la hipótesis de que los parámetros α y β son los mismos en ambos grupos.
- Contrastar la misma hipótesis utilizando únicamente la suma de los cuadrados de los residuos obtenidos en cada una de las regresiones anteriores, así como en la que se hizo con las diez observaciones, en la que se obtuvo $R^2 = 0,89$.

Problema 8.10. Para las variables macroeconómicas X e Y se dispone de observaciones anuales de cuatro países diferentes, de manera que:

	<i>País 1</i>	<i>País 2</i>	<i>País 3</i>	<i>País 4</i>
T_i	10	10	10	10
\bar{Y}_i	2	4	2	5
\bar{X}_i	2	1	2	5
ΣY_j^2	80	240	180	450
ΣX_j^2	60	100	80	300
$\Sigma X_j Y_j$	15	20	50	150

- a) Obtener, para cada país por separado, las estimaciones de los parámetros α y β :

$$Y_j = \alpha + \beta X_j$$

- b) Contrastar la hipótesis nula:

$$H_0: \alpha_i = \alpha, \quad i = 1, 2, 3, 4$$

$$\beta_i = \beta, \quad i = 1, 2, 3, 4$$

- c) Estimar el modelo bajo el supuesto: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$.

Problema 8.11. Se dispone de la siguiente información acerca de dos partes en que se ha dividido una muestra:

$$\mathbf{X}'_1 \mathbf{X}_1 = \begin{pmatrix} 50 & 25 \\ 25 & 100 \end{pmatrix} \quad \mathbf{X}'_2 \mathbf{X}_2 = \begin{pmatrix} 100 & 25 \\ 25 & 500 \end{pmatrix} \quad \mathbf{X}'_1 \mathbf{y}_1 = \begin{pmatrix} -40 \\ 160 \end{pmatrix} \quad \mathbf{X}'_2 \mathbf{y}_2 = \begin{pmatrix} 60 \\ 300 \end{pmatrix}$$

$$T_1 = 50, \quad T_2 = 100, \quad \mathbf{y}'_1 \mathbf{y}_1 = 450, \quad \mathbf{y}'_2 \mathbf{y}_2 = 2.700$$

- a) Obtener las estimaciones MCO de los parámetros a y b del modelo

$$y_t = \alpha + \beta x_t + u_t$$

en cada submuestra, así como sus matrices de covarianzas.

b) Contrastar la hipótesis de homogeneidad global. Justificar la utilización de MCG en la estimación del modelo restringido.

c) Comprobar que las estimaciones de α y β en el modelo restringido se hallan entre las estimaciones obtenidas para estos parámetros en cada una de las submuestras.

Problema 8.12. Probar que si se especifica el modelo

$$y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + u_t$$

donde las variables ficticias D_{1t} , D_{2t} , D_{3t} corresponden a cada una de las tres submuestras en que se ha dividido la muestra total, entonces las estimaciones MCO de los coeficientes α_i , $i = 1, 2, 3$ son los promedios de la variable endógena en cada una de las submuestras.

Problema 8.13. Probar que el estimador MCO del vector β en el modelo

$$y_{it} = \mu + \alpha_i + \mathbf{x}'_{it}\beta + u_{it}, \quad i = 1, 2, \dots, m, \quad t = 1, 2, \dots, T$$

bajo las restricciones $\sum_1^m \alpha_i = 0$, es igual al estimador MCO del vector β en el modelo

$$y_{it} = \gamma_i + \mathbf{x}'_{it}\beta + u_{it}, \quad i = 1, 2, \dots, m, \quad t = 1, 2, \dots, T$$

donde los parámetros γ_i , $i = 1, 2, \dots, m$ no están sujetos a ninguna restricción.

Problema 8.14. Suponga que dispone de datos de series temporales procedentes de dos países diferentes, con los que se pretende estimar el modelo formado por las ecuaciones:

$$\begin{aligned} y_{1t} &= x'_{1t}\beta_1 + u_{1t}, & t &= 1, 2, \dots, T \\ y_{2t} &= x'_{2t}\beta_2 + u_{2t}, & t &= 1, 2, \dots, T \end{aligned}$$

con:

$$\begin{aligned} E(u_{1t}) &= E(u_{2t}) = 0 \\ \text{Var}(u_{1t}) &= \sigma_1^2, \quad \text{Var}(u_{2t}) = \sigma_2^2, \quad \text{diferentes entre sí} \\ E(u_{1t} \cdot u_{2s}) &= E(u_{1t} \cdot u_{1s}) = E(u_{2t} \cdot u_{2s}) = 0 \quad \text{para todo } t \text{ distinto de } s \\ E(u_{1t} \cdot u_{2t}) &= \sigma_{12} \end{aligned}$$

Comente estos supuestos acerca de la estructura de autocorrelación y correlación cruzada de los términos de error.

Explique, con todo detalle analítico, cómo procedería a la contrastación de la hipótesis nula de homogeneidad de coeficientes entre las ecuaciones de ambos países, en el caso en que suponemos que $\sigma_{12} = 0$. ¿Cómo cambia su respuesta si consideramos el caso más general en que σ_{12} es distinto de cero?

Problema 8.15. Suponga que se ha estimado el modelo econométrico $y_t = \mathbf{x}'_t\beta + u_t$ con n submuestras distintas, cada una de las cuales consta de T observaciones. Halle los estadísticos de los multiplicadores de Lagrange, de razón de verosimilitudes y de Wald para el contraste de la hipótesis nula $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$.

Problema 8.16. Las funciones de demanda de dinero de dos países vienen dadas por los modelos (1) y (2) siguientes:

$$(1) \quad \left(\frac{M}{P}\right)_{1t} = \alpha_1 + \alpha_2 Y_{1t} + \varepsilon_{1t}$$

$$(2) \quad \left(\frac{M}{P}\right)_{2t} = \beta_1 + \beta_2 Y_{2t} + \varepsilon_{2t}$$

donde $(M/P)_t$ e Y_t denotan los saldos monetarios y la renta, ambos en términos reales.

Se dispone de la matriz de momentos cruzados calculados con 20 observaciones anuales:

	$(M/P)_1$	$(M/P)_2$	Y_1	Y_2
$(M/P)_1$	10	-1	1	-1
$(M/P)_2$		15	5	1
Y_1			11	1
Y_2				1

Contraste la $H_0: \alpha_2 = \beta_2$ bajo los siguientes supuestos alternativos:

1. $E(\varepsilon_{1t})^2 = E(\varepsilon_{2t})^2$.
2. $E(\varepsilon_{1t})^2 = 1$, $E(\varepsilon_{2t}) = 2$.
3. $E(\varepsilon_{1t})^2 = 1$, $E(\varepsilon_{2t}) = 2$, $E(\varepsilon_{1t}\varepsilon_{2t}) = 1$.

Problema 8.17. Considere la estimación de mínimos cuadrados del modelo M4

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix} = \begin{pmatrix} \mathbf{1}_T & \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1}_T & \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{1}_T & \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_m \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_m \end{pmatrix}$$

y muestre que no es posible particionar el problema de estimación, de modo que se reduzca la dimensionalidad del vector de coeficientes.

Pruebe que el estimador MCO de α es una media ponderada de las estimaciones que se tendrían para los términos independientes $\alpha_1, \alpha_2, \dots, \alpha_m$ si se supusiesen diferentes para cada submuestra.

Obtenga asimismo la expresión analítica para los estimadores de los parámetros $\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$.

Problema 8.18. Para analizar el efecto contractivo que sobre el mercado de crédito tiene la financiación del déficit público, se dispone de series temporales de datos anuales, para el periodo 1965-1986, procedentes de Estados Unidos, U.K. y Japón. Con ellos se especifica el modelo

$$r_{it} = \beta_{i1} + \beta_{i2}D_{it} + \beta_{i3}m_{it} + \beta_{i4}y_{it} + u_{it}$$

donde r_{it} , D_{it} , m_{it} , y_{it} denotan los tipos de interés reales, el déficit como porcentaje de PIB, el crecimiento de la oferta monetaria y la tasa de crecimiento del PIB. Por otra parte, $i = 1, 2, 3$ según que la observación de que se trate provenga de Estados Unidos, U.K. o Japón.

- a) ¿Cómo se estimaría el modelo anterior si se supone que $\beta_{11} = \beta_{21} = \beta_{31} = \beta_1$?
- b) Probar que el procedimiento de estimación sugerido, tanto para β_1 como para el vector $(\beta_{i2}, \beta_{i3}, \beta_{i4})$, es consistente con las ecuaciones normales para la estimación de mínimos cuadrados.
- c) ¿Cómo podría contrastarse la hipótesis $H_0: \beta_{11} = \beta_{21} = \beta_{31} = \beta_1$? Lleve a cabo esta discusión con el mayor detalle posible.

Problema 8.19. a) Obtenga estimaciones eficientes del modelo

$$y_{1t} = \alpha_{11}x_{1t} + \alpha_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \alpha_{32}x_{3t} + \alpha_{42}x_{4t} + u_{2t}$$

donde todas las variables se hallan en diferencias con respecto a la media, sabiendo que

$$\text{Var } u_{1t} = \sigma^2; \quad \text{Var } u_{2t} = 2\sigma^2; \quad \text{Cov}(u_{1t}, u_{2t}) = \sigma^2$$

y disponiendo de la matriz de *momentos muestrales*:

	y_1	y_2	x_1	x_2	x_3	x_4
y_1	2.000	500	-200	400	200	100
y_2		1.000	150	-200	30	-20
x_1			100	0	0	0
x_2				300	0	0
x_3					20	10
x_4						10

b) Proponga en todo detalle un estimador eficiente del parámetro σ^2 y calcule su valor numérico.

Problema 8.20. a) Estimar cada ecuación del modelo

$$Y_{1t} = \beta_{11}X_{1t} + \beta_{21}X_{2t} + u_{1t}$$

$$Y_{2t} = \beta_{12}X_{1t} + u_{2t}$$

por separado, donde todas las variables están en desviaciones con respecto a la media, sabiendo que

$$\text{Cov}(u_{1t}, u_{2t}) = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix}$$

y disponiendo de los momentos muestrales de segundo orden:

	Y_1	Y_2	X_1	X_2
Y_1	3.000	500	-200	400
Y_2		1.000	100	-200
X_1			100	0
X_2				300

b) Estimar ambas ecuaciones conjuntamente, utilizando el hecho de que $\text{Cov}(u_{1t}, u_{2t}) \neq 0$.

c) Contrastar la hipótesis nula $H_0: \beta_{11} = \beta_{12}$.

Problema 8.21. a) Describir y obtener, en todo detalle, las expresiones matriciales que conducirían a la obtención de un estimador eficiente del sistema

$$y_{1t} = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{12}x_{1t} + u_{2t}$$

si $E u_{1t} = E u_{2t} = 0$; $\text{Var } u_{1t} = \sigma_1^2$, $\text{Var } u_{2t} = \sigma_2^2$, $\text{Cov}(u_{1t}, u_{2t}) = \sigma_{12}$, siendo u_{1t} y u_{2t} ruidos blancos.

b) ¿Qué papel jugaría el estimador MCO ecuación por ecuación para obtener preestimaciones en la obtención numérica del estimador que ha propuesto en a)?

c) ¿Bajo qué condiciones coincidirían ambos estimadores? Demuestre su respuesta.

d) Describir las expresiones analíticas que generarían el estimador eficiente del sistema anterior, sujeto a la restricción $\beta_{11} = \beta_{12}$.

e) ¿A qué se reduce la expresión de dicho estimador restringido en el caso particular en que $\sigma_{12} = 0$? ¿Coincide con alguna versión de mínimos cuadrados?

Problema 8.22. Considere el modelo de regresiones aparentemente no relacionadas:

$$y_1 = X_1 \beta_1 + u_1$$

$$y_2 = X_2 \beta_2 + u_2$$

.....

$$y_m = X_m \beta_m + u_m$$

donde y_j , u_j , $j = 1, 2, \dots, m$ son vectores $T \times 1$, X_j , $j = 1, 2, \dots, m$ matrices $T \times k$ y β_j , $i \leq j \leq m$ vectores $k \times 1$, y suponga que las observaciones numéricas que configuran cada una de las matrices x_j coinciden entre submuestras: $X_1 = X_2 = \dots = X_m = Z$. Suponga asimismo que $\text{Var}(u_j) = \sigma_j^2 I_T$.

Demuestre que, en tales condiciones, la estimación MCG del sistema es equivalente a la estimación MCO ecuación por ecuación.

¿Se obtendría el mismo resultado si $\text{Var}(u_j) = \sigma_j^2 \Sigma$, $j = 1, 2, \dots, m$, con $\Sigma \neq I_T$?

Problema 8.23. Para analizar la estructura de la demanda en el sector del automóvil, se dispone de datos proporcionados por Ford España, General Motors y Seat acerca de sus volúmenes de ventas y_t , precios medios ponderados y características que, en promedio, tienen sus respectivos modelos.

Para intentar explicar el comportamiento de la variable y_t , se especifica el modelo

$$y_{it} = \alpha_i + x'_{it} \beta + u_{it}, \quad i = 1, 2, 3, \quad t = 1, 2, \dots, T$$

donde el vector x'_{it} incluye las variables antes mencionadas, además de variables de coyuntura como el IPI (o la renta disponible agregada). Nótese que el término independiente se supone constante a través del tiempo para los distintos fabricantes. Si se supone que el término de error del modelo tiene las siguientes propiedades:

1. $E(u_{it}) = \sigma_i^2$ para todo $t = 1, 2, \dots, T$.
2. $u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}$ para todo $i = 1, 2, 3$, $t = 1, 2, \dots, T$, con $E(\varepsilon_{it} \varepsilon_{js}) = 0$ para todo $i \neq j$ y para todo t, s .
3. $E(u_{it} u_{js}) = 0$ si $i \neq j$ para todo t, s .

Interprete estos tres supuestos y describa en *detalle* cómo llevaría a cabo la estimación *eficiente* del modelo anterior. ¿Cómo variaría su respuesta si la condición 3 anterior se cambiase por:

3'. $E(\varepsilon_{it}\varepsilon_{js}) = s_{ij}$ si $t = s$, e igual a cero en caso contrario, y $E(u_{it}u_{js}) = \sigma_{ij}$ si $t = s$, e igual a cero en caso contrario. ¿Qué relación existiría entre σ_{ij} y s_{ij} ?

Problema 8.24. Discutir en todo detalle cómo se llevaría a cabo el contraste de la hipótesis nula $\alpha_2 = \beta_2$ en el modelo de regresiones aparentemente no relacionadas:

$$y_{1t} = \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_{1t}$$

$$y_{2t} = \beta_1 x_{3t} + \beta_2 x_{4t} + u_{2t}$$

Problema 8.25. Los momentos muestrales de orden dos de las variables del modelo

$$y_{1t} = \beta_{11}x_{1t} + \beta_{12}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{21}x_{1t} + \beta_{22}x_{2t} + u_{2t}$$

que se hallan en diferencias con respecto a la media, son:

	y_1	y_2	x_1	x_2
y_1	500	200	200	150
y_2		1.000	100	300
x_1			100	100
x_2				200

calculados a partir de una muestra de 100 observaciones.

Estimar dicho modelo y contrastar la hipótesis $\beta_{11} = \beta_{22}$.

Problema 8.26. Estimar el modelo de regresiones aparentemente no relacionadas con variables en desviaciones respecto a la media:

$$y_{1t} = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{12}x_{1t} + \beta_{22}x_{2t} + u_{2t}$$

sabiendo que $\text{Var } u_{1t} = \text{Var } u_{2t}$ y $\text{Cov}(u_{1t}, u_{2t}) = 0$ y bajo la restricción $\beta_{21} = \beta_{12}$, disponiendo de los momentos muestrales:

	y_{1t}	y_{2t}	x_{1t}	x_{2t}
y_{1t}	120	80	20	45
y_{2t}		90	25	30
x_{1t}			10	5
x_{2t}				20

¿Coincide dicha estimación con la de MCO ecuación por ecuación?

Problema 8.27. Estimar el modelo de regresiones aparentemente no relacionadas con variables en desviaciones respecto a la media:

$$y_{1t} = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{32}x_{3t} + u_{2t}$$

bajo los supuestos: $\text{Var } u_{1t} = \text{Var } u_{2t} = \sigma^2$; $\text{Cov}(u_{1t}, u_{2t}) = 0$, y $\beta_{11} + \beta_{21} = \beta_{32}$.

Obtener las varianzas de las estimaciones de los tres coeficientes estimados. Los momentos muestrales de las variables incluidas en el modelo fueron:

	y_{1t}	y_{2t}	x_{1t}	x_{2t}	x_{3t}
y_{1t}	40	16	12	16	8
y_{2t}		80	8	8	28
x_{1t}			8	4	4
x_{2t}				12	4
x_{3t}					16

Problema 8.28. Estimar el modelo con variables en desviaciones respecto a la media:

$$y_{1t} = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{32}x_{3t} + u_{2t}$$

bajo los supuestos: $\text{Var } u_{1t} = \sigma_1^2$; $\text{Var } u_{2t} = 4\sigma_1^2$; $E(u_{1t}u_{2t}) = 2\rho\sigma_1^2$, $E(u_{1t}u_{1s}) = E(u_{2t}u_{2s}) = 0$ para $t \neq s$, disponiendo de los momentos muestrales:

	y_{1t}	y_{2t}	x_{1t}	x_{2t}	x_{3t}
y_{1t}	200	40	8	2	0
y_{2t}		300	0	0	2
x_{1t}			20	-4	0
x_{2t}				10	0
x_{3t}					10

¿Cuál es la peculiaridad de este modelo que permite su estimación aun desconociendo el valor del parámetro ρ ? ¿Coinciden las estimaciones obtenidas con las de MCO ecuación por ecuación?

Problema 8.29. Estimar, ecuación por ecuación, el modelo en desviaciones respecto a la media:

$$y_{1t} = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t}$$

$$y_{2t} = \beta_{12}x_{1t} + \beta_{22}x_{2t} + u_{2t}$$

sabiendo que a partir de una muestra de 100 observaciones se obtuvieron los momentos:

	y_{1t}	y_{2t}	x_{1t}	x_{2t}
y_{1t}	20	8	8	8
y_{2t}		40	4	12
x_{1t}			4	4
x_{2t}				8

Obtener la matriz de covarianzas de los coeficientes estimados y contrastar la hipótesis:

$$H_0: \beta_{21} = \beta_{12}$$

bajo el supuesto de Normalidad del vector (u_{1t}, u_{2t}) , ausencia de autocorrelación de todo tipo, y con $E(u_{1t}^2) = \sigma_1^2$, $E(u_{2t}^2) = \sigma_2^2$, $E(u_{1t}) = E(u_{2t}) = 0$, $E(u_{1t}u_{2t}) = \sigma_{12}$.

CAPITULO 9

MODELOS DINAMICOS

9.1. INTRODUCCION

En los capítulos anteriores hemos venido suponiendo que las variables que aparecen como explicativas en el modelo econométrico estaban contemporáneamente relacionadas con la variable endógena, por lo que sus índices temporales eran iguales. Sin embargo, la Teoría Económica sugiere que, en la mayoría de los casos, las relaciones entre variables son dinámicas. Ello puede ocurrir porque: *a)* el impacto de una variable sobre otra sólo se deja notar tras un cierto tiempo; *b)* el impacto, aunque sea instantáneo, se deja notar durante un cierto número de períodos, o *c)* ambas cosas a la vez.

Ello conduce, respectivamente, a considerar modelos como los siguientes:

$$\begin{aligned}y_t &= \beta_1 + \beta_2 x_{t-1} + u_t \\y_t &= \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2} + u_t \\y_t &= \beta_1 + \beta_2 x_{t-2} + \beta_3 x_{t-3} + u_t\end{aligned}$$

En el primer modelo, un cambio en el valor de x_t no se deja sentir sobre la variable endógena hasta el período siguiente; en el segundo modelo, un cambio en x_t tiene un efecto inmediato sobre y_t , medido por el coeficiente β_2 , pero también influirá sobre y_{t+1} e y_{t+2} , con efectos dados por los valores de β_2 y β_3 . En el tercer modelo, un cambio en x_t no influye sobre el valor de y_t , ni de y_{t+1} , pero sí sobre y_{t+2} e y_{t+3} .

Los ejemplos acerca de este tipo de relaciones son numerosos: *a)* la inversión productiva genera incrementos en el nivel de producción que, generalmente, se extienden a varios años y no comienzan de inmediato; *b)* el crecimiento de la oferta monetaria tiene efectos sobre los precios que, aparte de ser persistentes, se manifiestan sólo tras un retardo inicial.

Por otra parte, las variables económicas tienen bastante inercia, lo que hace que una variable dependa de su propio pasado, además de otras causas.

Así, por ejemplo, para tratar de explicar el comportamiento de la inflación π_t , tendría sentido introducir como variables explicativas, junto con las tasas de crecimiento monetario m_t , retardos de la propia tasa de inflación:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \beta_2 m_t + u_t$$

Es importante observar que la existencia de una relación dinámica entre variables, así como su mayor o menor persistencia (es decir, el número de retardos precisos para representarla), dependen crucialmente de cual sea la frecuencia de observación de los datos que se emplean en la estimación. Así, si una variable x_t influye sobre otra y_t no sólo contemporáneamente, sino también durante los dos meses siguientes, entonces la relación sería dinámica si el investigador utiliza datos mensuales, pero resultará estática si utilizase datos anuales. En este último caso, tan sólo el valor contemporáneo de x_t aparecería como variable explicativa en el modelo de regresión.

9.1.a. Primeras propiedades

Comencemos examinando las distintas propiedades dinámicas que revisten estos modelos. Consideremos a modo de ejemplo la relación:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2}$$

que, por simplificar, consideramos determinista, y supongamos que la economía está en equilibrio, es decir, que las observaciones de las variables x_t e y_t vienen siendo iguales en los últimos períodos a sus valores de equilibrio, x^* e y^* ; de acuerdo con el modelo, debe cumplirse que $y^* = \beta_1 + (\beta_2 + \beta_3 + \beta_4)x^*$.

Supongamos que, por alguna razón, en el instante t_0 la variable x_t se desvía de dicho equilibrio en una cantidad Δ , es decir: $x_{t_0} = x^* + \Delta$, pero vuelve al equilibrio a partir de entonces $x_t = x^*$ para todo $t > t_0$. Es fácil ver que $y_{t_0} = y^* + \beta_2 \Delta$, pero no todo acaba aquí; también tendremos $y_{t_0+1} = \beta_1 + \beta_2 x_{t_0+1} + \beta_3 x_{t_0} + \beta_4 x_{t_0-1} = \beta_1 + \beta_2 x^* + \beta_3 (x^* + \Delta) + \beta_4 x^* = \beta_1 + (\beta_2 + \beta_3 + \beta_4)x^* + \beta_3 \Delta = y^* + \beta_3 \Delta$. El lector puede comprobar que, de modo análogo, se tiene $y_{t_0+2} = y^* + \beta_4 \Delta$, así como $y_t = y^*$ para todo $t > t_0 + 2$. La función de respuesta al impulso es la sucesión de estos efectos que, en el caso de este modelo particular, resultan ser $\beta_2, \beta_3, \beta_4, 0, 0, \dots$, y resume el efecto que sobre y_t tiene una desviación puramente transitoria de x_t respecto de su valor de equilibrio inicial. Este efecto consiste en una desviación de y_t respecto de su equilibrio inicial durante tres períodos, para terminar volviendo al mismo a partir de entonces.

La función de respuesta al escalón es un estadístico que responde a una pregunta similar: ¿qué efecto tendría sobre y_t una desviación permanente de x_t respecto de su equilibrio inicial, es decir, si pasa a un nuevo valor de equilibrio? Denotando dicho valor de equilibrio por $x^* + \Delta$, supongamos que $x_t = x^*$ si $t < t_0$, $x_{t_0} = x^* + \Delta$, y $x_t = x^* + \Delta$ para $t > t_0$. Un razonamiento similar al anterior muestra que, en nuestro ejemplo, la función de respuesta

al escalón es β_1 , $\beta_1 + \beta_2$, $\beta_1 + \beta_2 + \beta_3$ y permanece constante en este valor a partir de tres periodos. En este caso, al pasar x_t a su nuevo valor de equilibrio, y_t pasa también a un nuevo valor de equilibrio: $y^* + (\beta_1 + \beta_2 + \beta_3)\Delta$, en el que permanece a partir de entonces. *La función de respuesta al escalón no es sino la sucesión de valores acumulados de la función de respuesta al impulso.*

Si normalizamos los coeficientes $\beta'_i = \beta_i / \sum_j \beta_j$, cada uno de los nuevos coeficientes β'_i nos da la proporción del efecto total que se deja sentir j periodos después del cambio en x_t . De este modo, puede responderse a preguntas como: ¿Cuántos periodos deben transcurrir antes de que se deje notar el 90 por 100 del efecto que el cambio en x_t tiene sobre y_t ? En particular, el *retardo mediano* es aquel periodo en que ya se ha dejado sentir sobre y_t el 50 por 100 del efecto producido por la desviación en x_t .

El *retardo medio* es el cociente $\sum_i i\beta'_i / \sum_j \beta'_j$. Un valor bajo del retardo medio significa que el ajuste se lleva a cabo rápidamente.

Utilizando el operador de retardos, $Lx_t = x_{t-1}$, podemos escribir un modelo general como:

$$y_t - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p} = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s} + u_t$$

en la forma $A(L)y_t = \delta + B(L)x_t + u_t$, donde $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$ y $B(L) = \beta_0 - \beta_1 L - \beta_2 L^2 - \dots - \beta_s L^s$, y la *ganancia* se define como el cociente $B(1)/A(1)$, donde se ha sustituido el operador L por la unidad en los polinomios $A(L)$ y $B(L)$.

Todos estos conceptos son diferentes cuando hay retardos de la variable endógena. Entonces hay que distinguir entre respuestas a corto y a largo plazo.

Por ejemplo, en el modelo

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_t + u_t$$

la respuesta a corto plazo de y_t a un impulso en x_t es simplemente β_2 , pero la aparición del retardo de la variable endógena hace dicho modelo equivalente a $(1 - \beta_1 L)y_t = \alpha + \beta_2 x_t + u_t$ que, bajo la *condición de estacionariedad* $-1 < \beta_1 < 1$, puede escribirse:

$$y_t = \frac{\alpha}{1 - \beta_1} + \sum_0^\infty \beta_2 (\beta_1)^s x_{t-s} + \sum_0^\infty (\beta_1)^s u_{t-s} = \frac{\alpha}{1 - \beta_1} + (\beta_2 x_t + \beta_2 \beta_1 x_{t-1} + \beta_2 \beta_1^2 x_{t-2} + \dots) + \sum_0^\infty (\beta_1)^s u_{t-s}$$

y, prescindiendo del componente aleatorio, nuestro análisis anterior conduce a una función de respuesta al impulso: $\beta_2, \beta_2 \beta_1, \beta_2 \beta_1^2, \beta_2 \beta_1^3, \dots$, mientras que la función de respuesta al escalón acumula ésta. Como puede verse, dicha respuesta es convergente tan sólo si el parámetro β_2 tiene valor absoluto inferior a la unidad. Esta es la *condición de estacionariedad del modelo*, que debe manifestarse en que las funciones de respuesta al impulso tienden a cero, y las de respuesta al escalón se estabilizan.

El tratamiento estadístico de estos modelos difiere, según que los valores retardados que aparecen como variables explicativas pertenezcan a variables exógenas, o que entre ellos haya algún retardo de la variable endógena.

9.1.b. Cuando todos los retardos corresponden a variables exógenas

Si el modelo es del tipo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t-1} + \dots + \beta_s x_{2t-s} + u_t \quad [9.1]$$

no se incumplen las hipótesis básicas del modelo lineal general, por cuanto las distintas variables explicativas del modelo de regresión son todas deterministas. En este modelo aparecen tan sólo dos posibles dificultades:

1. Los retardos consecutivos de una variable económica tienden a estar correlacionados entre sí, tanto más cuanto mayor sea la estructura de autocorrelación de dicha variable. Así, por ejemplo, si x_t sigue una estructura autorregresiva de primer orden con coeficiente 0,95:

$$x_t = 0,95x_{t-1} + \varepsilon_t$$

la correlación entre x_t y x_{t-s} será muy elevada, incluso para valores altos de s , siendo dicha correlación muy inferior si el coeficiente del modelo de x_{t-1} fuese igual a 0,40.

Cuanto mayor sea la correlación entre los retardos de x_t , más importante será la presencia de *multicolinealidad* en el modelo de regresión, cuyas implicaciones analizaremos en el Capítulo 10.

2. La segunda dificultad surge cuando $s = \infty$, es decir, cuando la estructura de retardos es de orden infinito. Entonces es imposible estimar directamente el modelo [9.1], por cuanto que no tendríamos observaciones suficientes para ello. Como veremos en la Sección 9.3, para estimar este modelo es imprescindible imponer a priori algún tipo de restricción entre los coeficientes, de modo que el modelo pueda transformarse en otro con un número reducido de variables explicativas.

9.1.c. Si aparecen valores retardados de la variable endógena

como variables explicativas, entonces dejaría de cumplirse uno de los supuestos bajo los que desarrollamos las teorías de estimación e inferencia del modelo econométrico, pues algunas de las variables explicativas serían ahora variables aleatorias (ya que y_t lo es). Sin embargo, si el término de error no tiene autocorrelación, el problema de estimación no es muy importante.

Consideremos, por ejemplo, el modelo:

$$y_t = \beta y_{t-1} + u_t, \quad |\beta| < 1 \quad [9.2]$$

donde u_t es un proceso de ruido blanco. El estimador de mínimos cuadrados del parámetro β es:

$$\hat{\beta}_{\text{MCO}} = \frac{\sum_2^T y_t y_{t-1}}{\sum_2^T y_{t-1}^2} = \sum_2^T \frac{(\beta y_{t-1} + u_t) y_{t-1}}{\sum_2^T y_{t-1}^2} = \beta + \frac{\sum_2^T y_{t-1} u_t}{\sum_2^T y_{t-1}^2}$$

de modo que el estimador será insesgado si y sólo si se cumple:

$$E\left(\frac{\sum_2^T y_{t-1} u_t}{\sum_2^T y_{t-1}^2}\right) = 0 \quad [9.3]$$

Si la distribución de u_t fuese independiente de y_s para todo par (t, s) , entonces se tendría para $s = 2, \dots, T$

$$E(y_{s-1} u_s / \sum_2^T y_{t-1}^2) = E(y_{s-1} / \sum_2^T y_{t-1}^2) \cdot E(u_s) = 0 \quad [9.4]$$

por lo que la condición [9.3] se cumpliría y el estimador MCO sería insesgado. Sin embargo, [9.2] muestra que las distribuciones de y_t y u_s no son independientes, puesto que si el valor absoluto de β es inferior a la unidad, entonces:

$$y_t = \sum_{s=0}^{\infty} \beta^s u_{t-s}$$

por lo que y_t depende de u_t y de valores retardados de u_t . Por consiguiente, y puesto que u_t es ruido blanco, y_t será independiente de u_{t+s} para $s \geq 1$; en particular, $E(y_{t-1} u_t) = 0$, por lo que el numerador de la fracción en [9.3] tendrá esperanza cero. Lo que ocurre es que el denominador incluye valores de y_t posteriores a u_s , por lo que no podemos efectuar el desarrollo en [9.4]. El estimador MCO del modelo [9.2] será, en general, sesgado. Sin embargo, veremos en la Sección 9.4.a que, bajo determinadas condiciones, dicho sesgo tiende a cero, por lo que el estimador MCO de [9.2] es consistente.

El problema se complica sustancialmente cuando *aparecen valores retardados de la variable endógena como variables explicativas y, además, el término de error tiene autocorrelación:*

$$y_t = \beta y_{t-1} + u_t, \quad |\beta| < 1 \quad [9.5.a]$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad [9.5.b]$$

En este caso, basta escribir el modelo [9.5.a] en el instante $t - 1$ para darse cuenta de que la variable explicativa y_{t-1} está correlacionada con u_{t-1} , quien, a su vez, está correlacionada con u_t (véase [9.5.b]). En consecuencia, una de las variables explicativas del modelo está correlacionada con el término de error, por lo que ya no se tiene $E(y_{t-1} u_t) = 0$.

En efecto, si u_t tiene una estructura autorregresiva como [9.5.b], entonces $E(u_t u_{t-s}) = \rho^s \sigma_u^2$, y teniendo en cuenta que $|\beta\rho| < 1$, se tiene en el modelo [9.5]:

$$E(y_{t-1} u_t) = E[(\sum_{i=0}^{\infty} \beta^i u_{t-1-i}) u_t] = \sum_{i=0}^{\infty} \beta^i \rho^{i+1} \sigma_u^2 = \frac{\rho \sigma_u^2}{1 - \beta\rho} \quad [9.6]$$

y, en general:

$$E(y_{t-s} u_t) = E[(\sum_{i=0}^{\infty} \beta^i u_{t-s-i}) u_t] = \sum_{i=0}^{\infty} \beta^i \rho^{i+s} \sigma_u^2 = \frac{\rho^s \sigma_u^2}{1 - \beta\rho}$$

por lo que $E(y_{t-s} u_t) \neq 0$ para todo $s \geq 1$. Ello hace que, como veremos en secciones posteriores, no podamos garantizar en este caso la consistencia del estimador MCO.

La explicación de la gran diferencia que existe entre estos dos casos es que todo lo que es preciso para que el estimador mínimo cuadrático de los coeficientes del modelo

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T$$

sea consistente es que se tenga $E[x_{t-s} u_t] = 0$ para todo $s \geq 0$ y para todas las variables explicativas del modelo. Una variable que satisface esta condición se llama *predeterminada*, a diferencia de las variables *exógenas*, que son las que satisfacen la condición $E[x_{t-s} u_t] = 0$ para todo s , ya sea positivo, negativo o cero. Toda variable exógena es, claramente, predeterminada. En particular, y_{t-1} juega el papel de una de las variables explicativas x_t , siendo *predeterminada cuando u_t es ruido blanco, pero no cuando u_t tiene autocorrelación*. Por otra parte, y_{t-1} no es exógena en ningún caso.

9.2. JUSTIFICACION TEORICA DE LOS MODELOS ECONOMETRICOS DINAMICOS

Ha sido tradicional en Econometría utilizar tres ideas alternativas para justificar una especificación dinámica. Veamos brevemente dichas especificaciones.

9.2.a. El modelo de expectativas adaptativas

En 1956, P. Cagan propuso un modelo analítico en el que la demanda de saldos monetarios reales se hacía depender del valor esperado de la tasa de inflación futura:

$$\frac{M_t}{P_t} = \beta_1 + \beta_2 E_t \pi_{t+1} + u_t$$

Este modelo encierra, en realidad, distintas especificaciones, según cual sea la hipótesis que se haga acerca del modo en que los agentes económicos forman sus expectativas acerca del valor futuro de la tasa de inflación. El mecanismo de *expectativas adaptativas*, utilizado por Cagan (así como por M. Friedman en su Teoría de Consumo), es:

$$E_t \pi_{t+1} = E_{t-1} \pi_t + \lambda (\pi_t - E_{t-1} \pi_t), \quad 0 < \lambda < 1$$

que postula que los agentes modifican la expectativa que acerca de π_t formaron el período anterior teniendo en cuenta únicamente el error de predicción cometido. Es decir, si la tasa de inflación está hoy por encima de lo que se esperaba, se corrige al alza dicha expectativa, mientras que se corrige a la baja en caso contrario.

El modelo de expectativas adaptativas puede también escribirse:

$$E_t \pi_{t+1} = \lambda \pi_t + (1 - \lambda) E_{t-1} \pi_t$$

mostrando que la expectativa de inflación futura que hoy se forma es una combinación lineal del valor actual de la tasa de inflación, y de la expectativa de inflación que se formó el período anterior.

Nótese que, en el caso extremo en que $\lambda = 0$, entonces $E_t \pi_{t+1} = E_{t-1} \pi_t$, es decir, que las expectativas de inflación son estáticas y no se hacen depender del error de predicción que se haya cometido. En el otro caso extremo en que $\lambda = 1$, se tiene $E_t \pi_{t+1} = \pi_t$, y las expectativas son totalmente adaptativas, ya que se adopta como valor esperado de la inflación futura el valor que la tasa de inflación ha tomado en este período. Se ignora así la información que condujo a formar las expectativas pasadas.

Aun puede obtenerse una nueva representación para la expectativa del valor futuro de la tasa de inflación, si iteramos en la última expresión:

$$E_t \pi_{t+1} = \lambda \pi_t + \lambda(1 - \lambda) \pi_{t-1} + \lambda(1 - \lambda)^2 \pi_{t-2} + \lambda(1 - \lambda)^3 \pi_{t-3} + \dots$$

es decir, una combinación lineal del valor actual y de todos los valores pasados de la tasa de inflación con coeficientes que decrecen a una tasa igual a $1 - \lambda$.

Si se incorporan las expectativas adaptativas al modelo de regresión, se tiene:

$$\frac{M_t}{P_t} = \lambda \beta_1 + \lambda \beta_2 \pi_t + (1 - \lambda) \left(\frac{M_{t-1}}{P_{t-1}} \right) + v_t$$

Aunque la estimación de este modelo no está exenta de dificultades (como pronto vamos a ver), una vez que se hubiesen estimado sus coeficientes, el parámetro λ podría obtenerse a partir del coeficiente estimado de $\frac{M_{t-1}}{P_{t-1}}$, mientras que el parámetro β_2 se obtendría dividiendo el coeficiente de π_t por la estimación del parámetro λ , y β_1 a partir del término independiente.

9.2.b. El modelo de ajuste parcial de Nerlove

Supongamos que el nivel de capital deseado en la economía, K_t^* , es una función del nivel de producto Y_t :

$$K_t^* = \beta_1 + \beta_2 Y_t + u_t \quad [9.7]$$

Si un investigador quisiera proceder a estimar cómo varía el stock de capital deseado (u óptimo) según la economía transcurre a través de una época de recesión o de expansión, tendría el grave problema de no disponer de observaciones K_t^* . Para evitar tal dificultad, se añade al modelo anterior una ecuación que describe el mecanismo por el que el stock de capital se ajusta a su nivel deseado. Supongamos que:

$$K_t - K_{t-1} = \delta(K_t^* - K_{t-1}), \quad 0 < \delta < 1$$

que postula que el stock de capital observado varía de un período a otro en una proporción de su distancia con respecto al stock deseado. Este es el modelo conocido en Econometría como de *ajuste parcial*. Si $\delta = 1$, entonces el stock de capital es igual, en cada período, a su valor deseado, lo que sólo sería razonable en una economía en que (al contrario de lo que ocurre en economías reales) el stock no está sujeto a importantes costes de ajuste. Por otra parte, si $\delta = 0$, entonces $K_t = K_{t-1}$, y el stock de capital no cambia, con independencia de lo lejos que se halle de su valor deseado, lo que, en general, no será óptimo⁽¹⁾.

El modelo puede también escribirse:

$$K_t = \delta K_t^* + (1 - \delta)K_{t-1}$$

que postula que, en cada período, el stock de capital es una combinación lineal convexa del valor deseado y de su valor previo. También puede iterarse la expresión anterior, para obtener:

$$K_t = \delta K_t^* + \delta(1 - \delta)K_{t-1}^* + \delta(1 - \delta)^2 K_{t-2}^* + \dots$$

Al introducir el mecanismo de ajuste parcial en el modelo econométrico se tiene:

$$K_t = \delta\beta_1 + \delta\beta_2 Y_t + (1 - \delta)K_{t-1} + \delta u_t \quad [9.8]$$

Una vez estimado el modelo, el parámetro δ se obtiene del coeficiente del K_{t-1} , mientras que β_2 se obtendría dividiendo el coeficiente de Y_t por el valor

⁽¹⁾ En ocasiones, la ecuación [9.7] de demanda a largo plazo se supone determinista, mientras que la ecuación de ajuste se supone aleatoria: $K_t - K_{t-1} = \delta(K_t^* - K_{t-1}) + u_t$. El modelo que se obtiene de este modo es indistinguible del descrito en esta sección.

de δ y β_1 a partir del término independiente estimado. La ecuación [9.8] suele calificarse como demanda de capital a corto plazo, frente a [9.7], que se interpreta como demanda a largo plazo. Esta distinción permite explicar la diferencia entre el valor observado K_t y el valor deseado K_t^* , que puede ahora obtenerse a partir del modelo [9.7].

Por último, en el Problema 9.6 al final del capítulo se pide al lector que combine en un modelo una especificación de costes de ajuste junto con una de expectativas adaptativas.

9.3. MODELOS DE RETARDOS INFINITOS

En ocasiones, especialmente con observaciones frecuentes, el investigador especifica una relación del tipo:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2} + \beta_4 x_{t-3} + \dots + u_t \quad [9.9]$$

en que la variable endógena se hace depender de un número infinito de retardos de una de las variables exógenas. Es claro que el modelo anterior no puede estimarse, puesto que para estimar los coeficientes de los retardos de x_t hay que perder observaciones iniciales y, en este caso, nos quedaríamos sin grados de libertad. Se hace preciso, por tanto, cambiar la parametrización del modelo de forma que aparezca un número finito (y reducido) de parámetros. Para conseguir esta reducción en el número de parámetros es preciso hacer supuestos acerca de la evolución de los coeficientes de los sucesivos retardos de la variable exógena.

9.3.a. El modelo de Koyck

Un supuesto muy habitual es el de Koyck:

$$\beta_i = \delta \beta_{i-1} \quad |\delta| < 1 \quad \text{para todo } i \geq 3 \quad [9.10]$$

que convierte el modelo [9.9] en:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + \delta \beta_2 x_{t-1} + \delta^2 \beta_2 x_{t-2} + \delta^3 \beta_2 x_{t-3} + \dots + u_t = \\ &= \beta_1 + \beta_2 (x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^3 x_{t-3} + \dots) + u_t \end{aligned} \quad [9.11]$$

que depende de tan sólo tres parámetros: β_1 , β_2 y δ . El supuesto $|\delta| < 1$ es necesario para evitar una especificación en que el pasado de x_t tuviese un impacto más importante sobre y_t cuanto más lejano en el tiempo fuese el valor considerado de x_t .

La ganancia del modelo de Koyck es β_2 y su retardo medio es $\frac{1}{(1-\delta)}$, menor cuanto mayor es δ . Para poder estimar [9.11] es preciso truncar

polinomio de retardos que aparece dentro del paréntesis en el primer período contenido en la muestra, de modo que se tiene:

$$y_t = \beta_1 + \beta_2(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \dots + \delta^{t-1} x_1) + \delta^t \beta_2(x_0 + \delta x_{-1} + \delta^2 x_{-2} + \dots) + u_t$$

Si consideramos una nueva variable z_t definida por el contenido del primer paréntesis: $z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \dots + \delta^{t-1} x_1$, entonces, dado un valor del parámetro δ , podríamos construir una serie temporal de observaciones de dicha variable, simplemente acumulando las observaciones pasadas de la variable x_t , multiplicadas por potencias de δ , del modo que indica la definición de z_t .

Sin embargo, no puede llevarse a cabo un procedimiento similar para el segundo paréntesis, puesto que no tenemos ninguna información acerca de los valores de las variables que en él aparecen. Dicho paréntesis puede, sin embargo, tratarse como un parámetro desconocido, puesto que no depende del tiempo. Así, definiendo

$$\gamma = \beta_2(x_0 + \delta x_{-1} + \delta x_{-2} + \delta^3 x_{-3} + \dots)$$

se tiene el modelo transformado:

$$y_t = \beta_1 + \beta_2 z_t + \gamma \delta^t + u_t \quad [9.12]$$

En cuanto a generar las observaciones muestrales de la variable auxiliar z_t que hemos introducido más arriba, es conveniente notar que se tiene:

$$z_1 = x_1; \quad z_2 = x_2 + \delta z_1; \quad z_3 = x_3 + \delta z_2; \quad \dots; \quad z_t = x_t + \delta z_{t-1}$$

de modo que se puede hacer recursivamente.

9.3.b. Estimación de máxima verosimilitud del modelo de Koyck

Si suponemos que el término de error u_t sigue una distribución Normal(0, σ_u^2), entonces el logaritmo de la función de verosimilitud de [9.12] resulta:

$$\ln L(y_t, z_t/\beta_1, \beta_2, \gamma, \delta) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_1^T (y_t - \beta_1 - \beta_2 z_t - \gamma \delta^t)^2 \quad [9.13]$$

por lo que maximizar la función de verosimilitud con respecto a β_1 , β_2 , γ y δ es equivalente a minimizar la suma de los cuadrados de los residuos del modelo [9.12] coincidiendo de este modo el estimador de máxima verosimilitud con el estimador de mínimo cuadrados ordinarios.

Puesto que el parámetro δ debe tomar valores en el intervalo $(-1, 1)$, es

posible hacer una partición de dicho intervalo, por ejemplo: $-1, -0,90, \dots, 0, 0,10, 0,20, \dots, 1$, y estimar el modelo [9.12] por mínimos cuadrados ordinarios bajo cada uno de estos valores de δ :

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \Sigma_1^T z_t & \frac{\delta(1 - \delta^T)}{1 - \delta} \\ & \Sigma_1^T z_t^2 & \Sigma_1^T \delta^t z_t \\ & & \frac{\delta^2(1 - \delta^{2T})}{1 - \delta^2} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_1^T y_t \\ \Sigma_1^T z_t y_t \\ \Sigma_1^T \delta^t y_t \end{pmatrix} \quad [9.14]$$

En general, no tiene mucho sentido suponer que los coeficientes β_i del modelo [9.9] alternan en signo, por lo que δ se supone inferior a la unidad en valor absoluto, pero positivo. En tal caso, es el intervalo $(0, 1)$ el que se particiona. Tras estimar el modelo [9.12] suponiendo que δ toma, alternativamente, cada uno de los valores de la partición, se escoge aquel valor de δ que generó una suma residual menor o, equivalentemente, un R^2 más alto. Las estimaciones de β_1 y β_2 son las que se obtuvieron con dicho valor de δ . Si se quiere afinar más en los valores numéricos estimados, puede hacerse una subdivisión de un intervalo alrededor del valor de δ inicialmente estimado, y repetir el proceso.

La matriz de covarianzas obtenida de la estimación de mínimos cuadrados finalmente escogida sólo sería apropiada si el verdadero valor del parámetro δ hubiese sido conocido de antemano, de modo que dicha regresión hubiese sido la única que hubiésemos estimado. Este no ha sido el caso, y la matriz de covarianzas apropiada es la inversa de la matriz de información, pues el estimador que hemos obtenido es, en realidad, el de máxima verosimilitud.

Para ello, habría que obtener la matriz de derivadas segundas de la función de verosimilitud con respecto al vector de parámetros $(\beta_1, \beta_2, \gamma, \delta, \sigma_u^2)$, puesto que ahora se estiman todos simultáneamente. Dicha matriz de covarianzas es:

$$I^{-1} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma} \\ \hat{\delta} \\ \hat{\sigma}_u^2 \end{pmatrix} = \sigma_u^2 \begin{pmatrix} T & \Sigma_1^T z_t & \frac{\delta(1 - \delta^T)}{1 - \delta} & \Sigma_1^T \left(\beta_2 \frac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma \right) & 0 \\ & \Sigma_1^T z_t^2 & \Sigma_1^T \delta^t z_t & \Sigma_1^T \left(\beta_2 \frac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma \right) z_t & 0 \\ & & \frac{\delta^2(1 - \delta^{2T})}{1 - \delta^2} & \Sigma_1^T \left(\beta_2 \frac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma \right) \delta^t & 0 \\ & & & \Sigma_1^T \left(\beta_2 \frac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma \right)^2 & 0 \\ & & & & \frac{T}{2\sigma_u^2} \end{pmatrix}^{-1}$$

acerca de la cual cabe hacer varias observaciones:

1. La varianza del estimador $\hat{\sigma}_u^2$ es igual a $\frac{2\sigma_u^4}{T}$, y es independiente de $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\gamma}$ y $\hat{\delta}$.

2. La submatriz superior de orden 3×3 coincide con la matriz que se utilizó en [9.14] para obtener el estimador MCO. Sin embargo, la matriz de covarianzas del estimador de máxima verosimilitud del vector $(\beta_1, \beta_2, \gamma)$ es la submatriz 3×3 que aparecería en la inversa de la matriz de información, y no coincide con la inversa de la matriz en [9.14]. Esta última ignoraría el hecho de que la última variable explicativa del modelo depende del parámetro desconocido δ .

9.4. ESTIMACION CON RETARDOS DE LA VARIABLE ENDOGENA

Como mencionamos antes, en la estimación de un modelo de este tipo es fundamental distinguir entre dos casos, según que el término de error del modelo tenga o no autocorrelación.

9.4.a. El término de error no tiene autocorrelación

Consideremos el modelo:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t, \quad |\beta_2| < 1$$

cuyas variables explicativas y término de error satisfacen las siguientes propiedades:

- $E(\mathbf{u}) = \mathbf{0}_T$, $E(\mathbf{u}\mathbf{u}') = \sigma_u^2 \mathbf{I}_T$, es decir, no existe autocorrelación.
- $E(x_t u_t) = 0$ para todo t , ya que x_t es determinista.
- $E(y_{t-1} u_t) = 0$, pues aunque y_{t-1} es estocástica, como se vio en la Sección 9.1, si $|\beta_2| < 1$, y_{t-1} depende de u_{t-1}, u_{t-2}, \dots , pero no de u_t , y si este proceso es un ruido blanco, entonces se tiene el resultado citado.

- $\text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right) = \Sigma_{xx}$, matriz simétrica, definida positiva, donde:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} T-1 & \Sigma_2^T y_{t-1} & \Sigma_2^T x_t \\ & \Sigma_2^T y_{t-1}^2 & \Sigma_2^T y_{t-1} x_t \\ & & \Sigma_2^T x_t^2 \end{pmatrix}$$

que a pesar de ser denotada por $\mathbf{X}'\mathbf{X}$ incluye también, en este caso, valores de la variable endógena, por aparecer un retardo suyo como variable explicativa. Esta condición *d*) se satisface, en general, bajo el supuesto $|\beta_2| < 1$,

siempre que existan las varianzas y covarianzas de las variables explicativas x_t e y_{t-1} . Con estos supuestos, el teorema de Mann-Wald (véase Capítulo 2) asegura que:

$$plim \left(\frac{\mathbf{X}'\mathbf{u}}{T} \right) = \mathbf{0}_k$$

y que:

$$\frac{\mathbf{X}'\mathbf{u}}{\sqrt{T}} \xrightarrow{D} N(\mathbf{0}, \sigma_u^2 \Sigma_{xx}) \quad [9.15]$$

de modo que se tiene:

$$\begin{aligned} plim \hat{\beta}_{MCO} &= plim \left[\beta + \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \left(\frac{\mathbf{X}'\mathbf{u}}{T} \right) \right] = \\ &= \beta + plim \left[\left(\frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \right] plim \left(\frac{\mathbf{X}'\mathbf{u}}{T} \right) = \beta + \Sigma_{xx}^{-1} \mathbf{0}_k = \beta \end{aligned}$$

y el estimador de mínimos cuadrados ordinarios es consistente. Además, como

$$\hat{\beta}_{MCO} = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}$$

entonces:

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \left(\frac{\mathbf{X}'\mathbf{u}}{\sqrt{T}} \right)$$

y utilizando [9.15] y la Proposición 2.16:

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{D} (\Sigma_{xx})^{-1} N(\mathbf{0}, \sigma_u^2 \Sigma_{xx}) = N(\mathbf{0}, \sigma_u^2 (\Sigma_{xx})^{-1})$$

Aunque esta distribución sólo es rigurosamente válida según tiende el tamaño muestral a infinito, en la práctica suele llevarse a cabo la siguiente aproximación: en primer lugar, pasando el factor \sqrt{T} , así como el vector constante β , al miembro de la derecha se tiene que, aproximadamente, $\hat{\beta}_{MCO}$ se distribuye en muestras grandes según una $N \left[\beta, \left(\frac{\sigma_u^2}{T} \right) (\Sigma_{xx})^{-1} \right]$. En segundo lugar, como la matriz Σ_{xx} es el límite de $\frac{\mathbf{X}'\mathbf{X}}{T}$, entonces, si la muestra es suficientemente grande, la matriz Σ_{xx} puede sustituirse por esta última

matriz $\frac{\mathbf{X}'\mathbf{X}}{T}$, por lo que, en definitiva, resulta como matriz de covarianzas aproximada de $\hat{\beta}_{MCO}$ la expresión habitual $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, a pesar de la presencia de la variable endógena retardada formando parte de la matriz $\mathbf{X}'\mathbf{X}$.

En resumen, en tanto en cuanto el término de error esté libre de autocorrelación, está justificado el uso de mínimos cuadrados en un modelo que incluye retardos de la variable endógena. Puede utilizarse la matriz de covarianzas habitual de dicho estimador, quien tiene además una distribución Normal en muestras grandes, por lo que los resultados de inferencia estadística del Capítulo 4 son aproximadamente válidos.

Debe notarse que el razonamiento anterior es válido con independencia del número de retardos de la variable endógena que aparecen como variables explicativas. Todo lo que es preciso es que, considerados separadamente del resto de las variables explicativas, los coeficientes de dichos retardos satisfagan las condiciones de estacionariedad que hemos supuesto aquí, y que discutiremos en mayor extensión en el Capítulo 13.

9.4.b. Estimación cuando el término de error tiene autocorrelación

Consideremos el modelo:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t & |\beta_2| < 1 \\ u_t &= \rho u_{t-1} + \varepsilon_t & |\rho| < 1 \end{aligned} \quad [9.16]$$

donde ε_t es ruido blanco.

La existencia de autocorrelación en el término de error hace que la propiedad c) del modelo discutido en 9.4.a no se satisfaga, ya que $E(y_{t-1}u_t) \neq 0$, como se vio en la Sección 9.1.c: y_{t-1} depende de u_{t-1} a través del modelo, pero u_{t-1} y u_t están relacionados con la estructura autorregresiva del término de error. En consecuencia, y_{t-1} y u_t están correlacionados. El estimador de mínimos cuadrados es sesgado, y además su sesgo no desaparece al aumentar el tamaño muestral.

Por ejemplo, en el caso más sencillo del modelo [9.16], con $\beta_1 = \beta_3 = 0$, se tiene:

$$plim \hat{\beta}_{2\ MCO} = \beta_2 + \frac{plim \ 1/T(\sum_2^T y_{t-1}u_t)}{plim \ 1/T(\sum_2^T y_{t-1}^2)}$$

y si los momentos muestrales convergen en probabilidad a sus análogos poblacionales, como $E(y_{t-1}u_t) = \frac{\rho\sigma_u^2}{1-\beta_2\rho}$ y $E(y_{t-1}^2)$ también es distinto de cero (véase Problema 9.8), entonces se tiene que el estimador mínimo cuadrático no es consistente. Si esto ocurre en un modelo sencillo, no debe esperarse que la situación sea más favorable en modelos más complejos. En efecto,

este resultado de inconsistencia es válido, en general, siempre que aparezcan conjuntamente retardos de la variable endógena y autocorrelación del término de error.

9.4.c. El estimador de variables instrumentales

El procedimiento para obtener estimaciones consistentes de un modelo de este tipo se conoce como *estimador de variables instrumentales*. Una variable instrumental es una variable z_t que satisface tres condiciones: a) no está incluida en el modelo como variable explicativa, b) está incorrelacionada con el término de error $E(z_t u_t) = 0$, y c) está correlacionada con la variable para la cual hace de instrumento, en este caso y_{t-1} .

Por ejemplo, en el modelo [9.16], el primer retardo de la variable exógena x_{t-1} satisface estas tres condiciones, pues aparte de que, por ser determinista: $E(x_{t-1} u_{t-1}) = 0$, también se tiene que x_{t-1} influye sobre y_{t-1} a través del propio modelo [9.16] escrito para el período $t - 1$.

En general, en el vector x_t tan sólo habrá unas variables que no satisfagan la condición $E(x_t u_t) = 0$, y son estas variables las que necesitan de variables instrumentales. Es decir, los vectores x_t y z_t tendrán en común aquellas variables que están incorrelacionadas con el término de error. Así, en el ejemplo anterior $x_t = (1, y_{t-1}, x_t)$, mientras que z_t podría ser el vector $z_t = (1, x_{t-1}, x_t)$. Finalmente, el estimador de variables instrumentales viene dado por la expresión matricial:

$$\hat{\beta}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} \quad [9.17]$$

donde \mathbf{Z} denota la matriz $T \times k$ de observaciones muestrales de las variables que componen el vector z_t , y suponemos que $\mathbf{Z}'\mathbf{X}$ es invertible. Volviendo al ejemplo anterior, el estimador de variables instrumentales se obtendría mediante.

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}_{VI} = \begin{pmatrix} T-1 & \Sigma_2^T y_{t-1} & \Sigma_2^T x_t \\ \Sigma_2^T x_{t-1} & \Sigma_2^T x_{t-1} y_{t-1} & \Sigma_2^T x_t x_{t-1} \\ \Sigma_2^T x_t & \Sigma_2^T x_t y_{t-1} & \Sigma_2^T x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_2^T y_t \\ \Sigma_2^T x_{t-1} y_t \\ \Sigma_2^T x_t y_t \end{pmatrix} \quad [9.18]$$

donde, como puede verse, la matriz $\mathbf{Z}'\mathbf{X}$ dista de ser simétrica.

En cuanto a la correlación que debe existir entre una variable instrumental y la variable explicativa para la que se utiliza como instrumento, cabe observar lo siguiente: a) es importante que dicha correlación exista, puesto que como muestra la expresión [9.18], la variable instrumental x_{t-1} sustituye parcialmente a la variable y_{t-1} en la estimación del modelo econométrico. pero, sin embargo, b) dicha correlación no puede ser muy importante, puesto que entonces también existiría una correlación apreciable entre la variable instrumental z_t y el término de error u_t , que es lo que motivó la necesidad de la variable instrumental.

En el modelo [9.16] también podría utilizarse otro retardo, por ejemplo x_{t-2} , como variable instrumental. La diferencia es que la relación entre esta variable e y_{t-1} se hace más indirecta; así, x_{t-2} está directamente correlacionada con y_{t-2} , quien a su vez está correlacionada con y_{t-1} a través del modelo [9.16], y ésta con y_t .

El estimador de variables instrumentales del modelo [9.16] es, en general, sesgado, pues la variable y_{t-1} aparece en la matriz $Z'X$. Para su insesgo debería ocurrir que $E[(Z'X)^{-1}Z'u] = \mathbf{0}_k$, pero ello envuelve en general funciones muy complejas cuya esperanza desconocemos. Afortunadamente, podremos afirmar que el estimador es consistente bajo las condiciones de la siguiente proposición:

Proposición 9.1. Sea Z una matriz $T \times k$ de observaciones de las variables z_1, z_2, \dots, z_k , quizá aleatorias. Sea z'_t la fila t de Z y supongamos que se tiene:

1. $E(z'_t u_t) = \mathbf{0}_k$ para todo t .
2. $plim \left(\frac{Z'X}{T} \right) = \Sigma_{zx}$, $plim \left(\frac{Z'Z}{T} \right) = \Sigma_{zz}$, ambas no singulares y finitas.

Entonces, el estimador de variables instrumentales es consistente.

Demostración:

$$\begin{aligned} plim(\hat{\beta}_{VI}) &= \beta + plim \left[\left(\frac{Z'X}{T} \right)^{-1} \left(\frac{Z'u}{T} \right) \right] = \\ &= \beta + \left[plim \left(\frac{Z'X}{T} \right) \right]^{-1} plim \left(\frac{Z'u}{T} \right) = \beta + \Sigma_{zx}^{-1} \mathbf{0}_k = \beta \end{aligned}$$

donde se ha utilizado que los instrumentos satisfacen: $plim \frac{Z'u}{T} = \mathbf{0}_k$.

Así pues, la consistencia de $\hat{\beta}_{VI}$ proviene de la ausencia de correlación entre instrumentos y término de error, con independencia de que éste tenga o no autocorrelación. En ausencia de autocorrelación, podemos caracterizar la distribución asintótica de $\hat{\beta}_{VI}$:

Proposición 9.2. Dado el modelo $y_t = x'_t \beta + u_t$, donde x_t es el vector de variables explicativas, que puede incluir algunos retardos de la variable endógena, y u_t , el término de error es un ruido blanco, sea Z la matriz $T \times k$ de observaciones de las variables z_1, z_2, \dots, z_k , y supongamos que:

1. $E(z'_t u_t) = \mathbf{0}_k$ para todo t .
2. $plim \left(\frac{Z'Z}{T} \right) = \Sigma_{zz}$, simétrica, definida positiva.
3. $plim \left(\frac{Z'X}{T} \right) = \Sigma_{zx}$, no singular.

Entonces el estimador $\hat{\beta}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ es consistente y se tiene:

$$\sqrt{T}(\hat{\beta}_{VI} - \beta) \xrightarrow{D} N(\mathbf{0}, \sigma_u^2(\Sigma_{zx})^{-1}\Sigma_{zz}(\Sigma_{zx}^{-1})')$$

Demostración. La consistencia del estimador ha sido el objeto de la proposición anterior. En segundo lugar, el teorema de Mann-Wald asegura que bajo los supuestos 1, 2, 3 se tiene:

$$plim\left(\frac{\mathbf{Z}'\mathbf{u}}{T}\right) = \mathbf{0}_k$$

y

$$\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \xrightarrow{D} N(\mathbf{0}_k, \sigma_u^2\Sigma_{zz})$$

y como $\sqrt{T}(\hat{\beta}_{VI} - \beta) = \left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1} \left(\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}}\right)$, esta función converge en distribución a $(\Sigma_{zx})^{-1}N(\mathbf{0}_k, \sigma_u^2\Sigma_{zz})$, que es una distribución $N(\mathbf{0}_k, \sigma_u^2\Sigma_{zx}^{-1}\Sigma_{zz}(\Sigma_{zx}^{-1})')$.

El resultado anterior justifica que en muestras grandes se utilice como matriz de covarianzas del estimador de variables instrumentales:

$$\text{Var}(\hat{\beta}_{VI}) = \frac{\sigma_u^2}{T} (\Sigma_{zx}^{-1})\Sigma_{zz}[(\Sigma_{zx})^{-1}]'$$

y si se utilizan las matrices de momentos muestrales $\frac{\mathbf{Z}'\mathbf{X}}{T}$, $\frac{\mathbf{Z}'\mathbf{Z}}{T}$ para aproximar sus límites respectivos Σ_{zx} , Σ_{zz} , entonces se tiene como matriz de covarianzas:

$$\text{Var}(\hat{\beta}_{VI}) = \left(\frac{\sigma_u^2}{T}\right) \left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1} \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right) \left[\left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1}\right]' = \sigma_u^2(\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Z})[(\mathbf{Z}'\mathbf{X})^{-1}]'$$

Por último, el parámetro σ_u^2 se estimaría dividiendo la suma residual por el número de grados de libertad, $T - k$. Sin embargo, es importante tener en cuenta que, una vez más, los residuos deben calcularse utilizando las variables originales del modelo, es decir:

$$\hat{\sigma}_u^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{VI})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{VI})}{T - k}$$

El resultado anterior no puede generalizarse fácilmente al caso en que el término de error tiene autocorrelación, por lo que suele utilizarse la matriz de covarianzas anterior incluso en tal caso, aun a sabiendas que no es sino un

aproximación. Un modelo de este tipo debe estimarse, a ser posible, por el método de máxima verosimilitud, que describimos en la Sección 9.7.

9.4.d. Contraste de exogeneidad de Hausman y Wu

La aparición de retardos de la variable endógena como variables explicativas en presencia de autocorrelación es un caso muy habitual en que se necesita utilizar un procedimiento de variables instrumentales si se pretende obtener estimaciones consistentes. Pero es aconsejable cuestionarse acerca de las propiedades de exogeneidad del resto de las variables explicativas, pues, de no satisfacerse, obtendríamos igualmente estimadores inconsistentes.

Hausman (1978) y Wu (1973) sugieren escribir el modelo a estimar, distinguiendo entre las r variables explicativas Y_1 que pueden estar correlacionadas con el término de error de aquellas $k - r$ variables Z_1 cuya ortogonalidad a u_i no se cuestiona:

$$y = X\beta + u = Y_1\alpha + Z_1\delta + u$$

y supongamos que se dispone de una lista de instrumentos para Y_1 , en caso de que se necesitasen.

El contraste consiste en estimar inicialmente el modelo por MCO y obtener la suma residual SR_0 . A continuación se estiman regresiones auxiliares de las variables en Y_1 sobre los instrumentos, y se sustituye Y_1 por la matriz de valores previstos \hat{Y}_1 en el modelo inicial, que se estima por MCO, obteniendo la suma residual SR_1 . El estadístico:

$$(\hat{\beta}_{MCO} - \hat{\beta}_{VI})' [\text{Var}(\hat{\beta}_{VI}) - \text{Var}(\hat{\beta}_{MCO})]^{-1} (\hat{\beta}_{MCO} - \hat{\beta}_{VI}) = \frac{SR_0 - SR_1}{\hat{\sigma}_u^2} \sim \chi_r^2$$

bajo la hipótesis nula de que todas las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría tal supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.

La lógica de este contraste puede utilizarse en situaciones distintas de la aquí analizada, constituyendo la filosofía general de «un test tipo Hausman». Supongamos que se dispone de dos estimadores: uno, $\hat{\beta}_1$, que bajo H_0 es consistente y eficiente, pero que es inconsistente bajo la hipótesis alternativa y, un segundo estimador, $\hat{\beta}_2$, que es consistente tanto bajo H_0 como bajo H_1 , aunque sea, posiblemente, ineficiente. Bajo la hipótesis nula, la diferencia $\hat{q} = \hat{\beta}_1 - \hat{\beta}_2$ converge en probabilidad a cero, pero no bajo la hipótesis alternativa. Hausman probó que si la distribución asintótica de $\sqrt{T}(\hat{\beta}_2 - \beta)$ está bien definida bajo H_0 , entonces la covarianza asintótica de $\sqrt{T}(\hat{\beta}_1 - \beta)$ y $\sqrt{T}\hat{q}$ es igual a cero. Como consecuencia, la matriz de covarianzas del vector \hat{q} es igual a la diferencia de matrices: $\text{Var}(\hat{q}) = \text{Var}(\hat{\beta}_2) - \text{Var}(\hat{\beta}_1)$.

Aun reemplazando estas matrices por sus estimaciones, se tiene que el estadístico

$$\sqrt{T}\hat{\mathbf{q}} \rightarrow N(\mathbf{0}, \text{Var}(\hat{\mathbf{q}}))$$

y, por tanto, $T\hat{\mathbf{q}}' \text{Var}(\hat{\mathbf{q}})\hat{\mathbf{q}}$, se distribuye como una variable χ_p^2 , siendo p el número de restricciones que se contrastan.

En el contraste de exogeneidad, el estimador MCO es el $\hat{\beta}_1$ en la notación del test de Hausman, mientras que el estimador de VI es el $\hat{\beta}_2$.

Ejemplo 9.1. A continuación analizamos diversas especificaciones dinámicas entre Consumo y PIB, con la misma muestra con que hemos trabajado en capítulos previos.

En ocasiones, la presencia de autocorrelación denota una mala especificación dinámica. La inicial función de consumo estática de la Sección 7.6 puede reformularse de diversas maneras:

$$(a) \quad C_t = 113,0 + 0,43 Y_t + 0,44 C_{t-1}$$

$$(48,6) \quad (0,04) \quad (0,05)$$

$$\bar{R}^2 = 0,999; \quad \hat{\sigma}_u = 101,4; \quad DW = 1,35$$

con matriz de covarianzas:

$$\text{Var}(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \begin{pmatrix} 2.366,3 & & \\ -0,4454 & 0,1592 \cdot 10^{-2} & \\ 0,3520 & -0,2079 \cdot 10^{-2} & 0,2742 \cdot 10^{-2} \end{pmatrix}$$

o también:

$$(b) \quad C_t = 173,0 + 0,38 Y_t + 0,39 Y_{t-1}$$

$$(74,3) \quad (0,09) \quad (0,09)$$

$$\bar{R}_2 = 0,998; \quad \hat{\sigma}_u = 145,8; \quad DW = 0,72$$

la primera de las cuales presenta una situación de autocorrelación mucho más favorable que la del modelo estático del Capítulo 7. De hecho, en (a) no rechazaríamos la hipótesis nula de ausencia de autocorrelación, por lo que la estimación que presentamos puede aceptarse como consistente. De acuerdo con este modelo, un incremento de 1.000 millones de pesetas de PIB genera un incremento de 430 millones de consumo en el mismo año, pero también un aumento de $1.000(0,43)(0,44) = 189,2$ millones de consumo el año siguiente, más otros $1.000(0,43)(0,44^2) = 83,2$ millones, luego

otros $1.000(0,43)(0,44^3) = 36,6$ millones, y así sucesivamente. El efecto a largo plazo es de:

$$\frac{(1.000)(0,43)}{1 - 0,44} = 767,86$$

por lo que podría hablarse de una propensión marginal a largo plazo de 0,768.

El modelo (b) tiene todavía importante autocorrelación, como ilustra su estadístico Durbin-Watson, por lo que la especificación (a) es preferible. Sin embargo, en tanto en cuanto se esté dispuesto a aceptar la exogeneidad de Y_t , la estimación de (b) es también consistente. La estimación de este modelo sugiere una propensión al consumo a largo plazo de 0,77, como la del modelo (a), pero con un efecto distribuido a partes iguales entre este año y el siguiente; es decir, una respuesta más corta en duración, aunque igual en magnitud, que la del modelo (a).

El investigador puede estar interesado en contrastar el grado en que la propensión marginal a consumir a largo plazo, $PMLP = \frac{\beta_2}{1 - \beta_3}$, se aparta de 1. Es decir, queremos contrastar $H_0: \beta_2 = 1 - \beta_3$, donde β_1 , β_2 y β_3 son los coeficientes del modelo (a) frente a la alternativa $\beta_2 < 1 - \beta_3$. La discrepancia de tal restricción es:

$$d = \hat{\beta}_2 - (1 - \hat{\beta}_3) = 0,43 - (1 - 0,44) = -0,13$$

y su varianza:

$$\begin{aligned} \text{Var}[\hat{\beta}_2 - (1 - \hat{\beta}_3)] &= \text{Var} \hat{\beta}_2 + \text{Var} \hat{\beta}_3 + 2 \cdot \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \\ &= 0,1592 \cdot 10^{-2} + 0,2742 \cdot 10^{-2} - 0,4158 \cdot 10^{-2} = 0,0176 \cdot 10^{-2} \end{aligned}$$

por lo que $\frac{(-0,13)^2}{0,0176 \cdot 10^{-2}} = 96,02$ debe compararse con el valor crítico para el contraste de una cola de una variable χ_1^2 , al que excede a cualquier nivel de significación, por lo que *rechazamos* la hipótesis nula de que la propensión marginal a consumir a largo plazo es igual a uno.

Incorporemos ahora como variable explicativa adicional el deflactor del consumo, obtenido a partir de los deflatores del consumo público y privado, ponderados de acuerdo con los niveles de las variables de consumo asociadas. La fuente es, de nuevo, Molinas, Sebastián, Zabalza (1991). Además del estadístico Durbin-Watson, presentamos también el estadístico h de Durbin en aquellos modelos en que resulta relevante. A diferencia de las regresiones anteriores, transformamos todas las variables en logaritmos.

Cabría esperar un efecto negativo sobre el consumo de los precios de consumo, medidos por el deflactor, pero éste no aparece en las estimaciones de la Tabla 9.1. Por el contrario, aparecen indicios evidentes acerca de las características dinámicas de la relación entre Consumo y PIB. Los estadísticos DW y Q (éste con 15 grados de libertad) sugieren la existencia de autocorrela-

ción en los modelos que no incluyen el valor retardado del consumo, pero sugieren que la columna (3) está estimada consistentemente.

En el contraste de una cola del estadístico h , se rechaza la hipótesis nula de ausencia de autocorrelación al 5 por 100 de significación cuando $h > 1.645$, si la alternativa es de autocorrelación positiva, y cuando $h < -1.645$ si la alternativa es de autocorrelación negativa. De nuevo, el modelo de la columna (3) parece correcto.

TABLA 9.1. Variable dependiente: Consumo agregado (privado + público)

Período muestral: 1954-1988
Variables en logaritmos

	(1)	(2)	(3)	(4)	(5)
<i>Constante</i>	0,757 (0,898)	0,161 (0,074)	0,102 (0,050)	0,052 (0,051)	-0,006 (0,060)
<i>PIB</i>	0,963 (0,010)	0,419 (0,125)	0,531 (0,052)	0,724 (0,095)	0,846 (0,143)
<i>Inflación</i>	0,165 (0,097)	0,015 (0,084)	0,017 (0,057)	0,057 (0,056)	0,079 (0,063)
<i>PIB</i> ₋₁	—	0,539 (0,123)	—	-0,398 (0,168)	-0,630 (0,276)
<i>Consumo</i> ₋₁	—	—	0,445 (0,054)	0,660 (0,103)	0,779 (0,161)
<i>Elasticidad a largo plazo</i>	0,963	0,958	0,957	0,959	0,977
<i>SR</i> * 10 ²	1,62	0,99	0,49	0,41	0,42
$\hat{\sigma}$ * 10 ²	2,29	1,82	1,28	1,19	1,24
\bar{R}^2	0,997	0,998	0,999	0,999	0,999
<i>DW</i>	0,41	0,74	1,68	0,74	2,24
<i>Durbin h</i>	—	—	0,98	4,6	—
<i>T</i>	34	34	34	34	32
<i>Q</i> (15)	123,7 (0,0)	57,5 (0,0)	4,7 (0,99)	4,4 (0,99)	9,2 (0,87)

Dadas las dificultades del modelo (4), estimamos nuevamente este modelo por variables instrumentales, utilizando como instrumentos: una constante, el PIB contemporáneo, así como dos retardos suyos; valores similares de la inflación, así como el consumo retardado dos períodos, obteniendo los resultados de la columna (5) de la tabla, que presenta variaciones algo

importantes en los coeficientes, como la presencia de autocorrelación residual habría hecho esperar. Actuando así, hemos supuesto exogeneidad del PIB.

Es importante observar la considerable estabilidad con que se estima la elasticidad a largo plazo entre el Consumo y la Renta, en torno a 0,960, siendo algo más elevada para la estimación de variables instrumentales.

Ejemplo 9.2. En la función de inversión de la Sección 7.6 tenemos otra situación en que la existencia de autocorrelación puede estar denotando una incorrecta especificación dinámica. En particular, la inversión puede responder al PIB con cierta inercia. Sin entrar a discutir en detalle las razones, la segunda columna de cada panel de la Tabla 9.2 recoge las estimaciones del modelo:

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \beta_3 r_t + u_t$$

donde I_t , Y_t y r_t denotan la inversión, el PIB y los tipos de interés reales del período t .

Para poder establecer comparaciones, la primera columna de la Tabla 9.2 contiene las estimaciones MCO que ya se presentaron en la Tabla 7.2. Como puede verse, el retardo del PIB aparece significativo en ambas regresiones. El efecto a largo plazo del PIB se mantiene similar al de las regresiones estáticas. En el modelo con tipo a corto, dicho efecto era de 15.900.000 pesetas de 1980 de aumento de la inversión por cada 100.000.000 de pesetas de 1980 de aumento del PIB. En el modelo dinámico, ese efecto se estima en 16.600.000 pesetas. Algo similar ocurre en el modelo con tipos a largo plazo.

TABLA 9.2. Variable endógena: Inversión pesetas constantes de 1980

<i>Constante</i>	8,344 (2,755)	4,750 (2,702)	5,404 (2,032)	10,546 (2,726)	7,887 (2,720)	6,289 (1,924)
<i>PIB</i>	0,159 (0,019)	0,652 (0,174)	0,536 (0,136)	0,155 (0,020)	0,593 (0,187)	0,467 (0,135)
<i>PIB₋₁</i>	—	-0,486 (0,171)	-0,384 (0,132)	—	-0,433 (0,184)	-0,319 (0,132)
<i>Tipo de interés a corto</i>	-0,311 (0,114)	-0,396 (0,103)	-0,314 (0,094)	—	—	—
<i>Tipo de interés a largo</i>	—	—	—	-0,303 (0,131)	-0,374 (0,122)	-0,247 (0,105)
\bar{R}^2	0,76	0,82	0,77	0,76	0,78	0,71
<i>DW</i>	0,61	0,86	1,87	0,58	0,75	1,63
$\hat{\sigma}$	2,69	2,33	1,57	2,80	2,54	1,60
<i>Q(12)</i>	35,3 (0,0)	27,0 (0,0)	5,41 (0,91)	37,7 (0,0)	27,9 (0,0)	8,5 (0,67)

En ambas regresiones, los estadísticos DW y Q (Ljung-Box) reflejaban claros indicios de autocorrelación, por lo que se procedió a estimar por MCG, utilizando el procedimiento de Cochrane-Orcutt tras ajustar un modelo autorregresivo de segundo orden a los residuos iniciales, con los resultados que aparecen en la tercera columna.

Como puede apreciarse al comparar con la Tabla 7.2, la incorrecta especificación dinámica suponía una subvaloración del efecto de los tipos de interés reales al estimar por MCG.

Si se transforman las variables inversión, PIB y deflactor en logaritmos y se estiman modelos análogos a los anteriores, se tienen los resultados de la Tabla 9.3, con una interpretación similar. Ahora puede hablarse de una *elasticidad* total de la Inversión respecto del PIB en torno a 0,80. Esta estimación es independiente de que se consideren los tipos reales a corto o largo plazo en el modelo, y de que se examine el modelo estático o su versión dinámica, si bien sólo esta última es aceptable.

TABLA 9.3. Variables en logaritmos

<i>Constante</i>	-0,05 (0,34)	-0,72 (0,39)	-0,20 (0,29)	0,06 (0,36)	-0,50 (0,41)	-0,02 (0,29)
<i>PIB</i>	0,70 (0,07)	2,77 (0,75)	2,26 (0,57)	0,69 (0,07)	2,51 (0,81)	1,94 (0,57)
<i>PIB₋₁</i>	—	-1,95 (0,70)	-1,52 (0,53)	—	-1,72 (0,76)	-1,24 (0,53)
<i>Tipo de interés a corto</i>	-0,010 (0,004)	-0,013 (0,003)	-0,011 (0,003)	—	—	—
<i>Tipo de interés a largo</i>	—	—	—	-0,010 (0,004)	-0,012 (0,004)	-0,009 (0,004)
\bar{R}^2	0,81	0,83	0,82	0,80	0,80	0,76
<i>DW</i>	0,62	0,96	1,96	0,58	0,85	1,70
$\hat{\sigma}$	0,080	0,069	0,049	0,084	0,075	0,051
<i>Q(12)</i>	35,1 (0,0)	24,4 (0,01)	4,7 (0,94)	38,2 (0,0)	22,6 (0,02)	8,2 (0,70)

9.5. EFICIENCIA RELATIVA DE LOS ESTIMADORES DE VARIABLES INSTRUMENTALES

Hasta aquí hemos desarrollado nuestra presentación del estimador de variables instrumentales *como si* se dispusiese de un número de instrumentos *igual* al de variables explicativas. En tal caso, no existe diferencia entre instrumentos *y* variables instrumentales. Más generalmente, se dispondrá de un número

mayor de instrumentos que de variables explicativas, situación que denominamos como de *sobreidentificación*⁽²⁾. En tal caso, habría muchas formas de construir las *variables instrumentales* que precisamos para obtener consistencia.

Debe observarse que, como muestra el sistema de ecuaciones [9.17] que define al estimador de variables instrumentales, éstas deben configurar un vector de k variables, igual al número de variables explicativas del modelo original. De otro modo, la matriz $\mathbf{Z}'\mathbf{X}$ no sería cuadrada y menos aún invertible⁽³⁾.

Pero acabamos de ver cómo la matriz de covarianzas del estimador de variables instrumentales depende de los valores de éstas, por lo que el modo en que los instrumentos se «combinan» para generar variables instrumentales influye sobre la eficiencia de un estimador de variables instrumentales respecto a otro estimador de su misma clase. Por ejemplo, consideremos el modelo:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \beta_3 x_{1t} + \beta_4 x_{2t} + \beta_5 x_{3t} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned} \quad [9.19]$$

en el que las variables x_{1t} , x_{2t} y x_{3t} , supuestas deterministas, están incorrelacionadas con el término de error, y son instrumentos válidos⁽⁴⁾. Pero sólo necesitamos una variable instrumental para y_{t-1} , y se trataría de buscar cuál de todas las posibles minimiza la varianza del estimador resultante. Nótese que cualquier combinación lineal $z_t = a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t}$ de los instrumentos es asimismo un instrumento válido, en particular, cualquiera de ellos por separado.

Una posibilidad consiste en generar la variable instrumental que presente mayor correlación con y_{t-1} . Para ello, se estima una regresión auxiliar de esta variable sobre los tres instrumentos de que disponemos, para obtener la variable generada, \hat{y}_{t-1} , que será una combinación lineal de x_{1t-1} , x_{2t-1} y x_{3t-1} y, como tal, una variable instrumental válida. La utilización del vector $\mathbf{z}'_t = (\hat{y}_{t-1}, x_{1t}, x_{2t}, x_{3t})$ genera el denominado *estimador de mínimos cuadrados en dos etapas*, $\hat{\beta}_{MC2E}$.

En general, la regresión de *todas* las variables explicativas sobre *todos* los instrumentos, recogidos en la matriz de observaciones \mathbf{W} , produce los coeficientes $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}$ y genera el vector de «variables explicadas» $\hat{\mathbf{X}} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}$, que utilizadas como variables instrumentales $\mathbf{Z} = \hat{\mathbf{X}}$ conducen finalmente al estimador de *mínimos cuadrados de dos etapas*:

$$\hat{\beta}_{VI} = [\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$$

⁽²⁾ Si se dispusiese de *menos* instrumentos que variables explicativas, el modelo no podría estimarse consistentemente.

⁽³⁾ Visto de otro modo, tendríamos de un número de ecuaciones normales inferior al número de coeficientes que queremos estimar.

⁽⁴⁾ Los retardos de las variables x_{it} son asimismo instrumentos válidos, y podrían incorporarse en la generación de variables instrumentales, aunque ignoramos este aspecto en nuestra discusión.

Por comparación con [9.17], es claro que las variables instrumentales utilizadas ahora son $\mathbf{Z} = \hat{\mathbf{X}}$ y, por consiguiente, su matriz de covarianzas es:

$$\text{Var}(\hat{\beta}_{\text{MC2E}}) = \sigma_u^2 (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1} (\hat{\mathbf{X}}' \hat{\mathbf{X}}) (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1} = \sigma_u^2 (\hat{\mathbf{X}}' \hat{\mathbf{X}})$$

donde la varianza del término de error se estima: $\hat{\sigma}_u^2 = \text{SR}_{\text{VI}}/(T - k)$. La última igualdad en la expresión de la matriz $\text{Var}(\hat{\beta}_{\text{VI}})$ es el motivo del Problema 9.1 al final del capítulo, donde también se solicita demostrar que el estimador MC2E puede también obtenerse sustituyendo *totalmente* a las variables que están correlacionadas con el término de error por sus variables instrumentales.

En el Capítulo 18 volvemos a encontrar este estimador en el contexto de modelos econométricos de ecuaciones simultáneas, y las propiedades que allí probaremos son aplicables con generalidad, por lo que pueden utilizarse en los modelos dinámicos que aquí estudiamos. En particular, puede probarse que *MC2E es el estimador lineal de variables instrumentales eficiente*, en el sentido de tener mínima matriz de covarianzas entre los estimadores que utilizan como variables instrumentales combinaciones lineales de los instrumentos disponibles⁽⁵⁾.

En un contexto más general debe permitirse la posibilidad de que el término de error tenga heteroscedasticidad. White (1982) consideró esta situación, incluso en el caso más general en que se desconoce la forma funcional de dicha heteroscedasticidad. En el trabajo citado se demuestra que el estimador que utiliza como variables instrumentales las componentes del vector \mathbf{z}_i , de dimensión $k \times 1$,

$$\hat{\beta}_{\text{VI}} = (\mathbf{Z}' \mathbf{X})^{-1} \mathbf{Z}' \mathbf{y} \quad [9.20]$$

tiene como matriz de covarianzas aproximada en muestras finitas:

$$\text{Var}(\hat{\beta}_{\text{VI}}) = (1/T \Sigma_1^T \mathbf{z}_i \mathbf{z}_i')^{-1} (1/T \Sigma_1^T \hat{u}_i^2 \mathbf{z}_i \mathbf{z}_i') [(1/T \Sigma_1^T \mathbf{z}_i \mathbf{z}_i')^{-1}]'$$

Por lo que, en el caso del estimador MC2E:

$$\text{Var}(\hat{\beta}_{\text{MC2E}}) = (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1} (1/T \Sigma_1^T \hat{u}_i^2 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i') [(\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1}]'$$

donde $\hat{\mathbf{X}} = \mathbf{W}(\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{X}$.

En esta situación, más general, se tiene que la diferencia $\text{Var}(\hat{\beta}_{\text{VI}}) - \text{Var}(\hat{\beta}_{\text{MC2E}})$ es semidefinida positiva, por lo que el estimador MC2E sigue siendo relativamente más eficiente que otro estimador de variables instrumentales.

⁽⁵⁾ La inclusión de retardos de las variables exógenas como instrumentos aumentaría el conjunto de información utilizado en la construcción del estimador del MC2E. Así hay un estimador de MC2E para cada conjunto de instrumentos que se considere. Al utilizar más información, el estimador MC2E resultante sería más eficiente que otro estimador similar que utilizase menos información; sin embargo, el uso de retardos obliga a prescindir de algunas observaciones muestrales, lo que disminuye algo la eficiencia.

Ahora bien, introduciendo más generalidad, si permitimos que las variables instrumentales disponibles no sean necesariamente independientes del término de error del modelo, White probó en el trabajo citado que existe un estimador aún más eficiente que el estimador MC2E. Este estimador recibe el nombre de estimador de *variables instrumentales en dos etapas*. En una primera etapa se estima el modelo por un procedimiento de variables instrumentales, por ejemplo, mediante MC2E. Tras obtener los residuos $\hat{\mathbf{u}}$, correspondientes a este estimador, se obtiene una segunda etapa del estimador:

$$\hat{\beta}_{\text{V12E}} = [\mathbf{X}'\mathbf{Z}(1/T\Sigma_1^T \hat{\mathbf{u}}_t^2 \mathbf{z}_t \mathbf{z}_t')^{-1} \mathbf{Z}'\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{Z}(1/\Sigma_u^T \hat{\mathbf{u}}_t^2 \mathbf{z}_t \mathbf{z}_t')^{-1} \mathbf{Z}'\mathbf{y}]$$

cuya matriz de covarianzas puede aproximarse por:

$$\text{Var}(\hat{\beta}_{\text{V12E}}) = [\mathbf{X}'\mathbf{Z}(1/T\Sigma_1^T \hat{\mathbf{u}}_t^2 \mathbf{z}_t \mathbf{z}_t')^{-1} \mathbf{Z}'\mathbf{X}]^{-1}$$

9.5.a. Contraste de Sargan de validez de instrumentos

La elección de los instrumentos a utilizar en la estimación no está exenta de elementos subjetivos por parte del investigador. Por ello es importante disponer de un contraste de validez de los posibles instrumentos.

Es fácil ver que el estimador MC2E minimiza la función objetivo: $S_{\text{MC2E}} = \hat{\mathbf{u}}_{\text{V1}}' \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'\hat{\mathbf{u}}_{\text{V1}}$, y Sargan demostró que el estadístico $S_{\text{MC2E}}/\hat{\sigma}_u^2$ obedece, bajo la hipótesis nula de validez de los instrumentos utilizados, una distribución chi-cuadrado con grados de libertad igual a $p - k$, siendo p el número total de instrumentos utilizados y k el número de variables explicativas del modelo original. Si este estadístico toma un valor demasiado elevado (superior a la variable χ_{p-k}^2 al nivel de confianza escogido), debe aceptarse que o bien el modelo está mal especificado o no todos los instrumentos utilizados son válidos, es decir, algunos de ellos están correlacionados con el término de error.

El valor numérico de S_{MC2E} puede obtenerse como la suma explicada (suma de valores previstos) en una regresión de los residuos de las variables instrumentales $\hat{\mathbf{u}}_{\text{V1}}$ sobre el vector de variables en \mathbf{W} . Conviene insistir en que el vector $\hat{\mathbf{u}}_{\text{V1}}$ se calcula con las variables del modelo original.

9.6. CONTRASTACION DE HIPOTESIS CON EL ESTIMADOR MC2E

Una vez obtenido el estimador MC2E, si se quiere contrastar un conjunto de hipótesis lineales como $\mathbf{R}\beta = \mathbf{r}$, donde \mathbf{R} es $q \times k$ y \mathbf{r} es $k \times 1$, el estadístico

$$\frac{(\mathbf{R}\hat{\beta} - \mathbf{r})' [\mathbf{R}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \mathbf{R}'] (\mathbf{R}\hat{\beta} - \mathbf{r})}{\hat{\sigma}_u^2}$$

se distribuye asintóticamente como una variable χ_q^2 , siendo q el número de restricciones y $\hat{\sigma}_u^2$ el estimador obtenido con los residuos MC2E.

Sería interesante encontrar una formulación en función de las sumas residuales restringida y sin restringir, pero hay que tener cierto cuidado en la definición de dichas sumas. Por comparación con el estimador MCO, que minimiza $(y - X\beta)'(y - X\beta)$, podemos interpretar el estimador MC2E como aquel que minimiza la expresión $S_{MC2E} = (y - X\beta)'W(W'W)^{-1}W'(y - X\beta)$, cuyas condiciones de primer orden son $X'W(W'W)^{-1}W'(y - X\beta) = 0_k$, que coinciden con las ecuaciones normales del estimador MC2E.

Si se minimiza ahora S_{MC2E} sujeta a las restricciones $R\beta = r$, se tiene:

$$\hat{\beta}_{MC2E}^R = \hat{\beta}_{MC2E} - (\hat{X}'\hat{X})^{-1}R'[R(\hat{X}'\hat{X})^{-1}R']^{-1}(R\hat{\beta}_{MC2E} - r)$$

y se tiene para los residuos la expresión:

$$\hat{u}_{MC2E}^R = y - X\hat{\beta}_{MC2E}^R = \hat{u}_{MC2E} + X(\hat{X}'\hat{X})^{-1}R'[R(\hat{X}'\hat{X})^{-1}R']^{-1}(R\hat{\beta}_{MC2E} - r)$$

y el numerador del estadístico χ_q^2 anterior es igual a:

$$S_{MC2E}^R - S_{MC2E} = \hat{u}_{MC2E}^R'W(W'W)^{-1}W\hat{u}_{MC2E}^R - \hat{u}_{MC2E}'W(W'W)^{-1}W\hat{u}_{MC2E}$$

por lo que, finalmente:

$$\frac{S_{MC2E}^R - S_{MC2E}}{\hat{\sigma}_u^2}$$

se distribuye χ_q^2 . Si los valores de S_{MC2E}^R y S_{MC2E} no se obtienen directamente del paquete estadístico utilizado en la estimación, pueden calcularse como la suma explicada de una regresión de los residuos MC2E correspondientes sobre la lista de variables en $W^{(6)}$.

9.7. EL ESTIMADOR DE MAXIMA VEROSIMILITUD

Consideremos el modelo:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

en el que suponemos que el término de error fundamental ε_t sigue una

⁽⁶⁾ Por otra parte, si el paquete calcula los «residuos incorrectos» $\hat{u} = y - \hat{X}\hat{\beta}_{IV}$, entonces los valores de S_{MC2E}^R y S_{MC2E} se obtendrían directamente de $SR = \hat{u}'\hat{u}$ utilizando los residuos de las estimaciones sin restringir y restringida, respectivamente.

distribución $N(0, \sigma_\varepsilon^2)$, independiente en el tiempo, por lo que podemos escribir su función de verosimilitud:

$$\begin{aligned} L(y_1, \dots, y_T) &= L(y_1, y_2) \cdot L(y_3, \dots, y_T | y_1, y_2) = \\ &= L(y_1, y_2) \cdot \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \right)^{\frac{T-2}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \varepsilon_t^2 \right\} \end{aligned}$$

donde:

$$\varepsilon_t = y_t - \beta_1(1 - \rho) - (\beta_2 + \rho)y_{t-1} + \beta_2\rho y_{t-2} - \beta_3x_t + \beta_3\rho x_{t-1}$$

Como aproximación a la función de verosimilitud, si ignoramos la función de densidad conjunta de las dos primeras observaciones⁽⁷⁾, entonces minimizar la función de verosimilitud es equivalente a minimizar la ruma residual que debe escribirse asimismo en términos de ε_t :

$$SR(\beta, \rho) = \sum_3^T [y_t - \beta_1(1 - \rho) - (\beta_2 + \rho)y_{t-1} + \beta_2\rho y_{t-2} - \beta_3x_t + \beta_3\rho x_{t-1}]^2$$

La minimización de SR con respecto a los coeficientes β y ρ proporciona el estimador de MCNL. Este estimador difiere del estimador MV en el tratamiento de las observaciones iniciales. Su obtención puede llevarse a cabo mediante alguno de los algoritmos que analizamos en el Capítulo 12, que proporcionarán no sólo las estimaciones de los coeficientes, así como de σ_ε^2 (de donde, a su vez, obtendremos la estimación de σ_u^2), sino también de sus varianzas y covarianzas.

Tiene especial interés considerar el algoritmo de Gauss-Newton, que consiste (como veremos en el Capítulo 12) en actualizar en cada iteración los coeficientes del modelo de regresión mediante los coeficientes estimados en la regresión auxiliar de los residuos del modelo sobre los componentes del vector gradiente de la función SR(β , ρ). En el caso aquí analizado, los residuos se obtienen como ya se ha visto, mientras que el vector gradiente está compuesto por:

$$\nabla \varepsilon_t = \left(-\frac{\partial \varepsilon_t}{\partial \beta_1}, -\frac{\partial \varepsilon_t}{\partial \beta_2}, -\frac{\partial \varepsilon_t}{\partial \beta_3}, -\frac{\partial \varepsilon_t}{\partial \rho} \right) = (1 - \hat{\rho}, \hat{y}_{t-1}^*, \hat{x}_t^*, \hat{u}_{t-1})$$

donde:

$$\begin{aligned} \hat{y}_{t-1}^* &= y_{t-1} - \hat{\rho}y_{t-2} \\ \hat{x}_t^* &= x_t - \hat{\rho}x_{t-1} \\ \hat{u}_{t-1} &= y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 y_{t-2} - \hat{\beta}_3 x_{t-1} \end{aligned}$$

⁽⁷⁾ Que podría también escribirse como la densidad de la primera observación, multiplicada por la densidad de la segunda observación, condicionada en la de la primera.

Así, los coeficientes estimados en cada una de estas tres variables auxiliares se *suman* a los valores de que disponíamos para los coeficientes β y ρ para obtener sus nuevas estimaciones. Una vez alcanzada la convergencia, el producto del vector gradiente por sí mismo, evaluado en los valores a los que se ha convergido, es una matriz 4×4 que, multiplicada por una estimación de $\hat{\sigma}_e^2$, proporciona la matriz de covarianzas de los coeficientes estimados. A su vez: $\hat{\sigma}_e^2 = SR/T$.

Este proceso de estimación no debe comenzarse a partir del estimador MCO, pues, siendo inconsistente, no garantizamos que mediante iteraciones sucesivas podamos converger al mínimo global que buscamos. Hay dos alternativas razonables para inicializar el algoritmo: a) puede utilizarse inicialmente un estimador de variables instrumentales, o b) puede llevarse a cabo una búsqueda en el espacio paramétrico correspondiente a ρ , escogiendo aquel valor que minimice SR, así como el vector β a él asociado. Si los instrumentos están bien escogidos, la primera alternativa es consistente y sugiere que si se produce la convergencia se alcanzará el mínimo global de SR. La segunda es una buena alternativa si la función SR es suficientemente suave, diferenciable y sin muchos puntos críticos.

Debe mencionarse que el algoritmo que acabamos de describir puede utilizarse en una versión algo diferente. Si definimos la variable $\hat{y}_t^* = y_t - \hat{\rho}y_{t-1}$, podemos estimar en cada iteración una regresión de esta variable sobre los componentes del gradiente que antes vimos. La constante, así como los coeficientes de \hat{y}_{t-1}^* y \hat{x}_t^* en esta regresión auxiliar son las nuevas estimaciones del vector β , sin necesidad de actualizar, mientras que el coeficiente de \hat{u}_{t-1}^* sería la corrección a introducir en la estimación de ρ .

Las sucesivas iteraciones que puedan llevarse a cabo en cualquiera de las dos versiones del procedimiento de Gauss-Newton no alteran sus propiedades asintóticas y tienen, en particular, la misma matriz de covarianzas teórica, por lo que no se gana eficiencia con ellas; sin embargo, las iteraciones pueden mejorar algo la estimación puntual en muestras finitas. El procedimiento descrito puede asimismo ser extendido al caso en que la correlación es de orden superior, sin más que utilizar filtros de orden superior, así como introducir residuos retardados adicionales en la regresión auxiliar.

El hecho de que, a diferencia de otros, este procedimiento trate simultáneamente la estimación de ρ con las del vector β , lo hace semejante al de MV. Hatanaka (1974) fue el primero en puntualizar que la inclusión de \hat{u}_{t-1} en la regresión auxiliar garantiza la eficiencia asintótica del estimador resultante.

9.8. ESTIMACION DE MODELOS CON EXPECTATIVAS RACIONALES

Una clase importante de modelos dinámicos la constituye aquellos en que la variable endógena depende, entre otros factores, de las expectativas que los agentes económicos tienen actualmente o tuvieron en el pasado acerca de los valores futuros de alguna variable exógena o de la propia variable endógena.

Es un principio generalmente aceptado que las decisiones económicas se toman en un contexto dinámico y bajo condiciones de incertidumbre. Por otra parte, las múltiples inercias de una economía real hacen que los efectos de las variables exógenas al decisor, como las posibles intervenciones de política económica, se dejan sentir durante un cierto tiempo, por lo que un esquema óptimo de toma de decisiones por parte de los agentes económicos debe incorporar tales efectos. Al hacerlo así, las decisiones que se toman en un período dependerán en parte de factores conocidos y en parte de otros factores sobre los que en el momento de decidir no disponemos sino de una previsión. Por ejemplo, el consumo de bienes duraderos, así como cualquier otra variable de inversión, puede depender de los tipos de interés actuales, pero también de las expectativas de tipos futuros. De modo análogo, la demanda de dinero puede depender de la tasa de inflación, pero también de la inflación esperada en el futuro.

En coherencia con tal visión, la especificación del modelo de determinación de una variable económica y_t puede ser del tipo:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 {}_t x_{t+1} + u_t \quad [9.21]$$

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 {}_{t-1} x_t + u_t \quad [9.22]$$

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 {}_{t-1} x_{t+1} + u_t \quad [9.23]$$

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 {}_t y_{t+1} + u_t \quad [9.24]$$

donde, por ejemplo, ${}_t x_{t+1}$ denota la expectativa que los agentes tienen en el período t acerca del valor de la variable x_t el próximo período. Pero la expectativa ${}_t x_{t+1}$ no es observable, por lo que tradicionalmente se ha adoptado en Econometría alguna de las siguientes estrategias:

1. Suponer que los agentes forman sus expectativas de modo adaptativo, como vimos en la Sección 9.2:

$${}_t x_{t+1} = {}_{t-1} x_t + (1 - \lambda) (x_t - {}_{t-1} x_t) \quad 0 < \lambda < 1 \quad [9.25]$$

2. Completar el modelo estructural con una ecuación de formación de expectativas:

$${}_t x_{t+1} = \delta_0 + \mathbf{z}'_t \boldsymbol{\delta}$$

utilizando esta ecuación para eliminar la expectativa de x_{t+1} .

3. Suponer que los agentes forman sus expectativas utilizando un mecanismo de retardos distribuidos del tipo:

$${}_t x_{t+1} = \sum_{s=0}^{\infty} \delta_s x_{t-s}$$

del que 1 no es sino un caso particular. Generalmente, se escoge:

$${}_t x_{t+1} = (1 - \delta) \sum_{s=0}^{\infty} \delta^s x_{t-s}$$

9.8.a. Expectativas racionales: primeras propiedades

Todas estas estrategias encierran cierto grado de arbitrariedad acerca del supuesto elegido. En esta sección vamos a considerar una alternativa, que consiste en suponer que los agentes forman las expectativas que aparecen en el modelo econométrico de modo racional. El supuesto de *racionalidad de expectativas* no impone ninguna forma funcional a la formación de las mismas, sino que postula tan sólo que los agentes explotan de modo óptimo la información de que disponen. Por ello entendemos que si Ω_t denota el conjunto de información de que disponen los agentes, éstos forman su *expectativa racional* acerca de x_{t+1} mediante la esperanza de dicha variable, *condicional* en el conjunto de información Ω_t , es decir, ${}_t x_{t+1} = E_t x_{t+1}$. El error racional de previsión queda definido por:

$$\varepsilon_{t+1} = x_{t+1} - {}_t x_{t+1} = x_{t+1} - E_t x_{t+1} \quad [9.26]$$

y tomando esperanzas condicionales en esta expresión se tiene:

$$E_t \varepsilon_{t+1} = E_t (x_{t+1} - E_t x_{t+1}) = E_t x_{t+1} - E_t x_{t+1} = 0$$

lo que viene garantizado por la utilización de la esperanza condicional en la formación de expectativas. En consecuencia, *los errores de previsión racionales en un período hacia el futuro son, necesariamente, impredecibles*, no pudiendo presentar autocorrelación.

La esperanza condicional es, como vimos en el Capítulo 2, el mecanismo de predicción óptimo de una variable, dado un conjunto de información⁽⁸⁾. Tiene, sin embargo, un inconveniente: para su cálculo es preciso especificar la distribución de probabilidad conjunta de *todas* las variables contenidas en Ω_t , junto con ε_{t+1} . Sólo entonces podría obtenerse la distribución de probabilidad de ε_{t+1} *condicional* en el resto del vector (las variables de Ω_t) y, por último, la esperanza matemática de dicha distribución condicionada. En la práctica, este ejercicio es poco menos que imposible.

⁽⁸⁾ Si se pretende resolver el problema:

$$\min_{x_{t+1}^*} E_t [x_{t+1} - x_{t+1}^*]^2$$

es decir, minimizar la varianza condicional del error de expectativas con respecto a un objetivo a escoger x_{t+1}^* , la solución óptima es:

$$x_{t+1}^* = E_t x_{t+1}$$

que sugiere utilizar como «proxy» de x_{t+1} en el período t la esperanza condicional $E_t x_{t+1}$.

El problema es mucho más simple cuando se supone Normalidad de la distribución de probabilidad conjunta; en tal caso, la esperanza condicional de x_{t+1} es una función *lineal* de las variables que condicionan (en este caso, las variables en Ω_t). Adoptamos este supuesto en lo sucesivo.

Supuesta la racionalidad de expectativas y utilizando [9.26] en el modelo [9.21], éste se convierte en

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t+1} + v_t \quad [9.27]$$

con

$$v_t = u_t - \beta_2 \varepsilon_{t+1} = u_t - \beta_2 (x_{t+1} - E_t x_{t+1}) \quad [9.28]$$

Como el error de predicción racional está libre de autocorrelación, entonces, si u_t era inicialmente ruido blanco, v_t estará libre de autocorrelación, por ser el agregado de dos ruidos blancos: u_t y el error de predicción ε_{t+1} . En consecuencia, se tiene en este modelo que $E_t v_t = u_t$. Nótese que, a pesar de su índice temporal, la variable v_t se realiza en el instante $t + 1$.

Sin embargo, tal ausencia de autocorrelación depende del horizonte temporal de las expectativas incluidas en el modelo. Por ejemplo, si imponemos racionalidad de expectativas en el modelo [9.23] se tiene:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t+1} + v_t$$

con

$$v_t = u_t - \beta_2 (x_{t+1} - E_{t-1} x_{t+1})$$

pero el error de predicción racional es ahora $\varepsilon_{t+1} = x_{t+1} - E_{t-1} x_{t+1}$, por lo que

$$E_t \varepsilon_{t+2} = E_t x_{t+2} - E_t (E_t x_{t+2}) = E_t x_{t+2} - E_t x_{t+2} = 0$$

mientras que

$$E_t \varepsilon_{t+1} = E_t (x_{t+1} - E_{t-1} x_{t+1}) = E_t x_{t+1} - E_{t-1} x_{t+1}$$

que será, en general, diferente de cero. Por consiguiente, en este modelo el error de predicción racional, que es de dos períodos hacia adelante, tiene una estructura MA(1), y lo mismo ocurrirá con el término de error compuesto $v_t = u_t + \varepsilon_{t+1}$.

Para comprender mejor este resultado, observemos que mediante la descomposición de Wald, si x_t es un proceso regular lineal, admite una representación de medias móviles, aunque quizá de orden infinito:

$$x_t = \sum_{s=0}^{\infty} \omega_s a_{t-s} = \omega_0 a_t + \omega_1 a_{t-1} + \omega_2 a_{t-2} + \dots$$

donde a_t es ruido blanco y $\sum_{s=0}^{\infty} |a_s| < \infty$, con $E_t a_{t+s} = 0$ para todo $s > 0$. En consecuencia:

$$x_{t+1} = \omega_0 a_{t+1} + \omega_1 a_t + \omega_2 a_{t-1} + \omega_3 a_{t-2} + \dots$$

$$E_t x_{t+1} = \omega_1 a_t + \omega_2 a_{t-1} + \omega_3 a_{t-2} + \dots$$

$$E_{t-1} x_{t+1} = \omega_2 a_{t-1} + \omega_3 a_{t-2} + \dots$$

por lo que los errores de previsión de x_{t+1} , uno y dos períodos hacia el futuro, son, respectivamente:

$$\varepsilon_t(1) = x_{t+1} - E_t x_{t+1} = \omega_0 a_{t+1}$$

$$\varepsilon_t(2) = x_{t+1} - E_{t-1} x_{t+1} = \omega_0 a_{t+1} + \omega_1 a_t$$

Por consiguiente, $\varepsilon_t(1)$ tiene, claramente, estructura de ruido blanco, al igual que a_t , pero, sin embargo, $\varepsilon_t(2)$ tiene una estructura MA(1), como se ve en la expresión anterior⁽⁹⁾.

9.8.b. Estimación por variables instrumentales

En cualquiera de las situaciones presentadas, el término de error de la versión estimable del modelo (nuestro v_t) está correlacionado con una de las variables explicativas, x_{t+1} . La correlación puede verse en la expresión [9.28] y puede calcularse, por ejemplo, en el caso de previsión un período hacia el futuro del modelo [9.27]:

$$\begin{aligned} \text{Cov}(v_t, x_{t+1}) &= \text{Cov}(u_t, x_{t+1}) - \beta_2 \text{Cov}(x_{t+1} - E_t x_{t+1}, x_{t+1}) = \\ &= 0 - \beta_2 \text{Cov}\left(\omega_0 a_{t+1}, \sum_{s=0}^{\infty} \omega_s a_{t-s+1}\right) = -\beta_2 \omega_0^2 \sigma_a^2 \end{aligned}$$

donde hemos utilizado el supuesto de ortogonalidad entre el término de error del modelo econométrico u_t y la innovación en la variable a predecir, a_t , es decir: $E(u_t a_s) = 0$ para todo t, s . En tales condiciones:

$$\text{Var}(v_t) = \sigma_u^2 + \beta_2^2 \sigma_a^2 \omega_0^2$$

⁽⁹⁾ Podrían aparecer en el modelo econométrico expectativas realizadas en diversas fechas sobre distinto horizonte, pero el argumento anterior puede extenderse sin mayor dificultad a tales situaciones.

Si normalizamos $\omega_0 = 1$, dado que es imposible identificarlo por separado de σ_a^2 , pero continuamos denotando por σ_a^2 la varianza del ruido blanco fundamental en x_t , tenemos:

$$\text{Corr}(v_t, x_{t+1}) = \frac{-\beta_2}{\sqrt{\lambda + \beta_2^2} \sqrt{\sum_0^{\infty} \omega_i^2}}$$

donde λ denota el coeficiente σ_u^2/σ_a^2 .

Esta correlación no nula entre perturbación y variable explicativa producirá la inconsistencia del estimador de mínimos cuadrados, por lo que habrá que estimar el modelo por un procedimiento de variables instrumentales, donde los instrumentos válidos son las variables ortogonales (es decir, incorrelacionadas) al término de error compuesto v_t .

Si la expectativa que aparece es con respecto a Ω_t , entonces, por definición de racionalidad, $E_t \varepsilon_{t+1} = 0$, lo que significa que cualquier variable z_t que forma parte de Ω_t está incorrelacionada con el error de predicción ε_{t+1} , uno de los dos componentes de v_t , puesto que

$$E(z_t \varepsilon_{t+1}) = E[E_t(z_t \varepsilon_{t+1})] = E[z_t E_t \varepsilon_{t+1}] = 0$$

Si podemos suponer que es también asintóticamente ortogonal al término de error original del modelo u_t , z_t será un instrumento válido. En particular, retardos de x_t , comenzando en el mismo instante temporal que el del conjunto de información utilizado en la previsión, son instrumentos válidos. Así, x_{t-s} es instrumento válido en la estimación de [9.21] para $s \geq 0$, mientras que es instrumento válido en la estimación de [9.22] y [9.23] tan sólo para $s \geq 1$ ⁽¹⁰⁾.

Por otra parte, ya hemos visto que ésta es una situación en que, teniendo variables explicativas aleatorias, el término de error del modelo a estimar puede presentar autocorrelación; en tal caso, algunos retardos de la variable endógena y_t son instrumentos válidos si, como ocurre aquí, la estructura de autocorrelación de v_t es de media móvil. Cuando v_t es MA(1), los retardos y_{t-s} son instrumentos válidos para $s > 1$, y lo son para $s > 2$ si v_t es MA(2).

Puede apreciarse, por tanto, que dispondremos generalmente de más instrumentos de los necesarios. El problema de reducir la lista de los q instrumentos que configuran un vector inicial w_t a los k instrumentos necesarios para la estimación, que incluiremos en el vector z_t , puede hacerse mediante una matriz A , de dimensión $k \times q$ y rango k .

Con el vector $z_t = Aw_t$, donde A es $k \times q$, siendo q el número de instrumentos y $q \gg k$, puede llevarse a cabo una estimación de variables instru-

⁽¹⁰⁾ Sin embargo, la aparición de x_t como variable explicativa de [9.21] desaconseja su utilización como único instrumento en la estimación de dicha ecuación, para evitar en lo posible que la matriz $Z'X$ sea singular.

mentales; la matriz de covarianzas del estimador resultante dependerá de la matriz A . Tiene sentido, por tanto, plantear el problema:

$$\min_A \|\text{Var}(\hat{\beta}_{V1}(A))\|$$

de minimización de la norma (es decir, el tamaño) de dicha matriz de covarianzas. Este es el problema que resuelve el *estimador de mínimos cuadrados en dos etapas* de la Sección 9.5: se estiman por mínimos cuadrados regresiones de las k variables explicativas sobre las q variables en w_t ; se tienen así k vectores, cada uno de dimensión q que, ordenados como filas, generan la matriz A óptima, es decir, el vector z_t óptimo, dado un vector inicial w_t de instrumentos⁽¹¹⁾.

9.9. EL ESTIMADOR GENERALIZADO DE MOMENTOS

Los instrumentos están, por definición, incorrelacionados en la población con el término de error del modelo para el que son válidos como tales instrumentos. Dado un vector de variables instrumentales, tiene por tanto interés preguntarse acerca del vector numérico de coeficientes β que hacen que las *correlaciones muestrales* entre instrumentos y residuos sean mínimas. Este es el fundamento del estimador generalizado de momentos.

Para analizar su utilización en la estimación del modelo [9.23], tomado como ejemplo, escribamos éste como:

$$E_{t-1}[y_t - \beta_0 - \beta_1 x_t - \beta_2 x_{t+1} - u_t] = 0 \quad [9.29]$$

por lo que, si denotamos por h_{t+1} la expresión dentro del corchete en [9.29], tenemos:

$$E_{t-1} h_{t+1} = 0$$

Si η_{t+1} denota el error de predicción de h_{t+1} en el instante $t-1$, se tiene que $h_{t+1} = E_{t-1} h_{t+1} + \eta_{t+1}$, donde η_{t+1} sigue una estructura MA(1). Para toda variable z_{t-1} en el conjunto de información Ω_{t-1} se tiene:

$$E[z_{t-1} h_{t+1}] = E[E_{t-1}(z_{t-1} h_{t+1})] = E[z_{t-1} E_{t-1} h_{t+1}] = 0$$

⁽¹¹⁾ Cuando el horizonte de previsión excede de un período, puede surgir un problema más delicado, como ocurre en el modelo [9.22], por cuanto el término de error del modelo resultante es:

$$v_t = u_t - \beta_2(x_{t+1} - E_{t-1} x_{t+1}) = u_t - \beta_2(\omega_0 a_{t+1} + \omega_1 a_t)$$

que está relacionado no sólo con la variable explicativa x_{t+1} , sino también con la otra variable explicativa, x_t . Por tanto, en este modelo se precisan variables instrumentales para ambas.

es decir,

$$0 = E[z_{t-1} E_{t-1} h_{t+1} + z_{t-1} \eta_{t+1}] = E(z_{t-1} \eta_{t+1})$$

donde hemos utilizado el hecho de que $E_{t-1} h_{t+1} = 0$. Ello muestra que toda variable en Ω_{t-1} está incorrelacionada con η_{t+1} ⁽¹²⁾. Dicho de otra manera, el supuesto de racionalidad de expectativas impone la *ausencia de correlación entre las variables del conjunto de información sobre el que se basan las expectativas y el error de previsión del modelo que se pretende estimar*.

Si basándonos en esta propiedad nos proponemos encontrar el valor numérico del vector β que minimice el *valor muestral* de tales correlaciones, trataremos de resolver el problema:

$$\min_{\beta} \|T^{-1} Z \eta(\beta)\| = \min_{\beta} \left\| T^{-1} \sum_1^T z_{t-1} \eta_{t+1}(\beta) \right\|$$

para un vector z_{t-1} dado de variables contenidas en el conjunto Ω_{t-1} , donde las barras verticales denotan algún concepto de norma del vector producto $Z \eta$, que es de dimensión $q \times 1$, puesto que Z es $q \times T$ y η es $T \times 1$. En la norma anterior sustuiremos η_{t+1} por los residuos generados por cada vector β , y se trata de encontrar el valor numérico de éste de modo que se minimice la norma escogida. En particular, dada una matriz A_T simétrica, definida positiva, que pudiera depender de información muestral, tenemos una norma:

$$\begin{aligned} J_T &= \|T^{-1} Z \eta\| = (T^{-1} Z \eta)' A_T (T^{-1} Z \eta) = \\ &= \left(T^{-1} \sum_1^T z_{t-1} \eta_{t+1} \right)' A_T \left(T^{-1} \sum_1^T z_{t-1} \eta_{t+1} \right) \end{aligned} \quad [9.30]$$

de cuya minimización obtendremos el estimador generalizado de momentos (GM) del vector β (también denominado estimador generalizado de variables instrumentales).

La matriz de covarianzas del estimador resultante depende de la elección que se haga de la matriz A_T . Hansen y Singleton (1982) demostraron que la elección óptima de matriz de ponderaciones A_T , en el sentido de minimizar la matriz de covarianzas del estimador resultante, se consigue utilizando una aproximación muestral a la inversa de la esperanza:

$$S_0 = E[(z_{t-1} \eta_{t+1})(z_{t-1} \eta_{t+1})'] \quad [9.31]$$

lo que se logra utilizando una matriz A_T :

$$A_T = \left[T^{-1} \sum_{j=-L}^L \sum_{t=j+1}^{T-1} (z_{t-1} \eta_{t+1})(z_{t-1-j} \eta_{t+1-j})' \right]^{-1} = \left(\sum_{-L}^L S_T(j) \right)^{-1}$$

(12) Esto corrobora asimismo que h_{t+1} y η_{t+1} son iguales entre sí.

de dimensión $q \times q$, donde $S_T(j)$ es cada una de las matrices de covarianzas retardadas:

$$S_T(j) = \frac{1}{T} \sum_{t=j+1}^T [(z_{t-1}\eta_{t+1})(z_{t-1-j}\eta_{t+1-j})']$$

y donde L es el orden de la autocorrelación del vector $z_{t-1}\eta_{t+1}$, que suele coincidir con la de η_{t+1} y que, en nuestro ejemplo, es igual a 1 por tener el error de predicción una estructura MA(1).

La matriz de covarianzas del *estimador generalizado de variables instrumentales* resultante de la minimización de la forma cuadrática J_T en [9.30] es:

$$\text{Var}(\hat{\beta}_{GM}) = (\mathbf{D}'_0 \mathbf{S}_0^{-1} \mathbf{D}_0)^{-1}$$

donde

$$\mathbf{D}_0 = E \left[z_{t-1} \left(\frac{\partial \eta_{t+1}}{\partial \beta} \right)' \right] = E \left[z_{t-1} \left(\frac{\partial \eta(\mathbf{x}_{t+1}, \beta)}{\partial \beta} \right)' \right]$$

y \mathbf{S}_0 es la matriz en [9.31]. La matriz \mathbf{D}_0 se aproxima en muestras finitas mediante:

$$\mathbf{D}_T = \frac{1}{T} \sum_1^T z_{t-1} \left(\frac{\partial \eta_{t+1}}{\partial \beta} \right)' = \frac{1}{T} \sum_1^T z_{t-1} \left(\frac{\partial \eta(\mathbf{x}_{t+1}, \hat{\beta})}{\partial \beta} \right)'$$

Para obtener el valor numérico del estimador $\hat{\beta}_{GM}$, cabe observar que las condiciones necesarias de optimalidad del problema de minimización de la norma de la matriz de covarianzas son:

$$\left[\frac{1}{T} \sum_1^T z_{t-1} \left(\frac{\partial \eta_{t+1}}{\partial \beta} \right)' \right]' \mathbf{A}_T \left(\frac{1}{T} \sum_1^T z_{t-1} \eta_{t+1} \right) = \mathbf{0}_k \quad [9.32]$$

donde los órdenes de los factores son $k \times q$, $q \times q$ y $q \times 1$, respectivamente. El sistema [9.32] será lineal en los β sólo si el gradiente $\partial \eta_{t+1} / \partial \beta$ lo es. Así, un modelo lineal como el de nuestro ejemplo se tiene

$$\partial \eta_{t+1} / \partial \beta = -(1, x_t, x_{t+1}) = \mathbf{x}_t$$

y se tiene el sistema de ecuaciones:

$$\left(\frac{1}{T} \sum_1^T z_{t-1} \mathbf{x}_t' \right)' \mathbf{A}_T \left(\frac{1}{T} \sum_1^T z_{t-1} \eta_{t+1} \right) = \mathbf{0}_k$$

que, suponiendo ausencia de autocorrelación y heteroscedasticidad, se convierte en:

$$\left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \mathbf{x}'_t\right) \left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \mathbf{z}'_{t-1}\right)^{-1} \left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \eta_{t+1}\right) = \mathbf{0}_k$$

cuya solución no es sino el estimador de mínimos cuadrados en dos etapas, posiblemente no lineal.

Salvo en el caso lineal, este procedimiento de estimación requiere un esquema iterativo numérico, que comienza a partir de una elección ineficiente de \mathbf{A}_T , por ejemplo: $\mathbf{A}_T = \mathbf{I}_q$, para obtener el estimador que minimiza:

$$J_T^0 = \left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \eta_{t+1}\right)' \left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \mathbf{z}'_{t-1}\right)^{-1} \left(\frac{1}{T} \sum_1^T \mathbf{z}_{t-1} \eta_{t+1}\right)$$

que es la norma euclídea de un vector $q \times 1$. A partir de este estimador se generan los residuos y se evalúan las matrices \mathbf{D}_T y \mathbf{A}_T del modo antes indicado para minimizar, en una segunda etapa del procedimiento, la forma cuadrática J_T .

PROBLEMAS

Problema 9.1. a) Escriba, con todo detalle, la expresión matricial para la estimación por variables instrumentales, del modelo [9.19], cuando se utiliza una variable \hat{y}_{t-1} como instrumento de y_{t-1} . Describa las condiciones de ortogonalidad (o ausencia de correlación) entre instrumentos y residuos en esta estimación.

b) Considérese, como instrumento de y_{t-1} , la «proyección» \hat{y}_{t-1} obtenida a partir de la regresión:

$$y_t = \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + v_t$$

mediante la expresión $\hat{y}_t = \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{\alpha}_3 x_{3t}$, donde $\hat{\alpha}_i$, $i = 1, 2, 3$, son las estimaciones de los coeficientes de esta regresión auxiliar. Pruebe que $\sum_1^T x_{it} \hat{v}_t = 0$, $i = 1, 2, 3$, así como $\sum_1^T \hat{y}_t \hat{v}_t = 0$.

c) Utilizar la expresión analítica obtenida en a) para probar que el estimador en b) coincide con el estimador de mínimos cuadrados del modelo

$$y_t = \beta_1 + \beta_2 \hat{y}_{t-1} + \beta_3 x_{1t} + \beta_4 x_{2t} + \beta_5 x_{3t} + u_t$$

en el que la variable \hat{y}_{t-1} ha sustituido totalmente a la variable y_{t-1} . Ambos estimadores satisfacen las mismas condiciones de ortogonalidad.

Problema 9.2. Discutir la afirmación: «El estimador MCO del parámetro β_2 en el modelo [9.16] es inconsistente, debido a la correlación existente entre la variable y_{t-1} y el término de error u_t . Sin embargo, la variable x_t es determinista, por lo que el estimador MCO de los parámetros β_1 y β_3 es consistente.

Problema 9.3. Probar que la matriz en pág. 306 es la expresión correcta para la matriz de covarianzas del estimador de máxima verosimilitud del modelo (9.12).

Problema 9.4. Probar que la matriz de covarianzas del estimador de máxima verosimilitud del modelo:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$|\beta_3| < 1 \quad ; \quad |\rho| < 1$$

supuesta una-distribución $N(0, \sigma_\varepsilon^2)$ para el proceso de ruido blanco fundamental ε_t , es:

$$\text{Var} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\rho} \\ \hat{\sigma}_u^2 \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} T(1-\rho^2) & (1-\rho)\Sigma x_t^* & (1-\rho)E(\Sigma y_{t-1}^*) & 0 & 0 \\ & \Sigma x_t^{*2} & E(\Sigma x_t^* y_{t-1}^*) & 0 & 0 \\ & & E(\Sigma y_{t-1}^{*2}) & \frac{T\sigma_\varepsilon^2}{1-\beta_2\rho} & 0 \\ & & & \frac{T\sigma_\varepsilon^2}{1-\rho^2} & 0 \\ & & & & \frac{T}{2\sigma_\varepsilon^2} \end{pmatrix}^{-1}$$

Problema 9.5. Si se quisiera utilizar la matriz de información anterior para estimar la matriz de covarianzas del estimador de máxima verosimilitud, habría que obtener las expresiones analíticas para las esperanzas matemáticas que en ellas aparecen. Probar que en el modelo

$$y_t^* = \beta_1(1-\rho) + \beta_2 y_{t-1}^* + \beta_3 x_t^* + \varepsilon_t \quad , \quad |\beta_2| < 1$$

se tiene:

- a) $E(y_t^*) = \frac{\beta_1(1-\rho)}{1-\beta_2} + \beta_3(\sum_{s=0}^{\infty} \beta_2^s x_{t-s})$
- b) $\text{Var}(y_t^*) = \frac{\sigma_\varepsilon^2}{1-\beta_2^2}$
- c) $E(y_t^{*2}) = \left(\frac{\beta_1(1-\rho)}{1-\beta_2}\right)^2 + 2\beta_3^2[\sum_{s=0}^{\infty} \beta_2^{2s+1} x_{t-s} x_{t-s-1}] + 2\frac{\beta_1\beta_3(1-\rho)}{1-\beta_2} \sum_{s=0}^{\infty} \beta_2^s x_{t-s} + \frac{\sigma_\varepsilon^2}{1-\beta_2^2}$

donde y_t^* , x_t^* están definidos en la Sección 9.7, y utilizar estas expresiones para discutir cómo se estimarían en la práctica las esperanzas matemáticas de los términos que aparecen en la matriz de covarianzas en el problema anterior.

Problema 9.6. Supongamos que en la ecuación de demanda de capital a largo plazo $K_t^* = \beta_1 + \beta_2 Y_{t+1}^* + u_t$, el stock de capital deseado, K_t^* , depende de las expectativas de renta futura Y_{t+1}^* . Supongamos asimismo que el stock de capital realizado K_t

aproxima a su valor deseado mediante una relación de ajuste parcial con parámetro δ , y que las expectativas acerca de la renta se forman mediante un mecanismo adaptativo con parámetro λ .

Probar que el modelo econométrico que resulta es de la forma:

$$K_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 K_{t-1} + \alpha_3 K_{t-2} + v_t$$

Encontrar los coeficientes α_i como función de β_1 , β_2 , δ y λ , así como las propiedades del término de error v_t . ¿Pueden obtenerse estimaciones de los parámetros β_1 , β_2 , λ y δ ?

¿Se resolvería el problema si se incluyese otra variable explicativa en la ecuación de K_t^* ?

Problema 9.7. Probar que el estimador de máxima verosimilitud del modelo [9.19] se reduce al procedimiento de Cochrane-Orcutt sugerido en el Capítulo 7.

Problema 9.8. Probar que en el modelo

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t, & |\beta_2| < 1 \\ u_t &= \rho u_{t-1} + \varepsilon_t, & |\rho| < 1 \end{aligned}$$

con $E(u_t) = 0$, $\text{Var}(u_t) = \sigma_u^2$, se tiene:

$$\begin{aligned} E(y_{t-1} u_t) &= \frac{\rho \sigma_u^2}{1 - \beta_2 \rho} \\ \text{Corr}(y_{t-1}, u_t) &= \rho \frac{\sqrt{1 - \beta_2^2}}{\sqrt{1 - \beta_2 \rho}} \end{aligned}$$

Interprete esta expresión cuando ρ tiende a cero, y también cuando β_2 tiende a cero. ¿Tiene sentido hacer una interpretación similar cuando β_2 o ρ tienden a 1?

Problema 9.9. a) Escribir en detalle la expresión matricial del estimador MCO del modelo de retardos distribuidos:

$$y_t = \omega \frac{1 - \theta L}{1 - \rho L} x_t + u_t, \quad t = 1, 2, \dots, T$$

donde $E(u_t u_s) = 0$ para todo $t \neq s$, e igual a σ_u^2 si $t = s$. ¿Qué características tendría este estimador? ¿Aconsejaría su utilización en este modelo?

b) Hacer lo mismo para un estimador de variables instrumentales suponiendo que sólo se dispone de series temporales de datos para las variables x_t e y_t . ¿Qué propiedades tiene este estimador? ¿Es insesgado? ¿Es consistente?

c) ¿Cuál sería la matriz de covarianzas del estimador de variables instrumentales que ha sugerido en b)? ¿Es dicho estimador el estimador de variables instrumentales de varianza mínima?

d) ¿Cómo se estimaría el parámetro σ_u^2 una vez obtenido el estimador de variables instrumentales que sugirió en b)?

e) ¿Cómo obtendría la varianza del parámetro θ y sus covarianzas con ω y ρ a partir de los elementos de la matriz de covarianzas que dio en c)?

Problema 9.10. Demostrar que, si para estimar el modelo $y_i = \beta x_i + u_i$ con las variables en desviaciones respecto a la media se utiliza un instrumento z_i , entonces la varianza relativa del estimador de variables instrumentales frente al estimador MCO viene dada por:

$$\frac{\text{Var}(\hat{\beta}_{VI})}{\text{Var}(\hat{\beta}_{MCO})} = (r_{xz}^2)^{-1}$$

siendo r_{xz} el coeficiente de correlación entre ambas variables, e interpretar este resultado.

Problema 9.11. Obtener el estimador de máxima verosimilitud condicionado del modelo:

$$\begin{cases} y_t = \phi y_{t-1} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{cases}$$

Probar que la matriz de covarianzas de dicho estimador cuando $\rho = 0$ es:

$$\Sigma = \text{Var} \begin{pmatrix} \hat{\phi} \\ \hat{\sigma}_\varepsilon^2 \end{pmatrix} = \begin{pmatrix} \frac{1 - \phi^2}{T} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{pmatrix}$$

Calcular la matriz de covarianzas en el caso $\rho \neq 0$.

Problema 9.12. Comparar los estimadores del modelo:

$$y_t = \beta_1 x_t + \beta_2 y_{t-1} + u_t$$

que se obtienen:

i) Resolviendo el sistema:

$$\begin{aligned} \Sigma_1^T y_t x_t &= \hat{\beta}_1 \Sigma_1^T x_t^2 + \hat{\beta}_2 \Sigma_1^T y_{t-1} x_t \\ \Sigma_1^T y_t x_{t-1} &= \hat{\beta}_1 \Sigma_1^T x_t x_{t-1} + \hat{\beta}_2 \Sigma_1^T y_{t-1} x_{t-1} \end{aligned}$$

ii) Mediante una regresión inicial de y_{t-1} sobre x_{t-1} que permita obtener \hat{y}_{t-1} , para luego estimar una regresión de y_t sobre x_t e \hat{y}_{t-1} .

iii) Mediante una regresión inicial de y_{t-1} sobre x_t y x_{t-1} para obtener \hat{y}_{t-1} , para luego estimar una regresión de y_t sobre x_t e \hat{y}_{t-1} .

iv) Mediante una regresión inicial de y_{t-1} sobre x_t , x_{t-1} , y_{t-2} para obtener \hat{y}_{t-1} , para luego estimar una regresión de y_t sobre x_t e \hat{y}_{t-1} .

Problema 9.13. Polinomio de Almon.

Supongamos que para estimar el modelo

$$y_t = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \beta_4 x_{t-4} + u_t$$

un investigador impone en los coeficientes β la estructura:

$$\beta_i = \gamma_0 + \gamma_1 i + \gamma_2 i^2, \quad i = 0, 1, 2, 3, 4$$

siguiendo la propuesta de Almon (1965).

a) Hallar el modelo transformado, que tiene $\gamma_0, \gamma_1, \gamma_2$ por coeficientes a estimar, y cuyas variables son transformaciones lineales de los retardos de x_t . ¿En qué medida puede reducirse la dimensionalidad del problema de estimación de un modelo de retardos distribuidos de orden s mediante un supuesto de este tipo?

b) ¿Cómo podría contrastarse el supuesto de que los coeficientes del modelo original siguen una estructura de Almon?

Problema 9.14. Considere la estimación de variables instrumentales del modelo:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + u_t, \quad |\beta_3| < 1$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

donde ε_t es ruido blanco.

Demuestre que el procedimiento de estimar una regresión inicial:

$$y_t = \delta_1 + \delta_2 x_t + \delta_3 x_{t-1} + v_t$$

para obtener la «predicción» \hat{y}_{t-1} y estimar a continuación:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 \hat{y}_{t-1} + v_t$$

equivale a la utilización de la expresión matricial:

$$\beta = (Z'X)^{-1} Z'y$$

donde Z es una matriz $T \times 3$ que contiene observaciones de las variables $(1, x_t, x_{t-1})$, mientras que X es la matriz $T \times k$ formada por las observaciones de $(1, x_t, y_{t-1})$.

Problema 9.15. Para estimar el modelo

$$y_t = \beta_1 y_{t-1} + \beta_2 x_t + u_t, \quad |\beta_1| < 1$$

$$u_t = \varepsilon_t - \theta \varepsilon_{t-1}, \quad |\theta| < 1$$

donde las dos variables x_t e y_t se hallan en diferencias con respecto a la media, se dispone de series temporales de datos. Se supone, además, que $E(x_{t-s} \cdot u_t) = 0$ para todo s . Con ellos se han obtenido las funciones de autocorrelación simple ρ_x y ρ_y :

k	0	1	2	3	4
$\rho_x(k)$	1	0,40	0,0	0,0	0,0
$\rho_y(k)$	1	0,80	0,60	0,40	0,20

con varianzas muestrales: $\sigma_x^2 = 2.500$, $\sigma_y^2 = 10.000$, y covarianza: $\sigma_{xy} = 4.000$. También se ha estimado la función de correlación cruzada muestral:

k	-2	-1	0	1	2
$\sigma_{xy}(k) = T^{-1} \sum_1^T x_t y_{t-k}$	0,0	0,40	0,40	0,80	0,60

- a) Obtener el estimador del modelo anterior, que utiliza x_{t-1} como instrumento, y calcular su matriz de covarianzas.
- b) Repetir el ejercicio, utilizando esta vez y_{t-2} como instrumento.
- c) ¿Son consistentes ambos estimadores? ¿Son eficientes?

Problema 9.16. Considere introducir la parametrización $\beta_0 = 1/\lambda$, $\beta_i = (1 - \lambda^i)/\lambda$, $i > 0$ en el modelo

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + u_t$$

y describa de qué modo llevaría a cabo la estimación del parámetro λ . ¿En qué medida es satisfactoria la utilización del cambio de parámetros propuesta para llevar a cabo la estimación del modelo de retardos infinitos original?

Problema 9.17. Discuta la estimación, por máxima verosimilitud condicionada, de los parámetros α , β , ρ , σ_ε^2 y σ_u^2 en el modelo:

$$y_t = \alpha y_{t-1} + \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

donde ε_t es un proceso de ruido blanco, $y: E(x_t \cdot \varepsilon_s) = 0$ para todo t, s .

Problema 9.18. Derive las ecuaciones para el cálculo del estimador de máxima verosimilitud del modelo

$$y_t = \alpha y_{t-1} + \beta x_t + u_t, \quad |\alpha| < 1$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

inicialmente, bajo el supuesto de que $\alpha = 0$ y ε_t es ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$, y posteriormente, con $\alpha \neq 0$ y ε_t ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$.

Problema 9.19. Considere el modelo:

$$y_t = \beta y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

donde tanto β como ρ son inferiores a la unidad en valor absoluto, y ε_t es un ruido blanco, y obtenga el sesgo asintótico del estimador MCO de ambos parámetros. Derive asimismo el sesgo asintótico de la suma de ambos.

Problema 9.20. Obtenga el estimador de máxima verosimilitud del modelo:

$$y_t = \alpha + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

donde ε_t es un ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$.

Problema 9.21. Encuentre el sesgo asintótico del estimador MCO de los coeficientes α y β del modelo:

$$y_t = \alpha + \beta y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

siendo ε_t un ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$.

Problema 9.22. a) Probar que el estimador MCO del coeficiente δ en el modelo

$$y_t = \delta y_{t-1} + u_t, \quad |\delta| < 1$$

$$u_t = \varepsilon_t + \alpha \varepsilon_{t-1}, \quad |\alpha| < 1$$

donde ε_t es un ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$, satisface:

$$plim \hat{\delta}_{MCO} = \delta + \frac{\phi(1 - \delta^2)}{1 + 2\delta\phi}, \quad \text{donde} \quad \phi = \frac{\alpha}{1 + \alpha^2}$$

b) Probar que

$$plim \left(\frac{1}{T} \sum_1^T \hat{u}_t^2 \right) = \sigma_\varepsilon^2 [1 + \alpha(\alpha - \alpha^*)], \quad \text{donde} \quad \alpha^* = \frac{\phi(1 - \delta^2)}{1 + 2\delta\phi}$$

Problema 9.23. A partir de una muestra de 52 observaciones, con variables en desviaciones respecto a la media, se han obtenido los siguientes productos cruzados:

	y_t	x_{1t}	x_{2t}	y_{t-1}	donde, por ejemplo,
y_t^2	100	-60	60	80	$\sum_{t=1}^{52} x_{1t} y_{t-1} = -40$
x_{1t}		80	50	-40	$\sum_{t=1}^{52} y_t^2 = 100$
x_{2t}			40	-10	
y_{t-1}				100	

a) Estimar el modelo $y_t = \beta_1 y_{t-1} + \beta_2 x_{1t} + u_t$, donde $u_t = \rho u_{t-1} + \xi_t$, y ξ_t es ruido blanco, por mínimos cuadrados ordinarios, obteniendo su matriz de covarianzas. ¿Qué problemas presenta este estimador?

b) Estimar el modelo utilizando x_{2t} como instrumento de y_{t-1} y obtener su matriz de covarianzas.

c) ¿Es el estimador calculado en b) eficiente? Si no lo es, construir un estimador de variables instrumentales más eficiente, obteniendo su matriz de covarianzas.

d) Obtener el estimador de mínimos cuadrados ordinarios del modelo:

$$y_t = \beta_1 \hat{y}_{t-1} + \beta_2 x_{1t} + u_t$$

donde \hat{y}_{t-1} es el instrumento utilizado en c). ¿Qué relación existe entre los estimadores obtenidos en c) y d)? ¿Puede demostrarse una proposición general en tal sentido?

Problema 9.24. Para analizar la estructura del sector del automóvil en España se especifica una función de producción con elasticidad de sustitución constante:

$$Q_t = \gamma[\delta K_t^{-\rho} + (1 - \delta)L_t^{-\rho}]^{-1/\rho}$$

donde Q_t , L_t , K_t son la producción, el empleo y el stock de capital productivo del sector en el período t , respectivamente.

a) Sugerir algún método para estimar los parámetros γ , ρ , δ , v .

b) Utilizar un argumento de equilibrio (en particular, las empresas maximizan el valor presente de sus beneficios) para obtener el stock de capital deseado K_t^* como función del precio relativo P_t/C_t y el nivel de producción Q_t que se desea alcanzar (P_t es el precio del producto, C_t el coste de uso de capital).

c) Justificar un modelo de ajuste dinámico del stock de capital, del tipo:

$$(1) \quad \Delta \log K_t = \mu(L) \cdot \Delta \log K_t^*$$

donde $\mu(L)$ es un polinomio de retardos, K_t el stock de capital realmente existente en el período t y K_t^* el stock deseado en ese período.

d) Explicar cómo se podría utilizar el modelo (1), junto con la determinación de K_t^* que se ha hecho en b), para estimar las elasticidades del precio relativo y del nivel de producto. ¿Qué distintas hipótesis acerca de los valores tomados por estas elasticidades sería relevante contrastar?

Problema 9.25. Supongamos que el valor deseado, cada período, K_t^* del stock de capital K_t responde al modelo de decisión:

$$K_t^* = \alpha + \beta \cdot Y_t^e$$

donde Y_t denota el valor del producto en la economía e:

$$Y_t^e = E(Y_{t+1}/\Omega_t)$$

Debido a la presencia de costes de ajuste de capital, éste se aproxima, pero no coincide, con su valor deseado, lo que puede representarse:

$$K_t = K_{t-1} + (1 - \lambda_1)(K_t^* - K_{t-1}) + u_t, \quad |\lambda_1| < 1$$

siendo $u_t \sim \text{NID}(0, \sigma_u^2)$.

Los agentes actualizan sus expectativas de Y_t de acuerdo con la expresión:

$$Y_t^e - Y_{t-1}^e = (1 - \lambda_2)(Y_t - Y_{t-1}^e)$$

Se pide:

a) Describir en todo detalle las restricciones que esta formulación impone sobre un modelo general de relación dinámica entre K_t e Y_t .

b) ¿Qué ocurriría en el caso particular $\lambda_2 = 1$?

c) Si el parámetro β resultara no significativo, ¿qué estructura ARIMA tendría K_t ?

d) ¿Qué estructura ARIMA sigue Y_t ?

Problema 9.26. Suponga que un investigador está interesado en analizar el modelo:

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

donde ambas variables se hallan en diferencias respecto a su media muestral, pero, por error, se omite x_{t-2} , especificando el modelo:

$$y_t = \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

- ¿Cuál es el sesgo de la estimación MCO de este último?
- ¿Cuál es el sesgo en la estimación del efecto a largo plazo de x_t sobre y_t ?
- ¿Cuál sería el sesgo asintótico si la variable x_t siguiera un esquema AR(1):

$$x_t = \rho x_{t-1} + \varepsilon_t$$

Problema 9.27. Un investigador pretende estimar el modelo:

$$(1) \quad y_t = \beta_1 x_t + \beta_2 \cdot E_{t-1} x_{t+1} + u_t$$

donde u_t es un ruido blanco, x_t una variable aleatoria, y tanto x_t como y_t son estacionarias. $E_{t-1} x_{t+1}$ denota la expectativa que los agentes formaron en el instante $t-1$ acerca del valor de x_{t+1} . Para ello, supondremos que las expectativas de los agentes son racionales.

a) Explique con todo rigor los inconvenientes que para la estimación de mínimos cuadrados presenta el hecho de que la expectativa $E_{t-1} x_{t+1}$ no sea observable.

b) Para evitar el problema, nuestro investigador considera sustituir la expectativa por el valor realizado *ex post* y estimar el modelo

$$(2) \quad y_t = \beta_1 x_t + \beta_2 x_{t+1} + v_t$$

por el procedimiento de variables instrumentales. Sin embargo, un colega suyo le comenta dos cuestiones: a) que el término de error del modelo (2) tiene autocorrelación, por lo que variables que aparentemente serían instrumentos válidos, como es el caso de x_{t-1} , no lo son, y b) que debido a la existencia de autocorrelación en v_t , ningún retardo de la variable endógena será instrumento válido. Discuta ambas afirmaciones.

c) ¿Es la presencia de la variable x_t como explicativa una fuente adicional de inconsistencia del estimador de mínimos cuadrados en este modelo?

Supongamos ahora que la especificación correcta no fuese (1) sino

$$(3) \quad y_t = \beta_1 x_{t-1} + \beta_2 \cdot E_{t-1} x_t + u_t$$

y que el investigador decide seguir una estrategia similar a la anterior y estimar por variables instrumentales el modelo resultante, utilizando como instrumentos los retardos x_{t-1} y x_{t-2} . ¿Obtendrá estimaciones consistentes?

Con independencia de su respuesta, obtenga numéricamente dicho estimador, sabiendo que en una muestra de T observaciones se estimó la función de autocorrelación de x_t :

$$\begin{array}{cccc} k = & 0 & 1 & 2 & 3 \\ \rho_x(k) = & 1 & 0,80 & 0,60 & 0,50 \end{array}$$

y $\rho_x(k) = 0$ para todo $k > 3$. La varianza muestral de x_t fue de 2.500. Además, la función de correlación cruzada estimada de x_t e y_t , en la dirección de la primera variable hacia la segunda, fue:

$$\rho_{xy}(k) = \begin{matrix} k = & 0 & 1 & 2 \\ & 0,40 & 0,80 & 0,60 \end{matrix}$$

donde $\rho_{xy}(k) = \text{Corr}(x_{t-k}, y_t)$. La varianza muestral de y_t fue de 6.400.

Problema 9.28. Supongamos que se quiere estimar el modelo:

$$(1) \quad y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

donde $\text{Corr}(x_{1t}, u_t) \neq 0$, mientras que $\text{Corr}(x_{2t}, u_t) = 0$. El investigador dispone de otras dos variables, x_{3t} y x_{4t} , tales que $\text{Corr}(x_{3t}, u_t) = \text{Corr}(x_{4t}, u_t) = 0$.

a) Proponga varios (más de uno) estimadores consistentes del modelo (1).

b) El investigador afirma: «Puesto que x_{3t} no está correlacionada con u_t y he calculado que la correlación muestral entre x_{3t} y x_{1t} es 0,70, puedo estimar por MCO el modelo:

$$(2) \quad y_t = \alpha + \beta_1 x_{3t} + \beta_2 x_{2t} + v_t$$

obteniendo estimaciones consistentes, pues, en definitiva, éste es un estimador de variables instrumentales».

c) ¿Es la afirmación hecha en b) cierta para todos, alguno o ninguno de los posibles estimadores de variables instrumentales? Demuestre su respuesta.

d) A partir de una muestra de 100 observaciones se obtuvieron los momentos muestrales de segundo orden siguientes:

	x_1	x_2	x_3	x_4	y
x_1	40	-20	10	0	20
x_2		20	-10	10	-20
x_3			10	0	-40
x_4				40	40
y					60

Calcule el estimador de MC2E del modelo (1) y su matriz de covarianzas.

Problema 9.29. Dado el modelo de comportamiento de la inflación:

$$(1) \quad \pi_t = \beta m_t + \lambda \pi_t^* + u_t$$

donde:

π_t : Tasa de variación de los precios en el instante t : $\pi_t = (p_t - p_{t-1})/p_{t-1}$.

m_t : Tasa de crecimiento del agregado que se utiliza como objetivo intermedio de la política monetaria.

π_t^* : Tasa de inflación del período t , esperada en el período $t - 1$.

u_t : Variable aleatoria, con $E(u_t) = 0$, $E(u_t^2) = \sigma_u^2$.

1. Discutir la estimación del modelo en cada uno de los casos siguientes:

- a) $\pi_t^* = \theta\pi_{t-1}$ y u_t sigue un proceso AR(1).
 b) Expectativas adaptativas: $\pi_t^* = \pi_{t-1}^* + \theta(\pi_{t-1} - \pi_{t-1}^*)$, $0 < \theta < 1$ y u_t es ruido blanco.
 c) Expectativas racionales, siendo u_t ruido blanco.

2. Encontrar la *solución* del modelo (1) bajo el supuesto de expectativas racionales [debe encontrarse una función $\pi_t = g(m_t, m_{t-1}, u_t)$]. Se supone que $m_t = m_{t-1} + \varepsilon_t$. Comparar dicha solución y su estimación directa, con la respuesta dada a 1.c).

Problema 9.30. Dado el modelo:

$$Y_t = \beta Y_{t-1} + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \varepsilon_t, \quad |\beta| < 1$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + a_t$$

donde a_t es un proceso de ruido blanco.

Proponga un estimador consistente para los parámetros de este modelo. Discuta su eficiencia.

Problema 9.31. Cuál es la distribución de Y_t condicionada en su historia pasada Y_{t-s} , $s = 1, 2, \dots$ si dicha variable tiene una estructura estocástica:

$$(1) \quad Y_t = \lambda Y_{t-1} + u_t, \quad |\lambda| < 1$$

$$u_t = N(0, \sigma_u^2), E(u_t \cdot u_s) = 0, \quad t \neq s$$

¿Cuál es la distribución marginal de Y_t para cualquier t ?

Escriba la función de verosimilitud exacta del modelo (1).

Problema 9.32. Supongamos que queremos predecir el nivel de consumo en función de la renta esperada, de forma que

$$(1) \quad C_t = \beta Y_{t+1}^e + \varepsilon_t$$

Dado que la renta esperada no es una variable observable, se utiliza la hipótesis de expectativas adaptativas:

$$(2) \quad Y_{t+1}^e - Y_t^e = (1 - \lambda)(Y_t - Y_t^e)$$

para que la estimación del modelo (1) sea posible.

a) Demostrar que bajo la hipótesis (2) Y_{t+1}^e puede escribirse como una media ponderada de los valores conocidos de Y_t .

b) Utilizando la hipótesis (2), derivar una ecuación de consumo que sea estimable y describir (detalladamente) un método que permita obtener estimadores consistentes de los parámetros del modelo. ¿Son eficientes los estimadores que acaba de obtener?

CAPITULO 10

DEFICIENCIAS MUESTRALES: MULTICOLINEALIDAD Y ERRORES DE MEDIDA

10.1. MULTICOLINEALIDAD: CONCEPTO Y CONSECUENCIAS

Es prácticamente imposible encontrar dos variables económicas cuyo coeficiente de correlación en una determinada muestra sea numéricamente cero. dicho coeficiente puede tomar valores pequeños, pero nunca llega a ser igual a cero y, en general, incluso es poco habitual encontrar coeficientes de correlación inferiores a 0,20 ó 0,25. Granger y Newbold (1974), entre otros autores, han ilustrado cómo basta introducir una tendencia lineal en dos series temporales independientes, para que su correlación aumente notablemente. Así, por ejemplo, que la mayoría de las series macroeconómicas agregadas tengan una apreciable tendencia creciente, hace que sus correlaciones sean importantes.

La *multicolinealidad* aparece cuando las variables explicativas de un modelo econométrico están correlacionadas entre sí, y tiene implicaciones negativas cuando se pretende estimar un modelo lineal por mínimos cuadrados. Consideremos el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t$$

Como sabemos, si existe la inversa de la matriz $X'X$, el estimador MCO de este modelo viene dado por $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$, y su matriz de covarianzas por $\text{Var}(\hat{\beta}) = \sigma_u^2(X'X)^{-1}$.

Supongamos que una de las variables explicativas x_{it} tiene una correlación elevada con respecto a las demás variables explicativas del modelo, es decir: que la regresión lineal

$$x_{it} = \delta_1 + \delta_2 x_{2t} + \dots + \delta_{i-1} x_{i-1,t} + \delta_i x_{i+1,t} + \dots + \delta_{k-1} x_{kt} + v_t \quad [10.1]$$

tiene un coeficiente de determinación alto. En tales condiciones, la variable x_{it} podría aproximarse por una combinación lineal de las restantes variables

explicativas, dada precisamente por la regresión estimada [10.1]. Como consecuencia, una de las columnas de la matriz X (la que contiene las observaciones de la variable x_{it}) es aproximadamente una combinación lineal de las demás y, por tanto, la matriz $X'X$ será aproximadamente singular.

En tanto en cuanto el determinante de $X'X$ no sea numéricamente igual a cero, existirá la inversa $(X'X)^{-1}$ y, por tanto, también existirá el estimador MCO. Este será todavía el estimador de mínima varianza entre los estimadores lineales e insesgados, pero se tendrían varias consecuencias:

a) La solución del sistema de ecuaciones normales está mal definida: En tanto en cuanto la dependencia de x_{it} con respecto a las demás variables no sea exacta, sino aleatoria como en [10.1], la matriz $X'X$ no será exactamente singular, existiendo un único estimador MCO, ya que habrá una única solución al sistema de ecuaciones normales, pero también habrá un número de vectores, β_1, β_2, \dots , que cuando fuesen sustituidos en el sistema serían aproximadamente una solución al mismo.

Este problema es similar al que ocurriría al tratar de resolver la ecuación $ax = b$, cuando a es un número muy pequeño en relación a b . Esta ecuación tiene una única solución: $x^* = \frac{b}{a}$, siempre que a sea distinta de cero. Sin

embargo, si a es muy próximo a cero, entonces una variación en el valor de x alrededor de x^* apenas genera cambios en el producto ax , por lo que un rango posiblemente amplio de valores de x genera valores del producto ax muy próximos a b . Si interpretamos la ecuación $ax = b$ como el sistema de ecuaciones normales $(X'X)\beta = X'y$, identificando el parámetro a con la matriz $X'X$, el parámetro b con $X'y$ y el vector de coeficientes β con x , podemos comprender el tipo de problemas que la multicolinealidad genera.

b) Como consecuencia, pequeñas variaciones muestrales producidas al incorporar o sustraer un número *reducido* de observaciones muestrales introducirían ligeros cambios en los productos $X'X$ y $X'y$ (los parámetros a y b de la ecuación anterior), pero podrían generar importantes cambios en la solución $\hat{\beta}$ del sistema de ecuaciones normales.

El problema está producido por las elevadas correlaciones entre las variables explicativas del modelo. Al variar el número de observaciones muestrales, dichas correlaciones seguirán siendo importantes, pero variará ligeramente su valor estimado en la muestra. Si la matriz $X'X$ es casi singular, esas pequeñas variaciones serán suficientes para alterar sustancialmente la estimación $\hat{\beta}$.

c) Por ser casi singular, la matriz $X'X$ será muy pequeña, lo que vendrá reflejado en su *norma*, ya sea obtenida como función de sus valores propios, o a través de su determinante, que es igual al producto de dichos valores propios. Como consecuencia, la matriz de covarianzas del estimador MCO, que es un múltiplo de $(X'X)^{-1}$, será grande, por lo que dicho estimador es muy poco preciso en esta situación.

Aunque hemos utilizado el ejemplo de datos macroeconómicos, también las variables de naturaleza microeconómica suelen estar correlacionadas entre sí. Por ello, al igual que con las cuestiones de heteroscedasticidad y autocorre-

lación, la cuestión relevante en el trabajo empírico no es la de discutir si existe o no multicolinealidad, sino *en qué medida existe multicolinealidad*. Se trata de debatir si el ignorar esta correlación entre variables explicativas es o no una aproximación suficientemente buena como para que el trabajo realizado bajo tal supuesto tenga validez. Si no lo es, entonces habrá que introducir las modificaciones necesarias para conseguir una especificación econométrica que esté libre de estos problemas.

En el *modelo lineal simple*: $y_t = \beta_1 + \beta_2 x_t + u_t$, la multicolinealidad equivaldría a que la variable explicativa x_t fuese aproximadamente constante. Dicho de otra manera, las fluctuaciones muestrales en x_t contribuyen a aumentar la precisión de las estimaciones de los coeficientes del modelo. Este resultado es válido con gran generalidad, y la idea intuitiva de que para conseguir precisión en las estimaciones mínimo-cuadráticas se necesita de un apreciable grado de fluctuación en las variables explicativas es, en general, cierta.

En el *modelo lineal general*, la multicolinealidad aparece porque las fluctuaciones de las variables explicativas se producen de forma relativamente sincronizada.

Multicolinealidad y pérdida de precisión: Un ejemplo. Para evaluar el efecto de la multicolinealidad sobre la precisión del estimador MCO, se consideró el modelo:

$$y_t = 8 + 5x_{2t} - 3x_{3t} + u_t \quad [10.2]$$

donde $E(u_t) = 0$, $E(u_t u_s) = 0$ si $t \neq s$ y $E(u_t^2) = 25$. Se generaron tres pares de series temporales para las variables (x_{2t}, x_{3t}, u_t) , cambiando en cada caso el grado de correlación entre x_{2t} y x_{3t} , de 0 a 0,90 y a 0,99. El número de observaciones fue de 112, representando 28 años de datos trimestrales, y la variable y_t se obtuvo siempre mediante el modelo [10.2]. Se repitió cada ejercicio cincuenta veces, de modo que las conclusiones no estuviesen afectadas por error muestral.

Los valores medios obtenidos a través de las 50 simulaciones que se llevaron a cabo en cada caso fueron:

	Correlación = 0,0	Correlación = 0,90	Correlación = 0,99
$\rho(x_{2t}, x_{3t})$	0,009 (0,091)	0,898 (0,023)	0,990 (0,002)
$ X'X $	0,992 (0,010)	0,194 (0,040)	0,021 (0,004)
$\hat{\beta}_1$	8,03 (0,441)	7,93 (0,427)	7,96 (0,444)
$\hat{\beta}_2$	5,10 (0,401)	5,13 (1,13)	5,25 (3,50)
$\hat{\beta}_3$	-3,00 (0,390)	-3,18 (1,02)	-3,28 (3,48)
$\hat{\sigma}_u^2$	25,3 (3,26)	26,20 (2,97)	24,13 (3,58)
$\text{Var}(\hat{\beta}_2)$	0,233 (0,047)	1,29 (0,206)	11,22 (2,37)
$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)$	-0,001 (0,021)	-1,05 (0,180)	-11,01 (2,31)
$\text{Var}(\hat{\beta}_3)$	0,232 (0,043)	1,06 (0,175)	11,03 (2,29)

donde los valores en paréntesis denotan las desviaciones típicas calculadas sobre las 50 evaluaciones que se obtuvieron de cada estadístico.

Como puede verse, los valores medios de los coeficientes de correlación muestrales entre x_{2t} y x_{3t} (primera fila) son muy similares a sus valores teóricos. En cuanto al efecto de la multicolinealidad, la primera llamada de atención surge del examen de los valores del determinante $|X'X|$, que queda dividido por 5 al aumentar la correlación de 0,0 a 0,90, y queda de nuevo dividido por 10 al pasar la correlación de 0,90 a 0,99⁽¹⁾.

En cuanto a las estimaciones de los coeficientes, en los tres casos se obtuvieron valores muy próximos a los verdaderos, 8, 5 y -3 . Esto es razonable, pues la multicolinealidad no introduce sesgos en el estimador MCO, aunque también es cierto que fueron siempre más próximos a los verdaderos valores en el caso de ortogonalidad. Las cifras en paréntesis que les acompañan indican la dispersión con que las estimaciones oscilaron alrededor de los valores medios a lo largo de las 50 simulaciones que se efectuaron en cada caso. Es evidente que dichas oscilaciones aumentan notablemente con el grado de correlación existente entre x_{2t} y x_{3t} ⁽²⁾.

La sexta fila muestra la estimación del parámetro σ_u^2 , en todos los casos próxima a su verdadero valor. Como sabemos, el cociente $\frac{(T-k)\hat{\sigma}_u^2}{\sigma_u^2}$ sigue una distribución chi-cuadrado con 109 grados de libertad (ya que $T-k=109$ y $\sigma_u^2=25$), por lo que el estimador $\hat{\sigma}_u^2$ tiene una esperanza matemática de 25 y una desviación típica de 3,39, similar a la observada en cada uno de los tres casos.

Las tres últimas filas de la tabla contienen las varianzas y covarianza de las estimaciones MCO de los coeficientes β_2 y β_3 . Como se muestra, éstas quedan multiplicadas por un factor de 5 al pasar de la ortogonalidad entre x_{2t} y x_{3t} a una correlación de 0,90 entre ambas variables, y quedan multiplicadas por un factor adicional de 10 al aumentar dicha correlación de 0,90 a 0,99. Por otra parte, dichas varianzas son, en cada caso, muy próximas a las obtenidas en el ejercicio de simulación correspondiente al caso de ortogonalidad entre x_{2t} y x_{3t} ; por ejemplo, en promedio, $\text{Var}(\hat{\beta}_2) = 0,233$, mientras que la desviación típica de la distribución empírica de $\hat{\beta}_2$ fue de 0,401.

Así, pues:

a) Un aumento de la correlación existente entre las variables x_2 y x_3 provoca una disminución en el valor del determinante de $X'X$, y esta disminución es mucho mayor cuanto mayor es dicha correlación inicialmente. Así, un 10 por 100 de aumento en la correlación (a partir de un coeficiente de correlación de 0,90) produce un descenso en dicho determinante aproximadamente en un factor de 10.

b) Como consecuencia, la varianza de las estimaciones MCO crece

⁽¹⁾ Las variables fueron normalizadas (se sustrajo su media y se dividió por su desviación típica), con lo que el determinante de la matriz $X'X$ es igual a 1 en el caso de ortogonalidad.

⁽²⁾ El coeficiente β_1 se estimó con una precisión muy similar en los tres casos. Ello no se debe a que sea el término independiente, sino a que es una variable no afectada por la multicolinealidad.

drásticamente al aumentar la correlación entre x_2 y x_3 . Al aumentar la correlación entre x_2 y x_3 , dichas varianzas aumentan en las mismas proporciones en que disminuye el determinante de la matriz $X'X$.

c) El aumento de la correlación entre x_2 y x_3 también genera un aumento en la covarianza de los respectivos coeficientes estimados β_2 y β_3 . Es interesante puntualizar que a pesar de que la correlación entre las variables es positiva, la correlación entre los coeficientes resulta ser negativa. Este resultado, a primera vista sorprendente, tiene sin embargo una clara interpretación: Si la correlación entre las variables es elevada, entonces es bastante difícil separar el efecto que cada una de ellas tiene sobre y_i . Si la estimación del coeficiente de x_{2i} es, por ejemplo, superior a su verdadero valor (desconocido), entonces la estimación de β_3 tenderá a estar por debajo de su verdadero valor para compensar la excesiva importancia que hemos asociado a la variable x_2 . Un argumento análogo puede hacerse para el caso en que la correlación entre ambas variables sea negativa.

d) Como analizamos teóricamente anteriormente, cuando la correlación entre las variables explicativas es importante, pequeñas variaciones numéricas en los momentos muestrales que aparecen en el sistema de ecuaciones normales pueden generar importantes cambios en las estimaciones mínimo-cuadráticas. Por ejemplo, consideremos el sistema de ecuaciones normales:

$$\begin{pmatrix} 1 & 0,90 \\ 0,90 & 1 \end{pmatrix} \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \end{pmatrix}$$

correspondiente a un modelo cuyas variables han sido transformadas en diferencias con respecto a la media. Como las posiciones en la diagonal de la matriz $X'X$ son ambas igual a 1, el elemento fuera de dicha diagonal, 0,90, es el coeficiente de correlación entre x_{2i} y x_{3i} . La solución a dicho sistema y, por tanto, las estimaciones MCO son $\hat{\beta}_2 = -4,21$ y $\hat{\beta}_3 = 15,79$.

Supongamos que el mismo vector $X'y = (10, 12)$ hubiese sido obtenido a partir de un par de variables explicativas con un coeficiente de correlación un 10 por 100 más elevado que el anterior:

$$\begin{pmatrix} 1 & 0,99 \\ 0,99 & 1 \end{pmatrix} \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \end{pmatrix}$$

Las estimaciones MCO serían ahora $\hat{\beta}_2 = -94,0$ y $\hat{\beta}_3 = 105,0$, como se ve bien diferentes de las obtenidas antes.

Supongamos ahora que con la misma correlación entre las variables explicativas (0,99) el vector $X'y$ hubiese sido $X'y = (10, 13)$, sólo ligeramente distinto del anterior:

$$\begin{pmatrix} 1 & 0,99 \\ 0,99 & 1 \end{pmatrix} \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 13 \end{pmatrix}$$

Pues bien, las estimaciones MCO serían $\hat{\beta}_2 = -143,5$, $\hat{\beta}_3 = 155,0$. Es nuevo sensiblemente distintas de las anteriores, a pesar del mínimo cambio.

introducido en el vector $\mathbf{X}'\mathbf{y}$. A pesar de las diferencias entre las estimaciones, es curioso observar que la suma $\hat{\beta}_2 + \hat{\beta}_3$ es muy similar en estos dos últimos casos, 11 y 11,15. Esto se debe a que la multicolinealidad es aproximadamente de la forma $x_{2t} - x_{3t} = 0$, ya que la correlación entre ambas es muy alta.

10.2. MULTICOLINEALIDAD EXACTA Y MULTICOLINEALIDAD APROXIMADA

La presencia de multicolinealidad en un modelo lineal puede revestir dos formas: la que se conoce como *multicolinealidad exacta*, que ocurre cuando una de las variables explicativas es combinación lineal determinista de todas las demás (o de algunas de ellas), y la *multicolinealidad aproximada*, que ocurre cuando una de las variables es «aproximadamente» igual a una combinación lineal de las restantes, como ocurría en [10.1]. Al contrario de lo que pudiera parecer, es más grave en la práctica la multicolinealidad aproximada que la exacta.

En primer lugar, la multicolinealidad exacta es fácilmente detectable, puesto que hace que la matriz $\mathbf{X}'\mathbf{X}$ sea singular. Por consiguiente, al intentar estimar tal modelo con cualquier paquete estadístico, el programa no podría ejecutarse, obteniéndose un mensaje que pondría de manifiesto la singularidad de dicha matriz. La siguiente etapa consistiría entonces en revisar todas las variables explicativas incluidas en el modelo para detectar cuál de ellas causa el problema. Para que haya multicolinealidad exacta, debe ocurrir que estemos incluyendo como explicativas un conjunto de variables que satisfacen una identidad contable, lo que debería ser fácilmente detectable. Puede pensarse en otros casos en que se produzca una dependencia lineal exacta, pero el mencionado es el habitual.

Por el contrario, el problema que surge con la multicolinealidad aproximada es que no es detectable por un ordenador del modo que lo es la multicolinealidad exacta. En efecto, a no ser que el grado de dependencia sea total, entonces la matriz $\mathbf{X}'\mathbf{X}$ será invertible y el paquete estadístico que se esté utilizando obtendrá el estimador MCO. El problema, como ya sabemos, es que habrá muchos otros vectores que «casi» satisfacen las ecuaciones normales y que podrían asimismo llamarse, aproximadamente, estimadores MCO del modelo. Sin embargo, el investigador puede no llegar a conocer nunca esta situación, puesto que el ordenador nos habrá proporcionado un único estimador MCO, y también porque casi ningún paquete estadístico proporciona información acerca de la situación de multicolinealidad, a no ser que el usuario la solicite explícitamente, o se construya su propia rutina para tal fin.

En segundo lugar, por ser la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ «grande», entonces la matriz de covarianzas estimada del vector $\hat{\beta}$ será también «demasiado grande». Como consecuencia, los intervalos de confianza para el contraste de hipótesis lineales serán asimismo de gran tamaño, por lo que dichos contrastes estarán sesgados en el sentido de aceptar cualquier hipótesis nula «demasiado frecuentemente», por lo que serán poco potentes. En particular, tendemos a creer demasiado a menudo que los coeficientes del modelo no son significativos. Esta observación

puede ayudarnos algo a detectar la existencia de multicolinealidad. En efecto, supongamos que al estimar un modelo econométrico se obtiene un R^2 elevado y prácticamente todos los coeficientes no significativos. Si el modelo contiene un término independiente, el alto valor del R^2 indica que la suma residual del modelo es pequeña en comparación con la suma total del modelo y, en consecuencia, el ajuste es bueno. Sin embargo, la multicolinealidad, al generar intervalos de confianza amplios para los coeficientes estimados, facilita que el origen caiga dentro de dicho intervalo, por lo que no se rechazará la hipótesis nula de no significación⁽³⁾.

Cuando existe multicolinealidad exacta, entonces ya hemos visto que existen infinitas soluciones al sistema de ecuaciones normales. Sin embargo, cualquiera de los vectores de parámetros que resuelven el sistema genera la misma suma residual. Esta propiedad es importante, porque implica que el R^2 no depende de la elección del vector β que se haga, por lo que todos estos vectores generan la misma medida de bondad de ajuste del modelo.

Ejemplo 10.1. Para intentar evaluar la importancia del problema, se consideraron las siguientes variables:

<i>Periodo</i>	y_t	x_{2t}	x_{3t}	x_{4t}
1	20	5	10	10
2	12	2	8	6
3	28	7	12	16
4	26	6	4	12
5	14	4	16	8
6	24	8	14	14
7	16	3	6	4

Las variables x_{3t} y x_{4t} constan de las mismas observaciones numéricas, sólo que en distinto orden, de modo que los coeficientes de correlación entre x_{2t} y estas dos variables son $\rho_{23} = 0,32$ y $\rho_{24} = 0,93$, como se ve sustancialmente diferentes entre sí. Una regresión de y_t sobre una constante, x_{2t} y x_{3t} , generó las siguientes estimaciones MCO:

$$y_t = 10,81 + 2,92x_{2t} - 0,54x_{3t} + \hat{u}_t$$

(2,6) (0,42) (0,21)

$$R^2 = 0,92, \quad \hat{\sigma}_u^2 = 2,09$$

donde los valores en paréntesis son las desviaciones típicas estimadas de los parámetros.

⁽³⁾ También es posible que, en presencia de multicolinealidad, se obtuviese un R^2 no muy alto, con coeficientes no significativos (debido a la multicolinealidad), y ello sirviese para concluir que las variables escogidas como explicativas no son adecuadas, a pesar de que, en realidad, el modelo podría no estar muy mal especificado. En este caso, no podríamos detectar la existencia del problema que ha llevado al rechazo del modelo.

Una regresión de y_t sobre una constante, x_{2t} y x_{4t} , produjo las estimaciones:

$$y_t = 6,67 + 1,33x_{2t} + 0,67x_{4t} + \hat{u}_t$$

$$(3,27) \quad (1,61) \quad (0,81)$$

$$R^2 = 0,83, \quad \hat{\sigma}_u^2 = 3,16$$

Ambas regresiones incluyen diferentes variables explicativas y no son, por tanto, directamente comparables entre sí. Sin embargo, llama la atención la no significación de las variables explicativas en la última regresión, a pesar de un alto coeficiente de determinación (los niveles críticos de la distribución t_4 son 2,78 y 4,60 al 95 y 99 por 100 de confianza, respectivamente). Ello refleja la existencia de importante multicolinealidad, producida por el elevado coeficiente de correlación entre las variables explicativas x_{2t} y x_{4t} .

Ejemplo 10.2. Un resultado curioso en la estimación de modelos dinámicos del tipo

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_k x_{t-k} + u_t \quad [10.3]$$

se refiere a la posible omisión, por error de la última variable explicativa, x_{t-k} . Esta es una situación frecuente porque la alta correlación que cabe esperar, generalmente, entre los sucesivos retardos de x_t genera un problema de multicolinealidad en el modelo anterior. Ello produce elevadas desviaciones típicas de los coeficientes y dificulta en gran medida la elección correcta del orden del modelo de retardos.

Cuando ello ocurre, el sesgo en la estimación de los restantes coeficientes depende, de manera complicada, de las correlaciones entre los distintos valores retardados de x_t (variables incluidas y omitidas del modelo). Un estadístico importante en el modelo [10.3] se refiere al efecto a largo plazo de x_t sobre y_t o ganancia: $ELP = \sum_{i=0}^k \beta_i$. Si se omite x_k del modelo, este estadístico estará asimismo sesgado.

Pues bien, como se muestra en un caso sencillo en el Problema 9.26, puede demostrarse que si x_t tiene una estructura $x_t = \rho x_{t-1} + u_t$, entonces el único coeficiente sesgado en la estimación de [10.3] es el último coeficiente incluido, β_{k-1} , que resulta *subestimado* en una magnitud igual a $(1 - \rho)\beta_k$.

En la Tabla 10.1 aparece un resumen de los resultados obtenidos al estimar el modelo [10.3] para el Consumo y el PIB españoles tomados de la Sección 7.6, ambos en pesetas constantes de 1980 y en logaritmos, y para valores $k = 0, 1, \dots, 6$.

Hay que observar, en primer lugar, la relativa estabilidad del efecto contemporáneo, a partir del modelo que incorpora un retardo del PIB, en torno a (0,50, 0,55). Por otra parte, el efecto a largo plazo que aparece en la columna «Suma» muestra una gran estabilidad, con una elasticidad a largo plazo en torno a 0,960.

Los sucesivos valores de \ln PIB, están muy correlacionados, por lo que el modelo presenta gran multicolinealidad, lo que hace que: a) no todos los coeficientes resulten significativos cuando se utiliza un simple contraste t , y b) que

TABLA 10.1. Modelo de retardos distribuidos para la función de consumo
(En logaritmos)

0	1	2	3	4	5	6	SR	Suma	$\hat{\sigma}$	\bar{R}^2
0,968							0,0183	0,968	0,024	0,9973
0,410	0,548						0,0099	0,958	0,018	0,9983
0,549	0,154	0,255					0,0083	0,958	0,017	0,9984
0,488	0,378	-0,139	0,233				0,0062	0,960	0,015	0,9986
0,541	0,255	0,050	-0,041	0,155			0,0054	0,960	0,015	0,9986
0,551	0,292	-0,066	0,171	-0,196	0,209		0,0044	0,961	0,014	0,9986
0,785	-0,054	0,194	-0,074	0,015	0,013	0,096	0,0034	0,975	0,013	0,9987

la imprecisión en la estimación de los coeficientes individuales haga que algunos de ellos tomen valores negativos.

10.3. ESTIMACION DE COEFICIENTES BAJO MULTICOLINEALIDAD EXACTA

Otra característica de un modelo con multicolinealidad es que, aunque el estimador MCO no está unívocamente definido, sin embargo, existen determinadas combinaciones lineales de los parámetros que pueden estimarse con toda precisión. Tomando un ejemplo de Johnston (1984), supongamos que en el modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

se tiene la relación funcional exacta, $x_{3t} = \lambda x_{2t}$, para todo instante $t = 1, 2, \dots, T$. Entonces, la matriz $X'X$ del modelo en diferencias con respecto a la media se convierte en:

$$X'X = \sum_1^T x_{2t}^2 \begin{pmatrix} 1 & \lambda \\ \lambda & \lambda^2 \end{pmatrix}$$

que tiene rango igual a 1 y es, por tanto, singular: $|X'X| = 0$. Debido a esta singularidad, el sistema de ecuaciones normales se reduce en este caso a una sola ecuación (puesto que el rango de la matriz $X'X$ es 1). Dicha ecuación es:

$$\sum_1^T x_{2t} y_t = \hat{\beta}_2 \sum_1^T x_{2t}^2 + \lambda \hat{\beta}_3 \sum_1^T x_{2t}^2$$

que tiene infinitas soluciones $(\hat{\beta}_2, \hat{\beta}_3)$. Sin embargo, existe una única solución para la combinación lineal $\beta_2 + \lambda \beta_3$, dada por:

$$\widehat{\beta_2 + \lambda\beta_3} = \frac{\sum_1^T x_{2t} y_t}{\sum_1^T x_{2t}^2}$$

es decir, que aunque no es posible estimar de modo único el par de parámetros (β_2, β_3) por separado, sí es posible estimar unívocamente la combinación lineal $\beta_2 + \lambda\beta_3$.

La explicación reside en que si utilizásemos la relación entre x_2 y x_3 , entonces se tendría el modelo:

$$y_t = \beta_1 + (\beta_2 + \lambda\beta_3)x_{2t} + u_t$$

que no presenta ningún problema de estimación y que genera un estimador perfectamente definido para su único coeficiente, que es la combinación: $\beta_2 + \lambda\beta_3$. Nótese que, además, el modelo puede estimarse para obtener predicciones acerca del comportamiento futuro de la variable y_t sin ningún problema (aparte de los habituales en cualquier ejercicio de predicción). En efecto, se tendría:

$$E_T y_{T+1} = \beta_1 + (\beta_2 + \lambda\beta_3)E_T x_{2,T+1}$$

que estará también definido de forma única. Todo ello *sugiere* que, como mencionamos en la sección anterior y se cuestiona en el Problema 10.2, la suma residual y el poder explicativo del modelo también estarán unívocamente definidos, a pesar de la presencia de multicolinealidad exacta.

Una discusión totalmente análoga sirve para generalizar estas propiedades al caso de un modelo con k variables explicativas. Así, consideremos el modelo

$$y = X\beta + u \quad [10.4]$$

y supongamos que presenta la particularidad de que $\text{Rango}(X) = r < k$. En consecuencia, también la matriz $X'X$ es de rango inferior a k y, por tanto, singular. Por ser el rango de la matriz X inferior al número de sus columnas, ello quiere decir que las variables explicativas del modelo econométrico son linealmente dependientes entre sí.

Ordenemos las variables explicativas de modo que las r primeras sean linealmente independientes entre sí. Particionemos la matriz X como: $X = [X_r; X_s]$, donde X_r es $T \times r$ y X_s es $T \times s$, con $r + s = k$. Como hay r variables linealmente independientes en X_r , y el rango de la matriz X es precisamente r , se concluye que cada una de las variables en la submatriz X_s puede escribirse como combinación lineal de las variables en X_r . Por tanto, el segundo bloque de la matriz X puede expresarse como función lineal del primero, es decir: $X_s = X_r W$, por lo que se tiene $X = X_r [I_r; W] = X_r Z$, donde Z es la matriz $r \times k$: $[I_r; W]$, y el modelo puede escribirse:

$$y = X\beta + u = X_r Z\beta + u = X_r \beta_r + u \quad [10.5]$$

donde β_r es el vector producto $\beta_r = Z\beta$. Las r combinaciones lineales de coeficientes β que forman el nuevo vector β_r , son perfectamente estimables por MCO y se tiene:

$$\hat{\beta}_r = (X_r' X_r)^{-1} X_r' y$$

De este modo, la idea es eliminar las variables X_s del modelo [10.4] utilizando su expresión como función lineal de las X_r y pasar a estimar los coeficientes resultantes de las variables X_r (las únicas que quedan en el modelo [10.5]). Esto es lo que hicimos en el ejemplo anterior al eliminar la variable x_{3t} , logrando estimar la combinación $\beta_2 + \lambda\beta_3$.

Por otra parte, en el ejemplo numérico de la Sección 10.1 se tenía la relación aproximada $x_{2t} = x_{3t}$, por lo que la suma $\beta_2 + \beta_3$ se puede estimar con bastante aproximación. La matriz Z es, en este ejemplo:

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \text{por lo que: } Z\beta = Z \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 + \beta_3 \end{pmatrix}$$

Resumiendo, hemos visto en esta sección que la presencia de multicolinealidad exacta no impide la estimación de algunas combinaciones lineales de los coeficientes del modelo. Lamentablemente, sin embargo, las combinaciones lineales que pueden estimarse carecerán, generalmente, de significado económico, aunque eso sí, medirán el poder explicativo de las variables X_r , una vez que hemos prescindido de las variables redundantes X_s . La predicción de valores futuros de y_t puede hacerse de modo análogo al del modelo más sencillo anterior. Sin embargo, además de los condicionantes usuales de la predicción, ésta queda sujeta al supuesto de que la misma dependencia lineal entre los vectores de variables X_s y X_r seguirá existiendo a lo largo del horizonte para el que se predice.

Con todo, el caso más delicado y frecuente es aquel en que alguna o varias de las variables del modelo tienen un cierto grado de dependencia lineal de las demás, sin que dicha dependencia sea exacta, pues, como ya hemos mencionado, los paquetes estadísticos no detectarán esta situación, y podemos llegar a utilizar los resultados de la estimación de modo inapropiado. Es, pues, muy importante elaborar un conjunto de reglas que permitan detectar la presencia de multicolinealidad, lo que pasamos a analizar en la sección siguiente.

10.4. DETECCIÓN DE LA MULTICOLINEALIDAD APROXIMADA

Al analizar los efectos de la multicolinealidad sobre el estimador de mínimos cuadrados en las secciones anteriores, hemos puesto de manifiesto los peligros de la multicolinealidad aproximada, por no ser inmediatamente detectable de forma numérica. Dedicamos esta sección precisamente a la exposición de algunos procedimientos de detección de la multicolinealidad aproximada:

unos, utilizando el análisis de relación entre todas las variables explicativas; otros, examinando directamente el condicionamiento de la matriz $\mathbf{X}'\mathbf{X}$.

Una primera posibilidad para detectar la presencia de multicolinealidad aproximada se deduce de los efectos de ésta que hemos examinado en las secciones previas. Así, sospecharemos que existe un importante grado de multicolinealidad en un modelo econométrico si se tiene alguna de las situaciones siguientes:

a) Pequeños cambios en los datos —como la exclusión o adición a la muestra de un reducido número de observaciones— produce importantes variaciones en las estimaciones mínimo cuadráticas.

b) Los coeficientes estimados tienen desviaciones típicas altas y resultan individualmente poco significativos, a pesar de ser conjuntamente significativos y tener el modelo un R^2 elevado.

Sin embargo, no siempre que hay multicolinealidad se producirán necesariamente estas dos situaciones, por lo que consideramos a continuación otros mecanismos de detección de multicolinealidad.

10.4.a. Métodos basados en la correlación entre variables explicativas

La selección de variables explicativas en un modelo de regresión pudiera ser adecuada, aunque debido a correlaciones importantes entre ellas, las varianzas de los coeficientes estimados pudieran ser excesivamente altas, los intervalos de confianza excesivamente grandes, y la hipótesis de no significación podría aceptarse.

En este apartado exploramos la relación existente entre las varianzas estimadas y el grado de relación entre las varianzas explicativas: si descomponemos la matriz \mathbf{X} como

$$\mathbf{X} = [\mathbf{x}_i; \mathbf{X}_i]$$

donde \mathbf{x}_i es el vector columna de observaciones de la i -ésima variable explicativa y \mathbf{X}_i la matriz $T \times (k - 1)$ de observaciones de las restantes variables, entonces la matriz $\mathbf{X}'\mathbf{X}$ puede escribirse:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{x}'_i\mathbf{x}_i & \mathbf{x}'_i\mathbf{X}_i \\ \mathbf{X}'_i\mathbf{x}_i & \mathbf{X}'_i\mathbf{X}_i \end{pmatrix}$$

por lo que, utilizando el lema de inversión de matrices particionadas, se obtiene que el elemento (1, 1) de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ es:

$$[(\mathbf{x}'_i\mathbf{x}_i) - \mathbf{x}'_i\mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}(\mathbf{X}'_i\mathbf{x}_i)]^{-1} = (\mathbf{x}'_i\mathbf{M}_i\mathbf{x}_i)^{-1}$$

donde \mathbf{M}_i es la matriz que ya ha aparecido en capítulos anteriores: $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i$, por lo que se tiene:

$$\text{Var}(\hat{\beta}_i) = \frac{\sigma_u^2}{\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i}$$

y, por tanto, la varianza estimada de cada coeficiente $\hat{\beta}_i$ será inversamente proporcional al valor de la forma cuadrática $\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i$. Pero como vimos tras la Proposición 3.9, ésta no es sino la suma residual de una regresión que tiene como variable dependiente x_i y como variables explicativas el vector formado por las restantes $k - 1$ variables. Esta observación sugiere, por tanto, un método de detección de multicolinealidad, que consistirá en estimar regresiones de cada una de las variables explicativas sobre las restantes. Un R^2 alto en estas regresiones auxiliares detectaría una variable cuya inclusión en el modelo puede producir importante multicolinealidad.

Si denotamos por ST_i la suma total correspondiente a la variable x_i : $\text{ST}_i = \sum_1^T (x_{it} - \bar{x}_i)^2$, por SR_i la suma residual de la regresión auxiliar de x_i : $\text{SR}_i = \mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i$ y por R_i^2 al coeficiente de determinación de dicha regresión, entonces, la expresión de la varianza de $\hat{\beta}_i$ puede escribirse:

$$\text{Var}(\hat{\beta}_i) = \frac{\sigma_u^2}{\text{SR}_i} = \frac{\sigma_u^2}{\text{ST}_i (1 - R_i^2)}$$

que depende de tres factores: a) la varianza del término de error de la regresión original, b) la suma total de la propia variable x_i , y c) el coeficiente de determinación R_i^2 . La varianza σ_u^2 es independiente del grado de correlación entre variables explicativas; la suma total depende sólo de la propia variable x_i , y no de las restantes variables explicativas, al contrario que el coeficiente R_i^2 . Por tanto, es este coeficiente el único determinante de $\text{Var}(\hat{\beta}_i)$ que cambia con el grado de multicolinealidad.

El valor más pequeño de dicha varianza se tiene cuando R_i^2 es cero. Ello ocurre cuando x_i es linealmente independiente de las restantes variables explicativas del modelo original, y entonces la varianza de la estimación de este coeficiente, que denotamos por $\hat{\beta}_i^0$, es:

$$\text{Var}(\hat{\beta}_i^0) = \frac{\sigma_u^2}{\text{ST}_i}$$

por lo que la relación entre las varianzas de la estimación de β_i en un caso de correlación entre variables explicativas y el caso de independencia lineal es

$$\frac{\text{Var}(\hat{\beta}_i)}{\text{Var}(\hat{\beta}_i^0)} = \frac{1}{1 - R_i^2}$$

Para distintos valores del coeficiente de determinación, la relación entre la varianza del estimador y su cota inferior, $\text{Var}(\hat{\beta}_i^0)$, resulta ser:

R_i^2	0,50	0,80	0,90	0,95	0,98	0,99	0,999
Cociente de varianzas	2	5	10	20	50	100	1000

De acuerdo con este análisis, los coeficientes de determinación obtenidos en las regresiones de cada variable explicativa sobre las restantes son un buen indicador de una posible situación de multicolinealidad. El cálculo de estos estadísticos debería ser algo habitual en todo trabajo empírico de regresión, por contraposición a la situación habitual de no prestarles excesiva atención en trabajos aplicados.

En el Ejemplo 10.1, el coeficiente de determinación de una regresión de x_{2t} sobre una constante y x_{3t} fue de 0,10, indicando que la correlación entre ambas variables explicativas es muy baja, mientras que el coeficiente de determinación de la regresión de x_{2t} sobre una constante y x_{4t} fue de 0,86, indicando que la multicolinealidad es un problema importante en esta segunda regresión.

10.4.b. Métodos basados en el tamaño de la matriz $X'X$

Por otra parte, si el problema numérico producido por la presencia de multicolinealidad es la singularidad aproximada de la matriz $X'X$, entonces cabe pensar en utilizar alguna medida del tamaño de esta matriz como indicador de posibles problemas de correlación entre variables. Una posibilidad sería examinar el valor numérico del determinante de dicha matriz, pero esto tiene el problema de que el determinante es sensible a cambios de unidades y, en general, a las unidades de medida utilizadas para las distintas variables, por lo que no sería posible utilizarlo como una medida absoluta de tamaño de $X'X$.

Sin embargo, ya vimos que el determinante de una matriz simétrica es igual al producto de sus valores propios y, por tanto, un examen de estos valores nos da una idea del «tamaño de la matriz». Uno o más valores propios «pequeños» tenderán a generar un valor pequeño para el determinante de la matriz $X'X$. De nuevo tenemos aquí el problema de que los tamaños de dichos valores propios dependen de las unidades de medida, por lo que antes de calcularlos se normaliza cada variable, dividiendo todas sus observaciones muestrales por su desviación típica, o por $\sqrt{\sum_1^T x_{it}^2}$.

Si examinamos el cociente entre el valor propio más grande, $\lambda_{\text{máx}}$, y el menor, $\lambda_{\text{mín}}$, entonces un valor grande de este cociente implicaría que el valor $\lambda_{\text{mín}}$ es realmente pequeño comparado con el $\lambda_{\text{máx}}$, indicando un potencial problema de multicolinealidad. A la raíz cuadrada de este cociente se le llama número de condición de la matriz X , y números de condición mayores de 25 suelen considerarse como problemáticos. Este concepto puede extenderse

calculando un número de condición para cada valor propio de la matriz $X'X$ (es decir, el cociente entre dicho valor propio y el menor de todos ellos) y examinando cuántos de ellos se hallan por encima de 25.

En el ejemplo que analizamos en la Sección 10.2, los valores propios de la matriz $X'X$ (en desviaciones con respecto a la media) fueron $\lambda_1 = 115,69$ y $\lambda_2 = 24,31$ cuando las variables explicativas fueron x_{2t} y x_{3t} (en diferencias con respecto a la media), y $\lambda_1 = 136,84$ y $\lambda_2 = 3,16$ cuando las variables explicativas fueron x_{2t} y x_{4t} . En el primer caso, el número de condición fue 2,18, indicando que la multicolinealidad no es un problema en dicha regresión. En la segunda regresión, dicho número es 6,58 que, sin ser excesivamente elevado, indica que la presencia de multicolinealidad es más importante.

El lector puede probar (Problema 10.4) que si las variables explicativas son ortogonales entre sí y si se normalizan sus unidades de modo que las varianzas muestrales sean todas igual a 1, entonces el número de condición de la matriz $X'X$ será 1, y éste es el menor valor posible. A partir de la situación de ortogonalidad de las variables explicativas, mayor correlación entre ellas generará valores mayores del número de condición.

Finalmente, cuando este método ha detectado multicolinealidad, hay que preguntarse en qué medida se verán afectadas por el problema las estimaciones de las varianzas de los coeficientes. Para ello, recordando la descomposición de una matriz (en este caso $X'X$) utilizando sus vectores propios: $V'(X'X)V = \Lambda$, donde Λ es la matriz diagonal de los valores propios de $X'X$ y V la matriz ortogonal que tiene por columnas sus vectores propios, se tiene:

$$\text{Var}(\hat{\beta}) = \sigma_u^2 (V\Lambda V')^{-1} = \sigma_u^2 (V\Lambda^{-1}V')$$

y por tanto:

$$\text{Var}(\hat{\beta}_i) = \sigma_u^2 \left(\frac{v_{i1}^2}{\lambda_1} + \frac{v_{i2}^2}{\lambda_2} + \dots + \frac{v_{ik}^2}{\lambda_k} \right), \quad i = 1, 2, \dots, k \quad [10.6]$$

que es una expresión alternativa para la varianza de cada coeficiente estimado por MCO. El cociente entre cada uno de los sumandos en el paréntesis que aparece en [10.6] y el valor de dicha suma constituye la proporción de la varianza de cada estimador atribuible a cada valor propio. La idea es que no se quiere tener una gran proporción de la varianza debida al valor propio (o conjunto de ellos), que han sido identificados como potencialmente peligrosos debido a su número de condición: ello querría decir que la multicolinealidad está teniendo un efecto importante sobre la estimación de β_i .

10.5. REMEDIOS CONTRA LA MULTICOLINEALIDAD

10.5.a. Mediante estimación del modelo propuesto

Los remedios que se han sugerido para resolver el problema de la multicolinealidad son variados, pero apenas satisfactorios:

1. *Prestimación.* Johnston (1984) sugiere la posibilidad de excluir alguna variable que pueda estar muy correlacionada con otras variables explicativas y asociar al coeficiente de dicha variable un valor estimado por otro procedimiento.

Por ejemplo, en un modelo de consumo especificado en logaritmos que incluyese varias variables explicativas correlacionadas podría prescindirse de la variable renta si se dispusiese de una elasticidad renta β estimada con datos de sección cruzada y utilizar las diferencias $C_t - \hat{\beta}Y_t$ como variable dependiente. Sin embargo, esta sugerencia tiene un grave problema: En datos de series temporales, la elasticidad renta mide la proporción en que aumentará el consumo cuando aumente la renta de la economía como un agregado. Es decir, que viene asociada con una idea de crecimiento o de ciclo económico. Por el contrario, en datos de sección cruzada, dicha elasticidad mide el grado en que aumentará el consumo al pasar de un grupo de familias a otro grupo con un nivel de renta superior en un mismo instante de tiempo, es decir, en una misma coyuntura económica.

2. *Regresión cresta.* Otra alternativa que se ha sugerido en la literatura es la de la *regresión cresta*: $\hat{\beta}_c = (X'X + cI_k)^{-1}X'y$, que al incrementar los elementos de la diagonal de $X'X$ en una constante cambia el «tamaño» de la matriz y evita el problema de su aproximada singularidad. Este estimador es *sesgado*, pero tiene la propiedad de que, si se elige la constante c de modo «adecuado», su matriz de varianzas $\sigma_u^2(X'X + cI_k)^{-1}X'X(X'X + cI_k)^{-1}$ puede ser menor que la del estimador MCO. Si al investigador no le importa utilizar un estimador sesgado, preocupándose más de que su error cuadrático medio sea pequeño, entonces puede ser que la menor varianza compense el mayor sesgo y el estimador cresta sea preferible al estimador MCO.

Desafortunadamente, el problema es que existe todo un continuo de valores de c con la propiedad citada, y no es trivial cuál debe seleccionarse. Hoerl y Kennard (1970) sugirieron comenzar con un valor muy pequeño de c e ir aumentándolo hasta que las estimaciones MCO se estabilicen. Pero éste parece un criterio un tanto arbitrario. Una sugerencia interesante fue la de Schmidt (1976), que aconsejaba elegir el valor de c que minimizase el ECM del estimador cresta. El grave problema es que tal elección de c depende del verdadero valor (desconocido) de los parámetros del modelo, cuya estimación es precisamente el problema que se quiere resolver. Incluso si se siguen otros criterios para la elección del valor de c , al investigador siempre le quedará la duda de cuáles son las variables explicativas que realmente ha utilizado en la estimación del modelo tras la suma de las constantes a la matriz $X'X$.

La solución que se tome ante la presencia de multicolinealidad debe depender de la finalidad que se busque con el modelo econométrico. Por ejemplo, si la finalidad fundamental de dicha especificación es predictiva, entonces ya hemos visto que la inclusión de las variables redundantes (aquellas que dependen linealmente de las demás) no impide conseguir un buen ajuste global y, con ello, buenas predicciones de la variable endógena. Si, por otra parte, se pretende hacer un análisis estructural que contribuya al conocimiento descriptivo de la economía, entonces la multicolinealidad es un problema grave. Como hemos visto, las estimaciones de mínimos cuadrados están mal

definidas, y cualquier interpretación de los valores estimados de los parámetros es bastante discutible.

3. *Otras soluciones.* En la literatura se han sugerido otras soluciones que no vamos a discutir en detalle:

a) Tratar de formalizar la relación entre las variables explicativas y pasar a un modelo con varias ecuaciones, como los que veremos en los Capítulos 17 y 18, aunque ello entraña un modelo de mayor complejidad.

b) Cuando la multicolinealidad se debe a la presencia, como variables explicativas, de varios retardos de una misma variable, puede especificarse una relación entre sus coeficientes para eliminar alguno de los retardos del modelo (como vimos en el Capítulo 9).

c) Utilizar componentes principales (véase Judge, Griffiths, Hill, Lütkepohl y Lee, 1985).

10.5.b. Mediante exclusión de variables

Parece, sin embargo, que lo más sensato en estas condiciones es realizar un análisis exhaustivo de las variables que están produciendo el problema de multicolinealidad y excluirlas del proceso de estimación, a no ser que con ello se pierda una parte importante del significado estructural del modelo. Con ello entramos inevitablemente en los problemas de mala especificación tratados en la Sección 3.11. Supongamos que en el modelo econométrico

$$y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad [10.7]$$

con las variables en desviaciones con respecto a la media, existen razones para pensar que x_{2t} y x_{3t} están altamente correlacionadas. Una posibilidad consistiría entonces en eliminar x_{3t} del modelo. Como ya hemos visto, cuanto mayor sea la correlación entre ambas variables explicativas, menor será el poder explicativo perdido con la exclusión de esta variable. Sin embargo, tal exclusión genera un estimador sesgado del parámetro β_2 (como vimos en la Sección 3.11). En efecto, si denotamos por $\tilde{\beta}_2$ la estimación MCO de dicho parámetro en el modelo

$$y_t = \beta_2 x_{2t} + v_t$$

entonces se tiene $E(\tilde{\beta}_2) = \beta_2 + \delta\beta_3$, donde δ es el coeficiente de la regresión: $x_{3t} = \delta x_{2t} + \xi_t$. Por otra parte, la varianza del estimador $\tilde{\beta}_2$ es:

$$\text{Var}(\tilde{\beta}_2) = \frac{\sigma_u^2}{\sum x_{2t}^2}$$

Como vimos en la sección anterior, la varianza del estimador MCO del parámetro β_2 en el modelo [10.7] es $\text{Var}(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum x_{2t}^2 (1 - R_{23}^2)}$, donde R_{23} es

el coeficiente de correlación entre x_{2t} y x_{3t} . Evidentemente, el parámetro β_2 es estimado con mayor precisión en el modelo reducido que en el modelo original [10.7]. Parece, pues, que el investigador pudiese estar dispuesto a conformarse con el estimador sesgado $\tilde{\beta}_2$, a cambio de una mayor precisión.

En el Problema 10.3, al final de este capítulo, se pide al lector que compruebe que el cociente de los errores cuadráticos medios (Varianza + Sesgo al cuadrado) de ambos estimadores ($\tilde{\beta}_2$ y $\hat{\beta}_2$) es igual a:

$$\frac{\text{ECM}(\tilde{\beta}_2)}{\text{ECM}(\hat{\beta}_2)} = 1 + R_{23}^2(t^2 - 1) \quad [10.8]$$

donde t denota el estadístico t de Student para el contraste de significación de β_3 en [10.7]. Así, en base al error cuadrático medio, el estimador sesgado $\tilde{\beta}_2$ debe preferirse al estimador insesgado $\hat{\beta}_2$ sólo cuando $|t| < 1$, pues sólo en tal caso será el cociente anterior inferior a la unidad. El estimador obtenido con esta estrategia se denomina *pretest*, pues se basa en contrastar una condición acerca del estadístico t de una regresión previa.

El problema fundamental al hacer esta evaluación en la práctica reside en que dicho estadístico t viene dado por:

$$t^2 = \frac{\beta_3^2}{\text{Var}(\beta_3)} = \frac{\beta_3^2}{\sigma_u^2} \sum_1^T x_{3t}^2 (1 - R_{23}^2) \quad [10.9]$$

y tanto el numerador como el denominador de la fracción son desconocidos, pues tanto β_3 como σ_u^2 se desconocen. Por tanto, una decisión como la citada debe basarse sobre una estimación del valor de t . En ocasiones se ha generalizado este criterio a la decisión de excluir un subconjunto de variables si el estadístico F correspondiente es menor que 1.

También se ha sugerido la utilización de una combinación lineal: $\tilde{\beta} = \alpha \hat{\beta} + (1 - \alpha) \tilde{\beta}$. El error cuadrático medio de tal combinación lineal depende del valor elegido para el parámetro α , y Feldstein (1973) ha probado

que el mínimo valor de dicho error cuadrático se alcanza cuando $\alpha = \frac{t^2}{1 + t^2}$,

donde t es el estadístico de Student de [10.9]. Sin embargo, nos enfrentamos de nuevo al problema de desconocimiento del valor del estadístico t , y la decisión basada en una estimación del mismo puede no ser óptima.

10.5.c. Estimadores restringido y no restringido: una opción delicada

En un contexto más general, con frecuencia se plantea al investigador la elección entre un modelo econométrico o una versión restringida del mismo, dado que difícilmente puede estar seguro de la validez de dichas restricciones. En una situación de multicolinealidad, la opción a considerar sería im-

ner las restricciones de exclusión de las variables que generan el problema de colinealidad.

Los estimadores sin restringir y restringido $\hat{\beta}_{MCO}$ y $\hat{\beta}_R$ pueden compararse de acuerdo con distintos criterios. Una primera posibilidad consistiría en escoger el estimador restringido $\hat{\beta}_R$ si la diferencia entre las matrices de ECM de $\hat{\beta}_{MCO}$ y $\hat{\beta}_R$ es definida positiva. Alternativamente, podría escogerse el estimador restringido si la traza de la primera es superior a la de la última, y el estimador no restringido en caso contrario.

Wallace y Toro-Vizcarrondo (1968 y 1969) demostraron que el estimador restringido es preferible, de acuerdo con el primer criterio, si se satisface la desigualdad:

$$\mu = \frac{(\mathbf{r} - \mathbf{R}\hat{\beta})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\beta})}{2\sigma_u^2} \leq \frac{1}{2}$$

Cuando las hipótesis lineales son ciertas, el estadístico que se obtiene al sustituir en μ los valores de $\hat{\beta}$ y σ_u^2 por sus estimaciones, sigue una distribución $F_{q, T-k}$ multiplicada por $q/2$. Cuando las hipótesis no son ciertas, el estadístico resultante sigue una distribución $F_{q, T-q}$ no centrada, con parámetro de no centralidad igual al valor del estadístico μ . Los autores citados sugieren comparar el valor de μ con los valores de una distribución $F_{q, T-k}$ con parámetro de no centralidad igual a $1/2$. Si el valor de μ supera el valor tabulado de esta distribución, ello es señal de que el parámetro de no centralidad es grande (por tanto, superior a $1/2$), y el estimador MCO sin restringir se considera preferible. Si el estadístico μ supera el valor de las tablas, pero es inferior a $1/2$, se entiende que las restricciones $\mathbf{R}\hat{\beta} = \mathbf{r}$ son falsas, pero su imposición genera un estimador que, aun siendo sesgado, tiene un ECM inferior al estimador sin restringir (que es insesgado).

El estimador restringido es preferible de acuerdo con el segundo criterio (comparación de las trazas de las matrices ECM) si $\mu \leq q/2$, donde q es el número de restricciones lineales que se contrasta, por lo que habría que comparar esta vez el valor de μ con una distribución $F_{q, T-k}$ con parámetro de no centralidad igual a $q/2$. El conjunto de restricciones lineales se rechaza si el estadístico μ supera al valor de las tablas de la distribución citada.

10.6. ERRORES DE MEDIDA

Una dificultad que subyace a casi todo trabajo empírico en Economía es la imposibilidad de disponer de observaciones muestrales de las variables que se pretende relacionar. Por ejemplo, las variables de Contabilidad Nacional, como el PIB, el stock de capital, o el consumo, no son sino estimaciones de conceptos teóricos que no se observan en la realidad. En otros casos, como ocurre con la Renta Permanente, la inteligencia, o la capacidad de un trabajador para realizar una determinada tarea, no disponemos ni siquiera de estimaciones, por lo que suelen utilizarse *variables proxy*, que aproximan los conceptos que se quieren utilizar. Así, por ejemplo, se utilizan los años de

experiencia en el puesto de trabajo como *proxy* de la *habilidad* de un trabajador.

Basándonos en los desarrollos de capítulos previos, cabe esperar que los *errores de medida*, así como el uso de *variables proxy*, introduzcan sesgos en muestras finitas en el estimador MCO. En realidad se trata de una combinación de variables omitidas (si bien no totalmente), a la vez que de inclusión de variables irrelevantes (aunque no completamente). El sesgo será menor: a) cuanto más se aproxime la variable que realmente se incluye en el modelo a la variable que teóricamente debió utilizarse, así como b) cuanto más independiente sea el error de medida de las restantes variables del modelo.

Consideremos el modelo lineal simple:

$$y_i = \beta x_i + u_i, \quad i = 1, 2, \dots, N \quad [10.10]$$

en el que sólo observamos una medida imperfecta de la variable endógena. Es decir, aunque disponemos de información muestral acerca de x_i , sólo observamos:

$$y_i^* = y_i + v_i, \quad i = 1, 2, \dots, N \quad [10.11]$$

donde suponemos que v_i se distribuye $N(0, \sigma_v^2)$, independiente de u_i y x_i .

Incorporando esta variable al modelo tenemos:

$$y_i^* = \beta x_i + (u_i + v_i) = \beta x_i + \varepsilon_i \quad [10.12]$$

que, bajo los supuestos mencionados, puede analizarse por los procedimientos habituales, sin que suponga ninguna dificultad. De hecho, el coeficiente β es el mismo en el modelo teórico [10.10] que en el modelo con las variables observables [10.12].

En consecuencia, los errores de medida en la variable endógena no producen ningún problema importante al estimar por mínimos cuadrados, por lo que en lo sucesivo suponemos que es x_i quien se observa con error:

$$x_i^* = x_i + \omega_i, \quad i = 1, 2, \dots, N \quad [10.13]$$

donde ω_i se distribuye $N(0, \sigma_\omega^2)$, independiente de u_i , de x_i y de y_i .

El modelo en las variables observables es:

$$y_i = \beta x_i^* + (u_i - \beta \omega_i) = \beta x_i^* + \varepsilon_i \quad [10.14]$$

en el que, a diferencia de lo que ocurría en [10.12], tenemos una dificultad: el término de error compuesto ε_i está correlacionado con la variable explicativa x_i^* , ya que:

$$\begin{aligned} \text{Cov}(\varepsilon_i, x_i^*) &= \text{Cov}(u_i - \beta \omega_i, x_i + \omega_i) = \text{Cov}(u_i, x_i) - \beta \text{Cov}(\omega_i, x_i) + \\ &+ \text{Cov}(u_i, \omega_i) - \beta \text{Cov}(\omega_i, \omega_i) = 0 - \beta \cdot 0 + 0 - \beta \sigma_\omega^2 \end{aligned}$$

donde el primer sumando es nulo por los supuestos habituales del modelo de regresión, mientras que los restantes se deben a las hipótesis efectuadas sobre el error de medida.

Esta correlación hace que el estimador MCO de [10.14] sea sesgado. Si calculamos su límite en probabilidad, tenemos:

$$\begin{aligned} \text{plím } \hat{\beta} &= \frac{\text{plím } \frac{1}{N} \sum_1^N x_i^* y_i}{\text{plím } \frac{1}{N} \sum_1^N x_i^{*2}} = \frac{\text{plím } \frac{1}{N} \Sigma (x_i + \omega_i) (\beta x_i + u_i)}{\text{plím } \frac{1}{N} \Sigma (x_i + \omega_i)^2} = \\ &= \beta + \frac{\text{plím } \frac{1}{N} \sum_1^N (x_i + \omega_i) (u_i - \beta \omega_i)}{\text{plím } \frac{1}{N} \sum_1^N (x_i + \omega_i)^2} = \beta + \frac{-\beta \sigma_\omega^2}{S_x^2 + \sigma_\omega^2} = \frac{\beta}{1 + \sigma_\omega^2 / S_x^2}. \end{aligned} \quad [10.15]$$

donde S_x^2 denota $S_x^2 = \text{plím } \frac{1}{N} \sum_1^N x_i^2$ que suponemos que existe. Para obtener [10.15] hemos utilizado nuevamente las hipótesis acerca del error de medida que antes hicimos.

El resultado general es que, *en presencia de errores de medida, el estimador MCO de β estará sesgado hacia el origen*, tanto si β es positivo como si es negativo. La magnitud del sesgo es tanto mayor cuanto mayor sea la volatilidad del error de medida. Un error de medida en x_i que fuese constante, no produciría ningún sesgo en la estimación del coeficiente β .

En el caso de un modelo de regresión múltiple, se tiene:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{X}^* &= \mathbf{X} + \boldsymbol{\omega} \end{aligned}$$

donde permitimos que *todas* las variables explicativas se midan con error. Extendiendo los resultados anteriores, se llegaría a:

$$\text{plím } \frac{\mathbf{X}^* \mathbf{X}^*}{N} = \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{\omega\omega}$$

donde $\boldsymbol{\Sigma}_{xx}$ es el límite: $\boldsymbol{\Sigma}_{xx} = \text{plím } \frac{\mathbf{X}\mathbf{X}}{N}$, supuesto existente.

Por otra parte:

$$\text{plím } \frac{\mathbf{X}^* \mathbf{y}}{N} = \boldsymbol{\Sigma}_{xx} \boldsymbol{\beta}$$

y, finalmente:

$$plim \hat{\beta}_{MCO} = [\Sigma_{xx} + \Sigma_{\omega\omega}]^{-1} \Sigma_{xx} \beta = \beta - [\Sigma_{xx} + \Sigma_{\omega\omega}]^{-1} \Sigma_{\omega\omega} \beta$$

que implica que un solo error basta para generar inconsistencias en todos los coeficientes del modelo.

10.7. ESTIMACION POR VARIABLES INSTRUMENTALES

La estimación consistente en presencia de errores de medida es posible si se dispone de *instrumentos*. Como vimos en la Sección 9.4, un instrumento es una variable, no incluida en el modelo, que: a) esté incorrelacionada, al menos asintóticamente, con el término de error, a la vez que: b) está correlacionada con la variable explicativa para la que actúa de instrumento.

Volviendo al modelo [10.14], el sesgo asintótico del estimador MCO surge por la correlación entre ε_i y x_i^* . Supongamos ahora que se dispone de una variable z_i tal que:

$$plim \frac{1}{N} \sum_1^N z_i \varepsilon_i = 0; \quad plim \frac{1}{N} \sum_1^N z_i x_i^* \neq 0$$

Entonces el estimador de variables instrumentales de [10.14] es:

$$\hat{\beta}_{VI} = \frac{\frac{1}{N} \sum_1^N z_i y_i}{\frac{1}{N} \sum_1^N z_i x_i^*} \quad [10.16]$$

Sin más que utilizar [10.14], se tiene:

$$plim (\hat{\beta}_{VI} - \beta) = plim \frac{\frac{1}{N} \sum_1^N z_i \varepsilon_i}{\frac{1}{N} \sum_1^N z_i x_i^*}$$

que, bajo los supuestos anteriores, es igual a cero.

Si hay varias variables explicativas con errores de medida, debemos utilizar una variable instrumental para cada una de ellas, definiendo el estimador como:

$$\hat{\beta}_{VI} = (Z'X)^{-1} Z'y$$

de lo que [10.16] es un caso particular. Bajo los supuestos anteriores, $\hat{\beta}_{VI}$ es consistente, con matriz de covarianzas asintótica:

$$\text{Var}(\hat{\beta}_{VI}) = \sigma_\varepsilon^2 (Z'X)^{-1} (Z'Z) (X'X)^{-1}$$

que, en el caso de [10.14], se reduce a:

$$\text{Var}(\hat{\beta}_{VI}) = \sigma_e^2 \cdot \frac{\sum_1^N z_i^2}{\left(\sum_1^N z_i x_i^*\right)^2}$$

El parámetro σ_e^2 se estima dividiendo la suma residual por el número de grados de libertad, y los residuos se obtienen mediante $\hat{u}_i = y_i - \mathbf{x}_i' \hat{\beta}_{VI}$.

10.8. UN CONTRASTE DE ESPECIFICACION PARA ERRORES DE MEDIDA

Bajo errores de medida, el estimador MCO es inconsistente, mientras que el estimador de variables instrumentales es consistente. Si, en realidad, no hubiese errores de medida, ambos estimadores serían consistentes y MCO es, además, eficiente, lo que no ocurre con cualquier estimador de variables instrumentales.

Por consiguiente, para contrastar la existencia de errores de medida puede utilizarse un test tipo Hausman como el considerado en la Sección 9.4.d, comparando el valor numérico de la diferencia $(\hat{\beta}_{MCO} - \hat{\beta}_{VI})$ con su matriz de covarianzas.

Por un argumento similar al de la sección mencionada, podemos considerar el estadístico de Wald:

$$W = (\hat{\beta}_{MCO} - \hat{\beta}_{VI})' (\hat{\Sigma}_{VI} - \hat{\Sigma}_{MCO})^{-1} (\hat{\beta}_{MCO} - \hat{\beta}_{VI})$$

que se distribuye asintóticamente como una χ_k^2 .

Es aconsejable utilizar en ambos casos el valor de σ_u^2 estimado a partir de los residuos de variables instrumentales. En tales condiciones, el estadístico W puede calcularse:

$$W = \frac{\mathbf{q}' \{[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\}^{-1} \mathbf{q}}{\hat{\sigma}_u^2} = \frac{\mathbf{q}' [(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}]^{-1} \mathbf{q}}{\hat{\sigma}_u^2}$$

donde las columnas de $\hat{\mathbf{X}}$ son los valores ajustados en la regresión de cada variable en \mathbf{X} sobre el vector de variables en \mathbf{Z} , y \mathbf{q} es el vector diferencia:

$$\mathbf{q} = \hat{\beta}_{MCO} - \hat{\beta}_{VI}$$

10.9. OBSERVACIONES INFLUYENTES

En ocasiones, existen en la muestra observaciones que pueden tener una importancia apreciable en la estimación de mínimos cuadrados. Cuando ello

ocurre, no resulta obvio cuál es el tratamiento más adecuado para dichas observaciones. Por ejemplo, si se trabaja con una sección cruzada de empresas, pueden resultar muy influyentes las observaciones que proceden de las empresas más importantes y es en cierto modo razonable que ellas condicionen el resultado de la estimación. Esto es algo más difícil de justificar con datos de series temporales y, en cualquier caso, es siempre interesante disponer de las estimaciones que se obtendrían sin utilizar las observaciones influyentes, de modo que se puedan comparar ambas. Para ello, necesitamos disponer de procedimientos que detecten la existencia de observaciones influyentes, cuya descripción es el objeto de esta sección.

En primer lugar, pueden examinarse los residuos de una estimación inicial. La intuición sugiere que si una observación es influyente, generarán un residuo de alto valor absoluto, es decir, «atípico». Si se pretenden comparar residuos entre sí, es importante normalizarlos adecuadamente antes de proceder a la comparación. Para ello, recordemos que si $\hat{\mathbf{u}}$ denota el vector de residuos MCO se tiene:

$$\text{Var}(\hat{\mathbf{u}}) = \sigma_u^2 \mathbf{M} = \sigma_u^2 (\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

por lo que cada residuo, normalizado por su desviación típica, sería:

$$\hat{u}_i^* = \frac{\hat{u}_i}{\hat{\sigma}_u \sqrt{m_{ii}}}$$

donde m_{ii} denota el elemento i -ésimo de la diagonal de \mathbf{M} .

Habitualmente se utiliza un procedimiento más completo, que consiste en estimar por MCO omitiendo la observación i -ésima:

$$\hat{\boldsymbol{\beta}}(i) = [\mathbf{X}(i)'\mathbf{X}(i)]^{-1}\mathbf{X}(i)'\mathbf{y}(i)$$

donde la notación (i) indica que se ha omitido dicha observación.

Así tendríamos unos residuos:

$$\hat{u}_i(i) = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}(i) = u_i + \mathbf{x}_i'[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(i)]$$

Como la observación i -ésima no se ha utilizado para obtener $\hat{\boldsymbol{\beta}}(i)$, ambos sumandos son independientes, y se tiene:

$$\text{Var}(\hat{u}_i(i)) = \sigma_u^2 [1 + \mathbf{x}_i'(\mathbf{X}(i)'\mathbf{X}(i))^{-1}\mathbf{x}_i]$$

que puede utilizarse para normalizar el residuo MCO. Valores del residuo normalizado por encima de 2 sugerirán posibles valores influyentes.

El lector debe notar que este tratamiento consiste en «predecir» el valor de y_i a partir de una estimación calculada con las demás observaciones muestrales. La expresión de $\text{Var}(\hat{u}_i(i))$ no es sino la expresión de la varianza del error de predicción MCO que ya vimos en el Capítulo 4.

PROBLEMAS

Problema 10.1. Comentar acerca de las siguientes afirmaciones relacionadas con la multicolinealidad en modelos econométricos:

a) El problema fundamental con la multicolinealidad exacta es que, debido a la singularidad de la matriz $X'X$, es imposible encontrar el estimador MCO.

b) La multicolinealidad aproximada, que consiste en que la matriz $X'X$ es casi singular, aunque todavía invertible, no presenta problemas de definición del estimador MCO, que es igual a $(X'X)^{-1}X'y$.

c) Cuando existe multicolinealidad, la hipótesis nula de no significación de uno o varios coeficientes tiende a aceptarse más frecuentemente de lo que se debiera.

d) La multicolinealidad hace que el estimador MCO sea muy impreciso, por lo que el modelo así estimado es esencialmente inservible a efectos de predicción.

Problema 10.2. Demostrar que si la matriz $X'X$ es singular, entonces todas las soluciones del sistema de ecuaciones normales generan la misma suma residual e igual coeficiente de determinación.

Problema 10.3. Probar que el cociente entre los errores cuadráticos medios de los estimadores $\hat{\beta}_2$ y $\hat{\beta}_3$ introducidos en la Sección 10.5 para el coeficiente β_2 en el modelo

$$y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

donde las variables están en desviaciones con respecto a la media, satisface la relación [10.8].

Problema 10.4. Demostrar que si las variables explicativas son ortogonales entre sí y se normalizan de modo que las varianzas muestrales sean igual a 1, entonces el número de condición de la matriz $X'X$ sería igual a 1.

Problema 10.5. Suponiendo que en el modelo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + u_t$$

se den las siguientes relaciones entre variables:

$$x_{2t} + x_{3t} = 2$$

$$x_{5t} = x_{4t}$$

$$3x_{4t} + 5x_{3t} - x_{2t} = 1$$

encontrar qué combinaciones lineales de los coeficientes $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ serían estimables.

Problema 10.6. (Relación entre el coeficiente \bar{R}^2 y el estadístico de Student.)

a) Probar que la suma residual de un modelo con k variables explicativas, una de ellas constante, satisface la relación:

$$\frac{SR_k}{y'y} = \frac{T - k - 1}{T - k} (1 - \bar{R}_k^2)$$

b) Demostrar que el estadístico t_{T-k-1} para el contraste de significación de una variable explicativa satisface la relación:

$$t_{T-k-1}^2 = \frac{SR_{k-1} - SR_k}{SR_k} (T - k - 1)$$

c) Utilizando los resultados anteriores, probar que:

$$\frac{1 - \bar{R}_{k-1}^2}{1 - \bar{R}_k^2} = 1 + \frac{t^2 - 1}{T - k}$$

d) Probar que $\bar{R}_k^2 \geq \bar{R}_{k-1}^2$ si y sólo si $t^2 \geq 1$ y concluir que para incrementar el valor de \bar{R}^2 de una regresión basta excluir del modelo las variables explicativas que tengan un estadístico t inferior a 1.

Problema 10.7. Utilice los resultados acerca del sesgo producido al omitir variables explicativas (Sección 3.11), para interpretar el sesgo producido por un error de medida en la variable x_i en el modelo lineal simple:

$$y_i = \beta x_i + u_i, \quad i = 1, 2, \dots, N$$

Problema 10.8. Demuestre la expresión [10.16] siguiendo las líneas que se sugieren en el texto.

Problema 10.9. Considere el modelo de regresión:

$$y_t = \beta x_t + u_t, \quad t = 1, 2, \dots, T$$

donde las variables se hallan en diferencias respecto a la media. La variable x_t se mide con error:

$$x_t^* = x_t + \omega_t, \quad t = 1, 2, \dots, T$$

y x_t es, en realidad, una variable aleatoria que evoluciona a través del tiempo de acuerdo con:

$$x_t^* = \rho x_{t-1}^* + v_t$$

Demuestre que, a diferencia del modelo de sección cruzada de la Sección 10.6, el coeficiente β puede estimarse consistentemente mediante el cociente:

$$\hat{\beta} = \frac{\sum_{t=2}^T x_{t-1}^* y_t}{\sum_{t=2}^T x_{t-1}^* x_t^*}$$

Problema 10.10. Considere el siguiente modelo de errores en las variables:

$$\begin{aligned} Y_{it}^* &= a + bX_{it}^*, & i &= 1, \dots, N, & t &= 1, 2 \\ X_{it} &= X_{it}^* + \varepsilon_{it} \\ Y_{it} &= Y_{it}^* + e_{it} \end{aligned}$$

donde:

$$\begin{aligned} E(\varepsilon_{it}) &= E(e_{it}) = E(X_{it}^* \varepsilon_{is}) = E(X_{it}^* e_{is}) = 0 \\ E(Y_{it}^* \varepsilon_{is}) &= E(Y_{it}^* e_{is}) = E(\varepsilon_{it} e_{is}) = 0 \\ E(\varepsilon_{it}^2) &= \sigma_e^2; \quad E(e_{it}^2) = \sigma_e^2; \quad E(\varepsilon_{i1} \varepsilon_{i2}) = \rho_e \sigma_e^2 \\ E(e_{i1} e_{i2}) &= \rho_e \sigma_e^2; \quad E(X_{it}^{*2}) = \sigma_x^2; \quad E(X_{i1}^* X_{i2}^*) = \rho_x \sigma_x^2 \\ \forall i, \quad t &= 1, 2 \quad \text{y} \quad s = 1, 2 \end{aligned}$$

Las variables X_{it} e Y_{it} son observables para $i = 1, \dots, N$ y $t = 1, 2$. Sea b_1 el estimador de variables instrumentales de b , obtenido en una regresión de sección cruzada usando datos del segundo periodo de tiempo y X_{i1} el instrumento. Sea b_2 el estimador de variables instrumentales de b obtenido usando la misma sección cruzada pero con Y_{i1} como instrumento. Probar que si

$$\rho_e > 0, \quad \rho_x > 0 \quad \text{y} \quad \rho_x > 0$$

entonces:

$$\text{plim}(b_1) \leq b \leq \text{plim}(b_2)$$

Problema 10.11. Dada la siguiente función: $r_t = a + bE_t D_{t+2} + \varepsilon_t$, donde:

r_t : Tipo de interés en el período t .

$E_t D_{t+2}$: Déficit público del período $t + 2$ esperado en el período t .

ε_t : Proceso de ruido blanco.

Si se estima el modelo $r_t = a + bD_{t+2} + u_t$, demuestre que:

- $E(u_t \cdot D_{t+2}) \neq 0$.
- u_t sigue un proceso MA(1).

Problema 10.12. El modelo de la renta permanente de M. Friedman sugiere la relación:

$$(1) \quad Y_i = \alpha + \beta X_i + u_i$$

donde:

Y_i : Gasto de consumo de una familia.

X_i : Renta permanente.

La dificultad práctica surge porque la renta permanente no es directamente observable, por lo que se utiliza en la estimación el promedio de la renta familiar de los últimos cinco años, por ejemplo, denotado por X_i^* .

El analista cree que la relación entre ambas variables puede representarse:

$$X_i^* = X_i + w_i$$

donde w_i es ruido blanco. De este modo, en vez de estimar (1) estimaremos:

$$(2) \quad Y_i = \alpha + \beta X_i^* + v_i$$

- a) Explicar la relación existente entre u_i y v_i .
- b) Calcular el sesgo asintótico del estimador MCO de (2).
- c) ¿Puede afirmarse si $\hat{\beta}^{\text{MCO}}$ será una subestimación o una sobrestimación de β ?

Problema 10.13. Un analista pretende estimar la propensión marginal al consumo de un determinado país utilizando datos anuales de series temporales y utilizando MCO en el modelo:

$$(1) \quad C_t = \alpha + \beta Y_t^d + u_t$$

donde se supone $E(Y_t^d u_t) = 0$.

El problema reside en que no observa la renta disponible, sino la renta agregada Y_t , que cree que se relaciona con la renta disponible por medio de la expresión $Y_t^d = Y_t + \varepsilon_t$, donde ε_t es ruido blanco, y $E(\varepsilon_t Y_t^d) = 0$ y $E(\varepsilon_t u_t) = 0$. ¿Por qué razón debía esperarse que la estimación mínimo cuadrática de (1) resulte inconsistente? ¿Es razonable esperar que ε_t sea ruido blanco? Encuentre la expresión analítica para el sesgo asintótico del estimador MCO de la propensión marginal al consumo. ¿Cuántas razones existen en este modelo para esperar la inconsistencia del estimador MCO?

Como alternativa, el investigador considera la posibilidad de utilizar datos agregados de renta e impuestos para estimar el modelo

$$(2) \quad C_t = \alpha + \beta_1 Y_t + \beta_2 T_t + v_t$$

y contrastar la hipótesis $H_0: \beta_1 = -\beta_2$. ¿Será inconsistente el estimador MCO del coeficiente β_1 en (2)? ¿Y los de los demás coeficientes del modelo? ¿Es cierto que, en la medida en que T_t dependa de la renta Y_t , la inconsistencia será mayor que antes?

En una segunda alternativa, el investigador considera estimar el modelo:

$$(3) \quad C_i = \alpha + \beta Y_i^d + u_i$$

donde $Y_i^d = Y_i - T_i$. ¿Tendría el valor estimado de β la misma interpretación que en el modelo (1)? Encuentre la expresión analítica para la inconsistencia del estimador MCO de β . ¿Cómo podría contrastarse la homogeneidad de la propensión marginal al consumo si se distribuyen las 4.000 familias que componen la muestra en dos grupos de renta alta y baja, con tamaños N_1 y $4.000 - N_1$, respectivamente?

CAPITULO 11

MODELOS NO LINEALES

11.1. INTRODUCCION

11.1.a. Especificaciones no lineales

La Teoría Económica propone modelos de relación entre variables económicas, pero deja generalmente indeterminada la forma funcional de dichas relaciones, lo que sugiere que éstas puedan ser, en ocasiones, no lineales. En tales casos, el modelo econométrico es del tipo:

$$y_t = f(\mathbf{x}_t, \boldsymbol{\beta}) + u_t, \quad t = 1, 2, \dots, T \quad [11.1]$$

donde $f(\mathbf{x}_t, \boldsymbol{\beta})$ es, en general, una función no lineal de las componentes de los vectores \mathbf{x}_t y $\boldsymbol{\beta}$. Como veremos en seguida, las dificultades surgen cuando, condicional en \mathbf{x}_t , la dependencia de y_t respecto de $\boldsymbol{\beta}$ es no lineal, por lo que es ésta la que nos preocupa. Como caso particular en que la dependencia respecto a $\boldsymbol{\beta}$ es lineal, *el modelo lineal general* aparece cuando $f(\mathbf{x}_t, \boldsymbol{\beta}) = \mathbf{x}_t' \boldsymbol{\beta}$.

Al hablar por primera vez de esta dificultad en el Capítulo 3, se proporcionaba como ejemplo la función de producción Cobb-Douglas. Es ésta, en efecto, una situación en que, de acuerdo con el modelo teórico, el nivel del producto depende, en forma no lineal, de los inputs del proceso productivo. Sin embargo, ya mostramos en aquel momento que, si dicha función de producción viene afectada por un término de error multiplicativo, entonces es equivalente a un modelo *lineal* que relaciona el *logaritmo* del producto con los *logaritmos* de las cantidades de los inputs utilizadas en producción.

Una especificación no lineal de un modelo econométrico puede estar indicando la incertidumbre del investigador acerca de la verdadera relación entre las variables del modelo. Por ejemplo, supongamos que para relacionar los gastos en bienes de consumo C_t , con la renta disponible Y_t , se especifica el modelo:

$$C_t = \beta_1 + \beta_2 Y_t^{\beta_3} + u_t \quad [11.2]$$

En dicho modelo, la estimación del parámetro β_3 permitiría contrastar la hipótesis de dependencia lineal o propensión marginal al consumo constante ($\beta_3 = 1$), frente a otras alternativas, como la de una menor sensibilidad del gasto en consumo a variaciones en la renta disponible ($\beta_3 < 1$). Este modelo sería, claramente, menos restrictivo que una especificación lineal; también puede interpretarse como una primera especificación, para pasar a estimar un modelo lineal si la hipótesis $\beta_3 = 1$ no se rechaza en una primera estimación del modelo [11.2].

Sin embargo, conviene distinguir entre varios tipos de no linealidades que pueden presentarse en la práctica. Consideremos los modelos:

$$a) \quad y_t = \beta_1 + \beta_2 e^{x_{2t}} + \beta_3 x_{3t} x_{4t} + u_t.$$

$$b) \quad y_t x_t + \beta_1 \ln y_t = \beta_2 x_t + u_t.$$

$$c) \quad y_t = \beta_1 + \beta_2 e^{\beta_3 x_{2t}} + u_t.$$

$$d) \quad y_t = \beta_1 + (\ln \beta_2) x_t + u_t.$$

$$e) \quad y_t = \beta_1 + \beta_2 x_t^{\beta_3} + u_t.$$

Estos cinco modelos tienen en común el proponer relaciones no lineales entre la variable endógena por un lado y las variables explicativas y los coeficientes por otro. Sin embargo, también presentan diferencias importantes que hacen que su tratamiento econométrico sea bien distinto. La no linealidad del modelo a) afecta únicamente a sus variables, pero no a sus coeficientes. En tal caso, basta definir unas nuevas variables: $z_{2t} = \exp(x_{2t})$, $z_{3t} = x_{3t} x_{4t}$, para obtener el modelo lineal:

$$a') \quad y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + u_t$$

que puede tratarse, sin ninguna dificultad, por los métodos desarrollados en los capítulos previos. En definitiva, siempre que la no linealidad del modelo afecte únicamente a sus variables explicativas, entonces dicho problema queda resuelto mediante una *transformación de datos*. La única excepción a esta afirmación la constituyen los modelos en que la no linealidad afecta *también a la variable endógena* de algún modo que haga imposible expresarla de modo explícito como función de los vectores \mathbf{x}_t y $\boldsymbol{\beta}$. Esto es lo que sucede en el modelo b). La forma funcional de tales modelos es una función implícita:

$$g(y_t, \mathbf{x}_t, \boldsymbol{\beta}) = u_t, \quad t = 1, 2, \dots, T \quad [11.3]$$

de la que el modelo [11.1] es, a su vez, un caso particular.

La no linealidad del modelo c) afecta tan sólo a sus coeficientes pero no a sus variables. Mediante los métodos estudiados en los capítulos anteriores podrían estimarse el término independiente y el coeficiente de la variable x_t , pero no podrían recuperarse estimaciones de los coeficientes β_2 y β_3 , a no ser que se contara con información adicional acerca de sus valores numéricos (por ejemplo, si su suma o cociente fuesen conocidos).

Por último, hay ocasiones en que el modelo es no lineal en los coeficientes sin que ello presente dificultades serias de estimación, como es el caso del

modelo *d*). En dicho modelo se estima un término independiente y un coeficiente de la variable x_t , y a continuación puede recuperarse el valor del parámetro β_2 mediante $\hat{\beta}_2 = \exp(\text{coeficiente de } x_t \text{ estimado})$. Sin embargo, el valor de β_2 así obtenido no heredaría las propiedades estadísticas que pudiera tener el estimador de e^{β_2} . El modelo *e*) es otro ejemplo de modelo no lineal que *no* puede tratarse por métodos lineales.

Finalmente, es importante observar que, a diferencia de los modelos lineales, en modelos no lineales el número de parámetros no coincide necesariamente con el número de variables explicativas, como ocurre en los modelos *b*), *c*) y *e*) anteriores.

11.1.b. Una aproximación lineal al modelo no lineal

Un primer tratamiento de la estimación del modelo [11.1] consistiría en obtener la mejor aproximación lineal (mediante un desarrollo en serie de Taylor) de la función $f(\mathbf{x}_t, \boldsymbol{\beta})$ alrededor de un estimador inicial $\hat{\boldsymbol{\beta}}$ y estimar el modelo lineal resultante, utilizando los procedimientos desarrollados en capítulos anteriores. Dicha aproximación es:

$$y_t \simeq f(\mathbf{x}_t, \hat{\boldsymbol{\beta}}) + \left(\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + u_t, \quad t = 1, 2, \dots, T$$

o, lo que es lo mismo:

$$y_t - f(\mathbf{x}_t, \hat{\boldsymbol{\beta}}) + \left(\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \hat{\boldsymbol{\beta}} = \left(\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T$$

obteniéndose el modelo lineal:

$$y_t^* \simeq \left(\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T \quad [11.4]$$

donde:

$$y_t^* = y_t - f(\mathbf{x}_t, \hat{\boldsymbol{\beta}}) + \left(\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \hat{\boldsymbol{\beta}}$$

En lo sucesivo, denotamos por $\frac{\partial f_t}{\partial \boldsymbol{\beta}}$ el vector gradiente en cada período: $\frac{\partial f(\mathbf{x}_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ de dimensión $k \times 1$ y por $\frac{\partial f_t(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}$ su valor en el punto $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(1)}$.

⁽¹⁾ A pesar de lo que pudiera parecer, la notación $\frac{\partial f_t(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}$ denota la operación de derivada parcial en primer lugar y la evaluación de ésta en el punto $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. Lógicamente, no tendría sentido efectuar estas operaciones en sentido inverso.

De este modo, dada una primera aproximación al estimador $\hat{\beta}$, se trataría de construir la variable y_t^* , así como las k variables que componen el valor del gradiente de la función $f(x_t, \beta)$ en el punto $\beta = \hat{\beta}$, es decir, $\frac{\partial f_t}{\partial \beta_i}$, $i = 1, 2, \dots, k$.

Las «observaciones muestrales» correspondientes a estas variables son función de las observaciones muestrales de y_t , x_t y del vector $\hat{\beta}$. A continuación se estima por mínimos cuadrados el modelo lineal que tiene por variable endógena y_t^* , y por variables explicativas los componentes del vector gradiente $\frac{\partial f_t(\hat{\beta})}{\partial \beta}$ para obtener la nueva estimación del vector β .

Si denotamos por $\frac{\partial f}{\partial \beta}$ y por $f(\mathbf{X}, \beta)$ las matrices:

$$\frac{\partial f}{\partial \beta} = \begin{pmatrix} \frac{\partial f_1}{\partial \beta_1} & \frac{\partial f_1}{\partial \beta_2} & \dots & \frac{\partial f_1}{\partial \beta_k} \\ \frac{\partial f_2}{\partial \beta_1} & \frac{\partial f_2}{\partial \beta_2} & \dots & \frac{\partial f_2}{\partial \beta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_T}{\partial \beta_1} & \frac{\partial f_T}{\partial \beta_2} & \dots & \frac{\partial f_T}{\partial \beta_k} \end{pmatrix}; \quad f(\mathbf{X}, \beta) = \begin{pmatrix} f(x_1, \beta) \\ f(x_2, \beta) \\ \dots \\ f(x_T, \beta) \end{pmatrix}$$

de dimensiones $T \times k$ y $T \times 1$ respectivamente, y por $\frac{\partial f(\hat{\beta})}{\partial \beta}$, $f(\mathbf{X}, \hat{\beta})$ sus evaluaciones en el punto $\beta = \hat{\beta}$, el estimador MCO del modelo lineal [11.4] sería:

$$\begin{aligned} \tilde{\beta} &= \left[\left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]^{-1} \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' y^* = \\ &= \left[\left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]^{-1} \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(y - f(\mathbf{X}, \hat{\beta}) + \frac{\partial f(\hat{\beta})}{\partial \beta} \hat{\beta} \right) = \\ &= \hat{\beta} + \left[\left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]^{-1} \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \hat{u} \end{aligned}$$

donde $\hat{u} = y - f(\mathbf{X}, \hat{\beta})$ es el residuo obtenido con la estimación inicial $\hat{\beta}$. Esta expresión proporciona la nueva estimación $\tilde{\beta}$ a partir de $\hat{\beta}$, mientras que la estimación del parámetro σ_u^2 puede obtenerse de modo similar a un modelo lineal:

$$\hat{\sigma}_u^2 = \frac{\tilde{u}'\tilde{u}}{T - k}, \quad \text{donde} \quad \tilde{u} = y - f(\mathbf{X}, \tilde{\beta})$$

Si existe la matriz inversa $\left[\left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]^{-1}$, entonces la distribución de probabilidad del estimador de mínimos cuadrados de esta aproximación lineal es:

$$N\left(\boldsymbol{\beta}, \sigma_u^2 \left[\left(\frac{\partial f(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right)' \left(\frac{\partial f(\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right) \right]^{-1}\right)$$

mientras que $\frac{(T-k)\hat{\sigma}_u^2}{\sigma_u^2}$ se distribuye, aproximadamente, como una variable chi-cuadrado con $T-k$ grados de libertad, independiente de la distribución Normal del vector $\tilde{\boldsymbol{\beta}}$. Todo ello permite llevar a cabo cualquier análisis de inferencia de modo análogo a como se discutió en el Capítulo 4.

Ejemplo 11.1. Consideremos la estimación del modelo no lineal:

$$y_t = f(x_t, \alpha, \beta) + u_t = \alpha e^{\beta x_t} + u_t \quad [11.5]$$

en el que se tiene el vector gradiente:

$$\left(\frac{\partial f_t}{\partial \alpha}, \frac{\partial f_t}{\partial \beta} \right) = (e^{\beta x_t}; \alpha x_t e^{\beta x_t})$$

por lo que el modelo puede aproximarse linealmente por:

$$\begin{aligned} y_t - f(x_t, \alpha_0, \beta_0) + \frac{\partial f_t(\alpha_0, \beta_0)}{\partial \alpha} \alpha_0 + \frac{\partial f_t(\alpha_0, \beta_0)}{\partial \beta} \beta_0 &= \\ &= \frac{\partial f_t(\alpha_0, \beta_0)}{\partial \alpha} \alpha + \frac{\partial f_t(\alpha_0, \beta_0)}{\partial \beta} \beta + u_t \end{aligned} \quad [11.6]$$

a partir de unos valores iniciales (α_0, β_0) , para estimar los parámetros α y β . Si tomamos como valores iniciales $\beta_0 = 0$, $\alpha_0 = \bar{y}$ y denotamos por y_t^* el miembro izquierdo de [11.6], tenemos:

$$y_t^* = y_t - \bar{y} + \bar{y} + \bar{y}x_t \cdot 0 = y_t$$

y el modelo linealizado [11.6] resulta:

$$y_t^* = y_t = \alpha + \beta x_t^* + u_t$$

donde $x_t^* = \bar{y}x_t$, $t = 1, 2, \dots, T$.

Hay que hacer hincapié en que este procedimiento sólo dará buenos resultados si las condiciones iniciales están próximas a los verdaderos valores de α y β , y eso es algo sobre lo que a priori no tenemos mucha información.

11.2. MINIMOS CUADRADOS NO LINEALES

Otra posibilidad es tratar de aplicar el procedimiento de minimización de la suma residual directamente al modelo no lineal [11.1]. En efecto, la idea que

o, en forma matricial:

$$\sum_1^T \frac{\partial f_t(\hat{\beta})}{\partial \beta} \hat{u}_t = \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \hat{u} = \mathbf{0}_k \quad [11.9]$$

que se reducen a la ortogonalidad entre residuos y variables explicativas cuando el modelo es lineal.

Ejemplo. El sistema de ecuaciones normales para el modelo del Ejemplo 11.1 es:

$$\begin{cases} \sum_1^T y_t e^{\beta x_t} = \hat{\alpha} \sum_1^T e^{2\beta x_t} \\ \sum_1^T y_t x_t e^{\beta x_t} = \hat{\alpha} \sum_1^T x_t e^{2\beta x_t} \end{cases}$$

Más generalmente, puede establecerse la siguiente:

Definición 11.1. El estimador de mínimos cuadrados del modelo $y_t = f(x_t, \beta) + u_t$, $t = 1, 2, \dots, T$ es aquel vector de coeficientes $\hat{\beta}$ que genera un vector de residuos ortogonal a cada uno de los componentes del vector gradiente de la función $f(x_t, \beta)$ evaluado en $\hat{\beta}$, como en [11.9].

Una diferencia muy importante con el modelo lineal es que, en modelos en que la función $f(x_t, \beta)$ no depende linealmente del vector β , sus derivadas parciales tampoco serán, en general, funciones lineales de las componentes del vector β . Esto se aprecia en el ejemplo anterior, donde se muestra que: a) la obtención del estimador de mínimos cuadrados requiere la resolución de un sistema de ecuaciones normales no lineales, y b) dicho estimador será, en general, una función no lineal del vector y .

Esta peculiaridad de los modelos no lineales genera, a su vez, una serie de dificultades:

1. La dependencia lineal del estimador MCO con respecto al vector \mathbf{u} era una característica fundamental para probar que dicho estimador era insesgado. El hecho de que el estimador de *mínimos cuadrados no lineales* dependa del vector \mathbf{y} , y en consecuencia del vector \mathbf{u} en forma no lineal, hace que dicho estimador no sea, en general, insesgado. En estas nuevas condiciones carecemos de un resultado que garantice la optimalidad de este estimador entre los de una determinada clase. Las propiedades del estimador de mínimos cuadrados no lineales vendrán de su posible relación con el estimador de máxima verosimilitud.

2. La solución a un sistema de ecuaciones no lineales puede no ser única y, por tanto, un modelo no lineal puede poseer varios estimadores mínimo cuadráticos. Por la misma razón, pudiera ser que el estimador de mínimos cuadrados no existiese, pues un sistema de ecuaciones no lineales no siempre tiene solución. Esto es bien distinto de lo que ocurre en el caso lineal, en el que siempre existe al menos una solución.

Por no poder obtener una expresión explícita para el estimador de mínimos cuadrados, hay que resolver el sistema de ecuaciones normales o el problema de optimización del que éstas proceden por métodos numéricos

(algoritmos) que veremos en el siguiente capítulo. Como ya veremos, cuando el tamaño muestral crece, el estimador de mínimos cuadrados obtenido por alguno de los algoritmos numéricos mencionados tiene una distribución Normal, con esperanza β y matriz de covarianzas:

$$\text{Var}(\hat{\beta}) = \sigma_u^2 \left[\Sigma_1^T \left(\frac{\partial f(\mathbf{x}_t, \hat{\beta})}{\partial \beta} \right) \left(\frac{\partial f(\mathbf{x}_t, \hat{\beta})}{\partial \beta} \right)' \right]^{-1}$$

donde el parámetro σ_u^2 se estima mediante⁽²⁾:

$$\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k}, \quad \text{donde} \quad \hat{\mathbf{u}} = \mathbf{y} - f(\mathbf{X}, \hat{\beta})$$

De nuevo puede verse que estas expresiones se reducen a las ya conocidas en el caso en que el modelo sea lineal. Como consecuencia de la distribución Normal que acabamos de ver, los habituales contrastes de hipótesis mediante estadísticos t o F son válidos, sin más que utilizar las expresiones anteriores en el cálculo de la matriz de covarianzas. Las condiciones bajo las que los resultados anteriores son válidos incluyen la existencia de un único mínimo global de la función $\text{SR}(\beta)$ y la no singularidad de la matriz límite:

$$p\lim_T \left[\frac{1}{T} \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]$$

Ejemplo 11.1 (continúa). Las condiciones de optimalidad para la obtención del estimador MCNL en el Ejemplo 11.1 serían:

$$\frac{\partial \text{SR}(\hat{\alpha}, \hat{\beta})}{\partial \alpha} = -2 \Sigma_1^T [(y_t - \hat{\alpha} e^{\beta x_t}) e^{\beta x_t}] = 0$$

$$\frac{\partial \text{SR}(\hat{\alpha}, \hat{\beta})}{\partial \beta} = -2 \Sigma_1^T [(y_t - \hat{\alpha} e^{\beta x_t}) x_t e^{\beta x_t}] = 0$$

que carecen de solución explícita. Suponiendo que el sistema pudiera resolverse, la matriz de covarianzas de las estimaciones sería:

$$\begin{aligned} \widehat{\text{Var}}(\hat{\alpha}, \hat{\beta}) &= \hat{\sigma}_u^2 [\Sigma_1^T \nabla f_t(\hat{\alpha}, \hat{\beta}) \nabla f_t(\hat{\alpha}, \hat{\beta})]^{-1} = \\ &= \hat{\sigma}_u^2 \begin{pmatrix} \Sigma_1^T e^{2\beta x_t} & \alpha \Sigma_1^T x_t e^{2\beta x_t} \\ \alpha \Sigma_1^T x_t e^{2\beta x_t} & \alpha^2 \Sigma_1^T x_t^2 e^{2\beta x_t} \end{pmatrix}^{-1} \end{aligned}$$

⁽²⁾ No es sorprendente que la aproximación lineal que antes discutimos tenga la misma distribución asintótica que el estimador de mínimos cuadrados, puesto que el modelo

$y_t^* = \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \beta + u_t$ coincide con el modelo $y_t = f(\mathbf{x}_t, \beta) + u_t$ cuando $\hat{\beta} = \beta$. Por tanto, según iteramos en el modelo linealizado, nos aproximamos al estimador de MCNL.

donde $\hat{\sigma}_u^2$ se obtendría de:

$$\hat{\sigma}_u^2 = \frac{\sum_1^T (y_t - \hat{\alpha} e^{\hat{\beta} x_t})^2}{T - 2}$$

11.3. EL ESTIMADOR DE MAXIMA VEROSIMILITUD

Como ya vimos en el caso del modelo lineal, la obtención del estimador de máxima verosimilitud precisa de un determinado supuesto acerca de la distribución de probabilidad del término de error, de modo que, en realidad, no existe «el» estimador de máxima verosimilitud, sino «un» estimador de máxima verosimilitud dependiendo del supuesto que se haga acerca de dicha distribución de probabilidad.

11.3.a. Condiciones necesarias

Supongamos, como es habitual, que tal distribución es $\mathbf{u} \sim N(\mathbf{0}_T, \sigma^2 \mathbf{I}_T)$, en cuyo caso la función de verosimilitud muestral es:

$$L(\boldsymbol{\beta}, \sigma_u^2) = \left(\frac{1}{2\pi\sigma_u^2} \right)^{T/2} \exp \left[\left(\frac{-1}{2\sigma_u^2} \right) \sum_1^T (y_t - f(\mathbf{x}_t, \boldsymbol{\beta}))^2 \right]$$

y su logaritmo, evaluado en $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2)$:

$$\ln L(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \hat{\sigma}_u^2 - \frac{1}{2\hat{\sigma}_u^2} \text{SR}(\hat{\boldsymbol{\beta}}) \quad [11.10]$$

donde $\text{SR}(\hat{\boldsymbol{\beta}}) = \sum_1^T (y_t - f(\mathbf{x}_t, \hat{\boldsymbol{\beta}}))^2$. La expresión [11.10] ilustra claramente la propiedad de que si el parámetro σ_u^2 no depende de ninguno de los parámetros $\boldsymbol{\beta}$, entonces escoger el vector de parámetros $\hat{\boldsymbol{\beta}}$ que maximice la función de verosimilitud (o su logaritmo) es equivalente a escoger el vector $\hat{\boldsymbol{\beta}}$ que minimice la suma residual $\text{SR}(\hat{\boldsymbol{\beta}})$. En consecuencia, al igual que en el modelo lineal, si el término de error sigue una distribución de probabilidad Normal y si su varianza es independiente de las componentes del vector $\boldsymbol{\beta}$, entonces los estimadores de máxima verosimilitud y de mínimos cuadrados, si existen, coinciden.

Las condiciones necesarias para la maximización de la función de verosimilitud son:

$$\frac{\partial \ln L(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2)}{\partial \hat{\boldsymbol{\beta}}} = -\frac{1}{2\hat{\sigma}_u^2} \frac{\partial \text{SR}(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = \frac{1}{\hat{\sigma}_u^2} \sum_1^T (y_t - f(\mathbf{x}_t, \hat{\boldsymbol{\beta}})) \frac{\partial f_t(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = \mathbf{0}_k \quad (k \text{ ecuaciones})$$

$$\frac{\partial \ln L(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2)}{\partial \hat{\sigma}_u^2} = -\frac{T}{2\hat{\sigma}_u^2} + \frac{1}{2\hat{\sigma}_u^4} \sum_1^T (y_t - f(\mathbf{x}_t, \hat{\boldsymbol{\beta}}))^2 = 0 \quad (1 \text{ ecuación})$$

cuyas soluciones proporcionan las estimaciones de máxima verosimilitud del vector β y el parámetro σ_u^2 bajo la hipótesis de Normalidad. De nuevo, estas condiciones necesarias muestran que el estimador de máxima verosimilitud del vector β coincide, bajo los supuestos hechos al comienzo de esta sección, con el estimador de mínimos cuadrados, ya que el primer bloque de ecuaciones son las condiciones de ortogonalidad [11.7] del estimador MCNL. Finalmente, la última ecuación genera la estimación de máxima verosimilitud de σ_u^2 , una vez que se haya obtenido la estimación del vector β :

$$\hat{\sigma}_u^2 = \frac{\text{SR}(\hat{\beta})}{T}$$

expresión igual a la obtenida en el caso del modelo econométrico lineal.

11.3.b. Matriz de covarianzas

Como vimos en el Capítulo 2, la matriz de covarianzas del estimador de máxima verosimilitud puede aproximarse, para muestras grandes, por la inversa de la matriz de información. Para calcular dicha matriz es preciso obtener previamente las derivadas de segundo orden del logaritmo de la función de verosimilitud y calcular su esperanza matemática.

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\frac{1}{2\sigma_u^2} \frac{\partial^2 \text{SR}(\beta)}{\partial \beta \partial \beta'}$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_u^2} = \frac{1}{2\sigma_u^4} \frac{\partial \text{SR}(\beta)}{\partial \beta}$$

$$\frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2} = \frac{T}{2\sigma_u^4} - \frac{\text{SR}(\beta)}{\sigma_u^6}$$

donde $\frac{\partial \text{SR}(\beta)}{\partial \beta}$ y $\frac{\partial^2 \text{SR}(\beta)}{\partial \beta \partial \beta'}$ denotan, respectivamente, el gradiente de la suma residual, y su matriz hessiana, formada por las derivadas de segundo orden. En las ecuaciones anteriores cabe observar lo siguiente:

$$a) \quad E\left(\frac{\partial \text{SR}(\beta)}{\partial \beta}\right) = -2\Sigma_1^T E\left[(y_t - f(\mathbf{x}_t, \beta))\left(\frac{\partial f_t}{\partial \beta}\right)\right] = -2\Sigma_1^T E\left[u_t\left(\frac{\partial f_t}{\partial \beta}\right)\right] = \mathbf{0}_k,$$

puesto que las derivadas parciales que en esta expresión aparecen son funciones de las variables explicativas (deterministas) y de los coeficientes β (constantes desconocidas). Como consecuencia:

$$E\left(-\frac{\partial^2 \ln L(y_t, \mathbf{x}_t; \beta)}{\partial \beta \partial \sigma_u^2}\right) = \mathbf{0}_k$$

y la matriz de información resulta ser diagonal a bloques (un bloque superior de dimensión $k \times k$, correspondiente al vector β , y otro inferior 1×1 , correspondiente al parámetro σ_u^2) y, por tanto, las estimaciones de máxima verosimilitud del vector β y del parámetro σ_u^2 son independientes según crece el tamaño muestral.

b) La matriz $\frac{\partial^2 \text{SR}(\beta)}{\partial \beta \partial \beta'}$ es igual a $2 \Sigma_1^T \left(\frac{\partial f_t}{\partial \beta} \right) \left(\frac{\partial f_t}{\partial \beta} \right)' - 2 \Sigma_1^T u_t \frac{\partial^2 f_t}{\partial \beta \partial \beta'}$, en la que ambos sumandos son matrices $k \times k$. El primer sumando es el producto del gradiente de la función $f(x_t, \beta)$, dispuesto como vector columna, por sí mismo, esta vez ordenado como vector fila. El segundo sumando es la matriz de derivadas segundas de la función $f(x_t, \beta)$, multiplicada por el término de error de dicho período, u_t . Ambas son matrices simétricas y, por tanto, también $\frac{\partial^2 \text{SR}(\beta)}{\partial \beta \partial \beta'}$ lo es. El elemento genérico (i, j) de la matriz $\frac{\partial^2 \text{SR}(\beta)}{\partial \beta \partial \beta'}$ es:

$$2 \left[\Sigma_1^T \frac{\partial f_t}{\partial \beta_i} \frac{\partial f_t}{\partial \beta_j} - \Sigma_1^T \frac{\partial^2 f_t}{\partial \beta_i \partial \beta_j} u_t \right]$$

y como $E \left[\frac{\partial^2 f_t}{\partial \beta_i \partial \beta_j} u_t \right] = 0$, se tiene:

$$E \left[- \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right] = \frac{1}{\sigma_u^2} \Sigma_1^T \frac{\partial f_t}{\partial \beta_i} \frac{\partial f_t}{\partial \beta_j}$$

c) Finalmente:

$$E \left(- \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2} \right) = - \frac{T}{2\sigma_u^4} + \frac{1}{\sigma_u^6} E[(\Sigma_1^T u_t^2)] = - \frac{T}{2\sigma_u^4} + \frac{T}{\sigma_u^6} \sigma_u^2 = \frac{T}{2\sigma_u^4}$$

Por tanto, la matriz de información es:

$$I(\beta, \sigma_u^2) = \begin{pmatrix} \frac{1}{\sigma_u^2} \left[\left(\frac{\partial f}{\partial \beta} \right)' \left(\frac{\partial f}{\partial \beta} \right) \right] & \mathbf{0}_k \\ \mathbf{0}_k & \frac{T}{2\sigma_u^4} \end{pmatrix}$$

Y, si invertimos y sustituimos los parámetros desconocidos por sus estimaciones, obtendremos la matriz de covarianzas del estimador de máxima verosimilitud en el caso de que el término de error siga una distribución de probabilidad Normal y que el tamaño de la muestra sea suficientemente grande:

$$\text{Var}(\hat{\beta}, \hat{\sigma}_u^2) = \begin{pmatrix} \sigma_u^2 \left[\left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta})}{\partial \beta} \right) \right]^{-1} & \mathbf{0}_k \\ \mathbf{0}_k & \frac{2\hat{\sigma}_u^4}{T} \end{pmatrix}$$

siempre que la matriz $\left(\frac{\partial f}{\partial \beta} \right)' \left(\frac{\partial f}{\partial \beta} \right)'$ sea no singular.

Este resultado, junto con la equivalencia entre el estimador de máxima verosimilitud y el de mínimos cuadrados no lineales, justifica la distribución Normal asintótica que para este último estimador propusimos en la Sección 11.2. En la práctica sólo se dispone de muestras finitas, por lo que la matriz anterior es sólo una aproximación a dicha matriz de covarianzas.

Ejemplo. Volviendo al Ejemplo 11.1, la matriz de covarianzas del estimador $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_u^2)_{\text{MV}}$ es:

$$\sigma_u^2 \begin{pmatrix} \left(\begin{array}{cc} \sum_1^T e^{2\beta x_t} & \alpha \sum_1^T x_t e^{2\beta x_t} \\ \alpha \sum_1^T x_t e^{2\beta x_t} & \alpha^2 \sum_1^T x_t^2 e^{2\beta x_t} \end{array} \right)^{-1} & 0 \\ 0 & 0 \\ 0 & 0 & \frac{2\sigma_u^4}{T} \end{pmatrix}$$

evaluada en las estimaciones máximo-verosímiles. La submatriz 2×2 superior coincide con la expresión analítica de $\text{Var}(\hat{\alpha}, \hat{\beta})$ que vimos antes.

También existe la posibilidad, por supuesto, de que la matriz de covarianzas del vector \mathbf{u} no sea diagonal, sino una matriz Σ simétrica, definida positiva cualquiera. En tal caso, el logaritmo de la función de verosimilitud es:

$$\ln L(y_t, x_t; \beta, \Sigma) = \text{constante} - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} [\mathbf{y} - f(\mathbf{X}, \beta)]' \Sigma^{-1} [\mathbf{y} - f(\mathbf{X}, \beta)]$$

y, por tanto, la condición necesaria de optimalidad que surge al tomar derivadas con respecto a los elementos de la matriz Σ^{-1} es:

$$\frac{\partial \ln L}{\partial \Sigma^{-1}} = \frac{1}{2} \Sigma - \frac{1}{2} [\mathbf{y} - f(\mathbf{X}, \beta)][\mathbf{y} - f(\mathbf{X}, \beta)]' = \mathbf{0}_{k \times k}$$

que tiene por solución:

$$\hat{\Sigma}_{\text{MV}} = [\mathbf{y} - f(\mathbf{X}, \beta)][\mathbf{y} - f(\mathbf{X}, \beta)]'$$

donde debe notarse que se utiliza la serie de residuos para estimar una matriz $T \times T$.

De este modo, habría que poner en práctica un procedimiento iterativo: Dada una estimación $\hat{\beta}$ del vector de coeficientes del modelo, pueden obtener-

se los residuos, que serán utilizados en la estimación de la matriz Σ . Una vez estimada esta matriz, debe maximizarse la función de verosimilitud, lo que equivale a minimizar $[y - f(\mathbf{X}, \boldsymbol{\beta})] \Sigma^{-1} [y - f(\mathbf{X}, \boldsymbol{\beta})]'$ con respecto al vector $\boldsymbol{\beta}$, y así sucesivamente. Sin embargo, este problema de minimización dista de ser trivial, y debe resolverse mediante algoritmos numéricos en un ordenador, como los que presentamos en el próximo capítulo.

Las propiedades estadísticas del estimador $\hat{\boldsymbol{\beta}}$ generado por este procedimiento son las mismas a partir de la primera iteración, y coinciden con las del estimador de máxima verosimilitud. En particular, dicho estimador es eficiente. Sin embargo, conviene iterar para tratar de minimizar el error muestral producido por la estimación de la matriz de covarianzas Σ .

Ejemplo 11.2. Si volvemos a la estimación de la función de consumo del Capítulo 7, podríamos considerar una versión algo más general de la relación entre el Consumo y el PIB:

$$C_t = \beta_1 + \beta_2 Y_t^\gamma + u_t \quad [11.11]$$

donde, si $\gamma = 1$, tenemos nuevamente el modelo lineal. Si, por el contrario, γ es un parámetro a estimar, entonces el modelo es no lineal. La función $f(x_t, \boldsymbol{\beta}) = \beta_1 + \beta_2 Y_t^\gamma$ tiene por gradiente:

$$\frac{\partial f(x_t, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (1; Y_t^\gamma; \beta_2 Y_t^\gamma \ln Y_t)$$

por lo que, de acuerdo con la linearización que efectuamos en la Sección 11.1:

$$C_t^* = C_t + \gamma \beta_2 Y_t^\gamma \ln Y_t$$

Se trata, por tanto, de dar valores iniciales a los parámetros β_1, β_2, γ , construir el vector gradiente así como C_t^* , y estimar una regresión de esta variable sobre las tres que integran el gradiente. En este modelo, $\hat{\beta}_1$ no juega ningún papel esencial en tal proceso de estimación.

Con los datos de la Tabla 7.1 obtuvimos el siguiente proceso iterativo:

Iteraciones	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}$	$\hat{\sigma}_u^2$
0	—	0,70	1,0	
1	1.676,5	-0,9767	1,2434	1.610.242 · 10 ⁴
2	1.377,2	0,07902	1,2448	3.120.269,6
3	1.377,7	0,07882	1,2262	38.633,0
4	1.376,8	0,08046	1,2221	16.288,6
5	1.376,7	0,08050	1,2210	16.288,4

con matriz de covarianzas:

$$\text{Var}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}) = \begin{pmatrix} 49.996,8 & & & \\ -7,83 & 0,13 \cdot 10^{-2} & & \\ & 9,60 & 0,16 \cdot 10^{-2} & \\ & & & 0,20 \cdot 10^{-2} \end{pmatrix}$$

Los coeficientes estimados en el modelo no lineal [11.11] no son comparables a los del modelo lineal del Capítulo 7. En particular, parece que nuestra estimación de $\hat{\beta}_2 = 0,080$ proporciona una propensión marginal al consumo muy reducida, aunque tal impresión es equivocada. En el modelo [11.11], la propensión marginal es $\frac{dC_t}{dY_t} = \beta_2 \gamma Y_t^{\gamma-1}$, que depende, por consiguiente, del valor de Y_t . Utilizando el valor muestral del año base: $Y_{1980} = 15.209,12$, se tiene:

$$\left(\frac{dC}{dY}\right)_{1988} = (0,080)(1,22)(15.209,1)^{0,22} = 0,81$$

11.4. TRANSFORMACION BOX-COX

La familia de transformaciones Box-Cox de la Sección 6.6 puede utilizarse para especificar una relación del tipo:

$$y_t = \alpha + \beta g(x_t) + u_t \quad [11.12]$$

donde:

$$g(x_t) = \frac{x_t^\lambda - 1}{\lambda}$$

Si $\lambda = 0$, entonces se tiene $g(x_t) = \log x_t$, mientras que si $\lambda = 1$, se tiene el modelo lineal; si $\lambda = -1$, entonces $g(x_t) = 1/x_t$. Si el valor del parámetro λ fuese conocido, entonces [11.12] se convierte en un modelo de regresión lineal simple; por el contrario, si λ es un parámetro a estimar, entonces se trata de un modelo intrínsecamente no lineal.

Un modo sencillo de tratar este modelo utiliza el hecho de que, fijado un valor de λ , el modelo se convierte en lineal tras un cambio de variables obvio. Por ello, podemos obtener diversas estimaciones de los coeficientes α y β condicionales en un valor dado del parámetro λ , que cabe esperar que se halle entre -1 y 1 . Aquel valor de λ para el que se obtenga una suma residual menor, junto con los valores de α y β asociados, constituyen el estimador de mínimos cuadrados del modelo [11.12]; bajo el supuesto de Normalidad de los residuos, serán también la estimación de máxima verosimilitud.

Sin embargo, puesto que se ha efectuado un proceso de búsqueda sobre el rango de valores admisibles del parámetro λ , la matriz de covarianzas del

par (α, β) que se obtiene en la regresión condicionada que se ha seleccionado como óptima no es válida y tenderá a subestimar las verdaderas varianzas.

Como alternativa, puede considerarse la estimación de máxima verosimilitud, suponiendo que $u_t \sim N(0, \sigma_u^2)$, teniéndose entonces:

$$\ln L = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_1^T u_t^2$$

El jacobiano de la transformación es:

$$\left| \frac{du}{dy} \right| = 1$$

de modo que se tiene la función de verosimilitud:

$$\ln L = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_1^T [y_t - \alpha - \beta g(x_t)]^2$$

y si denotamos por l_t el logaritmo de la verosimilitud del período t , tenemos el vector gradiente:

$$\frac{\partial l_t}{\partial \alpha} = \frac{u_t}{\sigma_u^2}$$

$$\frac{\partial l_t}{\partial \beta} = \frac{u_t}{\sigma_u^2} \cdot \frac{x_t^{\lambda-1}}{\lambda}$$

$$\frac{\partial l_t}{\partial \lambda} = \beta \frac{u_t}{\sigma_u^2} \frac{\partial g(x_t)}{\partial \lambda} = \beta \frac{u_t}{\sigma_u^2} \frac{\lambda x_t^{\lambda} \ln x_t - x_t^{\lambda} + 1}{\lambda^2}$$

$$\frac{\partial l_t}{\partial \sigma_u^2} = \frac{1}{2\sigma_u^2} \cdot \left(\frac{u_t^2}{\sigma_u^2} - 1 \right)$$

que habría que igualar a cero para tratar de obtener las estimaciones numéricas de MV de $\theta = (\alpha, \beta, \lambda, \sigma_u^2)$. Dicho sistema no tiene, sin embargo, solución analítica, y hay que recurrir a procedimientos numéricos como los que analizaremos en el próximo capítulo.

Multiplicando este vector gradiente por sí mismo y por una estimación de σ_u^2 :

$$\hat{\sigma}_u^2 = \frac{\sum_1^T [y_t - (\alpha + \beta g(x_t))]^2}{T}$$

se tiene una estimación de la matriz de covarianzas del vector $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\sigma}_u^2)$. Para ello habría que resolver previamente el sistema de condiciones de primer

orden: $\nabla \ln L(\hat{\theta}) = \mathbf{0}_4$, obtener los residuos \hat{u}_t y sustituirlos en la expresión mencionada de la matriz de covarianzas.

En algunas ocasiones es interesante someter a la variable endógena a la misma transformación que la exógena, considerando el modelo:

$$g(y_t) = \alpha + \beta g(x_t) + u_t$$

En tal caso, el jacobiano de la transformación de \mathbf{u} a \mathbf{y} es:

$$\left| \frac{d\mathbf{u}}{d\mathbf{y}} \right| = \prod_1^T y_t^{\lambda-1}$$

por lo que el logaritmo de la función de verosimilitud es:

$$\ln L = (\lambda - 1) \sum_1^T \ln y_t - \frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \cdot \sum_1^T (g(y_t) - \alpha - \beta g(x_t))^2 \quad [11.13]$$

Si se llevase a cabo un procedimiento de búsqueda sobre λ , habría que utilizar [11.13] como criterio a minimizar, y no simplemente $SR(\theta)$.

Si se pretende maximizar la función de verosimilitud directamente, se obtiene el gradiente, cuyos componentes $\frac{\partial l_t}{\partial \alpha}$, $\frac{\partial l_t}{\partial \beta}$ y $\frac{\partial l_t}{\partial \sigma_u^2}$ tienen la misma expresión analítica del modelo anterior. Sin embargo, ahora:

$$\frac{\partial l_t}{\partial \lambda} = \ln y_t - \frac{u_t}{\sigma_u^2} \cdot \left(\frac{\partial g(y_t)}{\partial \lambda} - \beta \frac{\partial g(x_t)}{\partial \lambda} \right)$$

donde:

$$\frac{\partial g(y_t)}{\partial \lambda} = \frac{\lambda y_t^\lambda \ln y_t - y_t^\lambda + 1}{\lambda^2} \quad ; \quad \frac{\partial g(x_t)}{\partial \lambda} = \frac{\lambda x_t^\lambda \ln x_t - x_t^\lambda + 1}{\lambda^2}$$

Nuevamente, el sistema de ecuaciones generado al igualar a cero el gradiente sólo tiene solución numérica, excepto por el estimador de σ_u^2 que, una vez calculadas las estimaciones de los demás parámetros, se obtendría por cociente entre la suma residual y el tamaño muestral. Sin embargo, las expresiones analíticas pueden utilizarse en alguno de los algoritmos numéricos que veremos en el próximo capítulo.

La matriz de covarianzas del estimador resultante se estima de modo similar al del caso anterior.

11.5. CONTRASTE DE RESTRICCIONES

11.5.a. Restricciones lineales

El contraste de un conjunto de *hipótesis lineales*: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, toma como elemento básico la *holgura* o *discrepancia*: $\mathbf{d} = \mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$, que se obtiene directamente a partir del estimador MCO, $\hat{\boldsymbol{\beta}}$. Incluso si las hipótesis que se contrastan fuesen ciertas, no debe esperarse que el vector discrepancia fuese exactamente igual a cero, al menos debido al error muestral. Por tanto, la tarea del investigador debe decidir si dicho vector discrepancia es suficientemente grande como para hacer imposible el mantenimiento de la hipótesis nula H_0 .

Como el vector *discrepancia* es una función lineal del estimador $\hat{\boldsymbol{\beta}}$, que tiene distribución Normal, también \mathbf{d} tendrá distribución Normal; bajo la hipótesis nula, es decir, si H_0 es cierta, se tiene:

$$E(\mathbf{d}) = E(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \mathbf{R}E(\hat{\boldsymbol{\beta}}) - \mathbf{r} = \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}_q$$

y:

$$\text{Var}(\mathbf{d}) = \text{Var}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \mathbf{R} \text{Var}(\hat{\boldsymbol{\beta}}) \mathbf{R}' = \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'$$

Estos resultados sugieren la realización del contraste de H_0 utilizando el criterio de Wald:

$$W = \mathbf{d}'[\text{Var}(\mathbf{d})]^{-1} \mathbf{d}$$

que utilizando la Sección 2.8 sobre formas cuadráticas de rango completo obtenidas de una variable Normal se distribuye como una χ_q^2 .

En la práctica se desconoce el valor de σ^2 , por lo que se divide W , que depende tan sólo de la estimación MCO del vector de coeficientes, por otra forma cuadrática que depende sólo de la estimación de σ ; como ambas estimaciones son independientes entre sí, el cociente de ambas formas cuadráticas se distribuye como una F .

11.5.b. Restricciones no lineales

El contraste de un conjunto de restricciones no lineales acerca de los coeficientes de un modelo de regresión: $H_0: \mathbf{R}(\boldsymbol{\beta}) = \mathbf{r}$, no presenta una diferencia conceptual importante con respecto al caso de restricciones lineales. Suponemos en lo sucesivo que la matriz $\partial \mathbf{R}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, de orden $q \times k$, con $q < k$, tiene rango igual a q , es decir, que tenemos menos restricciones que parámetros, y que las restricciones no son redundantes. El contraste se lleva a cabo nuevamente en función del tamaño del vector discrepancia, que evaluamos de igual modo a como hicimos en el caso de restricciones lineales: $\mathbf{d} = \mathbf{R}(\hat{\boldsymbol{\beta}}) - \mathbf{r}$. Se producen, sin embargo, algunas diferencias:

1. Al ser $\mathbf{R}(\cdot)$ una función no lineal (o un conjunto de ellas) ya no podemos afirmar que la esperanza matemática de $\mathbf{R}(\hat{\boldsymbol{\beta}})$ sea igual a $\mathbf{R}(E(\hat{\boldsymbol{\beta}}))$; afortunadamente, por la Proposición 2.12, podemos utilizar la consistencia del estimador MCO, $\hat{\boldsymbol{\beta}}$, para afirmar que $\text{plim } \mathbf{R}(\hat{\boldsymbol{\beta}}) = \mathbf{R}(\text{plim } \hat{\boldsymbol{\beta}})$.

2. Debido a la no linealidad, no puede mantenerse la distribución chi-cuadrado en muestras finitas para la forma cuadrática utilizada en la construcción de los estadísticos t o F que hemos venido utilizando hasta ahora.

Contraste F

Ya hemos dicho que la no linealidad del modelo impide justificar distribuciones t o F como las habituales. Si se estima el modelo por mínimos cuadrados mediante la aproximación lineal que hemos discutido en la Sección 11.1, la distribución del estadístico

$$F = \frac{[\text{SR}(\hat{\boldsymbol{\beta}}_R) - \text{SR}(\hat{\boldsymbol{\beta}})]/q}{\text{SR}(\hat{\boldsymbol{\beta}})/(T - k)}$$

no es conocida en muestras finitas, pero es una $F_{q, T-k}$ asintóticamente.

Contraste de Wald (W)

Una importante dificultad reside en el cálculo de la varianza de la diferencia $\mathbf{R}(\hat{\boldsymbol{\beta}}) - \mathbf{r}$, que es función no lineal del estimador $\hat{\boldsymbol{\beta}}$. Para calcularla, obtenemos la aproximación lineal de $\mathbf{R}(\hat{\boldsymbol{\beta}})$ alrededor del verdadero valor del vector de coeficientes $\boldsymbol{\beta}$:

$$\mathbf{R}(\hat{\boldsymbol{\beta}}) \cong \mathbf{R}(\boldsymbol{\beta}) + \left(\frac{\partial \mathbf{R}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

donde $\partial \mathbf{R}/\partial \boldsymbol{\beta}$ es una matriz $q \times k$, y la varianza de $\mathbf{R}(\hat{\boldsymbol{\beta}})$ puede aproximarse por la varianza del miembro derecho de la aproximación, es decir:

$$\text{Var}(\mathbf{R}(\hat{\boldsymbol{\beta}})) \cong \left(\frac{\partial \mathbf{R}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) \text{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \left(\frac{\partial \mathbf{R}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)'$$

donde todos los términos deben evaluarse en el estimador $\hat{\boldsymbol{\beta}}$. Como el verdadero vector $\boldsymbol{\beta}$ es constante —si bien desconocido—, se tiene:

$$\text{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

En definitiva, el estadístico $[\mathbf{R}(\hat{\boldsymbol{\beta}}) - \mathbf{r}] [\text{Var } \mathbf{R}(\hat{\boldsymbol{\beta}})]^{-1} [\mathbf{R}(\hat{\boldsymbol{\beta}}) - \mathbf{r}]$ sigue una distribución chi-cuadrado con un número de grados de libertad igual a q , el número de restricciones que se contrastan. Este es el estadístico de Wald, para cuyo cálculo sólo precisamos del estimador sin restringir, y que es asintóticamente equivalente a q veces el estadístico F .

Ejemplo 11.3. Consideremos nuevamente la contrastación de una propensión marginal al consumo unitaria a largo plazo, que ya analizamos en el Ejemplo 7.1. Así, se trata de contrastar $H_0: \beta_2/(1 - \beta_3) = 1$ en el modelo $C_t = \alpha_1 + \beta_2 Y_t + \beta_3 C_{t-1} + u_t$. De acuerdo con el análisis que acabamos de hacer, utilizaremos el resultado:

$$\left(\frac{\hat{\beta}_2}{1 - \hat{\beta}_3} - 1 \right)^2 \left(0; \frac{1}{1 - \hat{\beta}_3}; \frac{\hat{\beta}_2}{(1 - \hat{\beta}_3)^2} \right)' [\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}] \begin{pmatrix} 0 \\ 1 \\ \frac{1}{1 - \hat{\beta}_3} \\ \hat{\beta}_2 \\ \frac{1}{(1 - \hat{\beta}_3)^2} \end{pmatrix} \sim \chi_a^2$$

Contraste de razón de verosimilitudes (RV)

El logaritmo de la función de verosimilitud es:

$$\ln L = \text{jacobiano} - \frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_1^T u_t^2$$

y puesto que el estimador MV de σ^2 es $\hat{\sigma}^2 = \mathbf{SR}/T = (1/T) \sum_1^T \hat{u}_t^2$, se tiene la función de verosimilitud concentrada:

$$\ln L(\hat{\boldsymbol{\beta}}) = -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \hat{\sigma}^2 - \frac{T}{2}$$

que implica, si denotamos por L_R el valor de la función de verosimilitud, evaluada en el estimador restringido $\hat{\boldsymbol{\beta}}_R$:

$$\ln L_R - \ln L = \frac{T}{2} (\ln \hat{\sigma}^2 - \ln \hat{\sigma}_R^2) = \frac{T}{2} \ln \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_R^2} \right)$$

y el estadístico

$$-2(\ln L_R - \ln L) = T \ln \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_R^2} \right)$$

se distribuye asintóticamente como una variable aleatoria χ_a^2 .

La utilización de este estadístico de razón de verosimilitudes (RV) constituye, por tanto, un procedimiento alternativo de llevar a cabo el contraste de las hipótesis $H_0: \mathbf{R}(\boldsymbol{\beta}) = \mathbf{r}$, si bien requiere calcular tanto el estimador restringido como el estimador sin restringir.

En ocasiones existe la inclinación a utilizar el estadístico RV para el contraste de una hipótesis nula *simple* frente a una alternativa también *simple*; sin embargo, ello no está justificado teóricamente: los grados de libertad del estadístico RV vienen determinados por la reducción que, en la dimensión del espacio paramétrico, se consigue con la imposición de las restricciones. En un contraste de hipótesis nula *simple* frente a *simple*, dicha reducción es cero.

Contraste de los multiplicadores de Lagrange (ML)

La suma residual de un modelo restringido excede siempre a la del mismo modelo sin restringir, no importa cuáles sean el modelo y las restricciones. Esta es la base del contraste de los multiplicadores de Lagrange (ML) que se basa, precisamente, en la disminución que se produciría en el valor numérico de la suma residual si se abandonan las restricciones.

Supongamos que maximizamos la función de verosimilitud $L(\boldsymbol{\theta})$ sujeta a las restricciones $\mathbf{R}(\boldsymbol{\beta}) = \mathbf{r}$, donde el vector $\boldsymbol{\theta}$ incluye, junto a los coeficientes $\boldsymbol{\beta}$, los parámetros de la matriz de covarianzas del modelo. Si denotamos por λ el vector $q \times 1$ de multiplicadores de Lagrange para este problema de optimización restringida y por $L(\boldsymbol{\theta}_R)$ y $L(\boldsymbol{\theta})$ las funciones de verosimilitud restringida y sin restringir, el lagrangiano es:

$$\ln L(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) + \lambda'[\mathbf{R}(\boldsymbol{\beta}) - \mathbf{r}]$$

y las condiciones necesarias de maximización son:

$$\begin{aligned} \frac{\partial \ln L_R}{\partial \boldsymbol{\theta}_R} &= \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \mathbf{R}'\lambda = \mathbf{0}_k \\ \frac{\partial \ln L_R}{\partial \lambda} &= \mathbf{0}_q \end{aligned}$$

Si las restricciones son ciertas, su imposición apenas reducirá el valor máximo de la función de verosimilitud, por lo que las derivadas de $\ln L_R$ y $\ln L$ que aparecen en el primer bloque de ecuaciones deberían ser muy similares. Por consiguiente, el vector $\mathbf{R}'\lambda$ y, en particular, el valor óptimo de los multiplicadores de Lagrange debería ser muy reducido, es decir:

$$\frac{\partial \ln L(\hat{\boldsymbol{\theta}}_R)}{\partial \hat{\boldsymbol{\theta}}} \cong \frac{\partial \ln L(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = -\mathbf{R}'\lambda \cong \mathbf{0}_k$$

El mismo razonamiento utilizado al presentar el contraste de Wald sugiere el uso de la forma cuadrática:

$$\text{LM} = \left(\frac{\partial \ln L(\hat{\boldsymbol{\theta}}_R)}{\partial \hat{\boldsymbol{\theta}}_R} \right)' [\mathbf{I}(\hat{\boldsymbol{\theta}}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\boldsymbol{\theta}}_R)}{\partial \hat{\boldsymbol{\theta}}_R} \right)$$

Ejemplo. En el modelo de regresión no lineal, Judge y otros (1985) muestran que si denotamos los residuos obtenidos a partir de los estimadores restringidos: $\tilde{u}_i = y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_R)$, y por $\tilde{\mathbf{X}}$ la matriz de observaciones del gradiente de la función f , evaluado en los estimadores restringidos, se tiene:

$$\text{LM} = \frac{\tilde{\mathbf{u}}' \tilde{\mathbf{X}}' (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{u}}}{(\tilde{\mathbf{u}}' \tilde{\mathbf{u}})/T} = T\tilde{R}^2$$

se distribuye como una χ_q^2 . Por \tilde{R}^2 denotamos el coeficiente de determinación en la regresión del vector de residuos \tilde{u} sobre la matriz de variables explicativas \tilde{X} . Esta formulación del contraste ML es muy útil en el trabajo aplicado, siendo el modo en que más frecuentemente se lleva a cabo el contraste.

El contraste ML sólo precisa del estimador restringido, lo que lo hace especialmente interesante en aquellos casos en que la imposición de las restricciones conduce a un modelo más sencillo que aquel de que se partía.

La Figura 11.1 ilustra la lógica conceptual de cada uno de los tres procedimientos de contrastación de hipótesis que acabamos de presentar: El test de Wald se basa en que si H_0 es cierta, entonces el vector discrepancia evaluado en $\hat{\beta}_{MV}$ debería ser próximo a cero, puesto que el estimador MV es consistente. Así rechazamos la hipótesis nula si el vector discrepancia, evaluado en el estimador MV: $R(\hat{\beta}_{MV}) - r$, es grande. El contraste de razón de verosimilitudes se fundamenta en que si las restricciones que se contrastan fuesen ciertas, entonces su imposición no implicaría una reducción significativa en el valor máximo de la función de verosimilitud, por lo que la diferencia $\ln L - \ln L_R$, que es un escalar, debe ser pequeña.

Por último, el contraste de los multiplicadores de Lagrange se basa en que si las restricciones que se contrastan son válidas, el estimador restringido

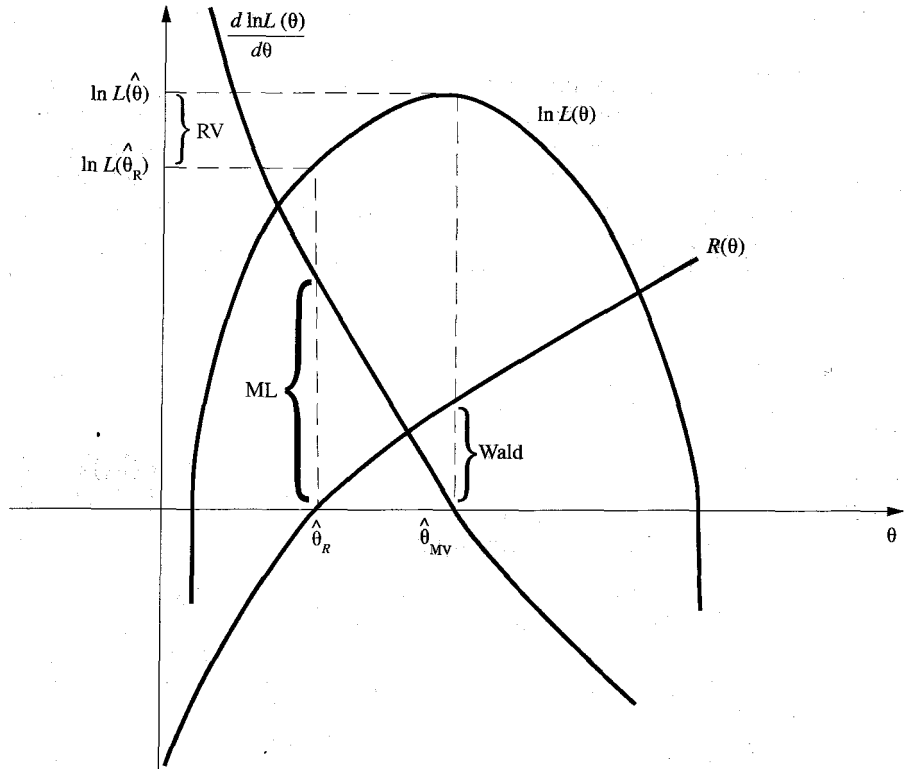


FIGURA 11.1.

debería estar próximo al punto que maximiza globalmente la función de verosimilitud: en consecuencia, la pendiente de la función de verosimilitud en el estimador restringido debería ser próxima a cero; así, este contraste se basa en el valor numérico de dicha pendiente.

Aunque los tres contrastes son equivalentes asintóticamente, se comportan de modo diferente en pequeñas muestras. Lamentablemente, su comportamiento depende de cada aplicación particular, lo que impide elaborar una teoría general acerca de sus interrelaciones. Por ello, la elección entre ellos se efectúa generalmente en base a su sencillez de computación en cada caso particular. En algunos casos particulares, como ocurre con el modelo lineal general, puede probarse la relación $ML \leq RV \leq W$, de manera que si un contraste ML rechaza la hipótesis nula, o si un test de Wald no rechaza H_0 , no se producirá contradicción entre los tres estadísticos.

Ejemplo 11.4. Consideremos nuevamente el modelo de consumo del Ejemplo 11.2:

$$C_t = \alpha + \beta Y_t^\gamma + u_t, \quad \gamma > 0$$

en el que queremos contrastar si la propensión marginal al consumo es independiente de la renta, lo que haría lineal al modelo: $H_0: \gamma = 1$.

El contraste *F habitual* se basa en las sumas residuales de la estimación mínimo-cuadrática del modelo sin restringir (Ejemplo 11.2), y del modelo restringido (Ejemplo 7.6.a). El estadístico:

$$\frac{(SRR - SRS)/q}{SRS/gdl(MSR)} = \frac{[1.079.364,7 - (16.288,4) \cdot 35]/1}{(16.288,4) \cdot 35/(35 - 3)} = 28,59$$

que se distribuye como una $F_{1,32}$.

El estadístico de Wald se basa en la discrepancia del estimador restringido:

$$W = \frac{(1,2221 - 1,0)^2}{(0,0443)^2} = 25,14$$

que se distribuye como una χ_1^2 .

El estadístico RV:

$$RV = -2(\ln L_R - \ln L) = T(\ln \hat{\sigma}_R^2 - \ln \hat{\sigma}^2) = 35(10,3365 - 6,1429) = 146,8$$

que se distribuye como una χ_1^2 .

Por último, para llevar a cabo el contraste ML, comenzamos estimando α y β en el modelo restringido: $C_t = \alpha + \beta Y_t + v_t$, para obtener los residuos restringidos. El gradiente de la función f que define el modelo anterior cuando se escribe en la forma $C_t = f(Y_t, v_t; \alpha, \beta, \gamma)$ es:

$$\frac{\partial f}{\partial \alpha} = 1; \quad \frac{\partial f}{\partial \beta} = Y_t; \quad \frac{\partial f}{\partial \gamma} = \beta Y_t^\gamma \ln Y_t$$

que, evaluado bajo la restricción $\gamma = 1$, resulta $(1, Y_t, z_t)$, donde $z_t = \beta Y_t \ln(Y_t)$, en el que β debe substituirse por su estimación MCO.

En una segunda etapa estimamos una regresión de los residuos previos sobre este vector de variables, comparando el producto $TR^2 = 35(0,474826) = 16,62$ con una distribución χ^2_1 .

Ejemplo 11.5. Volviendo al modelo dinámico de consumo del Capítulo 9, queremos contrastar la hipótesis de que la propensión marginal al consumo a largo plazo es igual a 1, es decir, en el modelo

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + u_t$$

queremos contrastar:

$$H_0: \frac{\beta_2}{1 - \beta_3} = 1$$

que ya tratamos en el Ejemplo 11.3.

Aunque ya vimos en dicho capítulo que, en realidad, esta restricción puede escribirse fácilmente como lineal, vamos a tratarla ahora como no lineal, con:

$$\mathbf{R}(\boldsymbol{\beta}) = \frac{\beta_2}{1 - \beta_3} \quad \text{y} \quad \mathbf{r} = 1$$

El gradiente es:

$$\frac{\partial \mathbf{R}}{\partial \boldsymbol{\beta}} = \left(\frac{1}{1 - \beta_3}, \frac{\beta_2}{(1 - \beta_3)^2} \right)$$

por lo que utilizando los resultados de la estimación del Capítulo 9 tenemos:

$$\frac{\partial \mathbf{R}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = (1,786; 1,371)$$

$$\begin{aligned} \text{Var}[\mathbf{R}(\hat{\boldsymbol{\beta}})] &= 1,786^2(0,1592 \cdot 10^{-2}) + 1,371^2(0,2742 \cdot 10^{-2}) - \\ &\quad - 2(1,786)(1,371)(0,2079 \cdot 10^{-2}) = 0,51 \cdot 10^{-4} \end{aligned}$$

cuya raíz cuadrada es $0,71 \cdot 10^{-2}$. Como

$$\mathbf{R}(\hat{\boldsymbol{\theta}}) - \mathbf{r} = \frac{0,43}{1 - 0,44} - 1 = -0,232$$

tenemos:

$$\frac{\mathbf{R}(\hat{\boldsymbol{\beta}}) - \mathbf{r}}{\sqrt{\text{Var}[\mathbf{R}(\hat{\boldsymbol{\beta}})]}} = \frac{-0,232}{0,71 \cdot 10^{-2}} = -32,68$$

que, comparada con una $N(0, 1)$, conduce a rechazar la hipótesis nula.

Sin embargo, en este caso parece razonable que la hipótesis nula hubiese sido:

$$H_0: \frac{\beta_2}{1 - \beta_3} < 1$$

y el contraste de una cola, en cuyo caso los valores críticos de la $N(0, 1)$ al 95 y 99 por 100 son de 1,645 y 2,36.

En este modelo es de sumo interés el valor de la ganancia g , en su caso, elasticidad a largo plazo: $g = \frac{\beta_2}{(1 - \beta_3)}$, y que podemos estimar a partir de $\hat{\beta}_2$ y $\hat{\beta}_3$. Utilizando los resultados del Capítulo 2, su varianza puede aproximarse por:

$$\widehat{\text{Var}}(\hat{g}) = \frac{1}{(1 - \hat{\beta}_3)^2} \cdot \text{Var}(\hat{\beta}_2) + \frac{\hat{\beta}_2^2}{(1 - \hat{\beta}_3)^4} \cdot \text{Var}(\hat{\beta}_3) + 2 \frac{\hat{\beta}_2}{(1 - \hat{\beta}_3)^3} \text{Cov}(\hat{\beta}_2, \hat{\beta}_3)$$

CAPITULO 12

ALGORITMOS NUMERICOS DE OPTIMIZACION

12.1. LA ESTIMACION DE MODELOS ECONOMETRICOS COMO SOLUCION A UN PROBLEMA DE OPTIMIZACION

En este capítulo analizamos diversos procedimientos para la obtención de soluciones numéricas a problemas de optimización que surgen en la estimación de modelos econométricos. En los problemas de optimización que vamos a tratar suponemos que tanto la función objetivo como las restricciones son funciones, varias veces diferenciables, de los parámetros que se pretenden estimar. Teóricamente, un problema de optimización se resuelve a partir de su lagrangiano, que se obtiene sumando a la función objetivo una combinación lineal de unas variables ficticias (multiplicadores de Lagrange) por las restricciones del problema. Las condiciones necesarias de optimalidad (o condiciones de primer orden) consisten en igualar a cero las derivadas parciales del lagrangiano tanto con respecto a los parámetros a estimar como a los multiplicadores de Lagrange⁽¹⁾. Un *punto estacionario* es una solución a este sistema de condiciones necesarias de optimalidad local.

Se dice que un problema de optimización tiene estructura lineal-cuadrática, cuando la función objetivo es un polinomio, a lo sumo de grado dos, en los parámetros a estimar, mientras que las posibles restricciones del problema son todas funciones lineales de los parámetros. Estos problemas se caracterizan porque sus condiciones necesarias de optimalidad local son lineales en los parámetros a estimar, por lo que generalmente pueden obtenerse analíticamente sus valores numéricos en un punto estacionario de la función objetivo. El carácter de la matriz de segundas derivadas de la función objetivo, evaluadas en un punto estacionario, determina el carácter de dicho punto (máximo local, mínimo local, o punto de silla).

⁽¹⁾ Para ser rigurosos, tales derivadas, que configuran el vector gradiente del lagrangiano, son iguales a cero en el óptimo sólo si éste es interior, es decir, si ningún parámetro del modelo alcanza la frontera de su soporte.

Un problema de estimación econométrica es, generalmente, un problema de optimización, en el que se trata de optimizar una función objetivo (por ejemplo, minimizar la suma de cuadrados de los residuos o maximizar la función de verosimilitud), posiblemente sujeta a unas restricciones. Sin embargo, rara vez adoptan estos problemas una estructura lineal-cuadrática; en tales circunstancias, las condiciones necesarias de optimalidad local no son lineales, y no es sencillo obtener los valores numéricos que deben tomar los parámetros del modelo para optimizar la función objetivo. Al contrario, en un gran número de casos de interés, es *imposible* obtener analíticamente los valores numéricos de dichas variables a partir de las condiciones de primer orden. En tales casos, la solución al problema de optimización no se obtiene resolviendo el sistema formado por las condiciones de optimalidad, haciéndose preciso utilizar procedimientos numéricos del tipo de los que presentamos en este capítulo.

De igual modo, prácticamente todos los problemas de estimación que consideramos en este texto pueden formalizarse como problemas de optimización *sin restricciones*, aunque las posibles restricciones podrían acomodarse en el lagrangiano correspondiente al problema de estimación, sin mayor dificultad.

Supondremos en lo sucesivo que, con objeto de estimar un modelo econométrico, nos vemos obligados a optimizar el valor numérico de una función $F(\theta)$, donde θ es un vector de los k parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ que determinan la suma de cuadrados que se pretende minimizar, o la función de verosimilitud que se pretende maximizar; ello incluye tanto los coeficientes del modelo como los parámetros que determinan la matriz de covarianzas del término de error.

En lo sucesivo, vamos a suponer que se trata de *minimizar* una función $F(\theta)$. Si se tratase de *maximizar* el valor numérico de $F(\theta)$, dicho problema es equivalente a minimizar la función $F^*(\theta) = -F(\theta)$, por lo que todo lo que decimos en este capítulo en relación con la minimización de funciones sería aplicable a la función F^* .

Describiremos distintos algoritmos de optimización (minimización, en este caso) y discutiremos sus ventajas e inconvenientes. Algunos de ellos precisan utilizar las expresiones analíticas del vector gradiente:

$$\nabla F(\theta) = (\partial F/\partial\theta_1, \partial F/\partial\theta_2, \dots, \partial F/\partial\theta_k)$$

y, en ocasiones, de la matriz hessiana $\nabla^2 F(\theta)$:

$$\nabla^2 F(\theta) = \begin{pmatrix} \frac{\partial^2 F}{\partial\theta_1^2} & \frac{\partial^2 F}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 F}{\partial\theta_1\partial\theta_3} & \dots & \frac{\partial^2 F}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 F}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 F}{\partial\theta_2^2} & \frac{\partial^2 F}{\partial\theta_2\partial\theta_3} & \dots & \frac{\partial^2 F}{\partial\theta_2\partial\theta_k} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\partial^2 F}{\partial\theta_k\partial\theta_1} & \frac{\partial^2 F}{\partial\theta_k\partial\theta_2} & \frac{\partial^2 F}{\partial\theta_k\partial\theta_3} & \dots & \frac{\partial^2 F}{\partial\theta_k^2} \end{pmatrix}$$

Estas expresiones analíticas pueden resultar difíciles, e incluso imposibles de obtener, en algunos casos concretos, por lo que es importante subrayar que pueden sustituirse por sus aproximaciones numéricas. En efecto, si la función $F(\theta)$ es derivable en el punto $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)'$, podemos utilizar

$$\frac{\partial F}{\partial \theta_j} = \frac{F(\theta^0 + \varepsilon e_j) - F(\theta^0)}{\varepsilon}$$

como aproximación a cada una de las coordenadas del vector gradiente de F , y

$$\frac{\partial^2 F}{\partial \theta_j \partial \theta_i} = \frac{\frac{\partial F}{\partial \theta_j}(\theta^0 + \varepsilon e_i) - \frac{\partial F}{\partial \theta_j}(\theta^0)}{\varepsilon}$$

como aproximación al elemento genérico (i, j) de la matriz hessiana de la función F , donde e_i denota un vector cuya única componente no nula es la i -ésima, que es igual a uno, y ε denota una cantidad positiva, infinitesimal, de modo que:

$$\theta^0 + \varepsilon e_i = (\theta_1^0, \theta_2^0, \dots, \theta_{i-1}^0, \theta_i^0 + \varepsilon, \theta_{i+1}^0, \dots, \theta_k^0)$$

Si el vector gradiente de la función F , ∇F , así como la matriz formada por sus derivadas de segundo orden, $\nabla^2 F$, existen y son continuas en un entorno de la solución θ_0 , entonces un conjunto de condiciones suficientes para que θ_0 sea un *mínimo local* de $F(\theta)$ son: a) $\nabla F(\theta_0) = \mathbf{0}_k$, y b) la matriz de derivadas segundas de F , evaluada en θ_0 , $\nabla^2 F(\theta_0)$ es definida positiva, es decir, la función F es estrictamente convexa en dicho punto del espacio paramétrico. Sin embargo, estas condiciones no garantizan que el vector θ_0 sea un mínimo global del problema de optimización, pues podría existir otro vector θ^* que cumpliera asimismo las condiciones a) y b), además de $F(\theta^*) < F(\theta_0)$; ello se debe a que las condiciones a) y b) se satisfacen en todos los *mínimos locales* de la función $F(\theta)$. Por el contrario, si se supiese que la función F es convexa en todo su dominio de definición, la condición a) sería necesaria y suficiente para la existencia de un mínimo de dicha función.

En las secciones siguientes examinamos los algoritmos habitualmente utilizados para resolver problemas de optimización y, por tanto, de estimación econométrica.

12.2. ALGORITMO DE BUSQUEDA

Este algoritmo es aplicable cuando k , el número de parámetros a estimar, es pequeño (uno o dos) y el rango de sus valores admisibles está acotado. Supongamos, por ejemplo, que estamos interesados en la estimación de un modelo con un solo coeficiente: $y_t = \beta x_t + u_t$, y que los valores admisibles de β están en el intervalo $(-1, 1)$. El algoritmo de búsqueda consiste en:

1. Construir una partición de dicho intervalo, por ejemplo: $-1,0$; $-0,90$; $-0,80$; ...; $0,80$; $0,90$; $1,0$.
2. Evaluar la función $F(\beta)$ en cada uno de los puntos de la partición.
3. Elegir como estimador de β aquel punto que proporciona un valor numérico más pequeño de la función $F(\beta)$.

No es muy frecuente que el investigador esté interesado tan sólo en uno o dos parámetros; sin embargo, en muchas ocasiones la estimación de un modelo econométrico resulta más sencilla si llevamos a cabo un procedimiento de búsqueda sobre un subvector de uno o dos coeficientes y efectuamos estimaciones condicionadas de los demás parámetros como en el siguiente ejemplo:

Ejemplo 12.1. *Procedimiento de Hildreth-Lu para la estimación de un modelo con autocorrelación.*

El procedimiento de Hildreth-Lu para la estimación, por mínimos cuadrados, del modelo de regresión con estructura de correlación autorregresiva de orden 1:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t \quad [12.1.a]$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad [12.1.b]$$

consiste en construir una partición del intervalo $(-1,1)$ para el parámetro ρ [o $(0,1)$ si se acepta a priori que sea positivo]. A continuación, cada uno de los valores de dicha partición se utiliza para filtrar las variables y_t y x_t :

$$y_t^* = y_t - \rho y_{t-1}; \quad \mathbf{x}_t^* = \mathbf{x}_t - \rho \mathbf{x}_{t-1}$$

y estas variables transformadas se utilizan en la estimación mínimo-cuadrática del modelo. Vimos en el Capítulo 7 que el valor numérico de ρ y los correspondientes valores del vector $\boldsymbol{\beta}$ que generen la menor suma residual: $F(\boldsymbol{\beta}/\rho) = \Sigma(y_t^* - \mathbf{x}_t^{*'} \boldsymbol{\beta})^2$, coinciden con el estimador por máxima verosimilitud del modelo.

Para llevar a cabo una estimación más precisa, podría construirse una partición alrededor del valor de ρ que minimizó la función $F(\boldsymbol{\beta}/\rho)$ en una primera aproximación y repetir el procedimiento. Sin embargo, este sencillo algoritmo está sujeto al posible error de que el valor de ρ que verdaderamente minimiza $F(\boldsymbol{\beta}/\rho)$ hubiese quedado lejos del punto que se escogió en la primera iteración. Si ello ocurriese, no llegaríamos a aproximarnos por este método a dicho mínimo absoluto.

Aunque el mismo procedimiento funcionaría si el intervalo no fuese $(-1,1)$, sin embargo es preciso que el conjunto de valores admisibles del parámetro sobre el que se lleva a cabo la búsqueda sea acotado. El método podría utilizarse de modo similar si fuesen dos los parámetros en los que buscar, siempre que ambos tuviesen un rango acotado de valores admisibles. Supongamos que la estructura de correlación fuese de orden 2:

$$y_t = \mathbf{x}_t^* \boldsymbol{\beta} + u_t \quad [12.2.a]$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \quad [12.2.b]$$

y tratamos de minimizar la suma residual $F(\boldsymbol{\beta}/\rho_1, \rho_2) = \sum (y_t^* - \mathbf{x}_t^* \boldsymbol{\beta})^2$, donde el filtro sería ahora, lógicamente, 12.2.b. El número de evaluaciones de la función $F(\boldsymbol{\beta}/\rho_1, \rho_2)$ que habría que llevar a cabo aumenta considerablemente con respecto al caso autorregresivo de orden 1. En efecto, ahora, para cada valor de ρ_2 , habría que evaluar la función F tantas veces como elementos hayamos considerado en la partición del rango admisible de ρ_1 .

12.3. ALGORITMO DEL DESCENSO MAS RAPIDO

Una función F de k variables representa una superficie en \mathbf{R}^k ; el gradiente de la función (un vector de dimensión k) en un punto proporciona la dirección perpendicular al plano tangente a la superficie en dicho punto. Si nos desplazamos de un punto P a otro P' , el valor numérico de la función cambiará: pues bien, el gradiente de F , evaluado en el punto P , representa la dirección de máxima variación de la función. Para aumentar el valor de F , basta desplazarse una distancia adecuada en la dirección del gradiente, moviéndonos en dirección opuesta al gradiente si queremos disminuir el valor de F , como es nuestra pretensión.

Así, una estrategia posible para tratar de minimizar el valor de la función $F(\boldsymbol{\theta})$ consiste en desplazarnos de un vector inicial $\boldsymbol{\theta}_0$ a otro $\boldsymbol{\theta}_1$, de acuerdo con la expresión:

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 - \lambda \nabla F(\hat{\boldsymbol{\theta}}_0), \quad \lambda > 0$$

donde la elección del parámetro $\lambda > 0$, que se conoce como *longitud de paso*, es crucial para reducir efectivamente el valor de $F(\boldsymbol{\theta})$. En efecto, si el valor de λ fuese excesivamente grande, entonces pudiera ser que $\hat{F}(\hat{\boldsymbol{\theta}}_1) > \hat{F}(\hat{\boldsymbol{\theta}}_0)$, lo cual implicaría que el algoritmo podría no converger. En el caso de maximización de una función que se ilustra en la Figura 12.1, partiendo del punto A , y moviéndose en la dirección correcta (hacia la izquierda), podríamos pasar al punto B , del que bien podríamos dirigirnos de nuevo hacia el máximo global P , encaminarnos hacia el máximo local, Q o, en general, hacia cualquier otro punto.

12.4. ALGORITMO DE NEWTON-RAPHSON

Supongamos que disponemos de una estimación $\hat{\boldsymbol{\theta}}_n$ del mínimo $\hat{\boldsymbol{\theta}}$ de una función continuamente diferenciable $F(\boldsymbol{\theta})$. Si consideramos un entorno pequeño del punto $\hat{\boldsymbol{\theta}}_n$, el valor numérico de F en un punto de dicho entorno puede aproximarse mediante un desarrollo en serie de Taylor de orden 2:

$$F(\boldsymbol{\theta}) \cong M(\boldsymbol{\theta}) = F(\hat{\boldsymbol{\theta}}_n) + [\nabla F(\hat{\boldsymbol{\theta}}_n)]' (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' [\nabla^2 F(\hat{\boldsymbol{\theta}}_n)] (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) \quad [12.3]$$

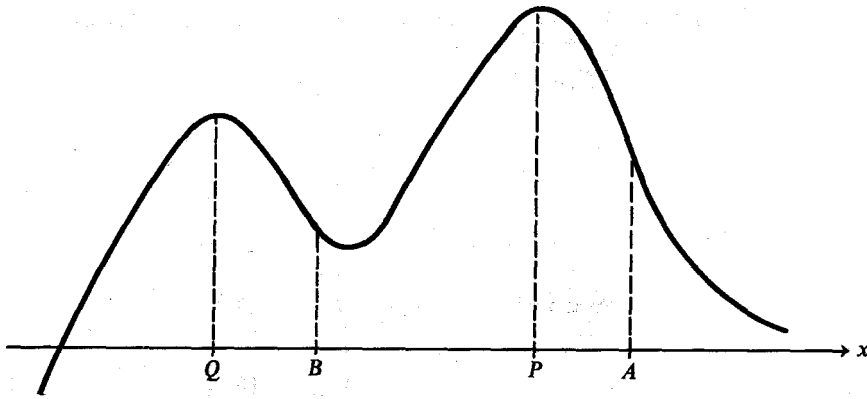


FIGURA 12.1.

donde $\nabla F(\hat{\theta}_n)$ y $\nabla^2 F(\hat{\theta}_n)$ denotan el gradiente (vector $k \times 1$) y matriz hessiana (matriz simétrica $k \times k$) de la función $F(\theta)$ evaluados en el punto $\hat{\theta}_n = \theta$. Podemos mejorar nuestra actual estimación, $\hat{\theta}_n$, reemplazándola por el vector que minimice la expresión cuadrática anterior. Ello ocurre en el punto en que:

$$\frac{\partial M(\theta)}{\partial \theta} = \nabla F(\hat{\theta}_n) + [\nabla^2 F(\hat{\theta}_n)] (\theta^* - \hat{\theta}_n) = 0 \quad [12.4]$$

es decir:

$$\hat{\theta}_{n+1} = \theta^* = \hat{\theta}_n - [\nabla^2 F(\hat{\theta}_n)]^{-1} \nabla F(\hat{\theta}_n) \quad [12.5]$$

expresión que sugiere cómo aproximarse al valor desconocido del vector $\hat{\theta}$ a partir de un vector $\hat{\theta}_n$ suficientemente próximo a él y que constituye el algoritmo de Newton-Raphson. Debe notarse que el punto θ^* que escogemos como nueva estimación minimiza realmente el valor de la función F en el entorno de $\hat{\theta}_n$ si la matriz hessiana $\nabla^2 F(\hat{\theta}_n)$ es definida positiva. Ello está garantizado si F es convexa en el punto $\hat{\theta}_n$, es decir, si dicho punto estaba ya suficientemente próximo a un mínimo local de la función F .

Este algoritmo se utiliza de un modo iterativo, utilizando la nueva estimación como punto de partida en cada etapa del algoritmo y llevando a cabo iteraciones hasta que se satisfagan los criterios de convergencia que el investigador haya estipulado.

Ejemplo 12.2. En el caso particular en que la función a minimizar (o maximizar) sea cuadrática, entonces el desarrollo anterior en serie de Taylor no sería una aproximación, sino exacto, y el mínimo (o máximo) de la función objetivo se alcanzaría en una sola iteración del procedimiento.

Consideremos, a modo de ejemplo, la función:

$$F(\beta_1, \beta_2) = 4\beta_1^2 - 5\beta_1\beta_2 + 3\beta_2^2 + 7\beta_1 - 2\beta_2 + 5$$

que tiene como vector gradiente y matriz hessiana, respectivamente:

$$\nabla F(\beta_1, \beta_2) = (8\beta_1 - 5\beta_2 + 7; -5\beta_1 + 6\beta_2 - 2)$$

$$\nabla^2 F(\beta_1, \beta_2) = \begin{pmatrix} 8 & -5 \\ -5 & 6 \end{pmatrix}$$

con lo que el algoritmo de Newton-Raphson resulta:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}_n = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}_{n-1} - \begin{pmatrix} \hat{\beta}_1 + \frac{32}{23} \\ \hat{\beta}_2 + \frac{19}{23} \end{pmatrix}_{n-1} = \begin{pmatrix} -\frac{32}{23} \\ -\frac{19}{23} \end{pmatrix} \quad [12.6]$$

que, con independencia de los valores iniciales escogidos para los parámetros β_1 y β_2 , proporciona el mínimo de la función $F(\beta_1, \beta_2)$ en tan sólo una iteración. Los valores iniciales de los parámetros son irrelevantes porque desaparecen del miembro derecho de la ecuación [12.6].

La utilización del algoritmo de Newton-Raphson descansa sobre los supuestos: a) *existencia* de las derivadas que en él aparecen, así como: b) *invertibilidad* del hessiano. Cuando esta matriz es singular, significa que la superficie que se pretende optimizar es plana en algunas direcciones, lo que equivale a la imposibilidad de identificar algunos de los parámetros que se están tratando de estimar.

Una situación de este tipo aparece en los Problemas 12.2 y 12.13 al final de este capítulo.

12.4.a. Estimación por mínimos cuadrados

El algoritmo de Newton-Raphson puede utilizarse para obtener numéricamente el estimador mínimo cuadrático en el caso general en que el modelo econométrico representa a y_i como una función *no lineal* de los coeficientes β . En tal caso, se tiene la función objetivo:

$$F(\theta) = SR(\beta) = \sum_1^T [y_i - f(x_i, \beta)]^2$$

y se trata de hallar aquel vector de coeficientes β que minimiza la suma residual (suma de cuadrados de los residuos) $SR(\beta)$; ésta juega el papel de la función genérica $F(\theta)$ en el desarrollo anterior, teniéndose:

$$\hat{\beta}_n \cong \beta_{n-1} - [\nabla^2 F(\hat{\beta}_{n-1})]^{-1} \nabla F(\hat{\beta}_{n-1}) \quad [12.7]$$

Para aplicar el algoritmo, utilizaremos las expresiones del gradiente y matriz hessiana de $SR(\beta)$ que ya vimos en el Capítulo 11:

$$\nabla F(\boldsymbol{\theta}) = \frac{\partial \text{SR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2 \sum_1^T u_t \frac{\partial f_t}{\partial \boldsymbol{\beta}}$$

$$\nabla^2 F(\boldsymbol{\theta}) = \frac{\partial^2 \text{SR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = 2 \sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right)' - 2 \sum_1^T u_t \frac{\partial^2 f_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$$

que deben sustituirse en el algoritmo de Newton-Raphson, para obtener:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n &= \hat{\boldsymbol{\beta}}_{n-1} - \left(\frac{\partial^2 \text{SR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right)_{\hat{\boldsymbol{\beta}}_{n-1}}^{-1} \left(\frac{\partial \text{SR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)_{\hat{\boldsymbol{\beta}}_{n-1}} = \\ &= \hat{\boldsymbol{\beta}}_{n-1} + \left\{ \sum_1^T \left[\left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right)' - \frac{\partial^2 f_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} u_t \right] \right\}^{-1} \left[\sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) u_t \right] \end{aligned} \quad [12.8]$$

donde f_t denota $f_t = f(\mathbf{x}_t, \boldsymbol{\beta})$.

Ejemplo 12.3. Consideremos la estimación de mínimos cuadrados del modelo $y_t = \alpha e^{\beta x_t} + u_t$. Si denotamos por $F(\boldsymbol{\theta}) = \text{SR}(\boldsymbol{\theta})$ su suma residual, donde $\boldsymbol{\theta} = (\alpha, \beta)$, tenemos:

$$\begin{aligned} F(\hat{\boldsymbol{\theta}}) &= \sum_1^T (y_t - \hat{\alpha} e^{\hat{\beta} x_t})^2 \\ \nabla F(\hat{\boldsymbol{\theta}}) &= -2 \sum_1^T [(e^{\hat{\beta} x_t}; \hat{\alpha} x_t e^{\hat{\beta} x_t}) \hat{u}_t] \\ \nabla^2 F(\hat{\boldsymbol{\theta}}) &= 2 \left[\sum_1^T \begin{pmatrix} e^{2\hat{\beta} x_t} & x_t e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) \\ x_t e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) & \hat{\alpha} x_t^2 e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) \end{pmatrix} \right] \end{aligned}$$

El algoritmo de Newton-Raphson adopta la forma:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_n = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_{n-1} + \left[\sum_1^T \begin{pmatrix} e^{2\hat{\beta} x_t} & x_t e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) \\ x_t e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) & \hat{\alpha} x_t^2 e^{\hat{\beta} x_t} (\hat{\alpha} e^{\hat{\beta} x_t} - \hat{u}_t) \end{pmatrix} \right]^{-1} \left[\sum_1^T \begin{pmatrix} e^{\hat{\beta} x_t} \\ \hat{\alpha} x_t e^{\hat{\beta} x_t} \end{pmatrix} \hat{u}_t \right]$$

Una vez que se haya logrado la convergencia del algoritmo, se toma como matriz de covarianzas del estimador obtenido el producto de la estimación de σ_u^2 por la inversa de la matriz hessiana. El estimador tiene distribución asintótica Normal:

$$N \left[\boldsymbol{\theta}; \sigma_u^2 \left(\frac{\partial^2 \text{SR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right)_{\hat{\boldsymbol{\beta}}_n}^{-1} \right] = N [\boldsymbol{\theta}; \sigma_u^2 (\nabla^2 F(\hat{\boldsymbol{\theta}}_n))^{-1}]$$

es decir, con matriz de covarianzas proporcional a la matriz que se invierte en cada etapa del algoritmo.

12.4.b. Estimación por máxima verosimilitud

El algoritmo de Newton-Raphson puede utilizarse también en la estimación de máxima verosimilitud de un modelo econométrico. En tal caso, $F(\boldsymbol{\theta})$ sería igual al logaritmo de la función de verosimilitud, cambiado de signo: $F(\boldsymbol{\theta}) = -\ln L(\boldsymbol{\theta})$, donde $\boldsymbol{\theta}$ incluye ahora tanto los coeficientes del modelo como los parámetros que determinan la matriz de covarianzas.

De este modo, el algoritmo resulta:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n &= \hat{\boldsymbol{\theta}}_{n-1} - \left(\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)_{n-1}^{-1} \cdot \left(\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{n-1} = \\ &= \hat{\boldsymbol{\theta}}_{n-1} - \left(\sum_1^T \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)_{n-1}^{-1} \cdot \left(\sum_1^T \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{n-1}\end{aligned}\quad [12.9]$$

donde l_t denota el logaritmo de la función de densidad de la observación t . El estimador obtenido, $\hat{\boldsymbol{\theta}}$, tiene como distribución asintótica:

$$\hat{\boldsymbol{\theta}}_n \sim N \left[\boldsymbol{\theta}; \left(\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)_{\theta_n}^{-1} \right]$$

donde la matriz de covarianzas se evalúa en la última estimación obtenida.

La utilización de un algoritmo numérico de optimización como el de Newton-Raphson, ya sea para la estimación de mínimos cuadrados o de máxima verosimilitud, hace innecesaria la resolución del sistema de condiciones necesarias de optimalidad que, en el caso de no ser lineales, es un problema complicado, o incluso imposible. Sin embargo, hay que tener en consideración una serie de importantes observaciones acerca de todo procedimiento de optimización numérica:

1. Como una función no lineal tiene, en general, múltiples máximos y mínimos locales, si el algoritmo iterativo converge, no hay una forma fácil de saber si el límite que se ha alcanzado es un mínimo local o global de la función, salvo en casos particulares en que se cuenta con información acerca de la localización del mínimo global, o si se sabe que la función que se minimiza es convexa en todo su dominio.

Por ello, es siempre conveniente intentar la convergencia desde puntos iniciales diferentes; si se alcanza convergencia a distintos mínimos, se escoge, lógicamente, aquel que proporciona un valor menor para la función objetivo. Si se converge al mismo punto, se tiene una mayor garantía —aunque nunca total, salvo en casos de convexidad global— de que dicho límite es el mínimo global de la función.

2. La variación que se introduce sobre el vector de parámetros a estimar en cada etapa del procedimiento iterativo debe conducir a un valor menor de la función objetivo, y para ello es preciso que el hessiano sea definido positivo. Sin embargo, como mencionamos en la Sección 12.3, bien pudiera ocurrir que, aun siendo en la dirección correcta, el incremento que se introduzca sea «excesivo», en el sentido de que el cambio en el estimador sea

tan grande como para llevarnos más allá del mínimo global de la función. Ello no impide que en la siguiente iteración volvamos a cambiar de dirección, convergiendo al punto deseado, pero también podríamos alejarnos cada vez más del mínimo global.

Para evitar este tipo de dificultades, es conveniente utilizar un parámetro λ de longitud de paso como el que consideramos en la Sección 12.3. Ello se basa en el hecho de que si $\nabla^2 F$ es una matriz positiva definida en el punto $\hat{\theta}_{n-1}$ y si dicho punto no es el mínimo global de la función, existe al menos un valor positivo del parámetro θ , tal que el punto $\hat{\theta}_n$ al que se llega satisface $F(\hat{\theta}_n) < F(\hat{\theta}_{n-1})$.

3. En una determinada iteración, la matriz $\nabla^2 F(\hat{\theta}_{n-1})$ podría no ser definida positiva. Ello haría aumentar el valor de la función $F(\theta)$ en vez de disminuirlo, o dejarlo invariante; es decir, en lugar de un mínimo local, podríamos haber alcanzado un máximo local o un punto de silla. Para resolver esta dificultad, se han propuesto diversas variantes del algoritmo de Newton-Raphson que garantizan que la matriz que premultiplica al gradiente de $F(\theta)$ sea siempre definida positiva.

El procedimiento conocido como *quadratic hill-climbing* consiste en sustituir en cada iteración la matriz hessiana de F por la matriz $\nabla^2 F(\hat{\theta}) + \mu \mathbf{I}_k$, donde el parámetro μ se elige de modo que la suma anterior sea una matriz definida positiva y, por tanto, invertible. Si $\nabla^2 F(\hat{\theta})$ es definida negativa, ello se consigue escogiendo μ de modo que sea superior, en valor absoluto, al mayor valor propio de $\nabla^2 F(\hat{\theta})$; si $\nabla^2 F(\hat{\theta})$ fuese singular, cualquier valor positivo de μ haría a la suma anterior definida positiva. Si el valor escogido para μ fuese muy grande, entonces la matriz identidad predomina en la suma anterior sobre el hessiano, y el algoritmo *quadratic hill-climbing* no es muy diferente del de «descenso más rápido» que vimos en la Sección 12.3. Según las sucesivas iteraciones nos fuesen aproximando al mínimo de la función $F(\theta)$, el hessiano pasaría a ser definido positivo, y el algoritmo se reduce al de Newton-Raphson.

Otra variante, conocida como *algoritmos quasi-Newton*, consiste en utilizar en el algoritmo de Newton-Raphson una sucesión de matrices definidas positivas, construidas de modo que converjan al hessiano de $F(\theta)$, en sustitución de éste. Dicha sucesión suele comenzar con la matriz identidad, en cuyo caso la primera iteración coincide con el algoritmo del «descenso más rápido».

4. Puesto que la falta de convergencia es una posible dificultad con todo procedimiento numérico de estimación, hay que establecer unos criterios que permitan decidir si el algoritmo está «suficientemente próximo» a un punto estacionario. Generalmente se utilizan, individual o simultáneamente, los siguientes criterios, según que se quiera poner más énfasis en que: a) el algoritmo haya llegado a un punto «estable», b) que el vector gradiente esté próximo a cero, o c) que el valor de la función objetivo no disminuya de modo apreciable:

$$(\hat{\theta}_n - \hat{\theta}_{n-1})' (\hat{\theta}_n - \hat{\theta}_{n-1}) < \varepsilon_1$$

$$\nabla F(\hat{\theta}_n)' \nabla F(\hat{\theta}_n) < \varepsilon_2$$

$$F(\hat{\theta}_{n-1}) - F(\hat{\theta}_n) < \varepsilon_3$$

12.5. ALGORITMO DE «SCORING»

Especialmente diseñado para el caso en que se pretende obtener el estimador de máxima verosimilitud, este algoritmo se basa en la propiedad de que la esperanza matemática de la matriz hessiana de la función de verosimilitud (es decir, la matriz de información cambiada de signo) tiene, generalmente, una expresión analítica más sencilla que la propia matriz de derivadas segundas.

Así se ha sugerido, como *aproximación*, sustituir la matriz de derivadas segundas que aparece en [12.9], por la matriz de información, teniéndose el llamado algoritmo de «scoring»:

$$\hat{\theta}_n = \hat{\theta}_{n-1} + [\mathbf{I}(\hat{\theta}_{n-1})]^{-1} \nabla \ln L(\hat{\theta}_{n-1}) \quad [12.10]$$

donde $\mathbf{I}(\theta)$ denota la matriz de información. Por utilizar una *aproximación* a la matriz de derivadas segundas de la función objetivo, este algoritmo es, en general, de convergencia más lenta que el propio algoritmo Newton-Raphson, pero es bastante más sencillo de utilizar, siempre que pueda obtenerse la representación analítica de la matriz de información. Otra ventaja que tiene es que la matriz de información es siempre definida positiva, por lo que no habrá problema en seguir una dirección inapropiada. Si se desea, puede utilizarse este algoritmo en conjunción con un parámetro λ de longitud de salto. Una vez lograda la convergencia, el estimador alcanzado tiene como matriz de covarianzas la inversa de la matriz de información.

Ejemplo 12.4. Si suponemos Normalidad del término de error del modelo del Ejemplo 12.3, así como que los parámetros de la matriz de covarianzas son diferentes de los coeficientes del modelo, entonces el hessiano $\partial^2 \ln L(\beta) / \partial \beta \partial \beta'$ coincide con $\partial^2 \text{SR}(\beta) / \partial \beta \partial \beta'$, salvo por el factor $-(1/2)\sigma_u^2$. Por consiguiente, la matriz de información puede calcularse cambiando de signo esta última matriz (es decir, la matriz $\nabla^2 F(\theta)$ en el Ejemplo 12.3) y tomando esperanzas, para obtener

$$\mathbf{I}(\alpha, \beta) = \left[\sigma_u^{-2} \sum_1^T \begin{pmatrix} e^{2\beta x_i} & \alpha x_i e^{2\beta x_i} \\ \alpha x_i e^{2\beta x_i} & \alpha^2 x_i^2 e^{2\beta x_i} \end{pmatrix} \right]$$

como matriz a utilizar en el algoritmo, que queda, finalmente:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_n = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_{n-1} + \left[\sum_1^T \begin{pmatrix} e^{2\beta x_i} & \alpha x_i e^{2\beta x_i} \\ \alpha x_i e^{2\beta x_i} & \hat{\alpha}^2 x_i^2 e^{2\beta x_i} \end{pmatrix}_{n-1} \right]^{-1} \left[\sum_1^T \begin{pmatrix} e^{\beta x_i} \\ \hat{\alpha} x_i e^{\beta x_i} \end{pmatrix}_{n-1} \hat{u}_i \right]$$

y la matriz de covarianzas del vector $\theta = (\alpha, \beta)$ se estima por:

$$\text{Var}(\hat{\theta}_n) = [\mathbf{I}(\hat{\theta}_n)]^{-1}$$

12.6. ALGORITMO DE GAUSS-NEWTON

12.6.a. Estimación de mínimos cuadrados

Este algoritmo es una variante del algoritmo de Newton-Raphson, especialmente útil cuando se trata de *estimar por mínimos cuadrados* un modelo no lineal, en el que la función objetivo es:

$$F(\boldsymbol{\theta}) = \text{SR}(\boldsymbol{\beta}) = \sum_1^T [y_t - f(\mathbf{x}_t, \boldsymbol{\beta})]^2$$

Volviendo a [12.8], el *algoritmo de Gauss-Newton* consiste en ignorar el término que contiene la segunda derivada de f_t en el hessiano. Si la función $f(\mathbf{x}_t, \boldsymbol{\beta})$ es aproximadamente cuadrática en el vector $\boldsymbol{\beta}$, entonces el término que contiene la segunda derivada es aproximadamente proporcional a la suma de los residuos, por lo que su contribución al hessiano es muy pequeña.

Con esta aproximación, sustituimos el hessiano por la matriz simétrica, definida positiva:

$$\left[\sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right)' \right]$$

por lo que el algoritmo de Gauss-Newton resulta:

$$\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_{n-1} + \left[\sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right)' \right]_{n-1}^{-1} \left[\sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) u_t \right]_{n-1} \quad [12.11]$$

Si se logra la convergencia del algoritmo, el estimador resultante tiene *distribución asintótica Normal*, con esperanza igual a $\boldsymbol{\beta}$, y matriz de covarianzas:

$$\text{Var}(\boldsymbol{\beta}) = \sigma_u^2 \left[\sum_1^T \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial f_t}{\partial \boldsymbol{\beta}} \right)' \right]^{-1}$$

donde el parámetro σ_u^2 se estima mediante $\hat{\sigma}_u^2 = \text{SR}(\hat{\boldsymbol{\beta}})/(T - k)$, donde k denota el número de coeficientes estimados.

El lector puede comprobar que, en el caso de un modelo de regresión lineal, la expresión anterior del algoritmo Gauss-Newton se reduce, como es lógico, a la que proporciona el estimador MCO:

$$\hat{\boldsymbol{\beta}} = (\sum_1^T \mathbf{x}_t \mathbf{x}_t')^{-1} (\sum_1^T \mathbf{x}_t y_t)$$

siendo \mathbf{x}_t el vector columna de las observaciones de las variables explicativas e y_t la observación de la variable endógena, ambos en el período t .

La representación [12.11] sugiere asimismo una forma muy simple de utilizar el algoritmo en la práctica: dado un estimador inicial $\hat{\boldsymbol{\beta}}_{n-1}$ del vector $\boldsymbol{\beta}$, se obtienen los residuos correspondientes a dicha estimación y se estima

la regresión auxiliar de \hat{u}_t sobre el vector gradiente $\partial f_t / \partial \beta$, evaluado en $\hat{\beta}_{n-1}$. Este vector gradiente tiene la misma dimensión que β , por lo que se tendrán k coeficientes δ estimados en la regresión auxiliar, que serán las correcciones a introducir en la estimación inicial $\hat{\beta}_{n-1}$, para obtener una nueva estimación $\hat{\beta}_n$:

$$\hat{\beta}_n = \hat{\beta}_{n-1} + \delta = \hat{\beta}_{n-1} + \left[\left(\frac{\partial f(\hat{\beta}_{n-1})}{\partial \beta} \right)' \left(\frac{\partial f(\hat{\beta}_{n-1})}{\partial \beta} \right) \right]^{-1} \left(\frac{\partial f(\hat{\beta}_{n-1})}{\partial \beta} \right)' \hat{u} \quad [12.12]$$

ecuación que reproduce [12.11]. El resultado puede utilizarse, a su vez, como estimación inicial en la siguiente etapa del algoritmo.

En un ejercicio al final del capítulo se pide al lector que compruebe que en una primera iteración el algoritmo de Gauss-Newton equivale a la aproximación lineal que se propuso en el Ejemplo 12.1. Puede asimismo comprobarse sin gran esfuerzo que el algoritmo de Gauss-Newton coincide, en el Ejemplo 12.3, con el algoritmo de scoring. Por último, es claro que, en un modelo de regresión *lineal*, los algoritmos de Gauss-Newton y de Newton-Raphson coinciden, ya que la matriz hessiana de la función f es en tal caso igual a cero.

El algoritmo de Gauss-Newton presenta dos ventajas con respecto al de Newton-Raphson:

a) Solamente utiliza derivadas de primer orden, lo que puede ser en algunos casos una ventaja importante, puesto que el cálculo de las diversas derivadas en una función no lineal puede ser complicado, e incluso pudieran no existir.

b) La matriz que se utiliza en sustitución del hessiano en el algoritmo de Gauss-Newton es siempre definida positiva, por construcción.

Los mismos comentarios que hicimos a propósito del esquema de Newton-Raphson son aplicables al algoritmo de Gauss-Newton, pues éste no es sino una versión simplificada del primero. En particular, una variante del algoritmo de Gauss-Newton, del tipo *quadratic hill-climbing*:

$$\hat{\beta}_n = \hat{\beta}_{n-1} + \left[\Sigma_1^T \left(\frac{\partial f_t}{\partial \beta} \right) \left(\frac{\partial f_t}{\partial \beta} \right)' + \mu \mathbf{I}_k \right]^{-1} \left[\Sigma_1^T \left(\frac{\partial f_t}{\partial \beta} \right) u_t \right]_{n-1}$$

es muy utilizada en el trabajo econométrico aplicado, y se conoce como *algoritmo de Marquardt*.

Ejemplo 12.4. Si quisiéramos utilizar el algoritmo de Gauss-Newton para la estimación del modelo [11.5] partiendo de condiciones iniciales: $\hat{\alpha}_0 = \bar{y}$, $\hat{\beta}_0 = 0$, obtendríamos los residuos $\hat{u}_t = y_t - \bar{y}$ y estimaríamos una regresión de ellos sobre las dos variables que componen el vector gradiente:

$$\nabla f(x_t, \hat{\alpha}_0, \hat{\beta}_0) = (e^{\hat{\beta}_0 x_t}; \hat{\alpha}_0 x_t e^{\hat{\beta}_0 x_t}) = (1; \bar{y} x_t)$$

por lo que la primera iteración coincidiría con el modelo linearizado [11.6].

12.6.b. Estimación de máxima verosimilitud mediante el algoritmo de Gauss-Newton

Berndt, Hall, Hall y Hausman (1974) han propuesto un algoritmo similar al de Gauss-Newton, pero aplicable en la estimación de máxima verosimilitud. Su propuesta nace del conocido resultado estadístico:

$$E \left[\left(\frac{\partial \ln L(y_t, \mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln L(y_t, \mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \right] = -E \left(\frac{\partial^2 \ln L(y_t, \mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \quad [12.13]$$

donde $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_u^2)$. Esta igualdad sugiere que en el caso de la estimación por máxima verosimilitud podría utilizarse una variante del algoritmo de Newton-Raphson que en vez de la matriz de información utilizase la matriz:

$$\left(\frac{\partial \ln L(y_t, \mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln L(y_t, \mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' = \sum_1^T \left(\frac{\partial l_t}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial l_t}{\partial \boldsymbol{\theta}} \right)'$$

teniéndose el algoritmo:

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} - \left[\sum_1^T \left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \right]_{n-1}^{-1} \cdot \left(\sum_1^T \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{n-1} \quad [12.14]$$

Esta variante del algoritmo de Gauss-Newton es especialmente útil en los casos en que la evaluación de las derivadas segundas de la función de verosimilitud es compleja, pero hay que hacer hincapié en que no es sino una aproximación, que sustituye una esperanza matemática (la que aparece a la izquierda de [12.13]) por su análogo muestral. Una vez más, el algoritmo puede utilizarse con un parámetro λ de longitud de salto.

La matriz de covarianzas del estimador resultante, una vez que se ha logrado la convergencia del algoritmo, viene dada por la matriz que se invierte en [12.14], evaluada en la última estimación alcanzada. Esta matriz incorpora ya el parámetro σ_u^2 .

El lector puede comprobar que en el Ejemplo 12.3 este algoritmo coincide, de nuevo, con el algoritmo de Gauss-Newton.

PROBLEMAS

Problema 12.1. Suponiendo que el término de error del modelo econométrico

$$y_t = \alpha + \beta_1 e^{\beta_2 x_t} + u_t, \quad t = 1, 2, \dots, T$$

sigue una distribución $N(0, \sigma_u^2)$, con $E(u_t, u_s) = 0$ para todo $t \neq s$, obtener las expresiones analíticas de los algoritmos de: a) descenso más rápido, b) Newton-Raphson, c) Gauss-Newton, d) scoring.

Mostrar cuál sería la primera iteración de cada método a partir de las condiciones iniciales: $\beta_1 = 1$, $\beta_2 = 0$.

¿Existe alguna diferencia entre la utilización de estos algoritmos para la estimación del modelo anterior por mínimos cuadrados y su utilización para la estimación del modelo por máxima verosimilitud? ¿A qué se debe tal propiedad?

Problema 12.2. Mostrar que en el modelo econométrico

$$y_t = \alpha + \beta_1 \beta_2 x_t + u_t, \quad t = 1, 2, \dots, T$$

existe multicolinealidad, por lo que no puede estimarse el vector $(\alpha, \beta_1, \beta_2)$. ¿Puede estimarse algún subvector? ¿Cómo podrían utilizarse los algoritmos discutidos en este capítulo?

Puesto que este modelo es lineal en las variables, ¿qué ventaja representa la utilización en este modelo de métodos de estimación de modelos no lineales frente a la aplicación de MCO?

Problema 12.3. Responder a las mismas preguntas del Problema 12.1, referidas ahora a la estimación del modelo:

$$y_t = \alpha \beta^{x_t} + u_t, \quad t = 1, 2, \dots, T$$

Problema 12.4. Responder a las mismas cuestiones, referentes a la estimación del modelo:

$$y_t = \frac{\beta_1}{1 - \beta_2} x_{1t} + \beta_3 e^{\beta_4 x_{2t}} + u_t, \quad t = 1, 2, \dots, T$$

Problema 12.5. Probar que si al aplicar el algoritmo de Gauss-Newton (o el de Newton-Raphson) a un modelo *lineal* con término independiente:

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t, \quad t = 1, 2, \dots, T$$

se utilizan como condiciones iniciales $\beta_1 = \bar{y}$; $\beta_i = 0$ para todo $i = 2, \dots, k$, se alcanza el estimador de mínimos cuadrados ordinarios en tan sólo una iteración.

(El ejercicio resulta trivial si se transforman las variables en diferencias con respecto a la media, pero tiene interés que el lector se convenza de que puede resolverlo sin dicha transformación.)

Problema 12.6. ¿Cómo podrían utilizarse los algoritmos presentados en el capítulo para la estimación del modelo:

$$g(y_t, \mathbf{x}_t, \boldsymbol{\beta}) = u_t, \quad t = 1, 2, \dots, T$$

donde g es una función implícita de la variable y_t y los vectores \mathbf{x}_t y $\boldsymbol{\beta}$?

Problema 12.7. Discutir las similitudes y diferencias entre el algoritmo de Gauss-Newton y el tratamiento del modelo no lineal que se propuso en la Sección 11.1.

Problema 12.8. Obtener las expresiones analíticas para los algoritmos de Gauss-Newton y de scoring correspondientes a la estimación de los parámetros de los modelos que aparecen en los Problemas 6.10, 6.12, 6.14, 6.15, 6.16, 6.17, 6.18 al final del Capítulo 6.

Problema 12.9. Relacionar el procedimiento de estimación de modelos univariantes de la Sección 13.10 con los algoritmos numéricos de estimación analizados en este capítulo.

Problema 12.10. Obtener la expresión analítica del algoritmo de Newton-Raphson para la maximización de la función de verosimilitud condicional en el modelo:

$$\begin{aligned}y_t &= \phi y_{t-1} + u_t, & t = 1, 2, \dots, T \\u_t &= \rho u_{t-1} + \varepsilon_t\end{aligned}$$

donde ε_t es ruido blanco.

Problema 12.11. Repetir el ejercicio anterior en el caso de que el término de error siga la estructura:

$$u_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

donde ε_t es ruido blanco.

Problema 12.12. Obtenga el test ML para el contraste global de significación de la estimación mínimo-cuadrática de un modelo econométrico no lineal.

Problema 12.13. Describa los algoritmos precisos para la estimación del modelo:

$$y_t = \beta_1 \beta_2 x_{1t} + e^{\beta_2} x_{2t} + u_t$$

Problema 12.14. Describa en detalle las expresiones analíticas que permitirían utilizar el uso del algoritmo Gauss-Newton para la estimación numérica del modelo:

$$y_t = \alpha e^{\beta x_t} + u_t$$

Problema 12.15. Explique en detalle cómo llevaría a cabo las sucesivas iteraciones de un procedimiento numérico para la obtención de estimadores de mínimos cuadrados de los modelos:

- a) $\ln y_t = \beta_0 + \beta_1 x_t + u_t$.
- b) $y_t = x_t^\beta + u_t$.

Problema 12.16. Pruebe que el contraste de los multiplicadores de Lagrange de la hipótesis $H_0: \gamma = 0$ en el modelo

$$y_t = \beta_1 x_{1t} + \beta_2 (x_{2t} - \gamma)^{-1} + u_t$$

consiste en una regresión de y_t sobre x_{1t} y $\frac{1}{x_{2t}}$, para luego estimar una segunda regresión de los residuos de la primera sobre x_{1t} , $\frac{1}{x_{2t}}$ y $\frac{1}{x_{2t}^2}$, comparando TR^2 de esta última con una distribución χ_1^2 .

Problema 12.17. Describa en detalle la utilización del algoritmo de Gauss-Newton para la estimación de máxima verosimilitud de los modelos con transformación Box-Cox de la Sección 11.4.

Problema 12.18. Describa en detalle la construcción del estadístico de los multiplicadores de Lagrange para el contraste de la hipótesis nula: $H_0: \lambda = 0$ en el modelo:

$$y_t = \beta_1 + \beta_2 \frac{x_t^\lambda - 1}{\lambda} + u_t$$

Problema 12.19. Describa en detalle la utilización del contraste de los multiplicadores de Lagrange para el contraste de la hipótesis $H_0: \lambda = 0$ en el modelo:

$$\frac{y_t^\lambda - 1}{\lambda} = \beta_1 + \beta_2 \frac{x_t^\lambda - 1}{\lambda} + u_t$$

Problema 12.20. Describa en detalle la utilización del algoritmo de Gauss-Newton para estimar el modelo:

$$y_t = \frac{\omega_0 + \omega_1 L + \omega_2 L^2}{1 - \delta L} x_t + u_t$$

$$u_t = \frac{1}{(1 - \delta L)(1 - \rho L)} a_t$$

donde a_t es un ruido blanco.

Problema 12.21. Explique en detalle la utilización del algoritmo de Gauss-Newton para la estimación de máxima verosimilitud y su matriz de covarianzas del modelo:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

$$u_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad |\theta| < 1$$

donde ε_t es un ruido blanco, distribuido $N(0, \sigma_\varepsilon^2)$.

Problema 12.22. a) Derive en detalle las condiciones necesarias para la obtención numérica del estimador de máxima verosimilitud condicionada del modelo AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

donde ε_t es un ruido blanco con distribución $N(0, \sigma_\varepsilon^2)$ y $E(u_t) = 0$, $\text{Var}(u_t) = \sigma_u^2$ para todo t .

b) Proporcione las expresiones analíticas de los elementos que integran la matriz de covarianzas de dicho estimador.

c) Explique en detalle cómo podría utilizarse el algoritmo de Newton-Raphson para la obtención numérica del estimador.

CAPITULO 13

MODELOS DE SERIES TEMPORALES

13.1. INTRODUCCION

En los capítulos anteriores hemos tratado de la estimación de modelos econométricos, en los que una lista de variables exógenas se utiliza para explicar el comportamiento de otra variable, endógena. Hemos tratado de la contrastación de hipótesis con dichos modelos, así como de su utilización con fines predictivos y, posteriormente, hemos considerado modelos dinámicos. Este desarrollo convencional del modelo econométrico de relación entre variables se basa en el principio de mínimos cuadrados, tradicional en la Teoría Estadística, que se ha aplicado habitualmente en la estimación de relaciones en todo tipo de ciencias, experimentales o no experimentales.

Pero en ciencias no experimentales como la Economía, el investigador debe incluso tratar de descubrir la especificación del modelo, pues no existe un diseño experimental que haya generado los datos de que dispone. Apenas hemos tratado hasta ahora de la especificación de dichos modelos, excepto por algunas referencias conceptuales, nunca suficientemente precisas, al aspecto de Teoría Económica en debate. En 1970, Box y Jenkins propusieron una metodología rigurosa para la identificación, estimación y diagnóstico de modelos dinámicos para datos de series temporales que, justificadamente, se ha convertido en una herramienta habitual en el análisis de series económicas, y que presentamos en este capítulo. En la primera parte analizamos modelos en los que el comportamiento de una variable se explica utilizando sólo su propio pasado, por lo que se conocen como *modelos univariantes*. Posteriormente, consideramos modelos dinámicos con variables explicativas, que se conocen como *modelos de función de transferencia*.

Conceptualmente, los modelos de función de transferencia no son distintos de los modelos econométricos que hasta ahora hemos considerado, con la diferencia de que los procedimientos de especificación y diagnóstico que vamos a introducir permiten que las propiedades dinámicas de las relaciones entre variables, así como la estructura estocástica del término de error, queden

perfectamente recogidas en los modelos estimados. Los modelos univariantes, por otra parte, son un complemento perfecto al estudio de relaciones entre variables. Un modelo univariante permite caracterizar adecuadamente muchas de las características de una variable económica en su dimensión temporal, y ello es importante, al menos por dos razones: *a)* porque el analista querrá que las variables que aparecen a ambos lados de un modelo de relación tengan en común algunas de sus características más importantes, como indicativo de que la relación que ha especificado es realmente relevante, *b)* porque constituye una referencia, relativamente sencilla de obtener, con la que comparar posibles modelos de relación que posteriormente pudieran estimarse.

En este capítulo vamos a definir y caracterizar una amplia familia de estructuras estocásticas lineales, así como una familia de estadísticos que nos permitan escoger una de tales especificaciones como la más adecuada para representar la estructura estocástica de la variable económica que se está analizando.

13.2. PRIMERAS DEFINICIONES

13.2.a. Proceso estocástico, ruido blanco, paseo aleatorio

Comenzamos introduciendo algunos conceptos fundamentales sobre los que se sustenta la teoría que se desarrolla en las secciones siguientes:

Definición 13.1. Llamamos *proceso estocástico* a una sucesión de variables aleatorias $\{y_t\}$, $t = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$.

Aunque el índice que describe la sucesión de variables aleatorias que configura un proceso estocástico no necesita tener una interpretación concreta, la utilización de este concepto de Econometría confiere al índice t una interpretación como del período al que corresponde la variable aleatoria y_t . Esta definición es muy general, y las variables aleatorias y_t no precisan satisfacer ninguna propiedad en particular. Podrían carecer de algunos momentos, como varianza o esperanza; incluso su distribución marginal (es decir, no condicional en un conjunto de información como el constituido por los valores de las variables previas y_s , $s < t$) podría no existir.

Según que las variables y_t satisfagan unas u otras propiedades, tenemos un proceso estocástico de uno u otro tipo, como los que comenzamos ya a presentar:

Definición 13.2. Se llama *ruido blanco* a una sucesión de variables aleatorias con esperanza cero, igual varianza, e independientes en el tiempo. En lo sucesivo, denotamos un ruido blanco por $\{\varepsilon_t\}$.

Definición 13.3. Un *paseo aleatorio* es un proceso estocástico $\{y_t\}$ cuyas primeras diferencias forman un proceso de ruido blanco, es decir:

$$\nabla y_t = \varepsilon_t$$

o, lo que es lo mismo:

$$y_t = y_{t-1} + \varepsilon_t, \quad t = -\infty, \dots, -2, -1, 0, 1, 2, \dots$$

donde ε_t es un ruido blanco.

Una forma de abordar el análisis estadístico de una serie temporal económica consiste en considerar que la serie temporal correspondiente a una variable como el consumo agregado es la *realización* de un proceso estocástico. Bajo este punto de vista, cada dato de la serie es una realización (es decir, una muestra de tamaño 1) de la variable C_t que compone el proceso estocástico de consumo $\{C_t\}_{t=1}^{\infty}$.

13.2.b. Estacionariedad

Definición 13.4. Un proceso estocástico $\{y_t\}$ es *estacionario en sentido estricto* si para toda m -tupla (t_1, t_2, \dots, t_m) y todo entero k el vector de variables $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$ tiene la misma distribución de probabilidad conjunta que el vector $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$.

Examinemos esta definición en detalle: ¿Qué ocurre cuando $m = 1$? En tal caso, el conjunto de variables se reduce a una sola (y_t) y, de acuerdo con la definición, en un proceso estocástico estacionario este conjunto ha de tener la misma distribución independientemente del valor del índice t . Es decir, que las *variables aleatorias que componen un proceso estocástico estacionario están idénticamente distribuidas*. En particular, la esperanza y la varianza de las variables y_t son independientes del tiempo.

Cuando $m = 2$, la distribución conjunta del par (y_t, y_{t-s}) en un proceso estacionario debe ser independiente del tiempo. Como consecuencia, la covarianza entre ellas, así como su coeficiente de correlación, debe ser invariante en t , dependiendo únicamente del valor del retardo, s . Es decir, en un proceso estocástico estacionario, la covarianza entre y_t e y_{t-2} es igual a la covarianza entre y_s e y_{s-2} , para todo t y s , si bien será generalmente diferente de la covarianza entre y_t e y_{t-1} o de la covarianza entre y_t e y_{t-3} , que serán a su vez constantes en el tiempo, aunque diferentes entre sí.

Sin embargo, el concepto de estacionariedad en sentido estricto implica el cumplimiento de un número de condiciones que es excesivo para nuestras necesidades prácticas. Generalmente, nos conformamos con un concepto menos exigente, como es el de *estacionariedad en sentido débil* o *de segundo orden*, que se produce cuando todos los momentos de primer y segundo orden del proceso estocástico son invariantes en el tiempo. Estos momentos incluyen la esperanza matemática y la varianza de las variables y_t , pero también las covarianzas y correlaciones entre diversos retardos a que antes hemos hecho referencia. En lo sucesivo, cuando hablamos de proceso estacionario nos referimos a un proceso estocástico estacionario de segundo orden.

Algunas de estas condiciones son fácilmente contrastables. Por ejemplo, si una serie temporal muestra una tendencia creciente, como cada observación es una realización de la variable aleatoria correspondiente, no podremos

mantener que la esperanza matemática de dichas variables es constante, sino que deberemos aceptar que crece con el tiempo. En otras ocasiones, una serie temporal económica experimenta fluctuaciones de amplitud creciente en el tiempo, en cuyo caso deberemos reconocer que la varianza de las variables y_t no es constante y el proceso no es estacionario de segundo orden. *El ruido blanco es un ejemplo de proceso estocástico estacionario de segundo orden.*

El paseo aleatorio no es un proceso estacionario, puesto que puede escribirse:

$$y_t = y_{t-1} + \varepsilon_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \dots$$

sin que esté definida su esperanza ni su varianza, ni tampoco la distribución marginal de cada y_t . Si está definida, sin embargo, su distribución condicional en el pasado: $\{y_{t-1}, y_{t-2}, \dots\}$, que será del mismo tipo que la de ε_t (es decir, Normal o binomial dependiendo de que ésta lo sea), con igual varianza, y esperanza igual a la de ε_t , aumentada en el valor realizado de y_{t-1} .

A veces se considera un paseo aleatorio al que se le ha impuesto una condición inicial, la de comenzar a partir de un valor numérico y_0 , de modo que puede escribirse:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \dots = \\ &= y_0 + \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_2 + \varepsilon_1 \end{aligned}$$

Este proceso, sobre el que hay que insistir que es *distinto del paseo aleatorio* que antes hemos definido, tiene distribución marginal, que será, por ejemplo Normal si la de ε_t lo es para todo t , con esperanza igual a y_0 y varianza igual a $t\sigma_\varepsilon^2$, creciente en el tiempo. Como consecuencia, tampoco es estacionario. Su distribución condicional es igual a la del paseo aleatorio.

Unos estadísticos fundamentales en la especificación de modelos univariantes son las *funciones de autocovarianza, de autocorrelación simple y de autocorrelación parcial* que a continuación introducimos:

Definición 13.5. La *función de autocovarianza* de un proceso estocástico $\{y_t\}$ es una función, a la que en lo sucesivo nos referimos como *fac.*, que para cada instante t y cada número entero k toma un valor, denotado por $\gamma_k(t)$, igual a la covarianza entre y_t e y_{t-k} , es decir:

$$\gamma_k(t) = \text{Cov}(y_t, y_{t-k})$$

Definición 13.6. La *función de autocorrelación simple* de un proceso estocástico $\{y_t\}$, a la que en lo sucesivo nos referimos por *fas.*, es una función que para cada instante t y cada entero k toma un valor $\rho_k(t)$ igual a la correlación entre y_t e y_{t-k} :

$$\rho_k(t) = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var } y_t} \sqrt{\text{Var } y_{t-k}}} = \frac{\gamma_k(t)}{\sqrt{\text{Var } y_t} \sqrt{\text{Var } y_{t-k}}}$$

Definición 13.7. La *función de autocorrelación parcial* de un proceso estocástico $\{y_t\}$, a la que en lo sucesivo nos referimos como *fap*, es una función que para cada instante t y cada entero k toma un valor igual a la correlación entre y_t e y_{t-k} , ajustada por el efecto de los retardos intermedios $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$.

El gran interés de un proceso estocástico estacionario reside en que las *fas*, *fas* y *fap* son independientes del tiempo t , por lo que puede omitirse dicho argumento temporal. Lo que es crucial es que dicha invarianza permite la estimación muestral de tales funciones, del modo que analizamos en la sección siguiente. Asimismo, aun antes de pasar a su estimación, el concepto de estacionariedad nos permite deducir algunas propiedades de estas funciones:

En general, parte de la correlación existente entre y_t e y_{t-2} estará producida por el hecho de que ambas están correlacionadas con y_{t-1} , y eso es lo que trata de corregir la *fap*. El primer valor de la *fap* es la correlación entre y_t e y_{t-1} , sin que haya que corregir por ningún retardo intermedio, puesto que no existen. Por eso es que *el primer valor de las fas y fap de cualquier proceso estocástico coinciden*. De manera análoga, como el lector verá fácilmente, *el valor inicial de la fas, ρ_0 , es igual a 1* en todo proceso estacionario, por ser el cociente de la varianza del proceso (constante en el tiempo) por sí misma. Por un razonamiento similar al anterior, concluimos que el valor inicial de la *fap* es asimismo igual a 1 en todo proceso estacionario. Por último, las *fas* y *fap* de un proceso de ruido blanco son iguales a cero, excepto en sus valores iniciales que, como hemos visto, son iguales a 1. Esto no es sino una manifestación de una propiedad más general que consiste en que *las fas y fap de todo proceso estocástico estacionario decrecen rápidamente hacia cero*.

13.2.c. Estimación de las funciones de autocorrelación de un proceso estacionario

Ya hemos expuesto cómo la estacionariedad de un proceso permite considerar la estimación de sus *fas* y *fap* que son las herramientas básicas de la especificación de su representación univariante. La estimación de los distintos valores ρ_k se lleva a cabo de la siguiente forma:

$$r_k = \frac{\frac{1}{T-k} \sum_{k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\frac{1}{T} \sum_1^T (y_t - \bar{y})^2} \sqrt{\frac{1}{T-k} \sum_{k+1}^T (y_{t-k} - \bar{y})^2}}$$

Distintas simplificaciones se consiguen en esta expresión cuando el tamaño muestral T es grande con respecto a k , pues entonces dividir por T o por $T-k$ es prácticamente lo mismo, en cuyo caso todos los cocientes de la forma $\frac{1}{T}$ y $\frac{1}{T-k}$ que aparecen en la expresión anterior pueden eliminarse.

Por otra parte, las medias muestrales de $(y_t - \bar{y})^2$ sobre las observaciones 1, 2, ..., T , o sobre las observaciones $k + 1, k + 2, \dots, T$ serán muy similares. Con estas aproximaciones se tiene el estimador:

$$r_k = \frac{\sum_{k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_1^T (y_t - \bar{y})^2}, \quad k = 0, 1, 2, \dots$$

La expresión que acabamos de presentar para la estimación de esta función garantiza que el valor estimado de \hat{r}_0 será siempre igual a 1.

Finalmente, en todo proceso estacionario, la función de autocovarianza es simétrica, es decir, $\gamma_{-k} = \gamma_k$, lo que proviene del hecho de que la covarianza entre y_t e y_{t-k} es igual a la covarianza entre y_t e y_{t+k} . Como consecuencia, la función de autocorrelación es también simétrica, por lo que en el trabajo práctico se analizan únicamente dichas funciones para valores de $k = 0, 1, 2, \dots$

El primer valor de la fap, que vamos a denotar por ϕ_{11} (luego veremos por qué utilizamos dos subíndices), puede estimarse transformando la serie y_t en desviaciones con respecto a su media muestral $\tilde{y}_t = y_t - \bar{y}$, y a continuación estimando una regresión de \tilde{y}_t sobre \tilde{y}_{t-1} .

Ya vimos en el Capítulo 3 que, en el caso de variables con media muestral igual a cero, el estimador MCO de β en la regresión $y_t = \beta x_t + u_t$ es precisamente igual al coeficiente de correlación parcial entre x e y , multiplicado por el cociente de sus desviaciones típicas. En este caso, las variables en la regresión son \tilde{y}_t e \tilde{y}_{t-1} , y si el proceso es estacionario, sus desviaciones típicas son iguales. Por tanto, la pendiente estimada de la regresión anterior coincide con el coeficiente de correlación entre y_t e y_{t-1} , es decir, ϕ_{11} .

Otra consecuencia de este comentario es que el primer valor de la fap es siempre igual al primer valor de la fas. Intuitivamente, la razón es que al estimar la correlación entre y_t e y_{t-1} , no hay que corregirla de ningún valor intermedio, por lo que las dos funciones, ϕ_{11} y r_1 , toman el mismo valor numérico, al igual que ocurre con sus valores teóricos.

El segundo valor de la fap se estima mediante una regresión de \tilde{y}_t sobre \tilde{y}_{t-1} e \tilde{y}_{t-2} ⁽¹⁾.

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + u_t$$

El coeficiente estimado ϕ_{22} mide la correlación entre \tilde{y}_t e \tilde{y}_{t-2} una vez que se ha tenido en cuenta el efecto común de \tilde{y}_{t-1} , incluida como otra variable explicativa adicional. Así, la fap puede estimarse mediante una serie de regresiones, cada una de las cuales contiene como variable explicativa un retardo más que la anterior, y de las que nos vamos quedando en cada caso con los coeficientes estimados en los retardos más altos: $\phi_{11}, \phi_{22}, \phi_{33}, \dots$. Otra posibilidad de obtener la fap estimada es mediante fórmulas recursivas.

⁽¹⁾ Rigurosamente hablando, las desviaciones con respecto a la media muestral alteran las propiedades estocásticas del proceso, pues para que no cambiasen se debería sustraer una constante. Restar la esperanza poblacional satisface esa condición, pero dicha esperanza es desconocida, y la corrección se hace con la media muestral, que es una variable aleatoria.

utilizando la fas previamente estimada, y utilizando las ecuaciones de Yule-Walker, como veremos en la sección siguiente.

13.3. MODELOS AUTORREGRESIVOS

Comenzamos a introducir en esta sección las estructuras estocásticas lineales que trataremos de asociar a una serie temporal de datos económicos. La primera de tales estructuras es la de los modelos autorregresivos.

Definición 13.8. Un *proceso autorregresivo* de orden 1, denotado por AR(1), viene definido por:

$$y_t = \phi y_{t-1} + \delta + \varepsilon_t \quad [13.1]$$

donde ϕ y δ son constantes y ε_t es un ruido blanco.

Si un modelo AR(1) es estacionario, entonces su esperanza y varianza son constantes en el tiempo, y se tiene que:

$$a) \quad E y_t = \phi E y_{t-1} + \delta + E \varepsilon_t = \phi E y_{t-1} + \delta, \text{ pero } E y_t = E y_{t-1}, \text{ por lo que}$$

$$E y_t = \mu_y = \frac{\delta}{1 - \phi}.$$

$$b) \quad \text{Var } y_t = \phi^2 \text{Var } y_{t-1} + \text{Var } \varepsilon_t. \text{ Ahora bien, si el proceso AR(1) es}$$

$$\text{estacionario, entonces } \text{Var } y_t = \text{Var } y_{t-1}, \text{ por lo que } \text{Var } y = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

Por otra parte, mediante sustituciones repetidas, puede verse que:

$$y_t = (\delta + \delta\phi + \delta\phi^2 + \dots) + (\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots) = \frac{\delta}{1 - \phi} + \sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s} \quad [13.2]$$

Esta expresión tendrá sentido sólo si la suma infinita que en ella aparece converge. Dicha suma es aleatoria, puesto que es una combinación lineal de las variables aleatorias ε_t y converge si y sólo si lo hace su varianza. Dado que las variables ε_t son independientes, se tiene:

$$\text{Var}(y_t) = \text{Var}\left(\sum_0^{\infty} \phi^s \varepsilon_{t-s}\right) = \sum_0^{\infty} \text{Var}(\phi^s \varepsilon_{t-s}) = \sum_0^{\infty} \phi^{2s} \text{Var } \varepsilon_{t-s} = \sum_0^{\infty} \phi^{2s} \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2}$$

por lo que la varianza de la combinación lineal es finita si y sólo si $|\phi| < 1$. Sólo si $|\phi| < 1$ tendrá sentido la expresión [13.2] y, en consecuencia, y_t podrá expresarse como función de ε_t y de todas las variables ε anteriores al instante t , pero de ninguna futura. Sólo entonces es el proceso AR(1) *estacionario*, para lo que es necesario y suficiente que $|\phi| < 1$. Es entonces cuando su esperanza y varianza están bien definidas por las expresiones que aquí hemos derivado.

El coeficiente de cada variable ε_{t-k} en dicha expresión es ϕ^k . Como el proceso ε_t es un ruido blanco, entonces se concluye que:

a) $E(y_{t-k}\varepsilon_t) = 0$ para todo $k > 0$, ya que y_{t-k} depende de ε_{t-k} y valores anteriores del proceso ε_t , pero no de sus valores futuros. Por tanto:

$$E(y_{t-k}\varepsilon_t) = E\left[\frac{\delta\varepsilon_t}{1-\phi}\right] + \sum_{s=0}^{\infty} \phi^s E(\varepsilon_{t-k-s}\varepsilon_t) = 0 + 0 = 0, \quad \forall k > 0$$

$$b) E(y_t\varepsilon_{t-k}) = E\left[\frac{\delta\varepsilon_{t-k}}{1-\phi}\right] + \sum_{s=0}^{\infty} \phi^s E(\varepsilon_{t-s}\varepsilon_{t-k}) = \phi^k \sigma_\varepsilon^2, \quad \forall k \geq 0.$$

Por otra parte, si se sustituye δ por su expresión equivalente $(1 - \phi)\mu_y$ en el modelo [13.1], se tiene:

$$y_t - \mu_y = \phi(y_{t-1} - \mu_y) + \varepsilon_t, \quad \text{es decir: } \tilde{y}_t = \phi\tilde{y}_{t-1} + \varepsilon_t$$

donde \tilde{y}_t denota la diferencia entre y_t y su esperanza poblacional. Dicha diferencia tiene los mismos momentos que la variable \tilde{y}_t . En particular: $\sigma_y^2 = \sigma_{\tilde{y}}^2$, y pueden aplicarse los resultados anteriores a) y b). Con todo ello, la función de autocovarianza de este proceso resulta ser:

$$\gamma_0 = \sigma_y^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2}, \text{ como vimos antes, y:}$$

$$\gamma_1 = E(\tilde{y}_t\tilde{y}_{t-1}) = E(\phi\tilde{y}_{t-1}^2 + \varepsilon_t\tilde{y}_{t-1}) = \phi\sigma_{\tilde{y}}^2 = \phi\gamma_0$$

$$\gamma_2 = E(\tilde{y}_t\tilde{y}_{t-2}) = E(\phi\tilde{y}_{t-1}\tilde{y}_{t-2} + \varepsilon_t\tilde{y}_{t-2}) = \phi\gamma_1 + E(\varepsilon_t\tilde{y}_{t-2}) = \phi^2\gamma_0 + 0 = \phi^2\gamma_0$$

$$\gamma_3 = E(\tilde{y}_t\tilde{y}_{t-3}) = E(\phi\tilde{y}_{t-1}\tilde{y}_{t-3} + \varepsilon_t\tilde{y}_{t-3}) = \phi\gamma_2 + E(\varepsilon_t\tilde{y}_{t-3}) = \phi^3\gamma_0$$

y así sucesivamente, de modo que: $\rho_0 = 1$; $\rho_1 = \phi$; $\rho_2 = \phi^2$; $\rho_3 = \phi^3$; ...; $\rho_k = \phi^k$ para todo $k > 0$, por lo que los valores de la fas son las sucesivas potencias del parámetro ϕ . Además, la condición $|\phi| < 1$ garantiza que los sucesivos valores de la fas convergen a cero. Esta función puede tener dos aspectos distintos, dependiendo del signo de tal parámetro. Así, se tiene que:

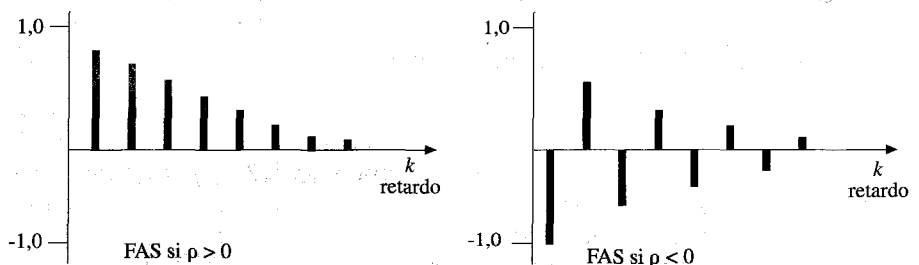


FIGURA 13.1. Función de autocorrelación simple de un proceso AR(1).

De la representación analítica [13.1] del proceso AR(1) se deduce que su fap tiene como primer valor $\phi_{11} = \phi$, y como restantes valores $\phi_{kk} = 0$. Por ejemplo, el segundo valor de dicha función sería el parámetro ϕ_{22} en la regresión:

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + \varepsilon_t$$

pero, de acuerdo con el modelo teórico [13.1], ϕ_{22} debe ser cero, y lo mismo ocurre para todo ϕ_{kk} con $k \geq 2$.

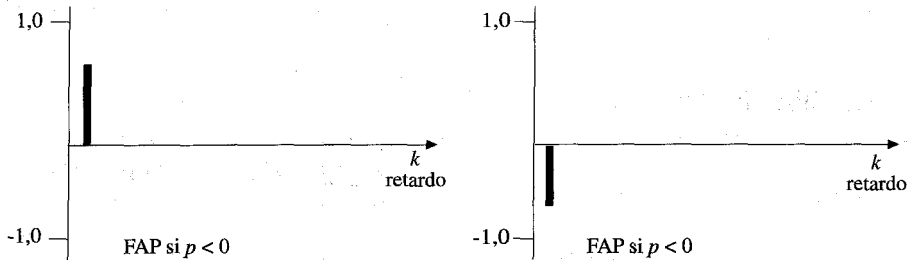


FIGURA 13.2. Función de autocorrelación parcial de un proceso AR(1).

Definición 13.9. Un proceso es *autorregresivo de orden 2*, que denotamos AR(2), si responde a la ley estocástica:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad [13.3]$$

y es estacionario si $|\phi_2| < 1$, $\phi_1 + \phi_2 < 1$ y $\phi_2 - \phi_1 < 1$. Como veremos más adelante, bajo estas condiciones, y_t puede escribirse como función de ε_t y sus valores previos. Calculemos la función de autocorrelación de un proceso AR(2) estacionario: para ello comencemos notando que $Ey_t = \mu_y = \frac{\delta}{1 - \phi_1 - \phi_2}$. Si ahora sustituimos δ en la ecuación del modelo por $(1 - \phi_1 - \phi_2)\mu_y$, se tiene:

$$y_t - \mu_y = \phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t$$

Multiplicando por $y_{t-k} - \mu_y$ para $k = 0, 1, 2$ y tomando esperanzas se tiene que:

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2 \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 \end{aligned} \quad [13.4]$$

y, en general, multiplicando [13.3] por $y_{t-k} - \mu_y$ para algún $k > 0$ y tomando esperanzas se tiene:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, \quad k \geq 1$$

donde hemos utilizado el hecho de que la función de autocovarianza es simétrica, de modo que $\gamma_k = \gamma_{-k}$. El sistema de ecuaciones [13.4] puede resolverse para obtener:

$$\begin{aligned}\text{Var}(y_t) = \gamma_0 &= \frac{(1 - \phi_2)\sigma_\varepsilon^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \\ \gamma_1 &= \frac{\phi_1\gamma_0}{1 - \phi_2} \\ \gamma_2 &= \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2} \gamma_0\end{aligned}$$

que implican:

$$\begin{aligned}\rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \frac{\phi_1^2}{1 - \phi_2} + \phi_2\end{aligned}$$

y en general:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \quad k \geq 1$$

Partiendo de estimadores de σ_y^2 , ϕ_1 , ϕ_2 y σ_ε^2 , las ecuaciones [13.4] pueden utilizarse para obtener estimaciones de r_0 , r_1 y r_2 o, recíprocamente, como es más usual en la práctica, podemos comenzar de los valores estimados en la muestra de σ_y^2 , ρ_0 , ρ_1 y ρ_2 , y obtener estimaciones de ϕ_1 , ϕ_2 y σ_ε^2 . Esta estrategia dual se puede extender a modelos autorregresivos de orden p , $\text{AR}(p)$, para cualquier $p > 0$. El sistema de ecuaciones [13.4] constituye las llamadas ecuaciones de Yule-Walker, que pueden extenderse a procesos $\text{AR}(p)$.

Un proceso $\text{AR}(2)$ con parámetro ϕ_2 negativo puede generar raíces complejas en su ecuación característica $1 - \phi_1 L - \phi_2 L^2 = 0$. En tal caso, y_t presentará ciclos de *período* T dado por:

$$\cos \frac{2\pi}{T} = \frac{\phi_1}{2\sqrt{-\phi_2}}$$

y *factor de amortiguamiento*: $d = \sqrt{-\phi_2}$.

La función de autocorrelación de los procesos $\text{AR}(2)$ también converge exponencialmente a cero. Sin embargo, dicha convergencia puede presentar una gran variedad de aspectos. Puede ser siempre con signo positivo, pero también puede alternar en signo. Estas dos eran las posibilidades en los modelos $\text{AR}(1)$. Sin embargo, ahora puede haber también convergencia a cero siguiendo una curva sinusoidal. Esta diversidad, junto con el hecho de que para algunos valores de ϕ_1 y ϕ_2 la fase es similar a la de los modelos $\text{AR}(1)$, hace que sea preciso alguna herramienta adicional para identificar un pro-

ceso AR(2). En cuanto a su fap, es claro que las estimaciones de los parámetros ϕ_{11} y ϕ_{22} en las regresiones

$$\begin{aligned}\tilde{y}_t &= \phi_{11}\tilde{y}_{t-1} + u_t \\ \tilde{y}_t &= \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + \varepsilon_t\end{aligned}$$

serán no nulas, pero, sin embargo, la estimación de ϕ_{33} en

$$\tilde{y}_t = \phi_{31}\tilde{y}_{t-1} + \phi_{32}\tilde{y}_{t-2} + \phi_{33}\tilde{y}_{t-3} + u_t$$

será estadísticamente igual a cero, al igual que ϕ_{44} , ϕ_{55} , ... Esto implica que la fap de un proceso AR(2) es cero para valores $k > 2$. Lo importante es que a pesar de la variedad de formas que puede adoptar la fas dependiendo de los valores de los parámetros ϕ_1 y ϕ_2 , sin embargo, la propiedad que acabamos de mencionar para la fap es independiente de dichos valores.

Una propiedad similar es válida para todo modelo AR(p), donde p es cualquier entero positivo p : su función de autocorrelación parcial es cero para valores $k > p$.

13.4. MODELOS DE MEDIAS MOVILES

Analizamos en esta sección una clase diferente de procesos estocásticos, los procesos de medias móviles.

Definición 13.10. Se llama *proceso de medias móviles de orden 1*, que denotamos MA(1), a la estructura

$$y_t = \delta + \varepsilon_t - \theta\varepsilon_{t-1}$$

donde ε_t es un ruido blanco. Como primeras propiedades de este proceso se tiene inmediatamente de su definición que $Ey_t = \delta$ y también que $\text{Var } y_t = (1 + \theta^2)\sigma_\varepsilon^2$. En cuanto a la función de autocovarianza, se tiene:

$$\begin{aligned}\gamma_0 &= \sigma_y^2 = (1 + \theta^2)\sigma_\varepsilon^2 \\ \gamma_1 &= -\theta\sigma_\varepsilon^2 \\ \gamma_2 &= 0 \\ \gamma_k &= 0 \text{ para todo } k \geq 2\end{aligned}$$

y por tanto: $\rho_1 = -\frac{\theta}{1 + \theta^2}$; $\rho_2 = 0$, $\rho_k = 0$ para todo $k \geq 2$. La función

$g(\theta) = -\frac{\theta}{1 + \theta^2}$ es monótona decreciente en θ y, como consecuencia, puede verse fácilmente que el máximo valor absoluto que puede tomar ρ_1 en un modelo MA(1) es 0,50, y éste es el *único valor no nulo de su función de autocorrelación simple*, siendo negativo si $\theta > 0$, y positivo en caso contrario.

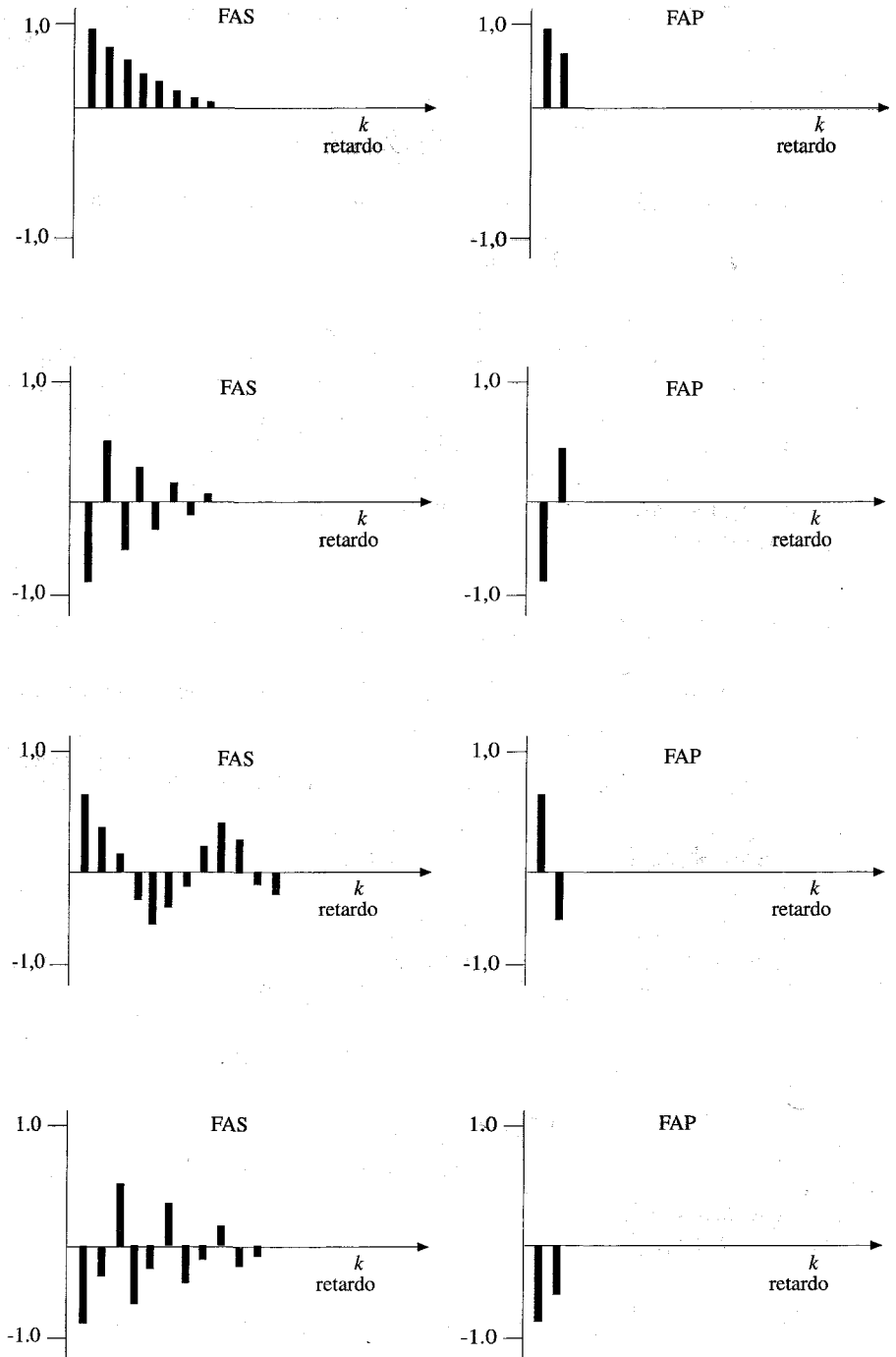


FIGURA 13.3. Algunos ejemplos de funciones de autocorrelación simple y parcial en modelos AR(2).

A partir del proceso MA(1) puede llevarse a cabo el siguiente esquema «de inversión»:

$$y_t = \delta + \varepsilon_t - \theta\varepsilon_{t-1}, \quad \text{que implica: } \varepsilon_t = y_t - \delta + \theta\varepsilon_{t-1}$$

Del mismo modo:

$$\varepsilon_{t-1} = y_{t-1} - \delta + \theta\varepsilon_{t-2}$$

$$\varepsilon_{t-2} = y_{t-2} - \delta + \theta\varepsilon_{t-3}$$

por lo que se tiene, sustituyendo:

$$\begin{aligned} \varepsilon_t &= y_t - \delta + \theta(y_{t-1} - \delta + \theta\varepsilon_{t-2}) = y_t + \theta y_{t-1} - \delta(1 + \theta) + \theta^2\varepsilon_{t-2} = \\ &= y_t + \theta y_{t-1} - \delta(1 + \theta) + \theta^2(y_{t-2} - \delta + \theta\varepsilon_{t-3}) = \\ &= y_t + \theta y_{t-1} + \theta^2 y_{t-2} - \delta(1 + \theta + \theta^2) + \theta^3\varepsilon_{t-3} \end{aligned}$$

Por tanto:

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} + \delta(1 + \theta + \theta^2) - \theta^3\varepsilon_{t-3} + \varepsilon_t \quad [13.5]$$

y el proceso continuaría, eliminando ahora el término ε_{t-3} .

Este procedimiento conduce, en el límite, a expresar y_t como función de su propio pasado, así como del valor contemporáneo del ruido blanco ε_t , pero sólo tiene sentido si $|\theta| < 1$. De otro modo, se tendría en [13.5] que el pasado de y_t tiene una gran importancia para determinar su comportamiento actual, y tanta más importancia cuanto más atrás en el tiempo. Cuando $|\theta| < 1$ se dice que el proceso MA(1) es *invertible* y podemos obtener su representación autorregresiva como límite del procedimiento descrito en [13.5]:

$$y_t = \frac{\delta}{1 - \theta} - \sum_{s=1}^{\infty} \theta^s y_{t-s} + \varepsilon_t$$

El hecho de que al invertir un proceso MA(1) se tengan coeficientes θ^s en los retardos de y_t sugiere, como así es, que la fap de un proceso MA(1) decae exponencialmente hacia cero, quizá alternando en signo.

Si el parámetro θ es negativo, entonces la fap converge a cero exponencialmente alternando en signo, y empezando de un valor positivo. Si, en cambio, el parámetro θ tiene signo positivo, entonces la convergencia va a ser con todos los valores de la fap. tomando signo negativo. Nótese, por tanto, que un proceso MA(1) no puede generar nunca una fap que sea siempre positiva, pero sí puede generar una función de autocorrelación parcial que es siempre negativa.

Definición 13.11. Un proceso de medias móviles de orden 2 es un proceso estocástico que sigue la ley:

$$y_t = \delta + \varepsilon_t + \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

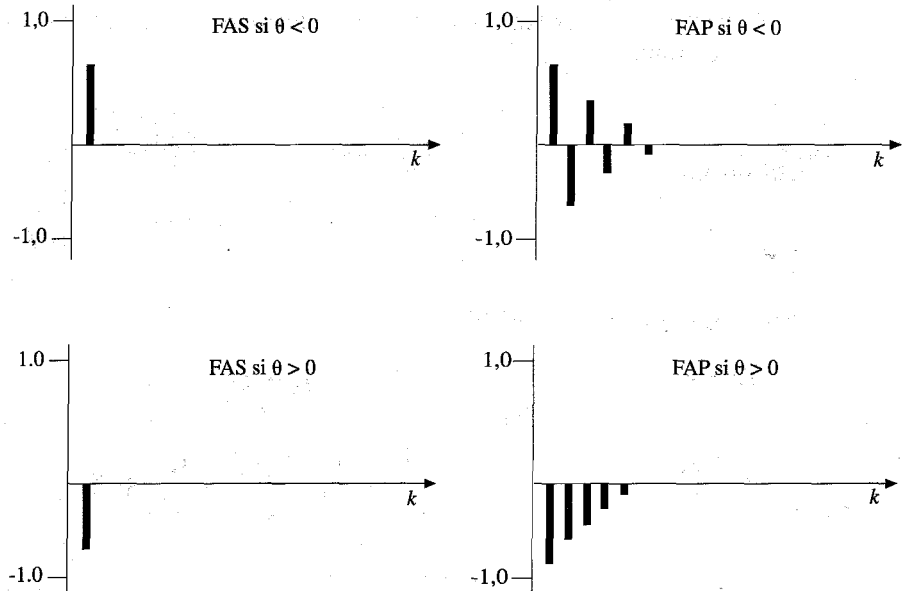


FIGURA 13.4. Ejemplos de funciones de autocorrelación simple y parcial de un proceso MA(1).

Siguiendo un proceso de inversión análogo al que hicimos con el proceso MA(1) se puede probar fácilmente que la fap de este proceso puede tener diversas formas, dependiendo de los signos y los valores relativos de los parámetros θ_1 y θ_2 . En cambio, la función de autocovarianza cumple:

$$\begin{aligned}
 * \quad \gamma_0 &= \sigma_y^2 = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2 \\
 \gamma_1 &= E[y_t y_{t-1}] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})] = \\
 &= E[\varepsilon_t \varepsilon_{t-1} - \theta_1 E \varepsilon_t \varepsilon_{t-2} - \theta_2 E \varepsilon_t \varepsilon_{t-3} - \theta_1 E \varepsilon_{t-1}^2 + \theta_1^2 E \varepsilon_{t-1} \varepsilon_{t-2} + \\
 &\quad + \theta_1 \theta_2 E \varepsilon_{t-1} \varepsilon_{t-3} - \theta_2 E \varepsilon_{t-1} \varepsilon_{t-2} + \theta_1 \theta_2 E \varepsilon_{t-2}^2 + \theta_2^2 E \varepsilon_{t-2} \varepsilon_{t-3}] = \\
 &= -\theta_1(1 - \theta_2)\sigma_\varepsilon^2
 \end{aligned}$$

donde se ha utilizado en repetidas ocasiones la ortogonalidad intertemporal del proceso de ruido blanco ε_t . Del mismo modo, se tiene:

$$\gamma_2 = E[y_t y_{t-2}] = E[-\theta_2 \varepsilon_{t-2} \varepsilon_{t-2}] = -\theta_2 \sigma_\varepsilon^2$$

mientras que $\gamma_k = 0$ para todo $k > 2$. Por tanto, la fas es:

$$\begin{aligned}
 \rho_1 &= -\frac{\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \\
 \rho_2 &= -\frac{\theta_2}{1 + \theta_1^2 + \theta_2^2} \\
 \rho_k &= 0 \quad \text{para todo } k > 2
 \end{aligned}$$

La identificación de un modelo MA(2) mediante su fap es bastante difícil [al igual que ocurre con las fas de un AR(2)], mientras que es muy sencillo identificar un modelo MA(2) por medio de su fas [igual que identificar un AR(2) mediante su fap]. Además los criterios de identificación son los mismos: la fas de un proceso MA(2) es cero para $k > 2$, de igual modo que la fap de un modelo AR(2) es cero para $k > 2$. Por otra parte, estos resultados son perfectamente generalizables para modelos de orden superior, AR(p) y MA(q), con p, q enteros positivos cualesquiera.

13.5. MODELOS ARMA

Hasta ahora hemos visto modelo de series temporales sencillos, en los que un proceso estocástico tenía una estructura autorregresiva «pura», o una estructura de medias móviles «pura». Sin embargo, en el análisis empírico de series temporales económicas es muy frecuente encontrar representaciones que tienen una componente autorregresiva así como una componente de medias móviles. Estos modelos se denotan como modelos ARMA(p, q), donde p y q denotan, respectivamente, los órdenes de los componentes autorregresivo y de medias móviles.

Definición 13.12. La estructura ARMA más sencilla es el modelo ARMA(1, 1):

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

que depende de cuatro parámetros desconocidos: ϕ , θ , σ_ε^2 y δ . El proceso ARMA(1, 1) es estacionario cuando $|\phi| < 1$, e invertible cuando $|\theta| < 1$. Distintos modelos pertenecientes a esta familia pueden escribirse sin más que variar los órdenes (p, q) de los componentes autorregresivo y de medias móviles.

La esperanza del proceso ARMA(1, 1) es $E(y_t) = \frac{\delta}{1 - \phi}$. Para determinar su varianza, supongamos que $\delta = 0$, lo que sin duda no afecta a la varianza del proceso, para obtener:

$$\begin{aligned} \text{Var}(y_t) = \gamma_0 &= E[(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})^2] = \phi^2 E(y_{t-1}^2) + 2\phi E(y_{t-1} \varepsilon_t) - \\ &- 2\theta \phi E(y_{t-1} \varepsilon_{t-1}) + E(\varepsilon_t^2) - 2\theta E(\varepsilon_t \varepsilon_{t-1}) + \theta^2 E(\varepsilon_{t-1}^2) = \\ &= \phi^2 \gamma_0 - 2\phi\theta E(y_{t-1} \varepsilon_{t-1}) + \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2 \end{aligned}$$

donde se han utilizado las siguientes propiedades: a) Si $|\phi| < 1$, el proceso ARMA(1, 1) es estacionario, y_{t-1} depende de ε_{t-1} y anteriores, pero no de sus valores futuros y, en particular, y_{t-1} es independiente de ε_t , y b) por ser ruido blanco, $E(\varepsilon_t \varepsilon_{t-1}) = 0$.

Como además, $E(y_{t-1} \varepsilon_{t-1}) = \sigma_\varepsilon^2$, se tiene de la expresión anterior que

$$\gamma_0(1 - \phi^2) = \sigma_\varepsilon^2(1 + \theta^2 - 2\phi\theta)$$

y, finalmente:

$$\gamma_0 = \frac{1 + \theta^2 - 2\phi\theta}{1 - \phi^2} \sigma_\varepsilon^2$$

Los distintos valores de la función de autocovarianza del proceso ARMA(1, 1) pueden obtenerse de un modo similar:

$$\gamma_1 = E[y_{t-1}(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})] = \phi \gamma_0 - \theta \sigma_\varepsilon^2 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 - \phi^2} \sigma_\varepsilon^2$$

$$\gamma_2 = E[y_{t-2}(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})] = \phi \gamma_1$$

y, en general, $\gamma_k = \phi \gamma_{k-1}$ para todo $k \geq 2$.

Como resultado, la función de autocorrelación del proceso ARMA(1, 1) es:

$$\rho_1 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

$$\rho_k = \phi \rho_{k-1} \quad \text{para todo } k \geq 2$$

En consecuencia, la función de autocorrelación de un proceso ARMA(1, 1) comienza del valor ρ_1 que acabamos de mostrar y, a partir de él, decrece a una tasa ϕ . Es decir, dicha función de autocorrelación se comporta, a partir de $k = 1$, como la función de autocorrelación de un proceso AR(1). Esta propiedad puede generalizarse: La función de autocorrelación de un proceso ARMA(p , q) se comporta como la función de autocorrelación del proceso AR(p), a partir de $k > q$.

Ello hace que la identificación de estos modelos por inspección de la serie temporal correspondiente no se ajuste a unas normas tan bien definidas como la identificación de modelos AR(p) o MA(q). Ello se debe a que tanto la función de autocorrelación como la función de autocorrelación parcial de los modelos ARMA heredan características de ambos componentes. Así puede probarse también que mientras que la función de autocorrelación parcial de un modelo AR(1) es cero para $k > 1$, sin embargo, la de un modelo ARMA(1, 1) no tendrá esta característica, pues a ella hay que superponer la de la componente MA(1), que genera, como sabemos, una función de autocorrelación parcial que converge exponencialmente a cero. Las mismas consideraciones pueden hacerse para la función de autocorrelación simple.

Sin embargo, en la práctica esta dificultad no resulta excesivamente importante, pues tampoco se trata de obtener la mejor identificación del modelo en un primer intento. Así, el modo más frecuente de terminar con una especificación ARMA(2, 1), por ejemplo, es haber comenzado con una especificación AR(2), y observar que los residuos obtenidos tras la estimación de tal modelo tienen una estructura MA(1) (o recíprocamente).

Es interesante hacer hincapié en que esta superposición de modelos tiene una fundamentación analítica. Si se ha especificado el modelo

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

y tras estimarlo e inspeccionar la función de autocorrelación de los residuos se observa que éstos parecen seguir una estructura MA(1):

$$\hat{u}_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

entonces estos dos modelos pueden unirse para obtener:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

es decir, una estructura ARMA(2, 1). En uno de los ejercicios al final del capítulo se pide al lector que lleve a cabo varios ejemplos similares a éste.

13.6. VARIABLES NO ESTACIONARIAS

Al presentar los modelos AR(p), supusimos que se satisfacían las condiciones necesarias para garantizar que el proceso en estudio era estacionario. Sin embargo, las series de datos económicos suelen caracterizarse por ser claramente no estacionarias. Como ya hemos citado, la observación de una tendencia lineal o cuadrática en el tiempo basta para rechazar el supuesto de estacionariedad.

Cuando esto ocurre en series económicas, es también cierto que si se toman primeras o segundas diferencias de la serie $\nabla y_t = y_t - y_{t-1}$, o $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$, se obtienen series transformadas que son estacionarias. Lo que se hace en tales situaciones es trabajar con las series en primeras o segundas diferencias, especificando y estimando un modelo para ellas. Como veremos más adelante, si se ha llevado a cabo un análisis de predicción para estas series, es bastante sencillo traducir las predicciones obtenidas para estas series diferenciadas en predicciones para las series originales, que son, en definitiva, aquellas en cuyo análisis estaba interesado el investigador.

Los procesos no estacionarios que pueden transformarse en estacionarios mediante sus diferencias de orden d se conocen como *procesos estocásticos no estacionarios, homogéneos de orden d* . Un ejemplo de tales procesos es el paseo aleatorio. Su varianza crece con el tiempo, lo que le hace no estacionario. Sin embargo, su primera diferencia es, por definición:

$$\nabla y_t = y_t - y_{t-1} = \varepsilon_t$$

un ruido blanco y, como tal, es un proceso estocástico estacionario.

Cuando se especifica un modelo ARMA(p, q) para la variable $\Delta^d y_t$, se dice que tenemos un modelo ARIMA(p, d, q) para y_t .

Si la evidencia de no estacionariedad proviene de que el gráfico de la serie temporal que se pretende analizar muestra fluctuaciones cuya amplitud cambia para distintos intervalos del período muestral, pensaremos igualmente que el proceso estocástico que generó la serie temporal no es estacionario. En este caso, la no estacionariedad surgiría porque la varianza de las diferentes va-

riables aleatorias que lo integran a lo largo del tiempo no son iguales entre sí.

El procedimiento habitual en esta situación consiste en transformar la variable tomando logaritmos, y pasar a analizar esta variable transformada. Veremos en la sección dedicada a predicción cómo recuperar las predicciones de la serie original a partir de las predicciones obtenidas para la serie en logaritmos. Es importante hacer notar que la transformación logarítmica no va a corregir el problema de heteroscedasticidad a que nos referimos, sino que simplemente lo va a amortiguar, hasta el punto de hacerlo apenas perceptible. En este sentido, esta transformación es conceptualmente diferente de la diferenciación en el caso de no estacionariedad en la esperanza matemática que hemos visto antes. Por otra parte, la transformación logarítmica persigue estabilizar la varianza de la variable, mientras que la diferenciación busca estabilizar su esperanza matemática.

También debe notarse que la transformación Box-Cox que analizamos en el Capítulo 11 incluye a la logarítmica como caso particular y que, en ocasiones, podría considerarse otra transformación de la familia Box-Cox.

13.7. ESTACIONARIEDAD E INVERTIBILIDAD

Hay varias razones importantes para pretender que un proceso estocástico con el que se va a efectuar trabajo empírico sea estacionario: En primer lugar, si no es estacionario en media, por ejemplo, ello quiere decir que la esperanza matemática de las variables del proceso cambia con el tiempo, y entonces habría que estimar un número infinito de parámetros. Aún más importante, un proceso puede ser no estacionario porque sus momentos, o su distribución, no existan. Lo mismo ocurriría con la varianza.

Un modelo AR(1) puede desarrollarse:

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t = \dots = \\ &= \phi^t y_0 + \phi^{t-1} \varepsilon_1 + \dots + \phi \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

Vimos anteriormente que el interés de que un proceso AR(1) sea estacionario se basa en que, de lo contrario, con $|\phi| > 1$ se tendría que la expresión anterior es divergente, como puede apreciarse en el hecho de que los coeficientes de las variables aleatorias en la serie son crecientes sin límite.

Estos comentarios pueden extenderse a cualquier proceso AR(k). Por ejemplo, un proceso AR(2), $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$, es estacionario si al descomponer su *ecuación característica* en factores

$$1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

donde L denota al operador de retardos, ambos parámetros, λ_1 y λ_2 , tienen módulo menor que 1. Bastaría, por el contrario, que uno de ellos fuese mayor que 1 en valor absoluto, para que la realización de y_t dependiese de valores futuros de las variables ε_t , y quizá también de sus valores pasados. Ya vimos

en la Sección 13.3 que las condiciones sobre los parámetros para que ello no ocurra son:

$$|\phi_2| < 1, \quad \phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1$$

Por otra parte, todo proceso MA de orden finito, por ser una combinación lineal de un número finito de variables aleatorias con distribución Normal, tiene asimismo una distribución Normal, con esperanza cero y una varianza que depende de los coeficientes del modelo, pero que es independiente del tiempo y es, por tanto, siempre un proceso estocástico estacionario.

En el caso de procesos de medias móviles, las condiciones similares a las de estacionariedad son las de invertibilidad, pero tienen una interpretación diferente: Cuando un proceso de medias móviles es invertible, entonces dicho modelo admite una representación autorregresiva, en que los valores pasados de la variable y_t reciben una ponderación (coeficientes) cada vez menor cuanto más alejados en el tiempo. Esta propiedad de un modelo de medias móviles es fundamental a efectos de predicción, para lo que, como veremos, es necesario invertir primero el proceso MA y obtener su representación AR.

Un proceso ARMA(p, q) es estacionario si lo es su componente AR, y es invertible si lo es su componente MA.

13.8. ESTACIONALIDAD

Las series temporales de datos económicos presentan generalmente características estacionales cuando se observan a una frecuencia inferior a la anual, ya sea mediante datos trimestrales o mensuales. Estas características se deben a que las decisiones tomadas por los agentes económicos en un determinado trimestre del año pueden estar correlacionadas con las decisiones tomadas en el mismo trimestre de otros años. Algo similar ocurre, en general, con datos mensuales.

Tales correlaciones pueden representarse analíticamente por modelos univariantes totalmente análogos a los vistos en las secciones anteriores. Así, la especificación de las características estacionales de una serie pudiera ser del tipo

$$y_t = \phi_{12}y_{t-12} + e_t$$

si además existen correlaciones entre las observaciones de meses consecutivos, se tendrá:

$$e_t = \phi e_{t-1} + \varepsilon_t$$

y, en resumen:

$$(1 - \phi L)(1 - \phi_{12}L^{12})y_t = (y_t - \phi_{12}y_{t-12}) - \phi(y_{t-1} - \phi_{12}y_{t-13}) = \varepsilon_t$$

La especificación del modelo univariante para la *estructura estocástica estacional* de una serie temporal se lleva a cabo de modo idéntico al descrito

en las secciones anteriores para la identificación de un modelo para su estructura estocástica regular, con la salvedad de que para ello se examinan los valores «estacionales» de las funciones de autocorrelación, $\rho_4, \rho_8, \rho_{12}, \dots$ si los datos son trimestrales, o $\rho_{12}, \rho_{24}, \rho_{36}, \dots$ si son mensuales.

Así, por ejemplo, una serie puede requerir diferencias de orden estacional si los valores estacionales de su función de autocorrelación no tienden a cero rápidamente. Si trabajando con datos trimestrales se tiene que tanto ρ_4 como ρ_8 y ρ_{12} son significativamente diferentes de cero, será en general apropiado transformar la variable mediante la diferencia:

$$z_t = \nabla_4 y_t = y_t - y_{t-4}$$

Así, una estructura posible para una serie mensual podría ser:

$$\nabla \nabla_{12} \ln y_t = (1 - \theta L) (1 - \Theta L^{12}) \varepsilon_t$$

que indica que la serie precisó de la transformación logarítmica, así como de una diferencia, tanto de carácter regular como estacional, esta última sugerida, sin duda, porque los valores de la fas para $k = 12, 24, 36$ no convergían a cero. Tras estas transformaciones, los únicos valores no nulos de la fas eran los correspondientes a $k = 1$ y $k = 12$, por lo que se especificó el modelo descrito.

13.9. PREDICCIÓN CON MODELOS ARIMA

Es un resultado conocido en Estadística que cuando se pretende que la predicción \hat{y}_{T+k}^T elaborada en el instante T acerca del valor de la variable endógena en el instante $T+k$ minimice la función de pérdida $E_T[(y_{T+k} - \hat{y}_{T+k}^T)^2]$, entonces la predicción debe ser $\hat{y}_{T+k}^T = E_T y_{T+k}$, es decir, precisamente la esperanza condicional de la variable y_{T+k} , calculada sobre la base de la información disponible en el instante T (véase Problema 13.4). Este es el criterio que utilizaremos en esta sección.

13.9.a. Modelos autorregresivos

En un modelo AR(1) se tiene:

$$E_T y_{T+1} = E_T(\delta + \phi y_T + \varepsilon_{T+1}) = \delta + \phi y_T + E_T \varepsilon_{T+1} = \delta + \phi y_T$$

donde hemos utilizado dos resultados:

a) Que y_T está en el conjunto de información disponible en el instante T , es decir, que la persona que realiza la predicción observa el valor y_T previamente al cálculo de la predicción, con lo que $E_T y_T = y_T$.

b) $E_T \varepsilon_{T+1} = 0$ por ser ε_t un ruido blanco. De modo similar:

$$\begin{aligned} E_T y_{T+2} &= E_T(\delta + \phi y_{T+1} + \varepsilon_{T+2}) = \delta + \phi E_T y_{T+1} + E_T \varepsilon_{T+2} = \\ &= \delta + \phi E_T y_{T+1} = \delta(1 + \phi) + \phi^2 y_T \\ E_T y_{T+k} &= \phi^k y_T + \delta(1 + \phi + \phi^2 + \dots + \phi^{k-1}) \end{aligned}$$

Notemos que:

$$1 + \phi + \phi^2 + \dots + \phi^{k-1} = \frac{1 - \phi^k}{1 - \phi}$$

y recordando que $Ey = \delta/(1 - \phi)$ se tiene:

$$\delta(1 + \phi + \phi^2 + \dots + \phi^k) = \left[\frac{1 - \phi^k}{1 - \phi} \right] (1 - \phi) Ey = (1 - \phi^k) Ey$$

por lo que se tiene finalmente:

$$E_T y_{T+k} = \phi^k y_T + (1 - \phi^k) Ey$$

de modo que la predicción óptima k periodos hacia el futuro es una combinación lineal convexa de dos términos:

- la última observación obtenida de la variable y_t ;
- su esperanza matemática.

Según avanzamos hacia el futuro, la última observación recibe una ponderación más pequeña, y la esperanza matemática recibe un peso más importante. Ello refleja el hecho de que, en un proceso estacionario, cuanto más lejos hacia el futuro queremos predecir, mayor será la incertidumbre bajo la que se obtiene la predicción: En un régimen de total incertidumbre, la predicción óptima del valor de una variable aleatoria es igual a su esperanza matemática. Por el contrario, cuando predecimos a horizontes cortos, la información muestral permite mejorar la predicción que se haría si se utilizase únicamente la esperanza matemática de la variable aleatoria. La propiedad característica del proceso AR(1) es que toda la información muestral relevante para la predicción quede resumida en el último valor observado de la variable; en consecuencia, dicho valor es todo lo que se precisa, junto con la esperanza matemática del proceso, para elaborar predicciones.

En un proceso AR(2) se tiene:

$$\begin{aligned} E_T y_{T+1} &= E_T(\delta + \phi_1 y_T + \phi_2 y_{T-1} + \varepsilon_{T+1}) = \delta + \phi_1 y_T + \phi_2 y_{T-1} \\ E_T y_{T+2} &= \delta + \phi_1 E_T y_{T+1} + \phi_2 y_T = \delta(1 + \phi_1) + (\phi_1^2 + \phi_2) y_T + \phi_1 \phi_2 y_{T-1} \\ E_T y_{T+3} &= \delta(1 + \phi_1 + \phi_2 + \phi_1^2) + \phi_1(\phi_1^2 + 2\phi_2) y_T + (\phi_1^2 \phi_2 + \phi_2^2) y_{T-1} \end{aligned}$$

y así sucesivamente. En procesos AR(p) de orden superior a 1 no existe una forma analítica sencilla para la predicción k periodos hacia el futuro. A pesar

de ello, es fácil observar una diferencia con respecto a la predicción con modelos AR(1), y es que la información muestral relevante para la predicción se resume en las p últimas observaciones de la variable a predecir.

13.9.b. Modelos de medias móviles

Un proceso MA(1) invertible puede escribirse:

$$y_{T+1} = -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} - \dots + \varepsilon_{T+1}$$

por lo que:

$$E_T y_{T+1} = -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} + \dots$$

$$E_T y_{T+2} = -\theta E_T y_{T+1} - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots =$$

$$= -(-\theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots) - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots = 0$$

$$E_T y_{T+k} = 0 \quad \text{para } k \geq 2$$

es decir, que la predicción de un proceso MA(1) para 2 o más períodos hacia el futuro es cero. Es fácil ver las razones que explican este resultado, puesto que el modelo

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

implica que

$$y_{T+1} = \varepsilon_{T+1} - \theta \varepsilon_T, \quad y_{T+2} = \varepsilon_{T+2} - \theta \varepsilon_{T+1}$$

y así sucesivamente.

Como ε_t es ruido blanco, se tiene $E_T \varepsilon_{T+1} = 0$, y más generalmente, $E_T \varepsilon_{T+k} = 0$ para todo $k > 0$. En consecuencia, $E_T y_{T+2} = 0$, y también $E_T y_{T+s} = 0$ para todo $s \geq 2$.

Análogamente, puede verse que en un modelo MA(2), $E_T y_{T+k} = 0$ para $k \geq 3$. Más generalmente, en un modelo MA(q), $E_T y_{T+s} = 0$ para todo $s > q$.

13.9.c. El modelo ARMA(1, 1)

Las expresiones analíticas correspondientes a las predicciones de valores futuros del modelo $y_t = \delta + \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$, son:

$$E_T y_{T+1} = \delta + \phi y_T - \theta \varepsilon_T$$

$$E_T y_{T+2} = \delta + \phi E_T y_{T+1} = (1 + \phi)\delta + \phi^2 y_T - \phi \theta \varepsilon_T$$

y, finalmente:

$$E_T y_{T+k} = (1 + \phi + \phi^2 + \dots + \phi^{k-1})\delta + \phi^k y_T - \phi^{k-1} \theta \varepsilon_T$$

donde puede verse que, como ocurre para todo proceso estacionario:

$$\lim_{k \rightarrow \infty} E_T y_{T+k} = \frac{\delta}{1 - \phi} = E(y_t)$$

la predicción converge a la esperanza matemática del proceso cuando el horizonte de predicción tiende a infinito.

13.9.d. Error de predicción

El error de predicción es la diferencia entre la realización de la variable aleatoria y la predicción hecha para dicho valor. Por supuesto, el error cometido en la predicción de y_{T+k} depende del periodo en que dicha predicción se hizo.

Denotamos por $e_T(k)$ el error de predicción k periodos hacia adelante, elaborada en el período T , es decir: $e_T(k) = y_{T+k} - E_T y_{T+k}$. En todos estos modelos, la esperanza del error de predicción es cero, precisamente, por ser el predictor óptimo la esperanza condicional de la variable a predecir. En efecto, se tiene:

$$E(e_T(k)) = E(y_{T+k} - E_T y_{T+k}) = E y_{T+k} - E[E_T y_{T+k}] = E y_{T+k} - E y_{T+k} = 0$$

Siendo el error de predicción $e_T(k)$ una variable aleatoria de esperanza cero, su varianza nos da una medida de su tamaño. En este sentido, una de las variables de más interés en el análisis de predicción es, por tanto, *la varianza del error de predicción*. Dicha varianza (más exactamente, su raíz cuadrada) es la cantidad que se utiliza para construir los intervalos de confianza de las predicciones puntuales obtenidas del modo descrito en los párrafos anteriores.

De acuerdo con dichas expresiones para las predicciones óptimas se tienen los siguientes errores y varianzas de error de predicción (que el lector debe asegurarse de entender):

a) Errores de predicción en un modelo AR(1):

$$e_T(1) = y_{T+1} - E_T y_{T+1} = (\delta + \phi y_T + \varepsilon_{T+1}) - (\delta + \phi y_T) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + \phi \varepsilon_{T+1}$$

.....

$$e_T(k) = \varepsilon_{T+k} + \phi \varepsilon_{T+k-1} + \phi^2 \varepsilon_{T+k-2} + \dots + \phi^{k-1} \varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \phi^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2(1 + \phi^2 + \phi^4)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)}) = \sigma_\varepsilon^2 \left[\frac{1 - \phi^{2k}}{1 - \phi^2} \right]$$

b) Errores de predicción en un modelo AR(2):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} + \phi_1 \varepsilon_{T+2} + (\phi_1^2 + \phi_2) \varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \phi_1^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2[1 + \phi_1^2 + (\phi_1^2 + \phi_2)^2]$$

c) Errores de predicción en un modelo MA(1):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} - \theta \varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} - \theta \varepsilon_{T+k-1} \quad \text{para todo } k > 1$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \theta^2)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \theta^2) \quad \text{para todo } k > 1$$

d) Errores de predicción en un modelo MA(2):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} - \theta_1 \varepsilon_{T+2} - \theta_2 \varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} - \theta_1 \varepsilon_{T+k-1} - \theta_2 \varepsilon_{T+k-2} \quad \text{para todo } k > 2$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \theta_1^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2) \quad \text{para todo } k > 2$$

e) Errores de predicción en un modelo ARMA(1, 1):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + (\phi - \theta)\varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} + (\phi - \theta)\varepsilon_{T+2} + \phi(\phi - \theta)\varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} + (\phi - \theta)\varepsilon_{T+k-1} + \phi(\phi - \theta)\varepsilon_{T+k-2} + \phi^2(\phi - \theta)\varepsilon_{T+k-3} + \dots + \phi^{k-2}(\phi - \theta)\varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = [1 + (\phi - \theta)^2]\sigma_\varepsilon^2$$

$$\text{Var}(e_T(3)) = [1 + (\phi - \theta)^2(1 + \phi^2)]\sigma_\varepsilon^2$$

$$\text{Var}(e_T(k)) = \left[1 + (\phi - \theta)^2 \frac{1 - \phi^{2(k-1)}}{1 - \phi^2} \right] \sigma_\varepsilon^2$$

que, al igual que en los modelos AR(p) y MA(q), tiende a la varianza del proceso ARMA(1, 1) cuando el horizonte de predicción k tiende a infinito.

Obsérvese que las amplitudes del intervalo de confianza para las sucesivas predicciones elaboradas con un modelo AR crecen continuamente con el horizonte de predicción, mientras que para un modelo MA(q) permanecen constantes a partir de q períodos hacia el futuro.

El lector puede notar asimismo que el error de predicción un período hacia el futuro es el mismo, ε_{T+1} , independientemente de cual sea el verdadero modelo de la serie temporal en estudio. Este es un resultado lógico, pues dicho error es la componente del valor futuro de la serie que no puede predecirse sobre la base de la información muestral, lo que en la literatura de series temporales se conoce como *innovación en la serie*.

Ello no quiere decir, sin embargo, que el error de predicción de una serie temporal un período hacia el futuro vaya a ser independiente del modelo utilizado para obtener dicha predicción. El error de predicción, por definición, es la diferencia entre la realización de la serie y su *predicción óptima*, obtenida con el *mejor* modelo que pueda elaborarse sobre la base de la información muestral. En la práctica, estas diferencias se reflejan en diferentes valores estimados del parámetro σ_ε^2 según se ajuste un modelo u otro. Dicho de otro modo, si no se utiliza el *mejor* modelo para la serie, entonces el error de predicción incluirá, junto con la innovación ε_t , un error de especificación ξ_t y, agregadamente, tendrán una varianza superior a la de la auténtica innovación.

13.9.e. Intervalos de confianza para las predicciones

La varianza del error de predicción puede utilizarse para obtener *intervalos de confianza de las predicciones* elaboradas, mediante la expresión:

$$E_T y_{T+k} \pm \lambda_\alpha \hat{\sigma}_{e_T(k)}$$

donde, si se supone que la innovación ε_t sigue una distribución Normal, el parámetro λ_α se obtendrá de las tablas de dicha distribución, al nivel de confianza α elegido. Así, por ejemplo, para obtener en el instante T un intervalo del 95 por 100 de confianza para el valor del proceso y_t en el instante $T+k$, supuesto que la innovación ε_{T+k} siga una distribución Normal, basta sumar y restar 1,96 veces el valor estimado de $\sigma_{e_T(k)}$ a la predicción $E_T y_{T+k}$. Como es habitual, este intervalo será tan sólo una aproximación, por haber estimado el parámetro $\hat{\sigma}_e$.

13.9.f. Predicción de una serie en diferencias

Si se ha estimado un modelo ARIMA con un número no nulo de diferencias, entonces será preciso recuperar las predicciones de la serie original a partir de las predicciones elaboradas para la serie en diferencias. Ello puede hacerse del siguiente modo: Supongamos que y_t denota la serie en cuyo análisis estamos interesados, y que se ha especificado y estimado un modelo univariante para la serie de primeras diferencias: $z_t = \nabla y_t$. Entonces, es claro que:

$$E_T z_{T+k} = E_T y_{T+k} - E_T y_{T+k-1}$$

por lo que:

$$\begin{aligned} E_T y_{T+k} &= E_T z_{T+k} + E_T y_{T+k-1} = \dots = \\ &= E_T z_{T+k} + E_T z_{T+k-1} + E_T z_{T+k-2} + \dots + E_T z_{T+1} + y_T \end{aligned}$$

y, consecuentemente, para obtener las predicciones de la serie y_t basta añadir a su última observación muestral las predicciones elaboradas para sus incrementos, z_{T+k} .

El lector puede comprobar, como se pide en uno de los ejercicios al final del capítulo, que la recuperación de las predicciones de una serie y_t a partir de las predicciones elaboradas para la serie de segundas diferencias $\nabla^2 y_t$ puede llevarse a cabo de un modo similar al que acabamos de describir.

Por último cabe citar que si el modelo univariante se ha estimado para la transformación logarítmica de la variable original $z_t = \ln(y_t)$, entonces el modo de recuperar las predicciones de los valores futuros de y_t es:

$$E_T y_{T+k} = \exp \left\{ E_T z_{T+k} + \frac{1}{2} \text{Var } e_T(k) \right\}$$

Dependiendo del tamaño de la varianza del error de predicción, la expresión anterior supondrá una diferencia significativa o no con respecto a la simple alternativa de hallar la función exponencial de los valores $E_T z_{T+k}$.

Sin embargo, los límites inferior y superior de los intervalos de confianza para $E_T y_{T+k}$ deben hallarse calculando los valores de la función exponencial en los límites inferior y superior del intervalo de $E_T z_{T+k}$. En consecuencia, el intervalo de confianza que se obtenga para la predicción $E_T y_{T+k}$ no será simétrico alrededor de dicha predicción.

Nótese, por otra parte, que las variaciones en las predicciones de z_{T+k} para valores sucesivos de k pueden interpretarse como variaciones porcentuales previstas en la variable original y_{t+k} .

13.10. ESTIMACION DE MODELOS ARIMA

13.10.a. Estimación de modelos autorregresivos

Un modelo autorregresivo presenta una diferencia notable con respecto a los modelos econométricos que hasta ahora hemos considerado. Las variables explicativas son ahora aleatorias, ya que son retardos de la variable y_t , que es aleatoria. Como hemos visto en el Capítulo 9, puede probarse que, en tales condiciones, el estimador MCO tiene buenas propiedades y, en particular, es un estimador consistente siempre que las variables explicativas x_{it} satisfagan la condición $E(x_{i,t-s}u_t) = 0$.

Si el término de error no tiene autocorrelación y si el modelo es estacionario, esta propiedad se satisface. En efecto, los valores $x_{i,t-s}$ de la condición anterior son ahora retardos de y_t , mientras que u_t es ahora ε_t . Como hemos visto en dicho capítulo, bajo el supuesto de estacionariedad, la variable y_t depende de ε_t y de sus valores anteriores, pero de ningún valor futuro de ε_t . Por tanto, las esperanzas $E(y_{t-s}\varepsilon_t)$ son cero para todo $s > 0$.

En tal caso, la estimación consistente del modelo autorregresivo

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

puede llevarse a cabo por MCO. Por tanto, todo depende de que la especificación sea correcta.

Si, por el contrario, el término de error del modelo tuviese autocorrelación, entonces la condición de ortogonalidad no se cumpliría y el estimador MCO dejaría de ser apropiado. En efecto, si ε_t no fuese ruido blanco sino que obedeciese, por ejemplo, al modelo $\varepsilon_t = \phi\varepsilon_{t-1} + \xi_t$, con ξ_t ruido blanco, entonces se tendría que: a) y_{t-1} estaría correlacionado con ε_{t-1} , a través del modelo univariante elaborado, y b) ε_t también estaría correlacionado con ε_{t-1} , a través de su modelo de autocorrelación. En estas condiciones, $E(y_{t-1}\varepsilon_t) \neq 0$, ya que ambos están correlacionados con ε_{t-1} , contradiciendo la condición de ortogonalidad necesaria para justificar el uso del estimador MCO.

La autocorrelación del término de error de un modelo univariante es un

indicio evidente de mala especificación de dicho modelo. Una especificación correcta debe generar un término de error con estructura de ruido blanco.

13.10.b. Estimación de modelos de medias móviles

Como ejemplo de la estimación de modelos de medias móviles, discutiremos las cuestiones importantes en el contexto de un modelo MA(2): $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$. Para ello, si suponemos que la innovación ε_t sigue una distribución Normal $(0, \sigma_\varepsilon^2)$, se tiene la siguiente expresión para el logaritmo de la función de verosimilitud:

$$\log L(\theta_1, \theta_2, \sigma_\varepsilon^2) = \text{Constante} - \frac{T}{2} \log \sigma_\varepsilon^2 - \frac{\text{SR}(\theta_1, \theta_2)}{2\sigma_\varepsilon^2}$$

donde $\text{SR}(\theta_1, \theta_2)$ denota la suma residual: $\text{SR}(\theta_1, \theta_2) = \sum_3^T \varepsilon_t^2$, de modo que, condicional en el valor del parámetro σ_ε^2 , la maximización de la función de verosimilitud coincide con la minimización de la suma residual. Se trataría, por tanto, de minimizar la suma de cuadrados de los residuos:

$$\text{mín } \text{SR}(\theta_1, \theta_2) = \sum_3^T \varepsilon_t^2 = \sum_3^T (y_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})^2$$

donde, como puede verse, se están ignorando los dos primeros períodos de la muestra, pues no se dispone para ellos de observaciones acerca del ruido blanco ε_t . La solución a este problema de minimización proporciona los valores estimados de los parámetros del modelo y, finalmente, la varianza σ_ε^2 se estima mediante:

$$\sigma_\varepsilon^2 = \frac{\text{SR}(\hat{\theta}_1, \hat{\theta}_2)}{T - p - q}$$

donde $p + q$ es el número de parámetros estimados en el modelo univariante. en este caso 2.

La minimización de la suma residual anterior podría llevarse a cabo mediante una red de búsqueda. Este procedimiento sería muy sencillo en el caso de un proceso MA(1), puesto que el espacio paramétrico admisible sería $(-1, 1)$, y bastaría hacer una partición de este intervalo, generar la serie de residuos $\{\hat{\varepsilon}_t\}$ tomando como dadas las observaciones y_1 y y_2 , y evaluando la función $\text{SR}(\theta)$. La certeza de haber hallado un mínimo global en vez de un mínimo local, aumenta si la partición que se lleva a cabo es suficientemente fina, o si se lleva a cabo una exploración bastante exhaustiva del espacio paramétrico. En el caso de un proceso MA(2), el proceso de búsqueda es más complejo, pues el espacio paramétrico tiene dos dimensiones. Sin embargo, dicho espacio queda limitado por las condiciones de invertibilidad del proceso que vimos anteriormente.

Como alternativa, dados unos valores iniciales de los parámetros θ_1 y θ_2 , la expresión

$$\varepsilon_t = y_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad [13.6]$$

podrá utilizarse para generar «observaciones» de ε_t para $t = 3, 4, \dots, T$. Claramente, los valores numéricos de los residuos $\hat{\varepsilon}_t \geq 3$ varían con los valores de los parámetros θ_1 y θ_2 , y se trata de encontrar el par de valores (θ_1, θ_2) para el que la suma de cuadrados de los residuos es mínima. Es imposible utilizar todos los pares de valores paramétricos posibles y calcular para cada uno de ellos la suma residual correspondiente, por lo que es preciso recurrir a un procedimiento analítico, que pasamos a describir.

Denotamos por (θ_1^0, θ_2^0) los valores iniciales de los parámetros y por ε_t^0 , $t \geq 3$ los residuos obtenidos con el par de valores (θ_1^0, θ_2^0) . El verdadero valor del residuo ε_t es, sin embargo, el correspondiente a los verdaderos valores de los parámetros θ_1 y θ_2 , y se puede llevar a cabo la siguiente aproximación por medio del desarrollo en serie de Taylor:

$$\begin{aligned} \varepsilon_t = & \varepsilon_t^0 + (\theta_1 - \theta_1^0) \left[\frac{\partial \varepsilon_t}{\partial \theta_1} \right]_{\theta = \theta^0} + (\theta_2 - \theta_2^0) \left[\frac{\partial \varepsilon_t}{\partial \theta_2} \right]_{\theta = \theta^0} + \\ & + \frac{1}{2} (\theta_1 - \theta_1^0)^2 \left[\frac{\partial^2 \varepsilon_t}{\partial \theta_1^2} \right]_{\theta = \theta^0} + \frac{1}{2} (\theta_1 - \theta_1^0)(\theta_2 - \theta_2^0) \left[\frac{\partial^2 \varepsilon_t}{\partial \theta_1 \partial \theta_2} \right]_{\theta = \theta^0} + \\ & + \frac{1}{2} (\theta_2 - \theta_2^0)^2 \left[\frac{\partial^2 \varepsilon_t}{\partial \theta_2^2} \right]_{\theta = \theta^0} + \dots \end{aligned} \tag{13.7}$$

donde ya hemos despreciado los términos de orden superior a 2, en la confianza de que tanto las potencias $(\theta_1 - \theta_1^0)^3$ como las derivadas parciales de esos órdenes serán pequeñas en valor absoluto. Si por idénticas razones despreciamos asimismo los términos de segundo orden que aparecen en [13.7], se llega a

$$\varepsilon_t = \varepsilon_t^0 + (\theta_1 - \theta_1^0) \left[\frac{\partial \varepsilon_t}{\partial \theta_1} \right]_{\theta = \theta^0} + (\theta_2 - \theta_2^0) \left[\frac{\partial \varepsilon_t}{\partial \theta_2} \right]_{\theta = \theta^0}$$

y puesto que en este modelo

$$\frac{\partial \varepsilon_t}{\partial \theta_1} = \varepsilon_{t-1} \quad \text{y} \quad \frac{\partial \varepsilon_t}{\partial \theta_2} = \varepsilon_{t-2}$$

se tiene:

$$\varepsilon_t^0 - \theta_1^0 \varepsilon_{t-1}^0 - \theta_2^0 \varepsilon_{t-2}^0 = -\theta_1 \varepsilon_{t-1}^0 - \theta_2 \varepsilon_{t-2}^0 + \varepsilon_t$$

que muestra que se pueden obtener estimaciones de los parámetros θ_1 y θ_2 sin más que estimar el modelo lineal:

$$w_t = \theta_1 x_{1t} + \theta_2 x_{2t} + \varepsilon_t \tag{13.8}$$

donde $w_t = \varepsilon_t^0 - \theta_1^0 \varepsilon_{t-1}^0 - \theta_2^0 \varepsilon_{t-2}^0$, $x_{1t} = -\varepsilon_{t-1}^0$, $x_{2t} = -\varepsilon_{t-2}^0$.

El lector debe notar que este procedimiento no es sino la aplicación del algoritmo Gauss-Newton a la minimización de la suma residual condicional.

La estimación del modelo [13.8] requiere de la previa construcción de estas variables. Los valores θ_i^0 , $i = 1, 2$ que en ellas aparecen son los valores iniciales de los parámetros que el investigador debe seleccionar. Pongamos, por ejemplo, que $\theta_i^0 = 0,10$, aunque más adelante daremos unas reglas más rigurosas para la elección de estos valores iniciales. Por otra parte, la relación [13.6] puede utilizarse para generar «datos» para la variable $\hat{\varepsilon}_t^0$, para $t = 3, 4, \dots, T$, utilizando los valores iniciales de los parámetros. La estimación del modelo [13.8] se llevaría a cabo con $T - 3$ observaciones.

El procedimiento descrito puede y debe iterarse, utilizando como valores (θ_1^0, θ_2^0) en una segunda etapa las estimaciones obtenidas en la primera etapa. El esquema iterativo podría terminar cuando las diferencias entre los valores paramétricos iniciales y finales de una etapa del proceso son pequeñas. En tal caso, se dice que el proceso de estimación numérica que hemos descrito ha convergido (supuestamente a los verdaderos valores de los parámetros). Como posible criterio de convergencia, podría decidirse detener el proceso si las variaciones en todos los parámetros son inferiores a 0,001, o al 5 por 100 de su valor inicial.

Un criterio de convergencia alternativo consistiría en detener el proceso si la variación producida en la suma residual (recordemos que ésta es la función objetivo del problema de optimización) es pequeña. De nuevo, habría que definir lo que se entiende por «pequeña», pero podría ser una diferencia inferior al 1 por 100. Un criterio más exigente sería el de aceptar la convergencia del proceso sólo cuando se cumplen simultáneamente los dos criterios que acabamos de citar.

Por supuesto, también existe la posibilidad de que el proceso anterior no converja. Ello puede deberse al hecho de que el procedimiento de estimación se basa en una aproximación lineal a la función que hace depender ε_t de los parámetros del modelo, y no en la verdadera relación entre ambos. Otra posibilidad es que los valores iniciales escogidos para los parámetros no hayan sido muy adecuados por lo que, cuando esto ocurre, debe repetirse el procedimiento con otros valores iniciales. Finalmente, la convergencia puede no producirse por una mala especificación del modelo, es decir, porque el modelo univariante que se está tratando de estimar no sea el modelo lineal que mejor representa la estructura del proceso estocástico que generó la serie temporal que está siendo objeto de análisis.

El procedimiento que acabamos de describir es válido, con las modificaciones obvias, para la estimación de cualquier modelo univariante. Al estimar el modelo

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

se ignoran los primeros q residuos, lo que equivale a suponerlos iguales a su esperanza, cero. Por ello, este procedimiento de estimación se conoce como *mínimos cuadrados condicionados* al supuesto de que los primeros residuos son iguales a cero. Si la muestra consta de pocas observaciones, el número

de grados de libertad (número de observaciones menos número de parámetros a estimar) puede ser muy pequeño, e ignorar las primeras observaciones en la estimación será una mala aproximación.

13.10.c. Obtención de valores iniciales para los parámetros del modelo

Para llevar a cabo el procedimiento analítico de estimación que hemos descrito en los párrafos anteriores es preciso haber identificado previamente un modelo para la serie temporal en estudio. Esta identificación se habrá llevado a cabo mediante el análisis de sus funciones de autocorrelación y de autocorrelación parcial. Los valores numéricos de las estimaciones muestrales de estas funciones pueden utilizarse para obtener estimaciones iniciales de los parámetros del modelo.

Por ejemplo, supongamos que el examen de la función de autocorrelación parcial de una serie sugiere que el proceso que la ha generado tiene una estructura autorregresiva pura. En tal caso, las ecuaciones de Yule-Walker permiten obtener valores de los parámetros ϕ_i a partir de los valores estimados de la función de autocorrelación. Nótese que si las estimaciones muestrales de la función de autocorrelación fuesen muy próximas a los verdaderos valores de dicha función, entonces la solución al sistema de Yule-Walker proporcionaría los valores de los parámetros del modelo, sin necesidad de utilizar el proceso analítico antes descrito. Sin embargo, la estimación de la función de autocorrelación no suele tener tan buenas propiedades, y las primeras iteraciones del procedimiento anterior producen, en general, variaciones significativas.

Si, por el contrario, se ha especificado un modelo MA(1), entonces, recordando que $\rho_1 = -\frac{\theta}{1 + \theta^2}$, esta ecuación puede invertirse para obtener el valor de θ . La ecuación que así resulta es de segundo grado y tiene, por tanto, dos soluciones; en general, una de las raíces es mayor, en valor absoluto, y la otra menor que 1. Para garantizar la invertibilidad del proceso, escogemos siempre la raíz inferior a la unidad en valor absoluto.

Para procesos de orden superior, la obtención de valores iniciales de los parámetros es más compleja, pero puede intentarse un procedimiento similar al descrito.

Una alternativa consistiría en obtener la representación autorregresiva del proceso MA, y estimarla. Este método tiene la gran ventaja de su sencillez. Sin embargo, ya hemos visto que la representación AR de cualquier proceso MA de orden finito es de orden infinito. Por tanto, habría que truncar la autorregresión. Ello no es un problema excesivo, puesto que los coeficientes en los sucesivos retardos son potencias de los coeficientes ϕ , que son menores que 1 en valores absoluto. Pero además, una vez truncada la regresión, los coeficientes restantes están sujetos a restricciones no lineales, que habría que imponer si se pretende obtener un estimador eficiente. Para ilustrar esta

problemática, consideremos el modelo MA(1), cuya representación autorregresiva es:

$$y_t = \varepsilon_t - \theta y_{t-1} - \theta^2 y_{t-2} - \theta^3 y_{t-3} - \dots$$

Si el parámetro θ fuese suficientemente pequeño como para que

$$y_t = -\theta y_{t-1} + u_t$$

fuese una buena aproximación, entonces el coeficiente de y_{t-1} sería una estimación bastante aproximada del parámetro θ . Hay que hacer notar, sin embargo, que tal regresión no dejaría de ser una mala especificación de un modelo que es, realmente, una autorregresión de orden infinito. De este modo, el término de error u_t incorporaría los retardos omitidos, y se produce el problema de correlación entre el regresor y_{t-1} y u_t al que antes hicimos mención. También es cierto que la trascendencia de tal correlación depende de la magnitud de los coeficientes de los retardos omitidos, es decir, de la calidad de la aproximación anterior.

Dicha aproximación puede mejorarse, y con ello aminorar el sesgo producido por la correlación entre y_{t-1} y u_t , si se incluye algún otro retardo, por ejemplo:

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} + u_t$$

pero como vemos, ya aparece una restricción no lineal entre los coeficientes de y_{t-1} e y_{t-2} .

13.11. DIAGNOSTICO DEL MODELO

Tras estimar un modelo ARIMA, es *esencial* llevar a cabo un análisis de los coeficientes y residuos del modelo, con el objetivo de detectar posibles indicios de mala especificación. Respecto a los coeficientes, éstos deben de satisfacer siempre las condiciones de estacionariedad e invertibilidad. Cuando se estiman términos AR(2) o MA(2), deben obtenerse siempre sus raíces, y factorizar dichos términos cuando las raíces sean reales, por la razón que en seguida veremos. También debe hacerse hincapié en que los coeficientes estimados sólo serán relevantes si el algoritmo numérico de estimación ha convergido. Si no lo ha hecho, puede concederse un número mayor de iteraciones, pero puede también ser un indicio de ausencia de estacionariedad o de invertibilidad y, por tanto, de un deficiente ajuste del modelo especificado a las series analizadas.

Asimismo, deben examinarse cuidadosamente los residuos resultantes hasta que se consiga, en la medida de lo posible, eliminar toda duda acerca de que éstos obedecen un proceso de ruido blanco. Esta verificación es crucial, pues sobre tal supuesto se habrá diseñado la estrategia de estimación y predicción, como hemos expuesto en las secciones anteriores. Cualquier evidencia en

contra de la hipótesis de ruido blanco para los residuos constituye un indicio de mala especificación del modelo.

13.11.a. Análisis de residuos

Ya vimos en la Sección 13.5 cómo unos residuos MA(1) obtenidos en un modelo AR(2) sugieren un modelo más correcto ARMA(2, 1). Como otro ejemplo, supongamos que se ha especificado y estimado el modelo:

$$y_t = \delta + \phi_1 y_{t-1} + u_t \quad [13.9]$$

cuyos residuos parecen obedecer el modelo:

$$u_t = \phi_2 u_{t-1} + \varepsilon_t \quad [13.10]$$

donde, a diferencia de u_t , ε_t es ruido blanco. Una simple sustitución del modelo [13.10] en [13.9] muestra que el verdadero modelo de y_t es:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$

donde $\beta_0 = \delta(1 - \phi_2)$, $\beta_1 = \phi_1 + \phi_2$ y $\beta_2 = -\phi_1 \phi_2$. Es decir, el proceso y_t tiene una estructura AR(2) aunque, por error, se había especificado un modelo AR(1).

Sin embargo, si el término de error del modelo [13.9] tuviese una estructura de media móvil: $u_t = \varepsilon_t + \theta \varepsilon_{t-1}$, entonces la sustitución directa de esta estructura en [13.9] muestra que, en realidad, y_t sigue una estructura ARMA(1, 1).

Supuesto que se ha identificado correctamente una estructura AR, la elección del orden se basa en contrastes de significación llevados a cabo con los valores estimados de la función de autocorrelación parcial. Cuando el verdadero modelo es AR(p), la distribución de los valores estimados de la fap es, aproximadamente: $\hat{\phi}_{jj} \sim N\left(0, \frac{1}{T}\right)$ para $j > p$, por lo que si $\hat{\phi}_{jj}$ está en el

intervalo $\pm \frac{2}{\sqrt{T}}$ para todos los valores de j superiores a un cierto p , entonces no se rechazará la hipótesis de que el orden del modelo AR es menor o igual que p .

Un procedimiento similar permite identificar el orden de un proceso MA(q). Como sabemos, un proceso MA(q) se caracteriza porque su función de autocorrelación simple teórica es cero para valores $k > q$. Se trata, por tanto, de efectuar contrastes de hipótesis de significación estadística acerca de los valores estimados de dicha función. En un proceso de ruido blanco, las estimaciones r_k se distribuyen asintóticamente $N\left(0, \frac{1}{T}\right)$. Así, se acepta la no

significación de las estimaciones r_k si caen dentro del intervalo $\pm \frac{2}{\sqrt{T}}$. Si ello

ocurre para todo k mayor que un cierto q , pensamos que el proceso es MA(q). Si ocurre para todo k , entonces se mantiene el supuesto de que la serie temporal de datos proviene de un proceso de ruido blanco. Si el proceso no es ruido blanco, entonces el cociente $\frac{1}{\sqrt{T}}$ puede no ser una aproximación suficientemente buena a la desviación típica de r_k . Bartlett [1946] ha probado que si ρ_k es distinto de cero para $k \leq q$ e igual a cero para $k > q$, entonces la varianza de r_k es aproximadamente igual a:

$$\text{Var}(r_k) = \frac{1 + 2(\rho_1^2 + \rho_2^2 + \dots + \rho_q^2)}{T}$$

En particular, para contrastar la hipótesis nula de que el proceso estocástico subyacente tiene estructura de ruido blanco, como ocurre cuando aplicamos este procedimiento a los residuos de un modelo estimado, la expresión anterior se reduce a:

$$\text{Var}(r_k) = \frac{1}{T}$$

que es la justificación para utilizar $\pm \frac{2}{\sqrt{T}}$ como intervalo de confianza del 95 por 100 para los valores estimados de la función de autocorrelación. Como acabamos de decir, tal práctica sólo está justificada si la hipótesis nula que se contrasta es la de ruido blanco. Si, por el contrario, ya se ha aceptado previamente que la serie en análisis proviene de un proceso con una cierta estructura estocástica, y se está considerando la posibilidad de añadir más estructura, entonces la práctica anterior sólo puede entenderse como una aproximación sencilla al contraste dado por la expresión anterior para la $\text{Var}(r_k)$.

La fórmula de Bartlett es únicamente una aproximación a la verdadera varianza de los valores estimados de la función de autocorrelación. Dicha aproximación tiende a sobrestimar el valor de la varianza, generando así unos intervalos de confianza de un tamaño mayor del que realmente debieran tener. En consecuencia, la utilización de la fórmula de Bartlett tiende a mantener la hipótesis nula de no significación de r_k más a menudo de lo que debiera y, con ello, a no detectar estructura estocástica en casos en que dicha estructura existe.

Conviene, por tanto, prestar especial atención a la posible detección de alguna regularidad en la función de autocorrelación que la asemeje a alguna de las vistas en este capítulo, *incluso* si sus valores no son significativos. En especial, conviene ser exigente con las estimaciones de los primeros valores de dicha función, y utilizar para ellos un intervalo de confianza de aproximadamente $\pm 1,25$ o $\pm 1,5$ desviaciones típicas, en vez de las habituales dos desviaciones típicas. En definitiva, $\frac{1}{T}$ no es sino una *cota superior* para la varianza de dichos valores.

En ocasiones, para formalizar el contraste de la hipótesis nula H_0 : «los residuos del modelo son ruido blanco», se resume toda su estructura de autocorrelación en el estadístico de Box-Pierce:

$$Q = T(r_1^2 + r_2^2 + \dots + r_k^2)$$

que se distribuye como una chi-cuadrado con $k - p - q$ grados de libertad, donde p y q son el número de parámetros AR y MA estimados. Con el objeto de ganar potencia, se ha propuesto también el estadístico modificado por Ljung y Box:

$$Q' = T(T+2) \sum_1^k \frac{r_j^2}{T-j}$$

que tiene distribución chi-cuadrado con k grados de libertad.

De igual modo, el gráfico de residuos puede ilustrar la presencia de posibles tendencias que no apareciesen obvias en la fas. Recordemos que, generalmente, en situaciones de no estacionariedad, la fas decrece hacia cero sólo lentamente, pero, en ocasiones, los residuos muestran una clara tendencia y, sin embargo, sólo los valores iniciales de la fas están fuera del intervalo de confianza del 95 por 100.

Cuando se analiza una variable estacional, debe prestarse especial atención a los primeros valores de las fas y fap, pero también a los primeros valores de orden estacional. Así, con una serie trimestral, el analista debe observar detenidamente los valores 1, 2, 3, 4 y 8 de ambas funciones, a los valores 1, 2, 3, 4, 5, 6, 12 y 24 si trabaja con una serie mensual. Todo ello sin menoscabo de examinar asimismo el resto de los valores.

13.11.b. Sobreparametrización y sobrediferenciación

Dos aspectos de la identificación de un modelo univariante están muy relacionados entre sí: a) la posible *sobreparametrización*, por existencia de factores comunes, y b) la posible *sobrediferenciación* del proceso. En primer lugar, debe notarse que los modelos:

$$\nabla^2 y_t = (1 - 0,97L)\varepsilon_t \quad [13.11]$$

$$\nabla y_t = \varepsilon_t \quad [13.12]$$

son prácticamente indistinguibles, puesto que como $\nabla = (1 - 1,0L)$, existe un factor común en ambos miembros de [13.11]. De modo similar, son modelos aproximados:

$$(1 - 0,95L)y_t = (1 - 0,20L)\varepsilon_t \quad [13.13]$$

$$\nabla y_t = (1 - 0,20L)\varepsilon_t \quad [13.14]$$

puesto que un término AR(1) con coeficiente positivo y próximo a 1 sugiere la no estacionariedad de la variable y , por ello, la conveniencia de diferenciarla. Por último, los modelos

$$(1 - 1,20L + 0,35L^2)\nabla y_t = (1 - 0,70L)\varepsilon_t \quad [13.15]$$

$$(1 - 0,50L)\nabla y_t = \varepsilon_t \quad [13.16]$$

son indistinguibles, por cuanto que la factorización del polinomio AR(2): $(1 - 1,20L + 0,35L^2) = (1 - 0,70L)(1 - 0,50L)$, muestra la existencia de un factor común en ambos miembros de [13.5] que no era evidente a simple vista. Como norma general, el analista siempre debe obtener las raíces del polinomio cuando estima un término AR(2).

En ocasiones, no resulta evidente cuál es el orden correcto de diferenciación de una serie. A este respecto, no hay una doctrina clara: algunos investigadores prefieren no diferenciar por miedo a eliminar información relevante; otros temen que no diferenciar deje en la serie aspectos de no estacionariedad que sesguen el proceso de elaboración del modelo. Esta última estrategia parece, sin embargo, dar mucho mejores resultados. El analista no debe temer introducir una diferencia adicional en la variable, especialmente si incorpora simultáneamente un término de media móvil. Por ejemplo, pasaría del modelo:

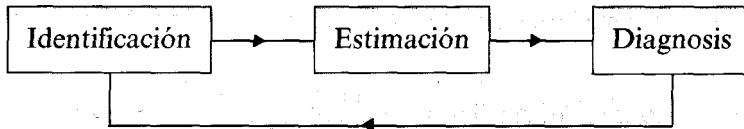
$$(1 + \phi_1 L + \phi_2 L^2)\nabla y_t = \varepsilon_t \quad [13.17]$$

Sobre cuyos residuos existen sospechas de no estacionariedad, al modelo:

$$(1 + \phi_1 L + \phi_2 L^2)\nabla^2 y_t = (1 - \theta L)\varepsilon_t \quad [13.18]$$

y si el parámetro θ se aproxima mucho a 1, puede volverse al modelo [13.7]. Sin embargo, siempre que el valor de θ no produzca un modelo claramente no invertible, conviene mantener la especificación [13.8], pues la combinación de diferencia y término de media móvil suele producir buenos resultados predictivos.

Cuando se han detectado factores comunes, no debe simplemente mantenerse el modelo tras la simplificación pertinente, sino que debe volverse a estimar, al igual que ocurre con cualquier evidencia de mala especificación que se obtuviese de los residuos o cuando se decide introducir una diferencia adicional. En definitiva, el proceso de elaboración de un modelo univariante tiene varias etapas:



que sólo finalizan cuando la diagnosis de un modelo estimado no sugiere indicación de mala especificación.

13.11.c. Valores influyentes y anomalías. Análisis de intervención

Otro aspecto importante en la elaboración de modelos univariantes lo constituye la posible existencia de valores influyentes. Estos son valores de la serie que se analiza que, por su magnitud, *distorsionan* el proceso iterativo de construcción del modelo. Piénsese que, en general, el modelo se especifica para la variable diferenciada al menos una vez, puesto que prácticamente ninguna variable stock, y tan sólo pocas variables flujo o precios, son estacionarias. Por consiguiente, un aumento o disminución mensual, trimestral o anual, según la frecuencia de observación de la serie, de importante magnitud, pueden resultar valores influyentes.

Hay dos posibles estrategias frente a este tipo de valores: Por una parte, pueden tratarse de igual modo que el resto de las observaciones de la serie. Por otro lado, puede *estimarse* su influencia para, en cierto modo, descontar su posible efecto en la elaboración del modelo univariante.

En algunos casos, una intervención de política económica tiene un importante efecto sobre la variación en la oferta monetaria o los tipos de interés que podría hacer dichos valores influyentes. Precisamente porque se sabe que han sido forzados por razones exógenas al proceso generador de la serie, que estamos tratando de elaborar, requieren un trato especial, por lo que debe estimarse su efecto. Algo análogo debe hacerse con el impacto de condiciones meteorológicas especialmente adversas sobre la cosecha de un producto agrícola, cuya producción se pretende modelizar, o con la ocurrencia de una huelga en un sector industrial cuya actividad productiva se quiere prever.

Por tanto, hay muchas instancias en que la estimación de tales efectos por separado del resto del modelo está justificada. A esta estrategia se le denomina *Análisis de intervención*. Cuando se desconoce la causa de un valor influyente, no procede, en general, llevar a cabo su *intervención*. Sin embargo, puede intervenir dicho valor con objeto de que no distorsione la especificación de la estructura estocástica de la serie, para comprobar posteriormente si la estructura que se ha identificado se mantiene al no intervenir el valor influyente, aunque haya que estimar nuevamente el modelo.

Las *intervenciones* efectuadas sobre un valor influyente son, generalmente, de dos tipos. Si y_{t_0} es un valor influyente, definimos una *variable impulso* en t_0 como:

$$\zeta_{t_0}^i = \begin{cases} 1 & \text{si } t = t_0 \\ 0 & \text{si } t \neq t_0 \end{cases}$$

y una *variable escalón* en t_0 como:

$$\xi_{t_0}^s = \begin{cases} 0 & \text{si } t < t_0 \\ 1 & \text{si } t \geq t_0 \end{cases}$$

El análisis de intervención consiste en introducir una de estas variables en el modelo univariante, y estimar su coeficiente. Se incluye una variable impulso cuando el valor influyente se ha producido tan sólo en t_0 . Una variable escalón es más adecuada cuando los valores de la serie (o de sus tasas de cambio) son sistemáticamente mayores o menores después de t_0 que antes de dicho instante o, dicho de otro modo, cuando se ha producido un cambio permanente en el valor medio de la variable, a diferencia del impulso, que corresponde a un cambio meramente transitorio. Cuando la media ha aumentado, el coeficiente asociado a $\xi_{t_0}^s$ será positivo, siendo negativo en caso contrario.

13.12. MODELOS DE FUNCION DE TRANSFERENCIA

La representación ARIMA univariante de una serie temporal correspondiente a una variable Y puede generalizarse para incorporar otras variables X como explicativas. El modelo resultante se conoce como *función de transferencia*, con las variables X como *input* e Y como *output*. Centramos nuestra presentación en el caso de un solo input, pues la extensión a múltiples inputs, tanto en términos de especificación como de estimación, es inmediata.

Conviene especificar un modelo de función de transferencia cuando:

1. Se espera que la relación entre input y output, tanto a través de sus componentes regulares como estacionales, tenga una *estructura dinámica* suficientemente rica como para que la representación econométrica habitual requiriese de un número elevado de parámetros.
2. Se cree que una representación adecuada de la estructura estocástica del *término de error* resultante de la relación de X hacia Y precisa de una modelización ARIMA, pues no estaría suficientemente recogida por los sencillos esquemas de autocorrelación utilizados en el modelo lineal general.

Un modelo de función de transferencia tiene varias *ventajas* adicionales:

1. Permite una representación *parsimoniosa*, es decir, basada en un número reducido de coeficientes, de relaciones dinámicas, incluso si son muy complicadas.
2. Dispone de una estrategia sencilla y gradual para la especificación de un modelo dinámico de relación que capture adecuadamente el efecto del input sobre el output.
3. Proporciona instrumentos para comprobar que se utilizan como input y output variables que tienen análogas características de estacionariedad, de modo que los residuos del modelo son estacionarios.

Esta propiedad, crucial para justificar el análisis de inferencia que con el modelo estimado pudiera llevarse a cabo, ha generado una considerable cantidad de estudios recientes, destinados a generar procedimientos de contraste de estacionariedad, así como de especificación, ambos en el contexto del modelo econométrico lineal, que serán objeto de análisis en el Capítulo 14.

La representación genérica del modelo de función de transferencia con un input es:

$$\begin{aligned} Y_t &= v(L)X_t + N_t = (v_0 + v_1L + v_2L^2 + \dots)L^b X_t + N_t = \\ &= \frac{\omega(L)}{\delta(L)} X_{t-b} + N_t = \delta^{-1}(L)\omega(L)X_{t-b} + N_t = \quad [13.19] \\ &= \frac{\omega_0 - \omega_1L - \dots - \omega_sL^s}{1 - \delta_1L - \dots - \delta_rL^r} X_{t-b} + N_t \end{aligned}$$

donde, al igual que en secciones previas, L denota el operador de retardos. Los polinomios $\delta(L)$ y $\omega(L)$ se denominan *autorregresivo* y de *medias móviles*, respectivamente, mientras que N_t es la *perturbación* del modelo. El polinomio $v(L)$ se denomina *función de respuesta al impulso*, pues sus sucesivos coeficientes describen el efecto que, a través del tiempo, tendría sobre Y un impulso (es decir, un cambio puramente transitorio) en la variable X . El parámetro b , ($b \geq 0$) que aparece como subíndice en la variable X se denomina *tiempo muerto* de la relación, y denota el número de períodos que deben transcurrir para que la variación en X comience a dejarse sentir sobre Y .

Por muy rico que sea el efecto de X sobre Y , tanto en intensidad temporal como en duración, es posible recogerlo adecuadamente por medio del polinomio $v(L)$, consiguiendo además que el término de error o perturbación N_t sea estacionario. Para ello, puede ser preciso, en ocasiones, utilizar un polinomio $v(L)$ de orden elevado, quizá infinito, pero su descomposición como cociente de dos polinomios de orden finito, incluso muy reducido, como aparece en [13.19], permite la representación parsimoniosa de la relación.

A su vez, la perturbación del modelo admitirá una representación univariante:

$$N_t = \psi(L)a_t = (1 + \psi_1L + \psi_2L^2 + \dots)a_t = \frac{\theta(L)\Theta(L)}{\phi(L)\Phi(L)} a_t$$

donde las mayúsculas hacen referencia a la naturaleza estacional de los polinomios respectivos, y donde los polinomios autorregresivos $\phi(L)$ y $\Phi(L)$ pueden contener raíces unitarias, aunque no raíces de módulo superior a la unidad, y a_t es un ruido blanco gaussiano, es decir, con distribución Normal.

La acumulación sucesiva de los coeficientes de la función de respuesta al impulso genera la *función de respuesta al escalón*, que proporciona el efecto gradual que sobre Y tendría una desviación *permanente* en el valor de X . Si la perturbación N_t es estacionaria, la función de respuesta al escalón es acotada; su límite se denomina *ganancia* de la relación, y mide el efecto a

largo plazo que sobre Y_t tiene una variación en X_t . Puede calcularse mediante $g = \omega(1)/\delta(1)$.

Un instrumento básico en la especificación del modelo de función de transferencia es la *función de correlación cruzada* (en lo sucesivo, fcc), similar a las funciones de autocorrelación, y definida como cociente entre la covarianza entre X e Y , a distintos retardos y en ambas direcciones, por el producto de sus desviaciones típicas:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{s_x s_y}$$

donde:

$$\gamma_{xy}(k) = \frac{1}{n} \sum_{t=1}^{T-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k = 0, 1, 2, \dots$$

$$\gamma_{xy}(k) = \frac{1}{n} \sum_{t=k+1}^T (x_{t-k} - \bar{x})(y_t - \bar{y}), \quad k = 0, -1, -2, \dots$$

A diferencia de las fas y fap de una serie, no es simétrica, y su valor central $\rho_{xy}(0)$ es el coeficiente de correlación habitual entre ambas variables.

Si las dos variables X e Y no estuviesen correlacionadas entre sí y una de ellas fuese ruido blanco, por ejemplo $Y_t = a_t$, se tendría:

$$\text{Cov}[\rho_{xa}(k), \rho_{xa}(k+s)] \cong \frac{\rho_{xx}(s)}{T-k}$$

$$\text{Var}[\rho_{xa}(k)] \cong \frac{1}{T-k}$$

de modo que:

$$\text{Corr}[\rho_{xa}(k), \rho_{xa}(k+s)] \cong \rho_{xx}(s)$$

En tales condiciones, el inverso del número de grados de libertad da la varianza de cada valor de la fcc, que reproduciría la estructura de la fas de X_t . Ello ocurrirá *a pesar* de la ausencia de correlación cruzada entre ambas variables. Por consiguiente, sólo puede esperarse una fcc con valores nulos fuera de la correlación contemporánea $\rho_{xy}(0)$, si *ambas series* fuesen ruido blanco.

Las expresiones anteriores muestran que los sucesivos valores de la fcc están generalmente correlacionados entre sí, lo que puede dificultar la identificación de la estructura dinámica de la relación entre X e Y . Lo que acabamos de ver es que la *fcc tenderá a sugerir más relación entre las variables de la que realmente existe*.

13.12.a. Identificación del modelo de función de transferencia

La identificación del modelo de transferencia consiste en obtener valores aproximados de los coeficientes de la función de respuesta al impulso $v(L)$, de modo que puedan utilizarse para inferir los órdenes r y s de los polinomios $\delta(L)$ y $\omega(L)$ en [13.19], así como el tiempo muerto b . Identificados r y s , los coeficientes de $v(L)$ pueden utilizarse asimismo para obtener preestimaciones de los coeficientes en $\delta(L)$ y $\omega(L)$.

Si escribimos [13.19] en la forma

$$\begin{aligned} (1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r)(v_0 + v_1 L + v_2 L^2 + \dots) = \\ = (\omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s) L^b \end{aligned} \quad [13.20]$$

se tiene el sistema:

$$\begin{aligned} v_j &= 0 && \text{para } j < b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_0 && \text{para } j = b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} - \omega_{j-b} && \text{para } j = b + 1, \dots, b + s \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} && \text{para } j > b + s \end{aligned} \quad [13.21]$$

de modo que los coeficientes de la función de respuesta al impulso tienen la siguiente estructura:

a) Los primeros b coeficientes son nulos, lo que permite identificar el tiempo muerto b : $v_0 = v_1 = \dots = v_{b-1} = 0$.

b) Desde el coeficiente v_b hasta el coeficiente v_{b+s} no se aprecia ninguna regla de formación.

c) Los coeficientes v_j para $j \geq b + s + 1$ se comportan de acuerdo con la ecuación en diferencias de orden r que aparece en [13.21], tomando como valores iniciales v_{b+s} , v_{b+s-1} , \dots , $v_{b+s-r+1}$.

Por ejemplo, en la función de transferencia:

$$Y_t = \frac{\omega_0}{1 - \delta_1 L - \delta_2 L^2} X_t$$

todos los coeficientes de la función de respuesta al impulso obedecen a una ecuación en diferencias de orden 2; lamentablemente, ello puede generar múltiples configuraciones, aunque una típica sería sinusoidal.

Por el contrario, en el modelo

$$Y_t = \frac{\omega_0 - \omega_1 L - \omega_2 L^2}{1 - \delta_1 L} X_{t-1}$$

se tendría un primer coeficiente v_0 nulo, seguido de v_1 y v_2 , que no obedecerían a ninguna regla, mientras que los sucesivos responderían a una ecuación en diferencias de orden 1, por lo que decaerían exponencialmente, aunque quizá alternando en signo.

Si las variables que se pretende relacionar no son estacionarias, su función de correlación cruzada, así como sus fas y fap, no decaerá rápidamente hacia cero. En tal caso, es preciso transformarlas mediante diferencias para lograr estacionariedad, lo que denotamos en lo sucesivo por $y_t = \nabla^d Y_t$, $x_t = \nabla^d X_t$, $n_t = \nabla^d N_t$, que serían estacionarias. Si es preciso tomar asimismo diferencias de orden estacional, así como quizá tomar logaritmos para atenuar la heteroscedasticidad, todo ello estaría incorporado en nuestra notación: y_t , x_t , n_t . La función de transferencia transformada resulta:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + n_t \quad [13.22]$$

que conserva los mismos coeficientes de la original. Supongamos que esta función de transferencia es tal que $v_j = 0$ para $j > q$. Si premultiplicamos en [13.22] por x_{t-k} para sucesivos valores $k \geq 0$ y tomamos esperanzas se tiene:

$$\begin{pmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \dots \\ \gamma_{xy}(q) \end{pmatrix} = \begin{pmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \dots & \gamma_{xx}(q) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \dots & \gamma_{xx}(q-1) \\ \dots & \dots & \dots & \dots \\ \gamma_{xx}(q) & \gamma_{xx}(q-1) & \dots & \gamma_{xx}(0) \end{pmatrix} \cdot \begin{pmatrix} v_0 \\ v_1 \\ \dots \\ v_q \end{pmatrix} \quad [13.23]$$

de modo que si se dispusiese de las estimaciones muestrales de la fcc entre x e y , lo que siempre es posible, podríamos resolver el sistema [13.23] para obtener estimaciones de los coeficientes de la función de respuesta al impulso v_0, v_1, \dots, v_n .

13.12.b. Identificación con preblanqueo

El proceso de identificación de la función de transferencia se simplifica notablemente si se filtran previamente ambas variables de manera adecuada. Supongamos que la variable x_t , siendo estacionaria, admite la representación ARMA:

$$\phi_x(L)\theta_x^{-1}(L)x_t = \alpha_t \quad [13.24]$$

donde α_t es ruido blanco y $\phi_x(L)$ y $\theta_x(L)$ son los polinomios autorregresivos y de medias móviles de dicho modelo ARMA. Una vez estimado [13.24], obtendríamos los residuos $\hat{\alpha}_t$, que no son sino el resultado de *filtrar* la variable x_t por su modelo ARMA. Supongamos que filtramos el output y_t con el mismo modelo ARMA del input que acabamos de estimar, obteniendo:

$$\beta_t = \phi_x(L)\theta_x^{-1}(L)y_t$$

Las variables así filtradas (no cabe esperar que b_t sea ruido blanco) satisfacen la relación:

$$\beta_t = v(L)\alpha_t + \varepsilon_t \quad [13.25]$$

donde:

$$\varepsilon_t = \phi_x(L)\theta_x^{-1}(L)n_t$$

pudiendo observarse que la función de respuesta al impulso en [13.25] es idéntica a la original. Si multiplicamos por α_{t-k} y tomamos esperanzas, se tiene:

$$\gamma_{\alpha\beta}(k) = v_k \sigma_\alpha^2$$

donde $\gamma_{\alpha\beta}(k)$ denota la covarianza entre α_t y β_t en el retardo k , por lo que:

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2} = \frac{\rho_{\alpha\beta}(k)}{\sigma_\alpha} \sigma_\beta, \quad k = 0, 1, 2, \dots \quad [13.26]$$

de modo que *tras preblanquear input y output con el modelo del input, la función de correlación cruzada es proporcional a la función de respuesta al impulso*. El efecto de preblanquear es transformar el sistema de ecuaciones [13.23] en un conjunto de ecuaciones como [13.26], que pueden resolverse por separado para obtener preestimaciones de la función de respuesta al impulso a partir de las desviaciones típicas y de la función de correlación cruzada muestral de α y β .

13.12.c. Identificación de un modelo para el ruido

Una vez identificada una función de respuesta al impulso, y habiendo obtenido estimaciones para sus coeficientes $\hat{v}(L)$, podemos estimar el ruido: $\hat{n}_t = y_t - \hat{v}(L)x_t$. Alternativamente, suele comenzarse especificando para n_t el modelo univariante de y_t , lo que sería correcto si no existiese relación entre y_t y x_t , en cuyo caso, la función de transferencia sería nula. Al ir estimando coeficientes de $v(L)$, el modelo de n_t se hace más sencillo⁽²⁾.

Ambas opciones se basan en modelos estimados previamente, por lo que todo lo que hacemos a continuación tiene un valor tan sólo aproximado. En definitiva, ha de ser el rigor del analista el que debe conducir a la especificación de un modelo univariante, como los estudiados en las secciones previas, para los residuos resultantes del modelo de función de transferencia.

⁽²⁾ Existen otros procedimientos analíticos más complejos para identificar un modelo univariante para la perturbación, que el lector interesado puede consultar en Box y Jenkins (1972). En cualquier caso, no es clara cuál es su utilidad práctica, más allá del buen análisis univariante de la serie de perturbaciones.

Al especificar un modelo de transferencia, Box y Jenkins sugieren:

1. Que un modelo con estructuras AR o MA de orden 1 ó 2 (además del posible tiempo muerto) será normalmente suficiente, si bien esto debe tenerse en cuenta tanto para la componente regular como para la estacional.
2. Que, aunque la estimación del modelo de transferencia es eficiente sólo si el modelo especificado es correcto, las estimaciones de los valores v_k de la función de respuesta al impulso son útiles en la identificación de la estructura dinámica de relación.
3. Que tampoco se gana mucho con tratar de obtener estimaciones eficientes de los v_k , pues son precisos muchos coeficientes de tal tipo para resumir el cociente de polinomios $\delta(L)/\omega(L)$. Recíprocamente, el objetivo último del analista es obtener estimaciones precisas de estos dos polinomios, y las estimaciones que ellas implican para los coeficientes v_k están altamente correlacionadas entre sí, y tienen una elevada varianza.

13.12.d. Estimación de un modelo de función de transferencia

Una vez identificado el modelo de transferencia, sin estacionalidad:

$$y_t = \delta^{-1}(L)\omega(L)x_{t-b} + \phi^{-1}(L)\theta(L)a_t$$

y dados unos valores iniciales (x_0, y_0, a_0) podemos obtener los residuos sucesivos: $a_t(b, \delta, \omega, \phi, \theta/x_0, y_0, a_0)$ a partir de $t_0 = \max\{r, s + b\} + 1$ y, bajo el supuesto de Normalidad, tratar de minimizar su suma de cuadrados.

Como ejemplo, consideremos la función de transferencia:

$$y_t = \frac{\omega_0 + \omega_1 L + \omega_2 L^2}{1 - \delta L} x_t + \frac{1}{1 - \rho L} a_t$$

en la que $\max\{r, s + b\} + 1 = 3$, y donde $n_t = a_t/(1 - \rho L)$, en la que podemos hacer:

$$\begin{aligned} n_3 &= y_3 - \delta y_2 - \omega_0 x_3 - \omega_1 x_2 - \omega_2 x_1 + \delta n_2 \\ n_4 &= y_4 - \delta y_3 - \omega_0 x_4 - \omega_1 x_3 - \omega_2 x_2 + \delta n_3 \\ &\dots \end{aligned}$$

comenzando de $n_2 = 0$ para después obtener las innovaciones a_t a partir de $t = 4$, por medio de:

$$\begin{aligned} a_4 &= n_4 - \rho n_3 \\ a_5 &= n_5 - \rho n_4 \\ &\dots \end{aligned}$$

y formar, por último, la suma residual:

$$S^2(b, \delta, \omega, \rho) \equiv \sum_4^T a_t^2(b, \delta, \omega, \rho/x_0, y_0, a_0)$$

donde las primeras innovaciones se han tomado iguales a su esperanza incondicional, que es cero.

La expresión genérica de cada a_t es:

$$a_t = y_t - \delta y_{t-1} - \omega_0 x_t - \omega_1 x_{t-1} - \omega_2 x_{t-2} + \delta n_{t-1} - \rho [y_{t-1} - \delta y_{t-2} - \omega_0 x_{t-1} - \omega_1 x_{t-2} - \omega_2 x_{t-3} + \delta n_{t-2}]$$

con vector gradiente:

$$\frac{\partial a_t}{\partial \delta} = -y_{t-1} + \rho y_{t-2} + n_{t-1} - \rho n_{t-2}$$

$$\frac{\partial a_t}{\partial \omega_0} = -x_t + \rho x_{t-1}$$

$$\frac{\partial a_t}{\partial \omega_1} = -x_{t-1} + \rho x_{t-2}$$

$$\frac{\partial a_t}{\partial \omega_2} = -x_{t-2} + \rho x_{t-3}$$

$$\frac{\partial a_t}{\partial \rho} = -(y_{t-1} - \delta y_{t-2} - \omega_0 x_{t-1} - \omega_1 x_{t-2} - \omega_2 x_{t-3} + \delta n_{t-2})$$

por lo que podemos comenzar un algoritmo iterativo del tipo Gauss-Newton a partir de estimaciones iniciales $\hat{\boldsymbol{\theta}} = (\hat{\delta}, \hat{\omega}_0, \hat{\omega}_1, \hat{\omega}_2, \hat{\rho})$ del modo que se discutió en el Capítulo 12. Como allí se vio, un modo de proceder consiste en obtener series temporales para cada una de las componentes del gradiente, así como para los propios residuos a_t , utilizando las estimaciones iniciales de los parámetros, para estimar la regresión:

$$\hat{a}_t = (\hat{\delta} - \delta) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \delta} + (\hat{\omega}_0 - \omega_0) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_0} + (\hat{\omega}_1 - \omega_1) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_1} + (\hat{\omega}_2 - \omega_2) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_2} + (\hat{\rho} - \rho) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \rho} + a_t$$

que se repite, tomando en cada iteración las últimas estimaciones como iniciales, hasta lograr la convergencia. La aparición del vector $\boldsymbol{\theta}$ hace referencia a que las componentes del gradiente de la función a_t están evaluadas en las preestimaciones. Este procedimiento está sujeto a todas las consideraciones que en el Capítulo 12 se hicieron en relación con la convergencia de un algoritmo numérico. Lograda la convergencia, la matriz de covarianzas de las estimaciones se aproxima por el producto $\hat{\sigma}_a^2 (\mathbf{V}\mathbf{a}\mathbf{V}')^{-1}$, donde $\hat{\sigma}_a^2$ se estima mediante el cociente del valor alcanzado por la suma residual en la

última iteración, y el número de observaciones, o de grados de libertad, si se prefiere.

13.12.e. Diagnóstico del modelo de función de transferencia

Uno de los análisis que debe hacerse de toda función de transferencia estimada es la búsqueda de posibles factores comunes que pudiesen quedar enmascarados en los polinomios AR y MA que la componen. Nótese que el modelo de función de transferencia, al igual que todo modelo lineal, es único excepto por operadores factoriales, posiblemente dinámicos. Es decir, si el «verdadero» modelo de transferencia es:

$$Y_t = \delta^{-1}(L)\omega(L)X_{t-b} + \phi^{-1}(L)\theta(L)a_t$$

también es válido el modelo:

$$D(L)Y_t = D(L)\delta^{-1}(L)\omega(L)X_{t-b} + D(L)\phi^{-1}(L)\theta(L)a_t$$

donde $D(L)$ es un polinomio cualquiera en el operador de retardos, aunque el analista preferirá el primero por su sencillez; en la práctica, el menor número de parámetros se traducirá en una mayor eficiencia en la estimación, así como en una ausencia de correlaciones entre sus valores estimados. El objetivo del investigador ha de ser siempre lograr un modelo tan sencillo como sea posible, comenzando siempre por especificaciones sencillas, para complicarlas posteriormente, sólo si es preciso. En todo caso, las correlaciones entre los parámetros estimados deben ser reducidas, pues valores elevados indican sobreparametrización.

Supongamos que el modelo correcto es $y_t = v(L)x_t + \psi(L)a_t$, pero el investigador, incorrectamente, especifica $y_t = v_0(L)x_t + \psi_0(L)a_{0t}$, que, una vez estimado, genera unos residuos que denotamos por a_{0t} . Estos quedan definidos por:

$$a_{0t} = \psi_0^{-1}(L)[v(L) - v_0(L)]x_t + \psi_0^{-1}(L)\psi(L)a_t$$

donde puede apreciarse que la mala especificación puede hacer que: *a)* los residuos tengan autocorrelación, y *b)* estén correlacionados con las variables x_t , y por tanto, con el ruido blanco α_t que genera las x_t .

Caso 1: El modelo de transferencia se especifica correctamente, pero el modelo del ruido incorrectamente: $v_0(L) = v(L)$, pero $\psi_0(L) \neq \psi(L)$. En tal caso: $a_{0t} = \psi_0^{-1}(L)\psi(L)a_t$, por lo que: *a)* la función de correlación cruzada entre los residuos y el input del modelo debe ser no significativamente distinta de cero, pero *b)* los residuos presentan autocorrelación.

Caso 2: El modelo de transferencia se especifica incorrectamente. En este caso se tiene $a_{0t} = \psi_0^{-1}(L)[v(L) - v_0(L)]x_t + a_t$, por lo que se tiene tanto

autocorrelación de los residuos, como una función de correlación cruzada significativamente distinta de cero con el input.

La conveniencia de preblanquear el input antes de obtener su función de correlación cruzada con los residuos proviene de que, como vimos al comienzo de esta sección, los sucesivos valores de dicha función de correlación cruzada están correlacionados, lo que no ocurre con la función de correlación cruzada entre el input preblanqueado y los residuos, si éstos son efectivamente ruido blanco.

Pero además, este mismo análisis sugiere el tipo de modificaciones que deben introducirse en la función de transferencia previamente estimada para obtener una mejor especificación. En efecto, considerando el modelo preblanqueado: $\beta_t = v(L)\alpha_t + \varepsilon_t$ y si denotamos por ε_{0t} los residuos del modelo «incorrecto», se tiene:

$$\varepsilon_{0t} = [v(L) - v_0(L)]\alpha_t + \varepsilon_t$$

se tiene:

$$v_k - v_{0k} = \rho_{\alpha, \varepsilon_0}(k) \frac{\sigma_{\varepsilon_0}}{\sigma_{\alpha}}, \quad k = 0, 1, 2, \dots$$

de modo que la función de correlación cruzada entre los residuos y el input preblanqueado mide la discrepancia entre la función de respuesta al impulso «verdadera» y la estimada.

Por último, puesto que es necesario para una correcta especificación de la función de transferencia que los residuos estén libres de autocorrelación, puede utilizarse los estadísticos de Box-Pierce o de Ljung-Box introducidos en la Sección 13.11. En este caso, $Q_g = T \sum_0^g \hat{r}_{\alpha}^2(k)$ se distribuye como una $\chi_{g+1-(r+s-1)}^2$, donde r y s son los órdenes de los polinomios AR y MA de la función de transferencia, y no dependen del modelo del ruido.

13.13. ALGUNOS EJEMPLOS

Ejemplo 13.1. Consideremos nuevamente las series de datos anuales de Consumo y PIB español, en pesetas constantes de 1980, de la Sección 7.6. La Figura 13.5 presenta ambas series, con evidencia obvia de no estacionariedad, ya que no podría mantenerse que la media de la serie fuese estable en el tiempo. En el análisis de series temporales, la transformación logarítmica es utilizada frecuentemente para amortiguar la heteroscedasticidad que suele aparecer cuando la varianza de la variable crece con la magnitud de ésta al avanzar en el período muestral.

Las funciones de autocorrelación simple (fas) de las variables transformadas (Figuras 13.6 y 13.8) son ejemplos típicos de ausencia de estacionariedad, amortiguándose sus valores con excesiva lentitud. Las funciones de autocorrelación parcial (fap, Figuras 13.7 y 13.9) presentan asimismo clara evidencia de ausencia de estacionariedad, con tan sólo su primer valor significativo, y muy próximo a 1.

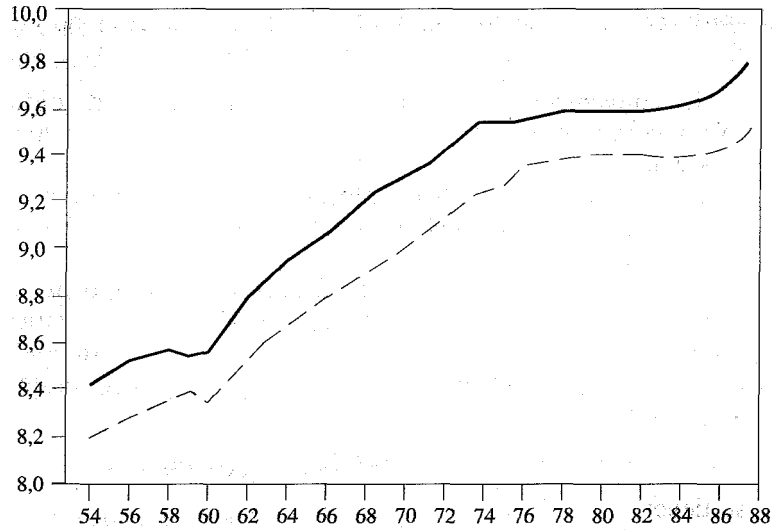


FIGURA 13.5. Consumo [---] y PIB [—]. Transformación logarítmica sobre las series en pesetas constantes base 1980, en miles de millones.

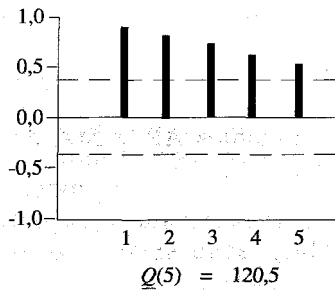


FIGURA 13.6. fas ln (PIB).

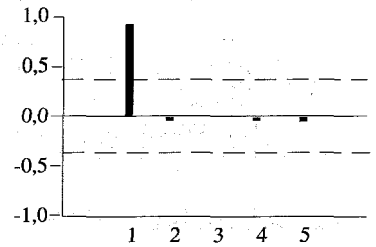


FIGURA 13.7. fap ln (PIB).

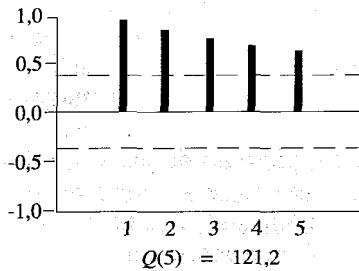


FIGURA 13.8. fas ln (Consumo).

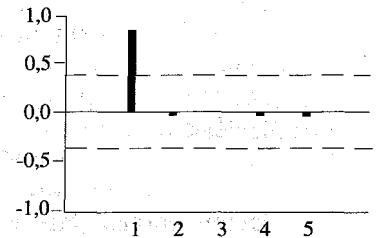


FIGURA 13.9. fap ln (Consumo).

Las series de diferencias del logaritmo (Figura 13.10) tienen la ventaja de poder interpretarse como las tasas de variación anual del Consumo y del PIB. Sus fas y fap en las Figuras 13.11 a 13.14 no presentan clara evidencia de ausencia de estacionariedad. Sin embargo, es importante observar que las series diferenciadas tienen medias muestrales positivas y significativas, con estadísticos t por encima de 8. La Figura 13.10 muestra la evidencia de tal media, junto con el hecho de que las series deambulan, sin oscilar alrededor de sus medias muestrales, como cabría esperar de una serie estacionaria.

Consideremos inicialmente la serie anual del Consumo en pesetas constantes de 1980. Si, a pesar de la evidencia apuntada sobre las primeras diferencias de su logaritmo, intentamos modelizar esta transformación de la variable, el primer valor de la fap, junto con el rápido amortiguamiento de su fas, sugieren un modelo ARIMA(1, 1, 0) para el logaritmo del Consumo. Otra alternativa consistiría en tomar una diferencia adicional de la variable, que aparece, junto con igual transformada del PIB, en la Figura 13.15. Las fas y fap (Figuras 13.6 a 13.19) sugieren ahora una posible estructura de media móvil de orden 1, por lo que deberíamos probar un ARIMA(0, 2, 1). Hay que notar que la duda acerca de una posible diferencia adicional debe ir acompañada siempre de un término MA(1) que pueda cancelarse (si se estima próximo a 1) con la diferencia adicional que estaba en duda; por ello, el modelo alternativo que proponemos tiene bastante sentido. Su evidencia en las fas y fap no es totalmente clara, pero hay que tener en cuenta que estas dos series tienen un número no muy elevado de observaciones. Por otra parte, la segunda diferencia elimina el problema de la media residual significativa.

La estimación de ambos modelos resultó en:

$$(1 - 0,857 L) \nabla \ln C_t = a_t \\ (0,088)$$

$$T = 33, \quad \hat{\sigma}_a = 2,75 \%, \quad \bar{a} = 0,0061 \\ (0,0046)$$

$$\nabla^2 \ln C_t = (1 - 0,545 L) a_t \\ (0,149)$$

$$T = 33, \quad \hat{\sigma}^2 = 2,61 \%, \quad \bar{a} = -0,0045 \\ (0,0261)$$

El modelo en primeras diferencias presenta tres dificultades:

1. El elevado coeficiente del término AR en el primer modelo sugiere la necesidad de una diferencia adicional, puesto que un término $(1 - \rho L)$ con ρ próximo a 1 es esencialmente igual a $(1 - L)$.
2. Lo que es más importante, su gráfico de residuos (Figura 13.20) presenta indicios claros de no estacionariedad, con rachas largas de valores sistemáticamente por encima o por debajo de su media.

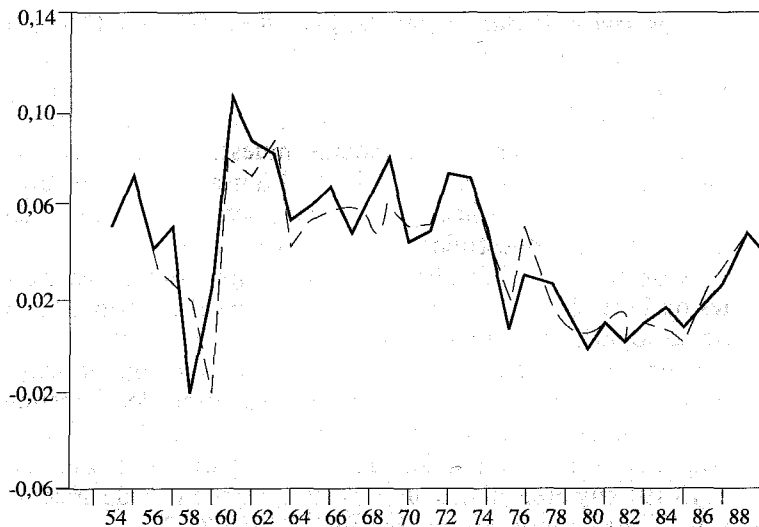


FIGURA 13.10. Consumo [---] y PIB [—]. Tasas de variación logarítmica.

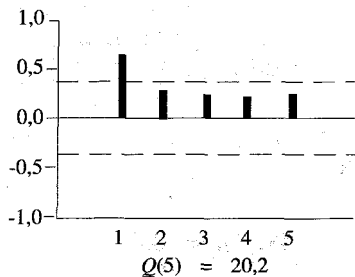


FIGURA 13.11. fas $\nabla \ln(\text{PIB})$.

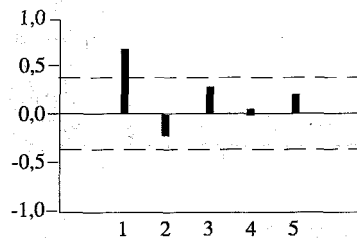


FIGURA 13.12. fap $\nabla \ln(\text{PIB})$.

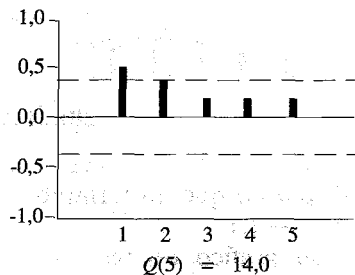


FIGURA 13.13. fas $\nabla \ln(\text{Consumo})$.

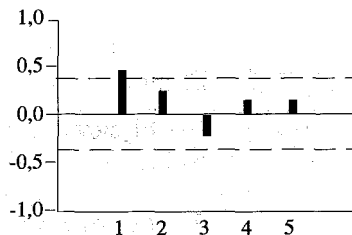


FIGURA 13.14. fap $\nabla \ln(\text{Consumo})$.

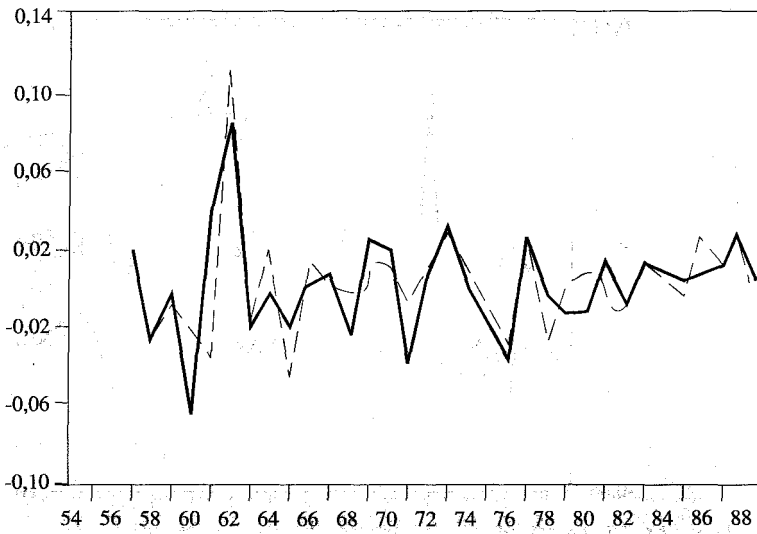


FIGURA 13.15. Consumo [---] y PIB [—]. Segundas diferencias del logaritmo.

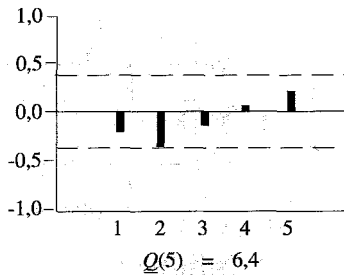


FIGURA 13.16. fas $\nabla^2 \ln(\text{PIB})$.

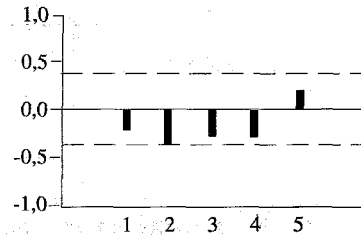


FIGURA 13.17. fap $\nabla^2 \ln(\text{PIB})$.

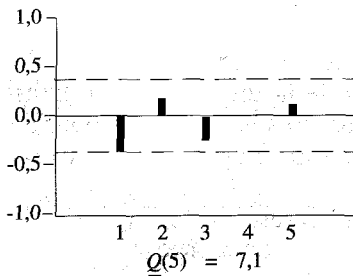


FIGURA 13.18. fas $\nabla^2 \ln(\text{Consumo})$.

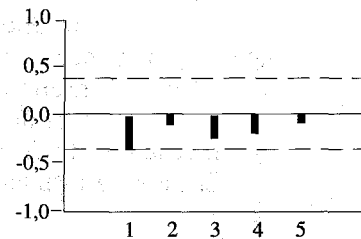


FIGURA 13.19. fap $\nabla^2 \ln(\text{Consumo})$.

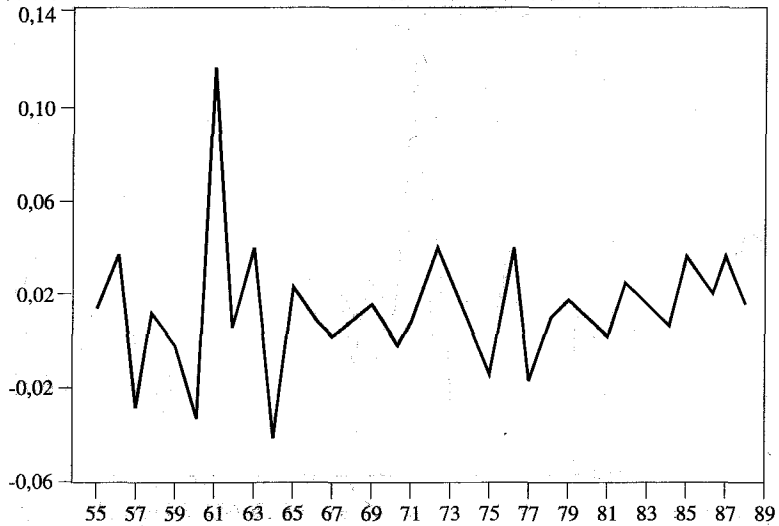


FIGURA 13.20. Residuos Consumo (Modelo en primeras diferencias).

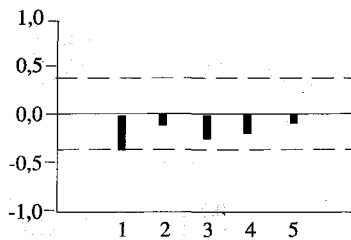


FIGURA 13.21. fas residuos.

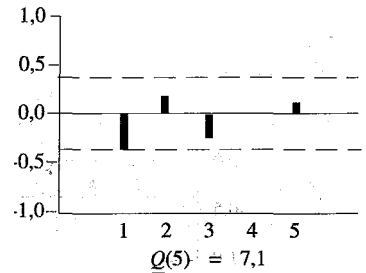


FIGURA 13.22. fap residuos.

3. Consistentemente con esta observación, las fas y fap (Figuras 13.21 y 13.22) de dichos residuos sugieren la existencia de estructura estocástica adicional, aunque en la forma de un término AR(1) que ya está incluido en el modelo.

El modelo en segundas diferencias no presenta ninguno de estos problemas: a) su parámetro MA está lejos de la unidad, lo que habría supuesto la cancelación con una de las diferencias, b) sus residuos (Figura 13.23) parecen estacionarios, y c) no hay indicios de estructura estocástica residual en las fas y fap (Figuras 13.26 y 13.27).

El modelo estimado del PIB resultó ser:

$$(1 + 0,16L + 0,38L^2)\nabla^2 \ln \text{PIB}_t = a_t$$

(0,16) (0,16)

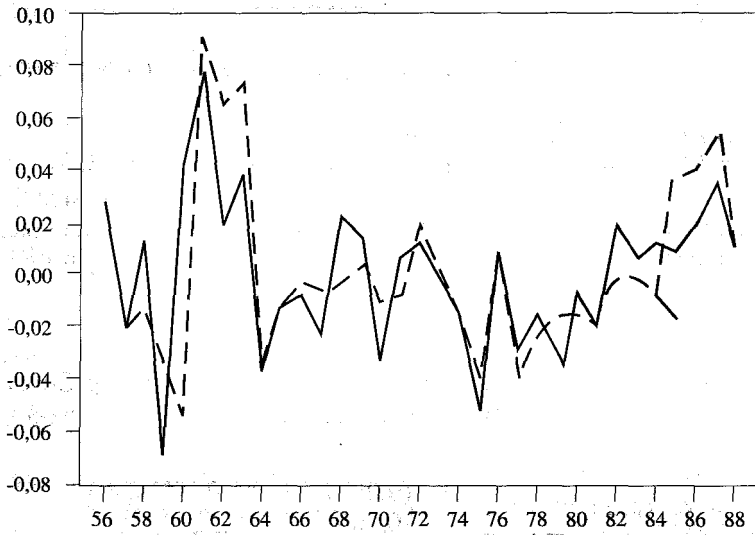


FIGURA 13.23. Residuos Consumo [---] y residuos PIB [—].

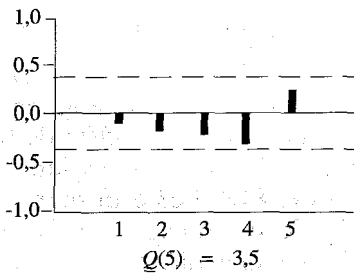


FIGURA 13.24. fas residuos PIB.

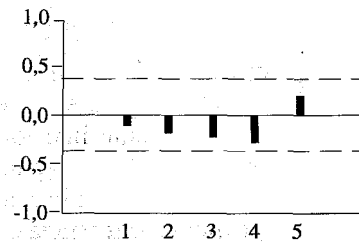


FIGURA 13.25. fap residuos PIB.

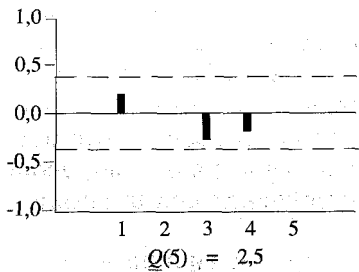


FIGURA 13.26. fas residuos Consumo.

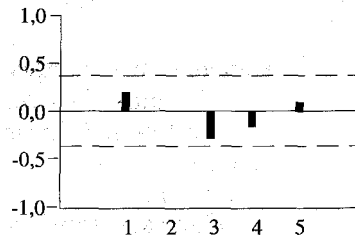


FIGURA 13.27. fap residuos Consumo.

con

$$T = 33, \hat{\sigma}_u = 2,58 \% ; a = -0,0002(0,0045)$$

sin que su gráfico de residuos (Figura 13.23) ni sus fas y fap (Figuras 13.24 y 13.25) mostrasen la necesidad de modificar el modelo. El signo positivo del coeficiente de L^2 en el polinomio AR(2) implica la existencia de una estructura cíclica en esta variable, que recoge la evolución del ciclo económico. De acuerdo con los coeficientes estimados, el período de dicho ciclo es de 4,37 años.

Posteriormente consideramos la posibilidad de establecer una relación dinámica entre ambas variables, tratando de recoger el efecto que variaciones en el PIB (como indicador de Renta) tienen sobre el Consumo. Tras preblanquear ambas variables y examinar la fcc resultante, estimamos el modelo:

$$\nabla \ln C_t = \frac{(0,062)}{0,766} \nabla \ln \text{PIB}_t + (1 - 0,292)a_t$$

$$\frac{(0,067)}{1 - 0,204L} \quad (0,156)$$

con

$$T = 33, \hat{\sigma} = 1,22 \% ; a = 0,0011(0,0021)$$

cuyos residuos, así como sus fas, fap, aparecen en las Figuras 13.28 a 13.30.

Como puede verse, la desviación típica de este modelo de transferencia es sensiblemente inferior a la obtenida con el modelo univariante del Consumo, lo que avala la capacidad explicativa del PIB. Por otra parte, a diferencia de los modelos univariantes, el modelo de transferencia se ha especificado en primeras diferencias. La cancelación de una diferencia en ambas variables sugiere la existencia de una «tendencia común» en ambas, lo que será objeto de análisis posterior en el próximo capítulo.

Ejemplo 13.2. Consideremos ahora la serie trimestral del número de ocupados en España, tomado de la Encuesta de Población Activa (EPA) para el período (1976, IV) a (1992, III) (Figura 13.31). Las fas y fap (Figuras 13.32 y 13.33) muestran una evidencia obvia de ausencia de estacionariedad. Las fas y fap de la primera diferencia de la variable (Figuras 13.34 y 13.35) sugieren su no estacionariedad estacional, puesto que los valores r_4, r_8, r_{12} son elevados. Los primeros valores de la fas hasta r_7 son muy altos, lo que sugiere tomar asimismo una diferencia regular adicional, acompañada de un término MA(1) que puede llegar a simplificarse con la misma, si se estima en torno a la unidad.

Con esta transformación, las fas y fap, que aquí no se muestran, sugirieron un posible modelo AR(2) estacional, si bien no de modo obvio. Al estimarse la estructura ARIMA(0, 2, 1) \times ARIMA₄(2, 1, 0) se detectó un valor atípico

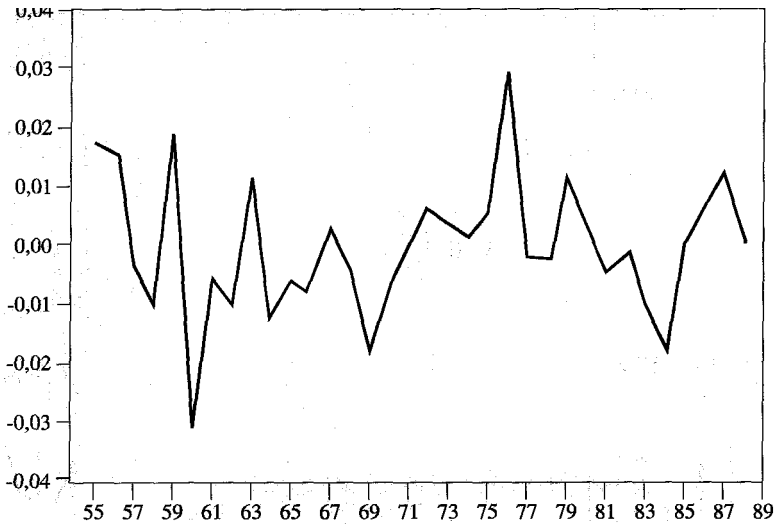


FIGURA 13.28. Residuos función transferencia (Consumo frente a PIB).

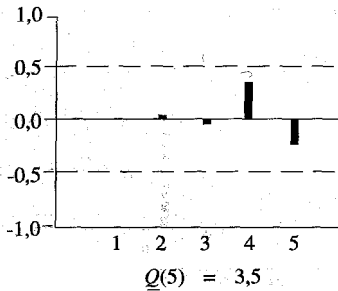


FIGURA 13.29. fas residuos.

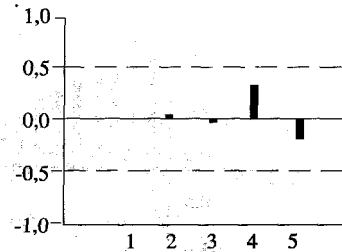


FIGURA 13.30. fap residuos.

elevado en el segundo trimestre de 1987, en que se llevó a cabo un cambio de metodología en la encuesta. Tras incorporar una intervención del tipo escalón en dicho trimestre, se llegó al modelo:

$$(1 + 0,62L^4 + 0,46L^8)\nabla^2\nabla_4 \text{OCUP}_t = 259,01 \xi_{87,2}^s + (1 - 0,53)a_t$$

(0,08)
(0,07)
(35,9)
(0,09)

con

$$T = 115, \quad \hat{\sigma} = 56,3; \quad a = -6,3(5,0)$$

sin que los residuos, ya sea directamente a través de la Figura 13.36 o de sus fas y fap (Figuras 13.37 y 13.38), mostrasen indicios de mala especificación.

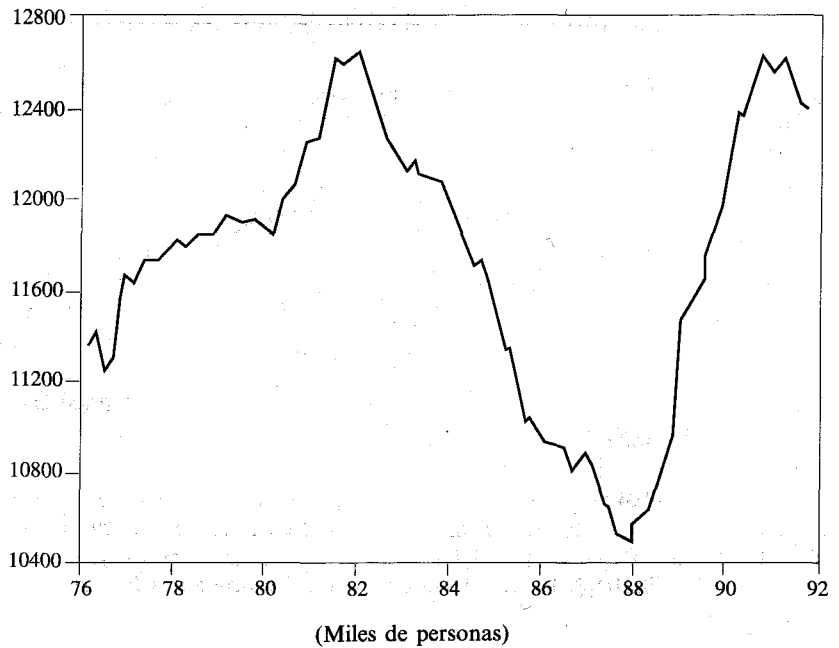


FIGURA 13.31. Ocupados EPA.

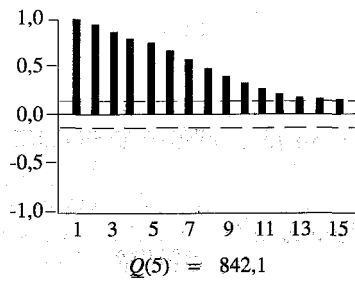


FIGURA 13.32. fas ocupados.

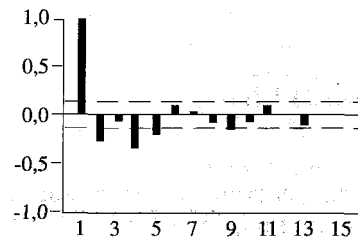


FIGURA 13.33. fap ocupados.

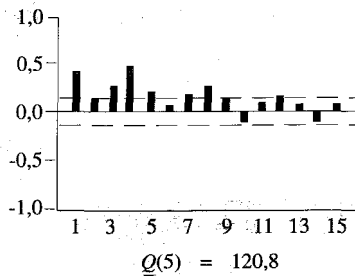


FIGURA 13.34. fas ∇ ocupados.

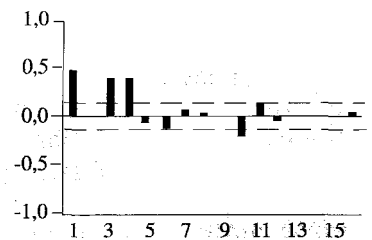


FIGURA 13.35. fap ∇ ocupados.

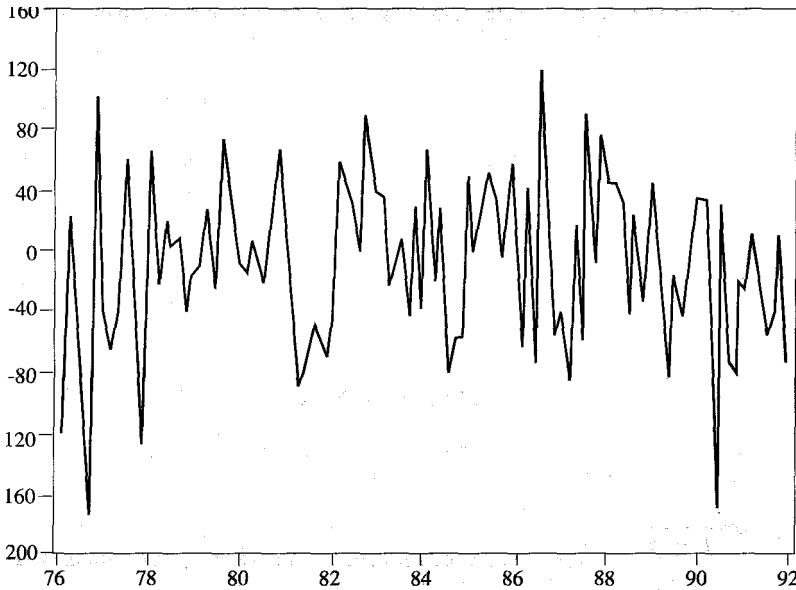


FIGURA 13.36. Residuos del modelo de ocupados.

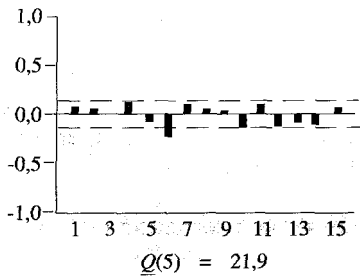


FIGURA 13.37. fas residuos ocupados.

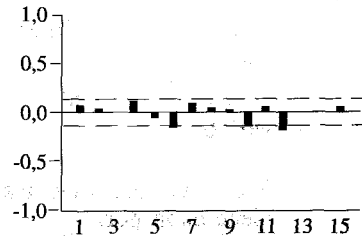


FIGURA 13.38. fap residuos ocupados.

El signo positivo del coeficiente de L^8 en el polinomio AR(2) estacional sugiere nuevamente una estructura cíclica en esta variable, en coherencia con la idea de que el número de ocupados evoluciona con variables del ciclo económico, como el PIB. De acuerdo con nuestras estimaciones, el período del ciclo en ocupados es de 5,75 años, algo por encima del estimado con observaciones anuales para el PIB.

PROBLEMAS

Problema 13.1. Supongamos que se ha especificado y estimado un modelo ARIMA(p, d, q) con $d = 2$, del que se han obtenido unas predicciones para los valo-

res futuros de z_t , donde $z_t = \Delta^2 y_t$. Probar que las predicciones correspondientes a la variable original y_t pueden recuperarse mediante la fórmula:

$$E_T y_{T+k} = y_T + k(y_T - y_{T-1}) + kE_T z_{T+1} + (k-1)E_T z_{T+2} + \dots + E_T z_{T+k}$$

Problema 13.2. Probar que las predicciones de los valores futuros de una variable y_t con estructura ARIMA(1, 1, 0) obedecen la relación:

$$\begin{aligned} E_T y_{T+1} &= \delta + (1 + \phi)y_T - \phi y_{T-1} \\ E_T y_{T+2} &= \delta(2 + \phi) + (1 + \phi + \phi^2)y_T - \phi(1 + \phi)y_{T-1} \end{aligned}$$

y, en general:

$$E_T y_{T+k} = E_T y_{T+k-1} + \phi E_T z_{T+k-1} + \delta$$

donde $z_t = \Delta y_t$. Utilizando este último resultado, probar que, según tiende el horizonte de predicción k a infinito, se espera que este proceso incremente a una tasa constante de $\frac{\delta}{1 - \phi}$.

Problema 13.3. Obtener las expresiones analíticas para las predicciones 1, 2, 3 y k periodos hacia adelante en un proceso ARIMA(1, 1, 1).

Problema 13.4. Demostrar que el valor de \hat{y}_{T+1}^T que minimiza la función objetivo $E_T[(y_{T+1} - \hat{y}_{T+1}^T)^2]$ es $\hat{y}_{T+1}^T = E_T y_{T+1}$.

Problema 13.5. Describir cómo se llevaría a cabo la estimación de un modelo ARMA(2, 1).

Explicar la utilización de la expansión en serie de Taylor en este proceso, las variables auxiliares a construir, y el modelo lineal a estimar.

Problema 13.6. Demostrar que la función de autocorrelación de un proceso MA(q) viene dada por:

$$\rho_k = 0 \begin{cases} \frac{-\theta_s + \theta_1 \theta_{s+1} + \dots + \theta_{q-s} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{si } k = 1, 2, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

Problema 13.7. Hallar la función de autocorrelación de un proceso ARMA(2, 1) como función de los parámetros de dicho proceso univariante.

Problema 13.8. Probar que las mismas expresiones para las predicciones de modelos univariantes que aparecen en la Sección 13.9 pueden obtenerse si se obtiene primero la «representación de medias móviles» del modelo univariante:

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} - \psi_3 \varepsilon_{t-3} - \dots$$

Problema 13.9. Probar que si en la estimación de un modelo MA(1) para la variable y_t los residuos u_t presentan evidencia de seguir, a su vez, un proceso MA(1), el modelo que debería especificarse para la variable y_t debería ser un modelo MA(2). ¿Qué re-

lación existirá entre los parámetros de los modelos previos para y_t y u_t y el modelo especificado finalmente para y_t ?

Problema 13.10. Explicar en detalle cómo puede utilizarse la función de autocorrelación simple para estimar los parámetros de un modelo AR(2), así como la varianza de su término de error.

Problema 13.11. Explicar la utilización de la función de autocorrelación simple para estimar el parámetro de un modelo MA(1). ¿Cuál es el rango de valores admisibles de dicho parámetro?

Problema 13.12. Probar que si $y_t = \delta + \rho y_{t-1} + u_t$ con $|\rho| > 1$, entonces y_t admite la representación:

$$y_t = -\frac{\delta}{\rho - 1} - \sum_1^{\infty} \left(\frac{1}{\rho^j}\right) u_{t+j}$$

Problema 13.13. Explicar cómo estimar el modelo $y_t = \varepsilon_t - \theta \varepsilon_{t-1}$ por mínimos cuadrados condicionados (o máxima verosimilitud condicionada).

Problema 13.14. El desarrollo de la Sección 13.10 acerca de la estimación de modelos univariantes es tan sólo una aproximación, puesto que ignora las derivadas $\frac{\theta \varepsilon_{t-1}}{\theta \theta_i}$, $\frac{\theta \varepsilon_{t-2}}{\theta \theta_i}$, $i = 1, 2$, en la expresión [13.7].

Probar que si se tienen en cuenta estas derivadas, se llega a la relación:

$$\varepsilon_t = \varepsilon_t^0 + (\theta_1 - \theta_1^0) \frac{y_{t-1}}{(1 - \theta_1^0 L - \theta_2^0 L^2)^2} + (\theta_2 - \theta_2^0) \frac{y_{t-2}}{(1 - \theta_1^0 L - \theta_2^0 L^2)^2}$$

¿Cómo podría llevarse a cabo la estimación del modelo MA(2) con esta formulación más completa?

Problema 13.15. Probar que la estimación del modelo ARMA(1, 1), de acuerdo con el desarrollo de la Sección 13.10, se reduce a la estimación del modelo:

$$\varepsilon_t^0 + \phi^0 y_{t-1} - \theta^0 \varepsilon_{t-1}^0 = \phi y_{t-1} - \theta \varepsilon_{t-1}^0 + \varepsilon_t$$

Problema 13.16. Probar que para llevar a cabo la estimación de un modelo ARMA(1, 1), de acuerdo con el Problema 13.14, sería preciso estimar el modelo:

$$w_t = \phi x_{1t} + \theta x_{2t} + u_t$$

donde:

$$\begin{aligned} x_{1t} &= \theta^0 x_{1,t-1} + y_{t-1} \\ x_{2t} &= 2\theta^0 x_{2,t-1} - (\theta^0)^2 x_{2,t-2} - y_{t-1} + \phi^0 y_{t-2} \\ w_t &= \varepsilon_t^0 + \phi^0 x_{1t} + \theta^0 x_{2t} \\ u_t &= \varepsilon_t \end{aligned}$$

Problema 13.17. Obtener las expresiones para los errores de predicción y sus varianzas, que aparecen en la Sección 13.9.

Problema 13.18. Explique cómo obtendría sus previsiones a partir de modelos univariantes que han precisado de dos diferencias de orden regular (no estacional).

Problema 13.19. En el análisis empírico de la relación entre dos series temporales estacionarias, $y_t = v(B)x_t + N_t$, se han obtenido los siguientes resultados numéricos:

$$y_t = (1 - 0,3L)a_{yt}$$

$$(1 - 0,3L)x_t = a_{xt}$$

$$a_{yt} = \frac{2L - 2,6L^2 + 0,6L^3}{1 - 0,3L} a_{xt} + \xi_t$$

donde $\xi_t = (1 - 0,5L)a_t$, $\hat{\sigma}_\beta = 0,04$, $\hat{\sigma}_{ax} = 0,03$, $\hat{\sigma}_a = 0,001$ y $\bar{a} = 0,0$.

a) Calcule la función de respuesta al impulso, la función de correlación cruzada y la ganancia correspondiente a $v(L)$.

b) De haber conocido previamente la función de correlación cruzada que ha obtenido en a), ¿qué función de transferencia había especificado y qué preestimaciones habría dado a sus parámetros?

c) ¿Qué proceso seguirá N_t ? Calcule una medida de la bondad de ajuste del modelo de función de transferencia.

d) A la vista de los resultados anteriores, ¿relacionaría las variables y_t y x_t en niveles? ¿Cómo cambiaría la especificación del modelo para lograr una mayor eficiencia en la estimación?

Problema 13.20. Tras estimar los siguientes modelos univariantes con observaciones anuales del IPC y los ALP:

a) $\nabla \text{IPC}_t = (1 - 0,45L)a_t$.

b) $(1 - 0,80L - 0,60L^2)\nabla \text{ALP}_t = (1 - 0,60L)a_t$.

Se utilizó el modelo b) para preblanquear tanto la serie del IPC como la de los ALP. A continuación se estimó la función de correlación cruzada entre las series transformadas, obteniéndose:

Correlación entre IPC_t^* y ALP_{t-k}^* :

$k =$	0	1	2	3	4	5	6	7	8	9	10
	-0,03	0,80	0,83	0,60	0,35	0,20	0,10	-0,02	0,01	0,03	0,02

Especificar posibles modelos de función de transferencia entre IPC y ALP (incluido el modelo del ruido).

Problema 13.21. Con objeto de prever el nivel de la Formación Bruta de Capital Fijo (I_t), se ha identificado el siguiente modelo de transferencia, que incluye como input la tasa de variación del stock nominal de ALP (M_t):

$$\ln I_t = (\omega_0 - \omega_1 L)\nabla \ln M_t + N_t$$

$$\nabla N_t = (1 - \phi L)^{-1} a_t$$

1. Describa, con todo detalle, las expresiones analíticas que conducirían a la estimación de máxima verosimilitud de dicho modelo, mediante el algoritmo de Gauss-Newton.

Si se define la siguiente transformación de variables:

$$\begin{aligned} X_{1t} &= (1 - 0,1L)\nabla^2 \ln M_t \\ X_{2t} &= (0,1L - 1)\nabla^2 \ln M_{t-1} \\ X_{3t} &= \nabla \ln I_{t-1} - (0,5 + 0,2L)\nabla^2 \ln M_{t-1} \end{aligned}$$

se obtienen, a partir de la información muestral, los siguientes valores:

$$\begin{aligned} n &= 100; & x_1 &= x_2 = x_3 = x_4 = a_0 = 0; \\ \Sigma x_{1t}^2 &= 2; & \Sigma x_{2t}^2 &= 1; & \Sigma x_{3t}^2 &= 2; \\ \Sigma x_{1t}x_{2t} &= \Sigma x_{2t}x_{3t} = 0; & \Sigma x_{1t}x_{3t} &= 1; \\ \Sigma x_{1t}a_{0t} &= 2,5; & \Sigma x_{2t}a_{0t} &= -0,5; & \Sigma x_{3t}a_{0t} &= 2; \\ \Sigma a_{0t}^2 &= 4,75 \end{aligned}$$

2. Obtenga estimaciones consistentes y eficientes de los parámetros ω_0 , ω_1 , ϕ , σ_a^2 y R^2 .

3. Contraste la hipótesis de Superneutralidad Monetaria: $g = 0$.

4. Discuta el efecto que sobre $\ln I_t$ tendrá un cambio permanente unitario sobre el nivel de $\ln M_t$.

Problema 13.22. a) Explique cómo pueden utilizarse los dos primeros valores estimados de la función de autocorrelación simple para obtener estimaciones de ϕ_1 y ϕ_2 en el modelo de AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

donde ε_t es ruido blanco.

b) Obtenga la expresión analítica de σ_y^2 como función de los parámetros ϕ_1 , ϕ_2 y σ_ε^2 en el modelo anterior.

c) Obtenga las expresiones analíticas que sirven para la predicción, varios períodos hacia el futuro, en un modelo ARMA(1, 1) con término constante. Halle una expresión genérica para $E_T y_{T+k}$, $k > 0$.

d) Suponga que para analizar la serie temporal y_t se ha estimado un modelo ARIMA(p, d, q) con $d = 2$. De dicho modelo se han obtenido predicciones: $E_T z_{T+1}$, $E_T z_{T+2}$, $E_T z_{T+3}$, donde $z_t = \nabla^2 y_t$. Obtenga expresiones para $E_T y_{T+1}$, $E_T y_{T+2}$, $E_T y_{T+3}$ como función únicamente de $E_T z_{T+i}$, $i = 1, 2, 3$ y de la información muestral acerca de la variable y_t .

e) Obtenga los cuatro primeros términos de la representación AR de un modelo MA(2) invertible.

f) ¿Por qué es interesante que un modelo MA(q) sea invertible? ¿Cuáles son las condiciones de invertibilidad de un modelo MA(2)? ¿Cuáles son las condiciones de estacionariedad de dicho modelo?

Problema 13.23. Para llevar a cabo el seguimiento mensual de las Disponibilidades Líquidas (M3) en España, se ha especificado el modelo:

$$\Delta^2 y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

donde y_t denota el logaritmo de M3 y se han obtenido las estimaciones:

$$\hat{\theta}_1 = 0,60; \quad \hat{\theta}_2 = 0,30$$

a) Utilizando este modelo, obtener previsiones para el logaritmo de M3 para los meses de junio a diciembre de 1988, si se dispone de la siguiente historia acerca de la realización de la serie y su predicción hecha con un mes de antelación:

Mes	Valor real de M3	Predicción hecha el mes anterior
5/87	22.913	22.881
6/87	22.434	22.615
7/87	23.852	23.450
8/87	23.686	23.782
9/87	23.932	23.901
10/87	23.753	23.842
11/87	23.747	23.810
12/87	24.781	24.805
1/88	24.255	24.310
2/88	24.357	24.342
3/88	24.918	24.902
4/88	24.936	24.950
5/88	25.115	25.164

b) ¿Cómo cambiaría su respuesta si el modelo especificado hubiese sido

$$\nabla \nabla^{12} y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

con las mismas estimaciones y datos anteriores?

Problema 13.24. Para estimar el modelo ARIMA siguiente:

$$(1 - \phi_1 L) \nabla Z_t = (1 - \theta_1 L) a_t$$

se dispone de la siguiente información muestral:

t	∇Z_t
1986	1
1987	0
1988	-1
1989	0
1990	-1
1991	1

tomándose como condiciones iniciales:

$$\begin{aligned} a_{1985} &= 0 \\ \nabla Z_{1984} &= \nabla Z_{1985} = 0 \\ \left(\frac{\partial a_t}{\partial \beta} \right)_{\beta_{0,1984}} &= \left(\frac{\partial a_t}{\partial \beta} \right)_{\beta_{0,1985}} = 0 \\ \beta_0 &= (\phi_1 = 0,8, \theta_1 = 0,5)' \end{aligned}$$

1. Complete la primera iteración del algoritmo de Gauss-Newton. Calcule la matriz de varianzas y covarianzas de las estimaciones y su R^2 .
2. Calcule previsiones para $t = 1993$ y 1994 y los riesgos asociados.

Problema 13.25. Derive las condiciones de optimalidad necesarias para obtener el estimador de máxima verosimilitud del modelo:

$$\begin{aligned} y_t &= \mathbf{x}_t' \beta + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, T \end{aligned}$$

Problema 13.26. Pruebe que el modelo

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned}$$

donde $\varepsilon_t, \eta_t, \zeta_t$ son ruidos blancos independientes entre sí, puede formularse como una ARIMA(0, 2, 2).

Problema 13.27. Igualando las funciones de autocovarianza de ambos procesos, halle el valor del parámetro θ de modo que el proceso estocástico $z_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}$, donde η_t es i. i. d., $N(0, \sigma_\eta^2)$ y ε_t es i. i. d., $N(0, \sigma_\varepsilon^2)$, admita la representación MA(1): $z_t = \xi_t - \theta \xi_{t-1}$, donde ξ_t es i. i. d., $N(0, \sigma_\xi^2)$. Halle asimismo la varianza de ξ_t .

Problema 13.28. Describa la aplicación del algoritmo Gauss-Newton a la estimación del modelo de función de transferencia:

$$\begin{aligned} y_t &= \frac{\omega_0 + \omega_1 L + \omega_2 L^2}{1 - \delta L} x_t + u_t \\ u_t &= \frac{1}{(1 - \delta L)(1 - \rho L)} a_t \end{aligned}$$

Problema 13.29. Suponga que se ha estimado un modelo ARIMA(0, 2, 1) con parámetro 0,30 para representar la serie trimestral de población activa española, cuyos valores durante el año 1992 fueron:

Trimestre	(1992, I)	(1992, II)	(1992, III)	(1992, IV)
Población activa (Miles de personas)	14.000	14.150	14.200	14.350

Elabore previsiones para la población activa durante los cuatro trimestres de 1993, sabiendo que la previsión que se había hecho en (1992, III) para (1992, IV) era de 14.500.

Problema 13.30. Tras estimar el modelo econométrico $y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$ con veinte años de observaciones trimestrales, por mínimos cuadrados, se calcularon las funciones de autocorrelación simple y parcial de los residuos, obteniéndose:

$k =$	0	1	2	3	4	5	6	7	8
<i>fas.</i>	1	0,50	0,20	0,04	-0,15	-0,21	-0,14	0,12	0,10
<i>fap.</i>	1	0,50	0,45	-0,15	0,18	0,12	-0,21	0,10	0,05

- a) Qué indican estas funciones acerca de la estructura estocástica del término de error? Obtenga estimaciones para los parámetros de dicha estructura.
- b) ¿Cómo debería estimarse eficientemente el modelo anterior con esta información adicional?
- c) ¿Qué relación existe entre el estimador que se acaba de proponer y el estimador de máxima verosimilitud?
- d) ¿Cómo puede calcularse una estimación para la varianza σ_u^2 ?

Problema 13.31. Tras estimar el modelo $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$ se ha identificado una estructura ARMA(1, 1) para los residuos MCO:

$$\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

y, como consecuencia, se decide estimar simultáneamente el sistema:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \end{cases}$$

Describa, en todo detalle, cómo podría obtener estimadores eficientes de $\beta_0, \beta_1, \beta_2, \rho$ y θ . ¿Cómo utilizaría sus series temporales para poner en práctica dicho procedimiento? ¿Podría pensarse que un procedimiento que estimase primero los coeficientes β y después ρ y θ conjuntamente pudiera ser totalmente eficiente? ¿Cuál sería la matriz de covarianzas de sus estimaciones en el procedimiento eficiente que ha propuesto?

Problema 13.32. Explique los detalles de la utilización práctica del método de «scoring» para la estimación del modelo

$$y_t = \phi y_{t-1} + u_t$$

donde u_t es *i. i. d.*, $N(0, \sigma_u^2)$. ¿Cuál será la matriz de covarianzas de los estimadores $\hat{\phi}$ y $\hat{\sigma}_u^2$?

CAPITULO 14

REGRESION CON VARIABLES NO ESTACIONARIAS

14.1. INTRODUCCION

14.1.a. Primeras definiciones

La observación de un agregado económico como la renta de un país o la cantidad de dinero en circulación muestra una *tendencia creciente* a lo largo del tiempo. Otras variables económicas, como los tipos de interés o la tasa de inflación, quizá no tienen tal tendencia, pero presentan una clara inclinación a permanecer durante largos períodos de tiempo por encima o por debajo de su valor central en la muestra. Decimos entonces que la variable *deambula* alrededor de dicho valor central. Ambas son situaciones que reflejan la potencial *no estacionariedad* de la variable objeto de análisis.

Desde el punto de vista estadístico, la presencia de una tendencia temporal, lineal o no, de naturaleza determinista, no representa una dificultad importante. Siendo determinista, y supuesta estable, puede estimarse con una razonable precisión para, posteriormente, ser sustraída de la serie temporal original. Así se obtiene el *componente puramente estocástico* de la variable económica que se pretende analizar. Cuando realizamos tales ejercicios, observamos con frecuencia que el componente estocástico de la variable *deambula* alrededor de su valor central, es decir, nos hemos reducido al segundo de los casos antes mencionados.

Tales rachas de valores sistemáticamente por encima o por debajo de la media, se deben a la *existencia de una raíz unitaria* en la estructura estocástica de la variable. Si existe dicha raíz en su representación autorregresiva⁽¹⁾: $A(L)y_t = \varepsilon_t$, ésta puede descomponerse:

$$A(L)y_t = A^*(L)(1 - L)y_t = A^*(L)\Delta y_t$$

⁽¹⁾ Que será teóricamente de orden infinito si el modelo ARIMA de y_t tiene un término invertible de medias móviles. Siguiendo con la convención habitual, denotamos por Δy_t la primera diferencia de la variable, que en el capítulo anterior denotamos por ∇y_t .

Si no hay más raíces unitarias en el polinomio $A^*(L)$, entonces la variable diferencia Δy_t es estacionaria, y oscilará de modo puramente aleatorio y, en particular, sin deambular, alrededor de su valor central.

Veamos por qué suele hablarse de las raíces unitarias como de *tendencias estocásticas*. El caso más sencillo de un proceso AR(1), cuando existe una raíz unitaria, puede representarse:

$$y_t = y_{t-1} + c + \varepsilon_t$$

que conduce a:

$$y_t = y_0 + ct + e_t$$

donde $e_t = \sum_1^t \varepsilon_s$, supuesto que el valor inicial de la variable y_0 sea una constante dada, aunque desconocida. En consecuencia, el proceso y_t admite una tendencia lineal, si bien en una representación en que el término de error es la acumulación temporal de un ruido blanco y, por consiguiente, *con una varianza creciente en el tiempo, sin límite*. Ello sugiere el tipo de comportamiento que cabe esperar de tal estructura estocástica, que definimos a continuación:

Definición 14.1. Una variable es *integrada de orden d* si su diferencia de orden d admite una representación ARMA estacionaria e invertible. Diremos en tal caso que x_t es $I(d)$.

Por ejemplo, una variable cuya representación autorregresiva tiene una única raíz unitaria es $I(1)$. La suma de variables $I(d)$ e $I(d')$ con $d < d'$ es $I(d')$; si x_t es $I(d)$, entonces $ax_t + b$ es también $I(d)$, siempre que $a \neq 0$.

Una variable $I(0)$ es estacionaria y: *a)* tiene varianza finita, *b)* dicha varianza es independiente del tiempo, *c)* una perturbación sobre su nivel tiene un efecto transitorio, pero no permanente, *d)* los valores sucesivos de su función de autocorrelación decrecen rápidamente hacia cero; además, *e)* no presentará intervalos en que deambule, de modo que el número de periodos que transcurren entre dos cruces consecutivos con su valor medio debe ser reducido.

Por el contrario, si x_t es $I(d)$ con $d \geq 1$: *a)* su varianza aumenta generalmente al transcurrir el tiempo, tendiendo a infinito, *b)* una perturbación sobre su nivel tiene efectos permanentes, *c)* los valores de su función de autocorrelación decrecen sólo muy lentamente hacia cero, *d)* el número esperado de periodos que deben transcurrir entre dos cruces consecutivos con su valor central muestral es infinito.

Consideremos, por simplicidad, que el modelo AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 2, \dots, T$$

evoluciona a partir de un valor inicial y_1 dado, determinista, y que $\varepsilon_t \sim N(0, \sigma^2)$ para todo t , independiente en el tiempo. En dicho modelo, el estimador MCO

de ρ es *siempre consistente*, cualquiera que sea el verdadero valor de ρ (Rubin, 1950).

Las dificultades aparecen cuando se pretende contrastar la hipótesis de existencia de raíz unitaria en y_t : $H_0: \rho = 1$, es decir, de no estacionariedad de y_t , mediante el estimador MCO $\hat{\rho}$ porque, bajo H_0 , el estadístico $(\hat{\rho} - 1)/\sqrt{\text{Var}(\hat{\rho})}$ no tiene la distribución « t » que suele asociársele. Sólo sigue una distribución t cuando el proceso y_t es estacionario, es decir, cuando $|\rho| < 1$.

En efecto, mientras que en el caso estacionario ($|\rho| < 1$) el estadístico t de Student converge a una Normal(0, 1) cuando el número de grados de libertad tiende a infinito (White, 1958):

$$\sqrt{\frac{\sum_2^T y_{t-1}^2}{\hat{\sigma}^2}} (\hat{\rho} - \rho) \xrightarrow{d} N(0, 1)$$

en el caso explosivo ($|\rho| > 1$) no se tiene tal resultado, si bien puede afirmarse que:

$$|\rho|^n \left(\frac{\hat{\rho} - \rho}{\rho^2 - 1} \right) \xrightarrow{d} \text{Cauchy}$$

El caso más interesante de existencia de raíz unitaria, $|\rho| = 1$, es más complejo y se tiene (véase Fuller, 1976) que $T(\hat{\rho} - 1)$ converge en distribución a un cociente de integrales de Wiener, que no estudiamos en este texto. A modo de ilustración, es interesante apuntar que $\text{Prob}(|\hat{\rho}| < 1/|\rho| = 1)$ tiende a 0,68 cuando el tamaño muestral tiende a infinito.

Puede, por tanto, caracterizarse la distribución de probabilidad asintótica del estimador MCO del modelo AR(1) en todos los casos, excepto cuando $|\rho| = 1$, que representa, por consiguiente, una discontinuidad. Tal discontinuidad se produce, sin embargo, estrictamente en el límite, ya que ligeramente por encima o por debajo de tal punto del espacio paramétrico la distribución asintótica pertenece a una determinada clase, Normal o Cauchy.

14.1.b. Regresión entre procesos no estacionarios

Existen muchas situaciones en Economía en que variables que son no estacionarias muestran una relación estable a través del tiempo, lo que sugiere la existencia de una relación de equilibrio a largo plazo entre ellas. Por ejemplo, para distintos países, el Consumo agregado C_t y la Renta disponible Y_t son variables $I(1)$; sin embargo, examinado a largo plazo, el Consumo viene a ser una proporción bastante estable de la Renta, por lo que existe una constante β para la que la diferencia $C_t - \beta Y_t$ es estacionaria.

La teoría de la Paridad del Poder Adquisitivo propone que, en ausencia de costes de transacción u otras fricciones, bienes análogos deberían tener el mismo precio efectivo en distintos países. Así, si P_t y P_t^* denotan el nivel de precios en España y Estados Unidos respectivamente y E_t denota el tipo

de cambio (ptas./\$), debería tenerse para bienes análogos que $P_t = E_t P_t^*$ y, tomando logaritmos:

$$p_t = e_t + p_t^*$$

que debería cumplirse exactamente, donde las letras minúsculas denotan los logaritmos de sus análogos mayores. Sin embargo, fricciones como las indicadas pueden impedir tal exactitud, por lo que una versión más débil de la teoría propone que la diferencia

$$z_t = p_t - e_t - p_t^*$$

sea estacionaria.

Todas estas son situaciones en que un conjunto de variables económicas están sujetas a una relación estable a largo plazo que, siendo lineal, puede representarse por un modelo de regresión: $y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$. Si las variables \mathbf{x}_t e y_t son no estacionarias, la regresión anterior sólo puede reflejar una relación de equilibrio si las posibles tendencias de las variables \mathbf{x}_t e y_t evolucionan conjuntamente, de modo que los residuos del modelo sean estacionarios, a pesar de que las variables que en él aparecen no lo sean.

Los comentarios de la sección anterior sugieren que los procedimientos de inferencia habituales dejan de ser válidos en un modelo autorregresivo cuando la variable que en él aparece no es estacionaria. Pero esta cuestión es aún más importante de lo que a primera vista puede parecer, por cuanto que los problemas pueden surgir también en modelos de regresión, cuando las variables que en él se incluyen no sean estacionarias. Un trabajo pionero en este área es el de Granger y Newbold (1974), en el que los autores generaron datos para dos variables, x_t e y_t , con estructura de camino aleatorio cuyas respectivas innovaciones eran independientes. Es decir, no existía ninguna relación entre las series temporales así generadas para ambas variables, a pesar de lo cual, al estimar por MCO un modelo de regresión lineal:

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T \quad [14.1]$$

estos autores observaron con mucha frecuencia que el estadístico t del coeficiente β conducía a rechazar la hipótesis nula de ausencia de relación entre ambas variables: $H_0: \beta = 0$, lo que comenzó a conocerse con el problema de *regresión espúrea*.

Más recientemente, se ha desarrollado la teoría estadística de regresión entre variables no estacionarias, que ha permitido explicar este tipo de resultados. En particular, ahora se sabe (Phillips, 1987) que en una regresión como [14.1], con variables no estacionarias, el estadístico «del tipo t de Student» que se obtiene dividiendo $\hat{\beta} - \beta_0$, donde β_0 es el valor de β que se contrasta, por la desviación típica de $\hat{\beta}$, no se distribuye como una t de Student. En realidad, en algunas ocasiones, los estadísticos «tipo t » tienden a infinito con el tamaño muestral, por lo que al aumentar éste tenderá a

rechazarse la hipótesis nula de ausencia de relación en [14.1]. Esto constituye una posible explicación al problema de la regresión espúrea.

Una situación así, en que los procedimientos de inferencia habituales están sesgados hacia aceptar la existencia de relación entre variables no estacionarias, viene generalmente acompañada de un R^2 aceptable e incluso, en ocasiones, muy elevado, junto con un estadístico Durbin-Watson muy bajo, por razones que luego veremos.

Es fácil entender que el problema es importante, porque una variable explicativa x_t en el modelo [14.1] que no sea estacionaria no satisfará la propiedad $plim(T^{-1}\Sigma_1^T x_t^2) = \text{constante}$, que es uno de los supuestos que justifican la utilización de los procedimientos de inferencia habituales, basados en el estimador MCO.

De este modo, la mayor parte de los resultados estadísticos habituales y, en particular, aquellos relacionados con las propiedades del estimador mínimo-cuadrático en modelos lineales dependen, precisamente, de supuestos, no siempre explícitos, acerca de la estacionariedad del término de error del modelo correspondiente. Tras puntualizar la importancia que tiene tal supuesto para llevar a cabo el análisis de inferencia estadística en un modelo de regresión, la segunda sección de este capítulo se dedica a la contrastación de la existencia de raíz unitaria en la serie temporal correspondiente a una variable. La Sección 14.7 aplica dichos resultados al análisis de estacionariedad de los residuos de un modelo de regresión, al objeto de garantizar que la utilización estadística de las estimaciones obtenidas es totalmente rigurosa. El capítulo termina con la discusión de algunas aplicaciones.

En este último contexto de existencia de raíz unitaria, los trabajos de Dickey y Fuller (1979) y (1981) han resultado enormemente influyentes. En ellos se establece una metodología sencilla para el contraste de existencia de raíz unitaria en un proceso estocástico y_t , que examinamos a continuación.

14.2. CONTRASTES DE RAZ UNITARIA DE DICKEY Y FULLER

Comenzamos pasando revista en esta sección a los contrastes de raíz unitaria en un proceso AR(1), para generalizarlos a otros tipos de estructuras estocásticas en secciones posteriores. Examinamos sucesivamente tres versiones del modelo, según cuáles sean sus componentes deterministas:

A) Modelo sin componentes deterministas:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.2]$$

con y_1 fijo y $\varepsilon_t \sim N(0, \sigma^2)$, independiente en el tiempo. Si no se rechaza la hipótesis nula, se dice que el proceso y_t es *estacionario en diferencias*.

B) Incluyendo un término constante en el modelo AR(1):

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.3]$$

y manteniendo el supuesto de que y_1 es fijo y $\varepsilon_t \sim N(0, \sigma^2)$, independiente en el tiempo. Si no se rechaza la hipótesis nula conjunta: $H_0: \alpha = 0, \rho = 1$, se tiene nuevamente que el proceso y_t es *estacionario en diferencias*.

C) Incluyendo, además de la constante, una tendencia lineal:

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.4]$$

con y_1 fijo y $\varepsilon_t \sim N(0, \sigma^2)$, independiente en el tiempo. Si rechazando la hipótesis nula concluimos que $\rho < 1$ y $\beta \neq 0$, se dice que el proceso y_t es *estacionario alrededor de una tendencia lineal*.

En la práctica, se ha popularizado la utilización de los contrastes sobre el modelo AR especificado en primeras diferencias. Para ello, basta observar que la contrastación de la hipótesis nula $H_0: \rho = 1$ en el modelo [14.4] equivale a la contrastación de la hipótesis $H_0: v = 0$ en el modelo:

$$\Delta y_t = \alpha + \beta t + v y_{t-1} + \varepsilon_t \quad [14.5]$$

mientras que las variantes A) y B) se obtienen al imponer en [14.5] algunas de las restricciones $\alpha = 0, \beta = 0$.

Para contrastar la hipótesis nula $H_0: v = 0$ en el modelo

$$\Delta y_t = v y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.6]$$

con y_1 fijo y $\varepsilon_t \sim N(0, \sigma^2)$, independiente en el tiempo, puede utilizarse el estadístico t construido a partir del estimador MCO de v :

$$\frac{\hat{v}}{[\text{Var}(\hat{v})]^{1/2}} = \frac{\hat{v}}{\hat{\sigma}} \left(\sum_2^T y_{t-1}^2 \right)^{1/2}$$

pero, como ya hemos mencionado, su distribución bajo H_0 ya no es del tipo « t de Student», y hay que comparar su valor numérico, que denotamos por τ , con los que aparecen en la Tabla A.11 al final de este texto [Tabla 8.5.2 de Fuller (1976), pág. 373].

En cualquiera de estos tres modelos existe una forma alternativa de llevar a cabo el contraste de la hipótesis nula de existencia de raíz unitaria, comparando los valores del producto $T(\hat{\rho} - 1)$ con los de la Tabla A.12 [Tabla 8.5.1 de Fuller (1976), pág. 371]. Si el valor numérico del estadístico excede al de las tablas, rechazamos la hipótesis nula $H_0: \hat{\rho} = 1$, mientras que la mantenemos en otro caso.

Como resumen de dichas tablas, que aparecen al final de este libro, si queremos contrastar la hipótesis nula de existencia de raíz unitaria, utilizaremos los valores críticos que se muestran en las Tablas 14.1 y 14.2.

TABLA 14.1. Contraste de raíz unitaria

Valor crítico c tal que $P(\tau < c) = x$ para $x = 0,01; 0,025; 0,05$						
	$T = 25$			$T = 100$		
Nivel de significación: x	0,01	0,025	0,05	0,01	0,025	0,05
Modelo [14.6]	-2,66	-2,26	-1,95	-2,60	-2,24	-1,95
Modelo [14.7]	-3,75	-3,33	-3,00	-3,51	-3,17	-2,89
Modelo [14.8]	-4,38	-3,95	-3,60	-4,04	-3,73	-3,45

Por ejemplo, si se ha estimado el modelo [14.3] con 25 observaciones y se aplica un contraste de hipótesis convencional, se rechazaría la hipótesis nula de existencia de raíz unitaria si el estadístico t del retardo y_{t-1} resultase inferior a $-1,71$, que es el valor crítico para el contraste de una cola al 95 por 100 de dicha distribución. Sin embargo, de acuerdo con la Tabla 14.1, debe rechazarse H_0 sólo si el estadístico t estuviese por debajo de $-3,0$, es decir, mucho menos a menudo de lo que sugiere el análisis tradicional.

Si el modelo incluye una constante

$$\Delta y_t = \alpha + \nu y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.7]$$

puede contrastarse la hipótesis nula $H_0: \alpha = 0$ mediante el estadístico t de dicho coeficiente, que denotamos como $\tau_{\alpha\mu}$, que debe compararse con los valores críticos de la Tabla A.13 al final de este texto (Tabla I en Dickey y Fuller, 1981). Si no se rechaza la hipótesis nula, estamos en el modelo [14.6]. Por otra parte, el contraste de la hipótesis conjunta $H_0: \alpha = 0, \nu = 0$, es decir, que y_t es un paseo aleatorio, puede llevarse a cabo a partir del estadístico F

TABLA 14.2. Contraste de raíz unitaria

Valor crítico c tal que $P[T(\hat{\rho} - 1) < c] = x$ para $x = 0,01; 0,025; 0,05$						
	$T = 25$			$T = 100$		
Nivel de significación: x	0,01	0,025	0,05	0,01	0,025	0,05
Modelo [14.6]	-11,9	-9,3	-7,3	-13,3	-10,2	-7,9
Modelo [14.7]	-17,2	-14,6	-12,5	-19,8	-16,3	-13,7
Modelo [14.8]	-22,5	-19,9	-17,9	-27,4	-23,6	-20,7

correspondiente, Φ_1 , sólo que comparando esta vez con la Tabla A.16 [la Tabla IV en Dickey y Fuller (1981)].

En el caso del modelo con constante y tendencia:

$$\Delta y_t = \alpha + \beta t + \nu y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.8]$$

con y_1 fijo y $\varepsilon_t \sim N(0, \sigma^2)$, independiente en el tiempo, la significación de la constante $H_0: \alpha = 0$ puede contrastarse mediante el estadístico t del estimador MCO de dicho parámetro, τ_{α} , utilizando la Tabla A.14 [Tabla II de Dickey y Fuller (1981)].

La significación estadística de la pendiente determinista $H_0: \beta = 0$ puede contrastarse mediante el estadístico t del estimador MCO de dicho parámetro, τ_{β} , utilizando la Tabla A.15 [Tabla III de Dickey y Fuller (1981)]. Si no se rechaza la hipótesis nula, estamos en la especificación [14.7] anteriormente discutida.

Por último, la existencia de una raíz unitaria puede contrastarse en una versión restringida: $H_0: \alpha = 0, \beta = 0, \nu = 0$ (es decir, y_t es paseo aleatorio), utilizando el estadístico F correspondiente, Φ_2 , y la Tabla A.17 [Tabla V de Dickey y Fuller (1981)], mientras que en su versión menos restrictiva: $H_0: \beta = 0, \nu = 0$ (y_t es paseo aleatorio con deriva), puede contrastarse utilizando el estadístico F correspondiente, Φ_3 —diferente del anterior—, y la Tabla A.18 [Tabla VI de Dickey y Fuller (1981)]. La utilización de estas tablas, construidas por Dickey y Fuller, en vez de las habituales, puede representar diferencias importantes. Así, para este contraste conjunto, los valores críticos de Dickey-Fuller para el estadístico F al 5 por 100 son 7,24; 6,73 y 6,49, según que el tamaño muestral empleado sea $T = 25, 50$ ó 100 . Por otra parte, los valores de la tablas habituales $F_{2,22}, F_{2,47}$ y $F_{2,97}$ son 3,44; 3,20 y 3,10, por lo que su utilización conduciría a rechazar la hipótesis nula de existencia de raíz unitaria *demasiado frecuentemente*.

Al efectuar estos contrastes de existencia raíz unitaria, es decir, de significación del parámetro ν en los modelos [14.2], [14.3] y [14.4], hay que tener en cuenta que puesto que la hipótesis alternativa es de estacionariedad, es decir, $\rho < 1$, el contraste debe ser de una cola. En segundo lugar, que es fundamental utilizar las tablas a las que nos hemos venido refiriendo, porque los estadísticos «de tipo t » y «de tipo F » para el contraste de las hipótesis individuales y conjuntas que hemos mencionado no tienen las distribuciones t o F que habitualmente les asignaríamos. Lamentablemente, ya hemos visto que los valores críticos para estos contrastes dependen tanto del tamaño muestral como de la especificación del modelo en que se llevan a cabo, por lo que hay que poner especial cuidado en utilizar la tabla adecuada en cada caso.

Es importante observar que los valores de los estadísticos de Dickey-Fuller que aparecen en estas tablas están obtenidos bajo los supuestos: *a*) el proceso sobre el que se contrasta la existencia de una raíz unitaria tiene una estructura AR(1), y *b*) no existe autocorrelación en su término de error. Por otra parte, las distribuciones empíricas que se presentan en las tablas al final del texto, y se resumen en nuestra Tabla 14.1, no cambian al introducir en el modelo retardos de la variable en diferencias: $\Delta y_{t-1}, \Delta y_{t-2}$, etc., lo que da lugar al

contraste conocido como *Dickey-Fuller ampliado* (que denotamos en lo sucesivo por DFA) en el que se estima una regresión que en el caso del modelo [14.3] anterior sería:

$$\Delta y_t = \alpha + \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \delta_3 \Delta y_{t-3} + \delta_4 \Delta y_{t-4} + u_t$$

y se compara el estadístico t obtenido para ρ con las tablas de Dickey y Fuller (las mismas de antes). El contraste DFA se utiliza para eliminar, en lo posible, la autocorrelación del término de error, pues, como hemos dicho, éste es un supuesto bajo el que fueron obtenidas las distribuciones empíricas del estadístico t . Esto es especialmente importante cuando se trabaja con datos trimestrales, por ejemplo, en cuyo caso la inclusión de cuatro diferencias retardadas de la variable endógena producirá, generalmente, un residuo prácticamente libre de correlación.

Al contrastar la posible existencia de una raíz unitaria, debe escogerse uno de los modelos [14.2], [14.3] o [14.4] como representación más adecuada de la variable en estudio, pues, como hemos visto, ello condiciona el modo de llevar a cabo el contraste. Tal elección debe hacerse tras un examen preliminar de la serie temporal correspondiente, que nos sugiera si hay evidencia de constante (si la media muestral de y_t es distinta de cero), posible tendencia determinista (si la media muestral de Δy_t es distinta de cero), etc.

14.3. CONTRASTACION EN MODELOS AUTORREGRESIVOS DE ORDEN SUPERIOR

El análisis que hasta ahora hemos llevado a cabo se refiere al caso en que el analista considera que la variable endógena obedece a un modelo autorregresivo del primer orden. Supongamos, alternativamente, que el modelo adecuado es de segundo orden:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t, \quad t = 1, 2, \dots, T$$

y supongamos que su polinomio característico acepta dos raíces reales, de modo que puede descomponerse: $1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$, donde $\lambda_1 + \lambda_2 = \phi_1$ y $\lambda_1 \lambda_2 = -\phi_2$. Para que el proceso sea estacionario, es preciso que las dos raíces, λ_1 y λ_2 , sean inferiores a la unidad en valor absoluto. Vamos a examinar el modo de contrastar una posible situación de no estacionariedad, en que una de las raíces, λ_1 , es igual a 1, siendo el valor absoluto de λ_2 inferior a la unidad. En función de los coeficientes estimados, debemos contrastar $H_0: \phi_1 + \phi_2 = 1$, condicionado a $|\phi_2| < 1$, puesto que $\lambda_1 = 1$ implica que $\lambda_2 = -\phi_2$, es decir: $1 - \phi_2 = \phi_1$. Para ello, lo más sencillo es escribir el modelo en la forma:

$$y_t = (\phi_1 + \phi_2)y_{t-1} - \phi_2(y_{t-1} - y_{t-2}) + u_t, \quad t = 1, 2, \dots, T$$

o, equivalentemente:

$$\Delta y_t = (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2 \Delta y_{t-1} + u_t, \quad t = 1, 2, \dots, T$$

en el que el contraste de significación del coeficiente de y_{t-1} equivale al contraste de existencia de una raíz unitaria (pero no más) en el modelo AR(2) original.

También podríamos añadir a este modelo una constante, así como una tendencia determinista. Dependiendo de cómo se formulara finalmente el modelo, el estadístico t de y_{t-1} se compararía con los valores críticos de una u otra de las filas en la Tabla 14.1 (o Tablas A.11 y A.12).

Más generalmente, sean λ_j , $j = 1, \dots, p$ los valores propios del polinomio característico de y_t del modelo:

$$y_t = \alpha_0 + \beta t + \sum_{i=1}^p \rho_i y_{t-i} + \varepsilon_t \quad [14.9]$$

es decir, las raíces de la ecuación $1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^p$, y supongamos que se quiere contrastar la existencia de exactamente una raíz unitaria: $H_0: \lambda_1 = 1, |\lambda_i| < 1, i = 2, \dots, p$. Para ello, se lleva a cabo un contraste del tipo DFA, estimando por MCO el modelo:

$$\Delta y_t = \alpha_0 + \beta t + \gamma_1 y_{t-1} + \sum_{i=1}^{p-1} \gamma_{2i} \Delta y_{t-i} + \varepsilon_t \quad [14.10]$$

que es una reparametrización del anterior, donde el orden p se ha escogido de modo que ε_t sea ruido blanco.

14.4. CONTRASTACION EN MODELOS CON ESTRUCTURA MA

Puesto que todo modelo ARIMA($p, 1, q$) puede aproximarse por una estructura AR finita, Said y Dickey (1984) propusieron utilizar un contraste del tipo DFA escogiendo s aproximadamente igual a la raíz cúbica del tamaño muestral y aplicando los contrastes anteriores a un modelo AR(s).

Si se conocieran los órdenes p y q de los polinomios autorregresivo y de medias móviles $\phi(L)$ y $\theta(L)$ del proceso y se pretendiera contrastar la hipótesis ARIMA($p, 1, q$): $H_0: \phi(L)\Delta y_t = \theta(L)\varepsilon_t$ frente a la hipótesis ARIMA($p, 0, q$): $H_1: \phi(L)(1 - \rho L)y_t = \theta(L)\varepsilon_t$, donde se supone que el polinomio $\phi(L)$ no tiene raíces unitarias, puede estimarse el modelo mediante una iteración del algoritmo Gauss-Newton a partir de estimaciones iniciales consistentes. El estadístico t del parámetro ρ resultante tiene la misma distribución que el estadístico τ de la Tabla A.11 [Tabla 8.5.2 en Fuller (1976), pág. 373]. Si se utilizase la variable y_t transformada en diferencias respecto a su media muestral, la distribución del estadístico t sería la misma que la de τ_μ en la misma tabla.

Molinas (1986) y Schwert (1987), (1989) han mostrado que la potencia de los tests de Dickey-Fuller pueden verse afectados en muestras finitas, incluso

si son largas, por la estructura del verdadero modelo generador de los datos. Por ejemplo, en un modelo ARIMA(0, 1, 1) con parámetro MA próximo a 1, los estadísticos DFA que hemos propuesto en las secciones previas tienen valores críticos muy por debajo de los valores tabulados por Dickey y Fuller, lo que, de no ser tenido en cuenta, haría que ambos contrastes sugirieran la estacionariedad de las series con excesiva frecuencia. La idea es que, como es sabido, un término MA(1) con coeficiente elevado aproxima una diferencia (es decir, una raíz unitaria) de la variable, por lo que la compensación aproximada de ambas hace que tal raíz unitaria pudiera no detectarse. En cambio, el estadístico de Said-Dickey que antes propusimos parece funcionar bien incluso en tales situaciones, con independencia del valor del parámetro de medias móviles.

14.5. CONTRASTE DE k RAICES UNITARIAS

Algunas variables económicas pueden contener más de una raíz unitaria, como sugiere el análisis del tipo Box-Jenkins de las mismas. Por ello, es interesante disponer de estrategias de contraste de varias raíces unitarias. Aunque los contrastes que hemos presentado en la sección anterior pueden utilizarse para las sucesivas diferencias de una variable, Dickey y Pantula (1987) sugieren una secuencia de contrastes, comenzando del máximo número posible (k) de raíces unitarias, descendiendo el valor de k cada vez que se rechaza H_0 , y parando en cuanto que no se rechace. Ello tiene en cuenta que contrastamos cada vez la presencia de una raíz unitaria y que, para aplicar los estadísticos descritos en las secciones anteriores, es preciso que el modelo especificado bajo la hipótesis alternativa sea estacionario.

Por consiguiente, si $k = 3$, comenzaríamos estimando el modelo $\Delta^3 y_t = \xi_0 + \xi_1 \Delta^2 y_{t-1} + \varepsilon_t$ para obtener el estadístico t de ξ_1 . Compararemos dicho estadístico con los valores críticos en Fuller (1976) y rechazamos H_3 : la variable tiene *exactamente* tres raíces unitarias, si $t \leq \tau_\mu$. A continuación estimaríamos el modelo $\Delta^3 y_t = \xi'_0 + \xi'_1 \Delta^2 y_{t-1} + \xi'_2 \Delta y_{t-1} + \varepsilon_t$ para obtener los estadísticos t de ξ'_1 y ξ'_2 . Rechazaremos H_2 : *exactamente* dos raíces unitarias, si además de $\hat{t}(\xi'_1) \leq \tau_\mu$ se tiene $\hat{t}(\xi'_2) \leq \tau_\mu$. Finalmente, estimaríamos la regresión $\Delta^3 y_t = \xi''_0 + \xi''_1 \Delta^2 y_{t-1} + \xi''_2 \Delta y_{t-1} + \xi''_3 y_{t-1} + \varepsilon_t$, rechazando H_1 : *exactamente* una raíz unitaria, si $t(\xi''_i) \leq \tau_\mu$ para $i = 1, 2, 3$.

14.6. INTEGRACION Y ESTACIONALIDAD

Un proceso y_t es *integrable de órdenes* (d, D), lo que se representa por $y_t \sim I(d, D)$, si $\Delta^d \Delta_s^D y_t$ admite una representación ARMA estacionaria, invertible, donde Δ_s denota la diferencia de orden estacional. En nuestra descripción suponemos que se trabaja con datos trimestrales, por lo que $s = 4$ y $\Delta_4 = 1 - L^4$.

14.6.a. Raíz unitaria estacional

Para contrastar $H_0: I(0, 1)$ frente a $I(0, 0)$, es decir: y_t tiene una raíz unitaria estacional pero ninguna regular, Dickey, Hasza y Fuller (1984) (DHF) han mostrado que la distribución límite del estadístico no varía al eliminar las medias estacionales de modelos autorregresivos, por lo que se sugiere comenzar con una regresión de y_t sobre variables ficticias estacionales, junto con una constante y una tendencia, con el objeto de eliminar la estacionalidad determinista, y utilizar los residuos \hat{u}_t de dicha regresión en:

$$\Delta_4 \hat{u}_t = \theta_0 + \sum_1^p \theta_i \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

definiendo después la variable filtrada:

$$z_t = \hat{\theta}(L)\hat{u}_t = (1 - \hat{\theta}_1 L - \hat{\theta}_2 L^2 - \dots - \hat{\theta}_p L^p)\hat{u}_t$$

y estimando, finalmente:

$$\Delta_4 \hat{u}_t = \xi_0 z_{t-4} + \sum_1^p \xi_i \Delta_4 \hat{u}_{t-1} + \varepsilon_t$$

El estadístico t del parámetro ξ_0 , denotado por $\tau_{\mu 4}$, se compara con los valores críticos:

TABLA 14.3. Valores críticos para el contraste DHF

Nivel de significación	$T = 40$	$T = 60$	$T = 80$	$T = 200$
0,01	-5,01	-4,85	-4,78	-4,67
0,05	-4,21	-4,14	-4,11	-4,06

La limitación de este tipo de contraste es que no considera todas las raíces unitarias que puede haber en un proceso estacional. En efecto, con datos trimestrales, el operador Δ_4 puede descomponerse: $(1 - L^4) = (1 - L)(1 + L)(1 - iL)(1 + iL)$, por lo que el proceso estacional puede tener cuatro raíces diferentes, todas ellas de módulo unidad: 1, -1 , i , $-i$. La *segunda* raíz, al actuar sobre y_t genera la ecuación en diferencias: $y_t = -y_{t-1}$, es decir: $y_t = y_{t-2}$, por lo que corresponde a un ciclo que se completa en dos períodos. por lo que, con datos trimestrales, generará dos ciclos por año. La *primera* raíz corresponde a la ecuación $y_t = y_{t-1}$, un ciclo de un período. Por último, es sencillo comprobar que las dos raíces complejas generan la condición: $y_t = y_{t-4}$ y por ello un ciclo de cuatro períodos, es decir, un ciclo por año. Estas dos raíces se consideran conjuntamente en el contraste de raíces unitarias estacionales.

Engle, Granger, Hylleberg y Yoo (1987) propusieron el siguiente procedimiento:

- a) Obtener el estimador $\hat{\theta}(L)$ como en el contraste DHF.
- b) Calcular las series filtradas:

$$\begin{aligned} z_{1t} &= \hat{\theta}(L)(1 + L + L^2 + L^3)\hat{u}_t \\ z_{2t} &= -\hat{\theta}(L)(1 - L + L^2 - L^3)\hat{u}_t \\ z_{3t} &= -\hat{\theta}(L)(1 - L^2)\hat{u}_t \end{aligned}$$

- c) Estimar:

$$\Delta_4 \hat{u}_t = \pi_1 z_{1t-1} + \pi_2 z_{2t-1} + \pi_3 z_{3t-2} + \sum_{i=1}^p \pi_{4i} \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

para analizar la significatividad estadística de los coeficientes π_1, π_2, π_3 , mediante sus estadísticos t . Al interpretar los resultados, debe observarse que el polinomio que define z_{1t} es el resultado de extraer de Δ_4 el término $1 - L$; por consiguiente, incluir z_{1t} como regresor es similar a incorporar y_t como variable explicativa en [14.5], [14.6] y [14.7] al contrastar la existencia de la raíz $1 - L$. Por otra parte, el polinomio que define z_{2t} , es el resultado de extraer $1 + L$, mientras que el que define z_{3t} , resulta de extraer las dos raíces complejas, es decir, $1 + L^2$.

Así, si y_t es $I(0, 0)$, los tres estadísticos deben ser conjuntamente significativos; si únicamente el estadístico t de π_1 es no significativo, entonces y_t es $I(1, 0)$, mientras que si alguno de los estadísticos t de π_2 o π_3 es no significativo, entonces y_t es $I(0, 1)$, pues sólo una raíz unitaria estacional podría explicar tal ausencia de significación. Los valores críticos aparecen en la referencia citada.

14.6.b. Raíz unitaria regular, junto con raíz unitaria estacional

Hasza y Fuller (1982) (HF) propusieron contrastar la hipótesis nula $H_0: y_t \sim I(1, 1)$, estimando un modelo autorregresivo para la variable doblemente diferenciada:

$$\Delta \Delta_4 \hat{u}_t = \psi_0 + \sum_{i=1}^p \psi_i \Delta \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

para, posteriormente, obtener las variables filtradas:

$$z_{4t} = \hat{\psi}(L)\Delta_4 \hat{u}_t = (1 - \hat{\psi}_1 L - \dots - \hat{\psi}_p L^p)\Delta_4 \hat{u}_t \quad y \quad z_{5t} = \hat{\psi}(L)\Delta u_t$$

y, finalmente, estimar el modelo:

$$\Delta \Delta_4 \hat{u}_t = \phi_1 z_{4t-1} + \phi_2 z_{5t-4} + \sum_{i=1}^p \phi_{3i} \Delta \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

Existe evidencia a favor de la hipótesis nula $H_0: y_t \sim I(1, 1)$ frente a la alternativa compuesta $H_1: y_t \sim I(0, 0), y_t \sim I(0, 1), y_t \sim I(1, 0)$ cuando se

TABLA 14.4. Valores críticos del contraste de Hasza y Fuller (1982)

$T = 100$			
<i>Nivel de significación</i>	$H_0: \phi_1 = 0$	$H_0: \phi_2 = 0$	$H_0: \phi_1 = \phi_2 = 0$
Constante y variables ficticias			
0,01	-2,61	-2,79	5,98
0,05	-1,95	-2,05	3,81
Sin constantes ni variables ficticias			
0,01	-2,48	-2,59	4,99
0,05	-1,87	-1,93	3,20

tiene: $\phi_1 = \phi_2 = 0$, lo que puede contrastarse mediante el estadístico F correspondiente. Los estadísticos t de ϕ_1 y ϕ_2 pueden utilizarse por separado para contrastar la misma H_0 frente a $H_1: y_t \sim I(0, 1)$, si: $\phi_1 = 0, \phi_2 < 0$, o a $H_1: y_t \sim I(1, 0)$, si: $\phi_1 < 0, \phi_2 = 0$, respectivamente.

Los valores críticos para estos contrastes son los que se muestran en la Tabla 14.4.

Alternativamente, si denotamos por Z_1, Z_2 y Z_3 los filtros

$$Z_1 = (1 + L + L^2 + L^3), Z_2 = -(1 - L + L^2 - L^3), Z_3 = -(1 - L^2)$$

de modo que, por ejemplo: $Z_3 y_t = -y_t + y_{t-2}$, el contraste de Hylleberg, Engle, Granger y Yoo (1990) (HEGY) de raíz unitaria estacional se basa en la regresión:

$$\Delta \Delta_4 y_t = \alpha_0 + \alpha_1 (Q_{1t} - Q_{4t}) + \alpha_2 (Q_{2t} - Q_{4t}) + \alpha_3 (Q_{3t} - Q_{4t}) + \pi_1 Z_1 y_{t-1} + \pi_2 Z_2 y_{t-1} + \pi_3 Z_3 y_{t-1} + \pi_4 Z_3 y_{t-2} + \sum_{i=1}^p \phi_i \Delta \Delta_4 y_{t-i} + \varepsilon_t$$

aceptándose la existencia de las raíces unitarias $1, -1, i$ y $-i$ si los coeficientes π_i son no significativos para $i = 1, 2, 3$ y 4 , respectivamente. La hipótesis nula es $y_t \sim I(1, 1)$ y se consideran como alternativas $y_t \sim I(2, 0)$ e $y_t \sim I(1, 0)$. Por ejemplo, la significación conjunta de π_2, π_3 y π_4 sugiere que $y_t \sim I(2, 0)$, mientras que si, además, π_1 también es significativa, entonces $y_t \sim I(1, 0)$. También en este contraste puede sustituirse y_t por \hat{u}_t .

Los valores críticos para estos contrastes (Hylleberg, Engle, Granger y Yoo, 1990) son los que se muestran en la Tabla 14.5.

14.7. ESTACIONARIEDAD Y COINTEGRACION

En las secciones anteriores hemos presentado diversos procedimientos para contrastar la hipótesis de existencia de raíces unitarias en una variable

TABLA 14.5. Valores críticos de Hylleberg, Engle, Granger y Yoo (1990) para 100 observaciones

	$\pi_1 = 0$	$\pi_2 = 0$	$\pi_3 = 0$	$\pi_4 = 0$	$\pi_3 = \pi_4 = 0$
Con constante					
0,01	-3,47	-2,61	-2,61	-2,38	4,77
0,05	-2,88	-1,95	-1,90	-1,68	3,08
Con constante y variables ficticias					
0,01	-3,55	-3,60	-2,56	-2,78	8,74
0,05	-2,95	-2,94	-1,89	-1,96	6,57
Con tendencia, constante y variables ficticias					
0,01	-4,09	-3,60	-4,12	-2,76	8,79
0,05	-3,53	-2,94	-3,48	-1,94	6,60

económica. Supongamos dos variables económicas como Consumo y Renta Disponible, entre las que creemos que existe una relación de equilibrio estable a largo plazo. Cabe esperar que, bajo tal supuesto, los residuos de la regresión que explica el gasto en Consumo de las familias mediante su Renta Disponible sean estacionarios, a pesar de que ninguna de las dos variables del modelo lo sean. Ello sería debido a que su evolución temporal es, en gran medida, común. Esta es la idea de *cointegración*.

Supongamos que las dos variables x_t e y_t que componen el vector $\mathbf{z}_t = (x_t, y_t)$ son $I(d)$, pero que existe una constante a tal que la diferencia $x_t - ay_t$ es $I(d - b)$ para un $b > 0$. En tal caso, se dice que x_t e y_t están cointegradas de orden (d, b) ; por ejemplo, si x_t e y_t son $I(1)$, pero $x_t - ay_t$ es estacionaria para una determinada constante a , entonces x_t e y_t son *cointegradas* de orden $(1, 1)$, lo que denotamos $CI(1, 1)$. El vector $\alpha = (1, -a)$ se denomina *vector de cointegración*, y el producto $\alpha' \mathbf{z}_t$ es estacionario. Es claro que el vector de cointegración está definido de forma única excepto por un factor, puesto que $\alpha^* = (\lambda, -a\lambda)$ es asimismo un vector de cointegración entre x_t e y_t . Más generalmente, \mathbf{z}_t puede ser un vector de dimensión k : $\mathbf{z}_t = (x_{1t}, x_{2t}, \dots, x_{kt})$, en cuyo caso puede existir más de un vector de cointegración; el número máximo de tales vectores es $k - 1$.

En definitiva, si las componentes del vector \mathbf{z}_t están ligadas entre sí por una condición lineal de equilibrio estable a largo plazo, existiría un vector α tal que $\alpha' \mathbf{z}_t = 0$. Período a período, pueden existir desviaciones respecto del equilibrio a largo plazo, por lo que se tendrá $\alpha' \mathbf{z}_t \neq 0$, pero a través del tiempo, el comportamiento de las variables en \mathbf{z}_t será tal que la relación $\alpha' \mathbf{z}_t$ tienda nuevamente hacia cero. Por ejemplo, si la propensión marginal a consumir en una economía es de 0,85 y si la Renta Disponible Y_t^d es el único determinante del gasto en Consumo C_t , se tendría que $C_t = 0,85 Y_t^d$ para

todo t , siendo la diferencia entre ambos miembros estacionaria. En tales condiciones, ambas variables, que quizá fuesen $I(1)$, estarían cointegradas.

La importancia estadística del concepto de cointegración estriba en que cuando las variables no estacionarias que aparecen a ambos lados de un modelo de regresión están cointegradas, entonces la estimación de MCO continúa teniendo buenas propiedades. Si, por ejemplo, las variables x_t e y_t son $CI(1, 1)$, entonces al estimar por MCO el modelo

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T$$

la estimación de β coincidirá con la constante de cointegración y los residuos serán estacionarios. El estimador MCO es consistente a pesar de ser las dos variables no estacionarias; más aún, el estimador MCO resulta ser *superconsistente*, puesto que converge al verdadero valor de los coeficientes α y β a una tasa $1/T$, en vez de a la tasa habitual⁽²⁾, $1/\sqrt{T}$.

La superconsistencia del estimador MCO en la regresión estática se debe a que, para valores del coeficiente β distintos de la constante de cointegración a , el residuo no es estacionario, y su varianza tiende a infinito mientras que, para $\beta = a$, su varianza es finita. Como el estimador MCO «busca» en el espacio paramétrico tratando de minimizar la varianza residual, no es sorprendente que escoja el valor a para el coeficiente β con más facilidad que si las variables no estuviesen cointegradas.

El concepto de cointegración responde a la idea de que un modelo que pretende explicar una variable y_t , que sea $I(1)$ utilizando variables explicativas x_t , todas $I(0)$, es decir, estacionarias, no puede dar un buen resultado estadístico, como tampoco podría darlo un modelo que pretendiera explicar una variable $I(2)$ mediante un conjunto de variables explicativas que fuesen $I(1)$ o $I(0)$. Por el contrario, puede aspirar a obtener un buen modelo para explicar una variable y_t que fuese $I(1)$ utilizando como explicativas x_t y z_t , ambas $I(2)$, siempre y cuando éstas fuesen $CI(2, 1)$ y su combinación lineal de cointegración $x_t - az_t$ estuviese, a su vez, cointegrada con y_t .

14.7.a. Cointegración y el modelo de corrección de error (MCE)

Supongamos que dos variables x_t e y_t son $C(1, 1)$. Un *modelo general de corrección de error* entre ambas es del tipo:

$$\begin{aligned} \Delta y_t &= -\gamma_1 w_{t-1} + A(L)\Delta y_{t-1} + B(L)\Delta x_{t-1} + u_{1t} \\ \Delta x_t &= -\gamma_2 w_{t-1} + C(L)\Delta y_{t-1} + D(L)\Delta x_{t-1} + u_{2t} \end{aligned} \quad [14.11]$$

⁽²⁾ Ello quiere decir que la diferencia $\hat{\beta} - \beta$ converge asintóticamente a una Normal, incluso cuando se le multiplica por T , lo que no aceptan otros resultados sobre distribuciones límite. Por ejemplo, en el contexto habitual de regresión con variables estacionarias se tiene que $\sqrt{T}(\hat{\beta} - \beta)$ converge a una Normal, por lo que $T(\hat{\beta} - \beta)$ no tiene una distribución asintótica bien definida.

donde $w_t = y_t - ax_t$ representa la combinación lineal de cointegración entre x_t e y_t y donde u_{it} , $i = 1, 2$ son términos de error estacionarios, aunque pueden tener autocorrelación.

En este modelo puede observarse que, por estar cointegradas y_t y x_t , entonces tanto Δy_t como Δx_t e $y_{t-1} - ax_{t-1}$ son $I(0)$, y como también lo son las diferencias que aparecen en [14.11], se tiene que las propiedades asintóticas habituales del estimador MCO son válidas para los coeficientes, si bien no los son para la constante de cointegración a . En particular, si u_t está libre de autocorrelación, entonces el estimador MCO es consistente. También es Normal en muestras finitas si u_t también lo es.

En general, se dice que un vector de variables z_t admite una *representación de corrección de error (MCE)* si puede expresarse como:

$$A(L)\Delta z_t = -\gamma'w_{t-1} + u_t, \quad \gamma \neq 0$$

siendo u_t un *vector de perturbaciones estacionario*, $A(0) = I_k$, teniendo la ecuación característica $|A(L)| = 0$ todas sus raíces fuera del círculo unidad. En esta representación, w_t es un vector de relaciones de cointegración entre las variables de z_t : $w_t = \alpha'z_t$. En el modelo [14.11], $z_t = (x_t, y_t)$.

Enlazando con nuestros comentarios al comienzo del capítulo, w_{t-1} representa el margen por el que las condiciones de equilibrio estable entre las variables que configuran el vector z_t dejaron de cumplirse en el período $t - 1$ y, aparte de su propio pasado, estas desviaciones son las únicas variables explicativas de la variación que experimenta el vector z_t en el instante t . El ajuste hacia el equilibrio será más o menos gradual, en función del número de retardos que aparezcan en [14.11], así como de sus coeficientes.

Si existe entre y_t y x_t una relación del tipo [14.11], entonces ambas variables están cointegradas con constante a , ya que todas las variables que aparecen en [14.11] aparte de dicha diferencia son estacionarias. El teorema siguiente prueba el recíproco: si su conjunto de variables está cointegrado, existe una representación suya en la forma de modelo de corrección de error.

Teorema de representación de Granger. Si existen r relaciones de cointegración entre las variables que componen el vector z_t con $d = 1$, $b = 1$, entonces:

a) Existe una *representación MCE* en función del vector $w_t = \alpha'z_t$, de dimensión $r \times 1$:

$$A^*(L)\Delta z_t = -\gamma'w_{t-1} + d(L)\varepsilon_t, \quad A^*(0) = I_k \quad [14.12]$$

donde las filas de α , de dimensión $r \times k$, contienen los coeficientes de cada relación de cointegración.

b) Además, existe una representación ARMA vectorial:

$$A(L)z_t = d(L)\varepsilon_t \quad [14.13]$$

siendo $d(L)$ el polinomio (escalar) de [14.12], rango $A(1) = r$ y $A(0) = I_k$. Si el polinomio $d(L) = 1$, entonces el vector z_t acepta una representación vectorial autorregresiva: $A(L)z_t = \varepsilon_t$, pero la matriz $A(1)$ es singular, pues es de rango igual al número de vectores de cointegración.

Un sistema bivalente cointegrado admite la representación:

$$\begin{aligned}\Delta x_t &= b(x_{t-1} - ay_{t-1}) + \text{retardos}(\Delta x_t, \Delta y_t) + u_t \\ \Delta y_t &= c(x_{t-1} - ay_{t-1}) + \text{retardos}(\Delta x_t, \Delta y_t) + v_t\end{aligned}$$

donde al menos uno de los coeficientes b o c es no nulo, y u_t es estacionario, aunque puede estar autocorrelacionado. En consecuencia, tal sistema debe tener una ordenación causal al menos en una dirección, puesto que w_t es una función de ambas variables. Sin embargo, los retardos de Δx_t e Δy_t pueden aparecer o no en las ecuaciones anteriores, por lo que puede tenerse todo tipo de estructuras de causalidad. En particular, si el término de corrección de error aparece en ambas ecuaciones, entonces ninguna variable es débilmente exógena con respecto a los parámetros de la otra ecuación, debido a las restricciones entre los coeficientes de las distintas ecuaciones.

Si dos variables están cointegradas $CI(1, 1)$, entonces existe un modelo del tipo anterior, pero no sabemos la dirección de causalidad. Esto es importante, porque se tendrán distintas estimaciones de los coeficientes del vector de cointegración según qué dirección de la relación se estime. Por ejemplo, si x_t e y_t son $CI(1, 1)$, entonces el producto de las estimaciones de a y $1/a$ en las regresiones de y_t sobre x_t y de ésta sobre la primera no es igual a 1 sino a R^2 .

14.7.b. Estimación de modelos de corrección de error

Engle y Granger (1987) sugieren una estimación en dos etapas: inicialmente, se estima el vector de cointegración a partir de una regresión entre los valores contemporáneos de las variables que componen el vector z_t . El residuo resultante es estacionario, de modo que el estimador MCO de dicho modelo es consistente. Condicionando en tal estimación del vector de cointegración α , se construye el vector de residuos w_t y se incorpora al modelo MCE, que pasa a estimarse por MCO, una vez que se han incorporado al mismo suficientes retardos de Δx_t y Δy_t como para que el término de error sea estacionario⁽³⁾.

La desviación típica del vector de cointegración en la primera etapa es inconsistente, porque la distribución del estimador MCO *no es Normal*. Ello implica que hay que escoger las variables a incluir en la regresión de largo

⁽³⁾ Debe prestarse atención al hecho de que otros procedimientos alternativos de estimar α no producen estimaciones consistentes: por ejemplo, una regresión de Δy_t sobre Δx_t no genera estimaciones consistentes. Un resultado similar se obtiene si, para evitar la casi inevitable autocorrelación en la regresión de cointegración, se utilizan transformaciones del tipo Cochrane-Orcutt.

plazo a priori, puesto que no se puede decidir su inclusión sobre la base de los contrastes habituales de significación. Por supuesto, uno de los criterios fundamentales para la inclusión de variables en dicha regresión es que estén cointegradas. Banerjee y colaboradores (1986) han probado que el sesgo en muestras finitas en el estimador MCO de los coeficientes de largo plazo, es decir, el vector de cointegración, puede ser apreciable y depende: a) del tamaño relativo de las varianzas de las ecuaciones de largo y corto plazo, y b) está inversamente relacionado con el R^2 de la regresión estática de cointegración.

Alternativamente, consideremos un modelo específico de corrección de error:

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{t-1} + \beta_3 \Delta y_{t-1} + \beta_4 (y_{t-1} - a x_{t-1}) + u_t \quad [14.14]$$

y veamos cómo puede estimarse simultáneamente el coeficiente de cointegración β y el vector $\beta = (\beta_1, \beta_2, \beta_4)$ del MCE. A primera vista, parece que éste debiera ser un problema de estimación no lineal sujeto a restricciones, pero si hacemos el cambio de parámetros:

$$\gamma_1 = \beta_1 \quad ; \quad \gamma_2 = \beta_2 \quad ; \quad \gamma_3 = -a\beta_4 \quad ; \quad \gamma_4 = \beta_4$$

basta estimar por MCO una regresión sin restricciones de Δy_t sobre: una constante, Δx_t , Δx_{t-1} e Δy_{t-1} , para obtener estimadores de los parámetros γ_i , $i = 1, 2, 3, 4$. La respuesta a largo plazo de y_t a x_t queda estimada mediante $a = -\hat{\gamma}_3/\hat{\gamma}_4$, aunque no tendremos una desviación típica para este parámetro. La estimación por MCNL nos habría proporcionado la desviación típica de a . Lo que ocurre es que, en presencia de no estacionariedad, la distribución de este estimador no es la habitual, por lo que no hubiéramos podido utilizar dicha desviación típica estimada.

14.7.c. Contrastes de cointegración

Un contraste de cointegración puede entenderse como equivalente a un contraste de raíz unitaria en los residuos de la ecuación de cointegración, ya que si las variables y_t y x_t , siendo ambas $I(1)$, estuviesen cointegradas, entonces los residuos del modelo $y_t = \alpha + x_t' \beta + u_t$ deberían ser estacionarios, no teniendo por tanto ninguna raíz unitaria.

Veamos dos contrastes propuestos por Engle y Granger (1987):

1. El estadístico DW de la ecuación de cointegración (Sargan y Barghava, 1983) conduce a rechazar la hipótesis nula de no cointegración, es decir, la existencia de raíz unitaria en los residuos, si el estadístico DW es significativamente mayor que cero. Este contraste, que parece bastante potente, se basa en que $DW = 2(1 - \rho)$, y los valores críticos para un tamaño muestral $T = 100$ son los que se muestran en la Tabla 14.6.

TABLA 14.6. Contraste de cointegración. Estadístico Durbin-Watson de la regresión de largo plazo o de cointegración

Nivel de significación	0,01	0,05	0,10
Valor crítico	0,511	0,386	0,322

Una dificultad con este contraste es que dichos valores críticos son función de la especificación adoptada para el modelo verdadero. El contraste se realiza mediante la existencia de cotas superior e inferior —como en un contraste DW de autocorrelación—, pero dichas cotas se separan entre sí mucho en cuanto aumenta el número de variables explicativas.

Una gran ventaja de este contraste es su invariancia frente a la posible inclusión de constantes y tendencias en el modelo, sin que por ello varíen sus valores críticos.

2. El segundo contraste, del tipo DF o DFA sobre la regresión de cointegración, consiste en estimar por MCO: $\Delta \hat{u}_t = \gamma_1 \hat{u}_{t-1} + \sum_{i=1}^p \gamma_{2i} \Delta \hat{u}_{t-i} + \varepsilon_t$, donde \hat{u}_t son los residuos MCO de la ecuación de cointegración, y el número de retardos p se escoge suficientemente grande como para que ε_t sea ruido blanco.

Los valores críticos con los que comparar el estadístico t del coeficiente γ_1 aparecen en Engle y Granger, y son diferentes de los de Fuller (1976). La razón para esta disparidad es que \hat{u}_t es una variable «generada» mediante un procedimiento que ha escogido $\hat{\alpha}$ y $\hat{\beta}$ de modo que se minimice la varianza residual, es decir, la varianza de \hat{u}_t y, por tanto, de modo que parezca tan estacionario como sea posible. En consecuencia, si utilizamos los valores críticos de los contrastes DF o DFA para este caso, rechazaríamos la hipótesis nula de no estacionariedad con demasiada frecuencia. Para evitar este sesgo, los valores críticos deben aumentarse ligeramente.

Otro factor que influye sobre los valores críticos apropiados es el número de variables explicativas, pues una combinación lineal de una larga lista de variables explicativas no estacionarias puede parecer estacionaria, incluso si las variables no están cointegradas. Engle y Granger (1987) presentaron los valores críticos correctos para una sola variable explicativa, lo que fue generalizado en Engle y Yoo (1987) hasta cinco variables explicativas.

TABLA 14.7. Contrastes de cointegración Dickey-Fuller y Dickey-Fuller ampliado. Valores críticos. $T = 100$ observaciones

	0,01			0,05			0,10		
	$N = 2$	$N = 3$	$N = 4$	$N = 2$	$N = 3$	$N = 4$	$N = 2$	$N = 3$	$N = 4$
DF	-4,07	-4,45	-4,75	-3,37	-3,93	-4,22	-3,03	-3,59	-3,89
DFA	-3,77	-4,22	-4,61	-3,17	-3,62	-4,02	-2,84	-3,32	-3,71

Ambos contrastes están diseñados para el caso en que tanto la hipótesis nula como la alternativa consideran modelos de primer orden, y presentan problemas para valores próximos a 1 del coeficiente ρ del modelo AR(1) de los residuos, es decir, para valores muy bajos del estadístico DW. A tales niveles, la potencia, tanto del contraste basado en el estadístico DW como de tipo DFA, es muy baja.

3. Otros contrastes se basan en la propiedad de que las variables cointegradas aceptan una representación en la forma de MCE y, por ello, como ya hemos visto, al menos una variable, digamos x_t , *causa* a la otra y_t , en el sentido de que sus retardos x_{t-s} , $s > 1$, contienen información relevante acerca de y_t que no está contenida en el pasado de esta última variable. Es decir, los retardos x_{t-s} deben aparecer en la regresión de y_t , incluso si se incluyen retardos y_{t-s} . Por consiguiente, podemos especificar:

$$\begin{aligned} \Delta x_t &= a_1 + b_1 \hat{u}_{t-1} + \varepsilon_{1t} \\ \Delta y_t &= a_2 + b_2 \hat{u}_{t-1} + c_2 \Delta x_t + \varepsilon_{2t} \end{aligned}$$

donde \hat{u}_t son los residuos mínimo-cuadráticos de la regresión de cointegración de y_t sobre x_t .

El modelo de corrección de error implica que b_1 y b_2 deben ser conjuntamente significativos. El estadístico para el contraste de tal hipótesis es la suma de cuadrados de los respectivos estadísticos «tipo *t*», $\tau_{b_1}^2 + \tau_{b_2}^2$. Si se rechaza la hipótesis nula $H_0: b_1 = b_2 = 0$, entonces x_t e y_t satisfacen un mecanismo de corrección de error y están cointegrados. Este es el contraste conocido como *VAR restringido* (RVAR). Un *test ampliado* (ARVAR) de esta hipótesis consiste en incluir retardos de Δx_t y Δy_t como variables explicativas de ambas ecuaciones.

TABLA 14.8. Contrastes de cointegración. Valores críticos

N = 2			
	0,01	0,05	0,10
RVAR	18,3	13,6	11,0
ARVAR	15,8	11,8	9,7
UVAR	23,4	18,6	16,0
AUVAR	22,6	17,9	15,5

Alternativamente, puede utilizarse un contraste de tipo *VAR sin restricciones* (UVAR), mediante la estimación del sistema:

$$\begin{aligned} \Delta x_t &= a_1 + b_1 y_{t-1} + c_1 x_{t-1} + \varepsilon_{1t} \\ \Delta y_t &= a_2 + b_2 y_{t-1} + c_2 x_{t-1} + d_2 \Delta x_t + \varepsilon_{2t} \end{aligned}$$

y contrastar la hipótesis nula conjunta $H_0: b_1 = c_1 = b_2 = c_2 = 0$ mediante el estadístico $2(F_1 + F_2)$, siendo F_1 el estadístico F para el contraste de la hipótesis conjunta $H_0: b_1 = c_1 = 0$ y F_2 el estadístico F para el contraste de $H_0: b_2 = c_2 = 0$. Si se rechaza dicha hipótesis, Δx_t e Δy_t dependen de sus niveles y, potencialmente, siguen un esquema MCE. Por último, puede construirse un *test ampliado* (AUVAR) incluyendo retardos adicionales de Δx_t e Δy_t como variables explicativas. Para este contraste ampliado se utiliza el mismo estadístico: $2(F_1 + F_2)$ anterior.

14.7.d. Contrastes de cointegración estacional

Los polinomios en el operador de retardos que utilizamos en la Sección 14.6 para la contrastación de la hipótesis de raíz unitaria bajo estacionalidad, pueden utilizarse asimismo para contrastar la cointegración de variables. Así, con datos trimestrales: a) existe *cointegración en el ciclo de un periodo* si existe un vector α_1 tal que la combinación lineal $u_t = \alpha'_1 Z_1 z_t$ es estacionaria, b) existe *cointegración en el ciclo de dos periodos* si existe un vector α_2 tal que la combinación lineal $v_t = \alpha'_2 Z_2 z_t$ es estacionaria, y c) existe *cointegración en el ciclo anual* si existen dos vectores α_3 y α_4 tal que la combinación lineal $w_t = (\alpha'_3 + \alpha'_4 L) Z_3 z_t$ es estacionaria.

Existe una versión estacional del modelo de corrección de error, que permite la incorporación al mismo de las combinaciones lineales de cointegración que puedan existir en las diversas raíces unitarias que ya hemos examinado. Teniendo en cuenta que todas las variables incluidas en el modelo deben ser $I(0)$, la forma general de dicho modelo es:

$$\begin{aligned} \Delta\Delta_4 y_t = & \sum_{i=1}^m \beta_i \Delta\Delta_4 y_{t-i} + \sum_{i=1}^n \delta_i \Delta\Delta_4 x_{t-i} + \\ & + \gamma_1 (Z_1 y_{t-1} - \alpha_{12} Z_1 x_{t-1}) + \gamma_2 (Z_2 y_{t-1} - \alpha_{22} Z_2 x_{t-1}) + \\ & + (\gamma_3 + \gamma_4 L) (Z_3 y_{t-1} - \alpha_{32} Z_3 x_{t-1} - \alpha_{41} Z_3 y_{t-2} - \alpha_{42} Z_3 x_{t-2}) + u_t \end{aligned} \quad [14.15]$$

donde se han incorporado suficientes retardos de $\Delta\Delta_4 y_t$ y $\Delta\Delta_4 x_t$ como para que u_{1t} sea ruido blanco. Permitimos que los vectores de cointegración de cada raíz unitaria (los α_{ij}) sean diferentes, así como también sus efectos sobre la corrección que experimenta en el periodo t la variable dependiente (los γ_i).

La estimación en dos etapas de este modelo de corrección de error estacional consiste en estimar los coeficientes α_{ij} , $i = 1, 2, 3, 4$, $j = 1, 2$ en cada regresión de cointegración, para luego incorporarlos a [14.15] y estimar por MCO. El número de retardos que se habrá incorporado a cada suma ha de ser suficiente como para eliminar la autocorrelación del término de error.

Hay dos hipótesis de interés que podrían imponerse al estimar el MCE estacional [14.15]:

1. Que se produce cointegración en todos los ciclos con igual parámetro: $\alpha_{12} = \alpha_{22} = \alpha_{32} = \alpha$ y $\alpha_{41} = \alpha_{42} = 0$, con lo que el MCE se reduce al expresado en las variables originales:

$$\Delta_4 y_t = \sum_{i=1}^m \beta_i \Delta_4 y_{t-i} + \sum_{i=1}^n \delta_i \Delta_4 x_{t-i} + \sum_{i=1}^3 \gamma_i Z_i (y_{t-1} - \alpha x_{t-1}) + \gamma_4 Z_3 (y_{t-2} - \alpha x_{t-2}) + u_t$$

2. Que existe cointegración en el ciclo de un período en las variables ajustadas estacionalmente, y en los demás ciclos con el mismo parámetro: $\alpha_{12} = \alpha$; $\alpha_{22} = \alpha_{32} = \alpha_s$; $\alpha_{41} = \alpha_{42} = 0$.

14.8. APLICACIONES DEL CONCEPTO DE COINTEGRACION

14.8.a. Una síntesis

Antes de examinar algunas aplicaciones de la teoría que hemos desarrollado en este capítulo, resumamos algunas ideas esenciales en relación a la importancia de la cointegración en la especificación y estimación de modelos de regresión:

1. No pueden mezclarse variables de distintos órdenes de integración en una regresión estática y esperar que los resultados estadísticos estén justificados rigurosamente.

2. Si se pretende generalizar una regresión estática incluyendo retardos para obtener su versión dinámica, es *esencial* incorporar retardos de todas las variables que entran en la regresión estática o, al menos, de todas las variables que parezcan relevantes en dicha regresión. Esto es lo que, de hecho, sugiere un MCE, por lo que no puede resultar sorprendente la estrecha conexión que existe entre MCE y cointegración.

3. Cuando se especifica una regresión dinámica, la propia riqueza de la estructura dinámica puede hacer que los residuos aparenten ser estacionarios en una muestra finita cuando en realidad las variables que aparecen en el modelo no estén cointegradas, en cuyo caso los procesos de inferencia habitual no serían válidos.

4. En la especificación de una relación dinámica entre variables económicas puede adoptarse una estrategia de seleccionar variables cointegradas, especificar su relación de largo plazo o regresión de cointegración, y estimar ésta, así como su modelo de corrección de error. Alternativamente, puede procederse especificando un modelo dinámico muy general en el que ir imponiendo restricciones para llegar a un modelo dinámico sencillo.

Teóricamente, en muestras infinitas, ambos enfoques deberían producir el mismo resultado. Sin embargo, en muestras finitas esto no es así; además de errores muestrales, al estimar un modelo dinámico general, puede que se incluyan variables de diverso orden de integración, con las dificultades que hemos expuesto en secciones anteriores, tanto en términos de cuál sea la distribución correcta para las estimaciones numéricas obtenidas y sus desviaciones típicas, así como para la estimación del efecto a largo plazo de las variables explicativas sobre la variable dependiente.

14.8.b. La eficiencia de un mercado financiero

Hay varias formas en que la cointegración de variables puede utilizarse para contrastar *la eficiencia de los mercados financieros*: la noción de eficiencia de mercado, fundamentada en la idea de que el precio actual resume toda la información disponible acerca del precio futuro, equivale a decir que la proyección del precio futuro p_{t+1} sobre el conjunto de información hoy disponible Ω_t es simplemente igual a p_t , sin que ninguna otra información resulte relevante, es decir:

$$p_{t+1} = p_t + u_{t+1}$$

con $E_t u_{t+1} = 0$, es decir, que u_t es ruido blanco y p_t un paseo aleatorio.

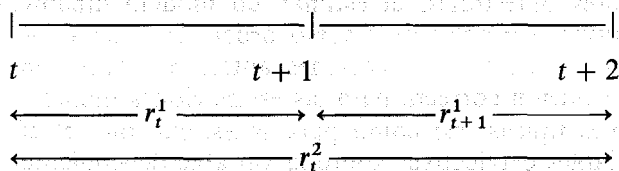
a) Cointegración entre precios de activos

El modelo de corrección de error implica que el cambio en y_t , Δy_t , se debe a un cambio contemporáneo en otra variable x_t , Δx_t , junto con el efecto a largo plazo que proviene del desajuste producido en el período anterior con respecto al equilibrio a largo plazo, $y_{t-1} - ax_{t-1}$. Por tanto, aun dejando aparte Δy_t , que era desconocido, el modelo sugiere que existía información conocida con anterioridad (en los valores de x_{t-1} e y_{t-1}) que podía ayudar a prever el cambio Δx_t .

Pero si $x_t \equiv p_t$ representa el precio de un activo cuyo mercado se supone eficiente, tal situación no es posible, puesto que la variación en su precio, Δp_t , debe ser impredecible a partir del conjunto de información Ω_{t-1} ; ello es debido a que el precio pasado, p_{t-1} , incorpora toda la información relevante acerca del precio futuro, desconocido, p_{t+s} , $s \geq 0$. En definitiva, todo ello nos sirve para afirmar que si x_t e y_t denotan los precios determinados en dos mercados financieros eficientes, por ejemplo, dos tipos de cambio, o dos tipos de interés forward, *no pueden estar cointegrados*.

b) El tipo forward como predictor del tipo spot futuro

Por otra parte, en muchos mercados financieros se intercambia un activo emitido a distintos plazos, por lo que podemos deducir tipos forward implícitos. En tal situación, la idea de eficiencia se traduce en el supuesto de que r_t^1 y r_t^2 , los rendimientos actuales de mercado a 1 y 2 períodos de horizonte de madurez, resumen toda la información acerca de r_{t+1}^1 , el tipo a un período que resultará vigente el próximo período:



Los tipos r_t^1 y r_t^2 determinan el tipo forward implícito F_t^1 definido como $F_t^1 = (2r_t^2 - r_t^1)$, siendo entonces la hipótesis de eficiencia la propiedad de que la proyección del tipo «spot» futuro r_{t+1}^1 sobre Ω_t sea igual a F_t^1 , es decir:

$$r_{t+1}^1 = F_t^1 + u_{t+1} = (2r_t^2 - r_t^1) + u_{t+1}$$

con u_t ruido blanco o, lo que es lo mismo: $E_t r_{t+1}^1 = F_t^1 = 2r_t^2 - r_t^1$.

Por esta razón, un buen número de trabajos se han dedicado a la contrastación de la hipótesis conjunta $H_0: a = 0, b = 1$ en el modelo:

$$r_{t+1}^1 = a + bF_t^1 + u_{t+1} \quad [14.16]$$

lo que se conoce en la literatura como que *el tipo forward es un predictor insesgado del tipo spot futuro*. La eficiencia del mercado precisa de una condición adicional: que el término de error u_t del modelo sea **ruido blanco**, pues de otro modo, la proyección de r_{t+1} sobre Ω_t incluiría, además de F_t^1 , valores retardados de ambas variables.

Desde el punto de vista de la cointegración de variables, afirmar que el tipo forward es un predictor insesgado del tipo spot futuro o que el modelo anterior es válido implica no sólo que ambas variables están cointegradas, sino también que su constante de cointegración es igual a 1 y que el término de error u_t es ruido blanco. Todo ello, conjuntamente, puede interpretarse como un contraste de eficiencia del mercado.

Respecto a la puesta en práctica de este contraste, cabe hacer dos observaciones:

1. La observación de la no estacionariedad de los tipos spot y forward ha llevado a especificar modelos del tipo:

$$r_{t+1}^1 - r_t^1 = a + b(F_t^1 - r_t^1) + u_{t+1} \quad [14.17]$$

en la confianza de que ambas variables transformadas resultasen estacionarias.

2. El contraste puede formularse de otro modo alternativo: si r_{t-1}^1 y F_t^1 están cointegradas, se tiene la representación en modelo de corrección de error:

$$r_{t+1}^1 - r_t^1 = A(F_t^1 - F_{t-1}^1) + B(r_t^1 - cF_{t-1}^1) + \varepsilon_{t+1} \quad [14.18]$$

donde c es la constante de cointegración. Sólo si $A = 1$ y $B = -1$ podrían compatibilizarse [14.17] y [14.18], supuesto que $a = 0$ y $b = 1$. Para ver este punto de otro modo, escribamos [14.18] como:

$$(1 - (1 + B)L)r_{t+1}^1 = -(A + Bc)F_t^1 + AF_{t-1}^1 + \varepsilon_{t+1} \quad [14.19]$$

que conduce a:

$$r_{t+1}^1 = AF_t^1 + B(A - c)\sum_{s=0}^{\infty} (1 + B)^s F_{t-1-s}^1 + \sum_{s=0}^{\infty} (1 + B)^s \varepsilon_{t-s+1} \quad [14.20]$$

supuesto que $B < 0$, pero esta expresión sólo podría ser compatible con [14.16] si $B = -1$ y $A = c = 1$. De este modo, los retardos de F_t^1 en la regresión r_{t+1}^1 sobre F_t^1 tendrían coeficientes igual a cero.

14.8.c. La cointegración del Consumo y el PIB españoles

Consideremos nuevamente los datos de Consumo y PIB españoles para el periodo 1954-1988, en pesetas constantes de 1980. Tras tomar logaritmos en ambas variables, estimamos modelos de regresión que nos permitieran contrastar la existencia de raíces unitarias en ambas variables. La Tabla 14.9 contiene los estadísticos t correspondientes al contraste de la hipótesis $H_0: I(2)$ frente a $I(1)$, mientras que en la Tabla 14.10 se recogen los estadísticos t para el contraste de $H_0: I(1)$ frente a $I(0)$. En cada caso, se presentan las estimaciones de los modelos [14.2] (primera y segunda filas de cada tabla) y [14.3] (tercera y cuarta filas); no se presenta el modelo [14.4] porque la variable de tendencia no resultó significativa en ninguno de los casos en que se incluyó. En ambos contrastes, la constante resultó significativa, por lo que son las dos últimas filas las que deberían utilizarse⁽⁴⁾.

Aunque nuestro tamaño muestral ($T = 35$) no coincide con los considerados en Fuller (1976), la comparación de los estadísticos t de la Tabla 14.9 con los valores críticos de la Tabla 14.1 permite mantener la hipótesis nula de dos raíces unitarias, frente a la alternativa de tan sólo una raíz unitaria. A la vista de estos resultados, el contraste de una raíz unitaria frente a estacionariedad ya es redundante, puesto que no hemos rechazado la existencia de dos raíces unitarias. Sin embargo, presentamos los resultados como ilustración al lector. Así, puede apreciarse en las Tablas 14.9 y 14.10 que, mientras el contraste $I(1)$ frente a $I(0)$ tiene una conclusión contundente, el contraste $I(2)$ frente a $I(1)$ es mucho más marginal, si bien puede mantenerse que ambas variables son $I(2)$.

TABLA 14.9. Contrastes de raíz unitaria $I(2)$ frente a $I(1)$

<i>Variable dependiente: $\Delta^2 y_t$</i>		
<i>Variables independientes</i>	<i>Consumo</i>	<i>PIB</i>
Δy_{t-1}	-1,71	-1,57
$\Delta y_{t-1} + \Delta^2 y_{t-1}$	-1,29	-1,54
Constante + Δy_{t-1}	-3,34	-2,93
Constante + $\Delta y_{t-1} + \Delta^2 y_{t-1}$	-2,46	-2,92

⁽⁴⁾ En el primer contraste, el retardo de la segunda diferencia no resultó significativo, por lo que en la Tabla 14.9 es la tercera fila la que debería tomarse, mientras que en el segundo contraste, la primera diferencia retardada resultó significativa, por lo que es la última fila de la Tabla 14.10 la que debería ser utilizada. Debe observarse, sin embargo, que ambos modelos no son totalmente consistentes entre sí: la presencia de una constante en el contraste de $I(2)$ frente a $I(1)$ debería conducir a una tendencia en el modelo utilizado en el contraste de $I(1)$ frente a $I(0)$.

TABLA 14.10. Contrastes de raíz unitaria $I(1)$ frente a $I(0)$

Variable dependiente: Δy_t		
Variables independientes	Consumo	PIB
y_{t-1}	8,30	7,98
$y_{t-1} + \Delta y_{t-1}$	2,62	2,26
Constante + y_{t-1}	-2,13	-2,56
Constante + $y_{t-1} + \Delta y_{t-1}$	-1,41	-1,54

La regresión de cointegración entre Consumo y PIB condujo a:

$$\ln C_t = 0,0408 + 0,969 \ln Y_t + \hat{u}_t$$

(0,80) (0,01)

$$\bar{R}^2 = 0,997 \quad DW = 0,36$$

y el estadístico DF para los residuos⁽⁵⁾ fue de $-1,86$, por lo que no se rechaza la hipótesis de que dichos residuos no son estacionarios. Sin embargo, al contrastar la presencia de dos raíces unitarias en los residuos frente a la alternativa de una sola raíz, se obtuvo un estadístico DF de $-6,66$, por lo que se rechaza dicha hipótesis nula. Los residuos tienen una sola raíz unitaria, y el Consumo y el PIB son variables $CI(2, 1)$.

Una evidencia similar se obtiene mediante la regresión en diferencias:

$$\Delta \ln C_t = 0,638 \cdot 10^{-2} + 0,818 \Delta \ln Y_t + \hat{u}_t$$

(0,41 10^{-2}) (0,08)

$$\bar{R}^2 = 0,762 \quad DW = 2,39$$

con un estadístico DF para los residuos (incluyendo una constante en dicho contraste) de $-6,79$, por lo que estos últimos residuos son estacionarios.

⁽⁵⁾ Los retardos que se incluyeron al realizar un contraste DFA resultaron no significativos, por lo que nos limitamos al contraste DF.

CAPITULO 15

DATOS DE PANEL

15.1. DESCRIPCION DEL PROBLEMA

En este capítulo discutimos la estimación de modelos econométricos utilizando paneles de datos; éstas son muestras formadas por las observaciones recogidas a N agentes económicos a lo largo de T instantes de tiempo⁽¹⁾. Suponemos a lo largo del capítulo que, como es el caso habitual en paneles de datos microeconómicos, el número N de agentes es mucho mayor que T . Ejemplos de este tipo de paneles en España son: la Central de Balances del Banco de España, en la que se recopila información acerca de los balances contables (anuales) de un gran número de empresas, la Encuesta Permanente de Consumo, y la Encuesta de Presupuestos Familiares, ambas recogidas por el Instituto Nacional de Estadística, en las que se sigue la evolución temporal de los gastos de consumo de un importante número de familias⁽²⁾. Los agentes económicos de los que provienen los datos en un panel microeconómico pueden ser personas, familias u hogares, o empresas, según cada encuesta particular.

Los tamaños relativos de las dos dimensiones del panel de datos influyen sobre la naturaleza de las preguntas que el investigador pretende analizar y, con ello, sobre el tratamiento otorgado al panel de datos. En el Capítulo 8 consideramos la estimación de un modelo econométrico con muestras temporales procedentes de distintas, pero similares, unidades muestrales, como son los países de un mismo entorno económico, o los fabricantes que operan dentro de una misma industria. Tales muestras también constituyen un panel

⁽¹⁾ Este capítulo se basa, en buena parte, en los desarrollos recientes sobre métodos de estimación de modelos con datos de panel en los que M. Arellano está teniendo una participación importante. Sus artículos de 1990 (con O. Bover), 1992 y 1993 han sido especialmente relevantes en la preparación de esta presentación.

⁽²⁾ Aunque en estos últimos casos existe cierta *atrición* o *desgaste* en la muestra, es decir, un cierto porcentaje de los hogares encuestados salen de la muestra cada trimestre, lo que requiere un tratamiento adecuado.

de datos, pero con una dimensión temporal bastante mayor que la dimensión de la sección cruzada (el número de países o fabricantes incluidos en la muestra). En tales casos, el investigador está generalmente preocupado por:

1. Contrastar una determinada hipótesis de comportamiento temporal, así como analizar la robustez de sus conclusiones a través de las distintas unidades muestrales de que dispone. Esta fue la estrategia analizada en la Sección 8.3 (modelos deterministas con paneles de datos).
2. Tratar las correlaciones que, en cada instante de tiempo, se producen entre las perturbaciones de las diferentes unidades muestrales. El Capítulo 8 (Parte II) (regresiones aparentemente no relacionadas) se dedicó en parte a tratar el modo eficiente de utilizar dichas correlaciones contemporáneas, con el objeto de ganar eficiencia en la estimación.

Por el contrario, en paneles microeconómicos, el investigador está interesado en analizar cómo varía el comportamiento de agentes económicos individuales frente a cuestiones como sus hábitos de consumo, su situación laboral, su nivel de estudios, etc. Estas son decisiones que dependerán de una lista de características socioeconómicas que el analista debe especificar como variables explicativas del modelo.

Sin embargo, no todos los agentes toman sus decisiones de igual modo: diferentes agentes, incluso si comparten las mismas características observables, toman decisiones distintas. Ello obliga a contemplar la existencia de *efectos latentes no observables*, específicos de cada agente encuestado, generalmente constantes en el tiempo, que inciden sobre el modo en que éste toma sus decisiones. Si estos efectos latentes existen y no se recogen explícitamente en el modelo, se producirá un problema de variables omitidas: los coeficientes estimados de las variables explicativas incluidas estarán sesgados, por recoger parcialmente los efectos individuales no observables. Este capítulo se dedica, precisamente, a presentar los procedimientos adecuados de utilización de paneles de datos con una dimensión de sección cruzada grande, con el objeto de estimar modelos econométricos que incluyen entre las variables explicativas los efectos individuales no observables.

Al disponer de un número reducido, T , de observaciones de cada uno de los N individuos de la muestra, podría pensarse en estimar un modelo econométrico con cada una de las T secciones cruzadas para luego comparar la evolución de los coeficientes del modelo a lo largo del tiempo. Hay dos tipos de razones que desaconsejan tal estrategia:

1. En primer lugar, la búsqueda de eficiencia sugeriría la estimación simultánea con las T secciones cruzadas, es decir, con todo el panel de datos, para explotar óptimamente el hecho de que los agentes (personas, familias o empresas, según la encuesta) en cada sección cruzada son los mismos,
2. Precisamente porque se dispone de T observaciones acerca de cada agente, es posible utilizar esta dimensión temporal para obtener estimaciones consistentes del modelo econométrico, a pesar de la presencia, entre sus variables explicativas, de los efectos latentes no observables. En efecto, un razonamiento intuitivo sugiere que si dichas variables latentes pueden supo-

nerse constantes en el tiempo, entonces la comparación de las decisiones de un mismo agente a lo largo del tiempo permitirá hacer inferencias relativas a sus determinantes, con independencia de la existencia de características individuales no observables. Este es el tipo de procedimientos que desarrollamos en este capítulo.

Ejemplo 15.1. Supongamos que se especifica un modelo de determinación de la rentabilidad de las empresas de una muestra:

$$R_{it} = \beta_0 + \beta_1 T_{it} + \alpha_i + u_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T$$

donde R_{it} y T_{it} denotan, respectivamente, la rentabilidad y la cuota de mercado de la empresa i -ésima; α_i recoge el efecto de variables no observables, características de cada empresa, estables en el tiempo, como pudieran ser la capacidad de los directivos para la gestión de la empresa, lo que podría tener un efecto sobre la rentabilidad de la misma, adicional a su cuota de mercado. Al estar, en general, correlacionados α_i y T_{it} , y pasar α_i a formar parte del término de error, tendríamos inconsistencia en la estimación de mínimos cuadrados.

Ejemplo 15.2 (Griliches, 1977). Otro ejemplo tradicional lo constituye la estimación de una función de salarios:

$$S_{it} = \beta_0 + \beta_1 E_{it} + \alpha_i + u_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T$$

en la que éstos, S_{it} , se hacen depender del nivel de estudios cursados por los individuos en la muestra, E_{it} (o por los años de educación recibidos). Es difícil pensar que las disparidades salariales que se observen entre individuos con igual nivel educativo sean puramente aleatorias, representadas por u_{it} ; más bien podría creerse que existen características individuales no medibles α_i , como la «habilidad» en el puesto de trabajo, que pueden explicar, al menos parcialmente, tales disparidades.

15.2. EL MODELO DE EFECTOS ALEATORIOS

La especificación general de un modelo econométrico de datos de panel es:

$$y_{it} = \mu + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\gamma} + \boldsymbol{\omega}'_i\boldsymbol{\delta} + \alpha_i + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad [15.1]$$

donde y_{it} es la variable endógena, en la que los subíndices i y t hacen referencia al agente económico del que proviene el dato y el instante de tiempo en que se recogió, respectivamente. \mathbf{x}_{it} es un vector de k variables explicativas, predeterminadas pero no necesariamente exógenas, es decir: $E(u_{it}/\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0$, aun que no necesariamente:

$$E(u_{it}/\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0$$

De hecho, en ocasiones x_{it} incluirá como componente algún retardo de la variable y_{it} , en cuyo caso diremos que el modelo es *dinámico*, lo que analizamos en la Sección 15.6.

El vector z_i está constituido por g variables exógenas *observables*, específicas de cada agente, pero invariantes en el tiempo, lo que las distingue de las variables que componen el vector x_{it} . Los agentes económicos que componen un panel microeconómico han sido generalmente escogidos al azar, mediante un proceso de muestreo previo y, en consecuencia, los efectos no observables o variables latentes deben tratarse como realizaciones de una distribución de probabilidad subyacente. La variable aleatoria α_i recoge tales efectos *no observables* específicos de cada agente en el panel; suponemos que, al igual que el vector z_i , es invariante en el tiempo. Al no ser observable, no resulta relevante asociar ningún coeficiente a α_i puesto que, en cualquier caso, no podría identificarse por separado de la propia variable. Por último, el vector ω_i representa una lista de variables que, evolucionando en el tiempo, afectan de igual manera a todos los individuos del panel; generalmente, cuando se incluye tal vector de variables, se hace tratando de corregir el comportamiento individual de los efectos puramente macroeconómicos como la inflación o el crecimiento agregado de la economía.

A efectos de los problemas que pretendemos ilustrar en este capítulo, nos basta sin embargo con considerar un modelo más sencillo:

$$y_{it} = \mu + x'_{it}\beta + z'_i\gamma + \alpha_i + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad [15.2]$$

al que nos referiremos en lo sucesivo, pudiendo generalizarse algunos de los resultados obtenidos al caso del modelo [15.1].

En lo sucesivo consideramos que, para formar las matrices globales, se disponen primero las T observaciones del primer individuo, seguidas de las T observaciones del segundo individuo, y así sucesivamente. Por ejemplo, el vector y , de dimensión $NT \times 1$, es $y' = (y_{11}, \dots, y_{1T}, y_{21}, \dots, y_{2T}, \dots, y_{N1}, \dots, y_{NT})$. El resto de las matrices y sus dimensiones son X , $NT \times k$, y Z $NT \times g$, con bloques de T filas iguales entre sí, mientras que los vectores α y u son $NT \times 1$, el primero de ellos con bloques de T coordenadas iguales entre sí; β es $k \times 1$ y γ es $g \times 1$. Las matrices propias de cada individuo se denotan por X_i y Z_i , que son respectivamente $T \times k$ y $T \times g$. El modelo [15.2] puede escribirse:

$$y = \mu \mathbf{1}_{NT} + X\beta + Z\gamma + (\alpha + u)$$

Ya hemos mencionado que las *variables latentes* α_i , específicas de cada individuo, deben tratarse como realizaciones de una variable aleatoria, y se trata de obtener estimaciones consistentes de los coeficientes del modelo en presencia de estos efectos. Al no ser observable α_i , pasa a formar parte del término de error del modelo estimado:

$$y_{it} = \mu + x'_{it}\beta + z'_i\gamma + v_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T \quad [15.3]$$

donde $v_{it} = u_{it} + \alpha_i$ para todo i, t , por lo que hay que distinguir entre dos situaciones radicalmente diferentes, según que el efecto individual α_i esté o no

correlacionado con las variables explicativas \mathbf{x}_{it} , \mathbf{z}_i : si, como en los Ejemplos 18.1 y 18.2, α_i está correlacionada con algunas variables en \mathbf{x}_{it} y \mathbf{z}_i , entonces las estimaciones de mínimos cuadrados ordinarios (MCO) de los coeficientes del modelo [15.3] serán inconsistentes; si, por el contrario, se tiene $\text{Corr}(\mathbf{x}'_{it}, \alpha_i) = \mathbf{0}_k$ y $\text{Corr}(\mathbf{z}'_i, \alpha_i) = \mathbf{0}_g$, entonces el problema de estimación es tan sólo un problema de eficiencia, debiendo diseñarse el modo óptimo de explotar la estructura de la matriz de covarianzas del término de error transformado v_{it} , lo que hacemos en la sección siguiente⁽³⁾.

15.3. ESTIMACION EFICIENTE EN AUSENCIA DE CORRELACIONES ENTRE LOS EFECTOS INDIVIDUALES NO OBSERVABLES Y LAS RESTANTES VARIABLES EXPLICATIVAS (El estimador de Balestra-Nerlove)

Suponemos en esta sección que en el modelo [15.3] se tiene:

$$\begin{aligned} E u_{it} &= E \alpha_i = E(\alpha_i u_{it}) = 0 \text{ para todo } i, t \\ E(\mathbf{x}'_{it} \alpha_i) &= E(\mathbf{x}'_{it} u_{it}) = \mathbf{0}_k \text{ para todo } i, t \\ E(\mathbf{z}'_i \alpha_i) &= E(\mathbf{z}'_i u_{it}) = \mathbf{0}_g \text{ para todo } i, t \\ E(\alpha_i \alpha_j) &= \sigma_\alpha^2 \text{ si } i = j, \text{ e igual a cero en caso contrario} \\ E(u_{it} u_{js}) &= \sigma_u^2 \text{ si } t = s \text{ e } i = j, \text{ e igual a cero en otro caso} \end{aligned}$$

Nótese que estos supuestos no contemplan la posibilidad de heteroscedasticidad ni autocorrelación en las observaciones procedentes de un mismo individuo⁽⁴⁾. Tampoco consideramos la existencia de correlaciones contemporáneas entre los términos de error correspondientes a individuos diferentes. Lo que distingue esta sección de las siguientes es el supuesto de que las variables latentes α_i no están correlacionadas con las restantes variables explicativas del modelo.

La varianza condicional de la variable endógena y_{it} para todo i, t es igual a $\sigma_y^2 = \sigma_v^2 = \sigma_u^2 + \sigma_\alpha^2$. Con ello, la matriz de covarianzas de los T valores correspondientes al individuo i -ésimo, $\mathbf{v}_i = (u_{i1} + \alpha_i, \dots, u_{iT} + \alpha_i)$, es:

$$\mathbf{V} = E(\mathbf{v}_i \mathbf{v}'_i) = \sigma_u^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}'_T = \begin{pmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \dots & \dots & \dots & \dots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_u^2 + \sigma_\alpha^2 \end{pmatrix}$$

⁽³⁾ Cuando $T = 1$, MCO es consistente y eficiente.

⁽⁴⁾ Aunque ambas podrían tratarse como una extensión del modelo que aquí discutimos.

donde $\mathbf{1}_T$ es un vector columna formado por T unos. La matriz \mathbf{V} , de dimensión $T \times T$, ilustra que la correlación entre dos observaciones provenientes de un mismo individuo es constante y no desaparece en el tiempo, debido al persistente efecto de α_i , lo que se refleja en la presencia del término σ_α^2 a lo largo de toda la matriz de covarianzas⁽⁵⁾. De esta manera, incluso si el término de error original u_{it} no tenía autocorrelación, el término de error agregado v_{it} sí la tiene.

La matriz \mathbf{V} es la misma para cada individuo en la muestra, por lo que $\text{Var } \mathbf{u} = \mathbf{I}_N \otimes \mathbf{V}$, de dimensión $TN \times TN$. Esta matriz es diagonal a bloques porque $E[(\alpha_i \mathbf{1}_T + \mathbf{u}_i)(\alpha_j \mathbf{1}_T + \mathbf{u}_j)'] = \mathbf{0}_{T \times T}$ para $i \neq j$. Utilizando el lema de inversión de matrices que vimos en la Sección 1.3 puede verse que la matriz de covarianzas \mathbf{V} tiene como inversa:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} \mathbf{1}_T \mathbf{1}_T' \right)$$

El problema de estimación en este contexto es un problema de eficiencia, debiendo diseñarse el modo óptimo de explotar la estructura de la matriz de covarianzas. El estimador de mínimos cuadrados generalizados (MCG) es el estimador lineal, insesgado de mínima varianza:

$$\begin{pmatrix} \hat{\mu} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left[\sum_1^N \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{W}_i \right]^{-1} \left[\sum_1^N \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{y}_i \right]$$

donde $\mathbf{W}_i = (\mathbf{1}_T; \mathbf{X}_i; \mathbf{Z}_i)$ es una matriz $T \times (1 + k + g)$. Para simplificar las expresiones analíticas, tiene interés introducir el parámetro $\psi = \sigma_\alpha^2 / (\sigma_u^2 + T\sigma_\alpha^2)$. Con él, la matriz \mathbf{V}^{-1} puede escribirse:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2} \left[\mathbf{Q} + \psi \frac{1}{T} \mathbf{1}_T \mathbf{1}_T' \right]$$

donde $\mathbf{Q} = \mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T'$ es la matriz que ya introdujimos en el Capítulo 1, y que transforma las variables correspondientes a cada individuo en diferencias con respecto a su media temporal.

Las ecuaciones normales del estimador MCG pueden escribirse⁽⁶⁾:

⁽⁵⁾ A diferencia de lo que ocurre con la matriz de covarianzas de un modelo de regresión con perturbaciones AR(1), cuyos elementos van disminuyendo de magnitud según se alejan de la diagonal principal.

⁽⁶⁾ Como se pide en un problema al final del capítulo.

correlacionado con las variables explicativas \mathbf{x}_{it} , \mathbf{z}_i : si, como en los Ejemplos 18.1 y 18.2, α_i está correlacionada con algunas variables en \mathbf{x}_{it} y \mathbf{z}_i , entonces las estimaciones de mínimos cuadrados ordinarios (MCO) de los coeficientes del modelo [15.3] serán inconsistentes; si, por el contrario, se tiene $\text{Corr}(\mathbf{x}'_{it}, \alpha_i) = \mathbf{0}_k$ y $\text{Corr}(\mathbf{z}'_i, \alpha_i) = \mathbf{0}_g$, entonces el problema de estimación es tan sólo un problema de eficiencia, debiendo diseñarse el modo óptimo de explotar la estructura de la matriz de covarianzas del término de error transformado v_{it} , lo que hacemos en la sección siguiente⁽³⁾.

15.3. ESTIMACION EFICIENTE EN AUSENCIA DE CORRELACIONES ENTRE LOS EFECTOS INDIVIDUALES NO OBSERVABLES Y LAS RESTANTES VARIABLES EXPLICATIVAS (El estimador de Balestra-Nerlove)

Suponemos en esta sección que en el modelo [15.3] se tiene:

$$\begin{aligned} Eu_{it} &= E\alpha_i = E(\alpha_i u_{it}) = 0 \text{ para todo } i, t \\ E(\mathbf{x}'_{it} \alpha_i) &= E(\mathbf{x}'_{it} u_{it}) = \mathbf{0}_k \text{ para todo } i, t \\ E(\mathbf{z}'_i \alpha_i) &= E(\mathbf{z}'_i u_{it}) = \mathbf{0}_g \text{ para todo } i, t \\ E(\alpha_i \alpha_j) &= \sigma_\alpha^2 \text{ si } i = j, \text{ e igual a cero en caso contrario} \\ E(u_{it} u_{js}) &= \sigma_u^2 \text{ si } t = s \text{ e } i = j, \text{ e igual a cero en otro caso} \end{aligned}$$

Nótese que estos supuestos no contemplan la posibilidad de heteroscedasticidad ni autocorrelación en las observaciones procedentes de un mismo individuo⁽⁴⁾. Tampoco consideramos la existencia de correlaciones contemporáneas entre los términos de error correspondientes a individuos diferentes. Lo que distingue esta sección de las siguientes es el supuesto de que las variables latentes α_i no están correlacionadas con las restantes variables explicativas del modelo.

La varianza condicional de la variable endógena y_{it} para todo i, t es igual a $\sigma_y^2 = \sigma_v^2 = \sigma_u^2 + \sigma_\alpha^2$. Con ello, la matriz de covarianzas de los T valores correspondientes al individuo i -ésimo, $v_i = (u_{i1} + \alpha_i, \dots, u_{iT} + \alpha_i)$, es:

$$\mathbf{V} = E(v_i v_i') = \sigma_u^2 \mathbf{1}_T + \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' = \begin{pmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \dots & \dots & \dots & \dots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_u^2 + \sigma_\alpha^2 \end{pmatrix}$$

⁽³⁾ Cuando $T = 1$, MCO es consistente y eficiente.

⁽⁴⁾ Aunque ambas podrían tratarse como una extensión del modelo que aquí discutimos.

donde $\mathbf{1}_T$ es un vector columna formado por T unos. La matriz \mathbf{V} , de dimensión $T \times T$, ilustra que la correlación entre dos observaciones provenientes de un mismo individuo es constante y no desaparece en el tiempo, debido al persistente efecto de α_i , lo que se refleja en la presencia del término σ_α^2 a lo largo de toda la matriz de covarianzas⁽⁵⁾. De esta manera, incluso si el término de error original u_{it} no tenía autocorrelación, el término de error agregado v_{it} sí la tiene.

La matriz \mathbf{V} es la misma para cada individuo en la muestra, por lo que $\text{Var } \mathbf{u} = \mathbf{I}_N \otimes \mathbf{V}$, de dimensión $TN \times TN$. Esta matriz es diagonal a bloques porque $E[(\alpha_i \mathbf{1}_T + \mathbf{u}_i)(\alpha_j \mathbf{1}_T + \mathbf{u}_j)'] = \mathbf{0}_{T \times T}$ para $i \neq j$. Utilizando el lema de inversión de matrices que vimos en la Sección 1.3 puede verse que la matriz de covarianzas \mathbf{V} tiene como inversa:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} \mathbf{1}_T \mathbf{1}_T' \right)$$

El problema de estimación en este contexto es un problema de eficiencia, debiendo diseñarse el modo óptimo de explotar la estructura de la matriz de covarianzas. El estimador de mínimos cuadrados generalizados (MCG) es el estimador lineal, insesgado de mínima varianza:

$$\begin{pmatrix} \hat{\mu} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left[\sum_1^N \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{W}_i \right]^{-1} \left[\sum_1^N \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{y}_i \right]$$

donde $\mathbf{W}_i = (\mathbf{1}_T; \mathbf{X}_i; \mathbf{Z}_i)$ es una matriz $T \times (1 + k + g)$. Para simplificar las expresiones analíticas, tiene interés introducir el parámetro $\psi = \sigma_\alpha^2 / (\sigma_u^2 + T\sigma_\alpha^2)$. Con él, la matriz \mathbf{V}^{-1} puede escribirse:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2} \left[\mathbf{Q} + \psi \frac{1}{T} \mathbf{1}_T \mathbf{1}_T' \right]$$

donde $\mathbf{Q} = \mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T'$ es la matriz que ya introdujimos en el Capítulo 1, y que transforma las variables correspondientes a cada individuo en diferencias con respecto a su media temporal.

Las ecuaciones normales del estimador MCG pueden escribirse⁽⁶⁾:

⁽⁵⁾ A diferencia de lo que ocurre con la matriz de covarianzas de un modelo de regresión con perturbaciones AR(1), cuyos elementos van disminuyendo de magnitud según se alejan de la diagonal principal.

⁽⁶⁾ Como se pide en un problema al final del capítulo.

$$\begin{pmatrix} \psi NT & \psi T \sum_1^N \bar{x}_i' & \psi T \sum_1^N z_i' \\ \psi T \sum_1^N \bar{x}_i & \sum_1^N \mathbf{X}_i' \mathbf{Q} \mathbf{X}_i + \psi T \sum_1^N \bar{x}_i \bar{x}_i' & \sum_1^N \mathbf{X}_i' \mathbf{Q} \mathbf{Z}_i + \psi T \sum_1^N \bar{x}_i z_i' \\ \psi T \sum_1^N z_i & \sum_1^N \mathbf{Z}_i' \mathbf{Q} \mathbf{X}_i + \psi T \sum_1^N z_i \bar{x}_i' & \sum_1^N \mathbf{Z}_i' \mathbf{Q} \mathbf{Z}_i + \psi T \sum_1^N z_i z_i' \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} =$$

$$= \begin{pmatrix} \psi NT \bar{y} \\ \sum_1^N \mathbf{X}_i' \mathbf{Q} \mathbf{y}_i + \psi T \sum_1^N \bar{x}_i \bar{y}_i \\ \sum_1^N \mathbf{Z}_i' \mathbf{Q} \mathbf{y}_i + \psi T \sum_1^N z_i \bar{y}_i \end{pmatrix}$$

donde \bar{x}_i , \bar{y}_i , $i = 1, 2, \dots, N$ denotan las medias muestrales de las características de cada individuo, así como de su variable endógena a través del tiempo.

La matriz de covarianzas de este estimador MCG es la inversa de la matriz que aparece en el miembro izquierdo de las ecuaciones normales. En el caso particular en que no aparecen las variables z_i , se tiene:

$$\text{Var}(\hat{\beta}_{\text{MCG}}) = \sigma_u^2 \left[\sum_1^N (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i') (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i') + \psi T \sum_1^N (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \right]^{-1}$$

En ocasiones es difícil mantener el supuesto de que no hay heteroscedasticidad ni autocorrelación en las observaciones procedentes de un mismo individuo. En tal caso, si no existe información suficiente como para modelizar en detalle tales características, puede ser preferible estimar por MCO, aunque utilizando la estimación *consistente* de la matriz de covarianzas propuesta por White (1980):

$$\text{Var}(\hat{\beta}_{\text{MCO}}) = (\mathbf{X}' \mathbf{X})^{-1} \left(\frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T (\mathbf{x}_{is} \hat{u}_{is}) (\hat{u}_{it} \mathbf{x}_{it}') \right) (\mathbf{X}' \mathbf{X})^{-1}$$

que implica operaciones con matrices de gran dimensión.

15.4. ESTIMACION CONSISTENTE EN PRESENCIA DE CORRELACIONES ENTRE LOS EFECTOS INDIVIDUALES NO OBSERVABLES Y LAS RESTANTES VARIABLES EXPLICATIVAS

Pasamos ahora a examinar el caso más verosímil en que existen correlaciones entre las variables latentes α_i , aleatorias y no observables, y las características observables x_{it} . Puesto que el problema está generado por la presencia de los efectos individuales, cabe preguntarse acerca de la existencia de alguna transformación del modelo [15.2] que elimine tales efectos. En realidad, existen dos transformaciones diferentes con esta propiedad, que examinamos sucesivamente.

15.4.a. El estimador intragrupos

Una primera posibilidad consiste en observar que si se promedian las ecuaciones [15.2] para cada individuo se tiene, para cada $i = 1, \dots, N$:

$$\bar{y}_i = \mu + \bar{x}_i' \beta + z_i' \gamma + \alpha_i + \bar{u}_i \quad [15.4]$$

donde, por ejemplo, $\bar{y}_i = T^{-1} \sum_1^T y_{it}$ y hemos recogido explícitamente el hecho de que $\bar{z}_i = z_i$, $\bar{\alpha}_i = \alpha_i$. Restando [15.4] de [15.2], se tiene:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + (u_{it} - \bar{u}_i) \quad [15.5]$$

donde no aparece α_i , por lo que si las variables x_{it} son estrictamente exógenas, entonces el estimador MCO del vector β será consistente⁽⁷⁾ incluso si α_i está correlacionada con las variables x_{it} ; el estimador de σ_u^2 que se obtenga con los residuos así generados será asimismo consistente. Este estimador se conoce como *estimador intragrupos*⁽⁸⁾. El requisito de que las variables x_{it} sean exógenas, y no sólo predeterminadas, para garantizar la consistencia del estimador se debe a la presencia de la media muestral de u_{i1}, \dots, u_{it} en el término de error del modelo transformado.

Este procedimiento tiene, sin embargo, una limitación: como puede verse en [15.5], el vector z_i también ha desaparecido del modelo, por lo que los coeficientes γ de dichos efectos individuales observables no pueden estimarse

⁽⁷⁾ El lector debería ver sin gran dificultad cuál sería el procedimiento análogo si los efectos son temporales e iguales para todos los individuos, en vez de ser específicos de los individuos de la muestra y constantes en el tiempo.

⁽⁸⁾ En contra de lo que inicialmente pudiera parecer por ser $\text{Var}(u_{it} - \bar{u}_i) = \sigma^2 Q$ el estimador intragrupos, que es del tipo MCO, es eficiente. Ello se debe a que la matriz Q puede descomponerse: $Q = A'A$, donde A es la matriz $(T-1) \times T$ que genera las diferencias: $y^* =$

$y - Ay = \left(\sqrt{\frac{t}{t-1}} (y_t - \bar{y}) \right)_{t=2}^T$, y es tal que $AA' = I_{T-1}$, y $\text{Var}(AQu_i) = \text{Var}(Au_i) = \sigma^2 I_{T-1}$.

Como el estimador intragrupos puede interpretarse como la aplicación de MCO a la regresión de Ay_i sobre Ax_i , se tiene que es eficiente.

en [15.5]; en la Sección 15.7 explicamos un procedimiento propuesto por Hausman y Taylor (1981) para obtener estimaciones de los coeficientes γ , una vez estimado el vector β .

La expresión analítica del estimador intragrupos, que consiste en utilizar MCO en [15.5], es:

$$\hat{\beta}_{IG} = \left[\sum_{i=1}^N (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i)' (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i) \right]^{-1} \left[\sum_{i=1}^N (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i)' (\mathbf{y}_i - \mathbf{1}_T \bar{y}_i) \right]$$

con matriz de covarianzas:

$$\text{Var}(\hat{\beta}_{IG}) = \sigma_u^2 \left[\sum_{i=1}^N (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i)' (\mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i) \right]^{-1}$$

Puede demostrarse que la diferencia entre las matrices de covarianzas de los estimadores MCG e intragrupos es definida negativa, ya que el parámetro ψ es positivo. Ello quiere decir que si los efectos latentes estuviesen incorrelacionados con el término de error, entonces es menos eficiente tratarlos como correlacionados con u_t . Sin embargo, si, efectivamente, dichos efectos son aleatorios y correlacionados con u_t , entonces el estimador MCG es inconsistente, y de poco sirve su eficiencia relativa.

Como en cualquier otro modelo estimado con variables transformadas, el parámetro σ_u^2 deberá estimarse a partir de los residuos que se obtienen con las estimaciones intragrupos obtenidas en el modelo transformado, pero aplicándolas a las variables originales, es decir, sin diferencias respecto a la media.

Como puede desprenderse de las expresiones anteriores, el estimador intragrupos utiliza únicamente la variación que se produce entre las observaciones procedentes de cada individuo, pero no a través de todo el panel de datos. Por no utilizar toda la información muestral, el estimador intragrupos no es, en general, eficiente; sí lo es cuando el modelo verdadero es de efectos deterministas, como ocurría en el Capítulo 8, donde ya propusimos este procedimiento de estimación para eliminar la heterogeneidad de los términos independientes de cada submuestra. El carácter determinista de los efectos latentes es más aceptable cuando la muestra incluye a *todas* las unidades de decisión del problema analizado, como por ejemplo: todas las empresas de los sectores productivos que se analizan, en un estudio comparativo de las estrategias de endeudamiento seguidas por las empresas de cada sector, o de series temporales procedentes de todos los fabricantes de automóviles de un país, en un estudio de las características de dicho sector.

En caso de existencia de heteroscedasticidad o autocorrelación en cada individuo, puede utilizarse la expresión de White para calcular la matriz de covarianzas del estimador intragrupos, como ya apuntamos en el caso del estimador MCG.

15.4.b. El estimador en primeras diferencias

Una transformación alternativa que permite eliminar del modelo la *variable latente* α_i consiste en tomar diferencias temporales en [15.2], para obtener:

$$\Delta y_{it} = (\Delta \mathbf{x}_{it})' \boldsymbol{\beta} + \Delta u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

aunque hay que notar que el término de error transformado ya no es ruido blanco, incluso si u_{it} lo era. La estimación MCO de este modelo será consistente aunque, como ocurría antes, no podemos estimar el vector $\boldsymbol{\gamma}$. La expresión analítica del estimador es:

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' (\Delta \mathbf{X}_i)]^{-1} [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' (\Delta \mathbf{y}_i)]$$

donde la matriz $\Delta \mathbf{X}_i$, de dimensión $(T-1) \times k$, tiene por elemento genérico: $x_{ijt} - x_{ijt-1}$. Los tres subíndices indican, respectivamente, el individuo del que procede la observación, la variable que se está considerando, y el periodo temporal correspondiente; el vector $\Delta \mathbf{y}_i$, de dimensión $(T-1) \times 1$, tiene por componente genérica $y_{it} - y_{it-1}$. La matriz de covarianzas del estimador en diferencias es:

$$\text{Var } \hat{\boldsymbol{\beta}} = [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' (\Delta \mathbf{X}_i)]^{-1} [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' \mathbf{H} (\Delta \mathbf{X}_i)] [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' \Delta \mathbf{X}_i]^{-1}$$

donde $\text{Var}(\Delta u_i) = \sigma^2 \mathbf{H}$, siendo \mathbf{H} una matriz simétrica, de dimensión $(T-1) \times (T-1)$, cuyos elementos son iguales a 2 en la diagonal principal, igual a -1 en las diagonales adyacentes a la principal, e igual a 0 en el resto.

La limitación de este estimador es que si, como ocurre con la variable educación en el Ejemplo 15.2 al comienzo de este capítulo, las variables \mathbf{x}_{it} del modelo [15.2] son constantes en el tiempo, dichas variables desaparecerían de la formulación en primeras diferencias, no pudiendo estimarse sus coeficientes.

Puesto que conocemos que $\text{Var}(\Delta u_i) = \sigma^2 \mathbf{H}$, podemos utilizar MCG:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' \mathbf{H}^{-1} (\Delta \mathbf{X}_i)]^{-1} [\boldsymbol{\Sigma}_1^N (\Delta \mathbf{X}_i)' \mathbf{H}^{-1} (\Delta \mathbf{y}_i)]$$

que coincide *numéricamente* con el estimador intragrupos.

En la práctica, las estimaciones en niveles (MCG) y en diferencias suelen proporcionar resultados muy diferentes, lo que sugiere la existencia de efectos individuales *no observables* que sesgan la estimación en niveles; en tal caso, el estimador en primeras diferencias es más fiable. Por otra parte, cuando la dimensión temporal T del panel es realmente pequeña, puede carecer de sentido considerar los promedios de las observaciones de cada individuo, teniendo en tales condiciones mucho más interés el estimador en primeras diferencias que el estimador intragrupos. Finalmente, cuando $T = 2$, los estimadores intragrupos y en primeras diferencias coinciden.

15.4.c. El estimador entre grupos

Un tercer estimador, el *estimador entre grupos*, consiste en aplicar MCO al modelo que resulta al relacionar las medias temporales de las observaciones correspondientes a los distintos individuos en la muestra, y que surge de promediar en [15.2], es decir:

$$\bar{y}_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + \mathbf{z}_i' \boldsymbol{\gamma} + (\alpha_i + \bar{u}_i), \quad i = 1, 2, \dots, N$$

donde, por ejemplo, $\bar{y}_i = T^{-1} \sum_1^N y_{it}$, y hemos utilizado nuevamente el hecho de que tanto α_i como \mathbf{z}_i son invariantes en el tiempo. Puesto que *no han desaparecido los efectos individuales no observables*, el estimador entre grupos es consistente *sólo* cuando α_i está incorrelacionado con *todas* las variables explicativas. Por otra parte, sus limitaciones en términos de eficiencia son análogas a las de los estimadores anteriores.

15.4.d. Relación entre estimadores

En el apéndice a este capítulo se explica cómo el estimador MCG no es sino una combinación lineal convexa de los estimadores intragrupos y entre grupos. Si denotamos por $\psi = \sigma_u^2 / (\sigma_u^2 + T\sigma_\alpha^2)$, se tiene que el estimador MCG es tanto más próximo a utilizar MCO directamente en [15.2] cuanto más próximo sea ψ a 1, y tanto más similar al estimador intragrupos, cuanto más próximo sea ψ a 0. Estos resultados se interpretan en el apéndice, donde también se explica cómo llevar a cabo en la práctica la estimación de *mínimos cuadrados generalizados*. En particular, para que este estimador sea factible, hay que obtener previamente estimaciones consistentes de los parámetros σ_u^2 y σ_α^2 .

15.5. CONTRASTES DE ESPECIFICACION

Puesto que la posible existencia de correlaciones entre los efectos latentes α_i y las variables explicativas $(\mathbf{x}_{it}, \mathbf{z}_i)$ juega un papel tan crucial en el tratamiento del modelo, es importante poder contrastar la hipótesis nula de ausencia de tales correlaciones, que representamos:

$$H_0: E(\alpha_i | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, \mathbf{z}_i) = 0$$

Bajo tal supuesto, el estimador MCG (o de Banestra-Nerlove) es consistente y también de mínima varianza, siendo inconsistente cuando las variables latentes están correlacionadas con las variables explicativas (es decir, cuando H_0 es falsa). Por el contrario, el estimador intragrupos es consistente, tanto si H_0 es cierta como si no lo es. En consecuencia, en ausencia de tales correlaciones, los valores numéricos de ambos estimadores serán muy simila-

res, tendiendo a diferir cuando efectos latentes y variables explicativas están correlacionados.

Aplicando el lema de Hausman (1978), podemos definir el vector $\hat{q} = \hat{\beta}_{IG}^* - \hat{\beta}_{MCG}^*$, donde $\hat{\beta}_{IG}^*$ y $\hat{\beta}_{MCG}^*$ denotan los vectores que se obtienen al eliminar el término independiente de los estimadores intragrupos y MCG, teniéndose: $\text{Var}(\hat{q}) = \text{Var}(\hat{\beta}_{IG}^*) - \text{Var}(\hat{\beta}_{MCG}^*)$ y, bajo la hipótesis nula, el estadístico de Wald: $N\hat{q}'[\text{Var}(\hat{q})]^{-1}\hat{q}$ se distribuye como una variable chi-cuadrado con $k - 1$ grados de libertad, el número de coeficientes estimados. Si \hat{q} resulta ser pequeño de acuerdo con esta medida, se mantiene la hipótesis nula de ausencia de correlaciones, rechazándose en caso contrario.

Si no se rechaza H_0 , debe utilizarse el estimador MCG, que es óptimo en tales condiciones. En caso contrario, hay que aceptar que existen correlaciones entre efectos individuales y variables explicativas, en cuyo caso $\hat{\beta}_{MCG}$ no es consistente y $\hat{\beta}_{IG}$ sí lo es, aunque es ineficiente⁽⁹⁾.

Probamos en el apéndice que, a medida que la dimensión temporal T del panel tiende a infinito, el estimador MCG se reduce al estimador intragrupos, por lo que este contraste carece de sentido, estando concebido para el caso habitual en que la dimensión «larga» del panel es la de la sección cruzada.

Es interesante puntualizar que, si existe heteroscedasticidad en la sección cruzada, o autocorrelación en los datos individuales, el estimador MCG deja de ser eficiente⁽¹⁰⁾, por lo que el contraste propuesto pierde nuevamente su significación. Arellano y Bover (1990) proponen la siguiente estrategia: consideremos un sistema de ecuaciones ampliado, en el que aparece un grupo de ecuaciones para los niveles, así como otro para las primeras diferencias; por ejemplo, cuando $T = 3$ tendríamos el sistema:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \Delta y_{i2} \end{pmatrix} = \begin{pmatrix} x'_{i1} \\ x'_{i2} \\ \Delta x'_{i2} \end{pmatrix} \beta + \begin{pmatrix} 0 \\ 0 \\ \Delta x'_{i2} \end{pmatrix} \alpha + \begin{pmatrix} v_{i1} \\ v_{i2} \\ \Delta v_{i3} \end{pmatrix}, \quad i = 1, 2, \dots, N$$

en el que tanto α como β pueden estimarse consistentemente por MCO. Un contraste de significación del coeficiente α en este modelo es un contraste de ausencia de correlación entre variables latentes y explicativas. Este procedimiento se desarrolla en Arellano (1993).

Hemos puesto de manifiesto repetidamente que la utilización fundamental de paneles de datos en la actualidad intenta explotar la variación temporal en el comportamiento de los agentes individuales para tratar adecuadamente la presencia de variables latentes en el modelo. En consecuencia, en muchos

⁽⁹⁾ La hipótesis nula podría también rechazarse debido a la inconsistencia de ambos estimadores como consecuencia, por ejemplo, de errores de medida en las variables. Si hay razones que justifiquen tal creencia, habría que estimar el modelo original por variables instrumentales para obtener un estimador consistente de la matriz de covarianzas; en una segunda etapa se compararían los estimadores MCG y de variables instrumentales, obtenido utilizando la matriz de covarianzas consistente.

⁽¹⁰⁾ Aunque podrían pensarse en extensiones de dicho estimador que tratase tales características adecuadamente, logrando la eficiencia.

casos se reconoce explícitamente el carácter dinámico de las decisiones económicas, lo que genera problemas adicionales a los que hasta ahora hemos considerado, y que son el objeto de la sección siguiente.

15.6. MODELOS DINAMICOS ⁽¹¹⁾

15.6.a. Estimación consistente de modelos autorregresivos

En nuestro tratamiento de modelos dinámicos con datos de panel ignoramos la presencia del vector \mathbf{z}_i . En realidad, un sencillo modelo AR(1) con efectos individuales α_i :

$$y_{it} = \rho y_{it-1} + \alpha_i + u_{it} \quad [15.6]$$

es suficiente para ilustrar los problemas que reviste la utilización en modelos dinámicos de los estimadores que hasta ahora hemos presentado:

a) En primer lugar, al igual que en otros modelos econométricos, la utilización del *estimador MCO* directamente en [15.6] es inconsistente, debido a la correlación entre α_i e y_{it-1} ; puede demostrarse que su sesgo asintótico no tiende a cero, y es negativo para valores $\rho > 0$.

b) El *estimador intragrupos* consiste en utilizar MCO con las variables del modelo [15.6] transformadas en desviaciones con respecto a sus promedios individuales, es decir, calculados a través del tiempo. A diferencia del caso estático, en el modelo dinámico [15.6] este estimador es inconsistente debido a la correlación entre las variables transformadas ⁽¹²⁾ $\tilde{y}_{it-1} = y_{it-1} - \bar{y}_{it-1}$ y $\tilde{u}_{it} = u_{it} - \bar{u}_{it}$. Como puede verse en Nickell (1981), o en Arellano y Bover (1990), bajo determinados supuestos ⁽¹³⁾, el sesgo asintótico es positivo para $\rho > 0$ y aumenta con σ_u^2 ; es de orden $1/T$, por lo que disminuye al aumentar la dimensión temporal del panel, pero, habitualmente, T es muy pequeño en paneles microeconómicos, por lo que *el sesgo del estimador intragrupos es importante* ⁽¹⁴⁾.

c) Por último, el *estimador MCO en primeras diferencias* es asimismo inconsistente. En efecto, el modelo [15.6] se convierte en:

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it} \quad [15.7]$$

del que ha desaparecido el efecto individual α_i ; sin embargo, ahora Δy_{it-1} y Δu_{it} están correlacionados, puesto que y_{it-1} y u_{it-1} lo están. El sesgo en muestras finitas de este estimador no depende del tamaño muestral y, por

⁽¹¹⁾ Esta sección se basa, en su práctica totalidad, en Arellano y Bover (1990).

⁽¹²⁾ A cada variable debe sustraerse la media muestral obtenida hasta dicho período.

⁽¹³⁾ a) El valor absoluto de ρ es inferior a la unidad y b) la perturbación u_{it} es ruido blanco.

⁽¹⁴⁾ Entre otras cosas, ello hace que el contraste de especificación que discutimos en la Sección 15.5 ya no sea válido.

consiguiente, no tiende a cero; de hecho, puede probarse que, al tender T a infinito, se tiene $plim(\hat{\rho} - \rho) = -(1 + \rho)/2$, que es negativo cuando $\rho > 0$, lo que implica que se subestima la estructura dinámica del modelo [15.6].

Para estimar consistentemente el modelo [15.7], Anderson y Hsiao (1981) observaron el hecho de que, si el término de error u_{it} no tiene autocorrelación, entonces y_{it-2} (que, como muestra [15.6], depende de u_{it-2} y anteriores) está claramente correlacionado con Δy_{it-1} , pero no con Δu_{it} (que sólo depende de u_{it} y u_{it-1}). Así, tanto y_{it-2} como los retardos previos son instrumentos válidos de Δy_{it-1} en [15.7]; con tres observaciones temporales en el panel ($T = 3$), y_{i1} podría utilizarse como instrumento de Δy_{i2} , en cuyo caso se estimaría tan sólo con la sección cruzada de primeras diferencias correspondiente a $T = 3$, teniéndose:

$$\hat{\rho} = \frac{\sum_{i=1}^N y_{i1}(y_{i3} - y_{i2})}{\sum_{i=1}^N y_{i1}(y_{i2} - y_{i1})}$$

Cuando la muestra consta de cuatro o más observaciones temporales, pueden utilizarse más secciones cruzadas en la estimación, resultando el estimador:

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{t=3}^T y_{i(t-2)} \Delta y_{it}}{\sum_{i=1}^N \sum_{t=3}^T y_{i(t-2)} \Delta y_{i(t-1)}} \quad [15.8]$$

que se conoce como *estimador de Anderson-Hsiao*⁽¹⁵⁾. Este estimador es consistente, tanto cuando T es fijo y $N \rightarrow \infty$, el caso habitual, como cuando N es fijo y $T \rightarrow \infty$.

Este estimador utiliza un instrumento por ecuación; sin embargo, el argumento anterior muestra que el número de instrumentos disponible va aumentando con el horizonte temporal de cada sección cruzada. Para ganar eficiencia, Arellano y Bond (1991) observaron que cuando T es pequeño en relación a N , si bien y_{i1} es el único instrumento válido en la ecuación en primeras diferencias correspondiente a $t = 3$, tanto y_{i1} como y_{i2} son instrumentos válidos en la estimación de la ecuación para $t = 4$, y así sucesivamente. De este modo, la matriz de observaciones de los instrumentos es:

(15) Finalmente, una variante de este procedimiento utiliza Δy_{it-2} en vez de y_{it-2} como instrumento, supuesto que se disponga al menos de cuatro observaciones temporales. Las sumas en la expresión analítica del estimador [15.8] oscilan entonces entre 4 y T .

$$\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{i(T-2)} \end{pmatrix}$$

que es de dimensión $(T-2) \times q$, siendo $q = (T-1)(T-2)/2$; este número de columnas proviene de la existencia de un número creciente de instrumentos para la ecuación de cada periodo, comenzando con un instrumento para la ecuación correspondiente a $t = 3$. En definitiva, la matriz \mathbf{Z}_i recoge todas las condiciones de ortogonalidad del tipo: $E[\mathbf{z}_i' \Delta \mathbf{u}_i] = 0$. Si los términos de error u_{it} , $t = 3, \dots, T$ son independientes, con varianza σ_u^2 , entonces se tiene $E[\Delta \mathbf{u}_i \Delta \mathbf{u}_i'] = \sigma_u^2 \mathbf{H}$, donde \mathbf{H} es la matriz que antes vimos, pero con dimensión $(T-2) \times (T-2)$. El estimador óptimo de variables instrumentales es:

$$\hat{\rho} = \frac{\left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{y}_{i(-1)} \right)' \left[\sum_{i=1}^N \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i \right]^{-1} \left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{y}_i \right)}{\left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{y}_{i(-1)} \right)' \left[\sum_{i=1}^N \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i \right]^{-1} \left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{y}_{i(-1)} \right)} \quad [15.9]$$

donde $\Delta \mathbf{y}_i = [\Delta y_{i3}, \dots, \Delta y_{iT}]$ e $\Delta \mathbf{y}_{i(-1)} = [\Delta y_{i2}, \dots, \Delta y_{i(T-1)}]$. El corchete central en el numerador y denominador de [15.9] estiman la matriz de covarianzas del vector $\mathbf{Z}_i' \Delta \mathbf{u}_i$: $\text{Var}(\mathbf{Z}_i' \Delta \mathbf{u}_i) = E[(\mathbf{Z}_i' \Delta \mathbf{u}_i)(\Delta \mathbf{u}_i \mathbf{Z}_i)] = \sigma_u^2 E(\mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i)$.

Si los errores u_{it} no son homoscedásticos, entonces el estimador de variables instrumentales sigue siendo consistente, pero se aumenta su eficiencia sustituyendo el corchete por una estimación más general de la matriz de covarianzas, como $\Sigma_i(\mathbf{Z}_i' \Delta \hat{\mathbf{u}}_i)(\Delta \hat{\mathbf{u}}_i' \mathbf{Z}_i)$, donde los $\hat{\mathbf{u}}_i$ provienen de una primera etapa en la que se ha utilizado la expresión [15.9] ignorando la presencia de heteroscedasticidad, es decir, sin incorporar la matriz \mathbf{H} en el término central del numerador y denominador. Este es el *estimador de variables instrumentales óptimo en dos etapas, o estimador generalizado de momentos en dos etapas* (EGM). Tal nombre proviene de que el estimador [15.9] minimiza la discrepancia entre los momentos muestrales $N^{-1} \sum_i^N \mathbf{Z}_i' \Delta \mathbf{u}_i$ y su valor poblacional, que es cero por las condiciones de ortogonalidad que caracterizan a los instrumentos \mathbf{Z}_i .

Aunque hemos propuesto el EGM en el contexto de modelos dinámicos, debe entenderse que su utilización es asimismo recomendable en el caso de modelos estáticos, en que las correlaciones existentes entre alguna de las variables explicativas y el término de error del modelo aconsejen el uso de variables instrumentales.

15.6.b. Contrastes de especificación en modelos dinámicos

Además de proporcionar estimaciones consistentes y relativamente eficientes, el procedimiento EGM sugiere diversos contrastes de especificación del modelo [15.6]:

a) Al utilizar retardos de y_{it-1} como instrumentos bajo el supuesto de que u_{it} es ruido blanco, este estimador sería inconsistente de no cumplirse tal hipótesis. Como el modelo a estimar [15.7] se ha formulado en primeras diferencias, entonces se introduce autocorrelación en los residuos, pero ésta ha de ser de primer orden si u_{it} era inicialmente ruido blanco y no superior. Bajo la hipótesis nula de ausencia de autocorrelación en u_{it} , el estadístico $\Sigma_1^N \Sigma_1^T [\Delta \hat{u}_{it} \Delta \hat{u}_{it-2}] / \hat{\sigma}^2$ se distribuye asintóticamente como una $N(0, 1)$, siendo $\hat{\sigma}^2$ un estimador consistente de la varianza asintótica de $(16) N^{-1/2} \Sigma_1^N \Sigma_1^T \Delta \hat{u}_{it} \Delta \hat{u}_{it-2}$.

b) En segundo lugar, se utiliza en la estimación un número de condiciones de ortogonalidad inferior a las disponibles, por lo que el modelo está sobreidentificado, es decir, existen más instrumentos disponibles que parámetros a estimar. Un contraste de dichas condiciones de sobreidentificación es un contraste de validez de la lista de instrumentos escogidos. El estadístico

$$(\Sigma_1^N \mathbf{Z}_i' \Delta \hat{\mathbf{u}}_i)' \mathbf{A}_N (\Sigma_1^N \mathbf{Z}_i' \Delta \hat{\mathbf{u}}_i)$$

donde $\mathbf{A}_N = [\Sigma_i (\mathbf{Z}_i' \Delta \hat{\mathbf{u}}_i) (\Delta \hat{\mathbf{u}}_i' \mathbf{Z}_i)]^{-1}$ se distribuye como una χ_r^2 , siendo r el número de condiciones de sobreidentificación, es decir, la diferencia entre el número de columnas en \mathbf{Z}_i y el número de coeficientes estimados. Un rechazo de la hipótesis nula sugeriría una inadecuada selección de instrumentos debida, por ejemplo, a una errónea caracterización de la situación de autocorrelación del modelo.

15.6.c. Modelos dinámicos con variables predeterminadas

Consideremos ahora un modelo autorregresivo que incorpora además un vector de variables explicativas \mathbf{x}_{it} :

$$y_{it} = \rho y_{it-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \alpha_i + u_{it} \tag{15.10}$$

donde suponemos que la correlación entre el vector \mathbf{x}_{it} y el efecto latente α_i es, en general, distinta de cero. Si u_{it} no presenta autocorrelación y si \mathbf{x}_{it} es predeterminada, es decir, $E(\mathbf{x}_{it-s} u_{it}) = \mathbf{0}_k$ sólo para $s \geq 0$, entonces el estimador de variables instrumentales óptimo, es decir, el estimador generalizado de momentos, es similar a [15.9]; en la estimación de la ecuación del período t en primeras diferencias son instrumentos válidos $\mathbf{z}_{it} = (y_{i1}, \dots, y_{it-2}, \mathbf{x}_{i1}, \dots, \mathbf{x}_{it-1})'$. Dicho estimador, que se obtiene en dos etapas, es:

$$(\hat{\rho}, \hat{\boldsymbol{\beta}})' = [(\Delta \mathbf{W} \mathbf{Z}) \mathbf{A}_N (\mathbf{Z}' \Delta \mathbf{W}')]^{-1} [(\Delta \mathbf{W} \mathbf{Z}) \mathbf{A}_N (\mathbf{Z}' \Delta \mathbf{y})] \tag{15.11}$$

donde: $\mathbf{w}'_{it} = (y_{it-1}, \mathbf{x}_{it})$, $\mathbf{W} = (\mathbf{w}_{12}, \dots, \mathbf{w}_{1T}, \dots, \mathbf{w}_{N2}, \dots, \mathbf{w}_{NT})$; $\mathbf{Z}_i = \text{diag}(z'_{it})$, donde los vectores \mathbf{z}_{it} son columnas a lo largo de la diagonal $\mathbf{Z}' = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_N)$, $\mathbf{y} = [y_{12}, \dots, y_{1T}, \dots, y_{N2}, \dots, y_{NT}]'$ y \mathbf{A}_N es la matriz antes introducida,

(16) La expresión analítica de dicha varianza (véase el apéndice de Arellano y Bond, 1991) es complicada.

siendo \hat{u}_i el vector columna de dimensión $(T-2) \times 1$ formado por los residuos de la primera etapa —es decir, ignorando la presencia de \mathbf{A}_N en [15.11]—, correspondientes al individuo i -ésimo, \mathbf{w}_{it} es un vector columna $(1+k) \times 1$; la matriz \mathbf{W} es $(1+k) \times N(T-1)$, por lo que $\Delta\mathbf{W}$ es $(1+k) \times N(T-2)$. El vector \mathbf{z}_i es $(T-1)k + (T-2) \times 1$, por lo que la matriz \mathbf{Z}_i es $(T-2) \times q$, donde $q = (1+k)T(T-1)/2 - (k+T-1)$, y \mathbf{Z} es $N(T-2) \times q$. Finalmente, la matriz \mathbf{A}_N es ahora de dimensión $q \times q$.

Si u_{it} presentase autocorrelación del tipo autorregresivo, se utilizarían las estimaciones de la primera etapa para estimar dicho modelo autorregresivo; a continuación, se filtra el modelo original [15.10] y se estima el modelo filtrado por el procedimiento generalizado de momentos en dos etapas, pudiendo contrastarse entonces las restricciones no lineales que como ocurre en la corrección de Cochrane-Orcutt (véase ecuación [7.3]) aparecen entre los parámetros del modelo.

Si u_{it} tiene estructura de autocorrelación MA(1), se puede obtener un estimador consistente similar al anterior, sin más que excluir de la lista de instrumentos el par $\mathbf{w}_{it-1} = (y_{it-2}, \mathbf{x}_{it-1})$, que ahora no serían instrumentos válidos. Si u_{it} es MA(q) con $q > T-3$, entonces el modelo no estaría identificado, al no haber suficientes instrumentos disponibles para proceder a su estimación; sólo si las variables \mathbf{x}_{it} fuesen no sólo predeterminadas, sino *exógenas*, podría identificarse el modelo, pues, en tal caso, las \mathbf{x}_{it} de todos los periodos serían instrumentos válidos.

15.7. IDENTIFICACION DE EFECTOS INDIVIDUALES EN EL ESTIMADOR INTRAGRUPOS

Hemos visto cómo al transformar el modelo para calcular el estimador intragrupos —al igual que ocurre cuando tomamos primeras diferencias—, tanto los efectos individuales no observables como los observables que no cambian con el tiempo α_i y $\mathbf{z}_i\gamma$ desaparecen del modelo. Se han propuesto algunas alternativas para la estimación de dichos efectos, una vez que se dispone de estimaciones de β y μ :

a) Basándose en la habitual fórmula de predicción mínimo-cuadrática, Lee y Griffith (1971) proponen utilizar:

$$\hat{\alpha}_i = \frac{\sigma_\alpha^2}{T\sigma_\alpha^2 + \sigma_u^2} \cdot \mathbf{1}'_T (y_i - \mathbf{X}_i \hat{\beta})$$

como estimador lineal e insesgado de los efectos latentes α_i . Este estimador puede interpretarse como una proporción del residuo MCG asignada a α_i , viniendo tal proporción dada por las varianzas relativas σ_α^2 y σ_u^2 .

Si no se rechazase la hipótesis nula $H_0: \alpha_i = 0, i = 1, 2, \dots, N$, que equivale a $H_0: \sigma_\alpha^2 = 0$, se tendría el modelo lineal general habitual. Esta hipótesis puede contrastarse mediante un test F que comparase las sumas residuales restringida SRR y sin restringir SRS del estimador intragrupos.

Como *alternativa* (propuesta por Breusch y Pagan, 1980), que requiere sólo los residuos restringidos, se estimaría una regresión de y sobre X , obteniéndose los residuos \tilde{u} . Entonces, bajo la hipótesis H_0 , se tendría el estadístico de los multiplicadores de Lagrange:

$$\lambda = \frac{TN}{2(T-1)} \left[\frac{\tilde{u}'(\mathbf{I}_N \otimes \mathbf{1}_T \mathbf{1}_T')\tilde{u}}{\tilde{u}'\tilde{u}} - 1 \right]^2 = \frac{TN}{2(N-1)} \left[\frac{\sum_{i=1}^N \left(\sum_{t=1}^T \tilde{u}_{it} \right)^2}{\tilde{u}'\tilde{u}} - 1 \right]^2$$

que se distribuye asintóticamente como una χ_1^2 .

b) Si algunas variables explicativas están incorrelacionadas con los efectos latentes α_i , no sólo puede ganarse eficiencia en la estimación, sino incluso identificar los coeficientes γ , siempre que podamos suponer que las variables z_i no están correlacionadas con los α_i . Una vez obtenido el estimador intra-grupos de β , podemos escribir:

$$\bar{y}_i - \bar{x}_i' \hat{\beta}_{IG} = z_i' \gamma + [\alpha_i + \bar{u}_i - \bar{x}_i' (\hat{\beta}_{IG} - \beta)]$$

Dada la ortogonalidad que suponemos entre los efectos individuales observables z_i y los no observables α_i , la estimación MCO de esta ecuación es consistente y nos permite evaluar el vector de coeficientes γ .

Si, como ocurrirá generalmente, no cabe sino admitir la existencia de correlaciones entre z_{it} y α_i , entonces la estimación consistente del vector γ requiere la utilización de un procedimiento de variables instrumentales, lo cual es sencillo sólo si se dispone de instrumentos externos incorrelacionados con α_i , pero correlacionados con z_{it} .

La situación mejora si hay al menos tantas variables x_{it} incorrelacionadas con α_i como variables z_i están correlacionadas con α_i ; Hausman y Taylor (1981) han propuesto una estrategia que permite identificar los efectos individuales en tal situación.

Apéndice 1: RELACIONES ENTRE LOS DISTINTOS ESTIMADORES

Como se pide demostrar en un problema al final de este capítulo, el estimador MCG del vector β puede escribirse:

$$\hat{\beta}_{MCG} = \Delta \hat{\beta}_{EG} + (\mathbf{I}_k - \Delta) \hat{\beta}_{IG}$$

donde Δ es la matriz:

$$\Delta = \psi T \left[\sum_1^N (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i)' (\mathbf{X}_i - \mathbf{1}_T \bar{x}_i) + \psi T \sum_1^T (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right]^{-1} \left[\sum_1^N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right]$$

mientras que el término constante $\hat{\mu}_{MCG} = \bar{y} - \bar{x}' \hat{\beta}_{MCG}$.

De acuerdo con esta expresión, el estimador MCG puede interpretarse

como una combinación lineal de los estimadores *intragrupos* y *entre grupos*. El parámetro ψ de la Sección 15.3 determina el tamaño de la matriz Δ de la combinación lineal, por lo que puede afirmarse:

- Proposición 15.1.** a) Para una dimensión temporal dada, el valor de ψ tiende a 1 según crece σ_u^2 en relación a σ_α^2 , y tiende a cero en caso contrario.
 b) El parámetro ψ tiende también a 0 cuando aumenta T , con independencia de los valores relativos de las dos varianzas σ_u^2 y σ_α^2 .

Demostración: Como ejercicio.

Proposición 15.2. a) Cuando ψ tiende a 1, el estimador de MCG se aproxima al estimador MCO.

b) Cuando ψ tiende a 0, el estimador MCG converge al estimador intragrupos.

Demostración. a) Basta hacer $\psi = 1$ en la expresión de V^{-1} de la Sección 15.3, que resulta entonces igual a la matriz identidad, y sustituirla en la expresión del estimador MCG en la Sección 15.3.

b) Cuando ψ se aproxima a 0, la matriz Δ también tiende a una matriz de ceros, lo que, llevado a la expresión del estimador MCG como combinación lineal de los otros dos estimadores, muestra su equivalencia con el estimador intragrupos.

Interpretemos estos resultados: En primer lugar, α_i proviene de suponer que el efecto latente es constante en el tiempo, si bien diferente entre individuos. Si la dimensión temporal es muy grande, es muy poco probable que la variable aleatoria latente fuese constante, salvo si fuese, en realidad, una constante diferente para cada individuo en el panel. Pero en tal caso, el estimador intragrupos sería eficiente, luego coincidiría con el estimador MCG.

En segundo lugar, al crecer el cociente $\sigma_u^2/\sigma_\alpha^2$, la componente estocástica de y_{it} viene dominada por u_{it} , mientras que la componente específica del individuo, α_i , pierde importancia. En relación a la aleatoriedad de u_{it} , la variable α_i es «casi» constante⁽¹⁷⁾. Pero si α_i fuese constante, entonces la matriz de covarianzas V sería escalar y, por tanto, el estimador MCG coincidiría con el estimador MCO.

En definitiva, el parámetro ψ puede interpretarse como el grado de relevancia que se otorga a la variación existente entre los distintos individuos incluidos en el panel. El estimador intragrupos (que es el límite del estimador MCG cuando ψ tiende a 0) ignora totalmente dicha variación, pues estima el vector de coeficientes de una sola vez, con todas las observaciones del modelo. El estimador MCO (que es el límite del estimador MCG cuando ψ tiende a 1) otorga a las variaciones entre individuos la misma relevancia que a las

⁽¹⁷⁾ De hecho, nótese que el estimador MCO se obtiene a partir del estimador MCG haciendo $\psi = 1$, lo cual es equivalente a $\sigma_\alpha^2 = 0$, es decir, α constante, en cuyo caso las distintas unidades muestrales serían inconfundibles entre sí.

variaciones entre las observaciones procedentes de un mismo individuo. En consecuencia, el estimador MCG, que trata el parámetro α_i como una variable aleatoria, adopta una posición intermedia en relación a la importancia otorgada a la variación entre individuos.

Apéndice 2: UTILIZACION PRACTICA DEL ESTIMADOR MCG

Los parámetros σ_u^2 y σ_α^2 son desconocidos, por lo que hay que utilizar el estimador MCG en dos etapas, dedicando la primera de ellas a obtener estimaciones de estos parámetros. Promediando [15.1] para las observaciones de cada individuo y restando posteriormente de [15.1], se obtienen las dos ecuaciones:

$$\begin{aligned} \bar{y}_i &= \mu + \bar{x}_i' \beta + \alpha_i + \bar{u}_i && (N \text{ observaciones}) \\ y_{it} - \bar{y}_i &= (x_{it} - \bar{x}_i)' \beta + (u_{it} - \bar{u}_i) && (NT \text{ observaciones}) \end{aligned}$$

En la segunda ecuación aparecen las variables en diferencias respecto a las medias de cada individuo, lo que recuerda al estimador intragrupos, y sugiere el estimador de σ_u^2 :

$$\hat{\sigma}_u^2 = \frac{\sum_1^N \sum_1^T [(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)' \hat{\beta}_{IG}]^2}{N(T-1) - (k-1)}$$

mientras que en la primera ecuación aparecen tan sólo las medias muestrales, lo que recuerda al estimador entre grupos, y sugiere el siguiente estimador de la suma $\sigma_\alpha^2 + \sigma_u^2/T$:

$$\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}_u^2}{T} = \frac{\sum_1^N (\bar{y}_i - \hat{\mu}_{EG} - \bar{x}_i' \hat{\beta}_{EG})^2}{N - k}$$

La estimación de σ_u^2 se obtiene directamente de la primera ecuación, mientras que la estimación de σ_α^2 surge de la segunda ecuación, una vez que hemos estimado σ_u^2 . La estimación de σ_α^2 podría resultar negativa, lo que para algunos autores sería una indicación de mala especificación del modelo; Maddala (1971) sugiere considerar en tal caso la posible omisión de efectos temporales relevantes.

Finalmente, Hausman y Taylor (1981) y Hsiao (1986) muestran que existe un sencillo procedimiento para computar el estimador MCG, observando que la matriz

$$P = \left[I_T - (1 - \sqrt{\psi}) \frac{1}{T} \mathbf{1}_T \mathbf{1}_T' \right]$$

donde ψ , el parámetro definido en la Sección 15.3, es tal que $\mathbf{P}\mathbf{P}' = \mathbf{V}^{-1}$. Por ello, premultiplicando por \mathbf{P} las matrices que contienen las observaciones muestrales, y estimando por MCO, se obtiene el estimador MCG del modelo original del modo habitual, es decir, utilizando MCO en el modelo con las variables transformadas⁽¹⁸⁾.

En realidad, tampoco es preciso llevar a cabo los productos matriciales citados, puesto que equivalen a sustraer de cada observación y_{it} el término $(1 - \sqrt{\psi})\bar{y}_i$, y de cada vector de observaciones \mathbf{x}_{it} el término $(1 - \sqrt{\psi})\bar{\mathbf{x}}_i$. La variable que acompaña al término independiente se sustituye simplemente por $(1 - \sqrt{\psi})y_{it}$.

PROBLEMAS

Problema 15.1. Con objeto de prever el margen de beneficios de un conjunto de empresas productoras de un mismo bien, un investigador propone el siguiente modelo:

$$(1) \quad \begin{aligned} Y_{it} &= \beta_1 Y_{it-1} + \beta_2 Y_{it-2} + \varepsilon_{it} \\ \varepsilon_{it} &= u_{it} + \alpha_i; \quad i = 1 \dots N, \quad t = 1 \dots T \\ E(u_{it}) &= E(\alpha_i) = 0, \quad \forall i, t; \quad E(u_{it}\alpha_j) = 0, \quad \forall i, t, j \\ E(u_{it}u_{js}) &= \sigma_u^2 \quad \text{si } i = j \text{ y } t = s \\ &= 0 \quad \text{en otro caso} \\ E(\alpha_i\alpha_j) &= \sigma_\alpha^2 \quad \text{si } i = j \\ &= 0 \quad \text{si } i \neq j \end{aligned}$$

donde Y_{it} es el margen de la empresa i en el período t , definido como:

$$\frac{V_{it} - W_{it}}{V_{it}}$$

V_{it} : Valor de la producción de la empresa i en el período t .

W_{it} : Costes variables de la empresa i en el período t .

Se pide:

- Demostrar por qué MCO no sería un método de estimación adecuado en este caso. ¿En qué caso el estimador intragrupos sería adecuado?
- Qué ventajas e inconvenientes plantea el aplicar el operador diferencias al modelo (1).
- Proponga un estimador consistente y eficiente de β_1 y β_2 en el caso de $T = 4$ años.
- ¿Cambiaría su respuesta en c), en el caso de $T = 5$ años?

⁽¹⁸⁾ Nótese asimismo la similitud entre las matrices \mathbf{P} y \mathbf{Q} que utilizamos con frecuencia.

Problema 15.2. Para tratar de analizar la política de financiación de las empresas industriales españolas, un investigador especifica el modelo:

$$(1) \quad y_{it} = \beta_{1i} + \beta_{2i}x_{it} + \beta_{3i}z_i + \beta_4 R_t + \varepsilon_{it}, \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T$$

donde:

- y_{it} = porcentaje de fondos propios sobre el pasivo neto de la empresa, representando la descomposición entre financiación mediante fondos propios y ajenos;
- x_{it} = gastos financieros/recursos ajenos, como indicador del coste de captación de capital ajeno a la empresa;

mientras que z_i es un efecto específico de cada empresa, constante en el tiempo, y R_t , un efecto propio de cada período, común para todas las empresas. Tanto z_i como R_t son observables.

Además, $E\varepsilon_{it} = 0, \quad \forall i, t$

$$E\varepsilon_{it}\varepsilon_{js} = \begin{cases} \sigma_\varepsilon^2 & \text{si } i = j \text{ y } t = s \\ 0 & \text{si } i \neq j \text{ o } t \neq s \end{cases}$$

a) Formule el modelo en notación matricial compacta bajo los dos supuestos alternativos:

- i) $\beta_{1i} = \beta_1$, constante; $\beta_{2i} = \beta_2$, constante;
- ii) $\left. \begin{matrix} \beta_{1i} = \beta_1 + \mu_i \\ \beta_{2i} = \beta_2 + \mu_i \end{matrix} \right\} \mu_i \sim N(0, \sigma_\mu^2)$,

siendo μ_i independiente de ε_{it} y serialmente incorrelacionado.

b) Discuta el procedimiento de estimación más adecuado bajo cada uno de los supuestos alternativos i) e ii), analizando sus ventajas e inconvenientes, y proponga expresiones analíticas concretas para el cálculo de dichas estimaciones para el modelo (1).

c) Formule el modelo «entre grupos» correspondiente, e indique su utilidad práctica.

Problema 15.3. Suponga que en el modelo de efectos latentes

$$y_{it} = \mu + \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T$$

el efecto no observable puede representarse para cada individuo:

$$\alpha_i = \bar{\mathbf{x}}'_i \mathbf{a} + \omega_i, \quad i = 1, 2, \dots, N$$

donde $\omega_i \sim N(0, \sigma_\omega^2)$. Las variables \mathbf{x}_{it} son exógenas.

a) Obtenga la representación de y_{it} como función de una constante, \mathbf{x}_{it} , $\bar{\mathbf{x}}_i$, y un término de error compuesto:

$$\mathbf{v}_i = \mathbf{u}_i + \mathbf{1}_T \omega_i, \quad i = 1, 2, \dots, N$$

b) Pruebe que la inversa de la matriz de covarianzas de v_t puede escribirse:

$$(\mathbf{V}^*)^{-1} = \frac{1}{\sigma_u^2} \cdot \left[\mathbf{Q} + \psi^* \frac{1}{T} \mathbf{1}_T \mathbf{1}'_T \right] \quad \text{donde} \quad \psi^* = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$$

c) Demuestre que:

$$\hat{\mu}_{\text{MCG}} = \bar{y} - \bar{x}' \hat{\beta}_{\text{EG}}$$

$$\hat{\beta}_{\text{MCG}} = \hat{\beta}_{\text{IG}}$$

$$\hat{\alpha}_{\text{MCG}} = \hat{\beta}_{\text{EG}} - \hat{\beta}_{\text{IG}}$$

Problema 15.4. Dado el modelo de datos de panel:

$$Y_{it} = \alpha_0 + \alpha_1 X_{1it} + \alpha_2 X_{2it} + \alpha_3 X_{3i} + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T$$

donde u_{it} se supone independiente en el tiempo, con distribución $N(0, \sigma_u^2)$, para todo i, t . La variable X_{3i} varía con los individuos en la muestra, pero es constante en el tiempo. Dicha variable es, además, *no observable*.

1. Dado que X_{3i} no es observable, un investigador estima por MCO una regresión de Y_{it} sobre X_{1it} y X_{2it} , junto con una constante. ¿Cuál será la magnitud del sesgo producido en dicha estimación mínimo-cuadrática del coeficiente α_2 ? ¿Qué información sería precisa para determinar el signo de dicho sesgo? ¿Sería suficiente conocer el signo de la correlación entre las variables X_{2it} y X_{3i} ?

2. Si el investigador estuviese dispuesto a suponer que la variable X_{3i} , no observable, *no es estocástica*, ¿qué procedimiento de estimación *consistente* y *eficiente* de los coeficientes α_1 y α_2 debería utilizar? Obtenga asimismo las expresiones algebraicas para las varianzas de dichos estimadores.

3. Supongamos ahora que, por el contrario, el investigador cree que debe modelizar X_{3i} (recuerde que no es observable) como variable *aleatoria*, aunque *incorrelacionada* con X_{1it} y X_{2it} . Supongamos asimismo que el término de error del modelo satisface las propiedades:

$$E(u_{it}^2) = \sigma_u^2, \quad E(u_{it}u_{jt}) = \sigma_{ij}, \quad i, j = 1, 2, \dots, N; \quad t = 1, 2, \dots, T$$

Derive la matriz de varianzas y covarianzas del vector

$$\mathbf{u} = \{(u_{it}), \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T\}$$

y discuta el modo de obtener un estimador *consistente* y *eficiente* de los coeficientes α_1 y α_2 .

¿Observa alguna posible dificultad en el cálculo numérico de este estimador?

¿Qué dificultades adicionales introduciría la existencia de correlación entre X_{3i} y las variables X_{1it} , X_{2it} ?

Problema 15.5. a) Probar:

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} \mathbf{1}_T \mathbf{1}'_T \right)$$

b) Introduciendo la expresión anterior en el sistema de ecuaciones normales del estimador MCG:

$$[\Sigma_1^m \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{W}_i] \boldsymbol{\beta} = [\Sigma_1^m \mathbf{W}_i' \mathbf{V}^{-1} \mathbf{y}_i]$$

obtener la expresión dada en el texto para dicho sistema de ecuaciones normales (Sección 15.3).

Problema 15.6. Efectúe el análisis de la varianza del modelo con efectos latentes individuales. Para ello, debe hallar los errores cuadráticos medios entre grupos e intragrupos.

Problema 15.7. Obtenga la expresión analítica del estimador $\hat{\boldsymbol{\beta}}_{\text{MCG}}$ de Balestra-Nerlove que aparece en la Sección 15.3.

a) Pruebe que, como se afirma en el Apéndice 2, la inversa de la matriz de covarianzas \mathbf{V} puede descomponerse: $\sigma_u^2 \mathbf{V}^{-1} = \mathbf{P}' \mathbf{P}$, donde $\mathbf{P} = \mathbf{I}_T - (1 - \sqrt{\psi}) \frac{1}{T} \mathbf{1}_T \mathbf{1}_T'$.

b) Pruebe asimismo que ello sugiere transformar las variables del modelo mediante:

$$y_{it}^* = y_{it} - (1 - \sqrt{\psi}) \bar{y}_i; \quad \mathbf{x}_{it}^* = \mathbf{x}_{it} - (1 - \sqrt{\psi}) \bar{\mathbf{x}}_i$$

Problema 15.8. Demostrar que el estimador MCO del vector $\boldsymbol{\beta}$ en el modelo:

$$y_{it} = \mu + \alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, N; \quad t = 1, 2, \dots, T$$

bajo las restricciones $\sum_1^N \alpha_i = 0$, es equivalente al estimador MCO de $\boldsymbol{\beta}$ en el modelo:

$$y_{it} = \gamma_i + \mathbf{x}_{it}' \boldsymbol{\beta} + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T$$

donde los coeficientes γ_i no están sujetos a ninguna restricción.

Explicar la relación que habría entre estos dos modelos y una tercera formulación:

$$y_{it} = \mu + \sum_{j=2}^k \alpha_j D_j + \mathbf{x}_{it}' \boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, N; \quad t = 1, 2, \dots, T$$

donde las D_j son variables ficticias para las unidades muestrales.

Problema 15.9. Probar que si la especificación correcta para la variable latente α_i como función del vector \mathbf{x}_i es la propuesta en el Problema 15.3, entonces:

$$E(\hat{\boldsymbol{\beta}}_{\text{EG}}) = \boldsymbol{\beta} + \mathbf{a}, \quad \text{mientras que} \quad E(\hat{\boldsymbol{\beta}}_{\text{IG}}) = \boldsymbol{\beta}$$

Problema 15.10. Obtenga la matriz de covarianzas del estimador de máxima verosimilitud del modelo lineal de regresión con efectos latentes no observables, específicos de cada unidad muestral.

Problema 15.11. Demuestre que el estimador Balestra-Nerlove, $\hat{\boldsymbol{\beta}}_{\text{MCG}}$, puede escribirse como combinación lineal de los estimadores intragrupos y entre grupos:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = (\mathbf{I}_k - \Delta) \hat{\boldsymbol{\beta}}_{\text{EG}} + \Delta \hat{\boldsymbol{\beta}}_{\text{IG}}$$

donde Δ es la matriz $k \times k$ que se menciona en el Apéndice 1.

Problema 15.12. Demuestre que en el caso de un modelo lineal que además de variables latentes individuales α_i contiene efectos individuales observables z_i :

$$y_{it} = \mu + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_i\boldsymbol{\gamma} + \alpha_i + u_{it}$$

se tiene:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{MCG} \\ \hat{\boldsymbol{\gamma}}_{MCG} \end{pmatrix} = \Delta \begin{pmatrix} \hat{\boldsymbol{\beta}}_{EG} \\ \hat{\boldsymbol{\gamma}}_{EG} \end{pmatrix} + (\mathbf{I} - \Delta) \begin{pmatrix} \hat{\boldsymbol{\beta}}_{IG} \\ \mathbf{0} \end{pmatrix}$$

Problema 15.13. Dado el modelo dinámico:

$$W_{it} = \beta_1 W_{it-1} + \beta_2 E_{it} + \alpha_i + \varepsilon_{it}, \quad t = 1, 2, 3, \dots, \quad i = 1, 2, \dots, 100$$

donde:

W_{it} = Salario del individuo i en el período t

E_{it} = Años de escolarización del individuo i en el período t

α_i = Habilidad del individuo i

ε_{it} = Variable aleatoria con los supuestos habituales

y utilizando la información muestral siguiente:

$$\sum_{i=1}^{100} \Delta W_{i2} W_{i1} = 10$$

$$\sum_{i=1}^{100} \Delta W_{i2} \Delta E_{i3} = 5$$

$$\sum_{i=1}^{100} \Delta E_{i3} W_{i1} = 10$$

$$\sum_{i=1}^{100} (\Delta E_{i3})^2 = 10$$

$$\sum_{i=1}^{100} (W_{i1})^2 = 20$$

$$\sum_{i=1}^{100} \Delta W_{i3} W_{i1} = 10$$

$$\sum_{i=1}^{100} \Delta E_{i3} \Delta W_{i3} = 2$$

obtenga estimaciones consistentes y eficientes de β_1 y β_2 . Calcule también la matriz de covarianzas de dichas estimaciones.

CAPITULO 16

VARIABLES DEPENDIENTES CUALITATIVAS Y LIMITADAS

Parte I

MODELOS DE ELECCION DISCRETA

16.1. INTRODUCCION

Aunque el rango de valores posibles de una variable económica es usualmente un intervalo de la recta real, sin embargo la teoría econométrica considera también modelos de regresión lineal en los que alguna de las variables explicativas toma valores en un conjunto discreto y finito. Un ejemplo de este tipo de variables lo constituye las llamadas variables ficticias que introdujimos en el Capítulo 4. Como allí se vio, dichas variables se han utilizado tradicionalmente para incorporar al modelo posibles variaciones estructurales ocurridas durante el período muestral, o efectos socioeconómicos que pudieran distinguir el comportamiento de unos individuos de otros.

En estos casos se introduce en el modelo una variable discreta para cada una de las características que se pretenden tomar en consideración. A cada una de las posibles modalidades que puede presentar la característica se le asocia un valor numérico, y la variable ficticia así definida se utiliza en la estimación del modelo como una variable explicativa más. Si se pretende recoger en el modelo la idea de un posible cambio estructural en el instante t_0 , entonces la variable ficticia correspondiente tomará el valor 0 para $1 \leq t < t_0$ y el valor 1 para $t_0 \leq t \leq T$. Si se pretendiese discriminar entre los gastos en educación de las familias de una muestra dependiendo de su pertenencia a un medio rural o urbano, podría definirse una variable ficticia que toma el valor 0 si la familia vive en un medio rural, y el valor 1 si vive en un medio urbano.

Existen otras variables discretas que podrían ser influyentes en el comportamiento de la variable endógena sin ser variables ficticias. Por ejemplo, el número de hijos es importante al analizar el gasto en educación de una determinada familia. Conviene asimismo recordar que variables continuas

como la edad o los ingresos de un individuo se transforman en ocasiones en variables discretas debido al diseño de la muestra utilizada en la recogida de datos.

Sin embargo, son muchos los problemas y cuestiones de interés en economía en los que es la variable endógena la que no toma en la muestra todos los valores de un intervalo real, sino sólo un número finito de ellos. A veces, esta variable ni siquiera es cuantificable, como ocurre cuando se pretende caracterizar el tipo de transporte que cada persona de un determinado colectivo toma diariamente cuando acude a su trabajo. En tal caso, si se consideran como alternativas: coche particular o transporte público, puede asignarse a la variable endógena valores 0 y 1, según que el medio utilizado sea privado o público, pero tal asignación es arbitraria. Este capítulo trata de los métodos apropiados de estimación en aquellos modelos econométricos en que la variable dependiente tiene esta naturaleza cualitativa.

Además de la elección del medio de transporte, existen otras situaciones que generan problemas de este tipo y a los que nos referiremos en lo sucesivo: *a)* la decisión de tomar parte en el mercado de trabajo frente a otra alternativa, como puede ser el realizar trabajo doméstico, *b)* el nivel de estudios de un individuo como función de sus características personales, *c)* el voto a depositar en una determinada elección política, *d)* la clase de vivienda a ocupar (alquiler frente a compra), *e)* el tipo de escuela al que enviar los hijos, si privada o pública, *f)* si un crédito concedido por una entidad bancaria será devuelto en la fecha de su vencimiento o no, etc.

El caso más frecuente de variables endógenas discretas surge, por tanto, cuando el investigador pretende utilizar un modelo econométrico para explicar la decisión tomada por un agente económico utilizando para ello un vector de características de dicho individuo. Cuando, como en los ejemplos citados, la decisión tomada es una variable de naturaleza cualitativa, pasa a representarse mediante una variable cuantitativa que toma un valor diferente para cada una de las posibles opciones dentro del conjunto de elección.

Debe resultar claro que esta situación surge únicamente cuando se dispone de datos recogidos al nivel en que se toman las decisiones objeto de estudio, ya sean individuos o familias, y no de variables macroeconómicas, puesto que la agregación de variables Y_i de tipo discreto para un número de individuos evitaría gran parte de las dificultades que aquí se mencionan.

16.2. MODELOS DE ELECCION BINARIA

Vamos a comenzar considerando el caso en que se pretende explicar la elección de una entre dos alternativas posibles. En esta situación, la variable dependiente puede tomar dos valores: $Y_i = \{0, 1\}$, según que el individuo escoja la primera o la segunda alternativas, y se pretende explicar la elección hecha por el decisor como función de unas variables que le caracterizan y que denotamos por x_i , un vector de dimensión k , añadiendo un término de error que explique las diferencias entre los valores observados de Y_i (cero o uno) y sus valores previstos.

Por ejemplo, en el caso de la elección de viviendas, podría definirse $Y_i = 1$ si la familia es propietaria de la vivienda que habita, e $Y_i = 0$ si dicha vivienda se ocupa en régimen de alquiler, y especificar un modelo:

$$Y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i, \quad i = 1, 2, \dots, N \quad [16.1]$$

Hay algún problema con una formulación de este tipo: Como Y_i sólo toma el valor 0 ó 1 para cada individuo o familia en la muestra, entonces, para cada observación \mathbf{x}_i , la perturbación u_i debe ser una variable aleatoria que solamente puede tomar valores $-\mathbf{x}'_i \boldsymbol{\beta}$ y $1 - \mathbf{x}'_i \boldsymbol{\beta}$. En el primer caso se tendría $Y_i = 0$ y en el segundo $Y_i = 1$. Pero, además, para que $E u_i = 0$, las probabilidades con que u_i debe tomar estos dos valores han de ser $1 - \mathbf{x}'_i \boldsymbol{\beta}$ y $\mathbf{x}'_i \boldsymbol{\beta}$, respectivamente. Esto crea, a su vez, varias dificultades:

1. A partir de la distribución de probabilidad anterior se tiene $\text{Var}(u_i) = (\mathbf{x}'_i \boldsymbol{\beta})^2 (1 - \mathbf{x}'_i \boldsymbol{\beta}) + (1 - \mathbf{x}'_i \boldsymbol{\beta})^2 (\mathbf{x}'_i \boldsymbol{\beta}) = (\mathbf{x}'_i \boldsymbol{\beta})(1 - \mathbf{x}'_i \boldsymbol{\beta})$ y, por tanto, el término de error tiene heteroscedasticidad. En particular, el estimador de mínimos cuadrados ordinarios será ineficiente en esta situación.

2. Los contrastes de significación usuales, que se basan en el supuesto de distribución Normal del término de error, no son aplicables ya que, como hemos visto, u_i tiene una distribución de probabilidad de tipo discreto, con tan sólo dos valores posibles. Las desviaciones típicas obtenidas para las estimaciones MCO mediante las expresiones discutidas en el Capítulo 4, son sesgadas y el cociente de determinación R^2 no es representativo de la calidad del modelo.

3. Debido a que el término de error sigue una distribución de probabilidad que no es Normal, métodos de estimación lineales como el de mínimos cuadrados (ordinarios o generalizados) pueden ser mejorados, en términos de eficiencia, por métodos no lineales.

4. Predecir es, como sabemos, uno de los usos fundamentales del modelo econométrico. En este caso, podríamos utilizar los valores del vector \mathbf{x}_i correspondientes a un individuo no incluido en la muestra, para predecir la decisión que tomaría en una determinada elección. Sin embargo, una vez estimado el vector $\boldsymbol{\beta}$, el valor implicado para dicha decisión: $\mathbf{x}'_i \boldsymbol{\beta}$, no será igual a 0 ó 1, y ello genera problemas de interpretación.

5. Algo se resuelve al observar que $P(Y_i = 1) = P[u_i = 1 - \mathbf{x}'_i \boldsymbol{\beta}] = \mathbf{x}'_i \boldsymbol{\beta}$, de modo que el valor previsto $\mathbf{x}'_i \boldsymbol{\beta}$ puede interpretarse como la *probabilidad* de que el individuo de características elija la opción $Y_i = 1$, con independencia de lo que le hayamos visto elegir en la muestra, si es que está incluido en ella.

16.3. EL MODELO LINEAL DE PROBABILIDAD

En el modelo anterior estaríamos tratando de explicar los valores tomados por una variable discreta Y_i mediante un vector \mathbf{x}_i formado por k variables que, en general, tendrán un soporte continuo. La implicación fundamental que de ello se deriva es que el modelo estimado generará valores de Y_i en un

cierto rango continuo y, por ello, fuera del rango de valores admisibles. Como vamos a ver, el problema se alivia bastante cuando la muestra consiste de observaciones repetidas hechas sobre un conjunto de individuos relativamente pequeño en número.

16.3.a. Observaciones repetidas

Por ejemplo, supongamos que para caracterizar los determinantes del tipo de transporte (público, privado) que se utiliza para ir al trabajo se dispone de observaciones diarias a lo largo de un mes para un conjunto de 100 personas. A cada individuo entrevistado se le pregunta su estado civil, la distancia desde su hogar a su lugar de trabajo, su nivel de ingresos y el tamaño de la ciudad en que vive. Se define la variable:

$$Y_i = \begin{cases} 1 & \text{si se utiliza el transporte público} \\ 0 & \text{si se utiliza el transporte privado} \end{cases}$$

El tamaño muestral es en este caso igual a 100 multiplicado por el número de días laborables de este mes (supongamos que son 20) y se trataría de explicar el valor tomado por Y_i como función de los valores tomados por las cuatro características citadas, junto con algunas características de los medios de transporte considerados (por ejemplo, el coste por kilómetro para el usuario).

Las veinte observaciones (el número de días laborables) correspondientes a un mismo individuo tendrán todas ellas igual vector de características. Sin embargo, el individuo habrá utilizado unos días el transporte público y otras veces el transporte privado. Si se calcula la frecuencia p_i (en términos porcentuales) con que dicho individuo ha utilizado el transporte público, se tendría una observación formada por dicha frecuencia muestral y el vector de características \mathbf{x}_i . La variable a explicar sería ahora p_i , la frecuencia observada con que el individuo con características \mathbf{x}_i escoge la opción $Y_i = 1$. Esta variable es ya continua, si bien todavía acotada.

Consideremos otra situación en que se quiere caracterizar la oferta de trabajo de la mujer casada en España, para lo que se recoge una muestra de 5.000 mujeres casadas y se define una variable Y_i que toma el valor 1 (por ejemplo) si la persona i -ésima en la muestra trabaja o busca trabajo, y el valor 0 si no toma parte activa en el mercado de trabajo. A continuación, las mujeres encuestadas se agrupan en clases dependiendo de los valores de un conjunto de variables discretas que representan su situación socioeconómica: número de hijos (uno, dos, tres o más), su nivel de renta (alto, medio, bajo), medio en el que se vive (rural, urbano), nivel de estudios (cuatro categorías: inferior a elemental, elemental, medio y superior).

En este caso habríamos clasificado las 5.000 observaciones correspondientes a las mujeres encuestadas en tan sólo 72 clases, por lo que habría un elevado número de observaciones en cada clase, todas aquellas mujeres con igual vector de características. Nótese que en este caso no hay observaciones

repetidas de los mismos individuos, pero aquellos con un mismo vector de características son indistinguibles a efectos del modelo.

De este modo, cada una de las 72 clases contendrá n_i mujeres, todas con igual vector x_i , de modo que $\sum_1^{72} n_i = 5.000$. Calcularíamos las frecuencias con que las mujeres casadas pertenecientes a una misma clase participan en el mercado de trabajo. Hecho este cálculo para cada una de las 72 clases en que se han dividido las mujeres entrevistadas nos dispondríamos a estimar el modelo con las 72 observaciones formadas por el par (frecuencia observada de Y_i ; vector de características x_i).

Una vez que dispongamos de estimaciones para los parámetros del modelo, la agrupación de los individuos en clases permite interpretar los valores \hat{Y}_i generados por la variable endógena como el porcentaje de mujeres casadas con vector de características igual a x_i que participan activamente o, equivalentemente, como la probabilidad de que una persona de dicho grupo decida participar en el mercado de trabajo. El vector estimado $\hat{\beta}$ podría utilizarse para calcular «pseudoeelasticidades».

Es importante observar que la distribución de las observaciones muestrales en clases puede también llevarse a cabo incluso cuando las variables del modelo son de tipo continuo. Ello implica cierta pérdida de información a cambio de una mayor sencillez en el proceso de estimación. Así, en el ejemplo de la elección del medio de transporte, incluso si los individuos hubiesen proporcionado valores numéricos precisos para la distancia a su centro de trabajo o sus ingresos, podríamos clasificarlos de acuerdo con distancias inferiores a 3 km, entre 3 y 4 km, o superiores a 5; de acuerdo con sus ingresos, en grupos de renta baja, media, alta y de acuerdo con el tamaño de la ciudad, según que ésta tenga menos de 250.000 habitantes, entre 250.000 y 1.000.000 de habitantes, o más de 1.000.000. Una vez elaborada esta clasificación, podríamos calcular las frecuencias observadas para la variable endógena Y_i dentro de cada clase, $p_i = n_i^{-1} \sum_{j \in I_i} Y_j$, y proceder a la estimación del modelo. En esta notación, I_i es el conjunto de índices (de 1 a n_i) de los individuos que pertenecen a la clase i , es decir, que tienen vector de características x_i .

El modelo lineal de probabilidad postula que la probabilidad teórica de que un individuo de la clase i escoja la opción $Y_i = 1$ viene dada por una función lineal del vector x_i : $P_i = x_i' \beta$. La frecuencia p_i observada en la muestra no es sino una aproximación a P_i con un error muestral u_i :

$$p_i = P_i + u_i = x_i' \beta + u_i, \quad i = 1, 2, \dots, M$$

Notemos además que, por la peculiaridad de ser Y_i una variable con valores cero o uno, P_i puede también interpretarse como el valor esperado de la variable Y_i , puesto que:

$$E(Y_i/x_i) = 1 \cdot P[Y_i = 1/x_i] + 0 \cdot P[Y_i = 0/x_i] = P[Y_i = 1/x_i] = P_i$$

propiedad que utilizaremos repetidamente.

16.3.b. Estimación por mínimos cuadrados generalizados

Supongamos que se dispone de una muestra de N observaciones clasificadas en M clases, en cada una de las cuales hay n_i observaciones, con $\sum_{i=1}^M n_i = N$.

Dadas las frecuencias muestrales $p_i = y_i/n_i$, donde y_i es el número de individuos que escogieron la opción $Y_i = 1$, tenemos el *modelo lineal de probabilidad*:

$$p_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i$$

Obsérvese que la probabilidad poblacional, P_i , aunque desconocida, es determinista, mientras que u_i representa el error muestral que hace que el valor de la proporción muestral observada p_i , que es una variable aleatoria, difiera de P_i . Cada coeficiente estimado proporcionaría una indicación del efecto que la variable explicativa correspondiente tiene sobre la probabilidad de escoger la acción $Y_i = 1$.

Al haber agrupado observaciones, la distribución del término de error del modelo cambia. Dicho término es ahora el error en la proporción de individuos que eligieron la decisión $Y_i = 1$, y no en la decisión tomada por un solo individuo. Como cada decisión es una variable aleatoria binomial, $Y_i \sim B(P_i)$, entonces $\sum_1^{n_i} Y_i \sim B(n_i, P_i)$ y $p_i = n_i^{-1} \sum_1^{n_i} Y_i$ se distribuye como $n_i^{-1} B(n_i, P_i)$, donde B denota la distribución binomial, lo que implica:

$$E(p_i) = n_i \frac{1}{n_i} P_i = P_i \quad \text{y} \quad \text{Var}(p_i) = n_i^{-1} P_i (1 - P_i)$$

y, consecuentemente, como $u_i = p_i - P_i$ se tiene:

$$Eu_i = 0 \quad \text{y} \quad \text{Var}(u_i) = \text{Var}(p_i) = n_i^{-1} P_i (1 - P_i)$$

Como es habitual en modelos con heteroscedasticidad, la varianza del término de error es desconocida y ha de ser estimada (mediante el producto $\hat{P}_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}}$) antes de que podamos utilizar el método de mínimos cuadrados generalizados. La matriz $\boldsymbol{\Sigma}$ tiene en la diagonal estas estimaciones de las $\text{Var}(u_i)$.

El estimador de mínimos cuadrados generalizados es:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{p}$$

donde cada fila de la matriz \mathbf{X} (de dimensión $M \times k$) es el vector de características \mathbf{x}_i' de una clase, $\boldsymbol{\Sigma}$ la matriz de covarianzas del vector \mathbf{u} (de dimensión $M \times M$) y \mathbf{p} el vector de las M frecuencias muestrales observadas, $p_i = \frac{y_i}{n_i}$. Es natural suponer que las perturbaciones de distintas clases son independientes, lo que convierte la matriz $\boldsymbol{\Sigma}$ en diagonal.

Con:

$$\text{Var}(\hat{\beta}_{\text{MCG}}) = \sigma^2 \mathbf{X}'\Sigma^{-1}\mathbf{X}$$

donde σ^2 se estimaría dividiendo la suma residual habitual por $M - k$.

Para ello se ha de estimar en una primera etapa por mínimos cuadrados ordinarios, ignorando la heteroscedasticidad:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{p}$$

donde \mathbf{p} es el vector formado por las M proporciones de los grupos de individuos. A continuación, se utilizarían los estimadores resultantes para «predecir» las probabilidades $\hat{P}_i = \mathbf{x}'_i\hat{\beta}$, y estimaríamos las varianzas de las u_i por medio de $\text{Var}(u_i) = \frac{\hat{P}_i(1 - \hat{P}_i)}{n_i}$, para dividir cada observación muestral (y_i, \mathbf{x}_i) por la raíz cuadrada de dicha varianza estimada. Entonces se vuelve a utilizar mínimos cuadrados ordinarios en el modelo corregido, lo que proporciona el estimador de mínimos cuadrados generalizados.

Obsérvese que en la estimación de la matriz de covarianzas no se utilizan las frecuencias observadas en la muestra, sino las previstas a partir de una estimación mínimo cuadrática previa. Ejercicios de simulación aconsejan el segundo modo de proceder, aunque en el caso en que no se dispone de observaciones repetidas de cada clase, éste sería el único procedimiento posible.

A pesar de ser un estimador consistente de la varianza de u_i , la estimación de la varianza condicional $E(u_i^2/\mathbf{x}_i)$ podría resultar negativa si $\mathbf{x}'_i\hat{\beta}$ es negativa o superior a uno, lo que estaría en contradicción con la propia especificación del modelo y no podríamos llevar a cabo la transformación de heteroscedasticidad que propusimos antes. Quizá lo mejor sea ignorar tales observaciones si se dispone de un tamaño muestral suficientemente grande.

El modelo lineal de probabilidad resuelve el problema de que la variable endógena pueda tomar cualquier valor en el intervalo $[0, 1]$, y no solamente los dos extremos de dicho intervalo. Precisamente por esto, la hipótesis de Normalidad del término de error, si no totalmente creíble, al menos comienza a ser más aceptable. Sin embargo, con el modelo lineal de probabilidad aún podría ocurrir que, al efectuar un ejercicio de predicción para un determinado vector de características \mathbf{x}^0 , la probabilidad obtenida, $\hat{P} = \mathbf{x}^0'\hat{\beta}$, caiga fuera del intervalo $[0, 1]$. Este problema suele evitarse truncando los valores posibles de P_i del modo que refleja la Figura 16.1. Si esto ocurre con una parte importante de la muestra, puede tener sentido atenerse exclusivamente al estimador de mínimos cuadrados ordinarios. En tal caso, hay que tener presente que, al igual que en todo modelo con heteroscedasticidad, la matriz de covarianzas de dicho estimador es $\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, donde Σ es la matriz antes descrita y que sería incorrecto utilizar $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, donde σ^2 se habría estimado como:

$$\frac{(\mathbf{p} - \mathbf{X}\hat{\beta})'(\mathbf{p} - \mathbf{X}\hat{\beta})}{M - k}$$

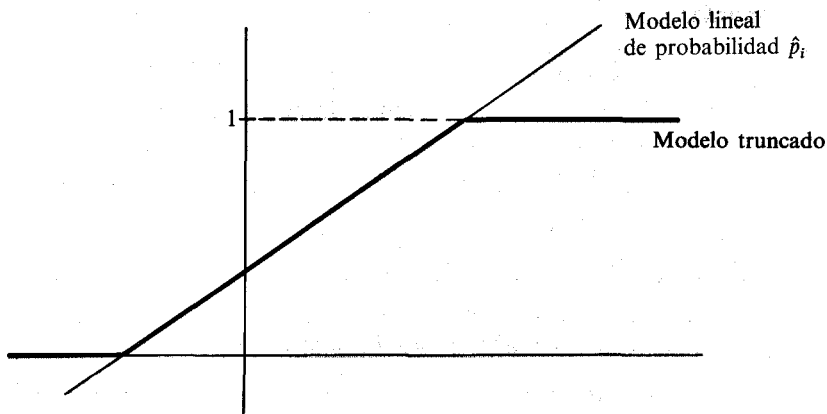


FIGURA 16.1.

Por su sencillez, es recomendable estimar este modelo siempre que se disponga de observaciones repetidas, siquiera sea para mantenerlo como base de comparación frente a posibles alternativas de mayor sofisticación estadística que pudieran utilizarse con la muestra disponible.

16.4. LAS DECISIONES DE LOS INDIVIDUOS POR MEDIO DE INDICADORES

Con el modelo de probabilidad lineal se trata de explicar la frecuencia (o la probabilidad) con la que los individuos en la muestra escogen la opción catalogada como de $Y_i = 1$, por medio del valor numérico del producto $\mathbf{x}_i'\boldsymbol{\beta}$, donde \mathbf{x}_i es un vector de variables que caracterizan a cada uno de los individuos en la muestra.

Supongamos ahora que en vez de utilizar una función lineal de las \mathbf{x}_i para caracterizar esta decisión utilizamos una función monótona creciente del producto $\mathbf{x}_i'\boldsymbol{\beta}$, $F(\mathbf{x}_i'\boldsymbol{\beta})$. Esta formulación es más general que la del modelo lineal de probabilidad y, en consecuencia, el procedimiento de estimación de los parámetros, así como la forma en que dichos valores deben interpretarse, es ahora diferente.

Con esta transformación se obtienen las siguientes ventajas: primero, basta tomar una función real F acotada entre 0 y 1 para que el problema que aparecía en el modelo de probabilidad lineal acerca del rango de valores de \hat{Y}_i desaparezca. En efecto, ahora $\hat{P}_i = F(\mathbf{x}_i'\boldsymbol{\beta})$ está siempre entre 0 y 1, con independencia del valor numérico del producto $\mathbf{x}_i'\boldsymbol{\beta}$. Es por ello que las funciones de distribución de variables aleatorias son candidatos importantes a ser elegidas para estas transformaciones.

En segundo lugar, supongamos que existe un indicador que depende de las características individuales: $I_i = \mathbf{x}_i'\boldsymbol{\beta}$, que determina la decisión tomada

por cada individuo. El individuo toma la decisión $Y_i = 1$ si el valor de su indicador I_i es superior a un cierto valor crítico I_i^* , y la decisión $Y_i = 0$ en caso contrario. Es decir, el indicador I_i refleja el sentimiento del decisor frente a la opción indicada por $Y_i = 1$, de modo que si su predisposición, indicada por I_i , es suficientemente grande (mayor que I_i^*), escoge dicha opción, y si no, escogerá la opción alternativa.

Por ser desconocido, consideramos el valor crítico I_i^* del indicador para cada individuo como una variable aleatoria. De acuerdo con esta interpretación, la probabilidad de que el individuo i -ésimo elija la acción $Y_i = 1$ viene dada por:

$$P_i = P(Y_i = 1) = P(I_i^* < I_i) = F(\mathbf{x}_i'\boldsymbol{\beta})$$

donde F es la distribución de probabilidad de la variable aleatoria I_i^* .

El lector puede observar que el modelo lineal de probabilidad pertenece a esta clase de modelos de decisión que acabamos de describir, en el caso particular en que se utiliza como $F(\cdot)$ la función de distribución de una variable aleatoria con distribución uniforme en el intervalo $[0, 1]$. En tal caso:

$$p_i = P_i + u_i = F(I_i) + u_i = F(\mathbf{x}_i'\boldsymbol{\beta}) + u_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$$

16.4.a. El modelo probit

Si se elige como función F la función de distribución Φ de una variable Normal $(0, 1)$, se tiene:

$$P_i = E(Y_i/\mathbf{x}_i) = P(Y_i = 1/\mathbf{x}_i) = P[I_i^* < I_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta})$$

de modo que:

$$\mathbf{x}_i'\boldsymbol{\beta} = \Phi^{-1}(P_i)$$

La probabilidad correspondiente a un vector \mathbf{x}_i es ahora:

$$P_i = \frac{1}{2\pi} \int_{-\infty}^{\mathbf{x}_i'\boldsymbol{\beta}} e^{-t^2/2} dt$$

que es una función creciente del valor numérico del indicador $I_i = \mathbf{x}_i'\boldsymbol{\beta}$.

Estimación de mínimos cuadrados con observaciones repetidas

El modelo original relaciona las frecuencias observadas p_i con las probabilidades teóricas P_i por medio de:

$$p_i = P_i + u_i, \quad \text{por lo que} \quad \Phi^{-1}(p_i) = \Phi^{-1}(P_i + u_i)$$

Como se muestra en el apéndice, esta expresión puede aproximarse por:

$$\Phi^{-1}(p_i) \cong \mathbf{x}'_i \boldsymbol{\beta} + \frac{1}{f(\mathbf{x}'_i \boldsymbol{\beta})} u_i$$

El modelo probit puede, por tanto, estimarse de modo aproximado, por una regresión de los llamados «probits» muestrales $\Phi^{-1}(p_i)$ sobre el vector \mathbf{x}_i . Se trata de calcular las frecuencias muestrales p_i , obtener los valores $\Phi^{-1}(p_i)$ a partir de las tablas de la distribución $N(0, 1)$ y estimar la regresión descrita.

Ahora bien, los residuos tienen heteroscedasticidad, puesto que:

$$\text{Var} \left(\frac{u_i}{f(\mathbf{x}'_i \boldsymbol{\beta})} \right) = \frac{P_i(1 - P_i)}{n_i [f(\mathbf{x}'_i \boldsymbol{\beta})]^2} \quad [16.2]$$

por lo que habría que utilizar mínimos cuadrados generalizados.

$$\boldsymbol{\beta} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\pi}$$

con una matriz $\boldsymbol{\Sigma}$ diagonal, con elementos genéricos dados por [16.2], donde $\boldsymbol{\pi}$ es el vector de probits muestrales. Como la matriz $\boldsymbol{\Sigma}$ es desconocida, hay que estimarla, para lo que se puede utilizar: a) las frecuencias observadas p_i , o bien b) las predicciones \hat{P}_i obtenidas a partir de un modelo de probabilidad lineal previamente estimado.

Estimación de máxima verosimilitud para observaciones individuales

El procedimiento de estimación por máxima verosimilitud es preciso cuando no es posible agrupar las observaciones según los valores del vector \mathbf{x}_i , como discutimos anteriormente. En tal situación, carece de sentido hablar de proporciones muestrales. En dichos casos, la estimación por máxima verosimilitud evita los problemas ya citados acerca de la estimación MCG del modelo lineal de probabilidad. Por otra parte, el estimador de máxima verosimilitud es eficiente, y se calcula sobre el modelo original, sin necesidad de ninguna aproximación.

En el caso del modelo probit, la función de verosimilitud es:

$$L = \prod_1^N [\Phi(\mathbf{x}'_i \boldsymbol{\beta})]^{Y_i} [1 - \Phi(\mathbf{x}'_i \boldsymbol{\beta})]^{1 - Y_i}$$

Nótese que para cada individuo i el término correspondiente en la función de verosimilitud es simplemente $\Phi(\mathbf{x}'_i \boldsymbol{\beta})$ o $1 - \Phi(\mathbf{x}'_i \boldsymbol{\beta})$, dependiendo de que $Y_i = 1$ o $Y_i = 0$. Por tanto, la función logaritmo neperiano de la verosimilitud es

$$\ln L = \sum_1^N Y_i \ln \Phi(\mathbf{x}'_i \boldsymbol{\beta}) + \sum_1^N (1 - Y_i) \ln [1 - \Phi(\mathbf{x}'_i \boldsymbol{\beta})]$$

y tomando derivadas con respecto al vector β se tienen las k condiciones necesarias de optimalidad:

$$\sum_1^N Y_i \frac{f(x_i' \beta)}{\Phi(x_i' \beta)} x_i + \sum_1^N (1 - Y_i) \frac{-f(x_i' \beta)}{1 - \Phi(x_i' \beta)} x_i = \mathbf{0}_k$$

$$S(\beta) = \sum_1^N \frac{Y_i - \Phi(x_i' \beta)}{\Phi(x_i' \beta) [1 - \Phi(x_i' \beta)]} f(x_i' \beta) x_i = \mathbf{0} \quad [16.3]$$

donde $S(\beta)$ denota el vector gradiente de la función de verosimilitud. Si derivamos de nuevo en la expresión [16.3] con respecto al vector β , se obtiene la matriz hessiana, y tomando esperanza en ésta y cambiando de signo se obtiene finalmente la matriz de información, $I(\beta)$:

$$I(\beta) = \sum_1^N \frac{[f(x_i' \beta)]^2}{\Phi(x_i' \beta) [1 - \Phi(x_i' \beta)]} x_i x_i'$$

Conviene hacer hincapié en que en las expresiones anteriores N denota el número total de observaciones, por lo que, prescindiendo de clasificaciones, hay que considerar un sumando para cada observación muestral. En particular, en estos problemas es más sencillo utilizar el método del «scoring» de la Sección 12.5, razón por la que hemos calculado directamente la matriz de información, a partir de la matriz de derivadas segundas de la función de verosimilitud con respecto al vector β . La inversa de la matriz de información será además la *matriz de covarianzas del estimador de máxima verosimilitud* del vector β . El procedimiento de estimación por máxima verosimilitud utilizaría las expresiones anteriores del modo indicado en el Capítulo 12, es decir:

$$\hat{\beta}_n = \hat{\beta}_{n-1} + [I(\hat{\beta}_{n-1})]^{-1} S(\hat{\beta}_{n-1})$$

que proporciona la corrección que hay que introducir en el estimador del vector β en cada iteración. Al sustituir las expresiones de $I(\beta)$ y $S(\beta)$ antes obtenidas puede verse fácilmente que si se hace el cambio de variables:

$$x_{ij}^* = \frac{x_{ij} f(x_i' \beta)}{\sqrt{\Phi(x_i' \beta) (1 - \Phi(x_i' \beta))}}, \quad j = 1, 2, \dots, k$$

que forma, para cada observación i , un vector de dimensión k , e:

$$y_i^* = \frac{y_i - \Phi(x_i' \beta)}{\sqrt{\Phi(x_i' \beta) (1 - \Phi(x_i' \beta))}}$$

entonces la corrección a introducir en el estimador $\hat{\beta}_{n-1}$ coincide con los coeficientes estimados por mínimos cuadrados ordinarios en una regresión que

utilizase y_i^* como variable a explicar, y \mathbf{x}_i^* como vector de variables explicativas, utilizando los β_{n-1} para calcular x_{ij}^* e y_i^* .

16.4.b. El modelo logit

Este modelo surge cuando, para representar la probabilidad de que un individuo escoja la opción $Y_i = 1$, se utiliza la función de distribución logística:

$$F(z) = \frac{e^z}{1 + e^z}, \quad -\infty < z < \infty$$

que tiene como función de densidad:

$$f(z) = \frac{1}{(1 + e^z)^2} = F(z)(1 - F(z)), \quad -\infty < z < \infty$$

y como inversa:

$$F^{-1}(w) = \ln \frac{w}{1 - w} \quad [16.4]$$

puesto que:

$$F^{-1}(F(z)) = \ln \left(\frac{\frac{e^z}{1 + e^z}}{1 - \frac{e^z}{1 + e^z}} \right) = \ln e^z = z$$

Bajo este supuesto tenemos:

$$P_i = P(Y_i = 1) = F(\mathbf{x}_i' \boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i' \boldsymbol{\beta}}} \quad [16.5]$$

de modo que:

$$p_i = P_i + u_i = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i' \boldsymbol{\beta}}} + u_i$$

- Estimación por mínimos cuadrados con observaciones repetidas

Utilizando la expresión de la aproximación lineal en serie de Taylor (véase apéndice):

$$\ln \frac{p_i}{1-p_i} = \ln \frac{P_i + u_i}{1 - (P_i + u_i)} \cong \ln \frac{P_i}{1 - P_i} + \frac{u_i}{P_i(1 - P_i)} = \mathbf{x}'_i \boldsymbol{\beta} + \frac{u_i}{P_i(1 - P_i)} \quad [16.6]$$

donde hemos utilizado [16.5].

La expresión [16.6] sugiere que en el caso de disponer de observaciones repetidas, es decir, de grupos de observaciones con iguales valores del vector de características \mathbf{x}_i , podrían calcularse las frecuencias muestrales p_i y estimar la regresión anterior. Es claro que, en un intento de ganar eficiencia, debería utilizarse el estimador MCG, con una matriz $\boldsymbol{\Sigma}$ diagonal cuyo elemento genérico sea igual a:

$$\text{Var} \left(\frac{u_i}{P_i(1 - P_i)} \right) = \frac{1}{n_i P_i(1 - P_i)}$$

ya que, como vimos con anterioridad, $\text{Var}(u_i) = \frac{P_i(1 - P_i)}{n_i}$. Al desconocer el valor de los parámetros P_i habría que obtenerlos de una estimación previa por MCO, o de la estimación de un modelo lineal de probabilidad.

Estimación de máxima verosimilitud con observaciones individuales

La función de verosimilitud muestral es:

$$L = \prod_{Y_i=1} F(\mathbf{x}'_i \boldsymbol{\beta}) \prod_{Y_i=0} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})] = \frac{e^{(\sum_1^N Y_i \mathbf{x}'_i \boldsymbol{\beta})}}{\prod_1^N [1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}]}$$

o, lo que es lo mismo:

$$\ln L = \sum_1^N Y_i (\mathbf{x}'_i \boldsymbol{\beta}) - \sum_1^N \ln (1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}) = (\sum_1^N Y_i \mathbf{x}'_i) \boldsymbol{\beta} - \sum_1^N \ln (1 + e^{\mathbf{x}'_i \boldsymbol{\beta}})$$

y denotando por $\mathbf{z}' = \sum_1^N Y_i \mathbf{x}'_i$ un vector fila $1 \times k$ se tiene:

$$\ln L = \mathbf{z}' \boldsymbol{\beta} - \sum_1^N \ln [1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}]$$

y

$$S(\boldsymbol{\beta}) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \mathbf{z} - \sum_1^N \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}} \mathbf{x}_i}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}} = \mathbf{0}_k$$

y este sistema de k ecuaciones no lineales debería, en principio, resolverse por procedimientos numéricos, para obtener el vector de estimadores β . La matriz de información es:

$$I(\beta) = \sum_1^N \frac{e^{x_i\beta} x_i x_i'}{[1 + e^{x_i\beta}]^2} = \sum_1^N x_i P_i (1 - P_i) x_i'$$

Para estimar el valor β por el algoritmo del «scoring» se comienza de un estimador β_0 y se actualiza por medio de:

$$\beta_1 = \beta_0 + [I(\beta_0)]^{-1} S(\beta_0)$$

En realidad, la matriz $S(\beta)$ puede escribirse también:

$$S(\beta) = \sum_1^N Y_i x_i - \sum_1^N \frac{x_i}{1 + e^{-x_i\beta}} = \sum_1^N (Y_i - P_i) x_i$$

donde $P_i = \frac{1}{1 + e^{-x_i\beta}}$, por lo que el algoritmo puede describirse como sigue:

1. A partir de un estimador inicial $\hat{\beta}_0$, calcular $\hat{P}_i(1 - \hat{P}_i)$.
2. Transformar las variables:

$$x_i^* = x_i \sqrt{\hat{P}_i(1 - \hat{P}_i)}$$

$$Y_i^* = \frac{(Y_i - \hat{P}_i)}{\sqrt{\hat{P}_i(1 - \hat{P}_i)}}$$

y el cambio a introducir en el vector $\hat{\beta}_0$ viene dado por los coeficientes estimados por mínimos cuadrados ordinarios en una regresión de Y_i^* sobre el vector x_i^* .

El algoritmo se itera hasta conseguir su convergencia, y se utiliza la inversa de la matriz de información evaluada en el último estimador obtenido, como estimación de la matriz de covarianzas de β . Por otra parte, los métodos de inferencia que discutimos en el Capítulo 4 son válidos utilizando esta *matriz de covarianzas*, ya que el estimador de máxima verosimilitud que resulta tiene *distribución Normal asintótica*. Las probabilidades de que un individuo con características x_i escoja la acción que hemos catalogado como $Y_i = 1$ se estiman mediante la expresión:

$$\hat{P}_i = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}$$

Ejemplo 16.1. Supongamos que se dispone de información procedente de 40 familias acerca de sus ingresos mensuales (en miles de pesetas). También se conoce si la vivienda que ocupan es de alquiler o, por el contrario, es propiedad de la familia. Se pretende analizar el efecto que los ingresos familiares tienen sobre la decisión de comprar un piso. Para ello se define la variable Y_i , que toma el valor 1 si la familia es propietaria de la vivienda, y el valor 0 si la familia ocupa la vivienda en régimen de alquiler. La variable x_i denota los ingresos mensuales, en miles de pesetas.

Y_i	x_i	Probabilidad		Y_i	x_i	Probabilidad		Y_i	x_i	Probabilidad	
		MLP	Logit			MLP	Logit			MLP	Logit
1	150	0,50	0,62	1	60	0,13	0,09	1	160	0,55	0,69
0	80	0,21	0,15	0	120	0,38	0,38	0	190	0,67	0,85
0	100	0,30	0,25	1	180	0,63	0,80	0	90	0,26	0,19
1	200	0,71	0,88	0	80	0,21	0,15	1	180	0,63	0,80
1	250	0,92	0,97	0	90	0,26	0,19	0	60	0,13	0,09
1	160	0,55	0,69	1	220	0,79	0,94	0	120	0,38	0,38
0	90	0,26	0,19	0	120	0,38	0,38	1	200	0,71	0,88
1	200	0,71	0,88	0	220	0,79	0,94	1	160	0,55	0,69
1	250	0,92	0,97	1	180	0,63	0,80	0	90	0,26	0,19
0	140	0,46	0,54	1	240	0,88	0,96	1	150	0,50	0,61
0	120	0,38	0,38	1	100	0,30	0,25	1	110	0,34	0,31
1	200	0,71	0,88	1	160	0,55	0,69	0	70	0,17	0,11
0	50	0,09	0,06	1	160	0,55	0,69	0	80	0,21	0,15
0	110	0,34	0,31								

La estimación del modelo lineal de probabilidad por mínimos cuadrados ordinarios y sus desviaciones típicas fue:

$$Y_i = -0,2589 + 0,0055 \text{ Ingresos}_i$$

(0,1755) (0,0011)

que puede interpretarse del siguiente modo:

1. La probabilidad estimada de que una familia sin ingresos sea propietaria de la vivienda es $-0,2589$. Claramente, dicha cifra no representa una probabilidad, y debe interpretarse en el sentido de que el suceso en cuestión es imposible.

2. Un incremento en el nivel de ingresos de 10.000 pesetas/mes incrementa en 0,055 la probabilidad de que la familia sea propietaria de la vivienda. Así, si una familia pasa de recibir ingresos de 100.000 pesetas/mes a unos ingresos de 110.000 pesetas/mes, la probabilidad de que sea propietaria aumenta de 0,29 a 0,34.

Como sabemos, sin embargo, el modelo lineal de probabilidad debe

estimarse por mínimos cuadrados generalizados, tras estimar la varianza del término de error por:

$$\hat{\sigma}_i^2 = \hat{P}_i(1 - \hat{P}_i), \quad \text{con} \quad \hat{P}_i = -0,2589 + 0,0055 \text{ Ingresos}_i$$

para ello, es preciso comprobar cuáles son las observaciones para las que dicha estimación de $\hat{\sigma}_i^2$ resulta negativa. Ello ocurrió con las familias números 5, 9 y 24, que fueron retiradas de la muestra en esta segunda estimación.

La estimación de mínimo cuadrados generalizados fue:

$$Y_i = -0,1160 + 0,0041 \text{ Ingresos}_i \\ (0,1518) \quad (0,0010)$$

donde, como puede verse, el método de mínimos cuadrados ordinarios sobrestimó el incremento que se produce en la probabilidad de poseer la vivienda al pasar de una familia con una determinada renta a una familia con una renta superior. En la tabla que contiene la información muestral aparecen también las probabilidades estimadas para cada familia, de acuerdo con este último modelo.

La estimación de máxima verosimilitud del modelo logit generó los siguientes resultados:

$$P_i = \frac{1}{1 + \exp(-\hat{\alpha}_{MV} - \hat{\beta}_{MV} \text{ Ingresos}_i)}$$

o, lo que es lo mismo: $\ln \frac{P_i}{1 - P_i} = \hat{\alpha}_{MV} + \hat{\beta}_{MV} \text{ Ingresos}_i$, con:

$$\hat{\alpha}_{MV} = -4,255 \quad \hat{\beta}_{MV} = 0,0315 \\ (1,35) \quad (0,0096)$$

Las probabilidades generadas por estas estimaciones aparecen asimismo en la tabla que contiene la información muestral. Una información adicional generada por estas estimaciones es la del nivel de ingresos a partir del cual es más probable que una familia sea propietaria de la vivienda. Ello es equivalente a encontrar el nivel de ingresos para el que $P_i > 0,50$, que en este caso corresponde a: Ingresos = 135.100 pesetas/mes.

El porcentaje de aciertos que se tendría con el modelo se obtiene comparando los casos en que el modelo genera una probabilidad mayor de 0,50 de ser propietario de la vivienda con los casos en que la familia realmente es propietaria de la vivienda. El modelo lineal de probabilidad registró 35 aciertos por 34 del modelo logit (interpretando $\hat{P}_i = 0,5$ como propietarios).

16.5. INFERENCIA EN MODELOS DE ELECCION DISCRETA

16.5.a. Interpretación de los coeficientes estimados

Al interpretar las estimaciones de los coeficientes del modelo, hay que tener en cuenta que en los modelos probit y logit los coeficientes estimados miden la relación lineal existente entre el índice I_i y las variables x_{ij} , es decir, que indican el efecto que las variables x_{ij} tienen sobre $F^{-1}(P_i)$. A diferencia del modelo lineal de probabilidad, la influencia que las variables explicativas tienen sobre la probabilidad de escoger la opción $Y_i = 1$ no son simplemente los valores de los coeficientes estimados, sino que dependen también de los valores de las variables explicativas.

En cualquier modelo, dicha influencia viene indicada por la derivada parcial de la variable endógena con respecto a las variables exógenas que, dependiendo de las especificaciones, son las siguientes:

a) En el modelo de probabilidad lineal: $\frac{\partial P_i}{\partial x_{ik}} = \beta_k$, independientemente del valor del vector de características x_i . Como vamos a ver a continuación, lo contrario ocurre en los modelos probit y logit.

b) En el modelo probit: $\frac{\partial P_i}{\partial x_{ik}} = f(x_i' \beta) \beta_k$, donde f es la función de densidad de una variable $N(0, 1)$. Como consecuencia, los cocientes entre valores estimados de dos parámetros $\hat{\beta}_j / \hat{\beta}_k$ miden la importancia relativa de los efectos que las variables x_j y x_k tienen sobre la probabilidad de escoger la alternativa $Y_i = 1$. Debido a esta propiedad, si bien los coeficientes de un modelo probit no son directamente interpretables, sus valores relativos sí lo son.

c) En el modelo logit: $\frac{\partial P_i}{\partial x_k} = \frac{e^{x_i' \beta} \beta_k}{[1 + e^{x_i' \beta}]^2}$, por lo que, de nuevo, el comentario anterior acerca de los valores relativos de los parámetros estimados $\hat{\beta}_j$ y $\hat{\beta}_k$ es también válido en este modelo.

Hay que tener presente que al representar estos modelos únicamente comparaciones bivariantes entre pares de alternativas, entonces las variables socio-económicas que toman el mismo valor para las distintas alternativas pierden todo poder explicativo.

Si los coeficientes estimados en los modelos probit y logit no admiten una interpretación inmediata, mucho menos son comparables entre sí o con las estimaciones de los coeficientes que se obtienen en el modelo lineal de probabilidad. Amemiya (1981) sugirió efectuar las siguientes transformaciones para llevar a cabo comparaciones entre las distintas estimaciones:

a)

$$\begin{aligned} \hat{\beta}_{mlp} &= 0,25 \hat{\beta}_{logit} \\ \hat{\alpha}_{mlp} &= 0,25 \hat{\alpha}_{logit} + 0,50 \end{aligned}$$

donde α denota el término independiente del modelo si lo hay. Los subíndices hacen referencia al procedimiento de estimación utilizado. Por otra parte:

b)

$$\hat{\beta}_{\text{logit}} \frac{\sqrt{3}}{\pi} = \hat{\beta}_{\text{probit}}$$

o

$$(0,625)\hat{\beta}_{\text{logit}} = \hat{\beta}_{\text{probit}}$$

16.5.b. La bondad de ajuste del modelo

Al no ser lineales, no deben utilizarse en los modelos probit y logit los estadísticos sugeridos en el Capítulo 4 para juzgar la bondad del ajuste del modelo. Para llevar a cabo dicha evaluación se utilizan los siguientes criterios:

1. En el caso de observaciones no repetidas, una vez estimado el modelo, puede utilizarse para predecir la probabilidad de que un individuo con características x_i elija cada alternativa. Hecho este ejercicio, escojamos la alternativa con mayor probabilidad, y compárese, para cada individuo, con la alternativa realmente escogida. El porcentaje de individuos que eligieron la alternativa predicha por el modelo puede utilizarse como indicador de la bondad del ajuste.

2. En el caso de observaciones repetidas debería llevarse a cabo una comparación, en términos del error cuadrático medio, entre las frecuencias muestrales observadas, p_i , y las frecuencias previstas por el modelo \hat{P}_i . El estadístico

$$S = \sum_1^M \frac{n_i(p_i - \hat{P}_i)^2}{\hat{P}_i(1 - \hat{P}_i)}$$

se distribuye aproximadamente, en muestras grandes, como una chi-cuadrado con $M - k$ grados de libertad. Un estadístico similar puede utilizarse incluso cuando no se dispone de observaciones repetidas, utilizando Y_i por p_i e \hat{Y}_i por \hat{P}_i , omitiendo el factor n_i .

El estadístico $\sum_1^{40} \frac{(P_i - \hat{P}_i)^2}{\hat{P}_i(1 - \hat{P}_i)}$ tomó en el ejemplo numérico de la Sección 16.4 un valor igual a 47,10. Este valor debe compararse con los valores críticos para una distribución chi-cuadrado con 38 grados de libertad que son de 55,8 si se utiliza un nivel de significación del 95 por 100 y de 63,7 si se utiliza el 99 por 100. Por tanto, en ambos casos se acepta el modelo como bueno.

3. Definiendo un «residuo» como $\hat{u}_i = Y_i - \hat{P}_i$ se tiene $SR = \sum_i^n \hat{u}_i^2$, $ST = \sum_1^n (Y_i - \bar{Y})^2$, y el cociente $\rho^2 = 1 - \frac{SR}{ST}$ puede utilizarse de modo análogo al coeficiente de determinación en un modelo lineal, para juzgar la bondad de ajuste de un modelo de elección discreta.

16.5.c. Contrastación de hipótesis

Si el modelo se ha estimado por máxima verosimilitud, para contrastar un conjunto de restricciones puede compararse el valor de la función de verosimilitud en el estimador restringido con el valor de dicha función de verosimilitud en el estimador obtenido ignorando las restricciones. El contraste de razón de verosimilitudes se basa en que el estadístico

$$-2 \ln \frac{L(\hat{\beta}_R)}{L(\hat{\beta}_{MV})} = -2(L(\hat{\beta}_R) - \ln L(\hat{\beta}_{MV}))$$

sigue una distribución chi-cuadrado con q grados de libertad (el número de restricciones).

En particular, puesto que la significación global del modelo es siempre una hipótesis de interés, debe calcularse siempre el valor L_0 de la función de verosimilitud que se obtiene cuando todos los coeficientes del modelo se hacen iguales a cero, excepto el término independiente, que se estima.

Puede probarse fácilmente que tanto en el modelo probit como en el logit se tiene:

$$\ln L_0 = N[p \ln p + (1 - p) \ln (1 - p)]$$

donde p es la proporción muestral de observaciones con $Y_i = 1$. El estadístico $-2(\ln L_0 - \ln L(\hat{\beta}_{MV}))$ se distribuye χ^2_{k-1} y se puede utilizar como contraste de significación global del modelo.

La función de verosimilitud puede utilizarse asimismo para obtener un estadístico que tiene cierta semejanza con el coeficiente de determinación R^2 :

$$\tilde{R}^2 = 1 - \frac{\ln L_0}{\ln L(\hat{\beta}_{MV})}$$

donde L_0 es el estadístico antes introducido, y $L(\hat{\beta}_{MV})$ denota el valor maximizado de la función de verosimilitud.

Si los coeficientes β son poco significativos, la capacidad explicativa del modelo será muy reducida, $L(\hat{\beta}_{MV})$ será similar en valor a L_0 y \tilde{R}^2 próximo a cero, al igual que ocurre con el coeficiente de determinación en el modelo lineal. Por el contrario, cuanto mayor sea la capacidad explicativa del modelo, mayor será $L(\hat{\beta}_{MV})$, siempre en exceso de L_0 , y más próximo estará \tilde{R}^2 a 1. Sin embargo, ni llega a valer 1 ni es fácil interpretar los valores de \tilde{R}^2 entre 0 y 1.

Para hipótesis sobre el valor de un coeficiente, puede utilizarse un estadístico t obtenido con la desviación típica proporcionada por la inversa de la matriz de información. Para restricciones lineales del tipo $R\beta = r$ puede utilizarse el estadístico de Wald:

$$W = (R\hat{\beta} - r)' [R\hat{\Sigma}R']^{-1} (R\hat{\beta} - r)$$

que debe compararse con una distribución chi-cuadrado con q grados de libertad, siendo q el número de filas de \mathbf{R} , es decir, el número de restricciones que se contrastan.

16.6. MODELOS DE ALTERNATIVAS MÚLTIPLES

En muchas ocasiones, el agente económico debe elegir una entre más de dos opciones alternativas. Los modelos que hemos presentado en las secciones anteriores pueden adaptarse, sin mucha dificultad, a este tipo de problemas de decisión. A modo de ejemplo, supongamos que cada individuo de una muestra puede escoger una entre tres alternativas posibles, y denotemos por Y_{ij} una variable que toma el valor 1 si el individuo i escoge la opción j , mientras que toma el valor 0 si el individuo escoge otra de las opciones. Sea P_{ij} la probabilidad de que la variable Y_{ij} tome el valor 1 y supongamos que las probabilidades P_{i1} , P_{i2} , P_{i3} dependen linealmente de un vector de características \mathbf{x}_i , de modo que:

$$P_{i1} = \mathbf{x}'_i \boldsymbol{\beta}_1; \quad P_{i2} = \mathbf{x}'_i \boldsymbol{\beta}_2; \quad P_{i3} = \mathbf{x}'_i \boldsymbol{\beta}_3 \quad [16.7]$$

donde $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ y $\boldsymbol{\beta}_3$ son vectores de dimensión k . Siguiendo un razonamiento análogo al de secciones anteriores, especificaríamos el modelo:

$$Y_{ij} = \mathbf{x}'_i \boldsymbol{\beta}_j + u_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, 3$$

donde N es el número de individuos en la muestra. Como $P_{i1} + P_{i2} + P_{i3} = 1$, se tiene $\beta_{j1} + \beta_{j2} + \beta_{j3} = 0$ para los coeficientes de la variable j -ésima en los tres modelos, excepto para el término independiente, para el que $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

De este modo, sólo es preciso estimar dos de las ecuaciones anteriores, pues los coeficientes de la tercera pueden obtenerse a partir de las igualdades que acabamos de mencionar. Si las observaciones muestrales pueden agruparse en clases de individuos con igual vector de características \mathbf{x}_i , entonces la especificación anterior consta de tres modelos lineales de probabilidad, que podrían estimarse, por separado, por mínimos cuadrados generalizados, siguiendo las directrices de las secciones anteriores. Todo esto supone que el vector de características \mathbf{x}_i es el mismo para las tres alternativas.

En realidad, el problema ahora es más complejo, puesto que los términos de error de los dos modelos lineales de probabilidad que estimaríamos en este ejemplo están, en principio, correlacionados entre sí, por lo que una estimación eficiente debería incorporar dichas correlaciones. Ello podría llevarse a cabo incorporando un método análogo al de los sistemas de regresiones aparentemente no relacionadas que vimos en el Capítulo 8 (véase Zellner y Lee, 1965).

También existen modelos probit y logit multinomiales, aunque vamos a remitirnos aquí a discutir tan sólo el modelo *logit multinomial*. Suponiendo que la función de distribución del término de error del modelo es la de

Weibull: $P(u_{ij} \leq a) = e^{-e^{-a}}$, $-\infty < a < \infty$, McFadden (1974) probó que las probabilidades de escoger cada alternativa posible vienen dadas por:

$$P_{ij} = \frac{e^{x_i' \beta_j}}{\sum_1^3 e^{x_i' \beta_j}}$$

por lo que se tiene:

$$\frac{P_{i1}}{P_{i2}} = \frac{e^{x_i' \beta_1}}{e^{x_i' \beta_2}} = \exp [x_i' (\beta_1 - \beta_2)]$$

y, por tanto:

$$\frac{P_{i2}}{P_{i1}} + \frac{P_{i3}}{P_{i1}} = \frac{P_{i2} + P_{i3}}{P_{i1}} = \frac{1 - P_{i1}}{P_{i1}} = \frac{1}{P_{i1}} - 1$$

y, en consecuencia:

$$P_{i1} = \left[1 + \sum_2^3 \frac{P_{ij}}{P_{i1}} \right]^{-1} = \left[1 + \frac{e^{x_i' \beta_2} + e^{x_i' \beta_3}}{e^{x_i' \beta_1}} \right]^{-1}$$

y, si se normaliza el vector β_1 como un vector formado por k ceros, se tiene:

$$P_{i1} = [1 + e^{x_i' \beta_2} + e^{x_i' \beta_3}]^{-1}$$

y

$$\frac{P_{ij}}{P_{i1}} = e^{x_i' \beta_j}$$

por lo que, finalmente:

$$P_{ij} = \frac{e^{x_i' \beta_j}}{1 + e^{x_i' \beta_2} + e^{x_i' \beta_3}}; \quad j = 1, 2$$

$$P_{i1} = \frac{1}{1 + e^{x_i' \beta_2} + e^{x_i' \beta_3}}$$

y, como fácilmente puede verse, su suma es igual a 1 para cada individuo i . A pesar de que no es difícil de estimar, este modelo tiene la dificultad de depender del axioma de independencia de alternativas irrelevantes (véase McFadden, 1974), cuya validez práctica es cuestionable cuando algunas de las alternativas son sustitutos cercanos.

Por otra parte, si existen observaciones repetidas, el modelo puede estimarse por mínimos cuadrados, como vimos en el modelo logit, puesto que:

$$\ln \frac{P_{i2}}{P_{i1}} = \mathbf{x}'_i \boldsymbol{\beta}_2$$

$$\ln \frac{P_{i3}}{P_{i1}} = \mathbf{x}'_i \boldsymbol{\beta}_3$$

aunque habremos de tener en cuenta, de nuevo, la presencia de heteroscedasticidad en el término de error de cada una de estas ecuaciones. Por otra parte, si no existen observaciones repetidas, será preciso estimar por máxima verosimilitud. La función de verosimilitud en este modelo es:

$$L(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3/\mathbf{x}_i) = \prod_{i \in I_1} \frac{1}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}} \prod_{i \in I_2} \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}} \prod_{i \in I_3} \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

Finalmente, una vez estimados los coeficientes $\boldsymbol{\beta}$, las probabilidades de que un individuo escoja cada una de las alternativas podrían estimarse a partir de las expresiones:

$$P(Y_1 = 1) = \frac{1}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

$$P(Y_2 = 1) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

$$P(Y_3 = 1) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

Parte II

VARIABLES DEPENDIENTES LIMITADAS

16.7. VARIABLES DEPENDIENTES TRUNCADAS

En ocasiones, la naturaleza de un proceso de recogida de datos hace que *sólo se observen las variables endógenas y exógenas* correspondientes a una unidad económica si el valor de la variable endógena excede de un cierto umbral a . Seguimos entonces interesados en explicar el valor esperado de una variable que consideramos endógena, y_i , *condicional* en los valores que toma un vector de

Por otra parte, si existen observaciones repetidas, el modelo puede estimarse por mínimos cuadrados, como vimos en el modelo logit, puesto que:

$$\ln \frac{P_{i2}}{P_{i1}} = \mathbf{x}'_i \boldsymbol{\beta}_2$$

$$\ln \frac{P_{i3}}{P_{i1}} = \mathbf{x}'_i \boldsymbol{\beta}_3$$

aunque habremos de tener en cuenta, de nuevo, la presencia de heteroscedasticidad en el término de error de cada una de estas ecuaciones. Por otra parte, si no existen observaciones repetidas, será preciso estimar por máxima verosimilitud. La función de verosimilitud en este modelo es:

$$L(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3/\mathbf{x}_i) = \prod_{i \in I_1} \frac{1}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}} \prod_{i \in I_2} \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}} \prod_{i \in I_3} \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

Finalmente, una vez estimados los coeficientes $\boldsymbol{\beta}$, las probabilidades de que un individuo escoja cada una de las alternativas podrían estimarse a partir de las expresiones:

$$P(Y_1 = 1) = \frac{1}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

$$P(Y_2 = 1) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

$$P(Y_3 = 1) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}_2} + e^{\mathbf{x}'_i \boldsymbol{\beta}_3}}$$

Parte II

VARIABLES DEPENDIENTES LIMITADAS

16.7. VARIABLES DEPENDIENTES TRUNCADAS

En ocasiones, la naturaleza de un proceso de recogida de datos hace que *sólo se observen las variables endógenas y exógenas correspondientes a una unidad económica si el valor de la variable endógena excede de un cierto umbral a* . Seguimos entonces interesados en explicar el valor esperado de una variable que consideramos endógena, y_i , *condicional* en los valores que toma un vector de

variables exógenas \mathbf{x}_i . Lo que distingue a este problema de los analizados en capítulos previos es que observamos el vector (y_i, \mathbf{x}_i) sólo si y_i excede de un umbral mínimo a . Suponemos que $u_i \sim N(0, \sigma_u^2)$.

Como se probó en la Sección 2.13, si la esperanza matemática de la variable y_i , sin truncamiento, viene representada por la combinación lineal $\mathbf{x}'_i\boldsymbol{\beta}$, donde $\boldsymbol{\beta}$ es un vector $k \times 1$ de coeficientes desconocidos, es decir, $E(y_i/\mathbf{x}_i) = \mathbf{x}'_i\boldsymbol{\beta}$, entonces:

$$E(y_i/\mathbf{x}_i, y_i > a) = \mathbf{x}'_i\boldsymbol{\beta} + \sigma_u \cdot \frac{\phi\left(\frac{a - \mathbf{x}'_i\boldsymbol{\beta}}{\sigma_u}\right)}{1 - \Phi\left(\frac{a - \mathbf{x}'_i\boldsymbol{\beta}}{\sigma_u}\right)} = \mathbf{x}'_i\boldsymbol{\beta} + \sigma_u \lambda(\alpha_i)$$

donde $\alpha_i = \frac{a - \mathbf{x}'_i\boldsymbol{\beta}}{\sigma_u}$ es una función no lineal de \mathbf{x}_i y $\boldsymbol{\beta}$.

Denotamos en lo sucesivo $\lambda_i = \lambda(\alpha_i) = \frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$, el ratio de Mills, y por $\delta_i = \delta(\alpha_i) = \lambda_i(\lambda_i - \alpha_i)$. Todas ellas son funciones que, evaluadas en cada observación muestral, se corresponden con los parámetros análogos de la Sección 2.13.

Bajo el supuesto de Normalidad de u_i , los efectos de las variables explicativas sobre la variable endógena vienen dados por:

$$\frac{\partial E(y_i/y_i > a)}{\partial \mathbf{x}_i} = \boldsymbol{\beta} + \sigma_u \left(\frac{d\lambda_i}{d\alpha_i}\right) \left(\frac{d\alpha_i}{d\mathbf{x}_i}\right) = \boldsymbol{\beta}[1 - \delta(\alpha_i)]$$

siendo inferiores, por tanto, a los que tienen sobre la variable y_i observada sin truncamiento. Esta atenuación de efectos también ocurre en la varianza, pues $\text{Var}(y_i/y_i > a)$ ya no es σ_u^2 , sino:

$$\text{Var}(y_i/y_i > a) = \sigma_u^2[1 - \delta(\alpha_i)] < \sigma_u^2$$

16.7.a. Estimación de máxima verosimilitud

El modelo de variable dependiente truncada puede escribirse:

$$y_i/y_i > a = E[y_i/y_i > a] + u_i = \mathbf{x}'_i\boldsymbol{\beta} + \sigma_u \lambda_i + u_i$$

donde u_i tiene heteroscedasticidad y $\lambda_i = \lambda(\alpha_i)$ no es observable.

La función de densidad truncada de u_i es:

$$f(u_i/u_i > a - \mathbf{x}'_i\boldsymbol{\beta}) = \frac{f(u_i)}{P(u_i > a - \mathbf{x}'_i\boldsymbol{\beta})} = \frac{f(u_i)}{1 - \Phi\left(\frac{a - \mathbf{x}'_i\boldsymbol{\beta}}{\sigma_u}\right)}$$

por lo que la función de verosimilitud del modelo es:

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_1^n (y_i - \mathbf{x}'_i \boldsymbol{\beta}) - \sum_1^n \ln \left[1 - \Phi \left(\frac{a - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma_u} \right) \right]$$

cuyas condiciones de maximización pueden escribirse:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_i \left[\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma_u^2} - \frac{\lambda_i}{\sigma_u} \right] \mathbf{x}_i = 0_k$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \sum_i \left[-\frac{1}{2\sigma_u^2} + \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{2\sigma_u^4} - \frac{\alpha_i \lambda_i}{2\sigma_u^2} \right] = 0$$

lo que puede utilizarse en un procedimiento de estimación del tipo Gauss-Newton.

Comenzando de una estimación inicial $(\hat{\boldsymbol{\beta}}_0, \hat{\sigma}_0^2)$ podríamos calcular α_i y λ_i e iniciar el algoritmo iterativo utilizando las expresiones del gradiente que hemos obtenido.

16.8. VARIABLES DEPENDIENTES CENSURADAS

Una variable dependiente se dice *censurada* cuando todos sus valores en un cierto rango son sustituidos por un valor fijo y_0 . Generalmente, son los valores por debajo (o por encima) de y_0 , que quedan sustituidos por éste. El ejemplo más conocido se debe a Tobin (1958), quien, analizando una muestra de gastos en consumo de un bien duradero (automóvil), observó que, por debajo del gasto mínimo y_0 necesario para adquirir el bien, a una familia que no hubiese adquirido coche se le asignaría una «demanda de gasto en consumo» de 0 pesetas cuando lo que ocurre es que su decisión de gasto habrá sido generalmente positiva, sólo que inferior a y_0 .

En una muestra censurada se observan los valores de las variables explicativas incluso cuando $y_i^* = 0$, a diferencia de lo que ocurre en una muestra truncada, en la que, cuando no se observa y_i , tampoco se observa el vector \mathbf{x}'_i .

Supongamos que el umbral de censura es 0, es decir, observamos y_i sólo si $y_i > 0$, y observamos el valor 0 si $y_i \leq 0$. Definamos una nueva variable:

$$y_i^* = 0 \quad \text{si} \quad y_i \leq 0$$

$$y_i^* = y_i \quad \text{si} \quad y_i > 0$$

Si y_i se distribuye $N(\mu, \sigma^2)$, entonces:

$$\begin{cases} P(y_i^* = 0) = P(y_i \leq 0) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right) \\ f(y_i^*) = f(y_i) \quad \text{si } y_i^* > 0 \end{cases}$$

Luego y_i^* tiene una distribución que mezcla una componente continua por encima de y_0 con otra discreta en y_0 . Además:

$$E(y_i^*) = 0 \cdot P(y_i^* = 0) + E(y_i^*/y_i^* > 0) \cdot P(y_i^* > 0) = \Phi\left(\frac{\mu}{\sigma}\right) \cdot (\mu + \sigma\lambda)$$

y

$$\text{Var}(y_i^*) = \sigma^2 \left[(1 - \delta) + (\lambda - \alpha)^2 \cdot \Phi\left(\frac{\mu}{\sigma}\right) \cdot \left(1 - \Phi\left(\frac{\mu}{\sigma}\right)\right) \right] \cdot \Phi\left(\frac{\mu}{\sigma}\right)$$

Consideremos ahora el problema de la censura en el contexto del modelo de regresión. Sea:

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i$$

$$\begin{cases} y_i^* = 0 & \text{si } y_i \leq 0 \\ y_i^* = y_i & \text{si } y_i > 0 \end{cases}$$

donde, de acuerdo con los resultados anteriores, se tiene, para una observación tomada al azar de la población, esté censurada o no:

$$E(y_i^*/\mathbf{x}_i) = \Phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}}{\sigma_u}\right) (\mathbf{x}'_i \boldsymbol{\beta} + \sigma_u \lambda_i) \quad [16.8]$$

donde:

$$\lambda_i = \frac{\phi(\mathbf{x}'_i \boldsymbol{\beta} / \sigma_u)}{\Phi(\mathbf{x}'_i \boldsymbol{\beta} / \sigma_u)}$$

mientras que, para las observaciones censuradas, se aplicarían los resultados de las secciones previas.

En cuanto a los efectos de las variables explicativas, su medida depende de la cuestión que se estudie. Así, se tiene:

$$\frac{\partial E(y/\mathbf{x})}{\partial \mathbf{x}} = \boldsymbol{\beta}$$

mientras que:

$$\frac{\partial E(y^*/\mathbf{x})}{\partial \mathbf{x}} = \boldsymbol{\beta} \cdot \Phi \left(\frac{\mathbf{x}'\boldsymbol{\beta}}{\sigma_u} \right)$$

16.8.a. Estimación de mínimos cuadrados

La esperanza condicional de la variable censurada y_i^* es no lineal:

$$E(y_i^*/\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma_u \phi_i + (1 - \Phi_i)\mathbf{x}_i'\boldsymbol{\beta}$$

por lo que la estimación MCO de y_i^* sobre \mathbf{x}_i incurre en un sesgo de variables omitidas. Un procedimiento consistente de estimación propuesto por Heckman (1976), (1979) introduce una variable ficticia z_i que toma el valor 1 si $y_i > 0$ (no hay censura), y 0 en caso contrario.

Si consideramos la esperanza condicional del modelo de regresión truncada, que es el que aplica a las observaciones no censuradas de y_i :

$$E[y_i/\mathbf{x}_i, z_i = 1] = \mathbf{x}_i'\boldsymbol{\beta} + \sigma_u \lambda_i$$

podríamos estimar consistentemente $\boldsymbol{\beta}$ y σ_u si λ_i fuese observable, pero no lo es. Ahora bien, el modelo de censura implica:

$$P[z_i = 1] = P[y > 0] = \Phi \left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma_u} \right)$$

y si denotamos $\boldsymbol{\gamma} = \frac{\boldsymbol{\beta}}{\sigma_u}$, tenemos:

$$a) \quad P[z_i = 1] = \Phi(\mathbf{x}_i'\boldsymbol{\gamma}); \quad P[z_i = 0] = 1 - \Phi(\mathbf{x}_i'\boldsymbol{\gamma}).$$

$$b) \quad \lambda_i = \frac{\phi(\mathbf{x}_i'\boldsymbol{\gamma})}{\Phi(\mathbf{x}_i'\boldsymbol{\gamma})}.$$

Como muestra *a)*, tenemos un modelo probit para explicar la censura de observaciones, en el que podemos estimar $\boldsymbol{\gamma}$. A partir de él obtenemos ϕ_i , Φ_i y λ_i y estimamos $\boldsymbol{\beta}$ y σ_u mediante una regresión de las observaciones *no censuradas* de y_i sobre \mathbf{x}_i y los valores ajustados $\hat{\lambda}_i$.

También podríamos obtener $\hat{t}_i = [\hat{\phi}_i - (1 - \hat{\Phi}_i)\hat{\alpha}_i]$ para *todas* las observaciones y estimar una regresión de y_i sobre \mathbf{x}_i y \hat{t}_i . Por último, como de [16.8] se tiene:

$$E[y_i/\mathbf{x}_i] = (\Phi_i \mathbf{x}_i)'\boldsymbol{\beta} + \sigma_u \phi_i$$

podríamos estimar una regresión de y_i sobre el vector $\Phi_i \mathbf{x}_i$ y la variable ϕ_i para obtener $\boldsymbol{\beta}$ y σ_u .

16.8.b. Estimación de máxima verosimilitud

Consideremos nuevamente el modelo:

$$y_i = \begin{cases} \mathbf{x}'_i \boldsymbol{\beta} + u_i & \text{si } \mathbf{x}'_i \boldsymbol{\beta} + u_i \geq 0 \\ 0 & \text{si } < 0 \end{cases}$$

Sea S el conjunto de observaciones en los que se observa $y_i = 0$. Suponemos que el término de error u_i es Normal $(0, \sigma_u^2)$ y denotamos por ϕ y Φ sus funciones de densidad y de distribución, respectivamente. La función de verosimilitud del problema es:

$$L(u_1, u_2, \dots, u_N) = \prod_S [1 - \Phi(\mathbf{x}'_i \boldsymbol{\beta}; \sigma_u)] \cdot \prod_{S^c} \phi(y_i - \mathbf{x}'_i \boldsymbol{\beta}; \sigma_u)$$

es decir:

$$\ln L = \text{Constante} + \sum_S \ln [1 - \Phi_i] - \frac{N - N_S}{2} \ln \sigma_u^2 - \sum_{S^c} \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{2\sigma_u^2}$$

donde Φ_i y ϕ_i denotan $\Phi(\mathbf{x}'_i \boldsymbol{\beta}; \sigma_u)$ y $\phi(\mathbf{x}'_i \boldsymbol{\beta}; \sigma_u)$ y N_S es el número de observaciones censuradas. Si utilizamos las expresiones de las derivadas

$$\frac{\partial \Phi_i}{\partial \boldsymbol{\beta}} = \phi_i \mathbf{x}_i$$

$$\frac{\partial \phi_i}{\partial \boldsymbol{\beta}} = -\phi_i \frac{\mathbf{x}'_i \boldsymbol{\beta}}{\sigma_u^2} \mathbf{x}_i$$

$$\frac{\partial \Phi_i}{\partial \sigma_u^2} = -\frac{1}{2\sigma_u^2} (\mathbf{x}'_i \boldsymbol{\beta}) \phi_i$$

$$\frac{\partial \phi_i}{\partial \sigma_u^2} = \frac{\mathbf{x}'_i \boldsymbol{\beta} - \sigma_u^2}{2\sigma_u^4} \phi_i$$

se tiene el gradiente:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = -\sum_S (1 - \Phi_i)^{-1} \phi_i \mathbf{x}_i + \frac{1}{\sigma_u^2} \sum_{S^c} (y_i - \mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}_i = \mathbf{0}_k \quad [16.9]$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{1}{2\sigma_u^2} \sum_S (\mathbf{x}'_i \boldsymbol{\beta}) \frac{\phi_i}{1 - \Phi_i} - \frac{1}{2\sigma_u^2} (N - N_S) + \frac{1}{2\sigma_u^4} \sum_{S^c} (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 = 0 \quad [16.10]$$

y derivadas segundas:

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \sum_S \frac{\phi_i}{1 - \Phi_i} \frac{\mathbf{x}_i' \boldsymbol{\beta}}{\sigma_u^2} \mathbf{x}_i \mathbf{x}_i' - \sum_S \frac{\phi_i^2}{(1 - \Phi_i)^2} \mathbf{x}_i \mathbf{x}_i' - \frac{1}{\sigma_u^2} \sum_{S^c} \mathbf{x}_i \mathbf{x}_i' = \\ &= - \sum_S \frac{\phi_i}{(1 - \Phi_i)^2} \left[\phi_i - \frac{1}{\sigma_u^2} (1 - \Phi_i) (\mathbf{x}_i' \boldsymbol{\beta}) \right] \mathbf{x}_i \mathbf{x}_i' - \frac{1}{\sigma_u^2} \sum_{S^c} \mathbf{x}_i \mathbf{x}_i' \end{aligned}$$

A partir de [16.10] podríamos obtener:

$$\hat{\sigma}_u^2 = \frac{\sum_{S^c} (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2}{(N - N_S) - \sum_S \frac{\phi_i}{1 - \Phi_i} (\mathbf{x}_i' \boldsymbol{\beta})} \quad [16.11]$$

para posteriormente utilizar un algoritmo Newton-Raphson puesto que, dado un $\boldsymbol{\beta}_i$ inicial en cada iteración, calcularíamos ϕ_i y Φ_i para obtener $\hat{\sigma}_i^2$ a partir de [16.11] y el algoritmo nos llevaría a la nueva estimación, $\boldsymbol{\beta}_{i+1}$.

16.8.c. Procedimiento de Fair

Un tercer procedimiento de estimación ha sido propuesto por Fair (1977), quien observó que multiplicando [16.9] por $\frac{\boldsymbol{\beta}}{2\sigma_u^2}$ y sumándolo a [16.10] se tiene:

$$\hat{\sigma}_{MV}^2 = \frac{\sum_{S^c} (y_i - \mathbf{x}_i' \boldsymbol{\beta}) y_i}{N - N_S} \quad [16.12]$$

mientras que multiplicando [16.9] por σ_u^2 se tiene:

$$- \sum_S \frac{\phi_i}{1 - \Phi_i} \sigma_u^2 \mathbf{x}_i + \sum_{S^c} (y_i - \mathbf{x}_i' \boldsymbol{\beta}) \mathbf{x}_i = \mathbf{0}_k$$

que puede escribirse:

$$\hat{\boldsymbol{\beta}}_{MV} = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} (\mathbf{X}'_1 \mathbf{Y}_1 + \mathbf{X}'_2 \mathbf{V}) \quad [16.13]$$

donde \mathbf{X}_1 , \mathbf{Y}_1 denotan las matrices de observaciones en los casos en que se

observa y_i , mientras que X_2 corresponde a las observaciones censuradas. El vector V , de dimensión $N_s \times 1$, tiene por componente genérica: $-\sigma_u^2 \frac{\phi_i}{1 - \Phi_i}$.

El sistema [16.12]-[16.13] puede utilizarse como un procedimiento de estimación en dos etapas. Para interpretar este método, obsérvese que las componentes de V son las esperanzas⁽¹⁾ $E(u_i/u_i \leq -x_i'\beta)$, es decir: $E(u_i/y_i = 0)$, de modo que el último sumando de [16.13] corrige el sesgo que podría aparecer si no se tuviera en cuenta dicha esperanza condicional. Si no hubiese ninguna observación censurada, se tendría $V = Y_2$ y [16.13] se convertiría en el estimador MCO obtenido con todas las observaciones:

$$(X_1' X_1 + X_2' X_2)\beta = X_1' Y_1 + X_2' Y_2$$

Al no observar $X_2\beta + u_2$, lo sustituimos por $X_2\beta + E[u_2/u_2 \leq -X_2\beta]$, obteniendo:

$$(X_1' X_1 + X_2' X_2)\beta = X_1' Y_1 + X_2' (X_2\beta + V)$$

es decir [16.13]. Fair sugirió comenzar este procedimiento iterativo a partir de $\beta^0 = 0$ si existe un elevado número de observaciones nulas, o del estimador MCO si existe un alto número de observaciones censuradas.

Si en vez de tener un umbral de censura igual a 0 se tuviese

$$y_i = \begin{cases} x_i'\beta + u_i & \text{si } y_i > c_i \\ c_i & \text{si } y_i \leq c_i \end{cases}$$

el cambio de variables $y_i^* = y_i - c_i$ y los vectores ampliados $x_i^* = (x_i, c_i)$, $\beta^* = (\beta, -1)$ nos ponen en la situación del modelo que aquí hemos analizado.

Si observamos y_i sólo si $y_i < 0$, multiplicando tanto y_i como el vector x_i por -1 , volvemos de nuevo al modelo propuesto.

⁽¹⁾ Puesto que:

$$\begin{aligned} E(u_i/u_i \leq -x_i'\beta) &= \sigma_u E\left(\frac{u_i}{\sigma_u} \middle| \frac{u_i}{\sigma_u} \leq -\frac{x_i'\beta}{\sigma_u}\right) = \sigma_u \left(-\sigma_u \frac{\phi\left(-\frac{x_i'\beta}{\sigma_u}\right)}{\Phi\left(-\frac{x_i'\beta}{\sigma_u}\right)} \right) = \\ &= -\sigma_u^2 \frac{\phi\left(\frac{x_i'\beta}{\sigma_u}\right)}{1 - \Phi\left(\frac{x_i'\beta}{\sigma_u}\right)} \end{aligned}$$

16.9. MODELOS CON SELECCION MUESTRAL

La selección muestral aparece cuando la inclusión de una unidad económica en la muestra depende de decisiones previamente tomadas por dicha unidad, por lo que la muestra no puede considerarse aleatoria. Este es el caso de un análisis del salario percibido por las mujeres trabajadoras, pues sólo podrán informar acerca de su salario aquellas mujeres que hayan decidido trabajar por recibir un salario superior a su «salario de reserva».

Este modelo consta de una ecuación de selección:

$$z_i = \mathbf{w}'_i \boldsymbol{\delta} + \varepsilon_i$$

y una ecuación en la que se centra el interés del investigador:

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i$$

en la que se observa y_i tan sólo si $z_i > 0$. Suponemos que ε_i y u_i son Normales, con coeficiente de correlación ρ .

Supongamos que las S primeras observaciones provienen de las personas cuya variable y_i fue observable. Tenemos:

$$E[y_i/\mathbf{x}_i, \text{regla de selección}] = \mathbf{x}'_i \boldsymbol{\beta} + E[u_i/\text{regla de selección}]$$

es decir:

$$E[y_i/\mathbf{x}_i, z_i \geq 0] = \mathbf{x}'_i \boldsymbol{\beta} + E[u_i/\varepsilon_i \geq -\mathbf{w}'_i \boldsymbol{\delta}] \quad [16.14]$$

Si ρ fuese cero, entonces la esperanza condicional de u_i sería cero y el proceso de selección sería puramente aleatorio, por lo que MCO aplicado a [16.14] sería consistente. En caso contrario, la esperanza condicional de u_i depende de las variables \mathbf{w}_i , que podrían ser significativas si se incluyesen en la regresión MCO de [16.14] como explicativas.

Al omitir estos factores, la estimación MCO de [16.14], utilizando los datos correspondientes a los casos en que y_i fue observada, será inconsistente.

Como

$$E[u_i/\varepsilon_i \geq -\mathbf{w}'_i \boldsymbol{\delta}] = \rho \sigma_u \lambda_i$$

donde

$$\lambda_i = \phi_i / (1 - \Phi_i) \quad \text{y} \quad \phi_i = \phi \left(\frac{\mathbf{w}'_i \boldsymbol{\delta}}{\sigma_\varepsilon} \right); \quad \Phi_i = \Phi \left(\frac{\mathbf{w}'_i \boldsymbol{\delta}}{\sigma_\varepsilon} \right)$$

basta incluir este término en [16.14] para tener consistencia en el estimador mínimo-cuadrático. Sin embargo, el modelo resultante tiene heteroscedastici-

dad y habría que estimar por MCG. En realidad, habría que tener en cuenta que

$$E[\varepsilon_i/\varepsilon_i \geq -\mathbf{w}'_i\boldsymbol{\delta}] = \sigma_\varepsilon\lambda_i$$

para especificar la ecuación:

$$z_i = \mathbf{w}'_i\boldsymbol{\delta} + \sigma_\varepsilon\lambda_i + \varepsilon_i \quad [16.15]$$

que se estimaría, simultáneamente con [16.14], por MCG, utilizando sólo las observaciones con $z_i > 0$.

Para ello, hay que llevar a cabo una primera etapa en la que se utiliza un modelo probit que discrimina las observaciones según que se haya observado y_i o no, del mismo modo que propusimos en el modelo con variable dependiente censurada. De este modo obtendríamos $\boldsymbol{\delta}/\sigma_\varepsilon$ y luego ϕ_i y λ_i , que llevadas a [16.14] generarán estimaciones consistentes del producto $\rho\sigma_u$ y el vector $\boldsymbol{\beta}$.

Apéndice: APROXIMACION LINEAL AL MODELO PROBIT

El desarrollo en serie de Taylor de la función $\Phi^{-1}(p_i)$ o $\Phi^{-1}(P_i + u_i)$ alrededor del punto P_i (probabilidad poblacional, desconocida) es:

$$\Phi^{-1}(p_i) \cong \Phi^{-1}(P_i) + \frac{d\Phi^{-1}(P_i)}{dP_i} u_i$$

donde hemos ignorado los términos de orden superior a 2. Ahora bien, como la función de distribución $\Phi: \mathbb{R} \longrightarrow [0,1]$ tiene una función inversa $\Phi^{-1}: [0,1] \longrightarrow \mathbb{R}$ bien definida, con $\Phi^{-1}(P_i) = \mathbf{x}'_i\boldsymbol{\beta}$, el teorema de la función inversa garantiza que:

$$\frac{d\Phi^{-1}(P_i)}{dP_i} = \frac{1}{\Phi'(\Phi^{-1}(P_i))}$$

donde Φ' denota la derivada de Φ , lo que sugiere una regla del tipo: «la derivada de la función inversa es el inverso de la derivada de la función original», excepto por el hecho de que el dominio de una función es el rango de la otra. Por eso es que en la expresión anterior, mientras el argumento de la derivada $d\Phi^{-1}/dP_i$ es P_i , con valores en $[0,1]$, el argumento de la derivada Φ' es $\Phi^{-1}(P_i)$, con valores en todo \mathbb{R} .

Por consiguiente:

$$\frac{d\Phi^{-1}(P_i)}{dP_i} = \frac{1}{f(\Phi^{-1}(P_i))} = \frac{1}{f(\mathbf{x}'_i\boldsymbol{\beta})}$$

donde f denota la función de densidad de la Normal (0,1) y se tiene finalmente:

$$\Phi^{-1}(p_i) \cong \Phi^{-1}(P_i) + \frac{1}{f(\mathbf{x}'_i\boldsymbol{\beta})} u_i = \mathbf{x}'_i\boldsymbol{\beta} + \frac{1}{f(\mathbf{x}'_i\boldsymbol{\beta})} u_i$$

APROXIMACION LINEAL AL MODELO LOGIT

De modo similar, podemos aproximar el valor de la inversa de la función logística en el punto $p_i = P_i + u_i$, alrededor del punto P_i :

$$F^{-1}(p_i) = F^{-1}(P_i + u_i) = F^{-1}(P_i) + \frac{dF^{-1}(P_i)}{dP_i} u_i$$

es decir:

$$\ln \frac{p_i}{1-p_i} = \ln \frac{P_i + u_i}{1-(P_i + u_i)} \cong \ln \frac{P_i}{1-P_i} + \frac{1}{P_i(1-P_i)} u_i$$

donde hemos sustituido la derivada de $F^{-1}(P_i)$ por el inverso de la derivada de F , evaluada en $F^{-1}(P_i)$, es decir:

$$\frac{dF^{-1}(P_i)}{dP_i} = \frac{1}{f(F^{-1}(P_i))} = \frac{1}{P_i(1-P_i)}$$

utilizando la propiedad ya vista en la Sección 16.4.b de que, en la distribución logit: $f(z) = F(z)(1-F(z))$.

PROBLEMAS

Problema 16.1. Demostrar que las estimaciones de mínimos cuadrados ordinarios de los modelos lineales de probabilidad en el caso de alternativas múltiples:

$$P_{i1} = \alpha_1 + \beta_1 x_i + u_{1i}$$

$$P_{i2} = \alpha_2 + \beta_2 x_i + u_{2i}$$

$$P_{i3} = \alpha_3 + \beta_3 x_i + u_{3i}$$

satisfacen $\sum_1^3 \alpha_i = 1$, $\sum_i^3 \beta_i = 0$.

Problema 16.2. ¿Qué ocurriría con los coeficientes del modelo lineal de probabilidad si la variable endógena se define como tomando los valores 0 ó 2, en vez de 0 ó 1? ¿Que sugiere esta observación acerca de la interpretación de los coeficientes estimados por mínimos cuadrados ordinarios?

Problema 16.3. ¿Por qué no se utilizan los residuos de mínimos cuadrados ordinarios para estimar la varianza residual σ_i^2 en el modelo lineal de probabilidad y así corregir la heteroscedasticidad del término de error?

Problema 16.4. Justificar las expresiones dadas en el capítulo acerca de la sensibilidad de las probabilidades de P_i ante variaciones en las variables explicativas, en el caso de los modelos lineal de probabilidad, probit y logit.

Problema 16.5. Obtener la matriz de información $I(\beta)$ del modelo probit que aparece en la Sección 16.4.a.

Problema 16.6. Demostrar que tanto en el modelo probit como en el logit, si denotamos por L_0 el valor de la función de verosimilitud que se obtiene cuando sólo se estima una constante, restringiendo los demás coeficientes a cero, se tiene:

$$-\ln L_0 = N[p \ln p + (1 - p) \ln (1 - p)]$$

donde p es la proporción de observaciones muestrales con $Y_i = 1$. ¿Puede extenderse este resultado a otros modelos?

Problema 16.7. Tras entrevistar a una muestra de familias con ingresos superiores a los 8 millones de pesetas, se pudo afirmar que «la típica familia rica española recibe unos ingresos de 10 millones de pesetas», habiendo sólo un 3 por 100 de dichas familias. Si se supone que la renta tiene una distribución lognormal, ¿qué podría inferirse acerca de los ingresos medios y su varianza para el total de las familias españolas?

Problema 16.8. Un estadio de fútbol, con capacidad para 60.000 espectadores, se ha llenado el 25 por 100 de los partidos en él celebrados a lo largo de la competición, con una cifra media de asistencia (incluidos los partidos en que se llenó el estadio) de 40.000 espectadores. ¿Cómo estimaría la media y varianza de la demanda de localidades?

¿Cómo cambiaría su respuesta si la cifra de 40.000 espectadores no incluyese los partidos en que se llenó el estadio?

Problema 16.9. Suponga que tiene un modelo lineal en el que la variable endógena se observa tan sólo si toma valores inferiores o iguales a un cierto umbral A .

Escriba la función de verosimilitud del problema y explique cómo podría llevar a cabo la estimación del modelo.

Problema 16.10. Un banco pretende caracterizar las empresas que cumplen puntualmente todos los plazos de devolución de los créditos que reciben. Tras crear una variable Y_i que toma valor «1» para las empresas que cumplieron dichos compromisos y «0» para aquellas que incumplieron algún pago, se estimó el siguiente modelo:

$$(1) \quad Y_i = 0,69 - 0,018x_{1i}^2 + 0,049x_{2i} + 0,057x_{3i}$$

$$(0,18) \quad (0,002) \quad (0,049) \quad (0,018)$$

$$R^2 = 0,69$$

donde:

$$x_{1i} = \frac{\text{Valor nominal de la deuda viva de la empresa}}{\text{Valor total del activo}} \times 100$$

$$x_{2i} = \frac{\text{Beneficios después de impuestos}}{\text{Valor total del activo}} \times 100$$

x_{3i} = Valor del activo, como indicador del tamaño de la empresa

a) Interprete los valores estimados en (1) para los parámetros del modelo, incluyendo el valor de R^2 .

b) ¿Cómo cree que habrá estimado el servicio de estudios del banco modelo (1)? ¿Qué propiedades tiene dicho estimador?

c) Posteriormente, dicho servicio de estudios estima por máxima verosimilitud una formulación logit del modelo, obteniendo:

$$(2) \quad \ln \frac{P_i}{1 - P_i} = -1,66 - 0,32x_{1i}^2 + 0,62x_{2i} - 0,90x_{3i}$$

(1,19) (0,06) (0,14) (0,22)

¿A qué puede deberse que varíen tanto los coeficientes estimados por ambos procedimientos?

d) Para una empresa con $x_{1i}^2 = 9,7\%$, $x_{2i} = 7,8\%$ y $x_{3i} = 0,6$ se tiene un valor del miembro derecho de la ecuación (2) de $-0,47$. ¿Qué interpretación tiene dicho valor numérico?

e) Explicar en detalle cómo se habría estimado la formulación logit y cómo se habrían obtenido las desviaciones típicas en (2).

Problema 16.11. En una encuesta realizada en junio de 1987 a diez alumnos de 5.º curso se les preguntó si aprobaron o no la asignatura de Macroeconomía, así como la calificación que obtuvieron en la asignatura de Econometría, con los siguientes resultados:

Aprobaron Macroeconomía	Calificación Econometría
Sí	8
Sí	8
No	6
Sí	6
No	6
No	5
Sí	5
Sí	4
No	4
No	4

a) Especificar y estimar por mínimos cuadrados ordinarios un modelo lineal que evalúe el efecto que la calificación de Econometría tiene sobre la probabilidad de aprobar Macroeconomía.

- b) Interpretar los valores estimados para cada uno de los coeficientes del modelo. ¿Cuál es la calificación que debe alcanzarse en Econometría para tener una probabilidad de 0,80 de aprobar Macroeconomía?
- c) Obtener una estimación eficiente del modelo especificado en a).
- d) Comprobar que la probabilidad de aprobar Macroeconomía a un alumno que obtuvo 9,5 en Econometría excede la unidad. ¿Existe alguna especificación que evite este tipo de problemas?

Problema 16.12. Para llevar a cabo un análisis que determine el grado de competitividad de las empresas españolas se ha hecho una encuesta acerca de sus gastos de investigación y desarrollo (clasificados como altos, medios o bajos) y de su tamaño, medido por el número de empleados (grande, mediana, pequeña), clasificando tales empresas como exportadoras (si venden en el exterior al menos un 20 por 100 de sus ventas totales) o no exportadoras (en caso contrario).

El valor medio de los gastos en inversión y desarrollo es del 20 por 100 de los gastos totales para las empresas con inversión alta, del 10 por 100 para aquellas con inversión media y del 5 por 100 para las de inversión baja, por lo que se toman dichos porcentajes como valores representativos de la variable «gastos en I + D» del grupo.

El número medio de empleados es de 50 en las empresas pequeñas, 250 en las medianas y 1.500 en las grandes que se toman, de nuevo, como valores representativos de la variable «tamaño» de cada empresa.

a) Si definimos una variable Y_i que toma el valor 1 para aquellas empresas que comercializan en el exterior más de un 20 por 100 de sus ventas totales y es igual a cero en caso contrario, estimar un modelo logit que analice la influencia del tamaño de las empresas y la importancia de sus gastos en «investigación y desarrollo» sobre su posible carácter exportador:

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

donde $Y_i = 1$ si exporta, $Y_i = 0$ si no exporta, x_{1i} = gastos en I + D, x_{2i} = tamaño.

Para ello se dispone de encuestas procedentes de 1.410 empresas, clasificadas en nueve grupos:

Grupo I [gastos altos ($x_1 = 0,20$), empresa grande ($x_2 = 1.500$)].

Grupo II (altos, mediana), Grupo III (altos, pequeña).

Grupo IV (medios, grande), Grupo V (medios, mediana).

Grupo VI (medios, pequeña), Grupo VII (bajos, grande).

Grupo VIII (bajos, mediana), Grupo IX (bajos, pequeña).

Con los siguientes resultados:

Grupo	I	II	III	IV	V	VI	VII	VIII	IX
N_i	100	200	50	100	400	260	50	50	200
n_i	75	100	25	20	40	20	5	10	8

donde N_i denota el número de empresas de cada grupo y n_i es el número de ellas que pueden clasificarse como exportadoras.

b) ¿Cómo podría utilizarse tal estimación para comenzar la estimación de máxima verosimilitud del modelo? Describir dicho proceso de estimación en todo detalle.

Problema 16.13. Obtener el estadístico para el contraste de multiplicadores de Lagrange de la hipótesis

$$H_0: \beta_2 = 0$$

en el modelo logit:

$$P_i = F(\mathbf{x}'_1 \beta_1 + \mathbf{x}'_2 \beta_2)$$

¿Puede utilizarse un estadístico similar para contrastar la misma hipótesis en un modelo probit?

Problema 16.14. Obtenga el estadístico de los multiplicadores de Lagrange para el contraste de la hipótesis nula de homoscedasticidad, frente a la alternativa

$$\text{Var } u_i = \sigma_i^2 = e^{\gamma' z_i}$$

en el modelo logit.

Problema 16.15. Explique en detalle cómo llevar a cabo contrastes de hipótesis sobre los coeficientes de variables explicativas en los modelos logit y probit.

CAPITULO 17

MODELOS DE ECUACIONES SIMULTANEAS

I. ESPECIFICACION E IDENTIFICACION

17.1. ESPECIFICACION DEL MODELO DE ECUACIONES SIMULTANEAS

En la mayoría de los modelos econométricos de los capítulos anteriores hemos mantenido el supuesto de que las variables explicativas eran exógenas. Dicho de otro modo, suponíamos que el estudio de su comportamiento podía llevarse a cabo con total independencia del análisis de la variable y_t . Estas variables están incorrelacionadas con el componente puramente aleatorio de y_t (el término de error del modelo), en ambas direcciones temporales.

En los Capítulos 9 y 13 relajamos algo tal supuesto, pues consideramos la inclusión de retardos de y_t como variables explicativas, y dichos retardos son de naturaleza aleatoria. Sin embargo, su valor realizado es conocido en el instante t y, por ello, reciben el nombre de variables *predeterminadas* con respecto a y_t . Estas variables están incorrelacionadas con los valores presente y futuros de la componente puramente aleatoria de y_t (aunque no con sus valores pasados) si ésta no presenta autocorrelación. Si u_t presenta autocorrelación, la mencionada ausencia de correlación entre variables no está garantizada, por lo que no todo retardo de y_t es entonces una variable *predeterminada*.

Tanto las variables exógenas como aquellas que son realmente predeterminadas en un cierto modelo, comparten la característica de que, aunque influyen sobre y_t , no son influidas por esta variable.

Esta característica se pierde cuando hay que recoger en un modelo econométrico la existencia de un conjunto de variables endógenas que se determinan mutuamente. Es preciso especificar entonces un *modelo de ecuaciones simultáneas*, formado por tantas ecuaciones como variables endógenas se determinan simultáneamente.

En él, cada una de las variables endógenas quedará determinada por las otras variables endógenas, así como por las variables exógenas del modelo y por los términos de error del mismo (tantos como variables endógenas se

determinan simultáneamente). Este es el tipo de modelo econométrico que vamos a analizar en este capítulo y el siguiente.

La principal dificultad que puede plantear la presencia de variables explicativas aleatorias es que el estimador de mínimos cuadrados ordinarios no sea insesgado. Dicha propiedad se basa en que $E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \mathbf{0}_k$. Como se vio en el Capítulo 3, el supuesto de que las variables explicativas sean deterministas hace que dichas variables no sean un problema para obtener la igualdad anterior.

Por otra parte, en el Capítulo 9 vimos que la presencia de retardos de la variable endógena es un obstáculo para obtener esta igualdad.

Como veremos en el Capítulo 18, la presencia como variables explicativas, de variables aleatorias que se determinan simultáneamente con la variable que se pretende analizar, hace asimismo que el estimador de mínimos cuadrados sea sesgado.

Consideremos una versión simplificada del modelo de consumo keynesiano:

$$\begin{aligned} C_t &= \beta_1 + \beta_2 Y_t + u_t \\ Y_t &= C_t + I_t \end{aligned} \quad [17.1]$$

donde C_t , Y_t , I_t denotan, respectivamente, el consumo agregado, el producto nacional y la inversión. La primera ecuación del modelo es la función de consumo, que explica a éste como función lineal del producto en la economía. La segunda ecuación de este modelo es la identidad contable del país, supuesta una economía cerrada y sin gasto público. El modelo considera la inversión como una variable determinada exógenamente, es decir, sin influencia del consumo o del nivel de producto. Las dos variables determinadas de forma endógena en el modelo son el consumo agregado y el nivel de producto.

Examinemos las implicaciones del modelo: Un alto valor de la perturbación de la ecuación de consumo en un período determinado implicará un flujo de consumo superior a su propio valor medio. A través de la identidad contable, éste implicará, a su vez, un valor del producto superior a su promedio. Este incremento en el producto generaría, a su vez, un incremento adicional en el consumo. Cuando se estima la función de consumo anterior, el coeficiente estimado β_2 atribuirá al producto los dos incrementos (el directo y el indirecto), que se producirían sobre el consumo, como consecuencia de un incremento en el término de error u_t . La consecuencia sería, pues, un sesgo al alza en la estimación de la propensión marginal al consumo.

Así vemos que en un modelo de ecuaciones simultáneas: a) se recogen interacciones entre variables, bastante más complejas que las que aparecen en un modelo uniecuacional, y b) la estimación aislada de una de las ecuaciones del modelo por los procedimientos habituales en los capítulos anteriores introducirá, en general, sesgos en los coeficientes estimados.

La representación analítica de un modelo genérico de ecuaciones simultáneas es:

$$\mathbf{y}_t'\boldsymbol{\Gamma} + \mathbf{x}_t'\mathbf{B} + \mathbf{u}_t' = \mathbf{0}_g, \quad t = 1, 2, \dots, T \quad [17.2]$$

Por último, la tercera condición admite la existencia de correlaciones contemporáneas (es decir, en el mismo instante de tiempo) entre los términos de error de dos ecuaciones diferentes, reflejadas en los parámetros σ_{ij} de [17.3]. La hipótesis incluye el supuesto adicional de que dichas correlaciones, si existen, no cambian en el tiempo, lo que constituye una generalización del concepto de homoscedasticidad. El hecho de que existan correlaciones contemporáneas no nulas es la esencia del modelo de ecuaciones simultáneas y, concretamente, una de las razones que nos lleva a considerar dicho conjunto de ecuaciones como un solo modelo econométrico, más que como una colección de modelos uniecuacionales aislados.

Una particularidad de un modelo de ecuaciones simultáneas es que un cambio en la perturbación aleatoria de una ecuación afecta a *todas* las variables endógenas del modelo, puesto que todas ellas se determinan simultáneamente. La simultaneidad proviene de dos características diferentes del modelo: a) que, en principio, *todas* las variables endógenas pueden aparecer en *todas* las ecuaciones del modelo y b) los términos de error de todas las ecuaciones están correlacionados contemporáneamente. En tanto en cuanto los términos de error de las distintas ecuaciones estén correlacionados, un cambio en la perturbación u_{jt} de la ecuación que define a la variable endógena y_{jt} vendrá asociado con otro cambio en la perturbación aleatoria u_{it} de la ecuación que define a la variable y_{it} .

Pudiera ocurrir que, en algún caso especial, el investigador estuviese dispuesto a imponer a priori la restricción de que en cada ecuación apareciese una sola variable endógena, aquella que está en su miembro izquierdo. Incluso en estas condiciones se tendría la interrelación a que nos hemos referido, siempre que $\Sigma \neq \mathbf{I}_g$. Este es el tipo de modelo estudiado en el Capítulo 8.

Debido a estas dos razones, un modelo de ecuaciones simultáneas es diferente de una colección de modelos uniecuacionales. Ni siquiera cuando los términos de error de las distintas ecuaciones son independientes entre sí se tiene tal equivalencia. Para que un modelo de ecuaciones simultáneas sea equivalente al conjunto de modelos uniecuacionales que lo componen, es preciso que tenga una estructura de *modelo recursivo*, que analizamos en el capítulo siguiente. La independencia de los términos de error es tan sólo una condición necesaria para que un sistema de ecuaciones tenga dicha estructura.

Denotemos por \mathbf{Y} la matriz $T \times g$ de observaciones de las variables endógenas y_{it} . Cada columna de esta matriz es análoga al **vector** \mathbf{y} del modelo econométrico uniecuacional, e incluye los valores muestrales de *una* de las variables endógenas del modelo. Del mismo modo, denotemos por \mathbf{X} la matriz $T \times k$ de las observaciones de todas las variables exógenas del modelo, y por \mathbf{U} la matriz $T \times g$ de los términos de error de todas las ecuaciones del modelo. Con esta notación, el modelo de ecuaciones simultáneas [17.2] puede escribirse:

$$\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B} + \mathbf{U} = \mathbf{0}_{T \times g} \quad [17.4]$$

A pesar de que esta notación matricial generaliza la notación similar del caso uniecuacional, cabe observar que la matriz de observaciones de las

variables exógenas es, como en el caso uniecuacional, de dimensión $T \times k$. Ello se debe a que todas las ecuaciones comparten la misma matriz de observaciones de las variables exógenas. Por otra parte, el lector debe comprobar que en el caso $g = 1$ el modelo que acabamos de describir se reduce al caso uniecuacional que hemos desarrollado en temas anteriores:

$$y_t \Gamma + x_t' \mathbf{B} + u_t = y_t \Gamma + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + u_t = 0 \quad [17.5]$$

donde Γ es ahora un escalar, \mathbf{B} un vector columna de k componentes, \mathbf{X} una matriz $T \times k$, mientras que \mathbf{U} y $\mathbf{0}$ son vectores columna de dimensión T .

Si suponemos que el vector u_t formado por los términos de error de las g ecuaciones en un cierto instante de tiempo sigue una distribución Normal g -variante: $N_g(\mathbf{0}_g, \Sigma)$, entonces la función de densidad conjunta del vector $\mathbf{u} = (u_1, u_2, \dots, u_T)$ es:

$$L(\mathbf{u}) = \prod_1^T f(u_t) = \prod_1^T \left(\frac{1}{2\pi} \right)^{g/2} |\Sigma|^{-1/2} \exp \left(-\frac{u_t' \Sigma^{-1} u_t}{2} \right)$$

Para construir la función de verosimilitud del vector $\mathbf{y} = (y_1, y_2, \dots, y_T)$ condicional en las observaciones muestrales de las variables endógenas, es preciso determinar el jacobiano de la transformación del vector u_t en el vector y_t . Del modelo se obtiene directamente que $\frac{\partial u_t}{\partial y_t} = \Gamma'$, por lo que se tiene la función de verosimilitud:

$$L(y_1, y_2, \dots, y_T) = \prod_{t=1}^T \frac{1}{(2\pi)^{g/2}} \cdot \frac{1}{|\Sigma|^{1/2}} \cdot |\Gamma| \cdot \exp \left[-\frac{1}{2} (y_t' \Gamma + x_t' \mathbf{B}) \Sigma^{-1} (y_t' \Gamma + x_t' \mathbf{B})' \right]$$

o también:

$$L(y_1, y_2, \dots, y_T) = (2\pi)^{-gT/2} |\Sigma|^{-T/2} |\Gamma|^T \exp \left[-\frac{1}{2} \text{traza } \Sigma^{-1} (\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B})' (\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B}) \right]$$

donde se ha utilizado la propiedad:

$$\Sigma_1' (y_t' \Gamma + x_t' \mathbf{B}) \Sigma^{-1} (y_t' \Gamma + x_t' \mathbf{B}) = \text{traza} [(\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B}) \Sigma^{-1} (\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B})']$$

17.2. FORMAS ESTRUCTURAL Y REDUCIDA

El problema que nos planteamos en este capítulo es el de estimación de los parámetros del modelo [17.4]. En total, tenemos g^2 parámetros en la matriz Γ , junto con kg parámetros en \mathbf{B} , y $\frac{g(g+1)}{2}$ parámetros diferentes en la matriz de covarianzas Σ . Ello da un total de $g \left(\frac{3g+1}{2} + k \right)$ parámetros a estimar. El

número de parámetros a estimar se reduce algo cuando observamos que, al igual que ocurría en el modelo uniecuacional, cada ecuación podría multiplicarse por una constante sin que pudiésemos distinguir entre ambas ecuaciones, puesto que la varianza del término de error de dicha ecuación es desconocida. Sólo cuando dicha varianza fuese conocida (lo que nunca ocurre en la práctica) estaríamos a salvo de este problema de cambio de escala.

Adoptamos aquí una actitud similar a la tomada en el caso del modelo uniecuacional, que fue la de *normalizar* cada ecuación de modo que una de las variables endógenas tenga un coeficiente igual a -1 en cada una de ellas. Obsérvese que en el caso uniecuacional de [17.5], este convenio equivale a dividir por la constante Γ cambiada de signo, con lo que resultaría el modelo:

$$y = X\beta + v \quad [17.6]$$

donde $\beta_i = \frac{b_i}{-\Gamma}$, $i = 1, \dots, k$ y $v_t = \frac{u_t}{-\Gamma}$, $t = 1, 2, \dots, T$.

Ya que hay tantas ecuaciones como variables endógenas, podemos asociar a cada variable endógena una ecuación diferente del modelo. Dicha ecuación se divide por el coeficiente de la variable endógena elegida, cambiado de signo, con lo que dicha variable podrá despejarse del mismo modo que hicimos en [17.6], con la única salvedad de que en [17.5] hay otras variables endógenas incluidas como explicativas en cada ecuación.

Esta asociación entre ecuaciones del modelo y variables endógenas puede venir naturalmente dada por el significado económico de dicha ecuación (por ejemplo, en «la función de consumo» el consumo debería ser la variable endógena con coeficiente unidad), o ser una asignación arbitraria en otros casos. De este modo, el número de parámetros a estimar pasa a ser de $\frac{g(3g-1)}{2} + gk$. En general, sin embargo, habrá que estimar menos parámetros

que el número aquí mencionado. Ello se debe a que, en general, al especificar las g ecuaciones del modelo, no habremos incluido en cada una de ellas las g variables endógenas, así como *todas* las variables predeterminadas. Dicho de otra forma, habremos supuesto ya en la especificación del modelo que algunos de los elementos, tanto de la matriz Γ como de la matriz B , son cero.

Consideremos la primera ecuación del modelo: Denotemos por Y_1 la submatriz $T \times (g_1 - 1)$ de Y formada por las observaciones correspondientes a las $g_1 - 1$ variables endógenas ($g_1 \leq g$) incluidas como variables explicativas en esta primera ecuación. Nótese que el número total de variables endógenas en la primera ecuación es g_1 . Una de ellas es la variable a explicar en dicha ecuación, mientras que las restantes $g_1 - 1$ variables endógenas aparecen como explicativas, junto con otras variables predeterminadas.

Sin pérdida de generalidad, supondremos en lo sucesivo que éstas son las g_1 primeras columnas de la matriz Y . Del mismo modo, denotemos por X_1 la submatriz $T \times k_1$ de X formada por las observaciones correspondientes a las variables predeterminadas incluidas en esta ecuación y supongamos asimismo que éstas son las correspondientes a las k_1 primeras columnas de la ma-

triz \mathbf{X} ($k \geq k_1$). Con esta notación, y tras haber normalizado de modo que el coeficiente de la variable y_1 sea igual a uno, la primera ecuación puede escribirse:

$$y_1 = \mathbf{Y}_1 \gamma_1 + \mathbf{X}_1 \beta_1 + \mathbf{u}_1 \quad [17.7]$$

donde γ_1 y β_1 denotan, respectivamente, las primeras columnas de Γ y \mathbf{B} , divididas por el coeficiente de la variable y_1 , cambiado de signo. De los supuestos anteriores acerca de las propiedades del vector formado por los términos de error de las distintas ecuaciones se tiene que para cada ecuación del modelo: $\text{Var}(\mathbf{u}_i) = \sigma_i^2 \mathbf{I}_T$.

Por consistencia con lo expuesto en los capítulos anteriores, dada la estructura que acabamos de ver para la matriz Σ , cabría ahora considerar la estimación por mínimos cuadrados ordinarios de la ecuación [17.7]. Sin embargo, dicha estrategia presenta un cierto número de problemas.

Por ejemplo, recordemos que la demostración que hicimos en el Capítulo 3 acerca de que el estimador MCO es insesgado se basó en el supuesto de que las variables explicativas en dicha ecuación eran deterministas. En cambio, la presencia de algunas variables endógenas como variables explicativas en [17.7] (es decir, su presencia en el vector \mathbf{Y}_1) impide mantener dicho supuesto, y con ello garantizar que el estimador de mínimos cuadrados ordinarios sea insesgado. Por no ser insesgado, el resultado de optimalidad que vimos para dicho estimador en el Capítulo 3 deja de ser válido.

Volvamos ahora a la expresión matricial del sistema de ecuaciones simultáneas [17.4] y supongamos que la matriz Γ es no singular, es decir, $|\Gamma| \neq 0$. En tal caso, existe su inversa Γ^{-1} , y postmultiplicando la ecuación [17.4] por Γ^{-1} se tiene:

$$\mathbf{Y} = -\mathbf{XB}\Gamma^{-1} - \mathbf{U}\Gamma^{-1} \quad [17.8]$$

o también:

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{V} \quad [17.9]$$

donde $\Pi = -\mathbf{B}\Gamma^{-1}$ es una matriz $k \times g$ y $\mathbf{V} = -\mathbf{U}\Gamma^{-1}$ la matriz $T \times g$ de los términos de error del sistema normalizado. Cada fila de la matriz \mathbf{V} es un vector $\mathbf{v}_t' = -\mathbf{u}_t' \Gamma^{-1}$ o, lo que es lo mismo, $\mathbf{v}_t = -(\Gamma^{-1})' \mathbf{u}_t$, que relaciona el vector formado por los términos de error de las g ecuaciones en el modelo original y en el normalizado. La matriz de covarianzas del vector \mathbf{v} , es:

$$\text{Cov}(\mathbf{v}_t) = E(\mathbf{v}_t \mathbf{v}_t') = (\Gamma^{-1})' E(\mathbf{u}_t \mathbf{u}_t') \Gamma^{-1} = (\Gamma^{-1})' \Sigma \Gamma^{-1} = \Omega$$

que es positiva definida, ya que Σ también lo es. Asimismo, una implicación de la ausencia de autocorrelación entre los términos de error de las distintas ecuaciones es que:

$$E(\mathbf{v}_t \mathbf{v}_s') = E[(\Gamma^{-1})' \mathbf{u}_t \mathbf{u}_s' \Gamma^{-1}] = (\Gamma^{-1})' E(\mathbf{u}_t \mathbf{u}_s') \Gamma^{-1} = \mathbf{0}_{g \times g}$$

Si además suponemos que \mathbf{u}_t sigue una distribución Normal $N(\mathbf{0}_g, \Sigma)$, entonces, como consecuencia, el vector \mathbf{v}_t tendrá asimismo distribución Normal $N(\mathbf{0}_g, \Omega)$. La relación entre las matrices de covarianzas del término de error del modelo original y del transformado puede también escribirse:

$$\Sigma = \Gamma' \Omega \Gamma \quad [17.10]$$

La representación alternativa [17.9] del modelo de ecuaciones simultáneas se conoce como *forma reducida* del modelo, por contraposición a la *forma estructural* [17.4] antes introducida. Al igual que la forma estructural, la forma reducida consta de tantas ecuaciones como variables endógenas. No es difícil comprobar que la función de verosimilitud muestral puede escribirse como función de los parámetros de la forma reducida:

$$L(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T / \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) = (2\pi)^{-gT/2} |\Omega|^{-T/2} \cdot \exp \left[-\frac{1}{2} \text{traza } \Omega^{-1} (\mathbf{Y} - \mathbf{X}\Pi)' (\mathbf{Y} - \mathbf{X}\Pi) \right]$$

Esta expresión ilustra la propiedad de que todos los parámetros de la forma reducida del modelo pueden obtenerse, previa estimación, a partir de los momentos de la distribución multivariante, propiedad que no es compartida por los parámetros de la forma estructural.

17.3. ESTIMACION DE LA FORMA REDUCIDA

El interés de la forma reducida de un modelo econométrico de ecuaciones simultáneas radica en que todas las variables explicativas de cada una de sus ecuaciones son exógenas. Estamos así en una situación semejante a la del modelo uniecuacional, y es perfectamente natural sugerir la estimación de este modelo por MCO. Como veremos en el Capítulo 18, dicho estimador es *insesgado*. La expresión del estimador MCO de los parámetros de la matriz Π en [17.9] viene dada por:

$$\hat{\Pi} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad [17.11]$$

donde puede comprobarse que las dimensiones del producto matricial son las de una matriz $k \times g$. Esta matriz producto contiene los coeficientes de las k variables explicativas, todas ellas exógenas, en cada una de las g ecuaciones de la forma reducida. Cada columna de la matriz Π proviene de postmultiplicar $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ por una de las columnas de \mathbf{Y} , por lo que la estimación que acabamos de sugerir es equivalente a estimar por MCO, separadamente, cada una de las ecuaciones de la forma reducida.

La matriz de covarianzas Ω , desconocida, puede estimarse generalizando el procedimiento seguido en el caso uniecuacional. En este caso, la estimación MCO de [17.11] permite generar una serie de residuos para cada una de las

ecuaciones del modelo. A su vez, estas series permiten estimar tanto las varianzas de los términos de error de cada una de las ecuaciones del modelo como las covarianzas entre cada dos de dichos términos de error. Tales estimaciones se obtendrían mediante los momentos muestrales análogos, calculados con las series de residuos mencionadas. En notación matricial:

$$\hat{\Omega} = \frac{\hat{V}'\hat{V}}{T-k} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\Pi})'(\mathbf{Y} - \mathbf{X}\hat{\Pi})}{T-k} = \frac{\mathbf{Y}'[\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}}{T-k}$$

Es decir, el elemento genérico ω_{ij} de la matriz Ω se estima mediante el cociente:

$$\hat{\omega}_{ij} = \frac{\sum_1^T \hat{u}_{it} \hat{u}_{jt}}{T-k}$$

A primera vista, y puesto que estamos suponiendo que existen correlaciones contemporáneas entre los términos de error de las diferentes ecuaciones, podría pensarse que la utilización del estimador MCG (es decir, la consideración de este modelo como uno de «regresiones aparentemente no relacionadas», de acuerdo con la terminología del Capítulo 18) generaría estimaciones más eficientes que la de MCO que hemos propuesto.

Sin embargo, notemos que en el modelo [17.10], todas las ecuaciones tienen las mismas matrices de observaciones para las variables explicativas, y recordemos que éste era uno de los casos en que la estimación conjunta del sistema de regresiones aparentemente no relacionadas no aportaba ninguna ganancia en términos de eficiencia. Otra cuestión diferente sería el caso en que existiesen restricciones de exclusión en algunas ecuaciones de la forma reducida. En tal caso, ya no sería cierto que todas las ecuaciones contendrían las mismas variables explicativas, por lo que la estimación conjunta podría ser más eficiente.

La diferencia conceptual entre la forma estructural del modelo y su forma reducida es que la primera, que es la equivalente a la especificación de modelos uniecuacionales discutida en capítulos anteriores, responde a una serie de conceptos basados en la Teoría Económica, que llevan al analista a especificar unas variables como endógenas y otras como exógenas, a la vez que decide qué variables entran en cada una de las ecuaciones, y cuáles no. ¿Cuándo nos va a convenir trabajar con una de las representaciones del modelo y no con la otra? Acabamos de ver que es posible obtener estimaciones insesgadas de los parámetros de la forma reducida, por lo que la cuestión anterior puede plantearse de nuevo en el sentido siguiente: ¿En qué casos necesitamos estimar la forma estructural, sin que la forma reducida nos proporcione toda la información deseada?

Si el interés del analista es obtener predicciones acerca de la evolución futura de las variables endógenas del sistema, entonces la forma estructural no es relevante, pues dichas predicciones podrían obtenerse a partir de los parámetros estimados de la forma reducida. Si, por el contrario, se pretende

llevar a cabo una serie de contrastes de hipótesis concernientes a los valores tomados por alguno de los coeficientes que tienen una interpretación de acuerdo con la Teoría Económica, resulta inevitable tener que estimar la forma estructural.

17.4. EL PROBLEMA DE IDENTIFICACION

Para obtener estimaciones mínimo-cuadráticas de la forma reducida no ha sido preciso hacer ningún supuesto adicional, por lo que esta representación del modelo puede estimarse siempre. Pero, y la forma estructural, ¿puede estimarse sin dificultad? Puesto que la forma reducida siempre puede estimarse sin problema, un modo de responder a esta pregunta es cuestionarse bajo qué condiciones sería posible recuperar los parámetros de la forma estructural a partir de las estimaciones de los parámetros de la forma reducida. Como hemos visto antes, tras llevar a cabo las normalizaciones ya discutidas, quedan $g(g - 1 + k)$ coeficientes por estimar en la forma estructural, mientras que hemos estimado gk parámetros en la matriz Π de coeficientes de la forma reducida. También querríamos estimar los $\frac{g(g + 1)}{2}$ parámetros de Σ a partir

de un mismo número de parámetros estimados en Ω . La ecuación [17.10] muestra que, si conseguimos estimar Γ , este segundo problema estará resuelto.

El número de parámetros que se pretenden recuperar en la forma estructural es mayor que el de parámetros estimados en la forma reducida, por lo que esta tarea de recuperación que nos hemos propuesto es, en general, imposible. En la práctica, sin embargo, la situación no es tan mala. Ya hemos comentado que, en general, la Teoría Económica sugerirá que determinadas variables no aparecen en ciertas ecuaciones del sistema, es decir, que el investigador estará dispuesto a imponer a priori una serie de restricciones que consisten en que algunos de los coeficientes de la forma estructural sean cero.

Mientras que estas restricciones no tienen, en general, ninguna influencia sobre las estimaciones a efectuar en la forma reducida, sin embargo, tendrán el efecto de reducir el número de coeficientes a estimar en la forma estructural. Por otra parte, incluso si finalmente terminamos con más coeficientes en la forma estructural que en la forma reducida, ello no necesariamente implica que no pueda estimarse ningún coeficiente de la forma estructural.

Por el contrario, podría ocurrir que a pesar de no poder obtener *todos* los coeficientes de la forma estructural, sin embargo, sí que se pueden recuperar los coeficientes de alguna ecuación. En tal caso, se dice que dicha ecuación está *identificada*. En caso contrario, se dice que la ecuación de la forma estructural *no está identificada*. A su vez, puede que haya un único modo de recuperar los coeficientes de dicha ecuación, en cuyo caso se dice que la ecuación está *exactamente identificada*, pero también podría ocurrir que hubiese varias formas, no equivalentes, de recuperar dichos parámetros. En este último caso, se dice que la ecuación está *sobreidentificada*.

Ejemplo 17.1. Consideremos el modelo en forma estructural:

$$\begin{aligned} \gamma_{11}^* y_{1t} + \gamma_{21}^* y_{2t} + \beta_{11}^* x_{1t} + \beta_{21}^* x_{2t} + u_{1t} &= 0 \\ \gamma_{12}^* y_{1t} + \gamma_{22}^* y_{2t} + \beta_{32}^* x_{3t} + u_{2t} &= 0 \end{aligned}$$

en el que hay diez parámetros a estimar: cuatro coeficientes de variables endógenas, tres coeficientes de variables predeterminadas y tres elementos de la matriz de covarianzas del vector (u_{1t}, u_{2t}) . Si normalizamos, dividiendo la primera ecuación por el coeficiente γ_{11}^* de la variable y_{1t} , y dividimos la segunda ecuación por γ_{22}^* , el coeficiente de y_{2t} , tenemos:

$$\begin{aligned} y_{1t} + \gamma_{21} y_{2t} + \beta_{11} x_{1t} + \beta_{21} x_{2t} + u_{1t} &= 0 \\ \gamma_{12} y_{1t} + y_{2t} + \beta_{32} x_{3t} + u_{2t} &= 0 \end{aligned}$$

donde, por ejemplo, $\beta_{21} = \beta_{21}^*/\gamma_{11}^*$ y $\gamma_{12} = \gamma_{12}^*/\gamma_{22}^*$.

La forma reducida del modelo puede obtenerse sustituyendo en la primera ecuación la variable y_{2t} por su expresión, obtenida a partir de la segunda, y llevando a cabo un ejercicio similar con la segunda ecuación. Con ello se llega a:

$$\begin{aligned} y_{1t} &= \pi_{11} x_{1t} + \pi_{21} x_{2t} + \pi_{31} x_{3t} + v_{1t} \\ y_{2t} &= \pi_{12} x_{1t} + \pi_{22} x_{2t} + \pi_{32} x_{3t} + v_{2t} \end{aligned}$$

donde:

$$\begin{aligned} \pi_{11} &= \frac{-\beta_{11}}{\Delta}, & \pi_{21} &= \frac{-\beta_{21}}{\Delta}, & \pi_{31} &= \frac{\gamma_{21}\beta_{32}}{\Delta}, & v_{1t} &= \frac{\gamma_{21}u_{2t} - u_{1t}}{\Delta}, \\ \pi_{12} &= \frac{\gamma_{12}\beta_{11}}{\Delta}, & \pi_{22} &= \frac{\gamma_{12}\beta_{21}}{\Delta}, & \pi_{32} &= \frac{-\beta_{32}}{\Delta}, & v_{2t} &= \frac{\gamma_{12}u_{1t} - u_{2t}}{\Delta} \end{aligned}$$

con $\Delta = 1 - \gamma_{12}\gamma_{21}$.

El lector debe comprobar que éste es el mismo resultado al que se llegaría si se utilizase directamente la expresión matricial que vimos para la matriz de coeficientes de la forma reducida: $\Pi = -\mathbf{B}\Gamma^{-1}$ y sus términos de error: $\mathbf{V} = -\mathbf{U}\Gamma^{-1}$. De igual modo, si se utilizan las expresiones anteriores del vector de términos de error de la forma reducida (v_{1t}, v_{2t}) para obtener sus varianzas y covarianza, se llega al mismo resultado que se alcanzaría si se utilizase la relación matricial $\Omega = \Gamma\Sigma\Gamma'$.

Una vez estimada la forma reducida del modelo, el estimador del coeficiente γ_{21} se obtendría de $\hat{\gamma}_{21} = \frac{-\hat{\pi}_{31}}{\hat{\pi}_{32}}$, mientras que los coeficientes β_{11} , β_{21} se calcularían mediante:

$$\begin{aligned} \hat{\beta}_{21} &= -(\hat{\pi}_{21} + \hat{\gamma}_{21}\hat{\pi}_{22}) \\ \hat{\beta}_{11} &= -(\hat{\pi}_{11} + \hat{\gamma}_{21}\hat{\pi}_{12}) \end{aligned}$$

Sin embargo, existen dos formas alternativas de obtener la estimación del coeficiente γ_{12} : $\hat{\gamma}_{12} = \frac{-\hat{\pi}_{12}}{\hat{\pi}_{11}}$, y también $\hat{\gamma}_{12} = \frac{-\hat{\pi}_{22}}{\hat{\pi}_{21}}$. Además, dichas estimaciones serán numéricamente diferentes, sin que se pueda escoger una sobre la otra. A su vez, generarán estimaciones diferentes de β_{32} mediante $\hat{\beta}_{32} = -(\hat{\pi}_{32} + \hat{\gamma}_{12} \hat{\pi}_{31})$.

Por consiguiente, la primera ecuación del modelo está *exactamente identificada*, mientras que la segunda ecuación está *sobreidentificada*.

Vamos a ver en las secciones siguientes cómo pueden utilizarse distintos tipos de restricciones para conseguir la identificación de las ecuaciones del modelo.

17.5. IDENTIFICACION MEDIANTE RESTRICCIONES DE EXCLUSION

Un modelo lineal que intenta representar el funcionamiento de un mercado competitivo constará, en general, de una ecuación de oferta del bien, una ecuación de demanda, y una condición de equilibrio.

Ejemplo 17.2. Consideremos el modelo:

$$q_t^o = \alpha_1 + \alpha_2 p_t + u_{1t}$$

$$q_t^d = \beta_1 + \beta_2 p_t + \beta_3 Y_t + u_{2t}$$

$$q_t^d = q_t^o$$

donde q_t^d , q_t^o , p_t , Y_t denotan, respectivamente, las cantidades demandada y ofrecida del bien, su precio y la renta agregada de la economía. Como puede verse, en la especificación del modelo se ha introducido ya la restricción, supuestamente basada en un conocimiento previo del mercado, de que, a diferencia de la demanda del bien en cuestión, su oferta es esencialmente inelástica con respecto a la renta.

En este modelo se supone que la renta agregada de la economía es una variable que se determina independientemente de las condiciones del mercado que se está analizando. Este es un supuesto muy razonable, salvo que el mercado en estudio fuese de una importancia muy considerable. Las variables endógenas son la cantidad y precio de equilibrio en dicho mercado. La forma reducida del modelo es:

$$q_t = \pi_{11} + \pi_{21} Y_t + v_{1t}$$

$$p_t = \pi_{12} + \pi_{22} Y_t + v_{2t}$$

donde:

$$\begin{aligned}\pi_{11} &= \frac{\alpha_2\beta_1 - \alpha_1\beta_2}{\alpha_2 - \beta_2}, & \pi_{12} &= \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} \\ \pi_{21} &= \frac{\alpha_2\beta_3}{\alpha_2 - \beta_2}, & \pi_{22} &= \frac{\beta_3}{\alpha_2 - \beta_2} \\ v_{1t} &= \frac{\alpha_2 u_{2t} - \beta_2 u_{1t}}{\alpha_2 - \beta_2}, & v_{2t} &= \frac{u_{2t} - u_{1t}}{\alpha_2 - \beta_2}\end{aligned}$$

Una vez obtenidas las estimaciones MCO de los coeficientes π_{ij} , $i, j = 1, 2$, podrían recuperarse algunos coeficientes de la forma estructural. Así:

$\hat{\alpha}_2 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{22}}$, $\hat{\alpha}_1 = \hat{\pi}_{11} - \hat{\alpha}_2 \pi_{12}$. Sin embargo, no podrían obtenerse estimaciones de ninguno de los coeficientes β_i , $i = 1, 2, 3$. Es cierto que, por ejemplo, $\hat{\beta}_1 = \hat{\pi}_{11} - \hat{\beta}_2 \hat{\pi}_{12}$, pero $\hat{\beta}_2$ no puede obtenerse a partir de las estimaciones de la forma estructural y, por tanto, tampoco $\hat{\beta}_1$.

Como consecuencia, decimos que la ecuación de oferta está identificada, ya que hemos podido recuperar estimaciones de sus coeficientes a partir de estimaciones de los coeficientes de la forma reducida. Por el contrario, la ecuación de demanda no está identificada. Veamos cuál es la interpretación intuitiva de este resultado:

Cuando se recogen datos de precios y cantidades intercambiadas en un mercado a lo largo del tiempo, dichas observaciones provendrán del equilibrio que el mercado ha registrado en cada instante de tiempo. En general, tanto la curva de demanda como la de oferta habrán sufrido desplazamiento a lo largo del período muestral, debido a diversas causas exógenas al mercado, con la situación que puede verse en la Figura 17.1. De este modo, la nube de puntos que se genera con todas las combinaciones posibles, aun siendo todos ellos puntos de equilibrio, no es una nube de puntos que describa ninguna de las curvas, demanda u oferta, sino una combinación lineal de ambas.

Alternativamente, supongamos que mientras que la curva de demanda se ha desplazado a lo largo del tiempo, quizá siguiendo las fluctuaciones en la renta agregada del país, la curva de oferta ha permanecido invariante ante dichas fluctuaciones. Este es el caso representado en el modelo anterior y que aparece en la Figura 17.2. La nube de puntos generada por la sucesión temporal de equilibrios en dicho mercado ahora describe la curva de oferta, que ha permanecido fija durante el período muestral. Estrictamente hablando, si la curva de oferta hubiese permanecido inmóvil, la nube de puntos estaría repartida a lo largo de una curva, precisamente la curva de oferta. Si la curva de oferta ha sufrido ligeros desplazamientos, sensiblemente inferiores a los de la curva de demanda, se obtendría una nube de puntos como en la Figura 17.3.

Si ajustamos una curva que relacione precios y cantidades a esta nube de puntos, se tendrá una estimación de la curva de oferta. Sin embargo, poco o

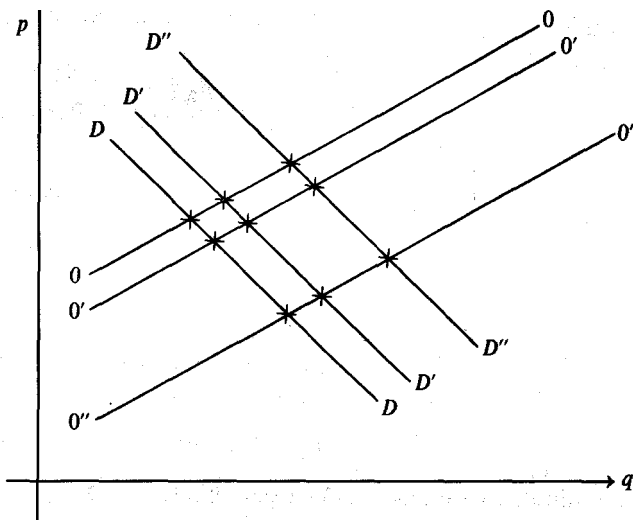


FIGURA 17.1.

nada podrá decirse acerca de la curva de demanda a partir de dicha estimación.

Uno de los orígenes de las especificaciones de modelos de ecuaciones simultáneas fue el análisis de mercados agrícolas. Estos mercados suelen caracterizarse porque su demanda es bastante estable, dependiendo únicamente de factores endógenos al mercado y, fundamentalmente, del precio del producto. Por el contrario, la oferta de estos bienes suele depender de factores exógenos como, por ejemplo, las condiciones climatológicas, ya sea la canti-

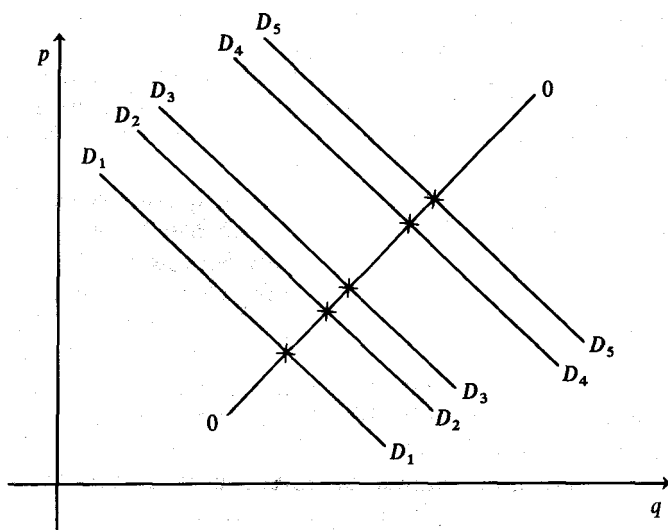


FIGURA 17.2.

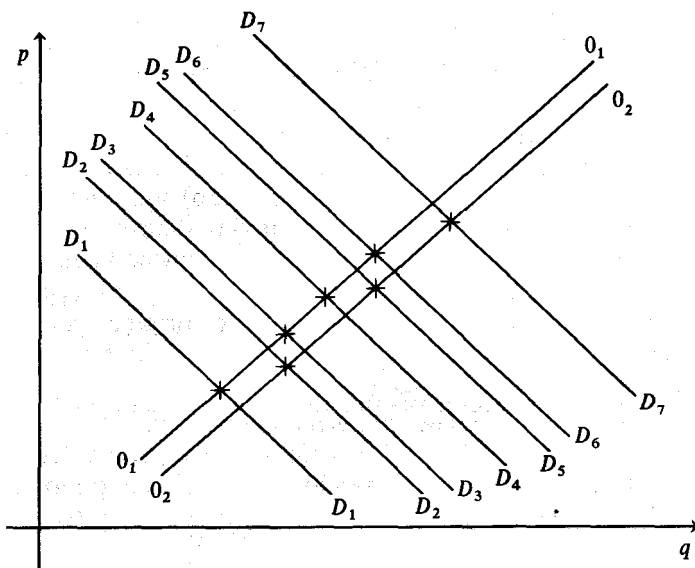


FIGURA 17.3.

dad de lluvia caída durante el año o la temperatura media, por ejemplo. Como consecuencia, éstos son ejemplos en que la curva ajustada a la nube de puntos procedente de dichos mercados proporcionará una idea bastante aproximada de las características de la curva de demanda de dichos productos.

Así pues, el problema de identificación surge cuando al pensar que se estima una sola ecuación de un sistema, en realidad, se está estimando una combinación lineal de las ecuaciones del mismo. Partiendo del sistema de ecuaciones:

$$y_t' \Gamma = -x_t' B - u_t'$$

y de su función de verosimilitud:

$$L(y_1, \dots, y_T/x_1, \dots, x_T) = |\Gamma|^T \prod_{t=1}^T f(u_t)$$

Consideremos una combinación lineal de sus ecuaciones, lo que se obtiene postmultiplicando el modelo por una matriz R' de dimensión $g \times g$:

$$y_t' \Gamma R' = -x_t' B R' - u_t' R' \quad [17.12]$$

o, lo que es lo mismo:

$$y_t' \Gamma^* = -x_t' B^* - u_t^{*'}$$

donde:

$$\Gamma^* = \Gamma R', \quad \mathbf{B}^* = \mathbf{B} R', \quad \mathbf{u}_t^* = \mathbf{u}_t' R', \quad \mathbf{u}_t^* = \mathbf{R} \mathbf{u}_t$$

Recordemos que, por las propiedades del producto de matrices, las columnas del producto $\Gamma^* = \Gamma R'$ (coeficientes de las variables endógenas en una determinada ecuación del modelo transformado) son combinaciones lineales de las columnas de la matriz Γ , y algo similar ocurre con la matriz \mathbf{B}^* . En consecuencia, cada ecuación del modelo transformado es una combinación lineal de las ecuaciones del modelo original.

Los términos de error del modelo original y el transformado están relacionados por $\mathbf{u}_t = \mathbf{R}^{-1} \mathbf{u}_t^*$, siendo el jacobiano de dicha transformación $\frac{\partial \mathbf{u}_t}{\partial \mathbf{u}_t^*} = \mathbf{R}^{-1}$. Por tanto, sus funciones de densidad multivariantes están relacionadas por $f(\mathbf{u}_t^*) = f(\mathbf{u}_t) |\mathbf{R}^{-1}|$. La función de verosimilitud del modelo transformado se obtiene aprovechando esta relación entre los vectores \mathbf{u}_t^* y \mathbf{u}_t :

$$\begin{aligned} L(\mathbf{y}_1, \dots, \mathbf{y}_T / \mathbf{x}_1, \dots, \mathbf{x}_T) &= |\Gamma^*|^T \prod_1^T f(\mathbf{u}_t^*) = |\Gamma R'|^T \prod_1^T f(\mathbf{u}_t^*) = \\ &= |\Gamma|^T |\mathbf{R}|^T |\mathbf{R}^{-1}|^T \prod_1^T f(\mathbf{u}_t) = |\Gamma|^T \prod_1^T f(\mathbf{u}_t) \end{aligned}$$

de modo que ambos modelos tienen la misma función de verosimilitud. En consecuencia, el mecanismo estadístico que genera las observaciones de uno y otro modelo son idénticos, por lo que las observaciones muestrales no son suficientes para distinguir entre el modelo original y el modelo obtenido tomando combinaciones lineales de sus ecuaciones.

Otra forma de ver esta equivalencia es observando que la maximización de la función de verosimilitud no permite estimar la matriz \mathbf{R} . En efecto, al maximizar $L(\mathbf{y}_1, \dots, \mathbf{y}_T / \mathbf{x}_1, \dots, \mathbf{x}_T)$ se estiman siempre las mismas matrices Γ y \mathbf{B} , con independencia de cual haya sido la matriz \mathbf{R} utilizada al formar las combinaciones lineales de las ecuaciones del sistema. Este problema surge cuando no existen restricciones acerca de los valores de los componentes de las matrices Γ , \mathbf{B} y Σ .

Sin embargo, si al margen de la muestra de datos se dispone de información en el sentido de que alguno de dichos componentes toman determinados valores (usualmente cero), entonces cabe la posibilidad de poder identificar los coeficientes de la forma estructural a partir de los coeficientes estimados de la forma reducida. Ello se debe a que si existen restricciones como las citadas sobre alguna de las ecuaciones, entonces combinaciones lineales de esa ecuación con las restantes ecuaciones del modelo pueden no satisfacer dicha restricción. Por ejemplo, una combinación lineal genérica de las ecuaciones del modelo

$$\begin{aligned} \gamma_{11} y_{1t} + \gamma_{21} y_{2t} + \beta_{11} x_{1t} &= u_{1t} \\ \gamma_{12} y_{1t} + \gamma_{22} y_{2t} + \beta_{12} x_{1t} + \beta_{22} x_{2t} &= u_{2t} \end{aligned}$$

que es un sistema de k_1 ecuaciones, y también:

$$\Pi_3 \gamma_1 = \Pi_3 \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \\ \dots \\ \gamma_{g_1,1} \end{pmatrix} = \mathbf{0}_{k-k_1} \quad [17.13]$$

que es un sistema de $k - k_1$ ecuaciones. El problema de identificación de una ecuación de la forma estructural de un modelo econométrico es el de resolver estos dos sistemas de ecuaciones para obtener estimaciones de los vectores γ_1 y β_1 a partir de estimaciones de la matriz Π . Un primer examen muestra que, puesto que el primer sistema consta de tantas ecuaciones, k_1 , como coeficientes hay en β_1 , se tiene que si pudiésemos resolver el segundo sistema de ecuaciones para obtener estimaciones del vector γ_1 , entonces tendríamos automáticamente estimado el vector β_1 .

Vayamos, pues, al último sistema de ecuaciones. Este es un sistema de $k - k_1$ ecuaciones lineales homogéneas en g_1 incógnitas, los elementos del vector γ_1 . Por ser homogéneo, los componentes del vector γ_1 podrían posiblemente obtenerse, salvo una constante de proporcionalidad. Este problema queda, sin embargo, resuelto con la normalización de ecuaciones que ya discutimos anteriormente, $\gamma_{11} = -1$.

Por ser un sistema lineal y homogéneo con g_1 ecuaciones, tendrá una única solución distinta del vector $\mathbf{0}_{g_1}$ si y sólo si se cumple la *condición de rango*⁽¹⁾:

$$\text{Rango}(\Pi_3) = g_1 - 1. \quad [17.14]$$

Supuesto que esta condición de rango se satisfaga, entonces el criterio de normalización antes discutido nos daría una única solución para el vector γ_1 . Si el rango de Π_3 fuese inferior a $g_1 - 1$, entonces el sistema de ecuaciones tendría infinitas soluciones incluso después de la normalización: en tal caso, la ecuación *no está identificada*. Finalmente, si el rango del sistema fuese g_1 , entonces el sistema tendría como única solución $\gamma_1 = \mathbf{0}_{g_1}$ (es decir, $\gamma_{11} = \gamma_{21} = \dots = \gamma_{g_1,1} = 0$). Es importante hacer hincapié en que la *condición de rango es necesaria y suficiente para la identificación de una ecuación*.

Por otra parte, cabe observar que como la submatriz Π_3 sobre la que está recayendo el peso de la identificación es de dimensión $(k - k_1) \times g_1$, sólo podrá tener rango igual a $g_1 - 1$ si se cumple la condición:

$$k - k_1 \geq g_1 - 1$$

Esta condición, llamada *condición de orden*, es una condición necesaria, aunque no suficiente, para que la primera ecuación del modelo esté identificada. La condición consiste en que el número de variables predeterminadas

⁽¹⁾ Un sistema homogéneo y completo de ecuaciones lineales tiene solución no nula tan sólo si la matriz de coeficientes es singular.

excluidas de la ecuación sea mayor o igual que el número de variables endógenas incluidas como explicativas en dicha ecuación. Si denotamos por $k_2 = k - k_1$ y $g_2 = g - g_1$ el número de variables exógenas y endógenas excluidas de la ecuación, podemos sumar g_2 a ambos miembros de la desigualdad para tener una expresión alternativa de la condición de orden:

$$k_2 + g_2 \geq g - 1$$

que consiste en que el número total de variables excluidas de la ecuación (predeterminadas y endógenas) sea mayor o igual al número de ecuaciones del modelo, menos uno. La condición de orden cataloga las ecuaciones en que $k - k_1 < g_1 - 1$ como de *subidentificadas*. Los coeficientes de dichas ecuaciones no pueden identificarse, es decir, dicha ecuación jamás puede satisfacer la condición de rango, puesto que la submatriz Π_3 no tiene suficientes filas. En efecto, puesto que $k - k_1 < g_1 - 1$, entonces se tiene que $\text{Rango}(\Pi_3) \leq k - k_1 < g_1 - 1$ y el sistema [17.13] tiene infinitas soluciones.

Por otro lado, las ecuaciones que satisfacen la condición de rango sí *pueden identificarse*. En tal caso, la condición de orden dirá si dichas ecuaciones están *sobreidentificadas* o *exactamente identificadas*. La discusión precedente conduce a la siguiente clasificación de cada una de las ecuaciones de un modelo, de acuerdo con su situación de identificación:

De acuerdo con la condición de orden, una ecuación está:

1. *Subidentificada*, si $k - k_1 < g_1 - 1$.
2. *Posiblemente identificada*, si $k - k_1 \geq g_1 - 1$, y dentro de éstas, de acuerdo con la condición de rango, la ecuación está:
 - 2.1. *Sobreidentificada*, si $k - k_1 > g_1 - 1$ y $\text{Rango}(\Pi_3) = g_1 - 1$.
 - 2.2. *Exactamente identificada*, si $k - k_1 = g_1 - 1$ y $\text{Rango}(\Pi_3) = g_1 - 1$.
 - 2.3. *No identificada*, si $k - k_1 \geq g_1 - 1$, pero $\text{Rango}(\Pi_3) < g_1 - 1$.

Conviene insistir en que la condición de orden es necesaria, pero no suficiente, por lo que una ecuación de un modelo de ecuaciones simultáneas pudiera satisfacer dicha condición y a pesar de ello no estar identificada.

Por el contrario, la condición de rango es necesaria y suficiente. La ventaja de la condición de orden es que es muy fácil de comprobar, puesto que consiste tan sólo en examinar si el número de variables exógenas excluidas de la ecuación en consideración es al menos igual al número de variables endógenas incluidas como explicativas.

La condición de rango es bastante más compleja de comprobar, puesto que requiere evaluar del rango de una submatriz de coeficientes de la forma reducida. Por tanto, es preciso comenzar calculando el modo en que los coeficientes de dicha forma reducida dependen de los coeficientes de la forma estructural. A continuación, se impondrían las restricciones de exclusión de que se disponga, y se pasaría a calcular el rango de la matriz resultante. La dificultad en la verificación de la condición de rango queda, sin embargo, resuelta por la siguiente proposición:

Proposición 17.1. La condición de rango [17.14] es *equivalente* a:

$$\text{Rango} \begin{pmatrix} \Gamma_2 \\ \mathbf{B}_2 \end{pmatrix} = g - 1$$

donde Γ_2 es la submatriz $g_2 \times g$ ($g_2 = g - g_1$) de Γ formada por los coeficientes, en todas las ecuaciones del sistema, de las variables endógenas excluidas de la ecuación en consideración, y \mathbf{B}_2 la submatriz $k_2 \times g$ ($k_2 = k - k_1$) de coeficientes, en todas las ecuaciones, de las variables exógenas excluidas de dicha ecuación. Nótese que por incluir tan sólo variables omitidas de la primera ecuación, en realidad la matriz $\begin{pmatrix} \Gamma_2 \\ \mathbf{B}_2 \end{pmatrix}$ debería constar de una primera columna de ceros, los coeficientes de las variables omitidas de dicha primera ecuación. Dicha columna se ha omitido, por lo que el número de columnas de la matriz anterior es $g - 1$.

Demostración. De acuerdo con la relación entre coeficientes de las formas estructural y reducida del modelo

$$\begin{array}{ccccc} g_1 & g_2 & 1 & g-1 & 1 & g-1 \\ k_1 & \begin{pmatrix} \Pi_1 & \Pi_2 \end{pmatrix} & \begin{pmatrix} \gamma_1 & \Gamma_1 \end{pmatrix} & g_1 & = & - \begin{pmatrix} \beta_1 & \mathbf{B}_1 \end{pmatrix} & k_1 \\ k_2 & \begin{pmatrix} \Pi_3 & \Pi_4 \end{pmatrix} & \begin{pmatrix} \mathbf{0}_{g_2} & \Gamma_2 \end{pmatrix} & g_2 & = & \begin{pmatrix} \mathbf{0} & \mathbf{B}_2 \end{pmatrix} & k_2 \end{array}$$

se tiene, en particular, $\Pi_3 \Gamma_1 + \Pi_4 \Gamma_2 = -\mathbf{B}_2$ (k_2 ecuaciones), y con éstas y [17.14] podemos formar el sistema:

$$\begin{array}{ccccc} g_1 & g_2 & 1 & g-1 & 1 & g-1 \\ g_2 & \begin{pmatrix} \mathbf{0}_{g_2 \times g_1} & \mathbf{I}_{g_2} \end{pmatrix} & \begin{pmatrix} \gamma_1 & \Gamma_1 \end{pmatrix} & g_1 & = & \begin{pmatrix} \mathbf{0}_{g_2} & \Gamma_2 \end{pmatrix} & g_2 \\ k_2 & \begin{pmatrix} -\Pi_3 & -\Pi_4 \end{pmatrix} & \begin{pmatrix} \mathbf{0}_{g_2} & \Gamma_2 \end{pmatrix} & g_2 & = & \begin{pmatrix} \mathbf{0}_{k_2} & \mathbf{B}_2 \end{pmatrix} & k_2 \\ (g_2 + k_2) \times g & g \times g & & & & (g_2 + k_2) \times g \end{array}$$

y como el rango de una matriz no varía al multiplicarla por una matriz no singular se tiene:

$$\begin{aligned} \text{Rango} \begin{pmatrix} \mathbf{0}_{g_2} & \Gamma_2 \\ \mathbf{0}_{k_2} & \mathbf{B}_2 \end{pmatrix} &= \text{Rango} \begin{pmatrix} \mathbf{0}_{g_2 \times g_1} & \mathbf{I}_{g_2} \\ -\Pi_3 & -\Pi_4 \end{pmatrix} = \text{Rango} \begin{pmatrix} \Pi_4 & \mathbf{I}_{k_2} \\ \mathbf{I}_{g_2} & \mathbf{0}_{g_2 + k_2} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{g_2 \times g_1} & \mathbf{I}_{g_2} \\ -\Pi_3 & -\Pi_4 \end{pmatrix} = \\ &= \text{Rango} \begin{pmatrix} -\Pi_3 & \mathbf{0}_{g_2 \times g_1} \\ \mathbf{0}_{g_2 \times (g_1 + 1)} & \mathbf{I}_{g_2} \end{pmatrix} = \text{Rango} (\Pi_3) + g_2 \\ & \quad (k_2 + g_2) \times g \end{aligned}$$

ya que la última matriz es diagonal a bloques. En consecuencia, se tiene en la condición de rango:

$$\begin{aligned} \text{Rango}(\Pi_3) = g_1 - 1 &\Leftrightarrow \text{Rango}(\Pi_3) + g_2 = g - 1 \Leftrightarrow \\ &\Leftrightarrow \text{Rango} \begin{pmatrix} \mathbf{0}_{g_2} & \Gamma_2 \\ \mathbf{0}_{k_2} & \mathbf{B}_2 \end{pmatrix} = \text{Rango} \begin{pmatrix} \Gamma_2 \\ \mathbf{B}_2 \end{pmatrix} = g - 1 \end{aligned}$$

Ejemplo 17.3. Consideremos el modelo:

$$\begin{aligned} \gamma_{11}y_{1t} + & & + \gamma_{31}y_{3t} & & + \beta_{11}x_{1t} + & & + u_{1t} = 0 \\ & \gamma_{22}y_{2t} + \gamma_{32}y_{3t} + \gamma_{42}y_{4t} + \beta_{12}x_{1t} + & & + u_{2t} = 0 \\ \gamma_{13}y_{1t} + \gamma_{23}y_{2t} + & & & & + \beta_{23}x_{2t} + \beta_{33}x_{3t} + u_{3t} = 0 \\ \gamma_{14}y_{1t} + \gamma_{24}y_{2t} + & & + \gamma_{44}y_{4t} + \beta_{14}x_{1t} + & & + u_{4t} = 0 \end{aligned}$$

En primer lugar, cabe observar que faltan de cada ecuación al menos tres variables, por lo que todas ellas satisfacen la condición de orden, necesaria para la identificación. Por tanto, todas las ecuaciones están *posiblemente identificadas*. Comprobamos ahora la condición de rango de cada una de las ecuaciones. La submatriz de coeficientes que de acuerdo con la discusión anterior corresponde a la primera ecuación es:

$$\begin{pmatrix} \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{42} & 0 & \gamma_{44} \\ 0 & \beta_{23} & 0 \\ 0 & \beta_{33} & 0 \end{pmatrix}$$

que, como puede apreciarse, tiene rango 3 y está *exactamente identificada*. Este rango podría ser aún inferior si existiesen determinadas restricciones entre los coeficientes de la ecuación. Como consecuencia, la *primera ecuación* de este modelo está *exactamente identificada*. La matriz de coeficientes correspondiente a la segunda ecuación es:

$$\begin{pmatrix} \gamma_{11} & \gamma_{13} & \gamma_{14} \\ 0 & \beta_{23} & 0 \\ 0 & \beta_{33} & 0 \end{pmatrix}$$

que tiene rango 2. *La segunda ecuación no está identificada*. De modo análogo, puede comprobarse que, mientras que *la tercera ecuación está exactamente identificada*, *la cuarta ecuación no está identificada*.

Como se dijo antes, un procedimiento alternativo para discutir la identificación de una ecuación consiste en examinar la clase de combinaciones lineales de las ecuaciones del modelo que satisfacen las restricciones incorporadas en la ecuación en estudio. Sin embargo, este procedimiento sólo ayuda

a discernir si una ecuación *posiblemente identificada* está identificada o no. Si lo está, habrá que remitirse a la condición de orden para saber si la ecuación está exactamente identificada o sobreidentificada.

Mención especial merece el tratamiento de las identidades que suelen aparecer en un modelo de ecuaciones simultáneas. Su presencia no crea ningún problema adicional a los ya tratados. Sin embargo, su carácter de identidad permite su utilización para eliminar alguna variable de todas las ecuaciones en que aparezca. Este procedimiento es el más recomendable en la práctica. Observemos que tampoco hay ninguna limitación en prescindir de la identidad tras la operación anterior, pues, por su carácter de identidad, no hay en ella ningún coeficiente que estimar.

17.6. IDENTIFICACION CON RESTRICCIONES LINEALES HOMOGENEAS

Hasta el momento, hemos discutido la identificación del modelo de ecuaciones simultáneas únicamente bajo el supuesto de que algunas variables faltan de ciertas ecuaciones de la forma estructural. Sin embargo, éste no es el único caso de interés. En efecto, en bastantes ocasiones, se dispone de restricciones lineales que relacionan los coeficientes de una ecuación. A veces, incluso se dispone de restricciones lineales sobre los coeficientes de diferentes ecuaciones. Discutimos el primero de estos casos en esta sección, dejando el segundo de ellos para la Sección 17.9. Denotemos por \mathbf{A} la matriz de dimensión $(g+k) \times g$ formada por los coeficientes de la forma estructural del modelo, es decir:

$$\mathbf{A} = \begin{pmatrix} \mathbf{\Gamma} \\ \mathbf{B} \end{pmatrix}$$

Si denotamos por \mathbf{z}_t el vector de dimensión $g+k$ definido por $\mathbf{z}_t' = [\mathbf{y}_t'; \mathbf{x}_t']$, entonces el modelo puede escribirse:

$$\mathbf{z}_t' \mathbf{A} + \mathbf{u}_t' = (\mathbf{y}_t'; \mathbf{x}_t') \begin{pmatrix} \mathbf{\Gamma} \\ \mathbf{B} \end{pmatrix} + \mathbf{u}_t' = 0, \quad t = 1, 2, \dots, T$$

La primera ecuación de la forma estructural puede escribirse: $\mathbf{z}_t' \boldsymbol{\alpha}_1 + u_{1t} = 0$, donde $\boldsymbol{\alpha}_1$ es la primera columna de la matriz \mathbf{A} . Consideremos la identificación de esta ecuación cuando existen restricciones lineales del tipo: $\lambda_1 \gamma_{11} + \lambda_2 \gamma_{21} + \dots + \lambda_g \gamma_{g1} + \lambda_{g+1} \beta_{11} + \dots + \lambda_{g+k} \beta_{k1} = 0$. Las condiciones de exclusión a las que hemos hecho referencia con anterioridad son un caso particular de éstas, donde tan sólo uno de los coeficientes de la ecuación tiene coeficiente no nulo en la restricción.

Por otra parte, vamos a mantener de momento el supuesto de que las restricciones son homogéneas, es decir, que carecen de término independiente. Dichas restricciones pueden representarse en forma matricial por:

$$\Phi \boldsymbol{\alpha}_1 = \mathbf{0}_q$$

donde la matriz Φ tiene tantas filas, g , como restricciones existen sobre los coeficientes de dicha ecuación, y tantas columnas como parámetros hay en α_1 , es decir, $k + g$. Los elementos de una fila dada de la matriz Φ son los $g + k$ coeficientes de los coeficientes $\gamma_{11}, \dots, \gamma_{g1}, \beta_{11}, \dots, \beta_{k1}$ en la restricción correspondiente.

Por ser de orden $(g + k) \times g$, el rango de la matriz A no podría ser nunca superior a g . Supondremos que $\text{Rango}(\Phi) = g$, es decir, que las restricciones son linealmente independientes entre sí. Por otra parte, la matriz producto ΦA tiene todos los elementos de su primera columna iguales a cero por ser igual a $\Phi \alpha_1$ y ser las restricciones homogéneas. Ello impone una cota más estricta sobre el rango de A , puesto que implica que dicho rango no puede superar a $g - 1$.

Estas no son, sin embargo, las únicas restricciones entre los coeficientes de la primera ecuación del modelo. En efecto, existen restricciones adicionales, provenientes de las relaciones existentes entre la forma estructural y la forma reducida. Como sabemos, la relación entre las matrices de coeficientes de ambas representaciones del modelo puede escribirse: $\Pi \Gamma + B = 0$, donde Π es la matriz de orden $k \times g$ de la forma reducida, es decir, que si definimos la matriz $W = [\Pi; I_k]$, de dimensión $k \times (g + k)$, se tiene:

$$WA = [\Pi; I_k] \begin{pmatrix} \Gamma \\ B \end{pmatrix} = 0_{k \times g}$$

Las restricciones que estas relaciones imponen sobre los coeficientes de la primera ecuación de la forma estructural son $W\alpha_1 = 0_k$, que junto con las anteriores pueden escribirse:

$$\begin{pmatrix} W \\ \Phi \end{pmatrix} \alpha_1 = 0_{k+g} \quad [17.15]$$

Una vez que se hubiera estimado la forma reducida del modelo (del modo que se vio en la Sección 17.3), todos los parámetros en la matriz $\begin{pmatrix} W \\ \Phi \end{pmatrix}$ son conocidos. Ello se debe a que los parámetros que aparecen en Φ son los coeficientes de las restricciones, mientras que los de la matriz W , aunque desconocidos, se han estimado. En el sistema anterior [17.15], las incógnitas son los valores de los $g + k$ coeficientes que integran el vector α_1 . La matriz $\begin{pmatrix} W \\ \Phi \end{pmatrix}$ es de dimensión $(k + g) \times (g + k)$ y, por tanto, $k + g$ es el número de ecuaciones en el sistema [17.15].

De modo análogo a la discusión que mantuvimos en el caso en que las únicas restricciones consideradas eran del tipo de exclusión, es fácil ver que la condición de orden (necesaria, aunque no suficiente para la identificación de la primera ecuación) es que el número de ecuaciones no sea inferior al número de incógnitas menos uno, es decir, $k + g \geq g + k - 1$, o equivalentemente: $g \geq g - 1$, es decir, que debe haber al menos tantas restricciones sobre los

coeficientes de una ecuación como ecuaciones hay en el modelo menos una. Si se cumple esta condición, se tiene:

$$\text{Rango} \begin{pmatrix} \mathbf{W} \\ \Phi \end{pmatrix} \leq \min \{k + q; g + k\} \leq g + k - 1$$

Ahora bien, por ser homogéneo, el sistema de ecuaciones anterior tendrá solución única tan sólo si el rango de la matriz $\begin{pmatrix} \mathbf{W} \\ \Phi \end{pmatrix}$ es precisamente igual a $g + k - 1$. Por ello, la *condición de rango* (necesaria y suficiente para la identificación) consiste en que se cumpla:

$$\text{Rango} \begin{pmatrix} \mathbf{W} \\ \Phi \end{pmatrix} = g + k - 1 \quad [17.16]$$

Al igual que en el caso de restricciones de exclusión, este modo de expresar analíticamente la condición de rango tiene el problema de que precisa conocer la forma en que los coeficientes de la forma reducida dependen de los de la forma estructural. Dicha dificultad desaparece una vez que se prueba (cosa que aquí no haremos) que la condición anterior es equivalente a $\text{Rango}(\Phi\mathbf{A}) = g - 1$, donde $\Phi\mathbf{A}$ es una matriz $q \times g$. A pesar de ser [17.15] un sistema de ecuaciones homogéneo, de nuevo la regla de normalización $\gamma_{11} = -1$ permite seleccionar una única solución al sistema, en caso de que dicha solución exista. Ello ocurrirá siempre que la ecuación esté identificada, es decir, siempre que se cumpla la condición [17.16].

De modo similar a la clasificación de las ecuaciones de un modelo, de acuerdo con su situación de identificación, que hicimos en la Sección 17.5, se tiene ahora:

De acuerdo con la condición de orden, una ecuación está:

1. *Subidentificada*, si $q < g - 1$.
2. *Posiblemente identificada*, si $q \geq g - 1$, y dentro de éstas, de acuerdo con la condición de rango, la ecuación está:
 - 2.1. *Sobreidentificada*, si $q > g - 1$ y $\text{Rango}(\Phi\mathbf{A}) = g - 1$.
 - 2.2. *Exactamente identificada*, si $q = g - 1$ y $\text{Rango}(\Phi\mathbf{A}) = g - 1$.
 - 2.3. *No identificada*, si $q \geq g - 1$, pero $\text{Rango}(\Phi\mathbf{A}) < g - 1$.

En el caso particular de que todas las restricciones fuesen de exclusión, entonces la matriz Φ constaría de unos y ceros, con *un solo* 1 por fila, y se tendría:

$$\Phi\mathbf{A} = \Phi \begin{pmatrix} \Gamma \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \Gamma_2 \\ \mathbf{0} & \mathbf{B}_2 \end{pmatrix}$$

que es exactamente la matriz cuyo rango debía ser igual a $g - 1$ para que se satisficiera la condición de rango. Por tanto, como cabía esperar, el caso de

las condiciones de exclusión puede derivarse como caso particular del caso de restricciones lineales homogéneas.

Hay que tener presente que, para discutir la identificación del sistema, hay que llevar a cabo este ejercicio con cada una de sus ecuaciones, y la matriz de coeficientes Φ diferirá de unas a otras.

Veamos cómo utilizar estas condiciones de identificación sobre un ejemplo concreto.

Ejemplo 17.4. Consideremos el sistema:

$$(y_{1t}, y_{2t}, y_{3t}) \begin{pmatrix} -1 & 0 & 0 \\ \gamma_{21} & -1 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & -1 \end{pmatrix} + (x_{1t}, x_{2t}, x_{3t}, x_{4t}) \begin{pmatrix} \beta_{11} & 0 & \beta_{13} \\ 0 & \beta_{22} & \beta_{23} \\ \beta_{31} & 0 & 0 \\ \beta_{41} & \beta_{42} & \beta_{43} \end{pmatrix} + (u_{1t}, u_{2t}, u_{3t}) = (0, 0, 0)$$

sujeto a la restricción $\beta_{31} = 3\beta_{11}$. La primera observación que conviene hacer es que hay dos restricciones sobre los coeficientes de la primera ecuación, tres restricciones sobre la segunda ecuación y dos restricciones sobre la tercera ecuación. Por tanto, de acuerdo con la condición de orden, las tres ecuaciones están «posiblemente identificadas». Si la ecuación de rango mostrase que, efectivamente, están identificadas, entonces la primera y tercera ecuaciones estarían exactamente identificadas, mientras que la segunda ecuación estaría sobredeterminada.

La matriz de los coeficientes de las restricciones sobre la primera ecuación es:

$$\Phi = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -3 & 0 & 1 & 0 \end{pmatrix}$$

de modo que:

$$\Phi A = \begin{pmatrix} 0 & \beta_{22} & \beta_{23} \\ 0 & 0 & -3\beta_{13} \end{pmatrix}$$

que tiene rango 2, el número de ecuaciones del sistema menos uno. Por tanto, la primera ecuación está *exactamente identificada*. Nótese que la restricción entre β_{31} y β_{11} permite identificar la primera ecuación que, sin ella, estaría subidentificada. La matriz Φ análoga para la segunda ecuación es:

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

de modo que:

$$\Phi A = \begin{pmatrix} -1 & 0 & 0 \\ \beta_{11} & 0 & \beta_{13} \\ 3\beta_{11} & 0 & 0 \end{pmatrix}$$

que tiene rango igual a 2, por lo que esta ecuación *no está identificada*. Finalmente, la matriz de coeficientes de las restricciones para la tercera ecuación es:

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

y

$$\Phi A = \begin{pmatrix} -1 & 0 & 0 \\ 3\beta_{11} & 0 & 0 \end{pmatrix}$$

que tiene rango igual a 1. La tercera ecuación del modelo *no está identificada*.

Consideremos ahora un sistema similar, cambiando únicamente el supuesto de que la variable exógena que falta de la tercera ecuación es x_{4t} , en vez de x_{3t} . El lector debe comprobar que, en tal sistema, la tercera ecuación está *exactamente identificada*. Ello se debe a que dicha ecuación sería la única que *no contiene a x_{4t}* , por lo que será fácilmente distinguible de una combinación lineal de las ecuaciones del sistema. Por el contrario, en el primero de los sistemas considerados, la variable x_{3t} estaba ausente tanto de la segunda como de la tercera ecuaciones. *Una ecuación de un sistema está identificada no sólo en función de cuantas variables faltan de dicha ecuación, sino también de si esas mismas variables están presentes o no en las restantes ecuaciones.*

17.7. IDENTIFICACION CON RESTRICCIONES LINEALES NO HOMOGENEAS

Sin duda que las restricciones entre los parámetros de un modelo no tienen por qué ser homogéneas. Consideremos ejemplos como $\beta_2 + 3\beta_3 = 5$ o $7\beta_1 - 4\beta_2 = 16$. Un ejemplo típico de esta situación sería el caso del contraste de rendimientos constantes a escala en una función de producción Cobb-Douglas, que se formula como la posibilidad de que la suma de unos determinados coeficientes del modelo de regresión sea igual a 1.

Las diferencias prácticas con el caso anterior son mínimas. No será cierta ahora nuestra afirmación anterior en el sentido de que la primera columna de la matriz producto ΦA sea cero. Por el contrario, ahora dicha columna será igual a c , el vector de términos independientes de las q restricciones. Por tanto, el rango de la matriz ΦA ya no tiene $g - 1$ como cota superior, sino que puede

llegar a ser g . La condición de rango consistirá en que el rango de dicha matriz sea precisamente igual a g :

$$\text{Rango}(\Phi\mathbf{A}) = g$$

y la clasificación de ecuaciones de acuerdo con su identificación puede reformularse del modo obvio.

17.8. RESTRICCIONES NO LINEALES

El tratamiento riguroso de este tipo de restricciones, aunque interesante, excede el ámbito de cuestiones que pueden tratarse al nivel de este texto. Baste sugerir que dichas restricciones se aproximen linealmente mediante desarrollos en serie de Taylor y se compruebe la identificación de las ecuaciones del modelo bajo estas restricciones aproximadas. La condición de rango se convierte en:

$$\text{Rango}(\Phi_0\mathbf{A}) = g - 1$$

donde Φ_0 es la matriz que contiene los coeficientes de las aproximaciones lineales a las restricciones.

La obtención de las aproximaciones lineales requiere de la elección de valores paramétricos alrededor de los cuales obtener el desarrollo de Taylor. Esta elección es un problema, máxime cuando la identificación es un proceso lógicamente anterior a la estimación del modelo. La identificación que así se contrasta se llama *local* (es decir, alrededor de los valores paramétricos escogidos), por contraposición a la que no requeriría aproximaciones, y que se llamaría *global*, pues puede depender de los valores paramétricos utilizados en la aproximación lineal.

17.9. IDENTIFICACION BAJO RESTRICCIONES ENTRE COEFICIENTES DE DISTINTAS ECUACIONES

En ocasiones, el investigador dispone de restricciones entre coeficientes que aparecen en distintas ecuaciones del modelo. Lógicamente, este tipo de restricciones también ayuda a conseguir la identificación de una ecuación; sin embargo, discutir las condiciones generales de identificación en tales situaciones es relativamente complejo, por lo que nos limitamos a la discusión individualizada de un ejemplo.

Ejemplo 17.5. Consideremos el modelo:

$$\begin{aligned} y_{1t} + \gamma_{21}y_{2t} + \beta_{11}x_t + u_{1t} &= 0 \\ y_{2t} + \beta_{12}x_t + u_{2t} &= 0 \end{aligned}$$

Claramente, la primera ecuación está subidentificada, mientras que la segunda ecuación está exactamente identificada. La forma reducida del modelo es:

$$\begin{aligned}y_{1t} + \pi_{11}x_t + v_{1t} &= 0 \\y_{2t} + \pi_{12}x_t + v_{2t} &= 0\end{aligned}$$

con

$$\begin{aligned}\pi_{11} &= \beta_{11} - \beta_{12}\gamma_{21} & v_{1t} &= u_{1t} - \gamma_{21}u_{2t} \\ \pi_{12} &= \beta_{12} & v_{2t} &= u_{2t}\end{aligned}$$

por lo que el coeficiente β_{12} se obtiene directamente de la estimación de π_{12} , aunque no se pueden estimar β_{11} ni γ_{21} .

Sin embargo, supongamos que se dispusiera de información adicional en el sentido de que los coeficientes de la variable predeterminada x_t en ambas ecuaciones son iguales: $\beta_{11} = \beta_{12} = \beta$. En tal caso, la estimación del coeficiente β seguiría obteniéndose de la estimación de π_{12} y, además, se tendría:

$$\hat{\pi}_{11} = \hat{\beta}(1 - \hat{\gamma}_{21}) = \hat{\pi}_{12}(1 - \hat{\gamma}_{21})$$

que permite obtener la estimación de γ_{21} , por lo que las dos ecuaciones estarían exactamente identificadas.

Ejemplo 17.6. Consideremos el modelo:

$$\begin{aligned}y_{1t} + \gamma_{21}y_{2t} + \beta_{11}x_t + u_{1t} &= 0 \\ \gamma_{12}y_{1t} + y_{2t} + \beta_{12}x_t + u_{2t} &= 0\end{aligned}$$

cuya forma reducida es:

$$\begin{aligned}y_{1t} + \pi_{11}x_t + v_{1t} &= 0 \\ y_{2t} + \pi_{12}x_t + v_{2t} &= 0\end{aligned}$$

con

$$\begin{aligned}\pi_{11} &= \frac{\beta_{11} - \beta_{12}\gamma_{21}}{\Delta}, & v_{1t} &= \frac{u_{1t} - \gamma_{21}u_{2t}}{\Delta} \\ \pi_{12} &= \frac{\beta_{12} - \beta_{11}\gamma_{12}}{\Delta}, & v_{2t} &= \frac{u_{2t} - \gamma_{12}u_{1t}}{\Delta}\end{aligned}$$

donde $\Delta = 1 - \gamma_{12}\gamma_{21}$.

La estimación de la forma reducida proporcionaría valores numéricos de los coeficientes π_{11} y π_{12} , pero sería imposible recuperar a partir de ellos estimaciones de ninguno de los cuatro coeficientes de la forma estructural. Ambas ecuaciones están *subidentificadas*.

Supongamos que se incorpora la restricción $\beta_{11} = \beta_{12} = \beta$. Entonces las relaciones entre coeficientes de las formas estructural y reducida serían:

$$\hat{\pi}_{11} = \frac{\hat{\beta}(1 - \hat{\gamma}_{21})}{\hat{\Delta}}$$

$$\hat{\pi}_{12} = \frac{\hat{\beta}(1 - \hat{\gamma}_{12})}{\hat{\Delta}}$$

de las que, de nuevo, sería imposible obtener estimaciones de β , γ_{12} y γ_{21} . Sería preciso disponer de información adicional para llegar a identificar las ecuaciones de la forma estructural. En el Problema 17.9, al final del capítulo, se pide discutir distintas alternativas de identificación para este modelo.

17.10. IDENTIFICACION CON RESTRICCIONES SOBRE LA MATRIZ DE COVARIANZAS

De nuevo, este análisis es suficientemente complejo para que analicemos tan sólo un ejemplo:

Ejemplo 17.7. Volvamos al modelo de equilibrio de mercado que consideramos en la Sección 17.5:

$$\begin{aligned} q_t^o &= \alpha_1 + \alpha_2 p_t + u_{1t} \\ q_t^d &= \beta_1 + \beta_2 p_t + \beta_3 Y_t + u_{2t} \\ q_t^d &= q_t^o \end{aligned}$$

del que vimos que la presencia de la renta en la ecuación de demanda permite identificar la ecuación de oferta, mientras que la primera está subidentificada. Veamos cómo contribuye a la identificación del modelo el supuesto de que los términos de error de ambas ecuaciones sean independientes: $\sigma_{u_1 u_2} = 0$. Repetimos aquí que la forma reducida del modelo es:

$$\begin{aligned} q_t &= \pi_{11} + \pi_{21} Y_t + v_{1t} \\ p_t &= \pi_{12} + \pi_{22} Y_t + v_{2t} \end{aligned}$$

con

$$\begin{aligned} \pi_{11} &= \frac{\alpha_2 \beta_1 - \alpha_1 \beta_2}{\alpha_2 - \beta_2} & \pi_{21} &= \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} \\ \pi_{12} &= \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} & \pi_{22} &= \frac{\beta_3}{\alpha_2 - \beta_2} \\ v_{1t} &= \frac{\alpha_2 u_{2t} - \beta_2 u_{1t}}{\alpha_2 - \beta_2} & v_{2t} &= \frac{u_{2t} - u_{1t}}{\alpha_2 - \beta_2} \end{aligned}$$

por lo que, bajo el supuesto $\sigma_{u_1 u_2} = 0$, se tiene:

$$\sigma_{v_1}^2 = \frac{\alpha_2^2 \sigma_{u_2}^2 + \beta_2^2 \sigma_{u_1}^2}{(\alpha_2 - \beta_2)^2} \quad \sigma_{v_2}^2 = \frac{\sigma_{u_1}^2 + \sigma_{u_2}^2}{(\alpha_2 - \beta_2)^2} \quad \sigma_{v_1 v_2} = \frac{\alpha_2 \sigma_{u_2}^2 + \beta_2 \sigma_{u_1}^2}{(\alpha_2 - \beta_2)^2} \quad [17.17]$$

Como se vio en la Sección 17.5, la identificación exacta de la primera ecuación implica que no hay ninguna dificultad en estimar α_1 y α_2 : $\hat{\alpha}_2 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{22}}$, $\hat{\alpha}_1 = \hat{\pi}_{11} - \alpha_2 \hat{\pi}_{12}$. Una vez estimados α_1 y α_2 , así como las varianzas y covarianza $\sigma_{v_1}^2$, $\sigma_{v_2}^2$ y $\sigma_{v_1 v_2}$, las tres igualdades [17.17] suponen tres ecuaciones, con incógnitas: β_2 , $\sigma_{u_1}^2$ y $\sigma_{u_2}^2$. Finalmente, las estimaciones de β_1 y β_3 se obtienen de $\hat{\beta}_1 = \hat{\pi}_{11} - \hat{\beta}_2 \hat{\pi}_{12}$ y $\hat{\beta}_3 = \hat{\pi}_{22}(\hat{\alpha}_2 - \hat{\beta}_2)$. Así, la condición de ortogonalidad entre los términos de error de ambas ecuaciones ha bastado para conseguir la *identificación exacta* de la segunda ecuación. Sin embargo, ello depende de que las ecuaciones [17.17], que son no lineales, tengan solución única para β_2 , $\sigma_{u_1}^2$ y $\sigma_{u_2}^2$.

PROBLEMAS

Problema 17.1. a) Encontrar los parámetros de la forma reducida como función de los parámetros de la forma estructural en el modelo:

$$\begin{aligned} y_{1t} &= \gamma_{21} y_{2t} + \beta_{11} x_{1t} + \beta_{21} x_{2t} + u_{1t} \\ y_{2t} &= \gamma_{12} y_{1t} + u_{2t} \end{aligned}$$

b) Supuesto que se dispusiese de estimaciones de los *cuatro* coeficientes que aparecen en la forma reducida, ¿sería posible, en general, encontrar un único estimador de los parámetros de la forma estructural? ¿De cuáles sí y de cuáles no? ¿Existe alguna condición especial entre los valores de los parámetros de la forma estructural que permitiese obtener estimaciones únicas de *todos* los parámetros de la forma reducida?

Problema 17.2. Un modelo *recursivo* es aquel en que la matriz Γ es triangular, mientras que la matriz de covarianzas Σ es diagonal.

- a) Probar que dicho modelo está siempre identificado.
- b) Probar que cada una de las dos condiciones citadas, por sí sola, no basta para conseguir la identificación de un modelo econométrico.

Problema 17.3. Mostrar que el modelo de ecuaciones simultáneas:

Función de consumo:	$C_t = \beta_0 + \beta_1 Y_t^d,$	$0 < \beta_1 < 1$
Impuestos:	$T_t = \alpha_0 + \alpha_1 Y_t,$	$0 < \alpha_1 < 1$
Función de inversión:	$I_t = \gamma_0 + \gamma_1 r_t,$	$\gamma_1 < 0$
Identidad:	$Y_t^d = Y_t - T_t$	
Gasto público:	$G_t = \bar{G}$	
Identidad contable:	$Y_t = C_t + I_t + G_t$	

Demanda de dinero: $M_t^d = a + bY_t + cr_t, \quad b > 0, c < 0$
 Oferta de dinero: $M_t^o = \bar{M}$
 Equilibrio monetario: $M_t^d = M_t^o$

puede reducirse al siguiente par de ecuaciones:

Curva IS: $Y_t = \delta_0 + \delta_1 r_t, \quad \delta_1 < 0$
 Curva LM: $Y_t = \lambda_0 + \lambda_1 M_t + \lambda_2 r_t, \quad \lambda_0 < 0, \quad \lambda_1, \lambda_2 > 0$

¿Cómo podrían incluirse perturbaciones aleatorias en el primer conjunto de ecuaciones para que el sistema IS-LM tuviera términos de error?

Problema 17.4. Considere el sistema de ecuaciones simultáneas:

$$\begin{aligned} C_t &= \beta_1 + \beta_2 Y_t - \beta_3 T_t + u_{1t} \\ I_t &= \alpha_0 + \alpha_1 Y_{t-1} + u_{2t} \\ T_t &= \gamma_0 + \gamma_1 Y_t + u_{3t} \\ Y_t &= C_t + I_t + G_t \end{aligned}$$

donde C_t, Y_t, T_t, I_t, G_t denotan, respectivamente, el consumo, el producto, los ingresos por impuestos, la inversión y el gasto público. Se consideran como endógenas C_t, I_t, T_t e Y_t . Discutir la identificación de cada una de las ecuaciones del modelo. ¿Cómo cambiaría su respuesta si incluyésemos el tipo de interés r_t , supuesto exógeno, en la función de inversión?

Problema 17.5. a) Discuta la identificación de las ecuaciones del sistema:

$$\begin{aligned} r_t &= \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t} \\ Y_t &= \alpha_0 + \alpha_1 r_t + u_{2t} \end{aligned}$$

donde r_t, M_t e Y_t denotan los tipos de interés, la oferta monetaria y el producto.

b) ¿Cambia la respuesta si se incluye Y_{t-1} como variable explicativa en la primera ecuación?

c) En el sistema a), introduzca la inversión I_t (supuesta exógena) como variable explicativa en la segunda ecuación y discuta la identificación del sistema.

d) Mantenga la estructura de c), pero añadiendo la ecuación $I_t = \gamma_0 + \gamma_1 Y_t + r_t + u_{3t}$, y vuelva a discutir la identificación del sistema.

Problema 17.6. Dado el modelo de ecuaciones simultáneas:

$$\begin{aligned} y_{1t} + \gamma_{21}y_{2t} + \beta_{21}x_{2t} + \beta_{31}x_{3t} + \beta_{41}x_{4t} &= u_{1t} \\ \gamma_{12}y_{1t} + y_{2t} + \beta_{12}x_{1t} &= u_{2t} \end{aligned}$$

a) Discutir las condiciones de orden y rango para la identificación de cada ecuación.

b) Probar que si alguna de las ecuaciones está exactamente identificada, las estimaciones de sus coeficientes pueden obtenerse de modo único a partir de las estimaciones de los coeficientes de la forma reducida.

c) Probar que si alguna de las dos ecuaciones está sobreidentificada, las estimaciones de sus coeficientes pueden obtenerse de más de un modo a partir de los coeficientes estimados de la forma reducida. ¿Qué dificultades genera este hecho acerca de la estimación de esta(s) ecuación(es)?

d) ¿Cómo afectaría a la identificación del sistema si la variable exógena incluida en la segunda ecuación fuese x_{2t} en lugar de x_{1t} ?

Problema 17.7. Discutir la identificación de cada ecuación en los siguientes modelos de ecuaciones simultáneas:

$$1. \quad \gamma_{11}y_{1t} + \gamma_{21}y_{2t} + \beta_{11}x_{1t} + \beta_{21}x_{2t} = u_{1t}$$

$$\gamma_{12}y_{1t} + \gamma_{22}y_{2t} + \beta_{12}x_{1t} + \beta_{22}x_{2t} = u_{2t}$$

1.a) sin restricciones

1.b) con las restricciones: $\beta_{12} = \beta_{21} = 0$

1.c) con las restricciones: $\beta_{21} = \beta_{22} = 0$

1.d) con las restricciones: $\beta_{11} = \beta_{12} = \beta_{22} = 0$

1.e) con las restricciones: $\beta_{11} = \beta_{21} = 0$

1.f) con las restricciones: $\beta_{11} = \beta_{21} = \beta_{22} = 0, \gamma_{12} + \beta_{12} = 0$

$$2. \quad y_{1t} + \gamma_{21}y_{2t} + \beta_{11}x_{1t} + u_{1t}$$

$$\gamma_{12}y_{1t} + y_{2t} + \beta_{12}x_{1t} + u_{2t}$$

con la restricción: $\beta_{11} + \beta_{12} = 0$

$$3. \quad y_{1t} + \beta_{11}x_{1t} = u_{1t}$$

$$\gamma_{12}y_{1t} + y_{2t} + \beta_{12}x_{1t} = u_{2t}$$

$$4. \quad y_{1t} + \beta_{11}x_{1t} = u_{1t}$$

$$\gamma_{12}y_{1t} + y_{2t} + \beta_{12}x_{1t} = u_{2t}$$

$$\gamma_{13}y_{1t} + \gamma_{23}y_{2t} + y_{3t} + \beta_{13}x_{1t} + u_{3t}$$

con las restricciones: $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$

Problema 17.8. Demostrar la igualdad:

$$\Sigma_1^T (y_i \Gamma + x_i' B) \Sigma^{-1} (y_i \Gamma + x_i' B) = \text{traza} [(Y \Gamma + X B) \Sigma^{-1} (Y \Gamma + X B)]$$

Problema 17.9. Comprobar que en el modelo

$$y_{1t} + \gamma_{21}y_{2t} + \beta x_t + u_{1t} = 0$$

$$\gamma_{12}y_{1t} + y_{2t} + \beta x_t + u_{2t} = 0$$

analizado en la Sección 17.9, se tiene, alternativamente:

a) Con la restricción adicional: $\gamma_{12} = 0$, ambas ecuaciones están exactamente identificadas.

b) Con la restricción $\gamma_{21} = a$, donde a es un parámetro conocido, ambas ecuaciones están exactamente identificadas.

c) Con la restricción $\gamma_{12} = \gamma_{21}$, las dos ecuaciones permanecen subidentificadas.

Problema 17.10. Considérese el siguiente modelo del mercado de un cierto producto:

$$q_t^d = \alpha_0 + \alpha_1 p_t + u_{1t}, \quad \alpha_1 < 0 \quad (\text{ecuación de demanda})$$

$$q_t^o = \beta_0 + \beta_1 p_t + u_{2t}, \quad \beta_1 > 0 \quad (\text{ecuación de oferta})$$

$$\Delta p_t = \gamma \Delta S_{t-1} + u_{3t} \quad (\text{ajuste de precio})$$

$$S_t = S_{t-1} + q_t^o - q_t^d \quad (\text{ecuación de existencias})$$

donde $\Delta p_t = p_t - p_{t-1}$ y S_t denota el nivel de existencias.

a) Obtener los valores del equilibrio estático del precio y la cantidad (prescindiendo de los términos de perturbación). ¿Qué restricciones adicionales sobre los parámetros sugieren estos valores?

b) Obtener el rango de valores de γ para que el sistema sea estable. Dar una interpretación económica de dicha condición en el contexto de una ecuación de ajuste de precio.

Problema 17.11. Discutir, mediante el procedimiento de formar combinaciones lineales de sus ecuaciones, la identificación del modelo

$$y_{1t} + \gamma_{21} y_{2t} + \beta_{11} x_t + u_{1t} = 0$$

$$\gamma_{12} y_{1t} + y_{2t} + \beta_{12} x_t + u_{2t} = 0$$

bajo las restricciones:

a) $\gamma_{21} = 0.$

b) $\gamma_{21} = \gamma_{12}; \beta_{11} = \beta_{12}.$

Problema 17.12. Discutir la identificación del modelo:

$$y_{1t} = \beta_1 + \gamma_{21} y_{2t} + \beta_{11} x_{1t} + u_{1t}$$

$$y_{2t} = \beta_2 + \gamma_{22} y_{2t} + \beta_{22} x_{2t} + \beta_{32} x_{3t} + u_{2t}$$

¿Cómo cambiaría la discusión anterior si se conociese que $\beta_{22} + \beta_{32} = 1$?

Problema 17.13. Demostrar que la condición $\text{Rango}(\Phi A) = g - 1$ de la Sección 17.6 es equivalente a la condición:

$$\text{Rango} \begin{pmatrix} W \\ \Phi \end{pmatrix} = g + k - 1$$

Problema 17.14. Demostrar que cuando las restricciones lineales son de exclusión (Sección 17.6) se tiene:

$$\Phi A = \begin{pmatrix} \mathbf{0} & \Gamma_2 \\ \mathbf{0} & \beta_2 \end{pmatrix}$$

de lo que se deduce la condición de rango de la Sección 17.5.

Problema 17.15. a) Discutir la identificación de cada ecuación del modelo:

$$\left. \begin{aligned} y_{1t} + \beta_{11}x_{1t} + \beta_{12}x_{2t} &= u_{1t} \\ \gamma_{21}y_{1t} + y_{2t} + \beta_{22}x_{2t} &= u_{2t} \end{aligned} \right\}$$

b) Discutir la identificación si se impone la restricción:

$$\gamma_{21} + \beta_{11} = 0$$

c) Discutir, en general, la siguiente afirmación: «Las únicas restricciones que ayudan a identificar una ecuación en un modelo de ecuaciones simultáneas son aquellas que conciernen a sus propios parámetros. En particular, una restricción que atañe únicamente a parámetros de otra ecuación no tiene nada que ver con la identificación de la primera de ellas».

CAPITULO 18

MODELOS DE ECUACIONES SIMULTANEAS

II. ESTIMACION

18.1. DIFICULTADES EN LA ESTIMACION POR MINIMOS CUADRADOS ORDINARIOS

Una primera aproximación al problema de estimación de una ecuación de un modelo de ecuaciones simultáneas podría consistir en la utilización del método de mínimos cuadrados, ordinarios o generalizados, según cual sea la estructura de la matriz de covarianzas de su término de error. Sin embargo, ya hemos mencionado en el capítulo anterior que este procedimiento presenta algunas dificultades que lo hacen desaconsejable. En particular, el estimador mínimo-cuadrático no sólo es sesgado, sino que su sesgo no desaparece al aumentar el tamaño muestral. Para ilustrar tales dificultades, consideremos la expresión que adoptaría el estimador MCO de los $g_1 + k_1 - 1$ coeficientes de la primera ecuación del modelo; una vez que se han normalizado de modo que el coeficiente de y_1 sea la unidad

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix}_{\text{MCO}} = \left[\begin{pmatrix} \mathbf{Y}'_1 \\ \mathbf{X}'_1 \end{pmatrix} (\mathbf{Y}_1; \mathbf{X}_1) \right]^{-1} \begin{pmatrix} \mathbf{Y}'_1 \\ \mathbf{X}'_1 \end{pmatrix} \mathbf{y}_1 = \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \mathbf{y}_1 \\ \mathbf{X}'_1 \mathbf{y}_1 \end{pmatrix}$$

y substituyendo el vector \mathbf{y}_1 , de dimensión $T \times 1$, por su expresión

$$\mathbf{y}_1 = (\mathbf{Y}_1; \mathbf{X}_1) \begin{pmatrix} \gamma_1 \\ \beta_1 \end{pmatrix} + \mathbf{u}_1$$

se tiene:

$$\begin{aligned} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \\ \mathbf{X}'_1 \end{pmatrix} \left[(\mathbf{Y}_1; \mathbf{X}_1) \begin{pmatrix} \gamma_1 \\ \beta_1 \end{pmatrix} + \mathbf{u}_1 \right] = \\ &= \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \\ \mathbf{X}'_1 \end{pmatrix} \mathbf{u}_1 = \\ &= \begin{pmatrix} \gamma_1 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \mathbf{u}_1 \\ \mathbf{X}'_1 \mathbf{u}_1 \end{pmatrix} \end{aligned}$$

y tomando esperanzas en ambos lados de la igualdad se tiene:

$$E \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \beta_1 \end{pmatrix} + E \left[\begin{pmatrix} \mathbf{Y}'_1 \mathbf{Y}_1 & \mathbf{Y}'_1 \mathbf{X}_1 \\ \mathbf{X}'_1 \mathbf{Y}_1 & \mathbf{X}'_1 \mathbf{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'_1 \mathbf{u}_1 \\ \mathbf{X}'_1 \mathbf{u}_1 \end{pmatrix} \right]$$

por lo que este estimador será sesgado a no ser que la esperanza del último corchete sea cero. Dicha esperanza no sólo no será cero, en general, sino además difícil de calcular sin imponer supuestos adicionales a los que habitualmente hemos considerado, pues depende de las correlaciones entre los vectores aleatorios \mathbf{Y}_1 y \mathbf{u}_1 , así como de los momentos superiores de la distribución de \mathbf{Y}_1 .

Ejemplo 18.1. En el caso del modelo IS-LM considerado en el Problema 17.3:

$$\text{IS: } Y_t = \alpha_0 + \alpha_1 r_t + u_{1t}, \quad \alpha_1 < 0$$

$$\text{LM: } Y_t = \beta_0 + \beta_1 M_t + \beta_2 r_t + u_{2t}, \quad \beta_0 < 0, \quad \beta_1, \beta_2 > 0$$

se tiene la forma reducida:

$$\begin{aligned} Y_t &= \frac{\alpha_0 \beta_2 - \alpha_1 \beta_0}{\beta_2 - \alpha_1} - \frac{\alpha_1 \beta_1}{\beta_2 - \alpha_1} M_t + \frac{\beta_2 u_{1t} - \alpha_1 u_{2t}}{\beta_2 - \alpha_1} \\ r_t &= \frac{\alpha_0 - \beta_0}{\beta_2 - \alpha_1} - \frac{\beta_1}{\beta_2 - \alpha_1} M_t + \frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1} \end{aligned}$$

de donde puede obtenerse:

$$E(r_t) = \frac{\alpha_0 - \beta_0}{\beta_2 - \alpha_1} - \frac{\beta_1 M_t}{\beta_2 - \alpha_1}$$

por lo que:

$$r_t - E(r_t) = \frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1}$$

y, finalmente:

$$\text{Cov}(r_t, u_{1t}) = \frac{\sigma_1^2 - \sigma_{12}}{\beta_2 - \alpha_1}$$

de modo que si se estima la curva IS por mínimos cuadrados ordinarios, con las variables en desviaciones con respecto a la media, es fácil ver que:

$$E(\hat{\alpha}_1) = E\left(\frac{\sum_1^T (Y_t - \bar{Y})(r_t - \bar{r})}{\sum_1^T (r_t - \bar{r})^2}\right) = \alpha_1 + E\left(\frac{\sum_1^T (r_t - \bar{r})(u_{1t} - \bar{u}_1)}{\sum_1^T (r_t - \bar{r})^2}\right)$$

por lo que el sesgo del estimador dependerá de la covarianza muestral entre el tipo de interés y la perturbación aleatoria de la curva IS. En ocasiones, podemos predecir cuál será el signo de este sesgo.

En primer lugar, la expresión anterior muestra que dicho sesgo tendrá el mismo signo que la covarianza entre r_t y u_{1t} . Supongamos que u_{1t} , la perturbación aleatoria de la curva IS, toma un valor positivo; ello implicaría un valor de la renta por encima de su promedio, lo que, a través de la curva LM, ha de venir asociado con un valor de r_t también por encima de su promedio. En consecuencia, la covarianza entre r_t y u_{1t} es positiva, y el coeficiente α_1 resulta *sobrestimado* si se utiliza el método de mínimos cuadrados ordinarios. Finalmente, como $\hat{\alpha}_0 = \bar{Y} - \hat{\alpha}_1 \bar{r}$, es claro que al sobrestimar la pendiente α_1 , también *subestimaremos* el término independiente α_0 , de modo que se tendrá una curva IS más horizontal de lo que se debiera (recordemos que $\alpha_1 < 0$).

La evaluación del sesgo asintótico del estimador puede también obtenerse

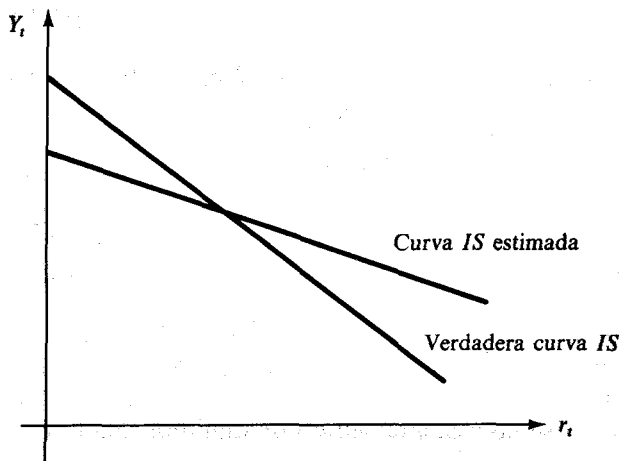


FIGURA 18.1.

analíticamente. En efecto, utilizando las propiedades del límite en probabilidad se tiene⁽¹⁾:

$$plim \hat{\alpha}_1 = \alpha_1 + \frac{plim \left[\left(\frac{1}{T} \right) \Sigma_1^T (r_t - \bar{r}) u_{1t} \right]}{plim \left[\left(\frac{1}{T} \right) \Sigma_1^T (r_t - \bar{r})^2 \right]}$$

por lo que el sesgo asintótico del estimador depende del cociente de los límites en probabilidad que aparecen en esta expresión. Su cálculo puede hacerse a partir de la forma reducida del modelo:

$$\begin{aligned} plim \Sigma_1^T \frac{(r_t - \bar{r}) u_{1t}}{T} &= \\ plim \left(\frac{1}{T} \right) \Sigma_1^T \left[\left(\frac{\alpha_0 - \beta_0}{\beta_2 - \alpha_1} - \frac{\beta_1}{\beta_2 - \alpha_1} (M_t - \bar{M}) + \frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1} \right) u_{1t} \right] &= \\ = \text{Cov} \left(\frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1}, u_{1t} \right) &= \frac{1}{\beta_2 - \alpha_1} (\sigma_1^2 - \sigma_{12}) \end{aligned}$$

donde se ha utilizado la propiedad de la convergencia en probabilidad de los momentos muestrales a los momentos análogos poblacionales. De este modo, el momento muestral del producto del término constante por u_{1t} , converge en probabilidad a cero, mientras que el momento muestral correspondiente al producto de $M_t - \bar{M}$ por u_{1t} converge a la covarianza de ambas variables, que es cero por ser M_t determinista.

De igual modo, se tiene:

$$plim \left[\left(\frac{1}{T} \right) \Sigma_1^T (r_t - \bar{r})^2 \right] = \text{Var } r_t = \text{Var} \left(\frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1} \right) = \frac{1}{(\beta_2 - \alpha_1)^2} (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$$

por lo que el sesgo asintótico del estimador mínimo cuadrático de α_1 es:

$$plim(\hat{\alpha}_1) - \alpha_1 = (\beta_2 - \alpha_1) \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

que, como puede verse, no desaparece incluso si la covarianza entre las dos perturbaciones aleatorias u_{1t} y u_{2t} es cero. Sin embargo, bajo el supuesto de independencia entre u_{1t} y u_{2t} , el sesgo en $\hat{\alpha}_1$ tiende a cero si la varianza σ_1^2 tiende a cero. En tal caso, la curva IS permanecería constante, por lo que la curva LM se desplazaría sobre la IS a través del tiempo; la nube de puntos muestral sería una descripción de la curva IS, cuyos coeficientes podrían estimarse consistentemente por mínimos cuadrados.

⁽¹⁾ Recordemos que $\Sigma_1^T (r_t - \bar{r}) (u_{1t} - \bar{u}_1) = \Sigma_1^T r_t (u_{1t} - \bar{u}_1) = \Sigma_1^T (r_t - \bar{r}) u_{1t}$.

Al igual que en este ejemplo, el estimador de mínimos cuadrados de una ecuación de la forma estructural del modelo de ecuaciones simultáneas es generalmente inconsistente. Ello hace que este procedimiento de estimación no sea recomendable. Veamos otras soluciones al problema de estimación de una ecuación del modelo de ecuaciones simultáneas.

18.2. ESTIMACION POR MINIMOS CUADRADOS INDIRECTOS

Este es un método de estimación válido únicamente para ecuaciones exactamente identificadas, aprovechando tal condición para recuperar los coeficientes de dicha ecuación de la forma estructural a partir de los coeficientes estimados de la forma reducida. La ventaja de este procedimiento reside en que, como sabemos, la estimación de mínimos cuadrados ordinarios de los parámetros de la forma reducida del modelo es insesgada.

Ejemplo 18.2. En la forma reducida del sistema IS-LM anterior:

$$\begin{aligned} Y_t &= \pi_{11} + \pi_{21}M_t + v_{1t} \\ r_t &= \pi_{12} + \pi_{22}M_t + v_{2t} \end{aligned}$$

con

$$\begin{aligned} \pi_{11} &= \frac{\alpha_0\beta_2 - \alpha_1\beta_0}{\beta_2 - \alpha_1}, & \pi_{21} &= -\frac{\beta_1\alpha_1}{\beta_2 - \alpha_1}, & v_{1t} &= \frac{\beta_2u_{1t} - \alpha_1u_{2t}}{\beta_2 - \alpha_1} \\ \pi_{12} &= \frac{\alpha_0 - \beta_0}{\beta_2 - \alpha_1}, & \pi_{22} &= -\frac{\beta_1}{\beta_2 - \alpha_1}, & v_{2t} &= \frac{u_{1t} - u_{2t}}{\beta_2 - \alpha_1} \end{aligned}$$

una vez obtenidas las estimaciones mínimo-cuadráticas de los coeficientes π_{ij} , se calcularía:

$$\hat{\alpha}_1 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{22}}, \quad \hat{\alpha}_0 = -\frac{\hat{\pi}_{12}\hat{\pi}_{21}}{\hat{\pi}_{22}} + \hat{\pi}_{11}$$

que son los estimadores de *mínimos cuadrados indirectos* de la curva IS.

El lector debe comprobar dos cuestiones: a) que utilizando las expresiones analíticas para el estimador de mínimos cuadrados de la forma reducida se tiene que el estimador de mínimos cuadrados indirectos de α_1 es:

$$\hat{\alpha}_1 = \frac{\sum_1^T (Y_t - \bar{Y})(M_t - \bar{M})}{\sum_1^T (r_t - \bar{r})(M_t - \bar{M})}$$

que es diferente del estimador de mínimos cuadrados ordinarios, y b) que no es posible obtener estimaciones de los coeficientes de la curva LM a partir

de los coeficientes estimados de la forma reducida del modelo, a diferencia de lo que hemos hecho con los coeficientes de la curva IS, ya que aquella no está identificada.

En general, supongamos que se pretende estimar la primera ecuación del modelo de ecuaciones simultáneas:

$$\mathbf{y}_1 = \mathbf{Y}_1 \boldsymbol{\gamma}_1 + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{u}_1 \quad [18.1]$$

es decir,

$$(\mathbf{y}_1; \mathbf{Y}_1; \mathbf{X}_1) \begin{pmatrix} -1 \\ \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_1 \end{pmatrix} = -\mathbf{u}_1$$

que puede también escribirse, incluyendo todas las variables del modelo:

$$(\mathbf{y}_1; \mathbf{Y}_1; \mathbf{Y}_2; \mathbf{X}_1; \mathbf{X}_2) \begin{pmatrix} -1 \\ \boldsymbol{\gamma}_1 \\ \mathbf{0}_{g_2} \\ \boldsymbol{\beta}_1 \\ \mathbf{0}_{g_2} \end{pmatrix} = -\mathbf{u}_1$$

donde \mathbf{Y}_1 y \mathbf{X}_1 son las matrices $T \times (g_1 - 1)$ y $T \times k_1$, de observaciones de las variables endógenas y predeterminadas incluidas como explicativas en la ecuación, e \mathbf{Y}_2 y \mathbf{X}_2 son, respectivamente, las matrices $T \times g_2$ y $T \times k_2$, de observaciones de las variables endógenas y predeterminadas excluidas de dicha ecuación.

El estimador de mínimos cuadrados indirectos consiste en utilizar mínimos cuadrados ordinarios en la forma reducida, como se vio en la Sección 17.3, para después tratar de recuperar las estimaciones de los coeficientes de la ecuación que se pretende estimar. Las estimaciones MCO de la forma reducida dan lugar a la matriz $k \times g$, $\hat{\Pi}$:

$$\hat{\Pi} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

donde \mathbf{Y} denota la matriz $T \times g$ que tiene por columnas los vectores de observaciones de cada una de las variables endógenas del sistema.

Proposición 18.1. El estimador MCO de la forma reducida es *insesgado*.

Demostración:

$$\begin{aligned} E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}] &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\Pi + \mathbf{V})] = \Pi + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}] = \\ &= \Pi + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{V}) = \Pi \end{aligned}$$

Proposición 18.2. El estimador MCO de la forma reducida del modelo de ecuaciones simultáneas es consistente, y se tiene la siguiente distribución asintótica:

$$\sqrt{T}(\hat{\Pi}_{MCO} - \Pi) \longrightarrow N(\mathbf{0}, \Omega \otimes \Sigma_{xx}^{-1})$$

donde $\Sigma_{xx} = \text{plim} (X'X/T)$ y $\Omega = (\Gamma^{-1})' \Sigma \Gamma^{-1}$. La matriz de covarianzas $\Omega \otimes \Sigma_{xx}^{-1}$ es de dimensión $kg \times kg$.

La demostración de este resultado es analíticamente compleja, y puede verse en Schmidt (1976).

Las ecuaciones $\Pi\Gamma = -\mathbf{B}$ relacionan los coeficientes de las formas estructural y reducida; por ejemplo, el producto de la matriz Π por la primera columna de la matriz Γ (coeficientes de las variables endógenas en la primera ecuación) es un vector de dimensión $k \times 1$, que debe ser igual a la primera columna de la matriz \mathbf{B} , que contiene los coeficientes de todas las variables predeterminadas en la primera ecuación.

Teniendo en cuenta la expresión $\hat{\Pi} = (X'X)^{-1}X'Y$, las relaciones $\hat{\Pi}\hat{\Gamma} = \hat{\beta}$ entre los coeficientes estimados de la primera ecuación pueden escribirse:

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} (y_1 Y_1 Y_2) \begin{pmatrix} -1 \\ \hat{\gamma}_1 \\ \mathbf{0}_{g_2} \end{pmatrix} = - \begin{pmatrix} \hat{\beta}_1 \\ \mathbf{0}_{k_2} \end{pmatrix}$$

donde g_2 y k_2 son, respectivamente, el número de variables endógenas y predeterminadas excluidas de la ecuación, es decir, $g_2 = g - g_1$ y $k_2 = k - k_1$. El estimador de mínimos cuadrados indirectos (MCI) es una solución del sistema anterior. Premultiplicando por la matriz $X'X$ se tiene:

$$\begin{aligned} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} (y_1 Y_1 Y_2) \begin{pmatrix} -1 \\ \hat{\gamma}_1 \\ \mathbf{0}_{g_2} \end{pmatrix} &= \begin{pmatrix} X_1'y_1 & X_1'Y_1 & X_1'Y_2 \\ X_2'y_1 & X_2'Y_1 & X_2'Y_2 \end{pmatrix} \begin{pmatrix} -1 \\ \hat{\gamma}_1 \\ \mathbf{0}_{g_2} \end{pmatrix} = \\ &= - \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \mathbf{0}_{k_2} \end{pmatrix} \end{aligned}$$

lo que da origen al sistema de ecuaciones:

$$\begin{aligned} (X_1'Y_1)\hat{\gamma}_1 + (X_1'X_1)\hat{\beta}_1 &= X_1'y_1 \quad (k_1 \text{ ecuaciones}) \\ (X_2'Y_1)\hat{\gamma}_1 + (X_2'X_1)\hat{\beta}_1 &= X_2'y_1 \quad (k_2 \text{ ecuaciones}) \end{aligned} \tag{18.2}$$

que forman un conjunto de k ecuaciones en $(g_1 - 1) + k_1$ incógnitas, las componentes de los vectores $\hat{\gamma}_1$ y $\hat{\beta}_1$. Puesto que la condición necesaria para identificación exacta es que $k - k_1 = g_1 - 1$, entonces se tiene el mismo número de ecuaciones que de incógnitas, por lo que, en general, el sistema

puede resolverse para obtener *soluciones únicas* para los parámetros γ_1 y β_1 .

La resolución de este sistema de ecuaciones es equivalente a obtener los parámetros de la ecuación del modo que hicimos con la curva IS en el Ejemplo 18.2. Como se ve, la condición de identificación es crucial para garantizar la existencia del estimador de mínimos cuadrados indirectos. Esta es la razón por la que no pudimos utilizar este procedimiento para estimar la curva LM del ejemplo, ya que dicha ecuación está subidentificada.

El estimador MCI es, en general, sesgado, a pesar de que los estimadores de la forma reducida eran insesgados. Ello se debe a que el estimador MCI es una función *no lineal* de las estimaciones de la forma reducida del modelo. Sin embargo, puede probarse que es un estimador consistente, ya que es una función continua del estimador MCO de la forma reducida, que es consistente. Así, en el caso de la curva IS se tiene:

$$plim \hat{\alpha}_1 = \frac{plim \hat{\pi}_{21}}{plim \hat{\pi}_{22}} = \frac{E \hat{\pi}_{21}}{E \hat{\pi}_{22}} = \frac{\pi_{21}}{\pi_{22}} = \alpha_1$$

El método MCI puede aplicarse también a ecuaciones sobreidentificadas. La única dificultad es que, dado que se basa en resolver directamente las ecuaciones que ligan a los parámetros de las formas estructural y reducida, en ecuaciones sobreidentificadas daría lugar a varias estimaciones diferentes. Por ello, en tales casos, se prefiere utilizar el método de estimación que introduciremos en la sección siguiente. Antes, veamos otro ejemplo.

Ejemplo 18.3. Consideremos el modelo de ecuaciones simultáneas:

$$\begin{aligned} y_{1t} &= \gamma_{21}y_{2t} + \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t} \\ y_{2t} &= \gamma_{12}y_{1t} + \beta_{32}x_{3t} + u_{2t} \end{aligned}$$

y supongamos que se han obtenido los siguientes momentos muestrales con respecto a la media:

	y_1	y_2	x_1	x_2	x_3
y_1	12	6	3	0	2
y_2	6	16	2	4	1
x_1	3	2	4	0	2
x_2	0	4	0	1	1
x_3	2	1	2	1	3

donde, por ejemplo: $T^{-1}\sum_1^T(x_{2t} - \bar{x}_2)(y_{2t} - \bar{y}_2) = 4$. Las estimaciones de mínimos cuadrados de cada una de las ecuaciones son:

Primera ecuación:

$$\begin{pmatrix} \hat{\gamma}_{21} \\ \hat{\beta}_{11} \\ \hat{\beta}_{21} \end{pmatrix} = \begin{pmatrix} y_2' y_2 & y_2' x_1 & y_2' x_2 \\ & x_1' x_1 & x_1' x_2 \\ & & x_2' x_2 \end{pmatrix}^{-1} \begin{pmatrix} y_2' y_1 \\ x_1' y_1 \\ x_2' y_1 \end{pmatrix} = \\ = \begin{pmatrix} 16 & 2 & 4 \\ 2 & 4 & 0 \\ 4 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} -9/2 \\ 3 \\ 18 \end{pmatrix}$$

Segunda ecuación:

$$\begin{pmatrix} \hat{\gamma}_{12} \\ \hat{\beta}_{32} \end{pmatrix} = \begin{pmatrix} y_1' y_1 & y_1' x_3 \\ & x_3' x_3 \end{pmatrix}^{-1} \begin{pmatrix} y_1' y_2 \\ x_3' y_2 \end{pmatrix} = \begin{pmatrix} 12 & 2 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$$

Como ya hemos visto, estas estimaciones no son muy fiables, pues no sólo son sesgadas, sino que además su sesgo no desaparece al aumentar el tamaño muestral. La primera ecuación del modelo está exactamente identificada y se presta, por tanto, a ser estimada por mínimos cuadrados indirectos. Para ello comencemos estimando por mínimos cuadrados ordinarios la forma reducida (recordemos que este estimador es insesgado):

$$\begin{pmatrix} \hat{\pi}_{11} & \hat{\pi}_{12} \\ \hat{\pi}_{21} & \hat{\pi}_{22} \\ \hat{\pi}_{31} & \hat{\pi}_{32} \end{pmatrix} = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 2 \\ 0 & 4 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & 5/2 \\ -1/2 & 8 \\ 1/2 & -4 \end{pmatrix}$$

El lector debe comprobar que las expresiones para recuperar las estimaciones de los parámetros de la primera ecuación son:

$$\hat{\gamma}_{21} = \frac{\hat{\pi}_{31}}{\hat{\pi}_{32}} = -\frac{1}{8}$$

$$\hat{\beta}_{11} = \hat{\pi}_{11} - \hat{\pi}_{12} \hat{\gamma}_{21} = \frac{13}{16}$$

$$\hat{\beta}_{21} = \hat{\pi}_{21} - \hat{\pi}_{22} \hat{\gamma}_{21} = \frac{1}{2}$$

Las estimaciones MCI que acabamos de obtener son consistentes, aunque sesgadas en muestras finitas. Como puede apreciarse, son bien diferentes de las estimaciones MCO que antes obtuvimos, que no son ni siquiera consistentes. La comparación entre ambas es, sin embargo, difícil de establecer, puesto que simplemente de acuerdo con sus propiedades en muestras finitas ninguno de los estimadores es estrictamente mejor que el otro. Su única similitud en este caso estriba en el signo de los coeficientes estimados.

La segunda ecuación, sin embargo, está sobreidentificada, por lo que hay más de una forma de obtener estimaciones de sus coeficientes a partir de las estimaciones de los coeficientes de la forma reducida. Así:

$$\hat{\gamma}_{12} = \frac{\hat{\pi}_{12}}{\hat{\pi}_{11}} = 5$$

pero también:

$$\hat{\gamma}_{12} = \frac{\hat{\pi}_{22}}{\hat{\pi}_{21}} = -16$$

Como vamos a ver en la sección siguiente, existe un estimador que es, en general, más eficiente que el estimador MCI.

18.3. ESTIMACION POR VARIABLES INSTRUMENTALES

El problema que presenta la utilización directa del estimador MCO en la estimación del modelo de ecuaciones simultáneas surge de la presencia de variables endógenas como explicativas. Sus correlaciones con el término de error de la ecuación hacen que dicho estimador sea sesgado e inconsistente. Estamos así en una situación similar a la del Capítulo 9, en que la aparición de retardos de y_t como variables explicativas producía este mismo tipo de dificultades.

Adoptamos la misma solución que allí, y que consiste en la utilización de *variables instrumentales*. Las estimaciones obtenidas serán tanto mejor cuanto mayores sean las correlaciones entre los instrumentos y las variables explicativas que los requieren. Sin embargo, dichas correlaciones no pueden ser muy elevadas, puesto que entonces los instrumentos tendrían una correlación apreciable con el término de error, conduciendo a inconsistencias.

A diferencia de un modelo uniecuacional, en un modelo de ecuaciones simultáneas tenemos candidatos naturales a ser utilizados como instrumentos: las variables exógenas del modelo que no aparecen como explicativas en la ecuación que queremos estimar.

En consecuencia, salvo que se dispusiera de instrumentos externos al modelo, será preciso disponer en el modelo de al menos tantas variables predeterminadas excluidas de la ecuación como variables endógenas estén incluidas como explicativas.

Esta condición puede escribirse: $k - k_1 \geq g - g_1$, que es, precisamente, la condición de orden para la identificación de la ecuación en estudio. Por tanto, una forma de interpretar la identificación de dicha ecuación es en el sentido de que exista un número suficiente de instrumentos disponibles para su estimación. La ecuación estará *exactamente identificada* si el número de instrumentos existentes es igual al de variables endógenas incluidas como explicativas, y estará *sobreidentificada* si el número de instrumentos disponibles es superior.

Este procedimiento de analizar la identificación de una ecuación del modelo es conveniente, sobre todo en el caso en que existan restricciones entre

los parámetros del modelo. Por ejemplo, consideremos la primera ecuación de un sistema:

$$y_{1t} = \gamma_{21}(y_{2t} + y_{3t}) + \beta_{11}x_{1t} + u_{1t}$$

y supongamos que en las ecuaciones correspondientes a las variables endógenas y_{2t} e y_{3t} existe otra variable predeterminada, x_{2t} . En tal caso, de esta primera ecuación falta tan sólo una variable, x_{2t} , por lo que $k - k_1 = 1$, que es inferior a 2, el número de ecuaciones del modelo menos una. Por tanto, parece que la ecuación está subidentificada. Sin embargo, la ecuación está exactamente identificada. En efecto, dado que se ha impuesto la restricción: $\gamma_{21} = \gamma_{31}$, hace falta un solo instrumento, para la variable $y_{2t} + y_{3t}$, y ese instrumento sería x_{2t} .

Si en el mismo modelo la primera ecuación fuese

$$y_{1t} = \gamma_{21}y_{2t} + \gamma_{31}(y_{3t} + x_{1t}) + u_{1t}$$

de nuevo, podría pensarse que está subidentificada, ya que la única variable predeterminada omitida es x_{2t} . Como hay dos variables endógenas explicativas (además con distintos coeficientes), podría creerse que no se dispone de suficientes instrumentos. Sin embargo, necesitamos un instrumento para y_{2t} y otro para $y_{3t} + x_{1t}$, y podemos utilizar x_{1t} y x_{2t} como instrumentos. La ecuación está exactamente identificada⁽²⁾.

Desde esta perspectiva, es fácil intuir el problema de sobreidentificación: si existe un exceso de instrumentos, entonces podrían utilizarse subconjuntos diferentes de los mismos para obtener el estimador de variables instrumentales. Cada subconjunto daría lugar a distintas estimaciones y, como consecuencia, el estimador de variables instrumentales no estaría unívocamente definido.

Por otra parte, cabe pensar que cuando existe tal exceso de información, entonces la utilización de tan sólo parte de ella para obtener uno de los posibles estimadores de variables instrumentales no sería eficiente. Tal intuición es correcta. De hecho, veremos en seguida que la matriz de covarianzas del estimador y, por tanto, su eficiencia, dependen del subconjunto de variables instrumentales utilizado; por ello, es natural preguntarse acerca de la elección del conjunto de variables instrumentales que hace que el estimador resultante tenga matriz de covarianzas mínima.

Tampoco es preciso recurrir a subconjuntos de variables instrumentales. Otra posibilidad consiste en utilizar $g_1 - 1$ combinaciones lineales de los $k - k_1$ instrumentos disponibles. Los casos antes citados, de elección de un subconjunto de $g_1 - 1$ instrumentos, son un caso particular de combinaciones lineales. La cuestión, finalmente, se reduce a la elección de las $g_1 - 1$ combinaciones lineales de los $k - k_1$ instrumentos disponibles que generen un estimador de variables instrumentales con una menor matriz de covarianzas. Pero para poder llevar a cabo tal elección, es preciso caracterizar primero las

⁽²⁾ Estas condiciones de identificación podrían asimismo analizarse mediante el procedimiento de la Sección 17.6.

propiedades estadísticas del estimador de variables instrumentales, en función de los instrumentos utilizados.

Definición 18.1. Sea $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{X}_1)$ el vector, de dimensión $g_1 + k_1 - 1$, de variables explicativas de la primera ecuación del sistema, y $\delta_1 = (\gamma_1, \beta_1)$ el vector de coeficientes de dicha ecuación. Un vector \mathbf{Z}_1^* de $g_1 + k_1 - 1$ variables es un *vector de variables instrumentales*, si se tiene:

$$a) \quad \text{plim} \frac{\mathbf{Z}_1^{*'} \mathbf{u}_1}{T} = \mathbf{0}_{g_1 + k_1 - 1}, \quad \text{que indica la ausencia (asintóticamente) de correlación entre variables instrumentales y término de error.}$$

$$b) \quad \text{plim} \frac{\mathbf{Z}_1^{*'} \mathbf{Z}_1}{T} = \Sigma_{z_1^* z_1}, \quad \text{no singular, lo que indica la existencia de correlación entre los vectores } \mathbf{Z}_1^* \text{ y } \mathbf{Z}_1.$$

$$c) \quad \text{plim} \frac{\mathbf{Z}_1^{*'} \mathbf{Z}_1^*}{T} = \Sigma_{z_1^* z_1^*}, \quad \text{definida positiva.}$$

Como puede verse, las variables que aparecen en el vector \mathbf{X}_1 , por ser predeterminadas, satisfacen la condición a) anterior, por lo que si existen sus momentos de segundo orden, pueden utilizarse como variables instrumentales de sí mismas. Así, en general, se trata de utilizar variables instrumentales tan sólo para las variables incluidas en el vector \mathbf{Y}_1 .

Definición 18.2. Un *estimador de variables instrumentales* es una solución al sistema de ecuaciones normales:

$$(\mathbf{Z}_1^{*'} \mathbf{Z}_1) \hat{\delta}_1 = \mathbf{Z}_1^{*'} \mathbf{y}_1 \quad [18.3]$$

De acuerdo con el comentario anterior, el vector \mathbf{Z}_1^* estará formado por $\mathbf{Z}_1^* = (\mathbf{Y}_1^*, \mathbf{X}_1)$, es decir, las variables en \mathbf{X}_1 hacen de instrumentos de ellas mismas. En consecuencia, el sistema de ecuaciones normales puede escribirse:

$$\begin{pmatrix} \mathbf{Y}_1^{*'} \mathbf{Y}_1 & \mathbf{Y}_1^{*'} \mathbf{X}_1 \\ \mathbf{X}_1' \mathbf{Y}_1 & \mathbf{X}_1' \mathbf{X}_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1^{*'} \mathbf{y}_1 \\ \mathbf{X}_1' \mathbf{y}_1 \end{pmatrix} \quad [18.4]$$

El estimador MCO, por otra parte, es una solución al sistema de ecuaciones normales:

$$\begin{pmatrix} \mathbf{Y}_1' \mathbf{Y}_1 & \mathbf{Y}_1' \mathbf{X}_1 \\ \mathbf{X}_1' \mathbf{Y}_1 & \mathbf{X}_1' \mathbf{X}_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1' \mathbf{y}_1 \\ \mathbf{X}_1' \mathbf{y}_1 \end{pmatrix}$$

Así, puede verse que, con respecto al estimador MCO, el estimador de variables instrumentales sustituye el vector \mathbf{Y}_1 por el vector de instrumentos \mathbf{Y}_1^* en algunas posiciones en la formulación de las ecuaciones normales.

La calificación que aparece en la definición anterior de «Un estimador de variables instrumentales», en vez de «El estimador de variables instrumentales», se debe a que podrían formularse tantos sistemas de ecuaciones normales como subconjuntos de variables instrumentales pudieran encontrarse.

El estimador de variables instrumentales satisface unas condiciones de ortogonalidad similares a las satisfechas por el estimador MCO:

Definición 18.3. El vector de residuos del estimador de variables instrumentales, que denotaremos por $\hat{\mathbf{u}}_{VI}$, es el vector de residuos generados por este estimador, pero con las variables originales del modelo (no los instrumentos utilizados para ellas):

$$\hat{\mathbf{u}}_{VI} = \mathbf{y}_1 - \mathbf{Z}_1 \hat{\boldsymbol{\delta}}_1^{VI} = \mathbf{y}_1 - \mathbf{Y}_1 \hat{\boldsymbol{\gamma}}_1^{VI} - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1^{VI}$$

Proposición 18.3. El vector de residuos del estimador de variables instrumentales es *ortogonal* a cada una de las variables instrumentales utilizadas en la estimación.

Demostración:

$$\mathbf{Z}_1^{*'} \hat{\mathbf{u}}_{VI} = \mathbf{Z}_1^{*'} (\mathbf{y}_1 - \mathbf{Z}_1 \hat{\boldsymbol{\delta}}_1^{VI}) = \mathbf{Z}_1^{*'} \mathbf{y}_1 - \mathbf{Z}_1^{*'} \mathbf{Z}_1 (\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*'} \mathbf{y}_1 = \mathbf{0}_{k_1 + g_1 - 1}$$

de acuerdo con [18.3].

En general, el estimador de variables instrumentales no es insesgado. Para que lo fuese, debería ocurrir que:

$$E[(\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*'} \mathbf{u}_1] = \mathbf{0}$$

pero ello implica conocer las propiedades, en muestras finitas, de las correlaciones entre \mathbf{Z}_1^* y \mathbf{u}_1 , así como sus momentos cruzados con \mathbf{Z}_1 ⁽³⁾. Dado que las correlaciones entre las componentes del vector \mathbf{Z}_1^* y \mathbf{u}_1 tienden a cero al aumentar el tamaño muestral, mientras que las correlaciones entre \mathbf{Z}_1^* y \mathbf{Z}_1 permanecen finitas, es fácil ver que el sesgo del estimador de variables instrumentales, aun existiendo, tiende a cero al aumentar el tamaño muestral:

Proposición 18.4. El estimador de variables instrumentales es consistente.

Demostración. Basta escribir

$$\hat{\boldsymbol{\delta}}_1^{VI} = (\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*'} (\mathbf{Z}_1 \boldsymbol{\delta}_1 + \mathbf{u}_1) = \boldsymbol{\delta}_1 + (\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1} \mathbf{u}_1$$

para obtener:

$$plim(\hat{\boldsymbol{\delta}}_1^{VI}) = \boldsymbol{\delta}_1 + plim \left[\left(\frac{\mathbf{Z}_1^{*'} \mathbf{Z}_1}{T} \right)^{-1} \right] plim \frac{\mathbf{Z}_1^{*'} \mathbf{u}}{T} = \boldsymbol{\delta}_1 + (\boldsymbol{\Sigma}_{z_1^* z_1})^{-1} \mathbf{0}_{g_1 + k_1 - 1} = \boldsymbol{\delta}_1$$

Proposición 18.5. Si la matriz de covarianzas del término de error de la ecuación es $\text{Var}(\mathbf{u}_1) = \sigma_{u_1}^2 \mathbf{I}_T$, entonces la matriz de covarianzas asintótica del estimador de variables instrumentales es:

$$\sigma_{u_1}^2 (\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1} (\mathbf{Z}_1^{*'} \mathbf{Z}_1^*) [(\mathbf{Z}_1^{*'} \mathbf{Z}_1)^{-1}]'$$

⁽³⁾ Que son imposibles de caracterizar sin añadir una muy detallada información acerca de las correlaciones entre variables explicativas y término de error.

Demostración. El estimador de variables instrumentales $\hat{\delta}_1^{VI}$ puede escribirse:

$$\hat{\delta}_1^{VI} = (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*\prime} \mathbf{y}_1 = (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*\prime} (\mathbf{Z}_1 \delta_1 + \mathbf{u}_1) = \delta_1 + (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1)^{-1} \mathbf{Z}_1^{*\prime} \mathbf{u}_1$$

por lo que:

$$\sqrt{T}(\hat{\delta}_1^{VI} - \delta) = \left(\frac{\mathbf{Z}_1^{*\prime} \mathbf{Z}_1}{T} \right)^{-1} \frac{\mathbf{Z}_1^{*\prime} \mathbf{u}_1}{\sqrt{T}}$$

Ahora bien, de acuerdo con nuestros supuestos, el teorema de Mann-Wald asegura que $\frac{\mathbf{Z}_1^{*\prime} \mathbf{u}_1}{\sqrt{T}}$ converge en distribución a una $N(\mathbf{0}_{g_1+k_1-1}, \sigma_{u_1}^2 \Sigma_{z_1^* z_1^*})$. En consecuencia, $\sqrt{T}(\hat{\delta}_1^{VI} - \delta_1)$ converge en distribución al producto:

$$(\Sigma_{z_1^* z_1^*})^{-1} N(\mathbf{0}_{g_1+k_1-1}, \sigma_{u_1}^2 \Sigma_{z_1^* z_1^*})$$

es decir, a una variable $N(\mathbf{0}_{g_1+k_1-1}, \sigma_{u_1}^2 (\Sigma_{z_1^* z_1^*})^{-1} \Sigma_{z_1^* z_1^*} (\Sigma_{z_1^* z_1^*})^{-1})$.

Este resultado equivale a utilizar como matriz de covarianzas del estimador de variables instrumentales en muestras finitas la matriz:

$$\frac{1}{T} \sigma_{u_1}^2 (\Sigma_{z_1^* z_1^*})^{-1} \Sigma_{z_1^* z_1^*} (\Sigma_{z_1^* z_1^*})^{-1}$$

Ahora bien, todas las matrices que aparecen como factores son desconocidas en la práctica, por lo que deben estimarse por sus momentos muestrales: $\frac{\mathbf{Z}_1^{*\prime} \mathbf{Z}_1}{T}$ y $\frac{\mathbf{Z}_1^{*\prime} \mathbf{Z}_1^*}{T}$, para obtener finalmente la matriz de covarianzas del estimador de variables instrumentales:

$$\sigma_{u_1}^2 (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1)^{-1} (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1^*) [(\mathbf{Z}_1^{*\prime} \mathbf{Z}_1)^{-1}]' \quad [18.5]$$

aunque no es sino una aproximación en muestras finitas.

La expresión [18.5] ilustra claramente el hecho de que la matriz de covarianzas del estimador de variables instrumentales depende del parámetro $\sigma_{u_1}^2$, del vector de variables \mathbf{Z}_1 incluidas en la ecuación, y del vector de instrumentos \mathbf{Z}_1^* . Los dos primeros son característicos de la ecuación que se pretende estimar, mientras que el vector \mathbf{Z}_1^* es elección del investigador, quien puede por tanto plantearse la posibilidad de escoger un vector \mathbf{Z}_1^* de modo que dicha matriz de covarianzas sea la menor posible, lo que discutimos en la próxima sección.

Ya hemos mencionado que el estimador de variables instrumentales no es, en realidad, un único estimador, sino toda una familia de ellos, dependiendo de cuáles sean los instrumentos elegidos. Tan sólo en el caso de una ecuación exactamente identificada está el estimador de variables instrumentales definido de modo único:

Proposición 18.6. El estimador MCI de una ecuación exactamente identificada es el estimador de variables instrumentales, que utiliza las variables predeterminadas excluidas de la ecuación como instrumentos de las variables endógenas incluidas en la misma.

Demostración. Por estar exactamente identificada, hay exactamente tantas variables predeterminadas excluidas de la ecuación como variables endógenas incluidas como explicativas, es decir: $g_1 - 1 = k - k_1$. Por tanto, a no ser que se utilizase información ajena al modelo, el estimador de variables instrumentales está unívocamente definido y utiliza como instrumentos de las $g_1 - 1$ variables endógenas Y_1 incluidas como explicativas, las $k - k_1$ variables predeterminadas X_2 excluidas de la ecuación. Veamos que dicho estimador coincide con el estimador MCI. En efecto, las ecuaciones normales para el estimador de variables instrumentales serán en este caso:

$$\begin{pmatrix} X_2' Y_1 & X_2' X_1 \\ X_1' Y_1 & X_1' X_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} X_2' y_1 \\ X_1' y_1 \end{pmatrix}$$

que, como puede apreciarse, es el mismo sistema de ecuaciones [18.2] cuya solución define al estimador de mínimos cuadrados indirectos. Por tanto, en ecuaciones exactamente identificadas, ambos estimadores coinciden.

Proposición 18.7. El estimador MCI de una ecuación exactamente identificada es consistente.

Demostración. Es consecuencia inmediata de las Proposiciones 18.4 y 18.6.

Una segunda implicación de la discusión anterior es que la matriz de covarianzas del estimador MCI de una ecuación exactamente identificada, obtenido a partir de una muestra finita, puede aproximarse por la expresión [18.5], donde ahora $Z_1 = (Y_1, X_1)$ y $Z_1^* = (X_2, X_1)$.

Una propiedad adicional de las ecuaciones normales anteriores es que podrían intercambiarse los vectores X_1 y X_2 , y se obtendría el mismo estimador de mínimos cuadrados indirectos. Es decir, supongamos que algunas de las variables en X_1 hacen de instrumentos de las variables en Y_1 , mientras que X_2 , junto con el resto de las variables en X_1 , hacen de instrumentos de las variables en X_1 . En tal caso, el estimador resultante sería:

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = \left[\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} (Y_1; X_1) \right]^{-1} (X_1'; X_2') y_1 = \begin{pmatrix} X_1' Y_1 & X_1' X_1 \\ X_2' Y_1 & X_2' X_1 \end{pmatrix}^{-1} \begin{pmatrix} X_1' y_1 \\ X_2' y_1 \end{pmatrix}$$

que coincide, efectivamente, con la expresión que antes obtuvimos para el estimador de mínimos cuadrados indirectos.

Debido a esta propiedad, puede darse la siguiente definición alternativa del estimador de mínimos cuadrados indirectos:

Definición 18.4. El estimador de mínimos cuadrados indirectos es el estimador de variables instrumentales que utiliza el vector \mathbf{X} formado por *todas* las variables predeterminadas de *todas* las ecuaciones del modelo como instrumentos del vector $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{X}_1)$ de variables explicativas de la ecuación a estimar.

Por tanto, puede obtenerse una nueva expresión alternativa de dicho estimador:

$$\hat{\delta}_1^{\text{MCI}} = (\mathbf{X}'\mathbf{Z}_1)^{-1}\mathbf{X}'\mathbf{y}_1 \quad [18.6]$$

Ejemplo 18.4. Volviendo al Ejemplo 18.3, el resultado que acabamos de ver muestra que las estimaciones de mínimos cuadrados indirectos de la primera ecuación pueden también obtenerse sin necesidad de estimar primero la forma reducida del modelo, pues dicho estimador es el que resulta de utilizar en la primera ecuación la variable predeterminada omitida x_{3t} como variable instrumental:

$$\begin{aligned} \begin{pmatrix} \hat{\gamma}_{21} \\ \hat{\beta}_{11} \\ \hat{\beta}_{21} \end{pmatrix} &= \begin{pmatrix} x'_3 y_2 & x'_3 x_1 & x'_3 x_2 \\ x'_1 y_2 & x'_1 x_1 & x'_1 x_2 \\ x'_2 y_2 & x'_2 x_1 & x'_2 x_2 \end{pmatrix}^{-1} \begin{pmatrix} x'_3 y_1 \\ x'_1 y_1 \\ x'_2 y_1 \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 4 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} -1/8 \\ 13/16 \\ 1/2 \end{pmatrix} \end{aligned}$$

que coinciden con las estimaciones que obtuvimos previamente a partir de los parámetros de la forma reducida.

La matriz de covarianzas de este estimador es:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 4 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 0 \\ 1 & 0 & 1 \end{pmatrix}^{-1} = (1/64) \begin{pmatrix} 4 & -2 & -16 \\ -2 & 17 & 8 \\ -16 & 8 & 128 \end{pmatrix}$$

Ahora podemos describir la dificultad que presenta el estimador MCI en una ecuación sobreidentificada; como dijimos en la Sección 18.2, en tal caso el sistema

$$\hat{\Pi} \begin{pmatrix} -1 \\ \gamma_1 \\ \mathbf{0}_{g_2} \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}_1; \mathbf{Y}_1; \mathbf{Y}_2) \begin{pmatrix} -1 \\ \gamma_1 \\ \mathbf{0}_{g_2} \end{pmatrix} = \begin{pmatrix} -\hat{\beta}_1 \\ \mathbf{0}_{k_2} \end{pmatrix}$$

tiene más ecuaciones (k) que incógnitas ($g_1 + k_1 - 1$), por lo que hay distintas formas de resolver el sistema, puesto que cada subconjunto posible de $k_1 + g_1 - 1$ ecuaciones nos proporcionaría una solución $(\hat{\gamma}_1, \hat{\beta}_1)$.

El sistema anterior puede también escribirse:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1\hat{\gamma}_1 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1\hat{\beta}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1$$

donde se ha utilizado la igualdad:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \begin{pmatrix} \hat{\beta}_1 \\ \mathbf{0}_{k_2} \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1\hat{\beta}_1$$

El lector debe asegurarse de que las dimensiones de las matrices que aparecen en este producto son compatible entre sí, y que de este modo se tiene un sistema de k ecuaciones con $k_1 + g_1 - 1$ incógnitas, las estimaciones $(\hat{\gamma}, \hat{\beta}_1)$.

La forma de seleccionar un subconjunto de $k_1 + g_1 - 1$ ecuaciones de este sistema es premultiplicando por una matriz \mathbf{A} de dimensión $(k_1 + g_1 - 1) \times k$ de la forma:

$$(\mathbf{0}_{k_1+g_1-1} \quad \mathbf{1}_{k_1+g_1-1} \quad \mathbf{1}_{k_1+g_1-1} \quad \mathbf{0}_{k_1+g_1-1} \quad \cdots \quad \mathbf{0}_{k_1+g_1-1} \quad \mathbf{1}_{k_1+g_1-1})$$

donde las columnas de unos corresponden a las ecuaciones que se seleccionan del sistema inicial de k ecuaciones. Así se obtiene el subsistema:

$$\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1\hat{\gamma}_1 + \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1\hat{\beta}_1 = \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1$$

con solución:

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\beta}_1 \end{pmatrix} = ([\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'](\mathbf{Y}_1; \mathbf{X}_1))^{-1}[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}_1$$

que es el estimador de variables instrumentales que utiliza como instrumentos el vector:

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'$$

Las matrices de covarianzas de cada uno de los posibles estimadores MCI de una ecuación sobreidentificada se obtendrían de acuerdo con lo que se expuso tras la Proposición 18.5. En la sección siguiente introducimos un estimador de variables instrumentales que, en ecuaciones sobreidentificadas, tiene una matriz de covarianzas inferior a la de cualquiera de los distintos estimadores MCI.

18.4. EL ESTIMADOR DE MÍNIMOS CUADRADOS EN DOS ETAPAS

El estimador de mínimos cuadrados en dos etapas puede utilizarse para estimar los coeficientes de ecuaciones exactamente identificadas o sobreidenti-

ficadas. En este último caso, el procedimiento genera un único estimador, que puede interpretarse como una combinación de los diversos estimadores que podrían generarse por el método de mínimos cuadrados indirectos (véase el Problema 18.4, al final del capítulo). Como veremos, no es una combinación lineal cualquiera, pues satisface una importante condición de optimalidad.

La estrategia seguida por el método de mínimos cuadrados en dos etapas consiste en construir una regresión auxiliar para cada una de las variables endógenas incluidas como explicativas. Cada una de estas regresiones auxiliares tiene como variable a explicar a una de dichas variables endógenas y como variables explicativas *todas las variables predeterminadas* del modelo de ecuaciones simultáneas. Supongamos que estimamos estas regresiones auxiliares y utilizamos dichas estimaciones para «generar» datos, a lo largo del período muestral, para la variable endógena correspondiente, ignorando la información muestral disponible acerca de dicha variable.

Si denotamos por \hat{Y}_1 la matriz $T \times (g_1 - 1)$ de observaciones generada de este modo por las regresiones estimadas de cada una de las variables en Y_1 sobre *todas las variables en X*, se tiene:

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

Estas series así generadas tienen una doble propiedad:

1. Por ser las «predicciones» muestrales de una variable endógena (una variable en Y_1) están correlacionadas con dicha variable.
2. Por ser una combinación lineal de las variables predeterminadas con matriz de coeficientes igual a $(X'X)^{-1}X'Y_1$, las variables \hat{Y}_1 estarán incorrelacionadas con el término de error de la ecuación en estudio.

En una segunda etapa, las predicciones obtenidas de estas regresiones auxiliares se utilizan en la ecuación de partida en vez de las variables endógenas que en ella aparecen como variables explicativas. Hecha esta sustitución, se estima la ecuación por mínimos cuadrados ordinarios, por lo que las estimaciones mínimo-cuadráticas de la segunda etapa son las soluciones al sistema de ecuaciones normales:

$$\begin{pmatrix} \hat{Y}_1' \hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1' \hat{Y}_1 & X_1' X_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1^{MC2E} \\ \hat{\beta}_1^{MC2E} \end{pmatrix} = \begin{pmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{pmatrix} \quad [18.7]$$

donde y_1 es el vector de observaciones de la variable endógena de la ecuación. El estimador de mínimos cuadrados ordinarios de dicha ecuación con estas nuevas variables explicativas se conoce como *el estimador de mínimos cuadrados en dos etapas* (MC2E) de la ecuación. La mención a las dos etapas proviene de los dos niveles en los que se utilizan mínimos cuadrados: a) una primera etapa, en que se utilizan mínimos cuadrados para estimar las regresiones auxiliares de las que se construirá la matriz \hat{Y}_1 , y b) la estimación por mínimos cuadrados de la regresión final, en la que aparecen como explicativas las variables \hat{Y}_1 .

Notemos que el vector de variables Y_1 puede escribirse:

$$Y_1 = \hat{Y}_1 + \hat{E}_1$$

donde \hat{E}_1 es la matriz $T \times (g_1 - 1)$ de residuos MCO de las regresiones auxiliares. En dicha descomposición se tienen las siguientes propiedades:

1. $X' \hat{E}_1 = 0$, debido a que en toda regresión mínimo-cuadrática, los residuos son ortogonales a las variables explicativas.

2. $\hat{Y}_1' \hat{E}_1 = 0$, como consecuencia de lo anterior, ya que $\hat{Y}_1' \hat{E}_1 = Y_1' X(X'X)^{-1} X' \hat{E}_1 = 0$.

Por tanto:

$$3. \hat{Y}_1' \hat{Y}_1 = Y_1' X(X'X)^{-1} X' \hat{Y}_1 = Y_1' X(X'X)^{-1} X' (Y_1 - \hat{E}_1) = \hat{Y}_1' Y_1.$$

$$4. \hat{Y}_1' X_1 = Y_1' X_1 - \hat{E}_1' X_1 = Y_1' X_1.$$

Con todo ello, las ecuaciones normales que generan el estimador de mínimos cuadrados en dos etapas quedan:

$$\begin{pmatrix} Y_1' X(X'X)^{-1} X' Y_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1^{MC2E} \\ \hat{\beta}_1^{MC2E} \end{pmatrix} = \begin{pmatrix} Y_1' X(X'X)^{-1} X' y_1 \\ X_1' y_1 \end{pmatrix} \quad [18.8]$$

Esta expresión matricial es importante, pues muestra que, en realidad, para obtener el estimador de mínimos cuadrados en dos etapas no es preciso estimar las regresiones auxiliares con las que introdujimos dicho estimador, sino que pueden utilizarse directamente las *ecuaciones normales* [18.8].

La siguiente proposición muestra que el estimador MC2E no es sino un estimador de variables instrumentales para una determinada elección del vector de instrumentos:

Proposición 18.8. El estimador de mínimos cuadrados en dos etapas es el estimador de variables instrumentales que utiliza el vector $Z_1^* = (\hat{Y}_1, X_1)$ como vector de instrumentos del vector de variables explicativas $Z_1 = (Y_1, X_1)$.

Demostración. Si sustituimos Y_1^* por \hat{Y}_1 en el sistema [18.4] de ecuaciones normales del estimador de variables instrumentales, se tiene:

$$\begin{pmatrix} \hat{Y}_1' Y_1 & \hat{Y}_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{pmatrix} \begin{pmatrix} \hat{\gamma}_1^{VI} \\ \hat{\beta}_1^{VI} \end{pmatrix} = \begin{pmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{pmatrix}$$

Ahora bien, como $\hat{Y}_1' \hat{Y}_1 = \hat{Y}_1' Y_1$ y $X_1' \hat{Y}_1 = X_1' Y_1$, este sistema de ecuaciones normales coincide con el sistema [18.7] correspondiente al estimador MC2E. Como consecuencia, ambos estimadores coinciden.

Ejemplo 18.5. A efectos ilustrativos, veamos el significado de la Proposición 18.8 en el caso de la primera ecuación de un modelo con dos variables endógenas:

$$y_{1t} = \gamma_{21} y_{2t} + \beta_{11} x_{1t} + \beta_{21} x_{2t} + u_{1t} \quad [18.9]$$

y supongamos que en la segunda ecuación existe otra variable predeterminada, x_{3t} . Un posible estimador de variables instrumentales de [18.9] se obtendría al utilizar x_{3t} como instrumento de y_{2t} ; éste sería, además, el estimador de mínimos cuadrados indirectos. Dicho estimador impondría las siguientes condiciones de ortogonalidad entre variables y residuos:

$$\Sigma_1^T x_{3t} \hat{u}_{1t} = \Sigma_1^T x_{1t} \hat{u}_{1t} = \Sigma_1^T x_{2t} \hat{u}_{1t} = 0$$

Si además hubiese un término independiente, se estaría exigiendo, adicionalmente, $\Sigma_1^T \hat{u}_{1t} = 0$.

Por otra parte, el estimador MC2E de esta ecuación comienza con una regresión auxiliar de y_{2t} sobre todas las variables predeterminadas del sistema, para obtener de ella la predicción \hat{y}_{2t} . Esta predicción está relacionada con la verdadera variable y_{2t} por $y_{2t} = \hat{y}_{2t} + \hat{e}_{2t}$, donde \hat{e}_{2t} es el residuo de la regresión auxiliar. En consecuencia, \hat{e}_{2t} es ortogonal a las variables explicativas de dicha regresión, que son todas las variables predeterminadas del sistema: x_{1t} , x_{2t} y x_{3t} . A continuación, se estima por MCO la ecuación:

$$y_{1t} = \gamma_{21} \hat{y}_{2t} + \beta_{11} x_{1t} + \beta_{21} x_{2t} + (u_{1t} + \gamma_{21} \hat{e}_{2t})$$

imponiéndose, por tanto, las condiciones de ortogonalidad:

$$\Sigma_1^T \hat{y}_{2t} (\hat{u}_{1t} + \gamma_{21} \hat{e}_{2t}) = \Sigma_1^T x_{1t} (\hat{u}_{1t} + \gamma_{21} \hat{e}_{2t}) = \Sigma_1^T x_{2t} (\hat{u}_{1t} + \gamma_{21} \hat{e}_{2t}) = 0 \quad [18.10]$$

Ahora bien, por ser el residuo \hat{e}_{2t} ortogonal a todas las variables predeterminadas del sistema, también es ortogonal a la predicción \hat{y}_{2t} , por ser una combinación lineal de las variables explicativas de la regresión auxiliar. Por tanto, se tiene:

$$\Sigma_1^T \hat{y}_{2t} \hat{e}_{2t} = \Sigma_1^T x_{1t} \hat{e}_{2t} = \Sigma_1^T x_{2t} \hat{e}_{2t} = \Sigma_1^T x_{3t} \hat{e}_{2t} = 0$$

con las que las condiciones de ortogonalidad [18.10] se reducen a:

$$\Sigma_1^T \hat{y}_{2t} \hat{u}_{1t} = \Sigma_1^T x_{1t} \hat{u}_{1t} = \Sigma_1^T x_{2t} \hat{u}_{1t} = 0 \quad [18.11]$$

que caracterizan al estimador MC2E de [18.9].

Por otra parte, si se utiliza directamente \hat{y}_{2t} como variable instrumental de y_{2t} en [18.11], entonces las condiciones de ortogonalidad de la Proposición 18.3 son precisamente las mismas [18.11], asimismo iguales a las del estimador de mínimos cuadrados indirectos, que son $\Sigma_1^T x_{3t} \hat{u}_{1t} = \Sigma_1^T x_{1t} \hat{u}_{1t} = \Sigma_1^T x_{2t} \hat{u}_{1t} = 0$. Los tres estimadores coinciden en la ecuación [18.11]. Esta equivalencia no es casual, y está producida por estar dicha ecuación exactamente identificada.

De paso, hemos comprobado que el término de error de la regresión estimada en la segunda etapa del método, $u_{1t} + \gamma_{21} \hat{e}_{2t}$, es diferente del término de error del modelo original, u_{1t} . Es por esta razón que *hay que evitar utilizar las desviaciones típicas estimadas en dicha segunda etapa*. De hecho, el vector

de residuos de tal regresión no tiene una interpretación clara. Por el contrario, como se vio en la Definición 18.3, el vector de residuos correspondiente al estimador MC2E es:

$$\hat{u}_1^{\text{MC2E}} = \mathbf{y}_1 - \mathbf{Y}_1 \hat{\gamma}_1^{\text{MC2E}} - \mathbf{X}_1 \hat{\beta}_1^{\text{MC2E}}$$

que utiliza las variables originales, y no los instrumentos, y son también los residuos que deben utilizarse en la estimación del parámetro σ_1^2 . Dicha estimación será el cociente entre la *suma residual* del modelo (definida como la suma de los cuadrados de los residuos anteriores) y el número de grados de libertad, $T - g_1 - k_1 + 1$.

Proposición 18.9. El estimador MC2E es consistente.

Demostración. Es consecuencia inmediata de la Proposición 18.4, por ser el estimador MC2E un estimador de variables instrumentales.

La matriz de covarianzas asintótica del estimador MC2E podría obtenerse directamente, a partir de la definición del estimador, o también aprovechando la propiedad de este estimador de ser un estimador de variables instrumentales. Así, su matriz de covarianzas asintótica (es decir, en muestras infinitas) coincidirá con la del estimador de variables instrumentales, cuando los instrumentos utilizados para las variables \mathbf{Y}_1 son el vector $\hat{\mathbf{Y}}_1$. Ahora bien, en tal caso, las propiedades antes vistas prueban las siguientes igualdades:

$$\mathbf{Z}_1^* \mathbf{Z}_1^* = \begin{pmatrix} \hat{\mathbf{Y}}_1' \hat{\mathbf{Y}}_1 & \hat{\mathbf{Y}}_1' \mathbf{X}_1 \\ \mathbf{X}_1' \hat{\mathbf{Y}}_1 & \mathbf{X}_1' \mathbf{X}_1 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{Y}}_1' \mathbf{Y}_1 & \hat{\mathbf{Y}}_1' \mathbf{X}_1 \\ \mathbf{X}_1' \mathbf{Y}_1 & \mathbf{X}_1' \mathbf{X}_1 \end{pmatrix} = \mathbf{Z}_1^* \mathbf{Z}_1$$

por lo que la matriz de covarianzas del estimador MC2E resulta ser:

$$\begin{aligned} \text{Var}(\hat{\delta}_1^{\text{VI}}) &= \sigma_1^2 (\mathbf{Z}_1^* \mathbf{Z}_1)^{-1} (\mathbf{Z}_1^* \mathbf{Z}_1^*) (\mathbf{Z}_1^* \mathbf{Z}_1)^{-1} = \\ &= \sigma_1^2 (\mathbf{Z}_1^* \mathbf{Z}_1^*)^{-1} = \sigma_1^2 (\mathbf{Z}_1^* \mathbf{Z}_1)^{-1} = \\ &= \sigma_1^2 \begin{pmatrix} \mathbf{Y}_1' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_1 & \mathbf{Y}_1' \mathbf{X}_1 \\ \mathbf{X}_1' \mathbf{Y}_1 & \mathbf{X}_1' \mathbf{X}_1 \end{pmatrix}^{-1} \end{aligned}$$

Es importante hacer hincapié en que la igualdad $\mathbf{Z}_1^* \mathbf{Z}_1^* = \mathbf{Z}_1^* \mathbf{Z}_1$ sólo es válida para la particular elección de instrumentos llevada a cabo al obtener el estimador MC2E, mientras que no es cierta, en general, para cualquier estimador de variables instrumentales. Del mismo modo, el lector debe observar que, sin embargo, $\mathbf{Z}_1^* \mathbf{Z}_1^* \neq \mathbf{Z}_1^* \mathbf{Z}_1$, incluso para el estimador MC2E.

Proposición 18.10. El estimador MC2E puede también expresarse como:

$$\hat{\delta}_1^{\text{MC2E}} = [\mathbf{Z}_1' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_1]^{-1} [\mathbf{Z}_1' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_1]$$

y su matriz de covarianzas puede aproximarse en muestras suficientemente grandes por:

$$\text{Var}(\hat{\delta}_1^{\text{MC2E}}) = \sigma_1^2 [\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1]^{-1}$$

Demostración. En efecto, utilizando la igualdad $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}_1 = \mathbf{X}_1$ (véase el Problema 18.19) se tiene:

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{Y}_1; \mathbf{X}_1) = (\hat{\mathbf{Y}}_1; \mathbf{X}_1) = \mathbf{Z}_1^*$$

y, por tanto,

$$\mathbf{Z}_1^{*\prime} \mathbf{Z}_1^* = \mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1 = \mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1$$

por lo que el estimador MC2E puede escribirse:

$$\hat{\delta}_1^{\text{MC2E}} = (\mathbf{Z}_1^{*\prime} \mathbf{Z}_1^*)^{-1} \mathbf{Z}_1^{*\prime} \mathbf{y}_1 = [\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1]^{-1} [\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_1] \quad [18.12]$$

con matriz de covarianzas:

$$\text{Var}(\hat{\delta}_1^{\text{MC2E}}) = \sigma_1^2 [\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1]^{-1}$$

Como es habitual, el parámetro σ_1^2 es desconocido, por lo que habrá que estimarlo. Por analogía con el modelo uniecuacional, su estimación viene dada por la expresión:

$$\hat{\sigma}_1^2 = \frac{(\mathbf{y}_1 - \mathbf{Y}_1 \hat{\gamma}_1 - \mathbf{X}_1 \hat{\beta}_1)' (\mathbf{y}_1 - \mathbf{Y}_1 \hat{\gamma}_1 - \mathbf{X}_1 \hat{\beta}_1)}{(T - g_1 - k_1 + 1)}$$

donde es importante observar que, una vez más, en el cálculo de la suma residual se utilizan las variables del modelo original y no sus instrumentos.

La expresión [18.12] puede utilizarse para demostrar que el estimador de mínimos cuadrados en dos etapas es sesgado. En efecto:

$$E(\hat{\delta}_1^{\text{MC2E}}) = \delta_1 + E([\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_1]^{-1} [\mathbf{Z}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_1]) \quad [18.13]$$

y poco puede afirmarse acerca de la esperanza de los productos de variables aleatorias que aparecen en el miembro derecho de esta expresión.

Una importante propiedad del estimador MC2E es que satisface una condición de optimalidad entre una subclase importante de todos los estimadores de variables instrumentales que podrían considerarse: Las combinaciones lineales de las variables predeterminadas utilizadas por el método de mínimos cuadrados en dos etapas son las que generan un estimador de variables instrumentales con menor varianza, es decir, de mayor precisión:

Proposición 18.11. El estimador MC2E es el de mínima varianza de entre todos los estimadores de variables instrumentales que utilizan como instrumentos combinaciones lineales de las variables predeterminadas del modelo.

Demostración. El estimador MC2E utiliza como instrumentos de las variables endógenas incluidas en la primera ecuación $\hat{Y}_1 = X(X'X)^{-1}X'Y_1$, que es un vector de $g_1 - 1$ combinaciones lineales de las k variables predeterminadas del sistema representadas por la matriz X , con matriz de coeficientes $(X'X)^{-1}X'Y_1$, de dimensión $k \times (g_1 - 1)$.

Consideremos ahora un estimador de variables instrumentales $\tilde{\delta}_1$, que utilice como instrumentos una combinación lineal diferente. Dicho estimador utilizará un vector de instrumentos que podrá escribirse genéricamente:

$$Z_1^* = XD$$

donde D es la matriz de dimensión $k \times (k_1 + g_1 - 1)$ formada por los coeficientes de las variables X en las $k_1 + g_1 - 1$ combinaciones lineales a utilizar como instrumentos. Por ejemplo, en el estimador de mínimos cuadrados en dos etapas, $D = (X'X)^{-1}X'(Y_1; X_1)$. Nótese que Z_1^* es una matriz de instrumentos válidos, pues siendo combinaciones lineales de las variables predeterminadas del sistema, son ortogonales al término de error de la ecuación que se pretende estimar (y de todas las demás ecuaciones, también).

Sustituyendo estos instrumentos en la expresión general de la matriz de covarianzas del estimador de variables instrumentales se tiene:

$$[\text{Var}(\tilde{\delta}_1)]^{-1} = \sigma_1^{-2}(D'X'Z_1)'(D'X'XD)^{-1}(D'X'Z_1)$$

mientras que, como vimos anteriormente,

$$[\text{Var}(\hat{\delta}_1^{\text{MC2E}})]^{-1} = \sigma_1^{-2}(Z_1'X(X'X)^{-1}X'Z_1)$$

Si descomponemos la matriz $X'X$ como el producto QQ' donde Q es una matriz no singular, de dimensión $k \times k$, entonces la diferencia entre ambas matrices inversas es:

$$\begin{aligned} & [\text{Var}(\hat{\delta}_1^{\text{MC2E}})]^{-1} - [\text{Var}(\tilde{\delta}_1)]^{-1} = \\ & = \sigma_1^{-2}[Z_1'X(QQ')^{-1}X'Z_1 - Z_1'X(Q^{-1})'Q'D(D'QQ'D)^{-1}D'Q(Q^{-1})X'Z_1] = \end{aligned}$$

y denotamos $C = Q'D$ y $A = Q^{-1}X'Z_1$, ambas matrices de dimensión $k \times (g_1 + k_1 - 1)$, entonces:

$$= \sigma_1^{-2}[A'A - A'C(C'C)^{-1}C'A] = \sigma_1^{-2}A'[I_{k-1} - C(C'C)^{-1}C]A$$

Este último corchete es una matriz definida positiva, como se pide demostrar en un ejercicio al final del capítulo. En consecuencia:

$$\text{Var}(\tilde{\delta}_1) > \text{Var}(\hat{\delta}_1^{\text{MC2E}})$$

La siguiente proposición muestra la equivalencia de los estimadores MC2E y MCI en ecuaciones exactamente identificadas:

Proposición 18.12. En una ecuación exactamente identificada, el estimador MC2E coincide con el estimador MCI.

Demostración. Con identificación exacta, $k = g_1 + k_1 - 1$, por lo que la matriz $\mathbf{X}'\mathbf{Z}_1$, que es de dimensión $k \times (g_1 + k_1 - 1)$, resulta en este caso cuadrada. Supuesto que no sea singular, entonces la expresión [18.12] para el estimador MC2E se convierte en:

$$\hat{\delta}_1^{\text{MC2E}} = (\mathbf{X}'\mathbf{Z}_1)^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{Z}_1'\mathbf{X})^{-1}\mathbf{Z}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1 = (\mathbf{X}'\mathbf{Z}_1)^{-1}\mathbf{X}'\mathbf{y}_1$$

que coincide con la expresión [18.12] para el estimador MCI.

Ejemplo 18.6. En el Ejemplo 18.3 la segunda ecuación está sobreidentificada, por lo que no admite *un* estimador de mínimos cuadrados indirectos. Se puede obtener un estimador similar al de mínimos cuadrados indirectos si se utiliza una de las variables predeterminadas omitidas, x_{1t} , como variable instrumental:

$$\begin{pmatrix} \hat{y}_{12} \\ \hat{\beta}_{32} \end{pmatrix} = \begin{pmatrix} x'_1 y_1 & x'_1 x_3 \\ x'_3 y_1 & x'_3 x_3 \end{pmatrix}^{-1} \begin{pmatrix} x'_1 y_2 \\ x'_3 y_2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4/5 \\ -1/5 \end{pmatrix}$$

con matriz de covarianzas:

$$\sigma_2^2 \begin{pmatrix} x'_1 y_1 & x'_1 x_3 \\ x'_3 y_1 & x'_3 x_3 \end{pmatrix}^{-1} \begin{pmatrix} x'_1 x_1 & x'_1 x_3 \\ x'_3 x_1 & x'_3 x_3 \end{pmatrix} \begin{pmatrix} x'_1 y_1 & y'_1 x_3 \\ x'_3 x_1 & x'_3 x_3 \end{pmatrix}^{-1} = \sigma_2^2 \frac{1}{25} \begin{pmatrix} 24 & -16 \\ -16 & 19 \end{pmatrix}$$

De modo similar, podría utilizarse la otra variable predeterminada omitida, x_{2t} , como variable instrumental, para obtener:

$$\begin{pmatrix} \hat{y}_{12} \\ \hat{\beta}_{32} \end{pmatrix} = \begin{pmatrix} x'_2 y_1 & x'_2 x_3 \\ x'_3 y_1 & x'_3 x_3 \end{pmatrix}^{-1} \begin{pmatrix} x'_2 y_2 \\ x'_3 y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} -11/2 \\ 4 \end{pmatrix}$$

con matriz de covarianzas:

$$\sigma_2^2 \begin{pmatrix} x'_2 y_1 & x'_2 x_3 \\ x'_3 y_1 & x'_3 x_3 \end{pmatrix}^{-1} \begin{pmatrix} x'_2 x_2 & x'_2 x_3 \\ x'_3 x_2 & x'_3 x_3 \end{pmatrix} \begin{pmatrix} x'_2 y_1 & y'_1 x_3 \\ x'_3 x_2 & x'_3 x_3 \end{pmatrix}^{-1} = \sigma_2^2 \begin{pmatrix} 3/2 & -1 \\ -1 & -1 \end{pmatrix}$$

Cualquiera de estos dos estimadores sería aceptable en el sentido de que, aun siendo sesgados en muestras finitas, su sesgo desaparece al tender el tamaño muestral a infinito. Ambos son estimadores de variables instrumentales y, como tales, son consistentes. Sin embargo, ninguno de ellos es eficiente, pues ignora la información acerca de la existencia de otro instrumento disponible. También puede comprobarse que el primero es más eficiente que el segundo, pues la diferencia de sus matrices de covarianzas es definida

negativa. (El coeficiente de correlación entre x_1 e y_2 es superior al de x_3 e y_2 .)

La suma $x_{1t} + x_{2t}$ es, por supuesto, otro instrumento válido (al igual que cualquier combinación lineal $ax_{1t} + bx_{2t}$) y genera como estimador:

$$\begin{pmatrix} (x_1 + x_2)'y_1 & (x_1 + x_2)'x_3 \\ x_3'y_1 & x_3'x_3 \end{pmatrix}^{-1} \begin{pmatrix} (x_1 + x_2)'y_2 \\ x_3'y_2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

con matriz de covarianzas:

$$\sigma_2^2 \begin{pmatrix} 1 & -1 \\ -2/3 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} 1 & -2/3 \\ -1 & 1 \end{pmatrix} = \sigma_2^2 \begin{pmatrix} 2 & -4/3 \\ -4/3 & 11/9 \end{pmatrix}$$

De acuerdo con la Proposición 18.11, en ecuaciones sobreidentificadas el estimador de mínimos cuadrados en dos etapas es eficiente⁽⁴⁾:

$$\begin{aligned} \begin{pmatrix} \hat{\gamma}_{12} \\ \hat{\beta}_{32} \end{pmatrix} &= \begin{pmatrix} y_1'X(X'X)^{-1}X'y_1 & y_1'x_3 \\ x_3'y_1 & x_3'x_3 \end{pmatrix}^{-1} \begin{pmatrix} y_1'X(X'X)^{-1}X'y_2 \\ x_3'y_2 \end{pmatrix} = \\ &= \begin{pmatrix} 5/2 & 2 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} -1/2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \end{aligned}$$

que, como puede verse, es una combinación lineal, con coeficientes $\frac{5}{7}$ y $\frac{2}{7}$, de las estimaciones que obtuvimos utilizando, alternativamente, x_{1t} y x_{2t} como variables instrumentales.

La matriz de covarianzas del estimador MC2E es:

$$\text{Var}(\hat{\delta}_2^{\text{MC2E}}) = \sigma_2^2 \begin{pmatrix} 5/2 & 2 \\ 2 & 3 \end{pmatrix}^{-1} = \sigma_2^2 \begin{pmatrix} 6/7 & -4/7 \\ -4/7 & 5/7 \end{pmatrix}$$

cuya diferencia con respecto a las dos matrices de covarianzas previas es semidefinida negativa, como puede comprobar el lector.

Si utilizamos este método de estimación en la primera ecuación, se tiene:

$$\begin{pmatrix} \hat{\gamma}_{21} \\ \hat{\beta}_{11} \\ \hat{\beta}_{21} \end{pmatrix} = \begin{pmatrix} y_2'X(X'X)^{-1}X'y_2 & y_2'x_1 & y_2'x_2 \\ x_1'y_2 & x_1'x_1 & x_1'x_2 \\ x_2'y_2 & x_2'x_1 & x_2'x_2 \end{pmatrix}^{-1} \begin{pmatrix} y_2'X(X'X)^{-1}X'y_1 \\ x_1'y_1 \\ x_2'y_1 \end{pmatrix} = \begin{pmatrix} -1/8 \\ 13/16 \\ 1/2 \end{pmatrix}$$

que coinciden, como afirma la Proposición 18.12, con las estimaciones de mínimos cuadrados indirectos, ya que esta ecuación está exactamente identificada.

⁽⁴⁾ El lector debe prestar especial atención al hecho de que la notación vectorial Y_1 , Y_2 es genérica y denota los subvectores de variables endógenas incluidas y excluidas de la ecuación que se estima. Dichos vectores difieren, por tanto, de una ecuación a otra. Por el contrario, la notación y_1 , y_2 denota siempre las variables y_{1t} , y_{2t} y no otras.

El estimador MC2E admite una interpretación alternativa que nos resultará de interés en lo sucesivo. Consideremos la estimación de la ecuación:

$$y_{1t} = \mathbf{Z}_{1t} \boldsymbol{\delta}_1 + u_{1t}$$

donde $\mathbf{Z}_{1t} = (\mathbf{Y}_{1t}, \mathbf{X}_{1t})$ es un vector de dimensión $g_1 + k_1 - 1$, $\boldsymbol{\delta}_1 = (\gamma_1, \boldsymbol{\beta}_1)$, y $E(u_{1t}^2) = \sigma_1^2$, $E(u_{1t}u_{1s}) = 0$ para todo $t \neq s$. Si premultiplicamos dicha ecuación por la matriz \mathbf{X}' ($k \times T$) de observaciones de todas las variables predefinidas del sistema, se tiene:

$$\mathbf{X}'\mathbf{y}_1 = \mathbf{X}'\mathbf{Z}_1\boldsymbol{\delta}_1 + \mathbf{X}'\mathbf{u}_1 \quad [18.14]$$

donde \mathbf{y}_1 y \mathbf{u}_1 son vectores de dimensión T y \mathbf{Z}_1 es la matriz $T \times (k_1 + g_1 - 1)$ de observaciones de las variables explicativas. La matriz de covarianzas del nuevo término de error es $E(\mathbf{X}'\mathbf{u}_1\mathbf{u}_1'\mathbf{X}) = \sigma_1^2(\mathbf{X}'\mathbf{X})$, por lo que procede estimar la ecuación [18.14] por mínimos cuadrados generalizados. Nótese que, de este modo, tenemos en cuenta que la matriz de covarianzas del término de error no es diagonal, pero no tratamos el hecho de que hay variables endógenas incluidas como explicativas en la ecuación. El estimador que resulta es:

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1)^{-1}(\mathbf{Z}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1)$$

que es, precisamente, el estimador MC2E de dicha ecuación.

Contraste de restricciones de sobreidentificación. Supongamos que la primera ecuación de un modelo de ecuaciones simultáneas está sobreidentificada, y consideremos el estadístico:

$$\frac{T - k}{k - (k_1 + g_1 - 1)} \cdot \frac{(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_1)'(\mathbf{M}_1 - \mathbf{M})(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_1)}{(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_1)'\mathbf{M}(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_1)}$$

donde $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, y $\hat{\boldsymbol{\gamma}}_1$ es el estimador MC2E del vector $\boldsymbol{\gamma}_1$. Basman (1960), que introdujo este estadístico, sugirió que tiene aproximadamente una distribución F , con grados de libertad igual a los dos términos de la primera fracción.

18.5. EL ESTIMADOR DE MAXIMA VEROSIMILITUD CON INFORMACION LIMITADA

Este estimador se obtiene maximizando la función de verosimilitud de la ecuación que se pretende estimar, supuesta la distribución Normal de su término de error, y sujeta a las restricciones que existen sobre sus coeficientes⁽⁵⁾. Sin embargo, puede probarse (véase Schmidt, 1986) que el método no

⁽⁵⁾ Dichas restricciones son $\boldsymbol{\Pi}_3\boldsymbol{\gamma}_1 = \mathbf{0}_{k-k_1}$.

precisa de la maximización analítica de dicha verosimilitud, y es equivalente al procedimiento que ilustramos en el siguiente ejemplo.

Consideremos la estimación de la ecuación:

$$y_{1t} = \gamma_{21}y_{2t} + \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t} \quad [18.15]$$

Tras hacer una partición del rango de valores admisibles del parámetro γ_{21} , se construye para cada uno de ellos la variable $y_{1t}^* = y_{1t} - \gamma_{21}y_{2t}$, y se estima, mediante mínimos cuadrados ordinarios, la regresión:

$$y_{1t}^* = \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t} \quad [18.16]$$

y se conserva el valor de la suma residual, SR_1 .

A continuación, se estima asimismo la regresión de la variable y_{1t}^* sobre todas las variables predeterminadas del modelo. Denotemos por SR_2 su suma residual. Si las restricciones incorporadas en la ecuación [18.15] son correctas, las sumas residuales de estas dos regresiones no debieran diferir en mucho. Finalmente, se escoge el valor del parámetro γ_{21} para el que el cociente $\frac{SR_1 - SR_2}{SR_1}$ es mínimo, o equivalentemente, el valor del parámetro γ_{21} para

el que el cociente $\frac{SR_1}{SR_2}$ sea mínimo. Las estimaciones de los coeficientes β son los que proceden de la estimación de [18.16] cuando se utiliza el valor de γ_{21} escogido de acuerdo con el criterio anterior.

Por comparación con este estimador, puede probarse que el estimador de mínimos cuadrados en dos etapas es aquel para el que la diferencia $SR_1 - SR_2$ es mínima. Si la ecuación incluye varias variables endógenas como explicativas, se particionan los rangos de todos sus coeficientes y se construye una variable y_{1t}^* similar a la anterior, dejando en el miembro derecho de la ecuación únicamente a las variables predeterminadas. Como se ve, el número de regresiones a estimar crece enormemente con el número de variables endógenas que aparecen como explicativas, aparte de que se requiere bastante cuidado al construir la partición para no perder el punto que minimizaría el criterio antes citado.

El estimador de máxima verosimilitud con información limitada (MVIL) coincide con el estimador MC2E (y, por tanto, con el estimador MCI), cuando la ecuación que se estima está exactamente identificada (véase Schmidt, 1976). Es un estimador consistente, y su distribución asintótica coincide con la del estimador MC2E, por lo que comparte la propiedad de eficiencia de este último.

18.6. ESTIMACION POR MINIMOS CUADRADOS EN TRES ETAPAS

Los métodos examinados en las secciones previas han tratado el problema de estimación de una ecuación de un modelo de ecuaciones simultáneas. Sin

embargo, en tanto en cuanto existan correlaciones entre los términos de error de las distintas ecuaciones del modelo, se puede ganar eficiencia en la estimación de cada una de ellas si se estiman simultáneamente. De ello trata esta sección, en la que vamos a analizar uno de estos métodos de estimación simultánea, el llamado de *mínimos cuadrados en tres etapas* (MC3E), válido para un bloque de ecuaciones exactamente identificadas o sobreidentificadas.

Este método generaliza el método de MC2E en el sentido de tomar en consideración las correlaciones entre los términos de error de ecuaciones diferentes. El método de MC2E, siendo uniecuacional, no entraba en tales consideraciones. El procedimiento por el que se van a incorporar dichas correlaciones es tratando la estimación simultánea del modelo como un problema de regresiones aparentemente no relacionadas, como los estudiados en el Capítulo 8. La única diferencia con respecto a lo que allí se discutió es que en esta ocasión algunas de las variables explicativas de las distintas ecuaciones pueden ser endógenas.

Estas consideraciones son relevantes en el tratamiento del modelo [18.14]. Aunque dicho modelo corresponde a una sola de las ecuaciones del sistema, podría llevarse a cabo una transformación similar con cada una de las ecuaciones, es decir:

$$\mathbf{X}'\mathbf{y}_i = \mathbf{X}'\mathbf{Z}_i\boldsymbol{\delta}_i + \mathbf{X}'\mathbf{u}_i$$

De este modo, se obtiene un conjunto de vectores correspondientes a los términos de error transformados: $\mathbf{X}'\mathbf{u}_1, \mathbf{X}'\mathbf{u}_2, \dots, \mathbf{X}'\mathbf{u}_g$, todos ellos de dimensión $k \times 1$. La matriz de covarianzas de estos vectores no es diagonal a bloques. Por el contrario:

$$\text{Cov}(\mathbf{X}'\mathbf{u}_i, \mathbf{X}'\mathbf{u}_j) = \mathbf{X}' \text{Cov}(\mathbf{u}_i, \mathbf{u}_j) \mathbf{X} = \sigma_{ij}(\mathbf{X}'\mathbf{X})$$

Como, por otra parte:

$$\text{Var}(\mathbf{X}'\mathbf{u}_i) = \mathbf{X}' \text{Var}(\mathbf{u}_i) \mathbf{X} = \sigma_i^2 \mathbf{X}'\mathbf{X}$$

entonces, si denotamos por $\boldsymbol{\varepsilon}$ el vector $kg \times 1$ de los términos de error de todo el sistema transformado, $\boldsymbol{\varepsilon}' = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_g)$, donde $\boldsymbol{\varepsilon}_i = \mathbf{X}'\mathbf{u}_i$, se tiene la matriz $kg \times kg$:

$$\text{Var}(\boldsymbol{\varepsilon}) = \text{Cov} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \dots \\ \boldsymbol{\varepsilon}_g \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \mathbf{X}'\mathbf{X} & \sigma_{12} \mathbf{X}'\mathbf{X} & \sigma_{13} \mathbf{X}'\mathbf{X} & \dots & \sigma_{1g} \mathbf{X}'\mathbf{X} \\ \sigma_{12} \mathbf{X}'\mathbf{X} & \sigma_2^2 \mathbf{X}'\mathbf{X} & \sigma_{23} \mathbf{X}'\mathbf{X} & \dots & \sigma_{2g} \mathbf{X}'\mathbf{X} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{1g} \mathbf{X}'\mathbf{X} & \sigma_{2g} \mathbf{X}'\mathbf{X} & \sigma_{3g} \mathbf{X}'\mathbf{X} & \dots & \sigma_g^2 \mathbf{X}'\mathbf{X} \end{pmatrix}$$

que, como se ve, no tiene estructura diagonal. Esta matriz puede escribirse:

$$\text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X}) = (\mathbf{I}_g \otimes \mathbf{X}') (\boldsymbol{\Sigma} \otimes \mathbf{I}_g) (\mathbf{I}_g \otimes \mathbf{X})$$

por las propiedades del producto de Kronecker vistas en el Capítulo 1. Si denotamos por \mathbf{X}^* la matriz diagonal a bloques:

$$\mathbf{X}^* = \mathbf{I}_g \otimes \mathbf{X}$$

de dimensión $Tg \times kg$, por \mathbf{Z} a la matriz $Tg \times \sum_1^g (g_i + k_i - 1)$:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & & & \\ & \mathbf{Z}_2 & & \\ & & \dots & \\ & & & \mathbf{Z}_g \end{pmatrix}$$

y por δ al vector de dimensión $\sum_1^g (g_i + k_i - 1)$ formado por los coeficientes de todas las ecuaciones se tiene:

$$\mathbf{X}^{*\prime} \mathbf{y} = \mathbf{X}^{*\prime} \mathbf{Z} \delta + \mathbf{X}^{*\prime} \mathbf{u}$$

donde $\mathbf{y}' = (y'_1, y'_2, \dots, y'_g)$; $\delta' = (\delta'_1, \delta'_2, \dots, \delta'_g)$ y, a la vista de la estructura de la matriz de covarianzas que acabamos de obtener, el estimador MCG viene dado por:

$$\begin{aligned} \hat{\delta}_{\text{MCG}} &= \{ \mathbf{Z}' \mathbf{X}^* [\boldsymbol{\Sigma} \otimes (\mathbf{X}' \mathbf{X})]^{-1} \mathbf{X}^{*\prime} \mathbf{Z} \}^{-1} \{ \mathbf{Z}' \mathbf{X}^* [\boldsymbol{\Sigma} \otimes (\mathbf{X}' \mathbf{X})]^{-1} \mathbf{X}^{*\prime} \mathbf{y} \} = \\ &= \{ \mathbf{Z}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \mathbf{Z} \} \{ \mathbf{Z}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \mathbf{y} \} = \hat{\delta}_{\text{MC3E}} \end{aligned} \quad [18.17]$$

donde se ha utilizado la representación $\mathbf{X}^* = \mathbf{I}_g \otimes \mathbf{X}$, junto con las propiedades del producto de Kronecker. Este es el *estimador de mínimos cuadrados en tres etapas* del modelo de ecuaciones simultáneas.

Comparando esta expresión para el estimador MC3E con la del estimador de variables instrumentales puede verse que el estimador de mínimos cuadrados en tres etapas es un estimador de variables instrumentales simultáneamente para todo el sistema, cuando los instrumentos elegidos vienen dados por el vector:

$$\mathbf{Z}^* = [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{Z}$$

Proposición 18.13. Si las ecuaciones de un sistema están identificadas, entonces el estimador MC3E es consistente.

Demostración. La consistencia se deriva de ser un estimador de variables instrumentales.

Proposición 18.14. La distribución asintótica de $\sqrt{T}(\hat{\delta}_{MC3E} - \delta)$ es Normal multivariante con esperanza $\mathbf{0}$, [donde $r = \sum_1^g (k_i + g_i - 1)$] y matriz de covarianzas⁽⁶⁾:

$$plim[\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}]^{-1} \quad [18.18]$$

Demostración. (Véase Schmidt, 1976.)

Por otra parte, utilizando de nuevo las propiedades del producto de Kronecker, el estimador MC2E de todo el sistema puede expresarse:

$$\hat{\delta}^{MC2E} = [\mathbf{Z}'\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Z}]^{-1}[\mathbf{Z}'\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}] \quad [18.19]$$

y como $\mathbf{X}^{*\prime}\mathbf{X}^* = \mathbf{I} \otimes \mathbf{X}'\mathbf{X}$, entonces:

$$\hat{\delta}^{MC2E} = \{\mathbf{Z}'[\mathbf{I}_g \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}\}^{-1}\mathbf{Z}'[\mathbf{I}_g \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \quad [18.20]$$

y comparando las expresiones [18.17] y [18.20] para $\hat{\delta}^{MC3E}$ y $\hat{\delta}^{MC2E}$, puede apreciarse claramente que la diferencia entre ambos estriba únicamente en la consideración, por parte del primero, de las correlaciones entre los términos de error de las distintas ecuaciones del modelo, que aparecen resumidas en la matriz de covarianzas $\boldsymbol{\Sigma}$ y que en [18.20] es sustituida por la identidad \mathbf{I}_g .

Precisamente en este punto surge una dificultad práctica ya familiar: La matriz $\boldsymbol{\Sigma}$, que juega un papel fundamental en la obtención del estimador MC3E, es desconocida. Por tanto, es preciso estimarla. Para ello, se obtienen primero las estimaciones MC2E de cada una de las ecuaciones del modelo por separado. A partir de ellas, se obtienen los vectores de residuos de cada ecuación. Dichos vectores se utilizan para estimar los $\frac{g(g+1)}{2}$ componentes de la matriz $\boldsymbol{\Sigma}$. Por ejemplo:

$$\hat{\sigma}_{ij} = \frac{\hat{u}_i' \hat{u}_j}{\sqrt{(T - g_i - k_i + 1)(T - g_j - k_j + 1)}}$$

$$\hat{\sigma}_i^2 = \frac{\hat{u}_i' \hat{u}_i}{T - g_i - k_i + 1}$$

Por último, esta matriz estimada se utiliza en [18.17]. De este modo, dos de las tres etapas de mínimos cuadrados a las que se refiere el nombre de este estimador se destinan a la obtención de los residuos necesarios para estimar $\boldsymbol{\Sigma}$. La tercera etapa es la estimación MCG de todas las ecuaciones del sistema simultáneamente.

⁽⁶⁾ La identificación de todas las ecuaciones del sistema es necesaria para que la matriz $plim[\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}]$ sea no singular (véase Schmidt, 1976). Ello impide la presencia de identidades, por lo que éstas deben utilizarse previamente para eliminar alguna de las variables del modelo, antes de utilizar MC3E.

Puesto que utiliza más información que el estimador MC2E, debe esperarse que el estimador MC3E sea más eficiente que el primero. En efecto, como puede verse en Theil (1970), se tiene el siguiente resultado:

Proposición 18.15. Para todo $i = 1, 2, \dots, g$ se tiene que la matriz de covarianzas de la distribución límite de $\sqrt{T}(\hat{\delta}_i^{MC2E} - \delta)$ excede a la matriz de covarianzas de la distribución límite de $\sqrt{T}(\hat{\delta}_i^{MC3E} - \delta)$ en una matriz semidefinida positiva, donde $\hat{\delta}_i^{MC3E}$ denota el subvector del estimador MC3E correspondiente a la ecuación i -ésima.

Hay situaciones en que el estimador MC3E no presenta ninguna ganancia en términos de eficiencia con respecto al estimador MC2E:

Proposición 18.16. Si $\sigma_{ij} = 0$ para todo $i \neq j$, entonces el estimador MC3E y el estimador MC2E coinciden.

Demostración. En estas condiciones $\Sigma = I_g$, por lo que es inmediato comprobar de las expresiones [18.17] y [18.20] la coincidencia de ambos estimadores.

Esta proposición no debe resultar en modo alguno sorprendente. Como hemos visto, el estimador MC3E no es sino una aplicación de la estimación MCG del modelo de regresiones aparentemente no relacionadas. Ahora bien, ya vimos en el Capítulo 8 que cuando los términos de error de las ecuaciones son independientes entre sí, entonces no se obtiene ningún beneficio de su estimación simultánea.

Debe notarse, sin embargo, que la utilización de ambos procedimientos de estimación generarían estimaciones numéricas diferentes, pues aunque Σ sea una matriz diagonal, su estimación $\hat{\Sigma}$, basada en los residuos MC2E ecuación, no lo será. Si se estimasen únicamente los elementos de la diagonal de Σ igualando los restantes a cero, entonces las estimaciones serían numéricamente iguales.

Otra situación que tiene las mismas consecuencias es la siguiente:

Proposición 18.17. Si todas las ecuaciones del modelo están exactamente identificadas, entonces los estimadores MC2E y MC3E coinciden.

Demostración. En tal caso se tiene, por la condición de orden que $g_i + k_i - 1 = k$, lo que hace que cada una de las matrices Z_i , que habitualmente tiene dimensión $T \times (g_i + k_i - 1)$, sea ahora de dimensión $T \times k$. En consecuencia, el producto $X'Z_i$ es $k \times k$ y, salvo singularidades, la inversa del producto que aparece en [18.17] puede obtenerse mediante la expresión habitual para la inversa de un producto de matrices, todas ellas invertibles:

$$\begin{aligned} \hat{\delta}^{MC3E} &= (X^*Z)^{-1} [\Sigma \otimes (X'X)] (Z'X^*)^{-1} (Z'X^*) [\Sigma \otimes (X'X)]^{-1} X^{*'}y = \\ &= (X^*Z)^{-1} X^{*'}y \end{aligned}$$

expresión idéntica a la del estimador MC2E de todo el sistema.

Corolario. Si todas las ecuaciones del sistema están exactamente identificadas, entonces los estimadores MC2E, MC3E, MVIL y MCI coinciden.

A efectos prácticos, si en un modelo de ecuaciones simultáneas hay un bloque de ecuaciones que están exactamente identificadas y todas las variables predeterminadas de éstas aparecen en las ecuaciones sobreidentificadas, las ecuaciones exactamente identificadas pueden omitirse cuando se estima por mínimos cuadrados en tres etapas el bloque de ecuaciones que estén sobreidentificadas. Las estimaciones de los coeficientes de estas últimas no variarán, con independencia de que se estimen conjuntamente con el bloque de ecuaciones exactamente identificadas.

El resultado dual no es, sin embargo, cierto. Las estimaciones del bloque de ecuaciones exactamente identificadas variarán, en general, según que se estimen conjuntamente con ellas las ecuaciones sobreidentificadas del modelo.

18.7. EL METODO DE MAXIMA VEROSIMILITUD CON INFORMACION COMPLETA

Este método maximiza la función de verosimilitud que se obtiene al suponer que el vector formado por los términos de error de todas las ecuaciones del modelo sigue una determinada distribución multivariante. A diferencia de los estimadores vistos hasta ahora, este estimador no admite una representación analítica como función de las observaciones muestrales. Por tanto, su obtención práctica no surge de la resolución de un sistema de ecuaciones, sino que, por el contrario, es preciso obtenerlo mediante la utilización de algoritmos numéricos de optimización.

Supongamos que el vector \mathbf{u} formado por los g términos de error del modelo sigue una distribución $N_{Tg}(\mathbf{0}_{Tg}, \Sigma \otimes \mathbf{I}_T)$. La función de probabilidad conjunta del vector \mathbf{u} es, por tanto:

$$\ln L(\mathbf{u}) = -\frac{Tg}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma \otimes \mathbf{I}_T| - \frac{1}{2} (\mathbf{y} - \mathbf{Z}\delta)' (\Sigma^{-1} \otimes \mathbf{I}_T) (\mathbf{y} - \mathbf{Z}\delta)$$

que puede expresarse en función del vector \mathbf{y} y teniendo en cuenta que:

$$\left| \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \right| = \begin{vmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \dots & \frac{\partial u_1}{\partial y_g} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \dots & \frac{\partial u_2}{\partial y_g} \\ \dots & \dots & \dots & \dots \\ \frac{\partial u_g}{\partial y_1} & \frac{\partial u_g}{\partial y_2} & \dots & \frac{\partial u_g}{\partial y_g} \end{vmatrix} = \begin{vmatrix} \Gamma & & & \\ & \Gamma & & \\ & & \dots & \\ & & & \Gamma \end{vmatrix} = |\Gamma|^T$$

y, en consecuencia:

$$\ln L(\mathbf{y}) = -\frac{Tg}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| + T \ln |\Gamma| - \frac{1}{2} (\mathbf{y} - \mathbf{Z}\delta)' (\Sigma^{-1} \otimes \mathbf{I}_T) (\mathbf{y} - \mathbf{Z}\delta)$$

donde se ha utilizado la propiedad $|\Sigma \otimes \mathbf{I}_T| = |\Sigma|^T$. Por otra parte, hay que observar que:

$$(\mathbf{y} - \mathbf{Z}\delta)' (\Sigma^{-1} \otimes \mathbf{I}_T) (\mathbf{y} - \mathbf{Z}\delta) = \sum_1^g \sum_1^g \sigma^{ij} (y_i - \mathbf{Z}_i \delta_i)' (y_j - \mathbf{Z}_j \delta_j)$$

donde σ^{ij} son los elementos de Σ^{-1} . Por tanto, las condiciones de optimalidad con respecto a cada elemento σ^{ij} son:

$$\frac{\partial \ln L(\mathbf{y})}{\partial \sigma^{ij}} = \frac{T}{2} \sigma_{ij} - \frac{1}{2} (\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_i)' (\mathbf{y}_j - \mathbf{Z}_j \hat{\delta}_j) = 0$$

donde hemos utilizado la propiedad *f*) de la Sección 1.3, que tienen por solución:

$$\hat{\sigma}_{ij} = \frac{(\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_i)' (\mathbf{y}_j - \mathbf{Z}_j \hat{\delta}_j)}{T} \quad ; \quad i, j = 1, 2, \dots, g$$

y, al sustituir en la función de verosimilitud, se tiene:

$$\ln L(\mathbf{y}) = -\frac{Tg}{2} \ln 2\pi - \frac{T}{2} \ln |\hat{\Sigma}| + T \ln |\Gamma| = \frac{T}{2}$$

y derivando esta última expresión con respecto al vector δ para obtener las restantes condiciones necesarias para la maximización de la función de verosimilitud se tiene:

$$\frac{\partial \ln L(\mathbf{y})}{\partial \delta} = -\frac{T}{2} \frac{\partial \ln |\Sigma|}{\partial \delta} + T \frac{\partial \ln |\Gamma|}{\partial \delta} = 0 \quad [18.21]$$

que es un sistema de ecuaciones no lineales que, en general, carece de solución analítica. Por ejemplo, la derivada parcial $\frac{\partial \ln |\Gamma|}{\partial \gamma^{ij}}$ es igual al elemento γ^{ij} de la matriz inversa Γ^{-1} , que depende, en modo nada trivial, de todos los elementos de la matriz Γ .

Puede probarse (véanse Schmidt, 1976, y Theil, 1970) que el estimador de máxima verosimilitud con información completa es consistente, y también que las matrices de covarianzas de las distribuciones límite de $\sqrt{T}(\hat{\delta}^{\text{MVIC}} - \delta)$ y $\sqrt{T}(\hat{\delta}^{\text{MC3E}} - \delta)$ coinciden (Rothenberg y Leenders, 1964). Ello implica que, puesto que el estimador MVIC tiene la propiedad de eficiencia inherente al estimador de máxima verosimilitud, entonces el estimador MC3E es también

asintóticamente eficiente. Así, el interés de este último es que siendo eficiente es mucho más sencillo de utilizar que el estimador MVIC.

Precisamente la dificultad en el cálculo del estimador MVIC condujo a la introducción del estimador de *máxima verosimilitud lineal* (MVL). Si denotamos el sistema de condiciones [18.21] necesarias para la maximización de la verosimilitud por $f(\delta) = 0$, entonces su aproximación lineal sería $f(\hat{\delta}) + F(\hat{\delta})(\delta - \hat{\delta}) = 0$, donde F denota el gradiente de f , sistema que tiene por solución:

$$\hat{\delta}^{\text{MVL}} = \hat{\delta} - [F(\hat{\delta})]^{-1} f(\hat{\delta})$$

donde $\hat{\delta}$ denota un estimador inicial, consistente, del vector δ . El candidato natural para utilizar como $\hat{\delta}$ es el estimador MC2E. Aunque el estimador de máxima verosimilitud lineal es más sencillo de obtener que el estimador MVIC, todavía es más complejo que el estimador MC3E.

Puede probarse (Theil, 1970) que $\sqrt{T}(\hat{\delta}^{\text{MVL}} - \delta)$ tiene la misma distribución asintótica que $\sqrt{T}(\hat{\delta}^{\text{MVIC}} - \delta)$, y también que $-F(\hat{\delta})^{-1}$ es una aproximación a la matriz de covarianzas de $\hat{\delta}^{\text{MVL}}$.

18.8. SISTEMAS RECURSIVOS

Una clase de modelos de ecuaciones simultáneas que no presentan las dificultades que hemos estado analizando en las secciones previas son los sistemas recursivos. Estos son modelos que satisfacen simultáneamente dos condiciones: a) la matriz de coeficientes de las variables endógenas es triangular, y b) la matriz de covarianzas de los términos de error de las distintas ecuaciones del sistema es diagonal:

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1g} \\ 0 & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{2g} \\ 0 & 0 & \gamma_{33} & \cdots & \gamma_{3g} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \gamma_{gg} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \cdots & \\ & & & & \sigma_g^2 \end{pmatrix}$$

Puede verse inmediatamente que las dos condiciones juntas son bastante restrictivas. Por una parte, los términos de error de las distintas ecuaciones son ortogonales entre sí, lo que de entrada elimina el principal aspecto de interés en la estimación conjunta del modelo. De todos modos, aún podría haber correlaciones entre las distintas variables endógenas, debido a que algunas de ellas aparecieran explícitamente en las ecuaciones de las demás.

Sin embargo, esta situación queda eliminada por la primera de las dos condiciones enunciadas. De acuerdo con ella, el sistema de ecuaciones puede ordenarse de modo que: La primera ecuación contenga tan sólo una variable endógena, la segunda ecuación contenga dos variables endógenas, pero una

de ellas sea la de la ecuación anterior; la tercera ecuación contenga sólo tres variables endógenas, pero dos de ellas sean las de las ecuaciones anteriores, y así sucesivamente. De este modo, las variables endógenas son predeterminadas con respecto a las ecuaciones que siguen a la suya propia: y_1 queda determinada tan sólo por las variables predeterminadas junto con u_1 (que es independiente de u_{2t}), y_2 queda determinada por las predeterminadas, junto con y_1 y u_2 . Sin embargo, y_1 es, a todos los efectos, predeterminada con respecto a y_{2t} ; y_3 queda determinada por las predeterminadas, junto con y_1 e y_2 así como u_3 , y así sucesivamente.

Los modelos recursivos están siempre exactamente identificados. El número de parámetros en dichos modelos es igual a $\frac{g(g+1)}{2} + gk$, precisamente igual al de parámetros en la forma reducida de todo modelo de ecuaciones simultáneas. Además, la disposición de tales restricciones es tal que la condición de rango siempre se satisface con igualdad, por lo que cada ecuación está exactamente identificada.

En cuanto a su estimación, dado que Γ es triangular, las condiciones de normalización que hemos adoptado garantizan que $|\Gamma| = 1$. En consecuencia, utilizando [18.21] se tiene que el vector $\hat{\delta}_{MV}$ es la solución del sistema:

$$\frac{\partial \ln |\hat{\Sigma}|}{\partial \hat{\delta}} = \mathbf{0}$$

Pero, como la matriz Σ también es diagonal, entonces:

$$\ln |\hat{\Sigma}| = \sum_{i=1}^g \ln \hat{\sigma}_i^2 = \sum_{i=1}^g \ln (\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_i)' (\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_i) - T_g$$

y, diferenciando con respecto a $\hat{\delta}_i$, se tiene finalmente:

$$\frac{\partial \ln |\hat{\Sigma}|}{\partial \hat{\delta}_i} = \sum_{i=1}^g \left(\frac{2\mathbf{Z}_i' (\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_i)}{T \hat{\sigma}_i^2} \right) = \mathbf{0}$$

que generan las ecuaciones normales del estimador de mínimos cuadrados ordinarios. Por tanto, en un modelo recursivo, el estimador de máxima verosimilitud con información completa coincide con el estimador de mínimos cuadrados ecuación por ecuación. En consecuencia, este último hereda las propiedades del estimador de máxima verosimilitud y, en particular, es eficiente.

El estimador de mínimos cuadrados ordinarios de un modelo recursivo es también insesgado, lo que puede resultar sorprendente a simple vista, dado que aparecen variables endógenas como explicativas. La característica crucial de tal modelo, sin embargo, es que dichas variables son independientes del término de error de la ecuación en la que aparecen como explicativas, aunque están correlacionadas con los términos de error de las ecuaciones previas. A diferencia de otros modelos, la ordenación de las ecuaciones de un modelo recursivo es, claramente, una característica importante del mismo.

18.9. COMPARACION ENTRE LOS DISTINTOS ESTIMADORES

Como hemos dicho repetidamente, las propiedades estadísticas de los estimadores de modelos de ecuaciones simultáneas en muestras finitas son, en su gran mayoría, desconocidas, por lo que su comparación ha de basarse en sus propiedades asintóticas. A este respecto, si existiesen restricciones sobre los coeficientes de las ecuaciones del sistema, dichas ecuaciones deberían estimarse simultáneamente, pues ignorar tal información conduciría a estimaciones ineficientes.

En el caso (poco frecuente) de que existan restricciones sobre los parámetros de la matriz de covarianzas, entonces el resultado de equivalencia asintótica antes citado no es válido, y el estimador MVIC es más eficiente que el estimador MC3E. En este caso, es aconsejable utilizar el estimador MVL.

En cuanto a la comparación entre estimadores simultáneos del sistema frente a estimadores ecuación por ecuación, ya hemos visto que la existencia de correlaciones entre los términos de error permite ganar eficiencia si se estima simultáneamente, incorporando dichas correlaciones. Sin embargo, debe tenerse en cuenta que, puesto que deben estimarse previamente, el error de estimación cometido hará que la ganancia de eficiencia no sea tan grande. También pudiera haberse cometido algún error al especificar una de las ecuaciones del modelo; si se estima simultáneamente, dicho error se propagará a las restantes ecuaciones.

La estimación MCO de la forma reducida genera estimaciones insesgadas de dichos coeficientes. Si el interés fundamental del investigador es la predicción, los valores numéricos de los coeficientes de la forma estructural no son necesarios, y la forma reducida es todo lo que se precisa. Parece natural preguntarse si, en tales condiciones, procedimientos más complicados como MC3E aportan algún beneficio.

La respuesta es que si hay alguna ecuación sobreidentificada, la estimación MCO de la forma reducida no incorpora las restricciones entre los coeficientes, a diferencia del estimador MC3E, lo que hace a éste más eficiente que el primero. El mismo comentario es válido cuando el investigador tiene interés en recuperar los coeficientes de la forma estructural de ecuaciones exactamente identificadas. Sin embargo, no es *necesariamente cierto* que las estimaciones MCO de la forma reducida sean menos eficientes que las obtenidas a partir de otros estimadores relativamente ineficientes, como MC2E.

PROBLEMAS

Problema 18.1. Calcular las expresiones analíticas para el sesgo del estimador de mínimos cuadrados ordinarios de la curva LM del modelo utilizado como ejemplo en el capítulo. ¿Puede decirse algo acerca de su signo?

Problema 18.2. Obtener las expresiones analíticas de los estimadores a) de mínimos cuadrados ordinarios, b) de mínimos cuadrados indirectos, c) de mínimos cuadrados en dos etapas, de la función de consumo en el modelo:

$$\begin{aligned}C_t &= \alpha + \beta Y_t + u_t \\Y_t &= C_t + I_t + G_t, \quad t = 1, 2, \dots, T \\G_t &= \bar{G}\end{aligned}$$

donde C_t , Y_t , I_t denotan el consumo, la renta, la inversión agregadas, y G_t es el gasto público.

Problema 18.3. Obtener la expresión analítica del estimador de mínimos cuadrados indirectos de la curva IS en el modelo del Ejemplo 18.1, cuando se utiliza una constante como instrumento. Obtener la expresión analítica para el mismo estimador cuando se utiliza la oferta monetaria M_t , como instrumento.

Problema 18.4. Considérese el modelo:

$$\begin{aligned}y_{1t} &= \gamma_{21}y_{2t} + \beta_{11}x_{1t} + \beta_{21}x_{2t} + u_{1t} \\y_{2t} &= \gamma_{12}y_{1t} + u_{2t}\end{aligned}$$

a) Obtener la expresión analítica del estimador de mínimos cuadrados indirectos del parámetro γ_{12} cuando se utiliza la variable x_{1t} como instrumento de y_{2t} .

b) Obtener la expresión analítica de dicho estimador cuando se utiliza la variable x_{2t} como instrumento.

c) Obtener la expresión analítica del estimador de mínimos cuadrados en dos etapas de dicho parámetro.

d) Justificar en qué sentido puede afirmarse que el estimador de mínimos cuadrados en dos etapas es una combinación lineal de los estimadores de mínimos cuadrados indirectos.

Problema 18.5. La ecuación que representa la curva IS en el Ejemplo 18.1 está exactamente identificada. Probar que la estimación de mínimos cuadrados indirectos y de mínimos cuadrados en dos etapas de los coeficientes de dicha ecuación coinciden, comparando sus expresiones analíticas.

Problema 18.6. Probar que si en la expresión matricial del estimador de mínimos cuadrados en tres etapas se supone que la matriz Σ es diagonal, se tiene el estimador de mínimos cuadrados en dos etapas ecuación por ecuación.

Problema 18.7. Demostrar que la matriz $A'[\mathbf{I}_{k-1} - C(C'C)^{-1}C]A$ es definida positiva.

Problema 18.8. Considérese el modelo de equilibrio de mercado:

$$\begin{aligned}q_t^d &= \beta p_t + u_t, \quad \beta < 0 \\q_t^o &= \alpha p_t + v_t, \quad \alpha > 0 \\q_t^o &= q_t^d\end{aligned}$$

Suponga que la covarianza $\sigma_{uv} = 0$ y encuentre condiciones bajo las cuales la estimación MCO de β es consistente. Hacer el mismo análisis para α e interpretar

ambas condiciones. Probar que si se estima una regresión de q_t sobre p_t , el parámetro estimado es una cota inferior de α y una cota superior de β .

Problema 18.9. Si se cambia la ecuación de demanda en el ejercicio anterior por

$$q_t^d = \beta p_t + \gamma x_t + u_t$$

con $\text{plím} \left(\frac{\sum_1^T x_t u_t}{T} \right) = \text{plím} \left(\frac{\sum_1^T v_t x_t}{T} \right) = 0$, y mantenemos el supuesto $\text{Cov}(u_t, v_t) = 0$, probar que el sesgo asintótico del estimador MCO del parámetro α es inferior al de antes.

Nota: La ecuación de demanda está ahora exactamente identificada, y puede estimarse consistentemente por MC2E.

Problema 18.10. Probar que en el modelo

$$\begin{aligned} q_t^d &= \beta p_t + \gamma x_t + u_t, & \beta < 0 \\ q_t^o &= \alpha p_t + \delta z_t + v_t, & \alpha > 0 \end{aligned}$$

el sesgo asintótico del estimador MCO del parámetro β es igual a:

$$(\beta - \alpha) \left[\frac{-\sigma_u^2}{\sigma_u^2 + \sigma_v^2 + \delta^2 \sigma_z^2 (1 - r_{zz})} \right]$$

y también que existe una expresión similar para el sesgo asintótico del estimador MCO de α . Mostrar condiciones bajo las cuales el sesgo asintótico es pequeño en ambos casos. En particular, justificar bajo qué condiciones la inclusión de las variables explicativas x_t y z_t contribuirán a disminuir el sesgo asintótico de los estimadores MCO frente a los obtenidos en el modelo del Problema 18.8.

Problema 18.11. Mostrar que en el modelo

$$\begin{aligned} C_t &= \alpha Y_t + \beta + u_t \\ Y_t &= C_t + I_t \end{aligned}$$

donde C_t , I_t , Y_t denotan el consumo, inversión y renta agregadas, y se supone que la inversión (variable aleatoria) satisface $\text{plím} \left(\frac{\sum_1^T I_t u_t}{T} \right) = 0$, el estimador MCO sobrestima la propensión marginal al consumo. Mostrar asimismo que el sesgo asintótico es menor cuanto mayor es el cociente $\frac{\sigma_I^2}{\sigma_u^2}$, e interpretar dicha condición.

Problema 18.12. Dado el modelo

$$\begin{aligned} y_{1t} &= a_{21} y_{2t} + b_{11} x_{1t} + b_{21} x_{2t} + u_{1t} \\ y_{2t} &= a_{12} y_{1t} + b_{32} x_{3t} + u_{2t} \end{aligned}$$

y la matriz de momentos muestrales de segundo orden:

	y_1	y_2	x_1	x_2	x_3
y_1	3,5	3	1	1	0
y_2		11,5	1	3	4
x_1			1	0	0
x_2				1	1
x_3					2

- a) Calcule el estimador MCO de los parámetros de la forma reducida sin restricciones.
- b) Sólo una de las dos ecuaciones del modelo puede estimarse por mínimos cuadrados indirectos. Discutir cuál de ellas y proceda a su estimación. ¿Es insesgado este estimador? ¿Es consistente?
- c) Estimar la otra ecuación por mínimos cuadrados en dos etapas. ¿Es insesgado este estimador? ¿Es consistente?
- d) Obtenga un estimador consistente de $\sigma_{12} = E(u_{1t}u_{2t})$.
- e) Encuentre las estimaciones de los parámetros de la forma reducida que se derivan de las estimaciones obtenidas en b) y c). ¿Coinciden estas estimaciones con las obtenidas en a)?

Problema 18.13. Dado el modelo de ecuaciones simultáneas:

$$y_{1t} + \gamma_{21}y_{2t} + \beta_{11}x_{1t} = u_{1t}$$

$$\gamma_{12}y_{1t} + y_{2t} + \beta_{22}x_{2t} + \beta_{32}x_{3t} = u_{2t}$$

y disponiendo de la matriz de momentos muestrales:

	y_1	y_2	x_1	x_2	x_3
y_1	100	200	30	20	40
y_2	200	900	0	50	160
x_1	30	0	100	0	0
x_2	20	50	0	50	0
x_3	40	160	0	0	40

- a) Estimar la segunda ecuación por MCI y por variables instrumentales, comprobando que ambos estimadores son iguales. Estimar sus respectivas matrices de covarianzas.
- b) Estimar la primera ecuación por MCO y por MC2E.

Problema 18.14. Dado el modelo

$$C_t = \alpha + \beta Y_t + u_t$$

$$Y_t = C_t + I_t$$

donde C_t , I_t e Y_t denotan el consumo, la inversión y la renta, y donde se supone que la inversión es una variable exógena, comparar las varianzas de los estimadores MCO y MCI de la propensión marginal al consumo.

Problema 18.15. Dado el modelo de ecuaciones simultáneas:

$$\begin{aligned} y_{1t} &= \gamma_{21}y_{2t} + \beta_{11}x_{1t} + u_{1t} \\ y_{2t} &= \gamma_{12}y_{1t} + \beta_{22}x_{2t} + \beta_{32}x_{3t} + u_{2t} \end{aligned}$$

y las matrices producto:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 5 & 10 \\ 40 & 20 \\ 20 & 30 \end{pmatrix}$$

Estimar la forma reducida del modelo a) por MCO, b) a partir de estimaciones MC2E o MCI de cada una de las ecuaciones. ¿Coinciden ambas estimaciones? ¿A qué se debe el resultado? ¿Cambia el procedimiento de estimación utilizado si γ_{21} fuese igual a cero y, a la vez, los términos de error u_{1t} y u_{2t} fuesen independientes?

Problema 18.16. Considere el modelo *recursivo* de ecuaciones simultáneas:

$$\begin{aligned} y_{1t} & & & + \beta_{11}x_{1t} & = & u_{1t} \\ \gamma_{12}y_{1t} + & y_{2t} & & + \beta_{12}x_{1t} & = & u_{2t} \\ \gamma_{13}y_{1t} + \gamma_{23}y_{2t} + & y_{3t} & & + \beta_{13}x_{1t} & = & u_{3t} \end{aligned}$$

donde $\text{Cov}(u_{it}, u_{jt}) = 0$ para todo $i \neq j$, $i, j = 1, 2, 3$, y demuestre que las tres ecuaciones están exactamente identificadas.

[*Sugerencia:* Muestre que la única transformación lineal de las ecuaciones que satisface las restricciones del sistema anterior es la que utiliza como matriz de transformación la identidad.]

Problema 18.17. Demostrar que para que exista el estimador MC2E es necesario que se satisfaga la condición de orden.

[*Sugerencia:* Seguir el razonamiento: Para que $\hat{\delta}_1^{\text{MC2E}}$ exista, es preciso que el rango de la matriz

$$\begin{pmatrix} \mathbf{Y}'_1\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1 & \mathbf{Y}'_1\mathbf{X}_1 \\ \mathbf{X}'_1\mathbf{Y}_1 & \mathbf{X}'_1\mathbf{X}_1 \end{pmatrix}$$

sea igual a $g_1 + k_1 - 1$. Esta matriz puede escribirse como $\mathbf{A}'\mathbf{A}$, donde

$$\mathbf{A} = [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1; \mathbf{X}_1]$$

y el rango de ésta es inferior al rango de \mathbf{X} , que suponemos igual a k . Por tanto, para la existencia del estimador MC2E es preciso que $g_1 + k_1 - 1 \leq k$, que es la condición de orden.]

Problema 18.18. La ecuación $y_{1t} = \gamma_{21}y_{2t} + \gamma_{31}y_{3t} + \beta_{11}x_{1t} + u_{1t}$ forma parte de un modelo que incluye otras tres variables predeterminadas. Si se dispone de los momentos muestrales

$$Y'Y = \begin{pmatrix} 80 & 60 & -20 \\ 60 & 240 & -180 \\ -20 & -180 & 280 \end{pmatrix}; \quad Y'X = \begin{pmatrix} 8 & 8 & 16 & 20 \\ 0 & 16 & 48 & -20 \\ 0 & -8 & -48 & 40 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 20 \end{pmatrix}$$

obtenidos a partir de treinta observaciones muestrales, estimar dicha ecuación por mínimos cuadrados en dos etapas y obtener las varianzas de sus coeficientes.

Problema 18.19. Demostrar que, como se afirma en la Sección 18.4,

$$X(X'X)^{-1}X'X_1 = X_1$$

Problema 18.20. Demostrar que si la ecuación $y_{1t} = \gamma_{21}y_{2t} + \beta_{11}x_{1t} + u_{1t}$ está exactamente identificada, es decir, se dispone de una sola variable exógena adicional x_{2t} , el estimador MCI puede obtenerse como el estimador que utiliza x_{1t} como variable instrumental de y_{2t} , y x_{2t} como variable intrumental de x_{1t} .

Problema 18.21. a) Analizar la situación de identificación de las ecuaciones del modelo:

$$C_t = \alpha_1 Y_t + \alpha_2 r_t + u_{1t} \text{ (función de consumo)}$$

$$I_t = \beta r_t + u_{2t} \text{ (función de inversión)}$$

$$r_t = \delta_1 m_t + \delta_2 Y_t + \delta_3 D_t + u_{3t} \text{ (determinación de tipos de interés)}$$

$$Y_t = C_t + I_t$$

donde las variables:

C_t : Gasto agregado en bienes de consumo

Y_t : Producto Nacional Bruto

r_t : Tasa promedio de rendimiento de la Deuda Pública

I_t : Inversión en bienes de equipo

m_t : Tasa de crecimiento de los Activos Líquidos en Manos del Público (ALP)

D_t : Stock de Deuda Pública viva

se hallan todas en desviaciones con respecto a sus medias muestrales. Se consideran exógenas m_t y D_t , y endógenas todas las variables restantes.

b) Proporcionar estimaciones eficientes de los parámetros α_1 , α_2 y β , disponiendo de los momentos muestrales:

	C_t	I_t	r_t	m_t	D_t
C_t	100	-80	-20	10	-40
I_t		80	-60	40	60
r_t			80	60	-60
m_t				100	50
D_t					50

Problema 18.22. Con objeto de estudiar la demanda de dinero para un determinado país, se propone el siguiente modelo:

$$\begin{aligned}
 M_{dt}^* &= a_1 + a_2 Y_t^e + a_3 R_t \\
 Y_t^e - Y_{t-1}^e &= d(Y_t - Y_{t-1}) \\
 M_{dt} - M_{dt-1} &= f(M_{dt}^* - M_{dt-1})
 \end{aligned}
 \tag{1}$$

donde:

M_d^* \equiv Stock deseado de saldos reales.

Y^e \equiv Renta real esperada.

R \equiv Tipo de interés.

N \equiv Población.

A partir de este modelo se han estimado por MCO las dos especificaciones siguientes:

$$\begin{aligned}
 \left(\frac{\hat{M}_d}{N}\right)_t &= 0,126 + 0,074 \left(\frac{Y}{N}\right)_t - 0,15 R_t + 0,002 R_{t-1} + 0,964 \left(\frac{M_d}{N}\right)_{t-1} \\
 &\quad (2,55) \quad (0,92) \quad (0,01) \quad (9,36)
 \end{aligned}
 \tag{2}$$

$$R^2 = 0,852$$

$$\begin{aligned}
 \left(\frac{\hat{M}_d}{N}\right)_t &= 0,127 + 0,074 \left(\frac{Y}{N}\right)_t - 0,014 R_t + 0,964 \left(\frac{M_d}{N}\right)_{t-1} \\
 &\quad (2,5) \quad (0,92) \quad (9,3)
 \end{aligned}
 \tag{3}$$

$$R^2 = 0,852$$

Se pide:

- Relacionar [1], [2] y [3].
- Determinar la validez del método de estimación empleado.
- Hacer un comentario crítico de la especificación del modelo.

Problema 18.23. Considere el modelo

$$y_{1t} = \alpha_1 + \beta_1 x_t + \varepsilon_{1t}$$

$$y_{2t} = \alpha_2 + \gamma_1 y_{1t} + \varepsilon_{2t}$$

- Si ε_1 y ε_2 no están correlacionados, ¿qué resultado proporcionaría la aplicación de MCO a las dos ecuaciones?
- ¿Qué resultado se obtendría aplicando MCO a la segunda ecuación en la que

se sustituye y_1 por \hat{y}_1 , siendo \hat{y}_1 el valor estimado de y_1 por MCO en la primera ecuación?

c) ¿Cómo variaría el resultado de b) si la segunda ecuación incorporara x_t como variable independiente?

Problema 18.24. Dado el modelo

$$\begin{aligned}C_t &= \alpha_1 + \alpha_2 Y_t + \varepsilon_t \\I_t &= \beta_1 + \beta_2 Y_t + \beta_3 G_{t-1} + \eta_t \\Y_t &= C_t + I_t + G_t\end{aligned}$$

se pide:

- Construir la forma reducida del modelo.
- Estudiar la identificación de las funciones de consumo e inversión.
- ¿Qué problema tendría el cálculo de la propensión marginal a consumir estimada aplicando MCO a la primera ecuación?

Problema 18.25. Dado el modelo de ecuaciones simultáneas:

$$[y_1 y_2] \begin{bmatrix} -1 & \gamma_2 \\ \gamma_1 & -1 \end{bmatrix} + [x_1 x_2 x_3] \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \\ 0 & \beta_3 \end{pmatrix} = [\varepsilon_1 \varepsilon_2]$$

en el que la primera ecuación está sobreidentificada y la segunda exactamente identificada y la información muestral en desviaciones respecto a la media. Encuentre las expresiones analíticas de:

- El estimador MCO de la primera ecuación.
- El estimador MCI de la segunda ecuación.
- El estimador MC2E de la primera ecuación.
- El estimador MC3E de ambas ecuaciones.

Problema 18.26. Dado el modelo macroeconómico de una economía abierta:

$$\begin{aligned}\text{Función de consumo:} & \quad C_t = \alpha_1 Y_t + \alpha_2 Y_{t-1} + u_t \\ \text{Función de importaciones:} & \quad M_t = \beta_1 C_t + \beta_2 C_{t-1} + \beta_3 E_t + v_t \\ \text{Identidad contable:} & \quad Y_t = C_t + E_t + M_t\end{aligned}$$

donde C_t denota el consumo agregado, E_t las exportaciones, M_t las importaciones e Y_t la producción de la economía, todas las variables están en desviaciones con respecto a la media y se considera que E_t es una variable exógena que tiene una estructura $MA(2)$.

Se pide:

- Discutir, con los procedimientos econométricos clásicos, la identificación del modelo, así como el procedimiento de estimación que se seguiría. Suponer que u_t y v_t son procesos de ruido blanco, con distribuciones Normales.
- ¿Qué restricciones implica en la representación matricial del proceso estocástico trivariante $Z_t = (C_t, Y_t, E_t)'$ la especificación anterior?
- ¿Qué implicaciones tiene la especificación propuesta respecto a las representaciones univariantes de cada una de las variables C_t e Y_t ?

Problema 18.27. a) Analice la situación de identificación de las ecuaciones del sistema:

$$\begin{aligned} y_{1t} &= a_{21}y_{2t} + b_{11}x_{1t} + b_{21}x_{2t} + b_{31}x_{3t} + u_{1t} \\ y_{2t} &= a_{12}y_{1t} + b_{22}x_{2t} + b_{32}x_{3t} + u_{2t} \end{aligned}$$

si u_{1t} y u_{2t} son ambos ruido blanco y $E(u_{1t}u_{2t}) = \sigma_{12}$.

b) ¿Cómo cambiaría su respuesta si $\sigma_{12} = 0$?

c) Demuestre que $b_{22} = -(\pi_{22} - \pi_{21}a_{12}) = \pi_{21} \frac{\pi_{12}}{\pi_{11}} - \pi_{22}$, donde π_{ij} denotan los coeficientes de la forma reducida.

d) Compruebe que utilizar π_{22} y π_{32} para obtener estimaciones de b_{22} y b_{32} da los mismos resultados que el procedimiento sugerido en a).

e) ¿Estaría identificada la primera ecuación si $\sigma_{12} \neq 0$ pero $b_{11} = b_{21}$?

f) Estime consistentemente la segunda ecuación a partir de la matriz de *momentos muestrales*:

	y_1	y_2	x_1	x_2	x_3
y_1	1.000	-50	-100	200	-50
y_2		800	-50	200	-40
x_1			800	200	0
x_2				100	0
x_3					200

y obtener su matriz de covarianzas como función del tamaño muestral T .

Problema 18.28. Considere el modelo:

$$\begin{cases} q_t = \alpha_1 p_t + \beta_1 I_t + \varepsilon_{dt} \\ q_t = \alpha_2 p_t + \beta_2 r_t + \varepsilon_{st} \end{cases}$$

donde:

$$E(I_v \varepsilon_{dt}) = E(I_v \varepsilon_{st}) = E(r_v \varepsilon_{dt}) = E(r_v \varepsilon_{st}) = 0 \quad \forall v, t$$

$$\begin{pmatrix} \varepsilon_{dt} \\ \varepsilon_{st} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} \sigma_d^2 & \sigma_{ds} \\ \sigma_{ds} & \sigma_s^2 \end{pmatrix} \right]$$

y donde todas las variables están en desviaciones con respecto a la media.

a) Encontrar $E(p_t \varepsilon_{dt})$ y $E(p_t \varepsilon_{st})$ como función de σ_d^2 , σ_{ds} , σ_s^2 y de los coeficientes del modelo.

b) Probar que la estimación MCO de la primera ecuación es inconsistente, evaluando el sesgo asintótico de los coeficientes estimados, como función de los mismos parámetros que en a).

c) Encontrar estimaciones consistentes de ambas ecuaciones, a partir de la matriz de momentos muestrales:

	q_t	p_t	I_t	r_t
q_t	100	60	40	-20
p_t		40	-40	20
I_t			60	0
r_t				40

d) Analice el caso en que $\beta_2 = 0$. ¿Se podrán obtener estimaciones consistentes de todos los parámetros del modelo?

Problema 18.29. El siguiente modelo muestra que la identificación exacta de una ecuación en un sistema no implica que el uso de MCO en dicha ecuación únicamente genera estimaciones consistentes.

Considere el sistema:

$$\begin{aligned} \text{Oferta: } & q_t^o = \beta_{11} + \gamma_{11}p_t + \beta_{21}x_t + \varepsilon_{1t} \\ \text{Demanda: } & q_t^d = \beta_{12} + \gamma_{12}p_t + \quad \quad \quad + \varepsilon_{2t}, \quad t = 1, 2, \dots, T \end{aligned}$$

con $(\varepsilon_{1t}, \varepsilon_{2t})$ i. i. d., $N(0, \Omega)$ y x_t exógena.

1. Pruebe que la primera ecuación está exactamente identificada.
2. Pruebe que si $\hat{\gamma}_{21}$ denota el estimador MCO de γ_{21} se tiene:

$$plim \hat{\gamma}_{21} = \gamma_{21} + \frac{\text{Cov}(p_t, \varepsilon_t)}{\text{Var } p_t}$$

3. Puesto que p_t es endógeno, se tiene que $\text{Cov}(p_t, \varepsilon_t) \neq 0$. Pruebe que:

$$\text{Cov}(p_t, \varepsilon_t) = \frac{\sigma_2^2 - \sigma_{12}}{\gamma_{11} - \gamma_{21}}$$

donde $\sigma_{12} = \text{Cov}(\varepsilon_{1t}, \varepsilon_{2t})$ y $\sigma_2^2 = \text{Var } \varepsilon_{2t}$.

4. Si Ω es diagonal, ¿qué signo debería esperarse para el sesgo asintótico de $\hat{\gamma}_{21}$?

APENDICE

TABLA A-1. Ordenadas de la función de densidad Normal (0, 1)

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0007
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

(Tomada de Mood y Graybill, *Introduction to the Theory of Statistics*, 3.ª ed., McGraw-Hill Book Company, Nueva York, 1973. Con permiso de los editores.)

TABLA A-2. Función de distribución de la variable Normal (0, 1)

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8989	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,8452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9116	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

x	1,282	1,645	1,960	2,326	2,576	3,090	3,291	3,891	4,417
$\Phi(x)$	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	0,99995	0,999995
$2[1 - \Phi(x)]$	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001	0,00001

(Tomada de Mood y Graybill, *Introduction to the Theory of Statistics*, 3.^a ed., McGraw-Hill Book Company, Nueva York, 1973. Con permiso de los editores.)

TABLA A-3. Función de distribución de la variable chi-cuadrado χ_n^2 *

$$F(u) = \int_0^u \frac{x^{(n-2)/2} e^{-x/2} dx}{2^{n/2} \Gamma(n/2)}$$

$n \backslash F$	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,995
1	0,0 ⁴ 393	0,0 ³ 157	0,0 ³ 982	0,0 ² 393	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7

* Esta tabla está resumida de «Tables of percentage points of the incomplete beta function and of the chi-square distribution», *Biometrika*, vol 32 (1941). Se publica con el permiso del autor, Catherine M. Thompson, y el editor de *Biometrika*.

TABLA A-4. Función de distribución de la variable t de Student*

$$F(t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{\pi n}\left(1 + \frac{x^2}{n}\right)^{(n+1)/2}} dx$$

$n \backslash F$	0,75	0,90	0,95	0,975	0,99	0,995	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	6,859
6	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	5,405
8	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,645
40	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,291

* Esta tabla está resumida de *Statistical Tables*, de R. A. Fisher y Frank Yates, editada por Oliver & Boyd, Ltd., Edinburgh and London, 1938. Se publica con el permiso de los autores y de los editores.

TABLA A-5. Función de distribución de la variable F , percentiles 95*

		Grados de libertad del numerador																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
Grados de libertad del denominador	1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254	
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53	8,53
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63	5,63
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37	4,37
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67	3,67
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23	3,23
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93	2,93
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71	2,71
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54	2,54
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40	2,40
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30	2,30
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21	2,21
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13	2,13
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07	2,07
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01	2,01
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96	1,96
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92	1,92
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88	1,88
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84	1,84
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81	1,81
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78	1,78
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76	1,76
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73	1,73
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62	1,62	
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51	1,51	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39	1,39	
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25	1,25	
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00	1,00	

La interpolación debe llevarse a cabo utilizando el inverso del número de grados de libertad.

* Esta tabla está tomada con el permiso de Biometrika Trustees from M. Merrington, C. M. Thompson, «Tables of percentage points of the inverted beta (F) distribution», *Biometrika*, vol. 33, pág. 73, 1943.

TABLA A-6. Función de distribución de la variable F , percentiles 99*

		Grados de libertad del numerador																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
Grados de libertad del denominador	1	4.052	5.000	5.403	5.625	5.764	5.859	5.928	5.982	6.023	6.056	6.106	6.157	6.209	6.235	6.261	6.287	6.313	6.339	6.366	
	2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5
	3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2	26,1	26,1
	4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7	13,6	13,5	13,5
	5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02	9,02
	6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88	6,88
	7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65	5,65
	8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86	4,86
	9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31	4,31
	10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91	3,91
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60	3,60
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36	3,36
	13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17	3,17
	14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00	3,00
	15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87	2,87
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75	2,75
	17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65	2,65
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57	2,57
	19	8,19	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49	2,49
	20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42	2,42
	21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36	2,36
	22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31	2,31
	23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26	2,26
	24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21	2,21
	25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,53	2,45	2,36	2,27	2,17	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01	2,01	
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80	1,80	
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60	1,60	
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38	1,38	
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00	1,00	

La interpolación debe llevarse a cabo utilizando el inverso del número de grados de libertad.

* Esta tabla está tomada con el permiso de Biometrika Trustees from M. Merrington, C. M. Thompson, «Tables of percentage points of the inverted beta (F) distribution», *Biometrika*, vol. 33, pág. 73, 1943.

TABLA A-7. Estadístico d de Durbin-Watson. Valores críticos de d_i y d_s al nivel de significación del 1 por 100*

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,42	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

n = número de observaciones.

k' = número de variables explicativas, excluyendo el término constante.

* Esta tabla está tomada de *Biometrika*, vol. 41, pág. 175, 1951, con el permiso de los administradores.

TABLA A-8. Estadístico d de Durbin-Watson. Valores críticos de d_i y d_s al nivel de significación del 5 por 100*

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

n = número de observaciones.

k' = número de variables explicativas, excluyendo el término constante.

* Esta tabla está tomada de *Biometrika*, vol. 41, pág. 173, 1951, con el permiso de los administradores.

TABLA A-9. Estadístico de Wallis para la autocorrelación de cuarto orden

Valores críticos de d_i y d_s al nivel de significación del 5 por 100, para regresiones sin variables ficticias estacionales ($k' = k - 1$).

n	k' = 1		k' = 2		k' = 3		k' = 4		k' = 5	
	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s
16	0,774	0,982	0,662	1,109	0,549	1,275	0,435	1,381	0,350	1,532
20	0,924	1,102	0,827	1,203	0,728	1,327	0,626	1,428	0,544	1,556
24	1,036	1,189	0,953	1,273	0,867	1,371	0,779	1,459	0,702	1,565
28	1,123	1,257	1,050	1,328	0,975	1,410	0,898	1,487	0,828	1,576
32	1,192	1,311	1,127	1,373	1,061	1,443	0,993	1,511	0,929	1,587
36	1,248	1,355	1,191	1,410	1,131	1,471	1,070	1,532	1,013	1,598
40	1,295	1,392	1,243	1,442	1,190	1,496	1,135	1,550	1,082	1,609
44	1,335	1,423	1,288	1,469	1,239	1,518	1,189	1,567	1,141	1,620
48	1,369	1,451	1,326	1,493	1,281	1,537	1,236	1,582	1,191	1,630
52	1,399	1,475	1,359	1,513	1,318	1,554	1,276	1,595	1,235	1,639
56	1,426	1,496	1,389	1,532	1,351	1,569	1,312	1,608	1,273	1,648
60	1,449	1,515	1,415	1,548	1,379	1,583	1,343	1,619	1,307	1,656
64	1,470	1,532	1,438	1,563	1,405	1,596	1,371	1,629	1,337	1,664
68	1,489	1,548	1,459	1,577	1,427	1,608	1,396	1,639	1,364	1,671
72	1,507	1,562	1,478	1,589	1,448	1,618	1,418	1,648	1,388	1,678
76	1,522	1,574	1,495	1,601	1,467	1,628	1,439	1,656	1,411	1,685
80	1,537	1,586	1,511	1,611	1,484	1,637	1,457	1,663	1,431	1,691
84	1,550	1,597	1,525	1,621	1,500	1,646	1,475	1,671	1,449	1,696
88	1,562	1,607	1,539	1,630	1,515	1,654	1,490	1,677	1,466	1,702
92	1,574	1,617	1,551	1,639	1,528	1,661	1,505	1,684	1,482	1,707
96	1,584	1,626	1,563	1,647	1,541	1,668	1,519	1,690	1,496	1,712
100	1,594	1,634	1,573	1,654	1,552	1,674	1,531	1,695	1,510	1,717

Valores críticos de d_i y d_s al nivel de significación del 5 por 100, para regresiones que incluyen un término constante y variables ficticias estacionales ($k'' = k - 4$).

n	k'' = 1		k'' = 2		k'' = 3		k'' = 4		k'' = 5	
	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s	d_i	d_s
16	1,156	1,381	1,031	1,532	0,902	1,776	0,777	2,191	0,693	2,238
20	1,228	1,428	1,123	1,556	1,013	1,726	0,899	1,954	0,806	2,042
24	1,287	1,459	1,199	1,565	1,107	1,694	1,011	1,856	0,928	1,949
28	1,337	1,487	1,261	1,576	1,181	1,679	1,099	1,803	1,025	1,889
32	1,379	1,511	1,312	1,587	1,243	1,673	1,171	1,773	1,104	1,850
36	1,414	1,532	1,355	1,598	1,293	1,672	1,230	1,755	1,170	1,824
40	1,445	1,550	1,391	1,609	1,336	1,674	1,279	1,745	1,225	1,807
44	1,471	1,567	1,422	1,620	1,373	1,677	1,321	1,739	1,272	1,795
48	1,494	1,582	1,450	1,630	1,404	1,681	1,357	1,737	1,312	1,788
52	1,514	1,595	1,474	1,639	1,432	1,686	1,389	1,736	1,347	1,782
56	1,533	1,608	1,495	1,648	1,456	1,691	1,416	1,736	1,377	1,779
60	1,549	1,619	1,514	1,656	1,478	1,696	1,441	1,737	1,404	1,777
64	1,564	1,629	1,531	1,664	1,497	1,700	1,463	1,739	1,429	1,776
68	1,577	1,639	1,546	1,671	1,515	1,705	1,482	1,741	1,450	1,775
72	1,590	1,648	1,560	1,678	1,531	1,710	1,500	1,743	1,470	1,776
76	1,601	1,656	1,573	1,685	1,545	1,714	1,517	1,746	1,488	1,776
80	1,611	1,663	1,585	1,691	1,559	1,719	1,531	1,748	1,504	1,777
84	1,621	1,671	1,596	1,696	1,571	1,723	1,545	1,751	1,519	1,778
88	1,630	1,677	1,607	1,702	1,582	1,727	1,558	1,753	1,533	1,779
92	1,639	1,684	1,616	1,707	1,593	1,731	1,570	1,756	1,546	1,781
96	1,647	1,690	1,625	1,712	1,603	1,735	1,580	1,759	1,558	1,782
100	1,654	1,695	1,633	1,717	1,612	1,739	1,591	1,761	1,569	1,784

TABLA A-10. Valores críticos de c_0 en el contraste CUSUM

m	α					m	α				
	0,10	0,05	0,025	0,01	0,005		0,10	0,05	0,025	0,01	0,005
1	0,40000	0,45000	0,47500	0,49000	0,49500	41	0,14916	0,17215	0,19254	0,21667	0,23331
2	0,35044	0,44306	0,50855	0,56667	0,59596	42	0,14761	0,17034	0,19050	0,21436	0,23081
3	0,35477	0,41811	0,46702	0,53456	0,57900	43	0,14611	0,16858	0,18852	0,21212	0,22839
4	0,33435	0,39075	0,44641	0,50495	0,54210	44	0,14466	0,16688	0,18661	0,20995	0,22605
5	0,31556	0,37359	0,42174	0,47692	0,51576	45	0,14325	0,16524	0,18475	0,20785	0,22377
6	0,30244	0,35522	0,40045	0,45440	0,48988	46	0,14188	0,16364	0,18295	0,20581	0,22157
7	0,28991	0,33905	0,38294	0,43337	0,46761	47	0,14055	0,16208	0,18120	0,20383	0,21943
8	0,27828	0,32538	0,36697	0,41522	0,44819	48	0,13926	0,16058	0,17950	0,20190	0,21735
9	0,26794	0,31325	0,35277	0,39922	0,43071	49	0,13800	0,15911	0,17785	0,20003	0,21534
10	0,25884	0,30221	0,34022	0,38481	0,41517	50	0,13678	0,15769	0,17624	0,19822	0,21337
11	0,25071	0,29227	0,32894	0,37187	0,40122	51	0,13559	0,15630	0,17468	0,19645	0,21146
12	0,24325	0,28330	0,31869	0,36019	0,38856	52	0,13443	0,15495	0,17316	0,19473	0,20961
13	0,23639	0,27515	0,30935	0,34954	0,37703	53	0,13330	0,15363	0,17168	0,19305	0,20780
14	0,23010	0,26767	0,30081	0,33980	0,36649	54	0,13221	0,15235	0,17024	0,19142	0,20604
15	0,22430	0,26077	0,29296	0,33083	0,35679	55	0,13113	0,15110	0,16884	0,18983	0,20432
16	0,21895	0,25439	0,28570	0,32256	0,34784	56	0,13009	0,14989	0,16746	0,18828	0,20265
17	0,21397	0,24847	0,27897	0,31489	0,33953	57	0,12907	0,14870	0,16613	0,18677	0,20101
18	0,20933	0,24296	0,27270	0,30775	0,33181	58	0,12807	0,14754	0,16482	0,18529	0,19942
19	0,20498	0,23781	0,26685	0,30108	0,32459	59	0,12710	0,14641	0,16355	0,18385	0,19786
20	0,20089	0,23298	0,26137	0,29484	0,31784	60	0,12615	0,14530	0,16230	0,18245	0,19635
21	0,19705	0,22844	0,25622	0,28898	0,31149	62	0,12431	0,14316	0,15990	0,17973	0,19341
22	0,19343	0,22416	0,25136	0,28346	0,30552	64	0,12255	0,14112	0,15760	0,17713	0,19061
23	0,19001	0,22012	0,24679	0,27825	0,29989	66	0,12087	0,13916	0,15540	0,17464	0,18792
24	0,18677	0,21630	0,24245	0,27333	0,29456	68	0,11926	0,13728	0,15329	0,17226	0,18535
25	0,18370	0,21268	0,23835	0,26866	0,28951	70	0,11771	0,13548	0,15127	0,16997	0,18288
26	0,18077	0,20924	0,23445	0,26423	0,28472	72	0,11622	0,13375	0,14932	0,16777	0,18051
27	0,17799	0,20596	0,23074	0,26001	0,28016	74	0,11479	0,13208	0,14745	0,16566	0,17823
28	0,17533	0,20283	0,22721	0,25600	0,27582	76	0,11341	0,13048	0,14565	0,16363	0,17604
29	0,17280	0,19985	0,22383	0,25217	0,27168	78	0,11208	0,12894	0,14392	0,16167	0,17392
30	0,17037	0,19700	0,22061	0,24851	0,26772	80	0,11079	0,12745	0,14224	0,15978	0,17188
31	0,16805	0,19427	0,21752	0,24501	0,26393	82	0,10955	0,12601	0,14063	0,15795	0,16992
32	0,16582	0,19166	0,21457	0,24165	0,26030	84	0,10835	0,12462	0,13907	0,15619	0,16802
33	0,16368	0,18915	0,21173	0,23843	0,25683	86	0,10719	0,12327	0,13756	0,15449	0,16618
34	0,16162	0,18674	0,20901	0,23534	0,25348	88	0,10607	0,12197	0,13610	0,15284	0,16440
35	0,15964	0,18442	0,20639	0,23237	0,25027	90	0,10499	0,12071	0,13468	0,15124	0,16268
36	0,15774	0,18218	0,20387	0,22951	0,24718	92	0,10393	0,11949	0,13331	0,14970	0,16101
37	0,15590	0,18003	0,20144	0,22676	0,24421	94	0,10291	0,11831	0,13198	0,14820	0,15940
38	0,15413	0,17796	0,19910	0,22410	0,24134	96	0,10192	0,11716	0,13070	0,14674	0,15783
39	0,15242	0,17595	0,19684	0,22154	0,23857	98	0,10096	0,11604	0,12944	0,14533	0,15631
40	0,15076	0,17402	0,19465	0,21906	0,23589	100	0,10002	0,11496	0,12823	0,14396	0,15483

Los valores de c_0 se utilizan para determinar el par de rectas: $s_r = \pm c_0 + \frac{r-k}{n-k}$. Para n observaciones, k variables explicativas (incluyendo la constante, si la hay) y nivel de significación α , c_0 se determina examinando la tabla para $m = \frac{1}{2}(n-k) - 1$ y $\frac{1}{2}\alpha$. Para contrastes de una cola se utiliza α . Si $n-k$ es impar, se interpola linealmente entre $m = \frac{1}{2}(n-k) - \frac{3}{2}$ y $m = \frac{1}{2}(n-k) - \frac{1}{2}$.

TABLA A-11. Función de distribución empírica de $\hat{\tau}$ para $\hat{\rho} = 1$

Tamaño muestral n	Nivel de significación $\alpha: \alpha = P(\hat{\tau} < x)$							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
$\hat{\tau}$: Modelo [14.2]								
25	-2,66	-2,26	-1,95	-1,60	0,92	1,33	1,70	2,16
50	-2,62	-2,25	-1,95	-1,61	0,91	1,31	1,66	2,08
100	-2,60	-2,24	-1,95	-1,61	0,90	1,29	1,64	2,03
250	-2,58	-2,24	-1,95	-1,62	0,89	1,29	1,63	2,01
500	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00
∞	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00
$\hat{\tau}_\mu$: Modelo [14.3]								
25	-3,75	-3,33	-3,00	-2,63	-0,37	0,00	0,34	0,72
50	-3,58	-3,22	-2,93	-2,60	-0,40	-0,03	0,29	0,66
100	-3,51	-3,17	-2,89	-2,58	-0,42	-0,05	0,26	0,63
250	-3,46	-3,14	-2,88	-2,57	-0,42	-0,06	0,24	0,62
500	-3,44	-3,13	-2,87	-2,57	-0,43	-0,07	0,24	0,61
∞	-3,43	-3,12	-2,86	-2,57	-0,44	-0,07	0,23	0,60
$\hat{\tau}_\tau$: Modelo [14.4]								
25	-4,38	-3,95	-3,60	-3,24	-1,14	-0,80	-0,50	-0,15
50	-4,15	-3,80	-3,50	-3,18	-1,19	-0,87	-0,58	-0,24
100	-4,04	-3,73	-3,45	-3,15	-1,22	-0,90	-0,62	-0,28
250	-3,99	-3,69	-3,43	-3,13	-1,23	-0,92	-0,64	-0,31
500	-3,98	-3,68	-3,42	-3,13	-1,24	-0,93	-0,65	-0,32
∞	-3,96	-3,66	-3,41	-3,12	-1,25	-0,94	-0,66	-0,33

FUENTE: Fuller (1976), *Introduction to Statistical Time Series*, Wiley, Nueva York, pág. 373.

TABLA A-12. Función de distribución empírica de $n(\hat{\rho} - 1)$ para $\rho = 1$

Tamaño muestral n	Nivel de significación $\alpha = P[n(\hat{\rho} - 1) < x]$							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
$\hat{\rho}$: Modelo [14.2]								
25	-11,9	-9,3	-7,3	-5,3	1,01	1,40	1,79	2,28
50	-12,9	-9,9	-7,7	-5,5	0,97	1,35	1,70	2,16
100	-13,3	-10,2	-7,9	-5,6	0,95	1,31	1,65	2,09
250	-13,6	-10,3	-8,0	-5,7	0,93	1,28	1,62	2,04
500	-13,7	-10,4	-8,0	-5,7	0,93	1,28	1,61	2,04
∞	-13,8	-10,5	-8,1	-5,7	0,93	1,28	1,60	2,03
$\hat{\rho}_\mu$: Modelo [14.3]								
25	-17,2	-14,6	-12,5	-10,2	-0,76	0,01	0,65	1,40
50	-18,9	-15,7	-13,3	-10,7	-0,81	-0,07	0,53	1,22
100	-19,8	-16,3	-13,7	-11,0	-0,83	-0,10	0,47	1,14
250	-20,3	-16,6	-14,0	-11,2	-0,84	-0,12	0,43	1,09
500	-20,5	-16,8	-14,0	-11,2	-0,84	-0,13	0,42	1,06
∞	-20,7	-16,9	-14,1	-11,3	-0,85	-0,13	0,41	1,04
$\hat{\rho}_\tau$: Modelo [14.4]								
25	-22,5	-19,9	-17,9	-15,6	-3,66	-2,51	-1,53	-0,43
50	-25,7	-22,4	-19,8	-16,8	-3,71	-2,60	-1,66	-0,65
100	-27,4	-23,6	-20,7	-17,5	-3,74	-2,62	-1,73	-0,75
250	-28,4	-24,4	-21,3	-18,0	-3,75	-2,64	-1,78	-0,82
500	-28,9	-24,8	-21,5	-18,1	-3,76	-2,65	-1,78	-0,84
∞	-29,5	-25,1	-21,8	-18,3	-3,77	-2,66	-1,79	-0,87

FUENTE: Fuller (1976), *Introduction to Statistical Time Series*, Wiley, Nueva York, pág. 371.

TABLA A-13. Distribución empírica de $\hat{t}_{\alpha\mu}$ para $(\alpha, \rho) = (0, 1)$ en $Y_t = \alpha + \rho Y_{t-1} + e_t$ (Distribución simétrica)

Tamaño muestral <i>n</i>	$P(\hat{t}_{\alpha\mu} < x)$			
	0,90	0,95	0,975	0,99
25	2,20	2,61	2,97	3,41
50	2,18	2,56	2,89	3,28
100	2,17	2,54	2,86	3,22
250	2,16	2,53	2,84	3,19
500	2,16	2,52	2,83	3,18
∞	2,16	2,52	2,83	3,18
s.e.	0,003	0,004	0,006	0,008

FUENTE: Dickey y Fuller (1981), pág. 1062.

En las Tablas A-13, A-14 y A-15, los valores para niveles de significación bajos pueden derivarse por la simetría de las distribuciones.

TABLA A-14. Distribución empírica de $\hat{t}_{\alpha t}$ para $(\alpha, \beta, \rho) = (0, 0, 1)$ en $Y_t = \alpha + \beta t + \rho Y_{t-1} + e_t$ (Distribución simétrica)

Tamaño muestral <i>n</i>	$P(\hat{t}_{\alpha t} < x)$			
	0,90	0,95	0,975	0,99
25	2,77	3,20	3,59	4,05
50	2,75	3,14	3,47	3,87
100	2,73	3,11	3,42	3,78
250	2,73	3,09	3,39	3,74
500	2,72	3,08	3,38	3,72
∞	2,72	3,08	3,38	3,71
s.e.	0,004	0,005	0,007	0,008

FUENTE: Dickey y Fuller (1981), pág. 1062.

TABLA A-15. Distribución empírica de $\hat{\tau}_{\beta t}$ para $(\alpha, \beta, \rho) = (\alpha, 0, 1)$ en $Y_t = \alpha + \beta t + \rho Y_{t-1} + e_t$ (Distribución simétrica)

Tamaño muestral <i>n</i>	$P(\hat{\tau}_{\beta t} < x)$			
	0,90	0,95	0,975	0,99
25	2,39	2,85	3,25	3,74
50	2,38	2,81	3,18	3,60
100	2,38	2,79	3,14	3,53
250	2,38	2,79	3,12	3,49
500	2,38	2,78	3,11	3,48
∞	2,38	2,78	3,11	3,46
s.e.	0,004	0,005	0,006	0,009

FUENTE: Dickey y Fuller (1981), pág. 1062.

TABLA A-16. Distribución empírica de Φ_1 para $(\alpha, \rho) = (0, 1)$ en $Y_t = \alpha + \rho Y_{t-1} + e_t$

Tamaño muestral <i>n</i>	$P(\Phi_1 < x)$							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
25	0,29	0,38	0,49	0,65	4,12	5,18	6,30	7,88
50	0,29	0,39	0,50	0,66	3,94	4,86	5,80	7,06
100	0,29	0,39	0,50	0,67	3,86	4,71	5,57	6,70
250	0,30	0,39	0,51	0,67	3,81	4,63	5,45	6,52
500	0,30	0,39	0,51	0,67	3,79	4,61	5,41	6,47
∞	0,30	0,40	0,51	0,67	3,78	4,59	5,38	6,43
s.e.	0,002	0,002	0,002	0,002	0,01	0,02	0,03	0,05

FUENTE: Dickey y Fuller (1981), pág. 1063.

TABLA A-17. Distribución empírica de Φ_2 para $(\alpha, \beta, \rho) = (0, 0, 1)$
 en $Y_t = \alpha + \beta t + \rho Y_{t-1} + e_t$

Tamaño muestral <i>n</i>	$P(\Phi_2 < x)$							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
25	0,61	0,75	0,89	1,10	4,67	5,68	6,75	8,21
50	0,62	0,77	0,91	1,12	4,31	5,13	5,94	7,02
100	0,63	0,77	0,92	1,12	4,16	4,88	5,59	6,50
250	0,63	0,77	0,92	1,13	4,07	4,75	5,40	6,22
500	0,63	0,77	0,92	1,13	4,05	4,71	5,35	6,15
∞	0,63	0,77	0,92	1,13	4,03	4,68	5,31	6,09
s.e.	0,003	0,003	0,003	0,003	0,01	0,02	0,03	0,05

FUENTE: Dickey y Fuller (1981), pág. 1063.

TABLA A-18. Distribución empírica de Φ_3 para $(\alpha, \beta, \rho) = (\alpha, 0, 1)$
 en $Y_t = \alpha + \beta t + \rho Y_{t-1} + e_t$

Tamaño muestral <i>n</i>	$P(\Phi_3 < x)$							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
25	0,74	0,90	1,08	1,33	5,91	7,24	8,65	10,61
50	0,76	0,93	1,11	1,37	5,61	6,73	7,81	9,31
100	0,76	0,94	1,12	1,38	5,47	6,49	7,44	8,73
250	0,76	0,94	1,13	1,39	5,39	6,34	7,25	8,43
500	0,76	0,94	1,13	1,39	5,36	6,30	7,20	8,34
∞	0,77	0,94	1,13	1,39	5,34	6,25	7,16	8,27
s.e.	0,004	0,004	0,003	0,004	0,015	0,020	0,032	0,058

FUENTE: Dickey y Fuller (1981), pág. 1063.

BIBLIOGRAFIA

- ALMON, S. (1965): «The Distributed Lag Between Capital Appropriations and Expenditures», *Econometrica*, 33, 178-196.
- AMEMIYA, T. (1981): «Qualitative Response Models: A Survey», *Journal of Economic Literature*, 19, 483-536.
- ANDERSON, T. W., y C. HSIAO (1981): «Estimation of Dynamic Models with Error Components», *Journal of the American Statistical Association*, 76, 598-606.
- ARELLANO, M. (1992): «Introducción al análisis econométrico con datos de panel», *Documento de Trabajo*, Servicio de Estudios, Banco de España.
- ARELLANO, M. (1993): «On the Testing of Correlated Effects with Panel Data», *Journal of Econometrics*.
- ARELLANO, M., y O. BOVER (1990): «La econometría de datos de panel», *Investigaciones Económicas*, 14, 3-45.
- ARELLANO, M., y S. BOND (1988): «Dynamic Panel Data Estimation Using DPD - A Guide for User», *Institute for Fiscal Studies*, working paper 88/15, Londres.
- ARELLANO, M., y S. BOND (1991): «Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations», *Review of Economic Studies*, 58, 277-297.
- ARNAIZ, G. (1978): *Introducción a la estadística teórica*, 3.^a edición, Lex Nova.
- AZNAR, A., y A. GARCÍA FERRER (1980): *Problemas de Econometría*. Ed. Pirámide.
- AZNAR, A., y J. TRIVEZ (1993): *Métodos de predicción en Economía*. Barcelona, Ariel.
- BANERJEE, A.; J. DOLADO; D. HENDRY, y G. SMITH (1986): «Exploring Equilibrium Relationships Through Static Models: Some Monte Carlo Evidence», *Oxford Bulletin of Economics and Statistics*, 48, 253-277.
- BARTLETT, M. S. (1946): «On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series», *Journal of the Royal Statistical Society*, 27.
- BASMANN, R. L. (1960): «On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics», *Journal of the American Statistical Association*, 55, 650-659.
- BERA, A. K., y C. M. JARQUE (1981): «An Efficient Large-Sample Test for Normality of Observations and Regression Residuals», *Working Paper in Econometrics*, 40, Australian National University, Canberra.
- BERNDT, E. K.; B. H. HALL; T. E. HALL, y J. A. HAUSMAN (1974): «Estimation and Inference in Nonlinear Structural Models», *Annals of Economic and Social Measurement*, 3, 653-666.

- BOX, G. E. P., y D. R. COX (1964): «An Analysis of Transformations», *Journal of the Royal Statistical Society*, Series B, 26, 211-243.
- BOX, G. E. P., y G. M. JENKINS (1970): *Time Series Analysis: Forecasting and Control*. San Francisco, Holden Day.
- BREUSCH, T. S. (1978): «Testing for Autocorrelation in Dynamic Linear Models: *Australian Economic Papers*, 17, 334-355.
- BREUSCH, T. S., y A. R. PAGAN (1979): «A Simple Test for Heteroscedasticity and Random Coefficient Variation», *Econometrica*, 47, 1287-1294.
- BREUSCH, T. S., y A. R. PAGAN (1980): «The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics», *Review of Economic Studies*, 47, 239-254.
- CAGAN, P. (1956): «The Monetary Dynamics of Hyperinflation» *Studies in the Quantity Theory of Money*, editado por M. Friedman. University of Chicago Press.
- CHAMBERLAIN, G. (1984): «Panel Data», en *Handbook of Econometrics*, vol. II, Elsevier Science, editado por Z. Griliches y M. D. Intriligator.
- DICKEY, D. A.; D. HASZA, y W. FULLER (1984): «Testing for Unit Roots in Seasonal Time Series», *Journal of the American Statistical Association*, 79, 355-367.
- DICKEY, D. A., y S. PANTULA (1987): «Determining the Order of Differencing in Autoregressive Processes», *Journal of Business and Economic Statistics*, 15, 455-461.
- DICKEY, D. A.; W. BELL, y R. MILLER (1986): «Unit Roots in Time Series Models: Tests and Implications», *The American Statistician*, 40, 12-26.
- DICKEY, D. A., y W. FULLER (1979): «Distribution of the Estimators for Autoregressive Time Series with a Unit Root», *Journal of the American Statistical Association*, 84, 427-431.
- DICKEY, D. A., y W. FULLER (1981): «Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root», *Econometrica*, 50, 1057-1072.
- ENGLE, R., y B. S. YOO (1987): «Forecasting and Testing in Cointegrated Systems», *Journal of Econometrics*, 35, 143-159.
- ENGLE, R., y C. W. GRANGER (1987): «Cointegration and Error Correction: Representation, Estimation and Testing», *Econometrica*, 55, 251-276.
- ENGLE, R.; C. W. GRANGER; S. HYLLEBERG, y B. S. YOO (1987): «Seasonal Integration and Co-integration», *Documento de Trabajo*, Departamento de Economía, Universidad de California en San Diego.
- ENGLE, R.; C. W. GRANGER; S. HYLLEBERG, y B. S. YOO (1987): «Seasonal Integration and Cointegration», Discussion Paper n.º 88-32, Universidad de San Diego.
- FAIR, R. (1977): «A Note on Computation of the Tobit Estimator», *Econometrica*, 45, 1723-1727.
- FELDSTEIN, M. (1973): «Multicollinearity and the Mean Square Error of Alternative Estimators», *Econometrica*, 41, 337-346.
- FRIEDMAN, M. (1956): *A Theory of the Consumption Function*. Princeton, New Jersey, Princeton University Press.
- FULLER, W. (1976): *Introduction to Statistical Time Series*. Nueva York, John Wiley.
- GODFREY, L. G. (1978): «Testing Against General Autoregressive and Moving Average Models when the Regressors Include Lagged Dependent Variables», *Econometrica*, 46, 1293-1302.
- GOLDBERGER, A. S. (1962): «Best Linear Unbiased Prediction in the Generalized Linear Regression Model», *Journal of the American Statistical Association*, 57, 369-375.
- GOLDBERGER, A. S. (1964): *Econometric Theory*. Nueva York, John Wiley.
- GOLDFELD, S. M., y R. E. QUANDT (1965): «Some Tests for Homoscedasticity», *Journal of the American Statistical Association*, 60, 539-547.

- GOLDFELD, S. M., y R. E. QUANDT (1972): *Nonlinear Methods in Econometrics*. North Holland.
- GRACIA DíEZ, M., y A. NOVALES (1993): «Una guía para la estimación de modelos ARCH», de próxima publicación en *Estadística Española*.
- GRANGER, C., y P. NEWBOLD (1974): «Spurious Regressions in Econometrics», *Journal of Econometrics*, 2, 111-120.
- GRILICHES, Z. (1977): «Estimating the Returns to Schooling: Some Econometric Problems», *Econometrica*, 45, 1-22.
- GRILICHES, Z., y J. HAUSMAN (1986): «Errors in Variables in Panel Data», *Journal of Econometrics*, 31, 93-118.
- HANSEN, L. P., y K. SINGLETON (1982): «Instrumental Variable Estimation of Non-linear Rational Expectations Models», *Econometrica*, 50, 1269-1287.
- HARVEY, A. C. (1976): «Estimating Regressions Models with Multiplicative Heteroscedasticity», *Econometrica*, 44, 461-465.
- HARVEY, A. C., y G. D. A. PHILLIPS (1974): «A Comparison of the Power of Some Tests for Heteroscedasticity in the General Linear Model», *Journal of Econometrics*, 2, 307-316.
- HASZA, D. P., y W. A. FULLER (1982): «Testing for Nonstationary Parameter Specifications in Seasonal Time Series Models», *The Annals of Statistics*, 10, 1209-1216.
- HATANAKA, M. (1974): «An Efficient Two-step Estimator for the Dynamic Adjustment Model with Autoregressive Errors», *Journal of Econometrics*, 2, 199-220.
- HAUSMAN, J. A. (1978): «Specification Tests in Econometrics», *Econometrica*, 46, 1251-1271.
- HAUSMAN, J. A., y W. E. TAYLOR (1981): «Panel Data and Unobservable Individual Effects», *Econometrica*, 49, 1377-1398.
- HECKMAN, J. (1976): «The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models», *The Annals of Economic and Social Measurement*, 5, 475-492.
- HECKMAN, J. (1979): «Sample Selection Bias as a Specification Error», *Econometrica*, 47, 153-161.
- HOERL, A. E., y R. W. KENNARD (1970): «Ridge Regression: Biased Estimation of Nonorthogonal Problems», *Technometrics*, 12, 55-67.
- HSIAO, C. (1986): *Analysis of Panel Data*. Cambridge University Press.
- HYLLEBERG, S.; R. ENGLE; C. W. GRANGER, y B. S. YOO (1990): «Seasonal Integration and Co-integration», *Journal of Econometrics*, 99, 215-238.
- JOHNSTON, J. (1984): *Econometric Methods*, 3.^a edición. McGraw-Hill.
- JUDGE, G.; W. GRIFFITHS; R. HILL; H. LÜTKEPOHL, y T. LEE (1985): *The Theory and Practice of Econometrics*, 2.^a edición. Nueva York, John Wiley.
- LEE, J. F. (1979): «Efficient Estimation of Dynamic Error Component Models with Panel Data», Discussion Paper n.º 79-118. Center for Economic Research, University of Minnesota.
- LINDGREN, B. W. (1976): *Statistical Theory*, 3.^a edición, MacMillan Co.
- LJUNG, T., y G. BOX (1979): «The Likelihood Function for a Stationary, Autoregressive. Moving Average Process», *Biometrika*, 66, 265-270.
- MADDALA, G. S. (1971): «The Use of Variance Components Models in Pooling Cross-Sections and Time Series Data», *Econometrica*, 39, 341-358.
- MADDALA, G. S. (1985): *Econometría*. McGraw-Hill.
- MALINVAUD, E. (1970): *Statistical Methods in Econometrics*, 2.^a edición. Amsterdam, New-Holland.
- MCFADDEN, D. (1974): «Conditional Logit Analysis of Qualitative Choice Behavior», en *Frontiers in Econometrics*, editado por P. Zarembka. Nueva York, Academic Press.

- MOLINAS, C. (1986): «A Note on Spurious Regressions with Integrated Moving Average Errors», *Oxford Bulletin of Economics and Statistics*, 48, 279-282.
- MOLINAS, C.; M. SEBASTIÁN, y A. ZABALZA (1991): «La Economía Española. Una Perspectiva Macroeconómica», A. Bosch, editor.
- MUNDLAK, Y. (1961): «On the Pooling of Time Series and Cross-Section Datas», *Econometrica*, 46, 69-85.
- NELSON, C. R. (1973): *Applied Time Series Analysis for Managerial Forecasting*. San Francisco, Holden Day.
- NEWBY, W., y K. WEST (1987): «A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix», *Econometrica*, 55, 703-708.
- NICKELL, S. (1981): «Biases in Dynamic Models with Fixed Effects», *Econometrica*, 49, 1417-1426.
- OSBORN, D. (1990): «A Survey of Seasonality in U.K. Macroeconomic Variables», Discussion Paper, University of Manchester.
- OSBORN, D.; A. P. L. CHUI; J. P. SMITH, y C. R. BIRCHENHALL (1988): «Seasonality and the Order of Integration for Consumption», *Oxford Bulletin of Economics and Statistics*, 50, 361-377.
- PEÑA, D. (1986): *Estadística. Modelos y Métodos*, 2 volúmenes. Alianza Universidad Textos.
- PEÑA SÁNCHEZ DE RIVERA, D. (1992): *Estadística, modelos y métodos*, 5.ª edición, Alianza Editorial, Madrid.
- PHILLIPS, P. (1987): «Time Series Regressions with Unit Roots», *Econometrica*, 55, 277-302.
- RAO, C. R. (1968): «Least Squares Theory Using an Estimated Dispersion Matrix and Its Application to the Measurement of Signals», *Proceedings of the Fifth Berkeley Symposium*, vol. I, 335-372.
- RAO, C. R. (1973): *Linear Statistical Methods and Its Applications*, John Wiley and Sons.
- Renta Nacional de España en 1989 y su distribución provincial* (1992), Banco Bilbao Vizcaya.
- ROTHENBERG, T. J., y C. T. LEENDERS (1964): «Efficient Estimation of Simultaneous Equation Systems», *Econometrica*, 32, 57-76.
- RUBIN, H. (1950): «Consistency of Maximum-Likelihood Estimates in the Explosive Case», en *Statistical Inference in Dynamic Economic Models*, T. C. Koopmans editor, John Wiley, Nueva York.
- SAID, S., y D. DICKEY (1984): «Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order», *Biometrika*, 71, 599-607.
- SALKEVER, D. (1976): «The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals», *Journal of Econometrics*, 4, 393-397.
- SARGAN, J., y A. BHARGAVA (1983): «Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk», *Econometrica*, 51, 153-174.
- SCHMIDT, P. (1976): *Econometrics*. Marcel Dekker, Inc., Nueva York.
- SCHWERT, G. (1987): «Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data», *Journal of Monetary Economics*, 20, 73-103.
- SCHWERT, G. (1989): «Tests for Unit Roots. A Montecarlo Investigation», *Journal of Business and Economic Statistics*, 7, 147-159.
- THEIL, H. (1961): *Economic Forecasts and Policy*. Amsterdam, North Holland, 1961.
- THEIL, H. (1971): *Principles of Econometrics*. Nueva York, John Wiley.
- TOBIN, J. (1958): «Estimation of Relationships for Limited Dependent Variables», *Econometrica*, 26, 24-36.
- TORO VIZCARRONDO, C., y T. D. WALLACE (1968): «A Test of the Mean Square Error

- Criterion for Restrictions in Linear Regression», *Journal of the American Statistical Association*, 63, 558-572.
- WALLACE, T. D., y C. E. TORO VIZCARRONDO (1969): «Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression», *Journal of the American Statistical Association*, 1649-1663.
- WHITE, J. (1958): «The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case», *Annals of Mathematical Statistics*, 29, 1188-1197.
- WHITE, H. (1980): «A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity», *Econometrica*, 48, 817-838.
- WHITE, H. (1982): «Instrumental Variables Regression with Independent Observations», *Econometrica*, 50, 483-499.
- WU, D. (1973): «Alternative Tests of Independence Between Stochastic Regressors and Disturbances», *Econometrica*, 41, 733-750.
- ZELLNER, A., y T. LEE (1965): «Joint Estimation of Relationships Involving Discrete Random Variables», *Econometrica*, 33, 383-394.

- Algoritmo
 - de búsqueda, 398
 - de descenso más rápido, 400
 - de Gauss-Newton, 407, 409, 457
 - de Marquardt, 409
 - de Newton-Raphson, 400, 402, 404
 - en modelos de variable dependiente censurada, 556
 - de «quadratic hill-climbing», 405, 408
 - de «scoring», 406
 - en la estimación del modelo logit, 542
 - en la estimación del modelo probit, 539
 - quasi-Newton, 405
- Análisis de covarianza, 257
- Análisis de intervención, 449
- Análisis de residuos en modelos univariantes, 445
- Aproximación lineal de un modelo lineal, 375
- ARCH, 208
 - ARCH(1) de regresión, 209
 - ARCH(p) de regresión, 209
 - ARCH en media, 210
 - contraste de estructura ARCH, 210
 - GARCH, 211
- Ausencia de sesgo
 - del estimador MCG, 166
 - del estimador MCO, 68
- Autocorrelación, 60, 161, 173, 224
 - autorregresiva de primer orden, 174, 237, 399
 - corrección de Newey-West en la matriz de covarianzas, 228
 - en modelos dinámicos, 309
 - estacional, 232
 - sus causas, 224
 - sus consecuencias, 227
- Bondad de ajuste
 - en el modelo econométrico lineal, 75
 - en modelos de elección discreta, 546
- Cambio de escala en el modelo lineal general, 96
- Cambio de variable en una distribución de probabilidad, 33
- Ciclo, 422
- Coefficiente de amortiguamiento, 422
- Coefficiente de asimetría, 24
- Coefficiente de correlación, 30
 - de rangos, 206
 - lineal en relaciones no lineales, 94
 - parcial, 96
- Coefficiente de curtosis, 24
- Coefficiente de determinación, 75
 - como bondad de ajuste del modelo, 76
 - con matriz de covarianzas no escalar, 168
 - corregido, 77
 - cota superior, 75
 - en el modelo lineal simple, 92
 - en modelos sin término constante, 75
- Cofactor de un elemento de una matriz, 5
- Cointegración, 491
 - estacional, 498
 - superconsistencia del estimador MCO, 492
 - vector de, 491
- Condición de orden para la identificación de una ecuación de un modelo de ecuaciones simultáneas, 582, 583, 587
- Condición de rango para la identificación de una ecuación de un modelo de ecuaciones simultáneas, 582, 583, 588
- Consistencia de un estimador, 38

- Contraste de cambio estructural, 139, 141, 142
 - a partir de errores de predicción, 151
 - de Chow, 139
- Contraste de cointegración, 481, 485, 495
 - de Dickey y Fuller, 481
 - mediante el estadístico Durbin-Watson, 495
- Contraste de exogeneidad de Hausman y Wu, 313
- Contraste de hipótesis, 45
 - en modelos de elección discreta, 547
 - inseguo del, 45
 - nivel de confianza, 45
 - nivel de significación, 45
 - potencia, 45
- Contraste de hipótesis lineales, 115, 180
 - con matriz de covarianzas no escalar, 180
 - en modelos de regresiones aparentemente no relacionadas, 281
 - estadístico F , 115, 119
 - mediante comparación de sumas residuales, 127
 - mediante sustitución de las hipótesis, 137
- Contraste de homogeneidad, 143, 144
 - como contraste de significación, 144
 - mediante variables ficticias, 143
- Contraste de igualdad de varianzas entre submuestras, 206
- Contraste de Normalidad, 81
- Contraste de raíz unitaria, 481
 - de Dickey y Fuller, 481
 - de Dickey y Fuller ampliado, 485
 - estacional, 488, 489, 490
- Contraste de razón de verosimilitudes, 390
 - en modelos logit y probit, 547
- Contraste de restricciones de sobreidentificación, 321, 623
- Contraste de significación, 123, 124
 - excepto término independiente, 123
 - global, 124
- Contraste de tipo Hausman, 313, 314
 - bajo errores de medida, 366
 - contraste de exogeneidad, 313
 - en modelos de datos de panel, 515
- Contraste de validez de instrumentos de Sargan, 321
- Contrastes de autocorrelación
 - Box y Pierce, 233, 447, 459
 - de Breusch y Godfrey, 232
 - de Durbin y Watson, 227
 - gráficos, 232
 - Ljung y Box, 233, 447, 459
- Contrastes de estabilidad, 139
- Contrastes de heteroscedasticidad, 199
 - de Breusch y Pagan, 201
 - de Glesjer, 203
 - de Goldfeld y Quandt, 200
 - de Harvey, 204
 - de Spearman, 206
 - de White, 205
- Convergencia de una sucesión de variables aleatorias
 - en distribución, 40
 - en probabilidad, 37
- Correlación entre los términos de error de distintas ecuaciones, 279, 280
- Correlación espacial, 162
- Correlación espúrea, 199, 479
 - producida por factores de escala, 199
 - producida por tendencias, 479
- Covarianza de variables aleatorias, 30
- Cuasivarianza, 26
- Descomposición canónica de una matriz, 14
- Descomposición de Wald, 327
- Desigualdad de Chebychev, 25
 - para intervalos de confianza de predicciones, 150
- Desviación típica, 24
- Determinante de una matriz, 4
 - como producto de sus valores propios, 14
- Distribución de probabilidad, 18
 - Bernouilli, 20
 - binomial, 20
 - bivariante, 26
 - chi-cuadrado, 34
 - chi-cuadrado no central, 34
 - condicional, 28
 - de Weibull, 549
 - exponencial, 23
 - F de Snedecor, 35
 - geométrica, 21
 - marginal, 27
 - normal, 23
 - normal multivariante, 35
 - Poisson, 21
 - t de Student, 34
 - uniforme, 23
- Distribución de probabilidad de una variable aleatoria, 18
 - condicional, 28
 - de formas cuadráticas de una Normal, 36
 - de tipo continuo, 21
 - de tipo discreto, 18
 - de una función diferenciable de una variable aleatoria, 43
 - de un subvector de una Normal multivariante, 36
 - marginal, 27
 - simétrica, 24
 - truncada, 46
- Ecuación característica de una matriz cuadrada, 12

- Ecuación de un modelo de ecuaciones simultáneas, 482, 608
 - exactamente identificada, 574, 582, 608
 - no identificada, 482, 574, 585
 - posiblemente identificada, 583, 586, 588
 - sobreidentificada, 574, 582, 588, 608
 - subidentificada, 583, 588
- Ecuaciones de Yule-Walker, 419, 422, 443
- Ecuaciones normales, 63, 377
 - del estimador de mínimos cuadrados no lineales, 376, 377
 - del estimador de variables instrumentales, 314
 - del estimador MC2E, 322
 - del estimador MCG, 167
 - del estimador MCO, 63
 - del estimador MCO en secciones cruzadas de series temporales, 256
 - en modelos de sección cruzada de series temporales, 259, 261, 262
- Efectos latentes en modelos de datos de panel, 505, 507
 - su eliminación, 511, 513
- Eficiencia
 - del estimador MCG, 165
 - del estimador MCO cuando la matriz de covarianzas es escalar, 70
- Elección binaria, 530
 - múltiple, 548
- Error cuadrático medio, 152
 - del estimador MCO con variables excluidas, 361
 - de un estimador, 33
 - medio de la predicción, 152
- Error de especificación, 100
- Error de estimación, 68
- Error de predicción, 148, 432
 - bajo errores de medida, 362
 - bajo expectativas racionales, 326
 - en modelos ARIMA, 432
 - su estructura bajo distintos horizontes de previsión, 327, 328
 - sus fuentes, 146
 - varianza del, 148
- Error de tipo I, 45
- Error de tipo II, 45
- Error porcentual medio, 152
- Errores de medida, 362
 - contraste de especificación, 366
 - estimador VI, 365
 - sesgo e inconsistencia del estimador MCO, 363, 364
- Escalón en una variable, 450
- Esperanza condicional, 28
- Esperanza matemática, 23
- Estacionalidad, 431
 - su modelo univariante, 431
- Estacionariedad, 298, 415, 430
 - ausencia de, 477
 - condición en un modelo AR(1), 419
 - condición en un modelo AR(2), 421
 - en un modelo MA(q), 431
- Estadístico
 - CUSUM, 141
 - CUSUMSQ, 142
 - de Chow, 139
 - de Durbin-Watson, 228
 - su sesgo en modelos dinámicos, 230
 - de los multiplicadores de Lagrange, 46
 - en el modelo lineal general, 391
 - para restricciones no lineales, 391
 - de razón de verosimilitudes, 46, 547
 - para restricciones no lineales, 390
 - de Wald, 45, 547
 - para restricciones no lineales, 389
 - de Wallis, 232
 - F para la contrastación de hipótesis lineales, 115, 119
 - para restricciones no lineales, 389
 - su expresión en función de sumas residuales, 128, 135
 - h de Durbin, 232
 - t de significación de un coeficiente, 123
 - U de Theil, 152
- Estimación, 32
- Estimación de la varianza del término de error, 78
 - ausencia de sesgo, 78
 - por máxima verosimilitud, 178
 - por MCG, 168
 - por MCO, 79
- Estimación de modelos ARIMA, 439
- Estimación de un modelo con retardos de las variables exógenas, 304
- Estimación de un modelo dinámico con retardos de la variable endógena, 307
 - bajo autocorrelación, 309
 - sin autocorrelación, 307
- Estimador, 32
 - consistente, 38
 - de mínima varianza, 32
 - eficiente, 32
 - insesgado, 32
- Estimador de Anderson y Hsiao, 517
 - variante de Arellano y Bond, 517
- Estimador de Hildreth y Lu para modelos con autocorrelación AR(1), 399
- Estimador de máxima verosimilitud (MV), 36
 - comparación con el estimador MC3E, 629
 - con autocorrelación, 237, 238
 - con heteroscedasticidad, 197
 - con información completa (MVIC), 630, 631
 - con información limitada (MVIL), 624, 625

- relación con MC2E y MCI en ecuaciones exactamente identificadas, 625
- con matriz de covarianzas escalar, 82
 - equivalencia con MCO, 83
- con matriz de covarianzas no escalar, 179
 - equivalencia con MCG, 180
- con retardos de las variables exógenas, 305
- del modelo de Koyck, 305
- de un modelo con autocorrelación AR(1), 237, 238
- de un modelo dinámico con autocorrelación, 322
- distribución asintótica, 44
- en modelos no lineales, 380
- lineal, 632
- Estimador de mínimos cuadrados condicionados, 442, 458
- Estimador de mínimos cuadrados en dos etapas (MC2E), 320, 613
 - como caso particular del EGM, 333
 - como estimador de variables instrumentales, 613
 - comparación con el estimador MC3E, 630
 - consistencia, 619
 - contraste de hipótesis con el estimador MC3E, 625, 626
 - ecuaciones normales, 615
 - eficiencia, 320
 - en modelos de expectativas racionales, 330
 - equivalencia con el estimador MCI en ecuaciones exactamente identificadas, 623
 - equivalencia con máxima verosimilitud con información completa, 631
 - matriz de covarianzas asintótica, 618
 - relación con el estimador MCI, 622
- Estimador de mínimos cuadrados en tres etapas (MC3E), 625, 626
 - como estimador de variables instrumentales, 625
 - consistencia, 625
 - equivalencia con el estimador MCI en ecuaciones, 629
 - equivalencia con máxima verosimilitud con información limitada, 629
 - exactamente identificadas, 629
 - matriz de covarianzas asintótica, 628
 - relación con el estimador MCI, 625
 - relación con estimador MC2E, 625
- Estimador de mínimos cuadrados generalizados (MCG), 165, 235
 - ausencia de sesgo, 166
 - coeficiente de determinación del, 169
 - condiciones para la equivalencia con MCO, 165
 - de la varianza del término de error, 167
 - del modelo lineal de probabilidad, 534
 - del modelo logit, 541
 - del modelo probit, 538
 - de un modelo con autocorrelación, 176, 178, 235
 - eficiencia, 168, 170
 - en secciones cruzadas de series temporales, 259
 - equivalencia con el estimador de máxima verosimilitud, 178
 - equivalencia con el estimador MV, 240
 - matriz de covarianzas, 166
 - sistema de ecuaciones normales, 166, 180
- Estimador de mínimos cuadrados indirectos (MCI), 603, 606, 614
 - como estimador de variables instrumentales, 613
 - consistencia, 605, 613
 - definición, 603
 - e identificación, 606
 - matriz de covarianzas asintótica, 612
 - sesgo, 604, 606
- Estimador de mínimos cuadrados no lineales (MCNL)
 - algoritmos para el cálculo del, 396
 - distribución asintótica del, 376
 - en modelos de corrección de error, 495
 - en modelos dinámicos con autocorrelación, 323
 - en modelos no lineales, 376, 377
 - no unicidad, 378
 - sesgo, 378
- Estimador de mínimos cuadrados ordinarios (MCO), 63, 65, 276
 - bajo heteroscedasticidad, 197
 - bajo restricciones, 133
 - con matriz de covarianzas no escalar, 161
 - del modelo en desviaciones con respecto a la media, 87
 - eficiencia, 70
 - inconsistencia en modelos dinámicos con autocorrelación, 309
 - independencia entre coeficientes y varianza del término de error, 114
 - ingreso, 68
 - matriz de covarianzas, 163, 169
 - matriz de covarianzas del estimador MCO, 68, 93, 134
 - propiedades con matriz de covarianzas escalar, 67
 - propiedades con matriz de covarianzas no escalar, 162
 - restringido, 133, 362
 - su matriz de covarianzas, 134
 - sin grados de libertad, 67
- Estimador de mínimos cuadrados ponderados, 172, 197, 198, 199, 204
 - en modelos de elección discreta, 535

- Estimador de variables instrumentales (VI),
 - 310, 311, 608
 - consistencia, 611
 - contraste de validez de instrumentos, 321
 - de una ecuación de un sistema, 608
 - de un modelo con retardos de la variable endógena, 307
 - distribución asintótica, 611
 - en modelos con errores de medida, 365
 - en modelos de expectativas racionales, 329
 - matriz de covarianzas asintótica, 611
 - multiplicidad, 318
 - sesgo, 318
- Estimador de variables instrumentales en dos etapas, 320, 518
- Estimador generalizado de momentos (EGM), 330
- Estimador generalizado de variables instrumentales, 331
- Estimador MCO de la forma reducida de un modelo de ecuaciones simultáneas, 603
- ausencia de sesgo, 603
- Estimador MCO de una ecuación de un sistema, 599
 - su sesgo e inconsistencia, 599
- Estimador pretest, 361
- Estimadores en modelos de datos de panel
 - de Anderson y Hsiao, 516
 - de Arellano y Bond, 516
 - de Balestra y Nerlove, 508
 - en primeras diferencias, 513
 - entre grupos (EG), 514
 - generalizado de momentos, 516
 - generalizado de momentos en dos etapas en modelos dinámicos de datos de panel, 516
 - intragrupos (IG), 511
- Expectativas
 - adaptativas, 302, 325
 - estáticas, 302
 - racionales, 326
 - errores de previsión con expectativas racionales, 326
- Factor de amortiguamiento, 422
- Forma estructural, 572
- Forma reducida, 572
- Frecuencias muestrales observadas, 532
- Función característica de una variable aleatoria, 42
- Función de autocorrelación, 416
 - de un ruido blanco, 417
 - estimación, 417
 - parcial, 416, 417, 427
 - simple, 416, 417, 426
 - de un modelo ARMA(1, 1), 428
- Función de autocovarianza, 416
- Función de correlación cruzada, 458
- Función de densidad, 21
- Función de distribución, 18
 - como indicador de decisiones individuales, 536
- Función de respuesta al escalón, 297, 451
 - y estacionariedad, 298
- Función de respuesta al impulso, 297, 451
- Función de verosimilitud
 - concentrada, 238
 - condicional, 239
- Ganancia, 298, 451
 - de Koyck, 304
- Gradiente de una función vectorial, 8
- Hessiano de una función vectorial, 8
- Heteroscedasticidad, 60, 161, 299
 - autorregresiva condicional (véase ARCH) causas, 192
 - corrección de White en la matriz de covarianzas, 196
 - en el modelo lineal de probabilidad, 534
 - en modelos de elección discreta, 531
 - multiplicativa, 203
- Holgura de una restricción, 120, 388
- Homoscedasticidad, 160, 199
- Identificación
 - bajo restricciones entre coeficientes de distintas ecuaciones, 591
 - bajo restricciones lineales no homogéneas, 590
 - bajo restricciones no lineales, 591
 - bajo restricciones sobre la matriz de covarianzas, 593
 - condición de orden, 582, 583, 587
 - condición de rango, 582, 583, 586
 - del modelo de función de transferencia, 454, 455
 - en función del número de instrumentos disponibles, 608
 - exacta, 583
 - global, 591
 - local, 591
 - mediante condiciones de exclusión, 576
 - para restricciones lineales, 586
 - sobreidentificación, 583
- Inclusión de variables irrelevantes, 99
- Independencia de variables aleatorias, 26
- Independencia del estimador MCO de β y σ_u^2 , 114
- Innovación, 328, 437, 456
- Instrumentos, 318
- Interpretación de coeficientes en modelos de elección discreta, 545
- Intervalo de confianza de un estimador, 32, 130

- Inversa por bloques de una matriz, 6
- Invertibilidad
 - de un modelo ARMA(p, q), 431
 - de un modelo MA(q), 431
- Límite en probabilidad, 37
- Logit muestrales, 540
- Longitud de salto, 400
- Matriz, 1
 - cuadrada, 1
 - definida negativa, 14
 - definida positiva, 14
 - descomposición canónica, 14
 - diagonal, 1
 - diagonal a bloques, 1
 - escalar, 1
 - idempotente, 2
 - identidad, 1
 - inversa, 6
 - de una matriz particionada, 7
 - número de condición, 357
 - ortogonal, 6
 - semidefinida negativa, 16
 - semidefinida positiva, 15
 - simétrica, 1
 - singular, 5
 - traspuesta, 1
 - triangular, 1
- Matriz de información, 44
 - del estimador MV en modelos no lineales, 381
 - en un modelo lineal con autocorrelación, 178
- Mediana, 24
- Mode, 24
- Modelo ARIMA(p, d, q), 4
 - ARIMA estacionario, 415, 430
 - ARIMA invertible, 430
 - error de predicción en modelos ARIMA, 435, 436
 - estacional, 431
 - estimación, 435, 440
 - estructura estacional de, 431
 - intervalos de confianza para la predicción, 438
 - predicción en modelos ARIMA, 432, 433, 434
 - su diagnóstico, 444
 - varianza del error de predicción, 435
- Modelo autorregresivo, 419
- Modelo con selección muestral, 558
- Modelo constante, 91, 92
 - estimador MCO, 92
 - su suma explicada, 92
 - su suma residual, 92
 - su suma total, 92
- Modelo de ajuste parcial, 304
- Modelo de alternativas múltiples, 548
- Modelo de corrección de error, 492
 - estimación, 494, 495
 - teorema de representación de Granger, 493
 - y cointegración, 493
- Modelo de datos de panel, 516, 519
 - dinámicos, 507, 516
 - estimador generalizado de momentos, 517
 - con variables predeterminadas, 519
 - inconsistencia del estimador en primeras diferencias, 516
 - inconsistencia de los estimadores IG, EG y MCO, 516
- Modelo de ecuaciones simultáneas, 565
 - estimación, 599
 - forma estructural, 572, 573
 - forma reducida, 572
 - eficiencia del estimador MCG, 572
 - insesgo del estimador MCO, 572
 - función de verosimilitud, 572
 - identificación, 574
 - número de parámetros a estimar, 574
- Modelo de elección discreta, 531, 533
 - distribución no Normal del término de error en el, 531
 - heteroscedasticidad en el, 533
- Modelo de expectativas adaptativas, 301
- Modelo de función de transferencia, 413, 450
 - diagnóstico, 458
 - estimación MV del, 456, 457
 - función de respuesta al escalón, 451
 - función de respuesta al impulso, 451
 - ganancia, 451
 - identificación del, 454, 455
 - tiempo muerto, 451
- Modelo de medias móviles, 423
 - inversión de un modelo, 424
 - representación autorregresiva, 424
- Modelo de regresión lineal, 54
 - características, 55
- Modelo de regresiones aparentemente no relacionadas, 272
 - como caso particular del modelo de ecuaciones simultáneas, 568
 - condiciones para la equivalencia entre los estimadores MCG y MCO, 279, 280
 - contrastación de hipótesis, 282
 - estimador de mínimos cuadrados, 276
 - importancia de las diferencias entre variables explicativas, 281, 282
 - relación con estimador MC3E, 625
- Modelo econométrico, 54
- Modelo econométrico lineal, 52
 - características, 52
 - término de error, 54

- Modelo en desviaciones con respecto a la media, 86
- Modelo lineal de probabilidad, 532
 - estimación MCG, 534
- Modelo lineal simple, 92
 - coeficiente de determinación, 93
 - con dos variables explicativas y una constante, 94
 - matriz de covarianzas del estimador MCO, 93
 - relación entre coeficiente de correlación y MCO, 93
 - sus sumas explicada, residual, total, 93
- Modelo logit, 540
 - aproximación lineal, 540, 560
 - estimación mediante algoritmo de «scoring», 539
 - estimación por mínimos cuadrados, 540
 - función de verosimilitud, 541
 - multinomial, 548
 - sus sumas explicada, residual, total, 546
- Modelo no lineal, 374, 376
 - aproximación lineal, 374
 - no unicidad del estimador de mínimos cuadrados, 376
 - sesgo del estimador MCNL, 376
- Modelo probit, 537, 559
 - aproximación lineal al, 538, 560
 - función de verosimilitud, 538
 - estimación MCG, 538
 - estimación mediante algoritmo «scoring», 539
 - estimación por máxima verosimilitud, 537
- Modelo recursivo, 568
- Modelos anidados, 76
- Modelos dinámicos, 60, 301
 - de retardos distribuidos de Koyck, 304
 - estimación MV, 305
 - sesgo del estimador MCO, 301
 - y autocorrelación, 309
- Modelos univariantes, 60, 413
 - con variables no estacionarias, 429
- Momentos de una distribución, 23
- Muestra de sección cruzada, 53
- Muestra de series temporales, 53
- Multicolinealidad, 60, 344
 - aproximada, 345
 - detección, 354
 - exacta, 143, 344
 - exacta y estimación, 352
 - soluciones a la, 358
 - tamaño de la matriz $X'X$, 357
 - y variables ficticias, 143
- Nivel de confianza de un contraste, 45
- Nivel de significación de un contraste, 45
- Número de condición de una matriz, 357
- Observaciones influyentes, 367
 - errores de predicción con, 367
- Omisión de variables relevantes, 100
- Paseo aleatorio, 414
 - con condición inicial, 416
- Pendientes del modelo de regresión, 54, 252
- Potencia de un contraste, 45
- Predicción con modelos ARIMA, 432, 433, 434
 - de una serie diferenciada, 438
 - de una serie en logaritmos, 438
 - de variables en diferencias, 438
 - de variables en logaritmos, 438
 - error, 435, 436
 - intervalo de confianza, 438
 - varianza del error, 435
- Predicción mínimo-cuadrática, 148
 - bajo autocorrelación, 184
 - bajo heteroscedasticidad, 184
 - bondad de la, 152
 - de la media, 151
 - de un conjunto de valores futuros, 150
 - error de, 148
 - fiabilidad de la, 148
 - intervalo de confianza de la, 148
 - con matriz de covarianzas no escalar, 183
 - mínimo cuadrática, 148
 - varianza del error de, 148
- Probit muestrales, 538
- Proceso estocástico, 414
 - AR(1), 419, 420
 - AR(2), 421, 422
 - ARMA(1, 1), 427
 - estacionario, 415, 431
 - estacionario de segundo orden, 415
 - estacionario en diferencias, 481
 - homogéneo de orden d , 429
 - MA(1), 423, 425
 - MA(q), 425
 - realización de un, 415
- Producto de Kronecker, 8
- Rango de una matriz, 9
 - como número de valores propios no nulos, 17
 - de una matriz idempotente, 15
- Ratio de Mills, 48, 551
- Red de búsqueda
 - para estimación del modelo de Koyck, 306
 - para la estimación de modelos MA(q), 440
 - para la estimación de un modelo con autocorrelación, 237
- Región de confianza, 132
- Regresión cresta, 359
- Regresión espúrea, 480
- Regresión particionada, 85

- Regresiones aparentemente no relacionadas, 274
 - contrastes de hipótesis, 282
 - estimación, 276, 279, 280
- Residuos, 63, 73, 445
 - del estimador de variables instrumentales, 311
 - recursivos, 141
 - recursivos normalizados, 141
- Respuesta
 - a corto plazo, 298
 - a largo plazo, 298
- Retardo
 - de Koyck, 304
 - mediano, 298
 - medio, 298
- Ruido blanco, 414, 417

- Sección cruzada de series temporales
 - con coeficientes variando en el tiempo, 267
 - contraste de homogeneidad entre unidades muestrales, 254
 - estimación con matrices de covarianzas diferentes, 264
 - estimación con matrices de covarianzas iguales, 256
- Sesgo asintótico, 38
- Sesgo de un estimador, 32
- Sistema de ecuaciones normales, 65, 615
 - del estimador de variables instrumentales, 311
 - del estimador MC2E, 615
 - del estimador MCG, 164
 - del estimador MCI, 603
 - del estimador MCO, 65, 87
- Sistema recursivo, 632
 - equivalencia de los estimadores MVIC y MCO, 632
 - variables predeterminadas, 632
- Sobrediferenciación, 447
- Sobreparametrización, 447
- Subespacio nulo de una matriz, 10
- Submatriz, 1
- Suma explicada, 73, 74
- Suma residual, 63, 73, 74
 - del estimador MCG, 169, 188
 - distintas expresiones, 72, 73
 - igual a cero, 66
 - restringida, 139
- Suma total, 73, 74

- Tamaño de un contraste, 45
- Tendencias estocásticas, 478
- Teorema Central del Limite, 42
- Teorema de Gauss-Markov, 70
- Teorema de Mann-Wald, 42
- Teorema de representación de Granger, 493
- Término de error del modelo econométrico, 54
- Término de error fundamental, 322
- Transformación Box-Cox, 207, 430
 - estimación MV, 386
 - estimador de mínimos cuadrados, 385
- Traslación de variables en el modelo econométrico, 96
- Traza de una matriz, 3
 - como suma de los valores propios, 13
 - de una matriz idempotente, 15

- Valor propio de una matriz, 12
 - de la inversa de una matriz, 13
- Valores influyentes, 449
- Variable
 - endógena, 54
 - exógena, 58, 302
 - en modelos de datos de panel, 506
 - explicativa, 54
 - explicativa no observable, 551
 - predeterminada, 59, 302, 565
 - en modelos de datos de panel, 506
 - proxy bajo errores de medida, 362
- Variable aleatoria, 18, 21
 - constante, 19
 - continua, 21
 - discreta, 18
- Variable dependiente censurada, 552
 - estimación de máxima verosimilitud, 535
 - estimación MCG, 554
 - estimador de Fair, 556
- Variable ficticia, 141
 - como indicador de una elección, 530
- escalón, 450
- impulso, 449
- y estacionalidad, 203
- Variable integrada, 478, 487
 - de orden d , 478
 - de orden d, D , 487
- Variables dependientes truncadas, 550
 - estimación de máxima verosimilitud, 551
- Variables instrumentales, 319
 - e identificación, 319
 - en modelos de ecuaciones simultáneas, 608
- Variables omitidas, 100
 - y autocorrelación, 100
 - y sesgo de estimador MCO, 100
- Varianza, 24
- Varianza de una restricción, 119
- Vector, 1
- Vector columna, 1
- Vector fila, 1
- Vector propio de una matriz, 12