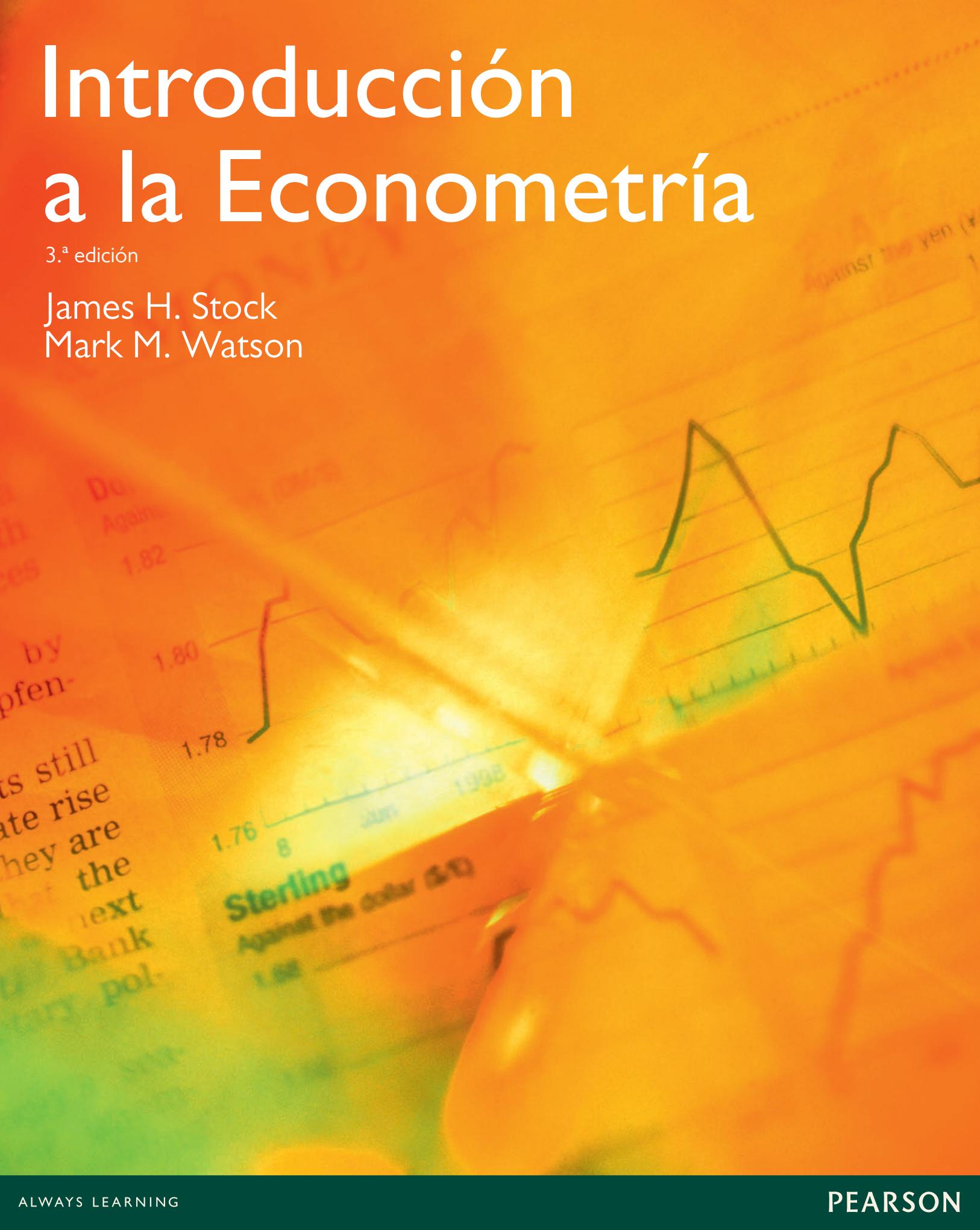


Introducción a la Econometría

3.^a edición

James H. Stock
Mark M. Watson



Introducción a la Econometría

Introducción a la Econometría

3.^a edición

James H. Stock

Harvard University

Mark W. Watson

Princeton University

Traducción

María Arrazola Vacas

Leticia Rodas Alfaya

Universidad Rey Juan Carlos

Traducción, coordinación de la traducción

y revisión técnica

Raúl Sánchez Larrión

Universidad Rey Juan Carlos

PEARSON

Datos de catalogación bibliográfica
Introducción a la Econometría, 3. ^a edición
James H. Stock y Mark W. Watson
PEARSON EDUCACIÓN, S.A., Madrid, 2012
ISBN: 9788483229675
Materia: 33. Economía
Formato: 215 × 270 mm
Páginas: 600

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser utilizada con la autorización de sus titulares, salvo excepción prevista por la ley. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sgts. Código penal).

Diríjase a CEDRO (Centro Español de Derechos Reprográficos —www.cedro.org), si necesita fotocopiar o escanear algún fragmento de esta obra.

Todos los derechos reservados.

© 2012 PEARSON EDUCACIÓN, S.A.
C/ Ribera del Loira, 28
28042 Madrid (España)

Authorized translation from the English language edition, entitled INTRODUCTION TO ECONOMETRICS, 3rd Edition by JAMES H. STOCK; MARK WATSON, published by Pearson Education, Inc, publishing as Prentice Hall, Copyright © 2011.

All rights reserved. No part of this book may be reproduced or transmitted in any form or any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

SPANISH language edition published by Pearson Educación, S.A., Copyright © 2012.

ISBN: 978-84-8322-967-5
Depósito Legal: M-10280-2012

Equipo de edición:

Editor: Alberto Cañizal
Técnico editorial: María Varela

Diseñadora: Elena Jaramillo

Equipo de producción:

Directora: Marta Illescas
Coordinadora: Tini Cardoso

Diseño de cubierta: Copobook, S.L.

Composición: Copobook, S.L.

Impreso por:

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

Nota sobre enlaces a páginas web ajena: este libro incluye enlaces a sitios web cuya gestión, mantenimiento y control son responsabilidad única y exclusiva de terceros ajenos a PEARSON EDUCACIÓN, S.A. Los enlaces u otras referencias a sitios web se incluyen con finalidad estrictamente informativa y se proporcionan en el estado en que se encuentran en el momento de publicación sin garantías, expresas o implícitas, sobre la información que se proporcione en ellas. Los enlaces no implican el aval de PEARSON EDUCACIÓN S.A. a tales sitios, páginas web, funcionalidades y sus respectivos contenidos o cualquier asociación con sus administradores. En consecuencia, PEARSON EDUCACIÓN S.A., no asume responsabilidad alguna por los daños que se puedan derivar de hipotéticas infracciones de los derechos de propiedad intelectual y/o industrial que puedan contener dichos sitios web ni por las pérdidas, delitos o los daños y perjuicios derivados, directa o indirectamente, del uso de tales sitios web y de su información. Al acceder a tales enlaces externos de los sitios web, el usuario estará bajo la protección de datos y políticas de privacidad o prácticas y otros contenidos de tales sitios web y no de PEARSON EDUCACIÓN S.A.

Este libro ha sido impreso con papel y tintas ecológicos

Contenido abreviado

PARTE I Introducción y repaso

CAPÍTULO 1	Cuestiones económicas y datos	1
CAPÍTULO 2	Repaso de probabilidad	11
CAPÍTULO 3	Repaso de estadística	47

PARTE II Los fundamentos del análisis de regresión

CAPÍTULO 4	Regresión lineal con regresor único	77
CAPÍTULO 5	Regresión con regresor único: contrastes de hipótesis e intervalos de confianza	103
CAPÍTULO 6	Regresión lineal con varios regresores	129
CAPÍTULO 7	Contrastes de hipótesis e intervalos de confianza en regresión múltiple	153
CAPÍTULO 8	Funciones de regresión no lineales	181
CAPÍTULO 9	Evaluación de estudios basados en regresión múltiple	223

PARTE III Otros temas relacionados con el análisis de regresión

CAPÍTULO 10	Regresión con datos de panel	249
CAPÍTULO 11	Regresión con variable dependiente binaria	275
CAPÍTULO 12	Regresión con variables instrumentales	303
CAPÍTULO 13	Experimentos y cuasi experimentos	339

PARTE IV Análisis de regresión con datos de series temporales económicas

CAPÍTULO 14	Introducción a la regresión de series temporales y predicción	373
CAPÍTULO 15	Estimación de efectos causales dinámicos	421
CAPÍTULO 16	Otros temas relacionados con la regresión en series temporales	455

PARTE V Teoría econométrica del análisis de regresión

CAPÍTULO 17	Teoría de regresión lineal con regresor único	483
CAPÍTULO 18	Teoría de regresión múltiple	503

Contenido

Prefacio	XXV
----------------	-----

PARTE I Introducción y revisión

CAPÍTULO 1	Cuestiones económicas y datos	1
1.1	Preguntas económicas a examen	1
	Pregunta #1 ¿Mejora la reducción del tamaño de las clases la educación en la escuela primaria?	1
	Pregunta #2 ¿Existe discriminación racial en el mercado de préstamos para la vivienda?	2
	Pregunta #3 ¿Cuánto reduce el tabaquismo los impuestos sobre los cigarrillos? ..	3
	Pregunta #4 ¿Cuál será la tasa de inflación del próximo año?	3
	Preguntas cuantitativas, respuestas cuantitativas	4
1.2	Efectos causales y experimentos ideales	4
	Estimación de los efectos causales	4
	Predicción y causalidad	5
1.3	Datos: fuentes y tipos	5
	Datos experimentales <i>versus</i> datos observacionales	5
	Datos de sección cruzada	6
	Datos de series temporales	6
	Datos de panel	7
CAPÍTULO 2	Repaso de probabilidad	11
2.1	Variables aleatorias y distribuciones de probabilidad	11
	Probabilidades, espacio muestral y variables aleatorias	11
	Distribución de probabilidad de una variable aleatoria discreta	12
	Distribución de probabilidad de una variable aleatoria continua	13
2.2	Esperanza, media y varianza	15
	La esperanza de una variable aleatoria	15
	La desviación típica y la varianza	16
	Media y varianza de una función lineal de variables aleatorias	16
	Otras medidas de forma de una distribución	17

2.3	Dos variables aleatorias	19
	Distribuciones conjunta y marginal	19
	Distribuciones condicionales	20
	Independencia	22
	Covarianza y correlación	23
	La media y la varianza de la suma de variables aleatorias	25
2.4	Las distribuciones normal, chi cuadrado, <i>t</i> de Student y <i>F</i>	26
	La distribución normal	26
	La distribución chi-cuadrado	28
	La distribución <i>t</i> de Student	28
	La distribución <i>F</i>	30
2.5	Muestreo aleatorio y distribución de la media muestral	31
	Muestreo aleatorio	31
	La distribución muestral de la media muestral	32
2.6	Aproximación para muestras grandes de las distribuciones muestrales	34
	La ley de los grandes números y la consistencia	34
	El teorema central del límite	36
	APÉNDICE 2.1 Obtención de los resultados del Concepto clave 2.3	45
CAPÍTULO 3	Repaso de estadística	47
3.1	Estimación de la media poblacional	48
	Los estimadores y sus propiedades	48
	Propiedades de \bar{Y}	49
	La importancia del muestreo aleatorio	50
3.2	Contrastes de hipótesis sobre la media poblacional	51
	Hipótesis nula y alternativa	51
	El <i>p</i> -valor	51
	Cálculo del <i>p</i> -valor con σ_Y^2 conocido	52
	La varianza muestral, la desviación típica muestral y el error estándar	53
	Cálculo del <i>p</i> -valor con σ_Y desconocido	54
	El estadístico <i>t</i>	54
	Contrastes de hipótesis con nivel de significación preestablecido	55
	Alternativas unilaterales	57
3.3	Intervalos de confianza para la media poblacional	57
3.4	Comparación de medias de diferentes poblaciones	58
	Contraste de hipótesis para la diferencia entre dos medias	58
	Intervalos de confianza para la diferencia entre dos medias poblacionales	59
3.5	Estimación de la diferencia de medias de los efectos causales mediante datos experimentales	60
	Los efectos causales como diferencia de las esperanzas condicionales	60
	Estimación de los efectos causales mediante las diferencias de medias	60
3.6	Utilización del estadístico <i>t</i> cuando el tamaño muestral es pequeño ..	62
	El estadístico <i>t</i> y la distribución <i>t</i> de Student	62
	La utilización de la distribución <i>t</i> de Student en la práctica	63

3.7	Diagramas de dispersión, covarianza muestral y correlación muestral	65
	Diagramas de dispersión	65
	Covarianza muestral y correlación	66
	APÉNDICE 3.1 La encuesta actualizada de población de EE.UU. (CPS)	74
	APÉNDICE 3.2 Dos pruebas de que \bar{Y} es el estimador de mínimos cuadrados de μ_Y	74
	APÉNDICE 3.3 Una prueba de que la varianza muestral es consistente	75
PARTE II	Los fundamentos del análisis de regresión	
CAPÍTULO 4	Regresión lineal con regresor único	77
4.1	El modelo de regresión lineal	77
4.2	Estimación de los coeficientes del modelo de regresión lineal	80
	El estimador de mínimos cuadrados ordinarios	82
	Estimaciones MCO de la relación entre calificaciones en los exámenes y ratio estudiantes-maestros	83
	¿Por qué utilizar el estimador MCO?	84
4.3	Medidas de ajuste	85
	El R^2	85
	El error estándar de la regresión	86
	Aplicación a los datos de las calificaciones en los exámenes	87
4.4	Los supuestos de mínimos cuadrados	87
	Supuesto #1: La distribución condicional de u_i dado X_i tiene media igual a cero ..	87
	Supuesto #2: (X_i, Y_i) , $i = 1, \dots, n$, son independientes e idénticamente distribuidas..	89
	Supuesto #3: Los datos atípicos elevados son improbables	90
	La utilización de los supuestos de mínimos cuadrados	91
4.5	Distribución muestral de los estimadores MCO	91
	La distribución muestral de los estimadores MCO	91
4.6	Conclusión	93
	APÉNDICE 4.1 La base de datos de las calificaciones en el examen de California ..	99
	APÉNDICE 4.2 Obtención de los estimadores MCO	99
	APÉNDICE 4.3 Distribución muestral del estimador MCO	100
CAPÍTULO 5	Regresión con regresor único: contrastes de hipótesis e intervalos de confianza	103
5.1	Contraste de hipótesis acerca de uno de los coeficientes de regresión 103	
	Hipótesis bilaterales acerca de β_1	104
	Hipótesis unilaterales sobre β_1	106
	Contraste de hipótesis acerca del término independiente β_0	107
5.2	Intervalos de confianza para un coeficiente de regresión	108
5.3	Regresión cuando X es una variable binaria	109
	Interpretación de los coeficientes de regresión	109
5.4	Heterocedasticidad y homocedasticidad	111
	¿Qué es la heterocedasticidad y la homocedasticidad?	111

	Implicaciones matemáticas de la homocedasticidad	112
	¿Qué significa esto en la práctica?	113
5.5	Fundamentos teóricos de mínimos cuadrados ordinarios	115
	Estimadores lineales condicionalmente insesgados y teorema de Gauss-Markov ...	115
	Estimadores de regresión alternativos a MCO	116
5.6	La utilización del estadístico t en regresión para muestras pequeñas	117
	El estadístico t y la distribución t de Student	117
	La utilización de la distribución t de Student en la práctica	118
5.7	Conclusión	118
	APÉNDICE 5.1 Fórmulas de los errores estándar MCO	124
	APÉNDICE 5.2 Las condiciones de Gauss-Markov y la demostración del teorema de Gauss-Markov	125
CAPÍTULO 6	Regresión lineal con varios regresores	129
6.1	Sesgo de variable omitida	129
	Definición del sesgo de variable omitida	130
	Fórmula del sesgo de variable omitida	131
	Solución del sesgo de variable omitida mediante la división de los datos en grupos	132
6.2	El modelo de regresión múltiple	134
	La recta de regresión poblacional	134
	El modelo de regresión múltiple poblacional	135
6.3	El estimador MCO en regresión múltiple	136
	El estimador MCO	136
	Aplicación a las calificaciones en los exámenes y la ratio estudiantes-maestros	137
6.4	Medidas de ajuste en regresión múltiple	139
	El error estándar de la regresión	139
	El R^2	139
	El « R^2 ajustado»	140
	Aplicación a las calificaciones en los exámenes	140
6.5	Los supuestos de mínimos cuadrados en regresión múltiple	141
	Supuesto #1: La distribución condicional de u_i dados $X_{1i}, X_{2i}, \dots, X_{ki}$ tiene media igual a cero	141
	Supuesto #2: $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i), i = 1, \dots, n$, son i.i.d.	141
	Supuesto #3: Los valores atípicos elevados son improbables	141
	Supuesto #4: Ausencia de multicolinealidad perfecta	141
6.6	La distribución de los estimadores MCO en regresión múltiple	142
6.7	Multicolinealidad	143
	Ejemplos de multicolinealidad perfecta	143
	Multicolinealidad imperfecta	145
6.8	Conclusión	146
	APÉNDICE 6.1 Obtención de la Ecuación (6.1)	151
	APÉNDICE 6.2 Distribución de los estimadores MCO en presencia de dos regresores y errores homocedásticos	152
	APÉNDICE 6.3 El teorema de Frisch-Waugh	152

CAPÍTULO 7	Contrastes de hipótesis e intervalos de confianza en regresión múltiple	153
7.1	Contrastes de hipótesis e intervalos de confianza para un único coeficiente	153
	Errores estándar de los estimadores MCO	153
	Contrastes de hipótesis para un único coeficiente	154
	Intervalos de confianza para un único coeficiente	155
	Aplicación a las calificaciones y la ratio estudiantes-maestros	155
7.2	Contraste de hipótesis conjuntas	156
	Contraste de hipótesis acerca de dos o más coeficientes	156
	El estadístico F	158
	Aplicación a las calificaciones y la ratio estudiantes-maestros	159
	El estadístico F válido con homocedasticidad	159
7.3	Contraste de una sola restricción sobre varios coeficientes	161
7.4	Conjuntos de confianza para varios coeficientes	162
7.5	Especificación del modelo en regresión múltiple	163
	Sesgo de variable omitida en regresión múltiple	164
	El papel de las variables de control en regresión múltiple	164
	La especificación del modelo en teoría y en la práctica	166
	Interpretación del R^2 y del R^2 ajustado en la práctica	166
7.6	Análisis de la base de datos de las calificaciones en los exámenes	167
7.7	Conclusión	171
	APÉNDICE 7.1 El contraste de hipótesis conjunta de Bonferroni	177
	APÉNDICE 7.2 Independencia en media condicional	178
CAPÍTULO 8	Funciones de regresión no lineales	181
8.1	Estrategia general para la modelización de funciones de regresión no lineales	182
	Calificaciones y renta del distrito	182
	El efecto sobre Y de un cambio en X con especificaciones no lineales	185
	Metodología general para la modelización no lineal mediante regresión múltiple	187
8.2	Funciones no lineales de una sola variable independiente	188
	Polinomios	188
	Logaritmos	190
	Modelos polinomiales y logarítmicos para calificaciones y renta del distrito	195
8.3	Interacciones entre variables independientes	196
	Interacciones entre dos variables binarias	197
	Interacciones entre una variable continua y una variable binaria	198
	Interacciones entre dos variables continuas	201
8.4	Efectos no lineales sobre las calificaciones de la ratio estudiantes-maestros	204
	Discusión de los resultados de la regresión	207
	Resumen de resultados	209
8.5	Conclusión	210
	APÉNDICE 8.1 Funciones de regresión que son no lineales en los parámetros	219
	APÉNDICE 8.2 Pendientes y elasticidades de funciones de regresión no lineales	221

CAPÍTULO 9 Evaluación de estudios basados en regresión múltiple	223
9.1 Validez interna y externa	223
Amenazas a la validez interna	224
Amenazas a la validez externa	224
9.2 Amenazas a la validez interna del análisis de regresión múltiple	225
Sesgo de variable omitida	226
Error de especificación de la forma funcional de la función de regresión	227
Sesgo de errores de medida y por errores en las variables	228
Datos perdidos y selección muestral	230
Causalidad simultánea	232
Origen de la inconsistencia de los errores estándar MCO	233
9.3 Validez interna y externa cuando la regresión se utiliza para predicción	234
Utilización de modelos de regresión para predicción	234
Evaluación de la validez de los modelos de regresión para predicción	235
9.4 Ejemplo: calificaciones y tamaño de las clases	235
Validez externa	235
Validez interna	240
Debate e implicaciones	242
9.5 Conclusión	242
APÉNDICE 9.1 Los datos de las calificaciones en las pruebas de educación primaria de Massachusetts	247

PARTE III **Otros temas relacionados con el análisis de regresión**

CAPÍTULO 10 Regresión con datos de panel	249
10.1 Datos de panel	249
Ejemplo: mortalidad en accidentes de tráfico e impuestos sobre el alcohol	250
10.2 Datos de panel con dos períodos temporales: comparaciones «antes y después»	252
10.3 Regresión de efectos fijos	254
El modelo de regresión de efectos fijos	254
Estimación e inferencia	255
Aplicación a la mortalidad en accidentes de tráfico	257
10.4 Regresión con efectos fijos temporales	257
Solamente efectos temporales	258
Efectos fijos individuales y temporales	258
10.5 Los supuestos de la regresión de efectos fijos y los errores estándar de la regresión de efectos fijos	260
Los supuestos de la regresión de efectos fijos	260
Errores estándar de la regresión de efectos fijos	261
10.6 Las leyes sobre conducción bajo los efectos del alcohol y la mortalidad por accidentes de tráfico	262

10.7 Conclusión	265
APÉNDICE 10.1 La base de datos estatales sobre mortalidad en accidentes de tráfico	270
APÉNDICE 10.2 Errores estándar de la regresión de efectos fijos	270
CAPÍTULO 11 Regresión con variable dependiente binaria	275
11.1 Variables dependientes binarias y modelo de probabilidad lineal	276
Variables dependientes binarias	276
El modelo de probabilidad lineal	277
11.2 Regresión probit y logit	279
Regresión probit	279
Regresión logit	283
Comparativa de los modelos de probabilidad lineal, probit y logit	284
11.3 Estimación e inferencia en los modelos logit y probit	284
Estimación por mínimos cuadrados no lineales	285
Estimación máximo verosímil	285
Medidas de ajuste	286
11.4 Aplicación a los datos HMDA de Boston	287
11.5 Conclusión	292
APÉNDICE 11.1 La base de datos HMDA de Boston	298
APÉNDICE 11.2 Estimación máximo verosímil	298
APÉNDICE 11.3 Otros modelos de variable dependiente limitada	300
CAPÍTULO 12 Regresión con variables instrumentales	303
12.1 El estimador VI con regresor único e instrumento único	303
El modelo VI y los supuestos	304
El estimador de mínimos cuadrados en dos etapas	304
¿Por qué funciona la regresión VI?	305
La distribución muestral del estimador MC2E	308
Aplicación a la demanda de cigarrillos	309
12.2 El modelo general de regresión VI	311
MC2E en el modelo general VI	312
Relevancia y exogeneidad de los instrumentos en el modelo general VI	313
Los supuestos de la regresión VI y la distribución muestral del estimador MC2E	313
Inferencia mediante el estimador MC2E	314
Aplicación a la demanda de cigarrillos	315
12.3 Verificación de la validez de los instrumentos	316
Supuesto #1: relevancia de los instrumentos	316
Supuesto #2: exogeneidad de los instrumentos	318
12.4 Aplicación a la demanda de cigarrillos	320
12.5 ¿De dónde provienen los instrumentos válidos?	324
Tres ejemplos	324
12.6 Conclusión	328
APÉNDICE 12.1 La base de datos de panel sobre consumo de cigarrillos	332
APÉNDICE 12.2 Obtención de la fórmula del estimador MC2 de la Ecuación (12.4)	332
APÉNDICE 12.3 Distribución del estimador MC2E para grandes muestras	333

APÉNDICE 12.4 La distribución del estimador MC2E para muestras grandes cuando el instrumento no es válido	334
APÉNDICE 12.5 Análisis de variables instrumentales con instrumentos débiles	335
APÉNDICE 12.6 MC2E con variables de control	336
CAPÍTULO 13 Experimentos y quasi experimentos	339
13.1 Variables respuesta, efectos causales y experimentos ideales	340
Variable respuesta y efecto causal promedio	340
Modelos econométricos para el análisis de datos experimentales	341
13.2 Amenazas a la validez de los experimentos	342
Amenazas a la validez interna	342
Amenazas a la validez externa	344
13.3 Estimaciones experimentales del efecto de la reducción del tamaño de las clases	345
Diseño experimental	346
Análisis de los datos STAR	346
Comparación de las estimaciones observacionales y experimentales de los efectos del tamaño de las clases	351
13.4 Cuasi experimentos	352
Ejemplos	352
El estimador de diferencias en diferencias	354
Estimadores de variables instrumentales	357
Estimadores de la regresión con discontinuidad	357
13.5 Problemas potenciales en cuasi experimentos	358
Amenazas a la validez interna	358
Amenazas a la validez externa	360
13.6 Estimaciones experimentales y cuasi experimentales en poblaciones heterogéneas	360
MCO con efectos causales heterogéneos	361
Regresión VI con efectos causales heterogéneos	361
13.7 Conclusión	363
APÉNDICE 13.1 La base de datos del proyecto STAR	370
APÉNDICE 13.2 Estimación VI con efectos causales que varían entre individuos	370
APÉNDICE 13.3 El marco de las variables respuesta para el análisis de datos procedentes de experimentos	371

PARTE IV Análisis de regresión con datos de series temporales económicas

CAPÍTULO 14 Introducción a la regresión de series temporales y predicción	373
14.1 Utilización de los modelos de regresión para predicción	374
14.2 Introducción a los datos de series temporales y correlación serial	375
Las tasas de inflación y desempleo en Estados Unidos	375
Retardos, primeras diferencias, logaritmos y tasas de crecimiento	375

Autocorrelación	378
Otros ejemplos de series temporales económicas	379
14.3 Modelos autorregresivos	380
El modelo autorregresivo de primer orden	381
El modelo autorregresivo de orden p	382
14.4 Regresión de series temporales con predictores adicionales y modelo autorregresivo de retardos distribuidos	385
Predicción de la variación de la tasa de inflación mediante los valores pasados de la tasa de desempleo	385
Estacionariedad	387
Regresión de series temporales con varios predictores	387
Incertidumbre de la predicción e intervalos de predicción	389
14.5 Selección de la longitud de los retardos mediante criterios de información	392
Determinación del orden de una autorregresión	392
Selección de la longitud de los retardos en una regresión de series temporales con varios predictores	394
14.6 Ausencia de estacionariedad I: tendencias	395
¿Qué es una tendencia?	395
Problemas ocasionados por las tendencias estocásticas	397
Detección de tendencias estocásticas: contraste de raíz unitaria AR	398
Resolución de los problemas originados por tendencias estocásticas	401
14.7 Ausencia de estacionariedad II: cambios estructurales	402
¿Qué es un cambio estructural?	402
Contrastes de cambio estructural	403
Predicción pseudo fuera de la muestra	407
Resolución de los problemas originados por cambios estructurales	411
14.8 Conclusión	411
APÉNDICE 14.1 Los datos de series temporales utilizados en el Capítulo 14	418
APÉNDICE 14.2 Estacionariedad en el modelo AR(1)	418
APÉNDICE 14.3 Notación del operador de retardos	419
APÉNDICE 14.4 Modelos ARMA	419
APÉNDICE 14.5 Consistencia del estimador de la longitud de los retardos BIC	420
CAPÍTULO 15 Estimación de efectos causales dinámicos	421
15.1 Un «primer gusto en boca» de los datos del zumo de naranja	422
15.2 Efectos causales dinámicos	424
Efectos causales y datos de series temporales	424
Dos tipos de exogeneidad	426
15.3 Estimación de efectos causales dinámicos con regresores exógenos	427
Los supuestos del modelo de retardos distribuidos	428
u_t autocorrelacionados, errores estándar e inferencia	428
Multiplicadores dinámicos y multiplicadores dinámicos acumulativos	429
15.4 Errores estándar consistentes en presencia de heterocedasticidad y autocorrelación	430

Distribución del estimador MCO con errores autocorrelacionados	430
Errores estándar HAC	431
15.5 Estimación de efectos causales dinámicos con regresores estrictamente exógenos	434
El modelo de retardos distribuidos con errores AR(1)	434
Estimación MCO del modelo ARD	436
Estimación MCG	437
El modelo de retardos distribuidos con retardos adicionales y errores AR(p)	438
15.6 Los precios del zumo de naranja y el frío	440
15.7 ¿Es creíble la exogeneidad?: algunos ejemplos	446
La renta de EE.UU. las exportaciones australianas	446
Los precios del petróleo y la inflación	446
La política monetaria y la inflación	447
La curva de Phillips	448
15.8 Conclusión	448
APÉNDICE 15.1 La base de datos del zumo de naranja	453
APÉNDICE 15.2 Modelo ARD y mínimos cuadrados generalizados en notación del operador de retardos	453
CAPÍTULO 16 Otros temas relacionados con la regresión en series temporales	455
16.1 Vectores autorregresivos	455
El Modelo VAR	456
Un modelo VAR para las tasas de inflación y desempleo	458
16.2 Predicciones multiperiodo	458
Predicciones multiperiodo iteradas	459
Predicciones multiperiodo directas	460
¿Qué método debe utilizarse?	462
16.3 Órdenes de integración y contraste DF-MCG de raíces unitarias	463
Otros modelos de tendencias y órdenes de integración	463
El contraste DF-MCG de raíces unitarias	464
¿Por qué los contrastes de raíz unitaria tienen distribuciones no normales?	467
16.4 Cointegración	468
Cointegración y corrección de error	468
¿Cómo se puede saber si dos variables aleatorias están cointegradas?	469
Estimación de los coeficientes de cointegración	471
Extensión a varias variables cointegradas	472
Aplicación a los tipos de interés	473
16.5 Volatilidad agrupada y heterocedasticidad condicional autorregresiva	475
Volatilidad agrupada	475
Heterocedasticidad condicional autorregresiva	476
Aplicación a la volatilidad de las cotizaciones de valores	477
16.6 Conclusión	477
APÉNDICE 16.1 Datos financieros de EE.UU. utilizados en el Capítulo 16	482

PARTE V Teoría econométrica del análisis de regresión

CAPÍTULO 17 Teoría de regresión lineal con regresor único	483
17.1 Los supuestos ampliados de mínimos cuadrados y el estimador MCO	483
Los supuestos ampliados de mínimos cuadrados	484
El estimador MCO	484
17.2 Fundamentos de teoría de distribución asintótica	485
La convergencia en probabilidad y la ley de los grandes números	485
El teorema central del límite y la convergencia en distribución	487
El teorema de Slutsky y el teorema de la función continua	488
Aplicación al estadístico t basado en la media muestral	488
17.3 Distribución asintótica del estimador MCO y del estadístico t	489
Consistencia y normalidad asintótica de los estimadores MCO	489
Consistencia de los errores estándar heterocedástico-robustos	489
Normalidad asintótica del estadístico t heterocedástico-robusto	490
17.4 Distribuciones muestrales exactas con errores normalmente distribuidos	490
Distribución de $\hat{\beta}_1$ con errores normales	491
Distribución del estadístico t válido con homocedasticidad	492
17.5 Mínimos cuadrados ponderados	492
MCP con heterocedasticidad conocida	493
MCP con heterocedasticidad de forma funcional conocida	493
¿Errores estándar heterocedástico-robustos o MCP?	495
APÉNDICE 17.1 La distribución normal y sus afines y los momentos de las variables aleatorias continuas	499
APÉNDICE 17.2 Dos desigualdades	501
CAPÍTULO 18 Teoría de regresión múltiple	503
18.1 El modelo lineal de regresión múltiple y el estimador MCO en forma matricial	504
El modelo de regresión múltiple en forma matricial	504
Los supuestos ampliados de mínimos cuadrados	504
El estimador MCO	506
18.2 Distribución asintótica del estimador MCO y del estadístico t	506
El teorema central del límite multivariante	507
Normalidad asintótica de $\hat{\beta}$	507
Errores estándar heterocedástico-robustos	508
Intervalos de confianza para los efectos previstos	508
Distribución asintótica del estadístico t	508
18.3 Contrastes de hipótesis conjuntas	509
Hipótesis conjuntas en notación matricial	509
Distribución asintótica del estadístico F	509
Conjuntos de confianza para varios coeficientes	510
18.4 Distribución de los estadísticos de regresión con errores normales ..	510
Representación matricial de los estadísticos de regresión MCO	510

Distribución de $\hat{\beta}$ con errores normales	511
Distribución de $s_{\hat{u}}^2$	511
Errores estándar válidos con homocedasticidad	512
Distribución del estadístico t	512
Distribución del estadístico F	512
18.5 Eficiencia del estimador MCO con errores homocedásticos	513
Las condiciones de Gauss-Markov para regresión múltiple	513
Estimadores lineales condicionalmente insesgados	513
El teorema de Gauss-Markov para regresión múltiple	513
18.6 Mínimos cuadrados generalizados	514
Los supuestos de MCG	515
MCG con Ω conocida	516
MCG cuando Ω contiene parámetros desconocidos	517
El supuesto de media condicional igual a cero y MCG	517
18.7 Variables instrumentales y estimación por el método generalizado de momentos	518
El estimador VI en forma matricial	519
Distribución asintótica del estimador MC2E	519
Propiedades de MC2E con errores homocedásticos	520
Estimación por el método generalizado de momentos en modelos lineales	522
APÉNDICE 18.1 Resumen de álgebra matricial	530
APÉNDICE 18.2 Distribuciones multivariantes	532
APÉNDICE 18.3 Obtención de la distribución asintótica de $\hat{\beta}$	533
APÉNDICE 18.4 Obtención de las distribuciones exactas de los estadísticos de contraste MCO con errores normales	534
APÉNDICE 18.5 Prueba del teorema de Gauss-Markov para regresión múltiple	535
APÉNDICE 18.6 Pruebas de algunos resultados seleccionados de la estimación VI y MGM	535
<i>Apéndice</i>	539
<i>Bibliografía</i>	547
<i>Glosario</i>	551
<i>Índice</i>	559

Conceptos clave

PARTE I Introducción y repaso

1.1	Datos de sección cruzada, series temporales y panel	9
2.1	Esperanza y media	15
2.2	Varianza y desviación típica	16
2.3	Medias, varianzas y covarianzas de la suma de variables aleatorias	25
2.4	Cálculo de probabilidades con variables aleatorias normales	27
2.5	Muestreo aleatorio simple y variables aleatorias i.i.d.	32
2.6	Convergencia en probabilidad consistencia, y ley de los grandes números	36
2.7	El teorema central del límite	37
3.1	Estimadores y estimaciones	48
3.2	Sesgo, consistencia y eficiencia	49
3.3	Eficiencia de \bar{Y} : \bar{Y} es ELIO	49
3.4	El error estándar de \bar{Y}	53
3.5	La terminología del contraste de hipótesis	56
3.6	El contraste de la hipótesis $E(Y) = \mu_{Y,0}$ frente a la alternativa $E(Y) \neq \mu_{Y,0}$	56
3.7	Intervalos de confianza para la media poblacional	58

PARTE II Los fundamentos del análisis de regresión

4.1	Terminología del modelo de regresión lineal con regresor único	80
4.2	El estimador MCO, valores estimados y residuos	83
4.3	Los supuestos de mínimos cuadrados	91
4.4	Distribuciones para grandes muestras de $\hat{\beta}_0$ y $\hat{\beta}_1$	92
5.1	Forma general del estadístico t	104
5.2	Contraste de la hipótesis $\beta_1 = \beta_{1,0}$ frente a la alternativa $\beta_1 \neq \beta_{1,0}$	105
5.3	Intervalo de confianza para β_1	108
5.4	Heterocedasticidad y homocedasticidad	112
5.5	El teorema de Gauss Markov para $\hat{\beta}_1$	116
6.1	Sesgo de variable omitida en la regresión con un único regresor	131
6.2	El modelo de regresión múltiple	136
6.3	Los estimadores MCO, valores de predicción y residuos en el modelo de regresión múltiple	137
6.4	Los supuestos de mínimos cuadrados en el modelo de regresión múltiple	142
6.5	Distribución para muestras grandes de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$	143

7.1	Contraste de la hipótesis $\beta_j = \beta_{j,0}$ frente a la alternativa $\beta_j \neq \beta_{j,0}$	154
7.2	Intervalos de confianza para un único coeficiente en regresión múltiple	155
7.3	Sesgo de variable omitida en regresión múltiple	164
7.4	R^2 y \bar{R}^2 qué nos dicen y qué no	167
8.1	El efecto esperado en Y de un cambio en X_1 en el modelo de regresión no lineal (8.3)	185
8.2	Logaritmos en la regresión: tres casos	194
8.3	Un método para la interpretación de los coeficientes en regresiones con variables binarias	198
8.4	Interacciones entre variables binarias y continuas	200
8.5	Interacciones en regresión múltiple	203
9.1	Validez interna y externa	224
9.2	Sesgo de variable omitida: ¿deberían incluirse más variables en la regresión?	227
9.3	Error de especificación de la forma funcional	228
9.4	Sesgo por errores en las variables	229
9.5	Sesgo de selección muestral	231
9.6	Sesgo por causalidad simultánea	233
9.7	Amenazas a la validez interna de un estudio de regresión múltiple	234

PARTE III Otros temas relacionados con el análisis de regresión

10.1	Notación para datos de panel	250
10.2	El modelo de regresión de efectos fijos	256
10.3	Los supuestos de la regresión de efectos fijos	260
11.1	El modelo de probabilidad lineal	278
11.2	El modelo probit, probabilidades estimadas y efectos estimados	281
11.3	Regresión logit	283
12.1	El modelo general de regresión de variables instrumentales y su terminología	312
12.2	Mínimos cuadrados en dos etapas	313
12.3	Las dos condiciones para la validez de los instrumentos	314
12.4	Los supuestos de la regresión VI	314
12.5	Una regla práctica para la verificación de instrumentos débiles	317
12.6	El contraste de sobreidentificación de restricciones (el estadístico J)	320

PARTE IV Análisis de regresión con datos de series temporales económicas

14.1	Retardos, primeras diferencias, logaritmos y tasas de crecimiento	377
14.2	Autocorrelación (correlación serial) y autocovarianza	378
14.3	Modelos autorregresivos	383
14.4	El modelo autorregresivo de retardos distribuidos	387
14.5	Estacionariedad	387
14.6	Regresión de series temporales con varios predictores	388
14.7	Contrastes de causalidad de Granger (contraste de contenido predictivo)	389
14.8	El contraste de Dickey-Fuller aumentado para raíz unitaria autorregresiva	400
14.9	El contraste QLR para la estabilidad de los coeficientes	405
14.10	Predicciones pseudo fuera de la muestra	407
15.1	El modelo de retardos distribuidos y la exogeneidad	427
15.2	Los supuestos del modelo de retardos distribuidos	428
15.3	Errores estándar HAC	433

15.4	Estimación de multiplicadores dinámicos con exogeneidad estricta	440
16.1	Vectores autorregresivos	456
16.2	Predicciones multiperiodo iteradas	461
16.3	Predicciones multiperiodo directas	462
16.4	Órdenes de integración, diferenciación y estacionariedad	464
16.5	Cointegración	468

PARTE V**Teoría econométrica del análisis de regresión**

17.1	Los supuestos ampliados de mínimos cuadrados para el modelo de regresión con regresor único	485
18.1	Los supuestos ampliados de mínimos cuadrados para el modelo de regresión múltiple	505
18.2	El teorema central del límite multivariante	507
18.3	El teorema de Gauss-Markov para regresión múltiple	514
18.3	Los supuestos MCG	516

Cuadros de interés general

La distribución de ingresos salariales en Estados Unidos en 2008	24
Un mal día en Wall Street	29
Diversificación financiera y carteras	33
¡Landon Gana!	50
La brecha de género en los ingresos salariales de los titulados universitarios en los Estados Unidos	61
Una nueva forma de fomentar el ahorro para la jubilación	64
El «beta» de una acción	85
El valor económico de un año de educación: ¿homocedasticidad o heterocedasticidad?	114
El efecto Mozart: ¿sesgo de variable omitida?	132
La rentabilidad de la educación y la brecha de género	202
La demanda de revistas de economía	205
¿Los fondos de inversión batén al mercado?	231
James Heckman y Daniel McFadden, ganadores del premio Nobel	293
¿Quién inventó la regresión de variables instrumentales?	306
Una regresión terrible	319
Las externalidades del consumo de tabaco	321
El efecto Hawthorne	344
¿Cuál es el efecto sobre el empleo del salario mínimo?	355
¿Se puede batir al mercado? Parte I	384
El río de sangre	391
¿Se puede batir al mercado? Parte II	408
Naranjos en movimiento	445
ÚLTIMA HORA: los operadores de materias primas hacen tiritar Disney World	447
Robert Engle y Clive Granger, ganadores del premio Nobel	471

Prefacio

La econometría puede ser una asignatura entretenida tanto para el profesor como para el estudiante. La realidad de la economía, los negocios, y el estado es un asunto complicado y confuso, repleto de ideas contrapuestas y preguntas que necesitan respuestas. ¿Resulta más efectivo abordar el problema de la conducción bajo los efectos del alcohol mediante leyes más severas o mediante un aumento de los impuestos sobre el alcohol? ¿Se podría ganar más dinero en bolsa comprando cuando los precios están históricamente bajos, en términos relativos a los salarios, o simplemente se debería no arriesgar tal y como sugiere la teoría del paseo aleatorio sobre el precio de los activos financieros? ¿Podría mejorarse la educación primaria reduciendo el número de alumnos por clase, o simplemente se debería poner a nuestros niños a escuchar a Mozart durante 10 minutos al día? La econometría nos ayuda a distinguir las buenas ideas de aquellas descabelladas y proporciona respuestas cuantitativas a importantes preguntas cuantitativas. La econometría abre una ventana en nuestro complicado mundo que permite ver las relaciones sobre las cuales las personas, las empresas y los gobiernos basan sus decisiones.

El libro *Introducción a la Econometría* está diseñado para un primer curso de econometría de grado universitario. De acuerdo con nuestra experiencia, para conseguir que la econometría sea pertinente en un curso introductorio, debe ocurrir que algunas aplicaciones interesantes consigan motivar la teoría y que la teoría acompañe a las aplicaciones. Este sencillo principio representa una significativa divergencia con la generación más antigua de libros de econometría, en los cuales los modelos teóricos y los supuestos no acompañan a las aplicaciones. No es extraño que algunos estudiantes cuestionen la relevancia de la econometría tras haber pasado una gran parte de su tiempo aprendiendo supuestos que posteriormente se revelan como poco realistas por lo que deben estudiar «soluciones» a «problemas» que aparecen cuando las aplicaciones no se corresponden con los supuestos. Creemos que es mucho mejor motivar la necesidad de herramientas con un ejemplo concreto y proporcionar posteriormente unos pocos y sencillos supuestos que se correspondan con esa aplicación. Al resultar la teoría inmediatamente relevante para las aplicaciones, este enfoque puede conseguir que la econometría cobre vida.

Nuevo en esta edición

- Tratamiento actualizado de los errores estándar en la regresión de datos de panel.
- Debate sobre cuándo y por qué los datos perdidos presentan un problema para el análisis de regresión.
- La utilización del diseño de la discontinuidad en la regresión como método para analizar cuasi experimentos.
- Tratamiento actualizado de los instrumentos débiles.

- Estudio sobre la utilización e interpretación de las variables de control integradas en el desarrollo del núcleo del análisis de regresión.
- Introducción del marco de análisis de las «Variables Respuesta» para datos experimentales.
- Cuadros de interés general adicionales.
- Ejercicios adicionales tanto escritos como empíricos.

Esta tercera edición se fundamenta tanto en la filosofía de la primera como de la segunda edición en cuanto a que las aplicaciones deben guiar la teoría, y no al revés.

Un cambio sustancial en esta edición ataña a la inferencia en la regresión con datos de panel (Capítulo 10). En datos de panel, los datos para una entidad individual habitualmente están correlacionados en el tiempo. Para que la inferencia sea válida, los errores estándar deben calcularse utilizando un método robusto a la presencia de esta correlación. El capítulo sobre datos de panel utiliza ahora este método los errores estándar agrupados, desde el comienzo. Los errores estándar agrupados son la extensión natural para datos de panel de los errores estándar heterocedástico-robustos introducidos en el tratamiento inicial del análisis de regresión de la Parte II. La investigación reciente muestra que los errores estándar agrupados poseen numerosas propiedades deseables, que se tratan en el Capítulo 10 y en un apéndice revisado del Capítulo 10.

Otro conjunto de cambios importante se refiere al tratamiento de los experimentos y cuasi experimentos en el Capítulo 13. El análisis de regresión de diferencias de las diferencias ha sido simplificado y está directamente inspirado en los principios de regresión múltiple introducidos en la Parte II. El Capítulo 13 trata el diseño de la discontinuidad en la regresión, que constituye un marco de análisis intuitivo e importante para el análisis de los datos cuasi experimentales. Además, el Capítulo 13 introduce el enfoque de variables respuesta, y relaciona esta cada vez más común terminología con los conceptos que se introducen en las Partes I y II.

Esta edición presenta otros cambios significativos. Uno de ellos es la incorporación de un tratamiento preciso, a la vez que accesible de las variables de control dentro del estudio inicial de la regresión múltiple. El Capítulo 7 trata ahora las condiciones bajo las que la inclusión de las variables de control resulte acertada en el sentido de que los coeficientes de las variables de interés sean insesgados incluso aunque los coeficientes de las variables de control en general no lo sean. Otros cambios son un tratamiento nuevo de los valores perdidos en el Capítulo 9, un nuevo apéndice del Capítulo 8 opcional basado en el cálculo de las pendientes y las elasticidades de las funciones de regresión no lineales, y un estudio actualizado en el Capítulo 12 sobre qué hacer si se tienen instrumentos débiles. Esta edición asimismo incluye nuevos cuadros de interés general, ejemplos empíricos actualizados, y ejercicios adicionales.

Características de este libro

Introducción a la Econometría difiere de otros libros de texto principalmente en tres aspectos distintos. Primero, integramos cuestiones del mundo real y datos en los desarrollos teóricos teniendo seriamente en cuenta los hallazgos principales del análisis empírico resultante. En segundo lugar, nuestra selección de temas refleja la teoría y la práctica modernas. En tercer lugar, proporcionamos la teoría y los supuestos que se corresponden con las aplicaciones. Nuestra intención es enseñar a los estudiantes a llegar a ser consumidores expertos de econometría y a serlo con un nivel matemático apropiado a un curso introductorio.

Cuestiones del mundo real y datos

Organizamos cada aspecto metodológico alrededor de una cuestión importante del mundo real que requiere una respuesta específica y numérica. Por ejemplo, enseñamos el análisis de la regresión con variable única, la regresión múltiple, y la forma funcional en el contexto de la estimación del efecto de los *inputs* educativos sobre los *outputs* educativos. (¿La reducción del tamaño de las aulas genera mejores calificaciones en los exámenes?). Enseñamos los métodos de datos de panel en el contexto del análisis del efecto de las leyes sobre conducción bajo los efectos del alcohol sobre la mortalidad en accidentes de tráfico. Utiliza-

mos la posible existencia de discriminación en el mercado de préstamos para la adquisición de vivienda como aplicación empírica para explicar la regresión con variable dependiente binaria (logit y probit). Enseñamos la estimación con variables instrumentales en el contexto de la estimación de la elasticidad de la demanda de cigarrillos. Aunque estos ejemplos implican un razonamiento económico, todos ellos pueden comprenderse con un curso introductorio de economía sencillo, y muchos de ellos sin ese curso previo. Así, el profesor puede centrarse en la enseñanza de econometría y no en la de microeconomía y macroeconomía.

Tratamos de forma seria todas nuestras aplicaciones empíricas, y de un modo que muestra a los estudiantes cómo pueden aprender de los datos y, al mismo tiempo, ser autocríticos y conscientes de las limitaciones del análisis empírico. Por medio de cada aplicación, enseñamos a los estudiantes a explorar especificaciones alternativas y, por tanto, a establecer si los resultados principales son robustos. Las cuestiones planteadas en las aplicaciones empíricas son importantes, a las que proporcionamos respuestas serias, y creemos que, creíbles. No obstante, emplazamos a los estudiantes y profesores a disentir, y les invitamos a reanalizar los datos, que están disponibles en la página web de la editorial (http://www.pearsonhighered.com/stock_watson).

Selección de temas actuales

La econometría ha recorrido un largo camino desde la década de los años 80. Los temas que tratamos reflejan lo mejor de la econometría aplicada actual. No pueden tratarse demasiadas cosas en un curso introductorio, por lo que nos centramos en los procedimientos y contrastes que se utilizan comúnmente en la práctica. Por ejemplo:

- **Regresión de variables instrumentales.** Presentamos la regresión de variables instrumentales como un método general para tratar la correlación entre el término de error y un regresor, lo cual puede aparecer por muchas razones, incluyendo las variables omitidas y la causalidad simultánea. A los dos requisitos para la validez de un instrumento —exogeneidad y relevancia— se les asigna la misma importancia. Continuamos la presentación con un tratamiento extenso acerca de cómo conocer el origen de los instrumentos y con los contrastes de sobreidentificación de restricciones y diagnóstico para instrumentos débiles, y explicamos qué hacer en el caso de que estos diagnósticos sugieran problemas.
- **Evaluación de programas.** Un número creciente de estudios económicos analizan tanto los experimentos aleatorizados controlados como los quasi experimentos, asimismo conocidos como experimentos naturales. Afrontamos estos temas, a los que a menudo comúnmente se conoce como evaluación de programas, en el Capítulo 13. Presentamos esta estrategia de investigación como un enfoque alternativo de los problemas de variables omitidas, causalidad simultánea y selección, y abordamos tanto las fortalezas como las debilidades de los estudios utilizando datos experimentales o quasi experimentales.
- **Predicción.** El capítulo sobre predicción (Capítulo 14) considera predicciones univariantes (autorregresivas) o multivariantes utilizando regresión de series temporales, no los grandes modelos estructurales de ecuaciones simultáneas. Nos centramos en herramientas sencillas y fiables, tales como los modelos autorregresivos y la selección de modelos mediante un criterio de información que funciona bien en la práctica. Este capítulo asimismo muestra un tratamiento de las tendencias estocásticas (raíces unitarias) orientado a la práctica, contrastes de raíz unitaria, contrastes de cambios estructurales (con punto de ruptura conocido y desconocido), y predicción pseudofuera de la muestra, todo en el contexto de modelos de predicción de series temporales estables y fiables.
- **Regresión de series temporales.** Hacemos una clara distinción entre dos aplicaciones muy diferentes de la regresión de series temporales: predicción y estimación de efectos causales dinámicos. El capítulo sobre inferencia causal mediante datos de series temporales (Capítulo 15) presta una cuidadosa atención acerca de si los diferentes métodos de estimación, incluyendo mínimos cuadrados generalizados, conducen o no a inferencias causales válidas, y a si es aconsejable estimar regresiones dinámicas mediante MCO con errores estándar consistentes a heterocedasticidad y autocorrelación.

Teoría que acompaña a las aplicaciones

Aunque las herramientas econométricas se motivan mejor mediante aplicaciones empíricas, los estudiantes necesitan aprender teoría econométrica suficiente como para comprender las fortalezas y las limitaciones de esas herramientas. Proporcionamos un tratamiento moderno en el cual el ajuste entre teoría y aplicaciones es tan estrecho como resulte posible, manteniendo las matemáticas a un nivel que solo requiere álgebra.

Las aplicaciones empíricas modernas comparten algunas características comunes: las bases de datos habitualmente son amplias (cientos de observaciones, a menudo más); los regresores no son fijos en muestras repetidas, sino más bien están recogidos mediante muestreo aleatorio (o algún otro mecanismo que lo hace aleatorio); los datos no están normalmente distribuidos; y no existe razón a priori para pensar que los errores sean homocedásticos (aunque a menudo existen razones para pensar que son heterocedásticos).

Esas observaciones llevan a diferencias importantes entre los desarrollos teóricos de este libro de texto y otros libros de texto.

- **Enfoque de muestras grandes.** Dado que las bases de datos son amplias, desde un principio utilizamos aproximaciones normales para muestras grandes para las distribuciones muestrales de los contrastes de hipótesis e intervalos de confianza. Nuestra experiencia es que se requiere menos tiempo para la enseñanza de los rudimentos de las aproximaciones para grandes muestras que para la enseñanza de las distribuciones exactas t de Student y F , la corrección por los grados de libertad, y demás. Este enfoque para muestras grandes asimismo ahorra a los estudiantes la frustración de descubrir que, debido a la no normalidad de los errores, la distribución teórica exacta que acaban de dominar es irrelevante. Una vez demostrado en el contexto de la media muestral, el enfoque de muestras grandes para la contrastación de hipótesis e intervalos de confianza conduce directamente al análisis de regresión múltiple, logit y probit, estimación por variables instrumentales, y los métodos de series temporales.
- **Muestreo aleatorio.** Debido a que los regresores raramente son fijos en las aplicaciones económicas, desde el principio consideramos los datos de todas las variables (dependientes e independientes) como el resultado de un muestreo aleatorio. Este supuesto encaja con nuestras aplicaciones iniciales para datos de sección cruzada, y es extendible fácilmente a datos de panel y de series temporales, y debido a nuestro enfoque de muestras grandes no plantea dificultades conceptuales o matemáticas.
- **Heterocedasticidad.** Los económetras aplicados utilizan de forma rutinaria errores estándar heterocedástico-robustos con el fin de eliminar preocupaciones acerca de si se presenta o no heterocedasticidad. En este libro, vamos más allá del tratamiento de la heterocedasticidad como una excepción o un «problema» a «resolver»; en su lugar, aceptamos la heterocedasticidad desde el principio y utilizamos sencillamente errores estándar heterocedástico-robustos. Presentamos la homocedasticidad como un caso especial que proporciona una motivación teórica para MCO.

Productores cualificados, consumidores expertos

Esperamos que los estudiantes que utilicen este libro se conviertan en consumidores expertos de análisis empírico. Para ello, deben aprender no solo cómo utilizar las herramientas del análisis de regresión, sino que además deben evaluar asimismo la validez del análisis empírico que se les presente.

Nuestro método para la enseñanza de cómo evaluar un estudio empírico es triple. En primer lugar, inmediatamente después de introducir las principales herramientas del análisis de regresión, dedicamos el Capítulo 9 a las amenazas a la validez interna y externa de un estudio empírico. Este capítulo trata los problemas en los datos y las cuestiones de la generalización de los resultados a otros entornos. Asimismo examina las principales amenazas al análisis de regresión, incluyendo variables omitidas, errores de especificación de la

forma funcional, errores en las variables, selección, y simultaneidad, y las distintas formas de reconocer estas amenazas en la práctica.

Segundo, aplicamos estos métodos para evaluar los estudios empíricos para el análisis empírico de los estudios desarrollados en este libro. Lo hacemos considerando especificaciones alternativas y considerando sistemáticamente las distintas amenazas a la validez de los análisis presentados en este libro.

Tercero, para llegar a ser consumidores expertos, los estudiantes necesitan experiencia de primera mano como productores. El aprendizaje activo supera al aprendizaje pasivo, y la econometría es una asignatura ideal para el aprendizaje activo. Por esta razón la página web del libro proporciona datos, *software*, y sugerencias para ejercicios empíricos de diferente alcance.

Modo de abordar las matemáticas y nivel de rigor

Nuestra intención es que los estudiantes desarrollen una comprensión especializada de las herramientas del análisis de regresión moderno, tanto si la asignatura se enseña a un nivel «alto» como a un nivel «bajo» de matemáticas. Las Partes I a IV del texto (que cubre el material principal) son accesibles para los estudiantes solamente con un poco de cálculo matemático. Las partes I a IV contienen menos ecuaciones, y más aplicaciones, que la mayoría de libros de econometría introductoria y muchas menos ecuaciones que los libros dedicados a las secciones matemáticas de las asignaturas de nivel universitario. No obstante más ecuaciones no implican un tratamiento más experto. En nuestra experiencia, un tratamiento más matemático no conduce a una comprensión más profunda para la mayoría de estudiantes.

Dicho esto, los distintos estudiantes aprenden de forma diferente, y para los estudiantes bien preparados en matemáticas el aprendizaje puede enriquecerse con un material matemático más explícito. Consecuentemente, la parte V contiene una introducción a la teoría econométrica apropiada para estudiantes con un bagaje matemático más potente. Creemos que, cuando los capítulos matemáticos de la Parte V sean usados junto con el material de las Partes I a IV, este libro resulta adecuado para cursos avanzados de grado universitario o máster.

Contenidos y organización

Introducción a la Econometría tiene cinco partes. Este libro de texto supone que el estudiante ha cursado una asignatura de probabilidad y estadística aunque repasamos este material en la Parte I. Cubrimos el material central del análisis de regresión en la Parte II. Las Partes III, IV y V presentan temas adicionales cimentados en el tratamiento central de la parte II.

Parte I

El Capítulo 1 introduce la econometría y pone de manifiesto la importancia de proporcionar respuestas cuantitativas a preguntas cuantitativas. Debate el concepto de causalidad en los estudios empíricos y proporciona una visión de conjunto de los diferentes tipos de datos a los que se enfrenta la econometría. El material de probabilidad y estadística se repasa en los Capítulos 2 y 3, respectivamente; la cuestión de si estos capítulos se explican en la asignatura o simplemente se toman como referencia dependerá de las circunstancias de los estudiantes.

Parte II

El Capítulo 4 introduce la regresión con regresor único y la estimación de mínimos cuadrados ordinarios (MCO), y el Capítulo 5 trata los contrastes de hipótesis y los intervalos de confianza del modelo de regresión con regresor único. En el Capítulo 6, los estudiantes aprenden cómo pueden tratar el sesgo de variables omitidas mediante regresión múltiple, estimando así el efecto de una variable independiente manteniendo constantes el resto de variables independientes. El Capítulo 7 trata los contrastes de hipótesis, incluyendo

los contrastes F , y los intervalos de confianza en regresión múltiple. En el Capítulo 8, se amplía el modelo de regresión lineal a funciones que son lineales en los parámetros (por lo que los parámetros pueden estimarse por MCO). En el Capítulo 9, los estudiantes vuelven sobre sus pasos y aprenden cómo identificar las fortalezas y las limitaciones de los estudios de regresión, viendo en el proceso cómo aplicar los conceptos de validez interna y externa.

Parte III

La Parte III presenta extensiones de los métodos de regresión. En el Capítulo 10, los estudiantes aprenden a utilizar los datos de panel para tener en cuenta variables no observadas que son constantes en el tiempo. El Capítulo 11 estudia la regresión con una variable dependiente binaria. El Capítulo 12 muestra cómo puede utilizarse la regresión de variables instrumentales para tratar la serie de problemas que genera la correlación entre el término de error y el regresor, y examina cómo pueden encontrarse y evaluarse los instrumentos válidos. El Capítulo 13 introduce a los estudiantes en el análisis de datos de los experimentos y cuasi experimentos, o experimentos naturales; estos temas se conocen a menudo como «evaluación de programas».

Parte IV

La Parte IV se ocupa de la regresión con datos de series temporales. El Capítulo 14 se centra en la predicción, e introduce varias herramientas modernas para analizar las regresiones de series temporales tales como los contrastes de raíces unitarias y los contrastes de estabilidad. El Capítulo 15 trata la utilización de datos de series temporales para estimar relaciones causales. El Capítulo 16 presenta algunas herramientas más avanzadas para el análisis de series temporales, que incluyen los modelos de heterocedasticidad condicionada.

Parte V

La Parte V es una introducción a la teoría econométrica. Esta parte es más que un apéndice que completa los detalles matemáticos omitidos en el texto. Además, es un tratamiento autocontenido de la teoría econométrica de estimación e inferencia en el modelo de regresión lineal. El Capítulo 17 desarrolla la teoría del análisis de regresión para un único regresor; la exposición no utiliza álgebra matricial, aunque requiere un mayor nivel de sofisticación matemática que el resto del texto. El Capítulo 18 presenta y estudia el modelo de regresión múltiple, la regresión de variables instrumentales y la estimación del modelo lineal mediante el método generalizado de momentos; todo ello en notación matricial.

Requisitos previos en el libro

Dado que a los diferentes profesores les gusta enfatizar diferentes aspectos, escribimos este libro teniendo en mente las diferentes preferencias de enseñanza. Hasta el máximo grado posible, los capítulos de las Partes III, IV y V son «sin ayuda», en el sentido que no requieren la impartición previa de todos los capítulos previos. Los requisitos previos de cada capítulo se describen en la Tabla I. Aunque consideramos que la secuencia de temas adoptada en este libro de texto funciona en nuestros cursos, los capítulos están escritos de un modo que permite a los profesores presentar los temas en un orden diferente si lo desean.

Muestra de planificación

Este libro se ajusta a diferentes estructuras de planificación del curso.

TABLA 1

Guía de requisitos previos para capítulos de temas especiales de las partes III, IV y V

	Requisitos previos de partes o capítulos							
	Parte I	Parte II	Parte III	Parte IV		Parte V		
Capítulo	1-3	4-7,9 8	10.1 10.2	12.1 12.2	14.1- 14.4	14.5 14.8	15	17
10	X ^a	X ^a X						
11	X ^a	X ^a X						
12.1, 12.2	X ^a	X ^a X						
12.3, 12.6	X ^a	X ^a X	X	X				
13	X ^a	X ^a X	X	X				
14	X ^a	X ^a ^b						
15	X ^a	X ^a ^b				X		
16	X ^a	X ^a ^b				X	X	X
17	X	X X						
18	X	X X		X				X

Esta tabla muestra los requisitos previos mínimos necesarios para cubrir el material en un capítulo dado. Por ejemplo, la estimación de los efectos causales dinámicos con datos de series temporales (Capítulo 15) requiere en primer lugar la Parte I (cuando sea necesario, dependiendo de la preparación de los estudiantes, y excepto como se recoge en la nota a pie de página a), la Parte II (excepto el capítulo 8; véase nota a pie b), y las Secciones 14.1 a 14.4.

^aLos Capítulos 10 a 16 utilizan exclusivamente aproximaciones para muestras grandes de las distribuciones muestrales, por lo que las Secciones opcionales 3.6 (la distribución *t* de Student para el contraste de los coeficientes de regresión) pueden saltarse.

^bLos Capítulos 14 a 16 (los capítulos de series temporales) puede explicarse sin haber impartido con anterioridad el Capítulo 8 (funciones de regresión no lineales) si el profesor se detiene a explicar el uso de transformaciones logarítmicas para aproximar los cambios porcentuales.

Econometría introductoria estándar

Este curso introduce la econometría (Capítulo 1) y repasa probabilidad y estadística si es necesario (Capítulos 2 y 3). Después se traslada a la regresión con regresor único, regresión múltiple, las bases del análisis de la forma funcional, y la evaluación de los estudios de regresión (todo en la Parte II). El curso pasa a cubrir la regresión con datos de panel (Capítulo 10), regresión con variable dependiente limitada (Capítulo 11), y la regresión de variables instrumentales (Capítulo 12), si el tiempo es suficiente. El curso concluye con los experimentos y quasi experimentos en el Capítulo 13, temas que proporcionan una oportunidad para retomar las cuestiones de estimación de efectos causales que surgen al principio del curso y recapitular los métodos de regresión centrales. Requisitos previos: *Álgebra II e Introducción a la Estadística*.

Econometría introductoria con series temporales y aplicaciones predictivas

Como curso introductorio estándar, este curso cubre toda la Parte I (si es necesario) y la Parte II. Opcionalmente, a continuación el curso proporciona una introducción breve a los datos de panel (Secciones 10.1 y 10.2) y se ocupa de la regresión con variables instrumentales (Capítulo 12, o solamente 12.1 y 12.2). El curso pasa a la Parte IV, para cubrir predicción (Capítulo 14) y la estimación de efectos causales dinámicos (Capítulo 15). Si se dispone de tiempo, el curso puede incluir algunos temas avanzados del análisis de series temporales tales como heterocedasticidad agrupada, y heterocedasticidad condicional (Sección 16.5). Requisitos previos: *Álgebra II e Introducción a la Estadística*.

Análisis aplicado de series temporales y predicción

Este libro puede usarse asimismo para un curso breve de series temporales aplicadas y predicción, para lo cual es requisito previo un curso de análisis de regresión. Se debe dedicar tiempo al repaso de las herramientas del análisis básico de regresión de la Parte II, dependiendo de la preparación de los estudiantes. El curso se traslada entonces directamente a la Parte IV pasando por la predicción (Capítulo 14), la estimación de los efectos causales dinámicos (Capítulo 15), y los temas avanzados de series temporales (Capítulo 16), incluyendo vectores autorregresivos y heterocedasticidad condicional. Una componente importante de este curso es la realización de ejercicios de predicción, disponibles para los profesores en la página web del libro. Requisitos previos: *Álgebra II* y *Econometría introductoria básica o equivalente*.

Introducción a la teoría econométrica

Este libro resulta asimismo adecuado para un curso universitario avanzado en el que los estudiantes posean una preparación matemática fuerte o para un curso de econometría a nivel de máster. El curso repasa brevemente la teoría de estadística y probabilidad si es necesario (Parte I). El curso introduce el análisis de regresión utilizando un tratamiento no matemático basado en las aplicaciones de la Parte II. Esta introducción continúa con el desarrollo teórico de los Capítulos 17 y 18 (hasta la Sección 18.5). El curso entonces trata la regresión con variable dependiente limitada (Capítulo 11) y la estimación máximo verosímil (Apéndice 11.2). Tras ello, el curso retorna opcionalmente a la regresión de variables instrumentales y al método generalizado de los momentos (Capítulo 12 y Sección 18.7), los métodos de series temporales (Capítulo 14), y la estimación de los efectos causales mediante datos de series temporales y mínimos cuadrados generalizados (Capítulo 15 y Sección 18.6). *Requisitos previos: Cálculo e introducción a la Estadística. El Capítulo 18 supone conocimiento previo de álgebra matricial.*

Características pedagógicas

Este libro de texto posee varias características pedagógicas con la intención de ayudar a los estudiantes a entender, retener y aplicar las ideas principales. Las *introducciones de los capítulos* proporcionan la motivación y un anclaje al mundo real, además de una breve hoja de ruta que destaca la secuencia del razonamiento. Los *Términos clave* se resaltan en negrita y se definen en su contexto a lo largo de cada capítulo, y los *Recuadros de los Conceptos clave* a intervalos regulares recogen las ideas centrales. Los *Cuadros de interés general* proporcionan excursiones interesantes a los temas relacionados y resaltan los estudios del mundo real que utilizan los métodos o los conceptos estudiados en el texto. Un *Resumen* concluye cada capítulo y sirve como marco útil para repasar los puntos principales del tema tratado. Las preguntas de la sección *Repaso de contenidos* evalúan la comprensión de los estudiantes acerca del contenido principal, los *Ejercicios* proporcionan un trabajo más intensivo desde el punto de vista práctico con los conceptos y técnicas introducidos en el capítulo, y los *Ejercicios empíricos* permiten a los estudiantes aplicar lo aprendido para resolver cuestiones empíricas del mundo real. Al final del libro, el *Apéndice* proporciona tablas estadísticas, la sección de *Bibliografía* enumera las fuentes para ampliar la lectura, y el *Glosario* define adecuadamente muchos de los términos clave del libro.

Complementos adjuntos al libro

Los complementos *online* que acompañan la tercera edición de *Introducción a la Econometría* incluyen el Manual de soluciones, *Archivo de preguntas de test* (de Manfred W. Keil del Claremont McKenna College), y las diapositivas de PowerPoint® con cuadros de texto, tablas y Conceptos clave. El Manual de soluciones incluye soluciones para todos los ejercicios planteados al final de cada capítulo, mientras el archivo de Preguntas de test, presentado en «Test Generator Software» (TestGen con QuizMaster), proporciona una

rica oferta editada fácilmente de los problemas y cuestiones de test de varios tipos para las demandas específicas del curso. Estos recursos están disponibles para descarga desde el Centro de Recursos para el profesor en http://www.pearsonhighered.com/stock_watson.

Además, la página web de la editorial, http://www.pearsonhighered.com/stock_watson, proporciona una extensa gama de recursos adicionales para estudiantes y profesores. Incluye bases de datos para los ejercicios empíricos planteados en el texto, bases de datos para replicar los resultados empíricos, réplicas de los archivos de los resultados empíricos proporcionados en el texto, cuestionarios prácticos, respuestas a las preguntas planteadas en las secciones de Repaso de conceptos del final de los capítulos y ejercicios, y tutoriales.

Agradecimientos

Un gran número de personas han contribuido a la primera edición de este libro. Nuestras mayores deudas de gratitud son para nuestros colegas de Harvard y Princeton que utilizaron las versiones preliminares de este libro en sus clases. Suzanne Cooper, de la Kennedy School of Government de Harvard, brindó valiosas sugerencias y comentarios detallados sobre múltiples borradores. Como profesora que comparte asignatura con uno de los autores (Stock), asimismo ayudó a examinar la mayoría del material de este libro mientras estaba siendo diseñado como un curso para los estudiantes de máster de la Kennedy School. Estamos del mismo modo en deuda con otros dos colegas de la Kennedy School, Alberto Abadie y Sue Dynarski, por sus pacientes explicaciones de los quasi experimentos y la disciplina de evaluación de programas y por sus comentarios detallados sobre los borradores preliminares del texto. De Princeton, Eli Taumer impartió clases a partir de un borrador preliminar y proporcionó asimismo útiles comentarios sobre el penúltimo borrador del libro.

Asimismo estamos en deuda con muchos de nuestros amigos y colegas de econometría que dedicaron su tiempo con nosotros discutiendo sobre el contenido de este libro y que colectivamente hicieron tantas sugerencias útiles. Bruce Hansen (Universidad de Wisconsin-Madison) y Bo Honore (Princeton) proporcionaron comentarios útiles sobre los bocetos preliminares y las primeras versiones del material central de la Parte II. Joshua Angrist (MIT) y Guido Imbens (Universidad de California, Berkeley) proporcionaron sesudas sugerencias sobre el tratamiento del material acerca de la evaluación de programas. Nuestra presentación del material de series temporales se benefició de los debates con Yacín Ait-Sahalia (Princeton), Graham Elliott (Universidad de California, San Diego), Andrew Harvey (Universidad de Cambridge), y Christopher Sims (Princeton). Finalmente, muchas personas han realizado sugerencias útiles sobre partes del manuscrito cercanas a su área de especialidad: Don Andrews (Yale), John Bound (Universidad de Michigan), Gregory Chow (Princeton), Thomas Downes (Tufts), David Drukker (StataCorp.), Jean Baldwin Grossman (Princeton), Eric Hanushek (Institución Hoover), James Heckman (Universidad de Chicago), Hang Hong (Princeton), Caroline Hoxby (Harvard), Alan Krueger (Princeton), Steven Levitt (Universidad de Chicago), Richard Light (Harvard), David Neumark (Universidad Estatal de Michigan), Joseph Newhouse (Harvard), Pierre Perron (Universidad de Boston), Kenneth Warner (Universidad de Michigan) y Richard Zeckhauser (Harvard).

Muchas personas fueron muy generosas proporcionándonos datos. Los datos de calificaciones en los exámenes de California se construyeron con la asistencia de Les Axelrod, de la Sección de Standars and Assessments del Departamento de Educación de California. Agradecemos a Charlie de Pascale, de los Servicios de Evaluación de los Estudiantes del Departamento de Educación de Massachusetts, su ayuda con algunos aspectos de la base de datos de calificaciones en los exámenes de Massachusetts. Christopher Ruhm (Universidad de California, Greensboro) nos proporcionó gentilmente su base de datos sobre conducción bajo los efectos del alcohol y mortalidad en accidentes de tráfico. El departamento de investigación del Banco de la Reserva Federal de Boston merece nuestro agradecimiento por la puesta en común de sus datos sobre discriminación racial en préstamos para la compra de vivienda; agradecemos particularmente a Geoffrey Tootell por proporcionarnos la versión actualizada de la base de datos utilizada en el Capítulo 9 y a Lynn Browne por explicarnos su contexto político. Agradecemos a Jonathan Gruber (MIT) por compartir

sus datos sobre venta de cigarrillos, que analizamos en el Capítulo 10, y a Alan Krueger (Princeton) por su ayuda con los datos STAR de Tennessee que analizamos en el Capítulo 11.

Estamos agradecidos a algunas personas por haber verificado cuidadosamente las páginas en busca de erratas, Kerry Griffin y Yair Listokin leyeron el manuscrito completo, y Andrew Fraker, Ori Heffetz, Amber Henry, Hong Li, Alessandro Tarozzi, y Matt Watson trabajaron en algunos capítulos.

En la primera edición, nos beneficiamos de la ayuda de una excepcional editora de desarrollo, Jane Tufts, cuya creatividad, trabajo duro y atención a los detalles mejoró este libro en muchos sentidos, grandes y pequeños. Addison-Wesley nos proporcionó apoyo de primer nivel comenzando con nuestra excelente editora, Sylvia Mallory, extensible a todo el equipo de publicación. Jane y Sylvia nos enseñaron pacientemente mucho sobre escritura, organización y presentación, y sus esfuerzos son evidentes en todas las páginas de este libro. Extendemos nuestros agradecimientos al soberbio equipo de Addison Wesley, quienes trabajaron con nosotros en la segunda edición: Adrienne D'Ambrossio (editor senior de adquisiciones), Bridget Page (productora asociada de medios), Charles Spaulding (diseñador senior), Nancy Fenton (editor gerente) y su selección de Nancy Freihofer y Thompson Steele Inc. quienes manejaron el proceso productivo totalmente, Heather McNally (coordinador de complementos), y Denise Clinton (editora jefe). Finalmente, tenemos la ventaja de la edición experta de Kay Ueno en la segunda edición. Agradecemos además al excelente equipo de Addison Wesley para la tercera edición formado por Adrienne D'Ambrosio, Nancy Fenton y Jill Kolongowski, así como Mary Sanger, el director del proyecto con Nesbitt Graphics.

Recibimos además gran cantidad de ayuda y sugerencias de los profesores, estudiantes e investigadores al preparar esta tercera edición. Los cambios realizados en la tercera edición incorporan o reflejan sugerencias, correcciones, comentarios, datos y ayuda proporcionada por numerosos investigadores y profesores: Donald Andrews (Universidad de Yale), James Cobbe (Universidad Estatal de Florida), Susan Dynarski (Universidad de Michigan), Nicole Eichelberger (Universidad Tecnológica de Texas), Boyd Fjeldsted (Universidad de Utah), Martina Grunow, Daniel Hammermesh (Universidad de Texas-Austin), Keisuke Hirano (Universidad de Arizona), Bo Honore (Universidad de Princeton), Guido Imbens (Universidad de Harvard), Manfred Keil (Claremont McKenna College), David Laibson (Universidad de Harvard), David Lee (Universidad de Princeton), Brigitte Madrian (Universidad de Harvard), Jorge Marquez (Universidad de Maryland), Karen Bennett Mathis (Departamento de Cítricos de Florida), Ulrich Müller (Universidad de Princeton), Serena Ng (Universidad de Columbia), Harry Patrinos (Banco Mundial), Peter Summers (Universidad Tecnológica de Texas), Andrew Vasnov (Universidad de Sydney) y Douglas Young (Universidad Estatal de Montana). Además nos beneficiamos de las aportaciones de los estudiantes F. Hoces de la Guardia y Carrie Wilson.

Se realizaron concienzudas revisiones de la tercera edición para Addison-Wesley por parte de Steve DeLoach (Universidad de Elon), Jeffrey De Simone (Universidad de Texas en Arlington), Gary V. Engelhardt (Universidad de Syracuse), Luca Flabbi (Universidad de Georgetown), Steffen Habermalz (Universidad de Northwestern), Carolyn J. Heinrich (Universidad de Wisconsin-Madison), Emma M. Iglesias-Vazquez (Universidad Estatal de Michigan), Carlos Lamarche (Universidad de Oklahoma), Vicki A. McCracken (Universidad Estatal de Washington), Claudiney M. Pereira (Universidad de Tulane), y John T. Warner (Universidad de Clemson). Además, recibimos aportaciones muy útiles de las revisiones de los borradores de los Capítulos 7 y 10 de John Berdell (Universidad De Paul), Janet Kohlhase (Universidad de Houston), Aprajit Mahajan (Universidad de Stanford), Xia Meng (Universidad de Brandeis), y Chan Shen (Universidad de Georgetown).

Sobre todo, estamos en deuda con nuestras familias por su paciencia en todo este proyecto. Escribir este libro ha llevado largo tiempo, y para ellos el proyecto ha debido parecer interminable. Ellos más que nadie han soportado el peso de este compromiso, y les estamos profundamente agradecidos por su ayuda y apoyo.

1

Cuestiones económicas y datos

Pregunte a media docena de económetras qué es la econometría y obtendrá media docena de respuestas diferentes. Alguien podría decirle que la econometría es la ciencia para la contrastación de teorías económicas. Un segundo podría decirle que la econometría es el conjunto de herramientas utilizadas para predicción de valores futuros de variables económicas, tales como las ventas de las empresas, el crecimiento de la economía en su conjunto, o el precio de las acciones. Otro podría decir que la econometría es el proceso de ajuste de modelos económicos matemáticos a los datos del mundo real. Un cuarto podría decirle que es la ciencia y el arte de utilizar los datos históricos para realizar recomendaciones numéricas, o cuantitativas, sobre las políticas a realizar por el gobierno en los negocios.

De hecho, todas esas respuestas son correctas. Desde un punto de vista amplio, la econometría es la ciencia y el arte de utilizar la teoría económica y las técnicas estadísticas para analizar los datos económicos. Los métodos econométricos son utilizados por distintas ramas de la economía, lo que incluye las finanzas, la economía laboral, la macroeconomía, la microeconomía, el *marketing* y la política económica. Los métodos econométricos son asimismo usados habitualmente en otras ciencias sociales, como las ciencias políticas y la sociología.

Este libro introduce el conjunto central de los métodos utilizados por los económetras. Utilizaremos esos métodos para responder a la gama de preguntas cuantitativas específicas, tomadas del mundo de los negocios y las políticas gubernamentales. Este capítulo plantea cuatro de estas cuestiones y estudia, en términos generales, el marco econométrico para resolverlas. El capítulo concluye con una visión de conjunto de los principales tipos de datos disponibles por parte de los económetras para responder esas y otras preguntas económicas cuantitativas.

1.1 Preguntas económicas a examen

Muchas decisiones económicas de los negocios y del gobierno versan acerca de la comprensión de las relaciones entre las variables del mundo que nos rodea. Esas decisiones requieren respuestas cuantitativas a preguntas cuantitativas.

Este libro examina algunas preguntas cuantitativas extraídas de la actualidad económica. Estas cuatro preguntas se refieren a la política educativa, los prejuicios raciales en los préstamos hipotecarios, el consumo de cigarrillos, y la predicción macroeconómica.

Pregunta #1 ¿Mejora la reducción del tamaño de las clases la educación en la escuela primaria?

Las propuestas para la reforma del sistema educativo de EE.UU. generan un debate acalorado. Muchas de las propuestas se refieren a los estudiantes más jóvenes, los de las escuelas primarias. La educación primaria tiene varios objetivos, tales como el desarrollo de las habilidades sociales, pero para muchos padres el objetivo más importante es el aprendizaje académico básico: leer, escribir, y las matemáticas básicas. Una propuesta importante para mejorar el aprendizaje básico es la de reducir el tamaño de las clases en las escuelas primarias. El argumento es que con pocos estudiantes en las clases, cada estudiante obtiene más atención del maestro, hay menos interrupciones de la clase, se mejora el aprendizaje y mejoran las calificaciones.

¿Pero cuál es, exactamente, el efecto sobre la educación primaria de la reducción del tamaño de las clases? Reducir el tamaño de las clases supone dinero: requiere contratar más maestros y, si la escuela está a su máxima capacidad, construir más aulas. Un decisor que considere contratar más profesores debe sopesar estos costes y beneficios. No obstante, para sopesar costes y beneficios, el decisor debe adquirir una comprensión cuantitativa acerca de los posibles beneficios. ¿El efecto beneficioso sobre el aprendizaje básico de tener clases más pequeñas es grande o pequeño? ¿Es posible que disponer de clases más reducidas no afecte en realidad al aprendizaje básico?

Aunque el sentido común y la experiencia diaria puedan sugerir que se produce más aprendizaje cuando hay menos estudiantes, el sentido común no proporciona una respuesta cuantitativa a la pregunta de cuál es exactamente el efecto sobre el aprendizaje básico de la reducción del tamaño de las clases. Para proporcionar tal respuesta, debemos examinar la evidencia empírica, es decir, la evidencia basada en los datos que relaciona el tamaño de las clases con el aprendizaje básico en las escuelas primarias.

En este libro, examinamos la relación entre el tamaño de las clases y el aprendizaje básico mediante la utilización de los datos recopilados sobre 420 distritos escolares de California en el año 1999. En los datos de California, los estudiantes de los distritos con tamaño de clases pequeño tienden a comportarse mejor en los exámenes estandarizados que los estudiantes de los distritos con clases mayores. Si bien este hecho es consistente con la idea de que las clases más pequeñas causan mejores calificaciones en los exámenes, esto puede reflejar simplemente que existen otras muchas ventajas que tienen los estudiantes de los distritos con clases menores sobre sus homólogos de los distritos con clases mayores. Por ejemplo, los distritos con tamaño de clases pequeño suelen tener habitantes más ricos que los distritos con clases grandes, por tanto, los estudiantes de los distritos con clases pequeñas podrían tener más oportunidades de aprendizaje fuera del aula. Podría suceder que fueran estas oportunidades extras de aprendizaje las que condujeran a calificaciones más altas en estos exámenes o pruebas estandarizadas, y no el menor tamaño de las clases. En la Parte II, utilizaremos el análisis de regresión múltiple para aislar el efecto de los cambios en el tamaño de las clases de los cambios en otros factores, tales como las circunstancias económicas de los estudiantes.

Pregunta #2 ¿Existe discriminación racial en el mercado de préstamos para la vivienda?

La mayoría de la gente compra sus casas con la ayuda de una hipoteca, un préstamo cuantioso asegurado por el valor de la casa. Por ley, las instituciones de préstamo en los EE.UU. no pueden tener en cuenta la etnia al decidir acerca de aprobar o denegar una solicitud para un préstamo hipotecario: los solicitantes, que son idénticos en todos los sentidos excepto en su etnia, deberían tener la misma probabilidad de que sus solicitudes fueran aprobadas. En teoría, por tanto, no deberían existir prejuicios raciales en la concesión de préstamos hipotecarios.

En contra de esta conclusión teórica, los investigadores del Banco de la Reserva Federal de Boston hallaron (utilizando datos desde el principio de los años 90) que al 28 % de los solicitantes negros se les denegaba la hipoteca, mientras que esta denegación solo la experimentaba el 9 % de los solicitantes blancos. ¿Estos datos indican que, en la práctica, existen prejuicios raciales en los préstamos hipotecarios? Si es así, ¿a cuánto asciende?

El hecho de que se les denieguen más solicitudes a los negros que a los blancos en los datos de la FED de Boston no proporciona por sí mismo evidencia acerca de la existencia de discriminación por parte de los prestamistas hipotecarios, puesto que las solicitudes de negros y blancos difieren en muchos más sentidos que su etnia. Antes de concluir que existen prejuicios raciales en el mercado hipotecario, estos datos deben ser examinados de forma más cuidadosa para comprobar si existe una diferencia en la probabilidad de denegación para los solicitantes *idénticos en todo lo demás* y, si es así, comprobar si esa diferencia es grande o pequeña. Para ello, en el Capítulo 11 se introducen los métodos estadísticos que hacen posible cuantificar el efecto de la etnia sobre la oportunidad de obtener un préstamo hipotecario, *manteniendo constantes* otras características del solicitante, principalmente su capacidad de devolver el préstamo.

Pregunta #3 ¿Cuánto reducen el tabaquismo los impuestos sobre los cigarrillos?

Fumar cigarrillos es una de las preocupaciones principales sobre salud pública a nivel mundial. Muchos de los costes del tabaquismo, tales como los gastos médicos de los cuidados de las enfermedades provocadas por el tabaquismo, así como los costes menos cuantificables que les supone a los no fumadores que prefieren no respirar el humo usado de los cigarrillos, son soportados por otros miembros de la sociedad. Debido a que estos costes son soportados por otras personas distintas al fumador, existe un papel reservado para la intervención pública que debe encargarse de intentar reducir el consumo de cigarrillos. Una de las herramientas más flexibles para reducir el consumo consiste en incrementar los impuestos sobre los cigarrillos.

La teoría económica básica dice que si el precio de los cigarrillos aumenta, el consumo se reducirá. ¿Pero en cuánto? Si el precio de venta aumenta en un 1 %, ¿cuál será el porcentaje en que disminuirá la cantidad vendida de cigarrillos? El cambio porcentual en la cantidad demandada que resulta de un aumento en un 1 % del precio es la *elasticidad precio de la demanda*.

Si se pretende reducir el consumo en una cierta cuantía, por ejemplo un 20 %, mediante el aumento de los impuestos, entonces resulta necesario conocer la elasticidad-precio con el fin de calcular el incremento necesario del precio que permita alcanzar esa reducción en el consumo. ¿Pero cuál es la elasticidad-precio de los cigarrillos?

Aunque la teoría económica nos proporciona los conceptos que nos ayudan a responder esta pregunta, esto no proporciona el valor numérico de la elasticidad-precio de la demanda. Para conocer la elasticidad, debemos examinar la evidencia empírica acerca del comportamiento de los fumadores y los fumadores potenciales; en otras palabras, resulta necesario analizar los datos sobre el consumo de cigarrillos y los precios.

Los datos que se examinan son las ventas de cigarrillos, los precios, los impuestos, y la renta personal en los EE.UU en los años 80 y 90. En esos datos, los estados con impuestos bajos, y los que tienen precios de cigarrillos bajos, presentan una alta proporción de fumadores, y los estados con precios altos presentan una baja proporción de fumadores. Sin embargo, el análisis de estos datos resulta complicado debido a que la causalidad se mueve en ambos sentidos: los impuestos bajos llevan a una demanda elevada, pero si en el estado existen muchos fumadores, entonces los políticos locales pueden intentar mantener bajos los impuestos sobre los cigarrillos con el fin de satisfacer a sus fumadores. En el Capítulo 12 se tratan los métodos de estimación de la elasticidad-precio de la demanda de cigarrillos.

Pregunta #4 ¿Cuál será la tasa de inflación del año próximo?

Parece que la gente siempre desea una anticipación del futuro. ¿Cuáles serán las ventas el año próximo de una empresa que se plantea invertir en nuevo equipamiento? ¿Subirá la bolsa el próximo mes, y si es así, cuánto? ¿Cubrirá la recaudación fiscal del próximo año los gastos planificados de los servicios ciudadanos? ¿Se centrará el examen de microeconomía de la próxima semana en las externalidades o en los monopolios? ¿Hará un buen día el sábado para ir a la playa?

Un aspecto del futuro en el que los microeconomistas y los economistas financieros tienen un interés particular es la tasa de inflación general de los precios a lo largo del año próximo. Un profesional financiero puede aconsejar a un cliente acerca de si efectuar un préstamo o solicitarlo, dado un tipo de interés, dependiendo de su mejor predicción sobre la tasa de inflación del año siguiente. Los economistas del banco central como los del Consejo de la Reserva Federal en Washington D.C., y los del Banco Central Europeo en Fráncfort, Alemania, son los responsables de mantener la tasa de inflación de los precios bajo control, por lo que sus decisiones acerca de cómo establecer los tipos de interés se basan en la perspectiva de inflación para el próximo año. Si piensan que la tasa de inflación aumentará en un punto porcentual, entonces podrían aumentar más los tipos de interés para ralentizar una economía que, bajo su punto de vista, corre el riesgo de sobrecalentamiento. Si su predicción es mala, se arriesgan tanto a causar una recesión innecesaria como un aumento no deseado de la inflación.

Los economistas profesionales que confían en los pronósticos numéricos precisos utilizan los modelos econométricos para efectuar esos pronósticos. Un trabajo predictivo consiste en predecir el futuro utilizando el pasado, y los económetras lo hacen mediante la teoría económica y las técnicas estadísticas para cuantificar las relaciones entre los datos históricos.

Los datos que utilizamos para predecir la inflación son las tasas de inflación y el desempleo en los Estados Unidos. Una relación empírica importante en los datos macroeconómicos es la «curva de Phillips», en la que un valor bajo de la tasa de desempleo en la actualidad está asociado con un incremento en la tasa de inflación del año siguiente. Una de las predicciones de inflación que desarrollamos y evaluamos en el Capítulo 14 está basada en la curva de Phillips.

Preguntas cuantitativas, respuestas cuantitativas

Cada una de esas cuatro preguntas requiere una respuesta numérica. La teoría económica proporciona las claves sobre la respuesta, el consumo de cigarrillos debería reducirse al aumentar el precio, pero el valor numérico efectivo debe averiguarse empíricamente, esto es, mediante el análisis de los datos. Debido a que se utilizan datos para responder a las preguntas cuantitativas, las respuestas siempre poseen algún grado de incertidumbre: un conjunto diferente de datos podría dar lugar a diferentes respuestas numéricas. Por tanto, el marco conceptual para el análisis debe proporcionar tanto una respuesta numérica a la pregunta como una medida acerca de su precisión.

El marco conceptual utilizado en este libro es el modelo de regresión múltiple, el pilar principal de la econometría. Este modelo, introducido en la Parte II, proporciona un método matemático para cuantificar cómo afecta un cambio en una variable a otra variable, manteniendo el resto de cosas constantes. Por ejemplo, ¿qué efecto tiene un cambio en el tamaño de la clase sobre las calificaciones en los exámenes, *manteniendo constante o controlando por* las características de los estudiantes (tales como la renta familiar) que un administrador del distrito escolar no puede controlar? ¿Qué efectos tiene su etnia sobre sus oportunidades de obtener aprobada una solicitud de hipoteca, *manteniendo constantes* otros factores como su capacidad de devolver el préstamo? ¿Qué efecto causa un aumento del 1 % en el precio de los cigarrillos sobre el consumo de cigarrillos, *manteniendo constante* la renta tanto de los fumadores como de los fumadores potenciales? El modelo de regresión múltiple y sus extensiones proporcionan un marco para responder esas preguntas utilizando los datos y para cuantificar la incertidumbre asociada con esas respuestas.

1.2 Efectos causales y experimentos ideales

Como muchas de las cuestiones que aborda la econometría, las tres primeras preguntas de la Sección 1.1 se refieren a las relaciones causales entre variables. Comúnmente, se dice que una acción causa un resultado si el efecto es el resultado directo, o consecuencia, de esta acción. Tocar una estufa caliente causa quemaduras; beber agua causa sentirse menos sediento; poner aire en los neumáticos causa que se hinchen; poner fertilizante en las parcelas de tomate causa mayor producción de tomates. La causalidad significa que una acción específica (aplicar fertilizante) conlleva una consecuencia específica, medible (más tomates).

Estimación de los efectos causales

¿Cómo podemos medir mejor los efectos causales en el rendimiento de los tomates (medido en kilogramos) de aplicar una cierta cantidad de fertilizante, por ejemplo, 100 gramos de fertilizante por metro cuadrado?

Un modo de medir este efecto causal es llevar a cabo un experimento. En este experimento, un investigador hortícola planta muchas parcelas de tomates. Cada parcela es atendida de forma idéntica, con una sola excepción: algunas parcelas reciben 100 gramos de fertilizante por metro cuadrado, mientras el resto no recibe nada. Además, el hecho de que una parcela sea fertilizada o no se determina aleatoriamente mediante un ordenador, asegurando que cualquier otra diferencia entre las parcelas no tenga relación con el hecho de recibir fertilizante. Al final de la época de crecimiento, el horticultor pesa la cosecha de cada parcela. La

diferencia entre el rendimiento medio por metro cuadrado de las parcelas tratadas y no tratadas es el efecto sobre la producción de tomates del tratamiento con fertilizante.

Esto es un ejemplo de **experimento aleatorizado controlado**. Es controlado en el sentido de que existe tanto un **grupo de control** que no recibe tratamiento (sin fertilizante) como un **grupo de tratamiento** que recibe el tratamiento (100 g/m^2). Es aleatorizado en el sentido de que el tratamiento se asigna aleatoriamente. Esta asignación aleatoria elimina la posibilidad de una relación sistemática entre, por ejemplo, cuánto sol recibe la parcela y si recibe fertilizante, por lo que la única diferencia sistemática entre el grupo de tratamiento y el de control es el tratamiento. Si este experimento se implementa adecuadamente a una escala suficientemente grande, arrojará una estimación del efecto causal sobre el resultado de interés (la producción de tomates) del tratamiento (aplicar 100 g/m^2 de fertilizante).

En este libro, el **efecto causal** se define como el efecto sobre un resultado de una acción dada o tratamiento medido en un experimento aleatorizado controlado ideal. En ese experimento, la única razón sistemática para la diferencia en los resultados entre los grupos de tratamiento y de control es el tratamiento en sí mismo.

Es posible imaginar un experimento aleatorizado controlado ideal para responder a cada una de las tres primeras preguntas de la Sección 1.1. Por ejemplo, para estudiar el tamaño de las clases se puede imaginar una asignación aleatoria de «tratamientos» o diferentes tamaños de clase para diferentes grupos de estudiantes. Si el experimento se diseña y se lleva a cabo de ese modo, la única diferencia sistemática entre los grupos de estudiantes es el tamaño de su clase, de esta forma, en teoría, este experimento estimaría el efecto sobre las calificaciones en los exámenes de la reducción del tamaño de las clases, manteniendo todo lo demás constante.

El concepto de experimento aleatorizado controlado ideal es útil porque proporciona una definición del efecto causal. En la práctica, sin embargo, no es posible llevar a cabo experimentos ideales. De hecho, los experimentos son escasos en econometría debido a que a menudo son inmorales, imposibles de ejecutar satisfactoriamente, o prohibitivamente caros. El concepto de experimento aleatorizado controlado ideal, sin embargo, proporciona una referencia teórica para el análisis econométrico de los efectos causales utilizando datos reales.

Predicción y causalidad

Aunque las tres primeras preguntas de la Sección 1.1 se refieren a efectos causales, la cuarta, la predicción de la inflación, no. No es necesario conocer una relación causal para realizar una buena predicción. Una buena manera de «predecir» si llueve es observar si los peatones están usando paraguas, pero el acto de usar un paraguas no causa que llueva.

Incluso aunque la predicción no necesite incluir relaciones causales, la teoría económica sugiere patrones y relaciones que pueden ser útiles para la predicción. Como veremos en el Capítulo 14, el análisis de regresión múltiple permite cuantificar las relaciones históricas que sugiere la teoría económica, para verificar si esas relaciones han sido estables en el tiempo, para realizar predicciones cuantitativas sobre el futuro, y para establecer la precisión de esas predicciones.

1.3 Datos: fuentes y tipos

En econometría, los datos provienen de dos fuentes: experimentos u observaciones del mundo no experimentales. Este libro examina tanto las bases de datos experimentales como las no experimentales.

Datos experimentales *versus* datos observacionales

Los **datos experimentales** provienen de experimentos diseñados para evaluar un tratamiento o política o investigar un efecto causal. Por ejemplo, el estado de Tennessee financió un gran experimento aleatorizado controlado que examinaba el tamaño de las clases en los años 80. En este experimento, que se examina en el Capítulo 13, miles de estudiantes fueron asignados aleatoriamente a clases de diferentes tamaños durante algunos años y se llevaron a cabo los exámenes estandarizados anuales.

El experimento del tamaño de las clases de Tennessee costó millones de dólares y necesitó la cooperación continuada de muchas administraciones, padres y maestros durante algunos años. Debido a que los experimentos en el mundo real con personas son difíciles de administrar y controlar, presentan fallos con respecto a los experimentos aleatorizados controlados ideales. Además, en algunas circunstancias los experimentos no solo son caros y difíciles de administrar, sino que además son poco éticos. (¿Sería ético ofrecer a adolescentes aleatoriamente seleccionados cigarrillos baratos para ver cuántos compran?). Debido a estos problemas financieros, políticos y éticos, los experimentos en economía son escasos. Por el contrario, la mayoría de los datos económicos se obtienen mediante observaciones del comportamiento del mundo real.

Los datos obtenidos mediante la observación del comportamiento real fuera de un marco experimental se denominan **datos observacionales**. Los datos observacionales se recopilan utilizando encuestas, tales como una encuesta telefónica a los consumidores, y registros administrativos, tales como los registros históricos sobre solicitudes de préstamos hipotecarios en manos de las instituciones de préstamo.

Los datos observacionales presentan un importante desafío a los intentos econométricos de estimar efectos causales, y a las herramientas econométricas que abordan estos desafíos. En el mundo real, los niveles de «tratamiento» (la cantidad de fertilizante en el ejemplo del tomate, la ratio estudiantes-maestros en el ejemplo del tamaño de las clases) no se asignan aleatoriamente, por tanto resulta difícil aislar el efecto del «tratamiento» del de otros factores relevantes. La mayor parte de la econometría, y la mayor parte de este libro, está dedicada a los métodos que existen para encarar los desafíos planteados cuando se utilizan datos del mundo real para estimar efectos causales.

Tanto si los datos son datos experimentales como si son observacionales, las bases de datos son de tres tipos principales: datos de sección cruzada, datos de series temporales y datos de panel. En este libro nos encontraremos con los tres tipos de datos.

Datos de sección cruzada

Los datos de individuos o entidades diferentes —trabajadores, consumidores, empresas, administraciones públicas, etc.— para un único periodo de tiempo se denominan **datos de sección cruzada**. Por ejemplo, los datos de las calificaciones en los exámenes estandarizados de los distritos escolares de California son datos de sección cruzada. Estos datos son para 420 individuos (distritos escolares) en un único periodo de tiempo (1999). En general, el número de individuos para los cuales se dispone de observaciones se expresa como n ; como por ejemplo, en la base de datos de California, $n = 420$.

La base de datos de las calificaciones en el examen de California contiene medidas de algunas variables diferentes para cada distrito. Algunas de ellas están tabuladas en la Tabla 1.1. Cada fila presenta los datos para un distrito diferente. Por ejemplo la calificación media para el primer distrito («distrito #1») es 690,8; esta es la media de las calificaciones en el examen de matemáticas y ciencias para todos los estudiantes de quinto curso en este distrito en 1999 en un examen estandarizado (La Prueba de prueba de logros de Stanford). La media de la ratio estudiantes-maestros en ese distrito es 17,89; es decir, el número de estudiantes del distrito #1 dividido por el número de maestros en el aula del distrito #1 es 17,89. La media de gasto por alumno en el distrito #1 es 6.385 \$. El porcentaje de estudiantes que en este distrito aún no saben inglés —esto es, el porcentaje de estudiantes para los que el inglés es una segunda lengua y que aún no dominan el idioma inglés— es 0 %.

Las filas restantes presentan los datos de otros distritos. El orden de la fila es arbitrario, y el número del distrito, que se denomina **número de observación**, es un número asignado arbitrariamente que sirve para organizar los datos. Como se puede observar en la tabla, todas las variables mostradas varían de forma considerable.

Con datos de sección cruzada, podemos aprender mucho sobre las relaciones entre las variables estudiando las diferencias entre personas, empresas, u otras entidades económicas durante un único periodo de tiempo.

Datos de series temporales

Los datos de series temporales son datos para un único individuo o entidad (persona, empresa, país) recogidos para múltiples periodos. Nuestra base de datos sobre la tasa de inflación y desempleo en los Esta-

TABLA 1.1 Observaciones seleccionadas sobre calificaciones en los exámenes y otras variables para los distritos escolares de California en 1999

Número de observación (distrito)	Calificación media del distrito en el examen	Ratio estudiantes-maestros	Gasto por estudiante (\$)	Porcentaje de estudiantes aprendiendo inglés
1	690,8	17,89	6.385	0,0
2	661,2	21,52	5.099	4,6
3	643,6	18,70	5.502	30,0
4	647,7	17,36	7.102	0,0
5	640,8	18,67	5.236	13,9
:	:	:	:	:
418	645,0	21,89	4.403	24,3
419	672,2	20,20	4.776	3,0
420	655,8	19,04	5.993	5,0

Nota: La base de datos de las calificaciones en el examen de California se describe en el Apéndice 4.1.

dos Unidos es un ejemplo de conjunto de datos de series temporales. La base de datos contiene observaciones sobre dos variables (las tasas de inflación y desempleo) para un único individuo (Estados Unidos) para 183 períodos. Cada periodo de tiempo en esta base de datos es un trimestre del año (el primer trimestre es enero, febrero y marzo; el segundo trimestre es abril, mayo y junio; etc.). Las observaciones de esta base de datos comienzan en el segundo trimestre de 1959, que se expresa como 1959:II, y termina en el cuarto trimestre de 2004 (2004:4). El número de observaciones (es decir, períodos de tiempo) de un conjunto de datos de series temporales se expresa como T . Como hay 183 trimestres desde 1959:II hasta 2004:IV, esta base de datos contiene $T = 183$ observaciones.

En la Tabla 1.2 se presentan algunas observaciones de esta base de datos. Los datos de cada fila corresponden a un periodo de tiempo diferente (año y trimestre). En el segundo trimestre de 1959, por ejemplo, la tasa de inflación fue del 0,7 % por año en tasa interanual. En otras palabras, si la inflación hubiera continua-

TABLA 1.2 Observaciones seleccionadas sobre la tasa de inflación del Índice de Precios al Consumo (IPC) y desempleo en los Estados Unidos: datos trimestrales, 1959-2004

Número de observación	Fecha (año:trimestre)	Tasa de inflación IPC (% de variación en un año en tasa interanual) (%)	Tasa de desempleo (%)
1	1959:II	0,7%	5,1%
2	1959:III	2,1	5,3
3	1959:IV	2,4	5,6
4	1960:I	0,4	5,1
5	1960:II	2,4	5,2
:	:	:	:
181	2004:II	4,3	5,6
182	2004:III	1,6	5,4
183	2004:IV	3,5	5,4

Nota: La inflación y el desempleo en EE.UU. se describen en el Apéndice 14.1.

do durante 12 meses en la tasa del segundo trimestre de 1959, el nivel global de precios (medido por el Índice de Precios al Consumo, IPC) habría aumentado un 0,7 %. En el segundo trimestre de 1959, la tasa de desempleo fue de 5,1 %, esto es, el 5,1 % de la fuerza laboral declara que no tuvo un trabajo aunque lo buscaba. En el tercer trimestre de 1959, la tasa de inflación del IPC fue del 2,1 %, y la tasa de desempleo fue del 5,3 %.

Rastreando un único individuo en el tiempo, los datos de series temporales pueden utilizarse para estudiar la evolución de las variables en el tiempo y para predecir los valores futuros de esas variables.

Datos de panel

Los **datos de panel**, asimismo denominados **datos longitudinales**, son datos sobre varios individuos en los que cada individuo se observa durante uno, dos o más períodos de tiempo. Nuestros datos sobre el consumo de cigarrillos y precios son un ejemplo de base de datos de panel, y en la Tabla 1.3 se muestran algunas variables seleccionadas, así como algunas observaciones. El número de individuos en una base de datos de panel se expresa como n , y el número de períodos de tiempo se expresa con T . En la base de datos de cigarrillos, tenemos observaciones sobre $n = 48$ estados continentales de EE.UU. (individuos) para $T = 11$ años (períodos de tiempo) desde 1985 a 1995. Por tanto existen un total de $n \times T = 48 \times 11 = 528$ observaciones.

Se recogen algunos de los datos de la base de datos de consumo de cigarrillos en la Tabla 1.3. El primer bloque de 48 observaciones recoge los datos para cada estado en 1985, ordenados alfabéticamente desde Alabama a Wyoming. El siguiente bloque de 48 observaciones presenta los datos para 1986, y sucesivamente, hasta 1995. Por ejemplo, en 1985, las ventas de cigarrillos en Arkansas fueron de 128,5 paquetes per cápita (el número total de paquetes de cigarrillos vendidos en Arkansas en 1985 dividido por la población total de Arkansas en 1985 es igual a 128,5). El precio medio de un paquete de cigarrillos en Arkansas en 1985, incluyendo impuestos, era de 1,105 \$. De los cuales 37 ¢ eran impuestos federales, estatales y locales.

Los datos de panel pueden utilizarse para conocer las relaciones económicas a partir de las experiencias de muchos individuos diferentes de la base de datos y de la evolución en el tiempo de las variables para cada individuo.

TABLA 1.3 Observaciones seleccionadas sobre la venta de cigarrillos, precios e impuestos, por estado y año para los estados de EE.UU., 1985-1995

Número de observación	Estado	Año	Ventas de cigarrillos (paquetes per cápita)	Precio medio por paquete (incluyendo impuestos) (\$)	Impuestos Totales (impuesto sobre cigarrillos + impuesto sobre ventas) (\$)
1	Alabama	1985	116,5	1,022	0,333
2	Arkansas	1985	128,5	1,015	0,370
3	Arizona	1985	104,5	1,086	0,362
⋮	⋮	⋮	⋮	⋮	⋮
47	Virginia occidental	1985	112,8	1,089	0,382
48	Wyoming	1985	129,4	0,935	0,240
49	Alabama	1986	117,2	1,080	0,334
⋮	⋮	⋮	⋮	⋮	⋮
96	Wyoming	1986	127,8	1,007	0,240
97	Alabama	1987	115,8	1,135	0,335
⋮	⋮	⋮	⋮	⋮	⋮
528	Wyoming	1995	112,2	1,585	0,360

Nota: La base de datos del consumo de cigarrillos se describe en el Apéndice 12.1.

**CONCEPTO
CLAVE**
1.1**Datos de sección cruzada, series temporales y panel**

- Los datos de sección cruzada consisten en múltiples individuos observados para un único periodo de tiempo.
- Los datos de series temporales consisten en un único individuo observado durante múltiples periodos de tiempo.
- Los datos de panel (asimismo conocidos como datos longitudinales) consisten en múltiples individuos, en los que cada individuo es observado durante dos o más períodos de tiempo.

Las definiciones de datos de sección cruzada, datos de series temporales, datos de panel están recogidas en el Concepto clave 1.1.

Resumen

1. Muchas de las decisiones en los negocios y en economía requieren estimaciones cuantitativas sobre cuánto afecta el cambio en una variable a otra variable.
2. Conceptualmente, el modo de estimar un efecto causal es un experimento aleatorizado controlado ideal, pero llevar a cabo tales experimentos en cada caso económico concreto o aplicación, resulta habitualmente poco ético, poco práctico, o demasiado costoso.
3. La econometría proporciona herramientas para estimar los efectos causales utilizando tanto datos observacionales (no experimentales) como los datos del mundo real, de experimentos imperfectos.
4. Los datos de sección cruzada se recopilan a partir de la observación de varios individuos en un único periodo de tiempo; los datos de series temporales se recogen a partir de la observación de un único individuo en distintos períodos de tiempo; y los datos de panel se recogen a partir de la observación de varios individuos, cada uno de los cuales se observa durante múltiples períodos de tiempo.

Términos clave

experimento aleatorizado controlado (5)
 grupo de control (5)
 grupo de tratamiento (5)
 efecto causal (5)
 datos experimentales (5)
 datos observacionales (6)

datos de sección cruzada (6)
 número de observación (6)
 datos de series temporales (6)
 datos de panel (8)
 datos longitudinales (8)

Revisión de conceptos

- 1.1** Diseñe un experimento aleatorizado controlado ideal hipotético para estudiar el efecto de las horas dedicadas al estudio sobre el desempeño en los exámenes de microeconomía. Sugiera algunos de los impedimentos que pueden surgir a la hora de llevar a cabo este experimento a la práctica.
- 1.2** Diseñe un experimento aleatorizado controlado ideal hipotético para estudiar el efecto sobre la mortalidad en accidentes de tráfico en carretera del uso del cinturón de seguridad. Sugiera algunos de los impedimentos que pueden surgir a la hora de llevar a cabo este experimento a la práctica.
- 1.3** Se le ha solicitado la elaboración de un estudio sobre el efecto causal de las horas dedicadas a la formación de los empleados (medido en horas por trabajador a la semana) en una fábrica sobre la productividad de sus trabajadores (producción por trabajador y hora). Describa:
 - a)** Un experimento aleatorizado controlado ideal para medir este efecto causal.
 - b)** Una base de datos observacionales de sección cruzada con los cuales podría estudiar este efecto.
 - c)** Una base de datos observacionales de series temporales para el estudio de este efecto.
 - d)** Una base de datos de panel para el estudio de este efecto.

2

Repasso de probabilidad

Este capítulo repasa las ideas centrales de la teoría de probabilidad que resultan necesarias para comprender el análisis de regresión y la econometría. Suponemos que con anterioridad se ha cursado una asignatura de probabilidad y estadística. Si se han olvidado los conocimientos sobre probabilidad, deberían ser refrescados con la lectura de este capítulo. Si se siente que se domina el material, debería igualmente echarse un vistazo al capítulo y a los términos y conceptos del final del capítulo para asegurarse de que se está familiarizado con las ideas y la notación.

La mayoría de los aspectos del mundo que nos rodea presenta algún elemento de aleatoriedad. La teoría de probabilidad proporciona herramientas matemáticas para cuantificar y describir esta aleatoriedad. La Sección 2.1 repasa las distribuciones de probabilidad para una única variable aleatoria, y la Sección 2.2 trata la esperanza matemática, la media, y la varianza de una única variable aleatoria. La mayoría de los problemas económicos de interés incluyen más de una variable, por lo que la Sección 2.3 introduce los elementos básicos de teoría de probabilidad para dos variables aleatorias. La Sección 2.4 estudia las tres distribuciones de probabilidad específicas que interpretan un papel central en estadística y en econometría: las distribuciones normal, chi cuadrado, y F .

Las dos secciones finales de este capítulo se centran en una fuente específica de aleatoriedad de importancia central en econometría: la aleatoriedad que aparece al realizar la selección al azar de una muestra de datos a partir de una población mayor. Por ejemplo, supongamos que se encuesta a diez recién graduados universitarios seleccionados aleatoriamente, se registran (u «observan») sus ingresos, y se calcula los ingresos promedio a partir de esas diez coordenadas (u «observaciones»). Dado que la muestra se elige aleatoriamente, podrían haberse elegido diez graduados diferentes por puro azar; una vez hecho, se habrían observado diez ingresos diferentes y se habría calculado una media muestral diferente. Al variar el promedio de los ingresos de una muestra aleatoriamente seleccionada respecto de la siguiente, la media muestral es en sí misma una variable aleatoria.

Por tanto, la media muestral posee una distribución de probabilidad, lo que se refiere a su distribución muestral debido a que esta distribución describe los diferentes valores posibles que puede tomar la media muestral habiendo seleccionado diferentes muestras.

La Sección 2.5 trata el muestreo aleatorio y la distribución muestral de la media muestral. La distribución muestral es, en general, complicada. Sin embargo, cuando el tamaño muestral es lo suficientemente grande, la distribución muestra de la media muestral es aproximadamente normal, un resultado conocido como el teorema central del límite, que se presenta en la Sección 2.6.

2.1 Variables aleatorias y distribuciones de probabilidad

Probabilidades, espacio muestral y variables aleatorias

Probabilidades y resultados. El género de la próxima persona nueva que conozca, su calificación en un examen y el número de veces que su ordenador se estropeará mientras redacta un trabajo, presentan todos ellos una componente de azar o aleatoriedad. En cada uno de esos ejemplos, existe algo que no es todavía conocido pero que a la larga se revelará.

Los resultados potenciales mutuamente excluyentes de un proceso aleatorio se denominan **resultados**. Por ejemplo, su ordenador puede no estropearse nunca, puede estropearse una vez, puede estropearse dos veces, etc. Solo uno de esos resultados puede ocurrir en la realidad (los resultados son mutuamente excluyentes), y los resultados no necesariamente son igualmente probables.

La **probabilidad** de un resultado es la proporción de veces que el resultado ocurre en el largo plazo. Si la probabilidad de que su ordenador no se estropee mientras redacta un trabajo es del 80 %, entonces durante el proceso de redactar muchos trabajos el 80 % de las veces terminará sin averías.

Espacio muestral y sucesos. El conjunto de todos los posibles resultados se denomina **espacio muestral**. Un **suceso** es un subconjunto del espacio muestral, esto es, un suceso es un conjunto de uno o más resultados. El suceso «mi ordenador no se estropeará más de una vez» es el conjunto que consta de dos resultados: «sin averías» y «una avería».

Variables aleatorias. Una variable aleatoria discreta es un resumen numérico de un resultado aleatorio. El número de veces que su ordenador se estropea mientras redacta un trabajo es aleatorio y toma un valor numérico, por lo que es una variable aleatoria.

Algunas variables aleatorias son discretas y otras son continuas. Como sus nombres sugieren, una **variable aleatoria discreta** toma valores solamente sobre un conjunto discreto, como 0, 1, 2, ..., mientras que una **variable aleatoria continua** toma valores en un continuo de posibles valores.

Distribución de probabilidad de una variable aleatoria discreta

Distribución de probabilidad. La **distribución de probabilidad** de una variable aleatoria discreta es una relación de todos los valores posibles de la variable junto con la probabilidad de que ocurra cada valor. Esas probabilidades suman 1.

Por ejemplo, sea M el número de veces que su ordenador se estropea mientras redacta un trabajo. La distribución de probabilidad de una variable aleatoria M es la relación de probabilidades de cada posible resultado: la probabilidad de que $M = 0$, que se expresa como $\Pr(M = 0)$, es la probabilidad de que el ordenador no se estropee; $\Pr(M = 1)$ es la probabilidad de una única avería en el ordenador; y así sucesivamente. Se ofrece un ejemplo de una distribución de probabilidad para M en la segunda fila de la Tabla 2.1; en esta distribución, si su ordenador se estropea cuatro veces, deberá abandonar y redactar el trabajo a mano. De acuerdo con esta distribución, la probabilidad de que no haya averías es del 80 %; la probabilidad de una avería es del 10 %; y la probabilidad de dos, tres, o cuatro averías es, respectivamente, 6 %, 3 %, y 1 %. Esas probabilidades suman el 100 %. Esta distribución de probabilidad está representada en la Figura 2.1.

Probabilidad de los sucesos. La probabilidad de un suceso puede calcularse a partir de la distribución de probabilidad. Por ejemplo, la probabilidad del suceso una o dos averías es la suma de probabilidades de los resultados de los que consta el suceso. Esto es $\Pr(M = 1 \text{ o } M = 2) = \Pr(M = 1) + \Pr(M = 2) = 0,10 + 0,06 = 0,16$, o 16 %.

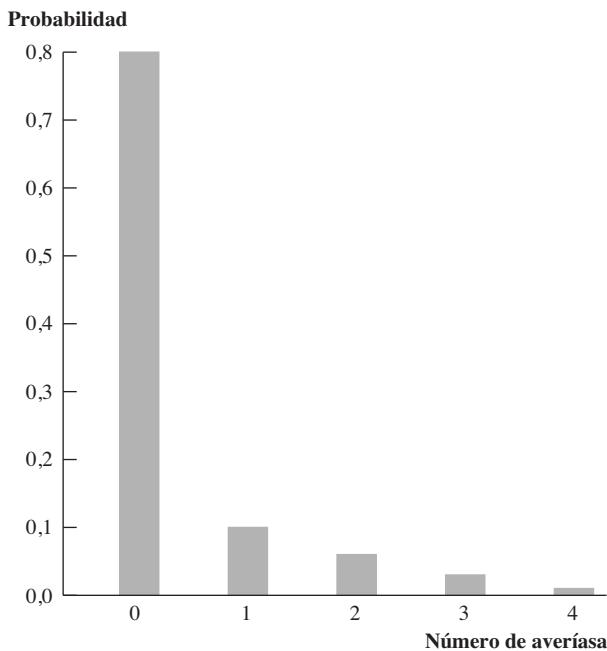
Distribución de probabilidad acumulada. La **distribución de probabilidad acumulada** es la probabilidad de que la variable aleatoria sea menor o igual a un valor concreto. La última fila de la Tabla 2.1 proporciona la distribución de probabilidad acumulada de la variable aleatoria M . Por ejemplo, la probabilidad de al menos una avería, $\Pr(M \leq 1)$, es del 90 %, que es la suma de las probabilidades de que no se averíe (80 %) y de una avería (10 %).

TABLA 2.1 Probabilidad de que su ordenador se averíe M veces

	Resultados (número de averías)				
	0	1	2	3	4
Distribución de probabilidad	0,80	0,10	0,06	0,03	0,01
Distribución de probabilidad acumulada	0,80	0,90	0,96	0,99	1,00

FIGURA 2.1 Distribución de probabilidad del número de averías de ordenador

La altura de cada una de las barras es la probabilidad de que el ordenador se averíe el número de veces indicado. La altura de la primera barra es 0,8, por tanto la probabilidad de 0 averías en el ordenador es del 80 %. La altura de la segunda barra es 0,1, por lo que la probabilidad de 1 avería en el ordenador es del 10 %, y lo mismo para el resto de barras.



La distribución de probabilidad acumulada se conoce además como **función de distribución acumulada, f.d.a., o distribución acumulada**.

La distribución de Bernouilli. Un caso particular importante de variable aleatoria discreta es cuando la variable aleatoria es binaria, es decir, los resultados son 0 o 1. Una variable aleatoria binaria se denomina **variable aleatoria de Bernouilli** (en honor al matemático y científico suizo del siglo XVII Jacob Bernouilli), y su distribución de probabilidad se denomina **distribución de Bernouilli**.

Por ejemplo, sea G el género de la próxima persona nueva que conozca, donde $G = 0$ indica que la persona será un varón y $G = 1$ indica que será mujer. Los resultados de G y sus probabilidades son

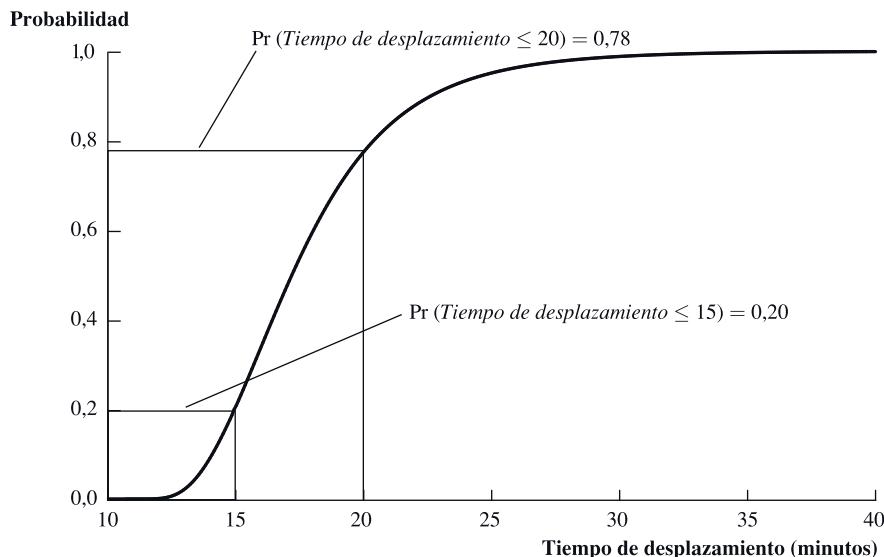
$$G = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p, \end{cases} \quad (2.1)$$

donde p es la probabilidad de que la próxima persona nueva que conozca sea mujer. La distribución de probabilidad de la Ecuación (2.1) es la distribución de Bernouilli.

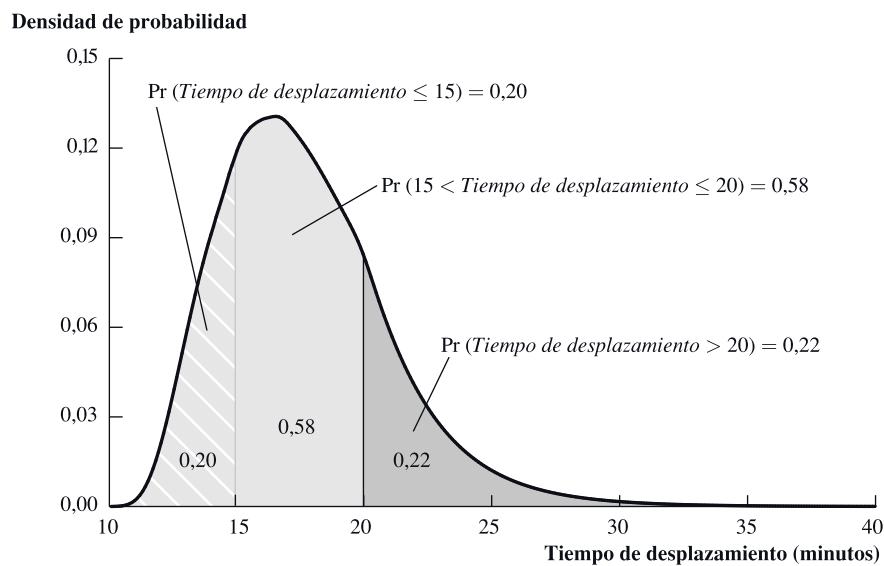
Distribución de probabilidad de una variable aleatoria continua

Distribución de probabilidad acumulada. La distribución de probabilidad acumulada de una variable aleatoria continua se define exactamente igual a como se hizo para una variable aleatoria discreta. Es decir, la distribución de probabilidad acumulada de una variable aleatoria es la probabilidad de que la variable aleatoria sea menor o igual que un valor concreto.

Por ejemplo, considere una estudiante que conduce desde su casa a la escuela. El tiempo que lleva este desplazamiento de la estudiante puede tomar un continuo de valores y, como depende de factores aleatorios tales como la meteorología y las condiciones del tráfico, es natural tratarlo como una variable aleatoria continua. La Figura 2.2a representa una distribución de probabilidad acumulada hipotética del tiempo de los desplazamientos. Por ejemplo, la probabilidad de que el desplazamiento de ida y vuelta dure menos de 15 minutos es del 20 % y la probabilidad de que dure menos de 20 minutos es del 78 %.

FIGURA 2.2 Funciones de distribución acumulada y de densidad de probabilidad del tiempo de desplazamiento

(a) Función de distribución acumulada del tiempo de desplazamiento



(b) Función de densidad de probabilidad del tiempo de desplazamiento

La Figura 2.2a muestra la función de distribución acumulada (o f.d.a.) del tiempo de desplazamiento. La probabilidad de que el tiempo de desplazamiento sea menos de 15 minutos es 0,20 (o 20 %), y la probabilidad de que sea menos de 20 minutos es 0,78 (78 %).

La Figura 2.2b muestra la función de densidad de probabilidad (o f.d.p.) del tiempo de desplazamiento. Las probabilidades vienen dadas por las áreas bajo la f.d.p. La probabilidad de que el tiempo de un desplazamiento se encuentre entre 15 y 20 minutos es 0,58 (58 %) y está dada por el área bajo la curva entre 15 y 20 minutos.

Función de densidad de probabilidad. Debido a que una variable aleatoria continua puede tomar sus valores posibles en un continuo, la distribución de probabilidad utilizada para las variables discretas, que presenta la probabilidad de cada posible valor de la variable aleatoria, no es aplicable a las variables continuas. En su lugar, la probabilidad viene recogida por la **función de densidad de probabilidad**. El área bajo la función de densidad de probabilidad entre cualesquiera dos puntos es la probabilidad de que la variable aleatoria se encuentre entre esos dos puntos. Una función de densidad de probabilidad se denomina asimismo como **f.d.p., función de densidad, o simplemente densidad**.

La Figura 2.2b representa la función de densidad de probabilidad del tiempo de ida y vuelta correspondiente a la distribución acumulada de la Figura 2.2a. La probabilidad de que el desplazamiento de ida y vuelta lleve entre 15 y 20 minutos, está dada por el área bajo la f.d.p. que es 0,58, o 58 %. De forma equivalente, esta probabilidad puede verse en la distribución acumulada de la Figura 2.2a como la diferencia entre la probabilidad de que el desplazamiento de ida y vuelta sea menor a 20 minutos (78 %) y la probabilidad de que sea menor a 15 minutos (20 %). Por tanto, la distribución de densidad de probabilidad y la función de probabilidad acumulada muestran la misma información en diferentes formatos.

2.2 Esperanza, media y varianza

La esperanza de una variable aleatoria

Esperanza. La **esperanza** de una variable aleatoria Y , denominada $E(Y)$, es el valor medio de largo plazo de la variable aleatoria a lo largo de muchos intentos repetidos o eventos. La esperanza de una variable aleatoria discreta se calcula como la media ponderada de los posibles resultados de la variable aleatoria, donde las ponderaciones son las probabilidades de esos resultados. La esperanza de Y se denomina asimismo **valor esperado** de Y o **media** de Y y se expresa mediante μ_Y .

Por ejemplo, suponga que presta a un amigo 100 \$ al 10 % de interés. Si el préstamo se devuelve, obtendrá 110 \$ (el principal de 100 \$ más los intereses de 10 \$), pero existe un riesgo del 1 % de que su amigo incumpla el pago y usted no reciba nada en absoluto. Por tanto, el montante que percibirá es una variable aleatoria que es igual a 110 \$ con probabilidad 0,99 e igual a 0 \$ con probabilidad 0,01. A lo largo de muchos préstamos así, el 99 % de las veces percibirá 110 \$, pero el 1 % de las veces no obtendrá nada, por tanto en media percibirá $110 \times 0,99 + 0 \times 0,01 = 108,90$ \$. Por tanto, el valor esperado del reembolso (o el «reembolso medio») es 108,90 \$.

Como segundo ejemplo, considérese el número de averías de ordenador M con la distribución de probabilidad dada en la Tabla 2.1. El valor esperado de M es el número medio de averías a lo largo de la realización de muchos trabajos, ponderado por la frecuencia con la que ocurre una avería de una determinada importancia. Por tanto,

$$E(M) = 0 \times 0,80 + 1 \times 0,10 + 2 \times 0,06 + 3 \times 0,03 + 4 \times 0,01 = 0,35. \quad (2.2)$$

Es decir, el número esperado de averías de ordenador durante la redacción de un trabajo es 0,35. Por supuesto, el número real de averías debe ser siempre un entero; ¡no tiene sentido decir que el ordenador se averió 0,35 veces durante la redacción de un trabajo! En cambio, el cálculo de la Ecuación (2.2) da a entender que el número medio de averías a lo largo de la redacción de muchos trabajos es 0,35.

La fórmula para la esperanza de una variable aleatoria discreta Y que puede tomar k valores diferentes viene dada en el Concepto clave 2.1. (El Concepto clave 2.1 utiliza la «notación en forma de sumatorios», que se repasa en el Ejercicio 2.25).

**CONCEPTO
CLAVE**

2.1

Esperanza y media

Suponga que la variable aleatoria Y puede tomar k posibles valores y_1, y_2, \dots, y_k , donde y_1 expresa el primer valor, y_2 expresa el segundo valor, y así sucesivamente, la probabilidad de que Y tome el valor y_1 es p_1 , la probabilidad de que Y tome y_2 es p_2 , etc. La esperanza o valor esperado de Y , expresado mediante $E(Y)$, es

$$E(Y) = y_1 p_1 + y_2 p_2 + \dots + y_k p_k = \sum_{i=1}^k y_i p_i, \quad (2.3)$$

donde la notación $\sum_{i=1}^k y_i p_i$ significa «la suma de $y_i p_i$ con i tomando valores de 1 a k ». La esperanza de Y se denomina asimismo media de Y o valor esperado de Y y se expresa mediante μ_Y .

Esperanza de una variable aleatoria de Bernouilli. Un caso particular importante de la fórmula general del Concepto clave 2.1 es la media de una variable aleatoria de Bernouilli. Sea G la variable aleatoria de Bernouilli con la distribución de probabilidad de la Ecuación (2.1). El valor esperado de G es

$$E(G) = 1 \times p + 0 \times (1 - p) = p. \quad (2.4)$$

Por lo que el valor esperado de una variable aleatoria de Bernouilli es p , la probabilidad de que tome el valor «1».

Esperanza de una variable aleatoria continua. El valor esperado de una variable aleatoria continua es asimismo la media ponderada por la probabilidad de los posibles resultados de la variable aleatoria. Debido a que una variable aleatoria continua toma los posibles valores en un continuo, la definición matemática formal de su esperanza implica cálculo, su definición se ofrece en el Apéndice 17.1

La desviación típica y la varianza

La varianza y la desviación típica miden la dispersión o «difusión» de una distribución de probabilidad. La **varianza** de una variable aleatoria Y , que viene expresada por $\text{var}(Y)$, es el valor esperado del cuadrado de la desviación de Y respecto de su media: $\text{var}(Y) = E[(Y - \mu_Y)^2]$.

A causa de que la varianza incluye el cuadrado de Y , las unidades de la varianza son las unidades de Y al cuadrado, lo que da lugar a que la varianza tenga una interpretación complicada. Por tanto es habitual medir la dispersión mediante la **desviación típica**, que es la raíz cuadrada de la varianza y se expresa mediante σ_Y . La desviación típica tiene las mismas unidades que Y . Estas definiciones se recogen en el Concepto clave 2.2.

**CONCEPTO
CLAVE**

2.2

Varianza y desviación típica

La varianza de la variable aleatoria discreta Y , expresada mediante σ_Y^2 , es

$$\sigma_Y^2 = \text{var}(Y) = E[(Y - \mu_Y)^2] = \sum_{i=1}^k (y_i - \mu_Y)^2 p_i. \quad (2.5)$$

La desviación típica de Y es σ_Y , la raíz cuadrada de la varianza. Las unidades de la desviación típica son las mismas que las unidades de Y .

Por ejemplo, la varianza del número de averías de ordenador M es la media ponderada por probabilidad de las diferencias al cuadrado entre M y su media, 0,35.

$$\begin{aligned} \text{var}(M) &= (0 - 0,35)^2 \times 0,80 + (1 - 0,35)^2 \times 0,10 + (2 - 0,35)^2 \times 0,06 \\ &\quad + (3 - 0,35)^2 \times 0,03 + (4 - 0,35)^2 \times 0,01 = 0,6475. \end{aligned} \quad (2.6)$$

La desviación típica de M es la raíz cuadrada de la varianza, así $\sigma_M = \sqrt{0,6475} \cong 0,80$.

Varianza de una variable aleatoria de Bernouilli. La media de una variable aleatoria de Bernouilli G con la distribución de probabilidad recogida en la Ecuación (2.1) es $\mu_G = p$ [Ecuación (2.4)], por lo que su varianza es

$$\text{var}(G) = \sigma_G^2 = (0 - p)^2 \times (1 - p) + (1 - p)^2 \times p = p(1 - p). \quad (2.7)$$

Por tanto la desviación típica de una variable aleatoria de Bernouilli es $\sigma_G = \sqrt{p(1 - p)}$.

Media y varianza de una función lineal de variables aleatorias

Esta sección aborda el estudio de variables aleatorias (digamos, X e Y) que están relacionadas mediante una función lineal. Por ejemplo, considérese un esquema impositivo sobre la renta bajo el cual un trabajador es gravado a un tipo del 20 % sobre sus ingresos y percibe posteriormente una subvención (libre de impues-

tos) de 2.000 \$. Bajo este esquema impositivo, los ingresos después de impuestos Y están relacionados con los ingresos antes de impuestos X , mediante la ecuación

$$Y = 2.000 + 0,8X. \quad (2.8)$$

Es decir, los ingresos después de impuestos Y son el 80 % de los ingresos antes de impuestos, más 2.000 \$.

Supongamos que los ingresos de un individuo antes de impuestos para el próximo año son una variable aleatoria con media μ_X y varianza σ_X^2 . Dado que los ingresos antes de impuestos son una variable aleatoria, lo son del mismo modo los ingresos después de impuestos. ¿Cuáles son las medias y las desviaciones típicas de los ingresos después de impuestos con estos impuestos? Tras los impuestos, sus ingresos son el 80 % de los ingresos originales antes de impuestos, más 2.000 \$. Por tanto la esperanza de los ingresos después de impuestos es

$$E(Y) = \mu_Y = 2.000 + 0,8\mu_X. \quad (2.9)$$

La varianza de los ingresos antes de impuestos es la esperanza de $(Y - \mu_Y)^2$. Debido a que $Y = 2.000 + 0,8X$, $Y - \mu_Y = 2.000 + 0,8X - (2.000 + 0,8\mu_X) = 0,8(X - \mu_X)$. Por tanto $E[(Y - \mu_Y)^2] = E\{[0,8(X - \mu_X)]^2\} = 0,64E[(X - \mu_X)^2]$. Esto implica que $\text{var}(Y) = 0,64\text{var}(X)$, por tanto, tomando la raíz cuadrada de la varianza, la desviación típica de Y es

$$\sigma_Y = 0,8\sigma_X. \quad (2.10)$$

Es decir, la desviación típica de la distribución de los ingresos después de impuestos es el 80 % de la desviación típica de la distribución de los ingresos antes de impuestos.

Este análisis puede generalizarse al caso de que Y dependa de X con un intercepto a (en lugar de 2.000 \$) y una pendiente b (en lugar de 0,8), por lo que

$$Y = a + bX. \quad (2.11)$$

Por tanto la media y la varianza de Y son

$$\mu_Y = a + b\mu_X \text{ y} \quad (2.12)$$

$$\sigma_Y^2 = b^2\sigma_X^2, \quad (2.13)$$

y la desviación típica de Y es $\sigma_Y = b\sigma_X$. Las expresiones de las Ecuaciones (2.9) y (2.10) son casos particulares de las fórmulas más generales de las Ecuaciones (2.12) y (2.13) con $a = 2.000$ y $b = 0,8$.

Otras medidas de forma de una distribución

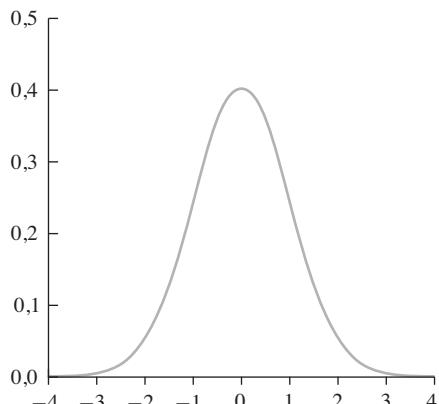
La media y la desviación típica miden dos características importantes de una distribución: su centro (la media) y su dispersión (la desviación típica). Esta sección aborda el estudio de las medidas de otras características de una distribución: la asimetría (o sesgo), que mide la falta de simetría de una distribución, y la curtosis, que mide el grosor o el peso de sus colas. La media, varianza, asimetría y curtosis están basadas toda ellas en lo que se denomina **momentos de una distribución**.

Asimetría. La Figura 2.3 representa cuatro distribuciones, dos de las cuales son simétricas (Figuras 2.3a y 2.3b) y otras dos que no lo son (Figuras 2.3c y 2.3d). Visualmente, la distribución de la Figura 2.3d aparece desviarse más de la simetría de lo que lo hace la distribución de la Figura 2.3c. La asimetría de una distribución proporciona un método matemático para describir cuánto se desvía una distribución de la simetría.

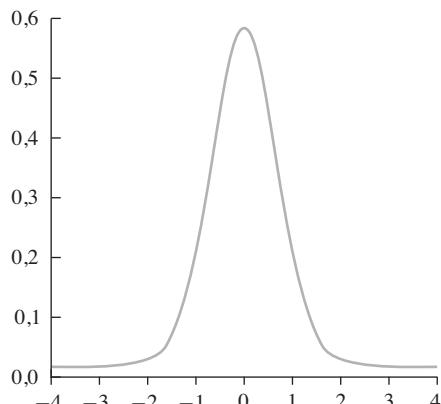
La **asimetría** de la distribución de una variable aleatoria Y es

$$\text{Asimetría} = \frac{E[(Y - \mu_Y)^3]}{\sigma_Y^3}, \quad (2.14)$$

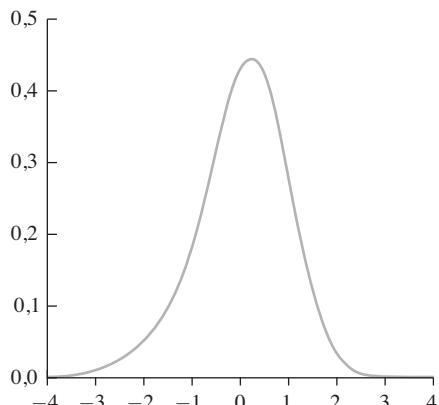
donde σ_Y es la desviación típica de Y . Para una distribución simétrica, un valor de Y de una determinada cuantía por encima de su media es exactamente igual de probable que un valor de Y de la misma cuantía por debajo de su media. Si es así, los valores positivos de $(Y - \mu_Y)^3$ se neutralizarán en media (en espe-

FIGURA 2.3 Cuatro distribuciones con diferentes asimetrías y curtosis

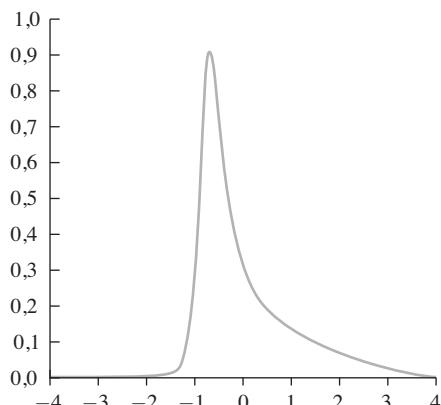
(a) Asimetría = 0, curtosis = 3



(b) Asimetría = 0, curtosis = 20



(c) Asimetría = -0,1, curtosis = 5



(d) Asimetría = 0,6, curtosis = 5

Todas estas distribuciones tienen una media de 0 y una varianza de 1. Las distribuciones con asimetría igual a 0 (a y b) son simétricas; las distribuciones con asimetría distinta de cero (c y d) no son simétricas. Las distribuciones con curtosis superior a 3 tienen colas gruesas.

ranza) mediante valores negativos igualmente probables. Por tanto, para una distribución simétrica, $E[(Y - \mu_Y)^3] = 0$, la asimetría de una distribución simétrica es cero. Si una distribución no es simétrica, entonces un valor positivo de $(Y - \mu_Y)^3$ en general no se neutralizará en media con un valor negativo igualmente probable, por lo que la asimetría es distinta de cero para una distribución asimétrica. Dividiendo por σ_Y^3 en el denominador de la Ecuación (2.14) las unidades de Y^3 del numerador se cancelan, por lo tanto la asimetría no tiene unidades; en otras palabras, cambiando las unidades de Y no cambia la asimetría.

Debajo de cada una de las cuatro distribuciones de la Figura 2.3 se presenta su asimetría. Si una distribución presenta una cola derecha larga, los valores positivos de $(Y - \mu_Y)^3$ no estarán totalmente compensados por valores negativos, y la simetría es positiva. Si una distribución presenta una cola izquierda larga, su asimetría es negativa.

Curtosis. La **curtosis** de una distribución es una medida de cuánta masa probabilística se encuentra en sus colas, por tanto, es una medida de cuánta varianza de Y proviene de los valores extremos. Un valor extremo de Y se denomina **atípico** (outlier). Cuanto mayor es la curtosis de una distribución, más probables son los atípicos.

La curtosis de la distribución de Y es

$$\text{Curtosis} = \frac{E[(Y - \mu_Y)^4]}{\sigma_Y^4}. \quad (2.15)$$

Si una distribución presenta una gran cantidad de masa probabilística en sus colas, entonces las desviaciones extremas de Y respecto de su media será probables, y esos valores altos conducirán a valores elevados, en media (en esperanza), de $(Y - \mu_Y)^4$. Por tanto, para una distribución con una gran cantidad de masa en las colas, la curtosis será elevada. Como $(Y - \mu_Y)^4$ no puede ser negativo, la curtosis no puede ser negativa.

La curtosis de una variable aleatoria normalmente distribuida es 3, por tanto una variable aleatoria con una curtosis mayor que 3 tiene más masa en sus colas que la variable aleatoria normal. Una distribución con una curtosis mayor que 3 se denomina **leptocúrtica** o más sencillamente, de colas anchas. Como la asimetría, la curtosis no tiene unidades, por tanto al cambiar las unidades de Y su curtosis no varía.

Debajo de cada una de las cuatro distribuciones de la Figura 2.3 se presenta su curtosis. Las distribuciones de las Figura 2.3b y c son de colas anchas.

Momentos. La media de Y , $E(Y)$, se denomina además momento primero de Y , y el valor esperado del cuadrado de Y , $E(Y^2)$, se denomina momento segundo de Y . En general, el valor esperado de Y^r se denomina momento r -ésimo de la variable aleatoria Y . Es decir, el **momento r -ésimo** de Y es $E(Y^r)$. La asimetría es una función del primero, segundo y tercer momento de Y , y la curtosis es una función del primer al cuarto momento de Y .

2.3 Dos variables aleatorias

La mayoría de las cuestiones económicas implican a dos o más variables. ¿Tienen los graduados universitarios más probabilidad de obtener un empleo que los no graduados? ¿Cómo es la distribución de la renta de las mujeres comparada con la de los hombres? Estas preguntas atan a las distribuciones de dos variables aleatorias, considerándolas de forma conjunta (educación y status laboral en el primer ejemplo, renta y género en el segundo). Las respuestas a esas preguntas requieren la comprensión de los conceptos de distribuciones de probabilidad conjunta, marginal y condicional.

Distribuciones conjunta y marginal

Distribución conjunta. La **distribución de probabilidad conjunta** de dos variables aleatorias discretas, X e Y , es la probabilidad de que las dos variables aleatorias tomen valores concretos de forma simultánea, x e y . Las probabilidades de todas las posibles combinaciones (x, y) suman 1. La distribución de probabilidad puede escribirse como la función $\Pr(X = x, Y = y)$.

Por ejemplo, las condiciones meteorológicas —si está lloviendo o no— afectan al tiempo de desplazamiento de la estudiante que se desplazaba en la Sección 2.1. Sea Y una variable aleatoria binaria que es igual a 1 si el desplazamiento es corto (menos de 20 minutos) e igual a 0 en otro caso y sea X una variable aleatoria binaria que es igual a 0 si llueve y 1 si no. Entre estas dos variables aleatorias, existen cuatro posibles resultados: lluvia y tiempo de desplazamiento largo ($X = 0, Y = 0$); lluvia y tiempo de desplazamiento corto ($X = 0, Y = 1$); sin lluvia y tiempo de desplazamiento largo ($X = 1, Y = 0$); y sin lluvia y tiempo de desplazamiento corto ($X = 1, Y = 1$). La distribución de probabilidad conjunta es la frecuencia con la que ocurre cada uno de estos sucesos a lo largo de muchas repeticiones de desplazamientos.

Se ofrece un ejemplo de una distribución conjunta de esas dos variables en la Tabla 2.2. De acuerdo con esta distribución, a lo largo de muchos desplazamientos, el 15 % de los días llueve y los desplazamientos son largos ($X = 0, Y = 0$); es decir, la probabilidad de un desplazamiento largo y con lluvia es del 15 %, o $\Pr(X = 0, Y = 0) = 0,15$. Además, $\Pr(X = 0, Y = 1) = 0,15$, $\Pr(X = 1, Y = 0) = 0,07$, y $\Pr(X = 1, Y = 1) = 0,63$. Estos cuatro posibles resultados son mutuamente excluyentes y constituyen el espacio muestral y por tanto las cuatro probabilidades suman 1.

TABLA 2.2 Distribución conjunta de condiciones meteorológicas y tiempo de desplazamiento

	Lluvia ($X = 0$)	Sin lluvia ($X = 1$)	Total
Desplazamiento largo ($Y = 0$)	0,15	0,07	0,22
Desplazamiento corto ($Y = 1$)	0,15	0,63	0,78
Total	0,30	0,70	1,00

Distribución de probabilidad marginal. La **distribución de probabilidad marginal** de una variable aleatoria Y es solo otro nombre para su distribución de probabilidad. Este término se utiliza para distinguir la distribución de Y en solitario (la distribución marginal) de la distribución conjunta de Y y otra variable aleatoria.

La distribución marginal de Y puede calcularse a partir de la distribución conjunta de X e Y sumando todas las probabilidades de todos los resultados posibles para los cuales Y toma un valor particular. Si X puede tomar l diferentes valores x_1, x_2, \dots, x_l , entonces la probabilidad marginal de que Y tome el valor y es

$$\Pr(Y = y) = \sum_{i=1}^l \Pr(X = x_i, Y = y) \quad (2.16)$$

Por ejemplo, en la Tabla 2.2, la probabilidad de un desplazamiento largo y lluvioso es del 15 % y la probabilidad de un desplazamiento largo sin lluvia es del 7 %, por tanto la probabilidad de un desplazamiento largo (lluvioso o no) es del 22 %. La distribución marginal del tiempo de desplazamiento se recoge en la columna final de la Tabla 2.2. De forma similar, la probabilidad marginal de lluvia es del 30 %, como se muestra en la última fila de la Tabla 2.2.

Distribuciones condicionales

Distribución condicional. La distribución de una variable aleatoria Y condicionada a que otra variable aleatoria X tome un valor específico se denomina **distribución condicional** de Y dado X . La probabilidad condicional de que Y tome el valor y cuando X toma el valor x se expresa como $\Pr(Y = y | X = x)$.

Por ejemplo, ¿cuál es la probabilidad de un desplazamiento largo ($Y = 0$) si se sabe que llueve ($X = 0$)? De la Tabla 2.2, se obtiene que la probabilidad conjunta de un desplazamiento corto lluvioso es del 15 % y la probabilidad conjunta de un desplazamiento lluvioso y largo es del 15 %, por lo tanto, si llueve, un desplazamiento largo y un desplazamiento corto son igualmente probables. Por tanto, la probabilidad de un desplazamiento largo ($Y = 0$), condicionado a que llueve ($X = 0$), es del 50 %, o $\Pr(Y = 0 | X = 0) = 0,50$. De forma equivalente, la probabilidad marginal de lluvia es del 30 %; es decir, si se toman muchos desplazamientos, llueve el 30 % de las veces. De este 30 % de desplazamientos, el 50 % de las veces el desplazamiento es largo (0,15/0,30).

En general, la distribución condicional de Y dado $X = x$ es

$$\Pr(Y = y | X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)} \quad (2.17)$$

Por ejemplo, la probabilidad condicional de un desplazamiento largo dado que llueve es $\Pr(Y = 0 | X = 0) = \Pr(X = 0, Y = 0) / \Pr(X = 0) = 0,15 / 0,30 = 0,50$.

Como segundo ejemplo, considérese una modificación del ejemplo de las averías en el ordenador. Supongamos que utiliza un ordenador en una biblioteca para redactar un trabajo y el bibliotecario le asigna aleatoriamente un ordenador de los disponibles, la mitad de los cuales son nuevos y la otra mitad son viejos. Por habersele asignado un ordenador de forma aleatoria, la edad del ordenador que usa, A (= 1 si el ordenador es nuevo, = 0 si es viejo), es una variable aleatoria. Supongamos que la distribución conjunta de las variables aleatorias M y A viene dada por la Parte A de la Tabla 2.3. Entonces la distribución condicional de las

TABLA 2.3 Distribuciones conjunta y condicional de averías en ordenador (M) y edad del ordenador (A)

A) Distribución conjunta		$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	Total
Ordenador viejo ($A = 0$)		0,35	0,065	0,05	0,025	0,01	0,50
Ordenador nuevo ($A = 1$)		0,45	0,035	0,01	0,005	0,00	0,50
Total		0,80	0,10	0,06	0,03	0,01	1,00
B) Distribución condicional de M dado A		$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	Total
Pr($M A = 0$)		0,70	0,13	0,10	0,05	0,02	1,00
Pr($M A = 1$)		0,90	0,07	0,02	0,01	0,00	1,00

averías de ordenador, dada la edad del ordenador, viene recogida en la Parte B de la tabla. Por ejemplo, la probabilidad conjunta $M = 0$ y $A = 0$ es 0,35; como la mitad de los ordenadores son viejos, la probabilidad condicionada de ausencia de averías, dado que se usa un ordenador viejo, es $\text{Pr}(M=0|A=0)=\text{Pr}(M=0, A=0)/\text{Pr}(A=0)=0,35/0,50=0,70$, o del 70 %. Por contra, la probabilidad condicional de ausencia de averías dado que se le haya asignado un ordenador nuevo es del 90 %. De acuerdo con las distribuciones condicionales de la Parte B de la Tabla 2.3, los ordenadores más nuevos son menos proclives a averiarse que los viejos; por ejemplo, la probabilidad de tres averías es del 5 % con un ordenador viejo, pero solamente del 1 % con un ordenador nuevo.

Esperanza condicional. La **esperanza condicional** de Y dado X , asimismo denominada **media condicional** de Y dado X , es la media de la distribución condicional de Y dado X . Es decir, la esperanza condicional es el valor esperado de Y , calculado mediante la distribución condicional de Y dado X . Si Y toma k valores y_1, y_2, \dots, y_k , entonces la media condicional de Y dado $X = x$ es

$$E(Y|X=x) = \sum_{i=1}^k y_i \Pr(Y=y_i|X=x). \quad (2.18)$$

Por ejemplo, en base a las distribuciones condicionales de la Tabla 2.3, el número esperado de averías de ordenador, dado que el ordenador es viejo, es $E(M|A=0) = 0 \times 0,70 + 1 \times 0,13 + 2 \times 0,10 + 3 \times 0,05 + 4 \times 0,02 = 0,56$. El número esperado de averías de ordenador dado que el ordenador es nuevo, es $E(M|A=1) = 0,14$, menor que para los ordenadores viejos.

La esperanza condicional de Y dado $X = x$ es exactamente el valor medio de Y cuando $X = x$. En el ejemplo de la Tabla 2.3, el número medio de averías es 0,56 para los ordenadores viejos, por lo que la esperanza condicional de Y dado que el ordenador es viejo es 0,56. De forma similar, entre los ordenadores nuevos, el número medio de averías es 0,14, es decir, la esperanza condicional de Y dado que el ordenador es nuevo es 0,14.

La ley de esperanzas iteradas. La media de Y es la media ponderada de la esperanza condicional de Y dado X , ponderada por la distribución de probabilidad de X . Por ejemplo, la altura media de los adultos es la media ponderada de la altura media de los hombres y la altura media de las mujeres, ponderando por la proporción de hombres y mujeres. Matemáticamente, si X toma l valores x_1, x_2, \dots, x_l , entonces

$$E(Y) = \sum_{i=1}^l E(Y|X=x_i) \Pr(X=x_i). \quad (2.19)$$

La Ecuación (2.19) se desprende de las Ecuaciones (2.18) y (2.17) (véase Ejercicio 2.19).

Expresado de otro modo, la esperanza de Y es la esperanza de la esperanza condicional de Y dado X ,

$$E(Y) = E[E(Y|X)], \quad (2.20)$$

donde la esperanza interior de la parte derecha de la Ecuación (2.20) se calcula utilizando la distribución condicional de Y dado X y la esperanza exterior se calcula utilizando la distribución marginal de X . La Ecuación (2.20) se conoce como la **ley de esperanzas iteradas**.

Por ejemplo, el número medio de averías M ; es la media ponderada de la esperanza condicional de M dado que es viejo y la esperanza condicional de M dado que es nuevo, por lo que $E(M) = E(M|A = 0) \times \Pr(A = 0) + E(M|A = 1) \times \Pr(A = 1) = 0,56 \times 0,50 + 0,14 \times 0,50 = 0,35$. Esta es la media de la distribución marginal de M , tal y como se calcula en la Ecuación (2.2).

La ley de esperanzas iteradas implica que si la media condicional de Y dado X es cero, entonces la media de Y es cero. Esto es una consecuencia inmediata de la Ecuación (2.20): si la $E(Y|X) = 0$, entonces $E(Y) = E[E(Y|X)] = E[0] = 0$. Dicho de otra forma, si la media de Y dado X es cero, entonces debe ocurrir que la media ponderada por probabilidad de esa media condicional sea cero, es decir, la media de Y debe ser cero.

La ley de esperanzas iteradas además es aplicable a las esperanzas que son condicionales sobre varias variables aleatorias. Por ejemplo, sean X , Y , y Z variables aleatorias distribuidas de forma conjunta. Por tanto, la ley de esperanzas iteradas dice que $E(Y) = E[E(Y|X, Z)]$, donde $E(Y|X, Z)$ es la esperanza condicional de Y dados tanto X como Z . Por ejemplo, en el ejemplo concreto de las averías de ordenador de la Tabla 2.3, sea P el número de programas instalados en el ordenador; por tanto $E(M|A, P)$ es el número esperado de averías para un ordenador con edad A que tiene instalados P programas. El número esperado de averías en total, $E(M)$, es la media ponderada del número esperado de averías de un ordenador con edad A y un número instalado de programas P , ponderado por la proporción de ordenadores con ese valor tanto de A como de P .

El Ejercicio 2.20 proporciona algunas propiedades adicionales de las esperanzas condicionales con varias variables.

Varianza condicional. La varianza de Y condicionada a X es la varianza de la distribución condicional de Y dado X . Matemáticamente, la **varianza condicional** de Y dado X es

$$\text{var}(Y|X = x) = \sum_{i=1}^k [y_i - E(Y|X = x)]^2 \Pr(Y = y_i | X = x). \quad (2.21)$$

Por ejemplo, la varianza condicional del número de averías dado que el ordenador es viejo es $\text{var}(M|A = 0) = (0 - 0,56)^2 \times 0,70 + (1 - 0,56)^2 \times 0,13 + (2 - 0,56)^2 \times 0,10 + (3 - 0,56)^2 \times 0,05 + (4 - 0,56)^2 \times 0,02 \cong 0,99$. La desviación típica de la distribución condicional de M dado que $A = 0$ es por tanto $\sqrt{0,99} = 0,99$. La varianza condicional de M dado que $A = 1$ es la varianza de la distribución de la segunda fila del Panel B de la Tabla 2.3, que es 0,22, por lo que la desviación típica de M para ordenadores nuevos es $\sqrt{0,22} = 0,47$. Para la distribución condicional de la Tabla 2.3, el número esperado de averías para los ordenadores nuevos (0,14) es menor que para los ordenadores viejos (0,56), y la dispersión de la distribución del número de averías medida por la desviación típica condicional es más pequeña para los ordenadores nuevos (0,47) que para los viejos (0,99).

Independencia

Dos variables aleatorias X e Y están **independientemente distribuidas**, o son **independientes**, si el conocimiento del valor de una de las variables no proporciona información sobre la otra. En concreto, X e Y son independientes si la distribución condicional de Y dado X es igual a la distribución marginal de Y . Es decir, X e Y están independientemente distribuidas si, para todos los valores de x e y ,

$$\Pr(Y = y | X = x) = \Pr(Y = y) \quad (\text{independencia de } X \text{ e } Y). \quad (2.22)$$

Al sustituir la Ecuación (2.22) en la Ecuación (2.17) se obtiene una expresión alternativa para las variables aleatorias independientes en términos de su distribución conjunta. Si X e Y son independientes, entonces

$$\Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y). \quad (2.23)$$

Es decir, la distribución conjunta de dos variables aleatorias independientes es el producto de sus distribuciones marginales.

Covarianza y correlación

Covarianza. Una medida del grado al que dos variables aleatorias evolucionan conjuntamente es su covarianza. La **covarianza** entre X e Y es la esperanza $E[(X - \mu_X)(Y - \mu_Y)]$, donde μ_X es la media de X y μ_Y es la media de Y . La covarianza se expresa mediante $\text{cov}(X, Y)$ o σ_{XY} . Si X puede tomar l valores e Y puede tomar k valores, la covarianza viene dada por la fórmula

$$\begin{aligned}\text{cov}(X, Y) &= \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_{i=1}^k \sum_{j=1}^l (x_j - \mu_X)(y_i - \mu_Y) \Pr(X = x_j, Y = y_i).\end{aligned}\tag{2.24}$$

Para interpretar esta fórmula, supongamos que cuando X es mayor que su media (por tanto $X - \mu_X$ es positivo), entonces Y tiende a ser mayor que su media (por lo que $Y - \mu_Y$ es positivo), y cuando X es menor que su media (entonces $X - \mu_X < 0$), entonces Y tiende a ser menos que su media (por lo que $Y - \mu_Y < 0$). En ambos casos, el producto $(X - \mu_X) \times (Y - \mu_Y)$ tiende a ser positivo, por lo que la covarianza es positiva. Por contra, si X e Y tienden a evolucionar en sentido opuesto (es decir si X es grande cuando Y es pequeña, y viceversa), entonces la covarianza es negativa. Finalmente, si X e Y son independientes, entonces la covarianza es cero (véase Ejercicio 2.19).

Correlación. Debido a que la covarianza es el producto de X e Y , en desviaciones respecto de sus medias, sus unidades son, desafortunadamente, las unidades de X multiplicadas por las unidades de Y . Este problema de «unidades» puede ocasionar que los resultados numéricos de la covarianza sean difíciles de interpretar.

La correlación es una medida alternativa de la dependencia entre X e Y que resuelve el problema de las «unidades» de la covarianza. En concreto, la **correlación** entre X e Y es la covarianza entre X e Y dividida por sus desviaciones típicas.

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}\tag{2.25}$$

Al ser las unidades del numerador de la Ecuación (2.25) las mismas que las del denominador, las unidades se cancelan y la correlación no tiene unidades. Las variables aleatorias X e Y se dice que están **incorelacionadas** si $\text{corr}(X, Y) = 0$.

La correlación siempre toma valores entre -1 y 1 ; es decir, como se demuestra en el Apéndice 2.1,

$$-1 \leq \text{corr}(X, Y) \leq 1 \quad (\text{desigualdad de la correlación})\tag{2.26}$$

Correlación y media condicional. Si la media condicional de Y no depende de X , entonces Y y X están incorrelacionadas. Es decir,

$$\text{si } E(Y|X) = \mu_Y, \text{ entonces } \text{cov}(Y, X) = 0 \text{ y } \text{corr}(Y, X) = 0.\tag{2.27}$$

Demostramos ahora este resultado. Primero supongamos que Y y X tienen media cero por lo que $\text{cov}(Y, X) = E[(Y - \mu_Y)(X - \mu_X)] = E(YX)$. Por la ley de esperanzas iteradas [Ecuación (2.20)], $E(YX) = E[E(YX|X)] = E[E(Y|X)X] = 0$ debido a que $E(Y|X) = 0$, entonces $\text{cov}(Y, X) = 0$. La Ecuación (2.27) se obtiene de sustituir $\text{cov}(Y, X) = 0$ en la definición de las correlaciones de la Ecuación (2.25). Si Y y X no tienen media igual a cero, se les resta primero sus medias, y entonces es aplicable la prueba precedente.

Sin embargo, *no es* necesariamente cierto que si X e Y están incorrelacionadas, entonces la media condicional de Y dado X no depende de X . Dicho de otra forma, es posible que la media condicional de Y dado X sea una función de X y aun así que Y y X estén incorrelacionados. Se ofrece un ejemplo en el Ejercicio 2.23.

La distribución de ingresos salariales en Estados Unidos en 2008

Algunos padres cuentan a sus hijos que serán más capaces de alcanzar un trabajo mejor, y mejor pagado, si obtienen un título universitario que si se saltan la educación superior. ¿Están estos padres en lo cierto? La distribución salarial difiere entre trabajadores que son titulados universitarios y trabajadores que solamente tienen un diploma de escuela secundaria?, y si es así, ¿cuánto? Entre los trabajadores con una educación similar, es distinta la distribución salarial para hombres y mujeres? Por ejemplo, ¿gana la mujer mejor pagada y con mejor educación tanto como el hombre mejor pagado y mejor educado?

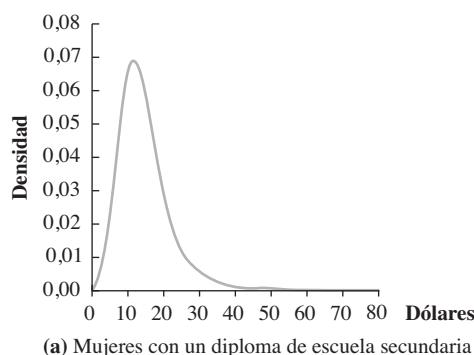
Una forma de responder a estas preguntas es mediante el examen de la distribución de salarios de los trabajadores a tiempo completo, condicionada al grado más alto de educación alcanzado (diploma de escuela secundaria o grado universitario) y al género. Estas cuatro distribuciones condicionales se muestran en la Figura 2.4, y la media, la desviación típica, y algunos percentiles de las distribuciones condicionadas se presentan en la Tabla 2.4¹. Por ejemplo, la media con-

dencial de los ingresos salariales de las mujeres cuyo título más elevado es el diploma de escuela secundaria —es decir, $E(\text{Ingresos Salariales} | \text{Títulación más elevada} = \text{diploma de escuela secundaria}, \text{Género} = \text{mujer})$ — es 14,73 \$ por hora.

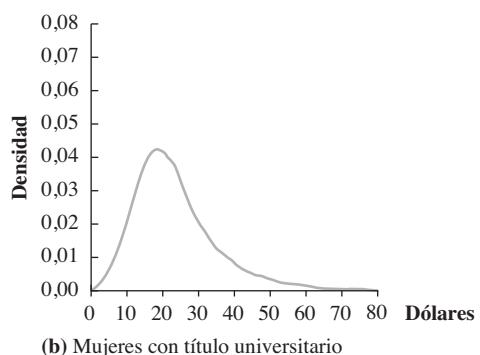
La distribución de los ingresos medios por hora para mujeres con título universitario (Figura 2.4b) se encuentra desplazada a la derecha de la distribución de las mujeres con tan solo un título de escuela secundaria (Figura 2.4a); el mismo desplazamiento puede observarse para los dos grupos de hombres (Figura 2.4d y Figura 2.4c). Tanto para mujeres como para hombres, los ingresos salariales medios son mayores para aquellos con título universitario (Tabla 2.4, primera columna numérica). Resulta interesante que la dispersión de la distribución de ingresos, medida por la desviación típica, es mayor para aquellos con un título universitario que para aquellos con

¹ Las distribuciones fueron estimadas utilizando los datos de la Encuesta Actualizada de Población, que se trata en detalle en el Apéndice 3.1.

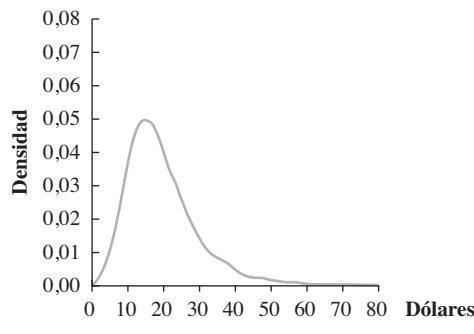
FIGURA 2.4 Distribuciones condicionales de los ingresos salariales medios por hora de los trabajadores a tiempo completo en EE.UU. en 2008, dados nivel educativo y género



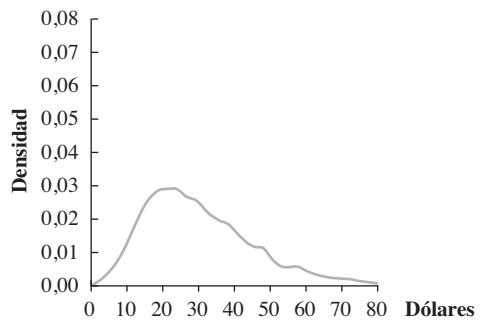
(a) Mujeres con un diploma de escuela secundaria



(b) Mujeres con título universitario



(c) Hombres con un diploma de escuela secundaria



(d) Hombres con título universitario

Las cuatro distribuciones de salarios son para mujeres y hombres, para aquellos con solo un diploma de escuela secundaria (a y c), y para aquellos que tienen como máximo un título universitario de 4 años (b y d).

TABLA 2.4

Resumen de la distribución condicional de ingresos salariales medios por hora de los trabajadores a tiempo completo en EE.UU. en 2008 dados nivel educativo y género

	Medio	Desviación típica	Porcentaje			
			25 %	50 % (mediana)	75 %	90 %
a) Mujeres con diploma de escuela secundaria	14,73	7,77	9,62	13,19	17,50	23,96
b) Mujeres con título universitario de cuatro años	23,93	12,59	15,38	21,63	28,85	39,42
c) Hombres con diploma de escuela secundaria	19,64	10,21	12,50	17,48	24,04	32,69
d) Hombres con título universitario de cuatro años	30,97	16,08	19,23	28,45	39,34	52,88

Los ingresos salariales medios por hora son la suma de los sueldos, salarios, propinas y bonus anuales antes de impuestos dividida por el número de horas anuales trabajadas. Las distribuciones se calcularon a partir de la Encuesta Actualizada de Población (C.P.S.) de marzo de 2009, que está descrita en el Apéndice 3.1.

un diploma de escuela secundaria. Además, tanto para hombres como para mujeres, el percentil 90 de ingresos es mucho mayor para los trabajadores con título universitario que para los trabajadores con solo un diploma de escuela secundaria. Esta última comparación es consistente con la advertencia paternal de que un título universitario abre puertas que permanecen cerradas a los individuos con tan solo un diploma de educación secundaria.

Otra característica de estas distribuciones es que la distribución de ingresos para hombres se encuentra desplazada hacia la derecha de la distribución de ingresos de mujeres. Esta «brecha de género» en ingresos salariales es un importante —y, muchas veces, preocupante— aspecto de la distribución de ingresos salariales. Volveremos a este tema en capítulos posteriores.

La media y la varianza de la suma de variables aleatorias

La media de la suma de dos variables aleatorias, X e Y , es la suma de sus medias:

$$E(X + Y) = E(X) + E(Y) = \mu_X + \mu_Y. \quad (2.28)$$

CONCEPTO CLAVE

2.3

Medias, varianzas y covarianzas de la suma de variables aleatorias

Sean X , Y y V variables aleatorias, sean μ_X y σ_X^2 la media y la varianza de X , sea σ_{XY} la covarianza entre X e Y (lo mismo igualmente para las otras variables), y sean a , b y c constantes. Las Ecuaciones (2.29) a (2.35) se derivan de las definiciones de la media, varianza y covarianza:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \quad (2.29)$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \quad (2.30)$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \quad (2.31)$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \quad (2.32)$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{YV}, \quad (2.33)$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y, \quad (2.34)$$

$$|\text{corr}(X, Y)| \leq 1 \text{ y } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (desigualdad de la correlación).} \quad (2.35)$$

La varianza de la suma de X e Y es la suma de sus varianzas más dos veces su covarianza:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}. \quad (2.36)$$

Si X e Y son independientes, entonces la covarianza es cero y la varianza de su suma es la suma de sus varianzas:

$$\begin{aligned} \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) = \sigma_X^2 + \sigma_Y^2 \\ &\quad (\text{si } X \text{ e } Y \text{ son independientes}) \end{aligned} \quad (2.37)$$

En el Concepto clave 2.3 se recogen expresiones útiles para las medias, varianzas y covarianzas que incluyen sumas ponderadas de variables aleatorias. Los resultados del Concepto clave 2.3 se deducen en el Apéndice 2.1.

2.4 Las distribuciones normal, chi-cuadrado, t de Student y F

Las distribuciones de probabilidad con las que más frecuentemente nos encontramos en econometría son la normal, la chi-cuadrado, la t de Student, y la F .

La distribución normal

Una variable aleatoria continua con **distribución normal** posee la conocida densidad de probabilidad con forma de campana que se muestra en la Figura 2.5. La función que define la densidad de probabilidad normal se ofrece en el Apéndice 17.1. Como muestra la Figura 2.5, la densidad normal con media μ y varianza σ^2 es simétrica respecto de su media y tiene el 95 % de su probabilidad entre $\mu - 1,96\sigma$ y $\mu + 1,96\sigma$.

Para la distribución normal se ha desarrollado una notación y una terminología particular. La distribución normal con media μ y varianza σ^2 se expresa de forma concisa como « $N(\mu, \sigma^2)$ ». La **distribución normal estándar** es la distribución normal con media $\mu = 0$ y varianza $\sigma^2 = 1$ y se expresa como $N(0, 1)$. Las variables aleatorias con distribución $N(0, 1)$ a menudo se expresan como Z , y la función de distribución acumulada normal estándar se escribe con la letra griega Φ ; así, $\Pr(Z \leq c) = \Phi(c)$, donde c es una constante. Los valores de la función de distribución acumulada normal estándar están tabulados en la Tabla 1 del Apéndice.

Para buscar probabilidades de una variable normal con cualquier media y varianza, debemos **estandarizar la variable** restando primero la media, y posteriormente dividiendo el resultado por la desviación típica. Por ejemplo, supóngase que Y está distribuida como $N(1, 4)$ —es decir, Y está normalmente distribuida con

FIGURA 2.5 La densidad de probabilidad normal

La función de densidad de probabilidad normal con media μ y varianza σ^2 es una curva con forma de campana, centrada en μ . El área bajo la f.d.p. normal entre $\mu - 1,96\sigma$ y $\mu + 1,96\sigma$ es 0,95. La distribución normal se expresa como $N(\mu, \sigma^2)$.

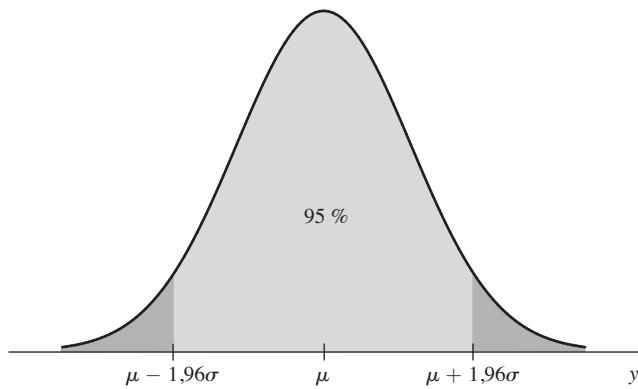


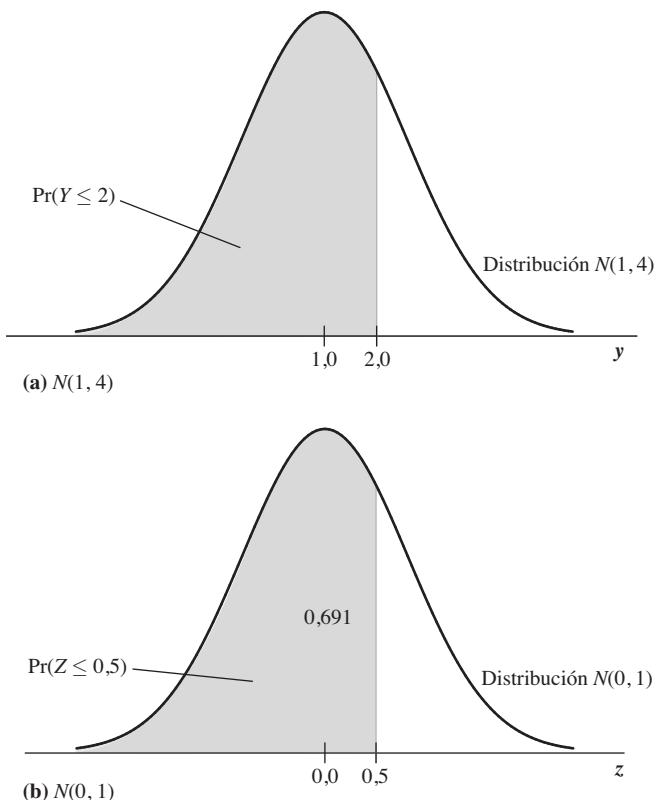
FIGURA 2.6 Cálculo de la probabilidad de que $Y \leq 2$ si Y es $N(1, 4)$

Para el cálculo de $\Pr(Y \leq 2)$, Y se estandariza, posteriormente se utilizan las tablas de la distribución normal estándar. Y se estandariza restándole su media ($\mu = 1$) y dividiendo por su desviación típica ($\sigma = 2$).

La probabilidad de que $Y \leq 2$ se muestra en la Figura 2.6a, y la probabilidad correspondiente tras estandarizar Y se muestra en la Figura 2.6b.

Como la variable aleatoria estandarizada $(Y - 1)/2$, es una variable aleatoria normal estándar (Z), $\Pr(Y \leq 2) = \Pr\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = \Pr(Z \leq 0,5)$.

De la Tabla 1 del Apéndice, $\Pr(Z \leq 0,5) = \Phi(0,5) = 0,691$.



CONCEPTO CLAVE

2.4

Cálculo de probabilidades con variables aleatorias normales

Supongamos que Y está normalmente distribuida con media μ y varianza σ^2 ; en otras palabras, Y está distribuida como $N(\mu, \sigma^2)$. Por tanto, Y se estandariza restándole su media y dividiendo por su desviación típica, es decir, calculando $Z = (Y - \mu)/\sigma$.

Sean c_1 y c_2 dos números con $c_1 < c_2$ y sea $d_1 = (c_1 - \mu)/\sigma$ y $d_2 = (c_2 - \mu)/\sigma$. Entonces

$$\Pr(Y \leq c_2) = \Pr(Z \leq d_2) = \Phi(d_2), \quad (2.38)$$

$$\Pr(Y \geq c_1) = \Pr(Z \geq d_1) = 1 - \Phi(d_1), \quad (2.39)$$

$$\Pr(c_1 \leq Y \leq c_2) = \Pr(d_1 \leq Z \leq d_2) = \Phi(d_2) - \Phi(d_1). \quad (2.40)$$

La función de distribución normal acumulada Φ está tabulada en la Tabla 1 del Apéndice.

una media de 1 y una varianza de 4. ¿Cuál es la probabilidad de que $Y \leq 2$ —es decir, cuál es el área sombreada de la Figura 2.6a? La versión estandarizada de Y es Y menos su media, dividida por su desviación típica, es decir $(Y - 1)/\sqrt{4} = \frac{1}{2}(Y - 1)$. En consecuencia, la variable aleatoria $\frac{1}{2}(Y - 1)$ está distribuida como una normal con media cero y varianza uno (véase el Ejercicio 2.8); posee la distribución normal estándar mostrada en la Figura 2.6b. Ahora $Y \leq 2$ es equivalente a $\frac{1}{2}(Y - 1) \leq \frac{1}{2}(2 - 1)$ —es decir, $\frac{1}{2}(Y - 1) \leq \frac{1}{2}$. Por lo tanto

$$\Pr(Y \leq 2) = \Pr\left[\frac{1}{2}(Y - 1) \leq \frac{1}{2}\right] = \Pr\left(Z \leq \frac{1}{2}\right) = \Phi(0,5) = 0,691, \quad (2.41)$$

donde el valor 0,691 se toma de la Tabla 1 del Apéndice.

El mismo método puede aplicarse para calcular la probabilidad de que una variable aleatoria normalmente distribuida supere un valor o que se encuentre en un cierto rango. Estos pasos se recogen en el Concepto clave 2.4. El cuadro «Un mal día en Wall Street» presenta una aplicación poco habitual de la distribución normal acumulada.

La distribución normal es simétrica, por lo que su asimetría es cero. La curtosis de la distribución normal es 3.

La distribución normal multivariante. La distribución normal puede generalizarse para describir la distribución conjunta de un conjunto de variables. En este caso, la distribución se denomina **distribución normal multivariante**, o bien si solo se están considerando dos variables, **distribución normal bivariante**. La fórmula para la f.d.p. normal bivariante se ofrece en el Apéndice 18.1.

La distribución normal multivariante presenta cuatro propiedades importantes. Si X e Y presentan una distribución normal bivariante con covarianza σ_{XY} y a y b son dos constantes, entonces $aX + bY$ posee distribución normal:

$$aX + bY \text{ está distribuida } N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}) \quad (2.42)$$

(X, Y normal bivariante)

De forma más general, si n variables aleatorias presentan una distribución normal multivariante, entonces cualquier combinación lineal de esas variables (como puede ser su suma) se distribuye normalmente.

Segundo, si un conjunto de variables poseen una distribución normal multivariante, entonces la distribución marginal de cada una de esas variables es normal [lo que se deduce de la Ecuación (2.42) haciendo $a = 1$ y $b = 0$].

Tercero, si las variables que presentan una función de distribución normal multivariante tienen covarianzas iguales a cero, entonces las variables son independientes. Así, si X e Y tienen una distribución normal bivariante y $\sigma_{XY} = 0$, entonces X e Y son independientes. En la Sección 2.3 se ha establecido que si X e Y son independientes, entonces, sea cual sea su distribución conjunta, $\sigma_{XY} = 0$. Si X e Y se distribuyen conjuntamente como una normal, entonces lo recíproco es igualmente cierto. Este resultado —la covarianza cero implica independencia— es una propiedad particular de la distribución normal multivariante que no se cumple en general.

Cuarto, si X e Y presentan una distribución normal multivariante, entonces la esperanza condicional de Y dado X es lineal en X ; es decir $E(Y|X = x) = a + bx$, donde a y b son constantes (Ejercicio 17.11). La normalidad conjunta implica linealidad de las esperanzas condicionales, pero la linealidad de las esperanzas condicionales no implica normalidad conjunta.

La distribución chi-cuadrado

La distribución chi-cuadrado se utiliza para contrastar cierto tipo de hipótesis en estadística y econometría.

La **distribución chi-cuadrado** es la distribución de la suma de m variables aleatorias normales estándar independientes al cuadrado. Esta distribución depende de m , que se denomina grados de libertad de la distribución chi-cuadrado. Por ejemplo, sean Z_1 , Z_2 , y Z_3 variables aleatorias normales estándar independientes. Entonces $Z_1^2 + Z_2^2 + Z_3^2$ presenta una distribución chi-cuadrado con 3 grados de libertad. El nombre de esta distribución procede de la letra griega utilizada para denominarla: una distribución chi-cuadrado con m grados de libertad se expresa mediante el símbolo χ_m^2 .

En la Tabla 3 del Apéndice se ofrecen algunos percentiles seleccionados de la distribución χ_m^2 . Por ejemplo, la Tabla 3 del Apéndice muestra que el percentil 95 de la distribución χ_3^2 es 7,81, por lo que $\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7,81) = 0,95$.

La distribución t de Student

La **distribución t de Student** con m grados de libertad se define como la distribución del cociente entre una variable aleatoria normal estándar y la raíz cuadrada de una variable aleatoria chi-cuadrado inde-

Un mal día en Wall Street

En un día típico el valor conjunto de las acciones negociadas en la bolsa de EE.UU. puede aumentar o caer en un 1 % o incluso más. Esto es mucho —pero nada comparado con lo que ocurrió el lunes, 19 de octubre de 1987. En el «Lunes Negro», el Dow Jones Industrial (una media de 30 acciones de grandes empresas) cayó un 22,6 %. Desde el 1 de febrero de 1980, hasta el 31 de diciembre de 2009, la desviación típica de la variación porcentual diaria en los precios del Dow fue del 1,13 %, por lo que la caída del 22,6 % representaba una rentabilidad negativa de 20 (= 22,6/1,13) desviaciones típicas. La enormidad de esta caída puede verse en la Figura 2.7, un gráfico de la rentabilidad diaria en el Dow durante los años 80.

Si la variación porcentual diaria de los precios se distribuye normalmente, entonces la probabilidad de un cambio de al menos 20 desviaciones típicas es $\Pr(|Z| \geq 20) = 2 \times \Phi(-20)$. No encontrará este valor en la Tabla 1 del Apéndice, pero puede de calcularse utilizando un ordenador (¡inténtelo!). Esta probabilidad es $5,5 \times 10^{-89}$, esto es 0,0000 ... 00055, ¡donde hay un total de 88 ceros!

¿Cómo de pequeño es $5,5 \times 10^{-89}$?

- La población mundial se acerca a los 7 mil millones de habitantes, por lo que la probabilidad de ganar una lotería

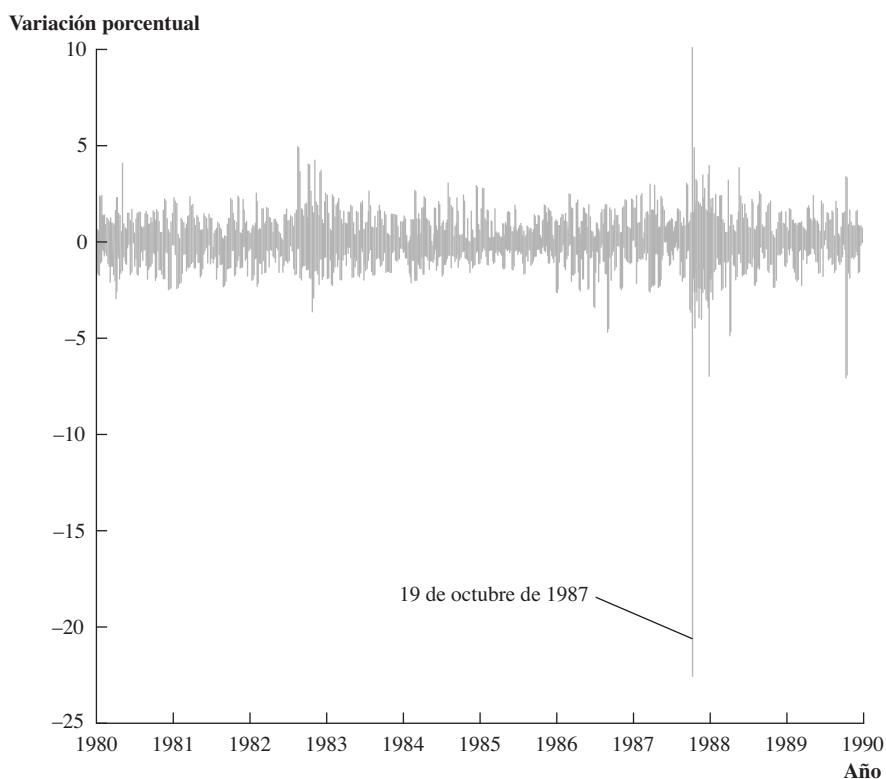
aleatoria entre todas las personas vivas es de alrededor de una sobre 7 mil millones, o $1,4 \times 10^{-10}$.

- Se cree que el universo existe desde hace 14 mil años, o lo que es lo mismo aproximadamente 5×10^{17} segundos, por lo que la probabilidad de elegir un segundo concreto de forma aleatoria desde el inicio de los tiempos es de 2×10^{-18} .
- Hay aproximadamente 10^{43} moléculas de gas en el primer kilómetro que envuelve a la superficie terrestre. La probabilidad de elegir una aleatoriamente es 2×10^{-43} .

Aunque Wall Street *tuviera* un mal día, el hecho de que sucediera no sugiere en absoluto que su probabilidad sea mayor que $5,5 \times 10^{-89}$. De hecho, ha habido muchos días —buenos y malos— con cambios en los precios demasiado grandes como para ser consistentes con una distribución normal con varianza constante. La Tabla 2.5 presenta las diez mayores variaciones porcentuales diarias en los precios del Dow Jones a lo largo de los 7.571 días hábiles entre el 1 de enero de 1980, y el 31 de diciembre de 2009, junto con la variación estandarizada mediante la media y la varianza de este periodo. Las diez variaciones superan las 6,4 desviaciones típicas, un suceso extremadamente escaso incluso si el precio de las acciones estuviera distribuido normalmente.

FIGURA 2.7 Variación porcentual diaria en el «Dow Jones Industrial Average» en la década de 1980

Durante los años 80, la variación porcentual diaria media del índice «Dow» fue del 0,05% y su desviación típica fue del 1,16 %. El 19 de octubre de 1987 —el «Lunes Negro»— el índice cayó un 25,6 % o más de 22 desviaciones típicas.



(continúa)

TABLA 2.5 Las diez mayores variaciones porcentuales diarias en el índice Dow Jones, 1980-2009, y la probabilidad normal de una variación al menos tan elevada

Fecha	Variación porcentual (x)	Variación estandarizada $z = (x - \mu)/\sigma$	Probabilidad normal de una variación al menos tan grande $\Pr(Z \geq z) = 2\Phi(-z)$
19 de octubre de 1987	-22,6	-20,1	$5,5 \times 10^{-89}$
13 de octubre de 2008	11,1	9,8	$1,1 \times 10^{-22}$
28 de octubre de 2008	10,9	9,6	$6,5 \times 10^{-22}$
21 de octubre de 1987	10,1	9,0	$2,8 \times 10^{-19}$
26 de octubre de 1987	-8,0	-7,2	$7,4 \times 10^{-13}$
15 de octubre de 2008	-7,9	-7,0	$2,1 \times 10^{-12}$
1 de diciembre de 2008	-7,7	-6,9	$6,3 \times 10^{-12}$
9 de octubre de 2008	-7,3	-6,5	$5,9 \times 10^{-11}$
27 de octubre de 1997	-7,2	-6,4	$1,4 \times 10^{-10}$
17 de septiembre de 2001	-7,1	-6,4	$2,0 \times 10^{-10}$

Claramente, las variaciones porcentuales del precio de las acciones tienen una distribución con colas más anchas que las de la distribución normal estándar. Por esta razón, los profesionales financieros utilizan otros modelos para las variaciones del precio de las acciones. Uno de esos modelos considera las variaciones del precio de las acciones como una distribución normal con varianza que evoluciona en el tiempo, por lo que períodos como los de octubre de 1987 y la

crisis financiera en otoño de 2008 presentan mayor volatilidad que otros (los modelos con varianza que varía en el tiempo se tratan en el Capítulo 16). Otros modelos abandonan la distribución normal a favor de distribuciones con colas más anchas, una idea popularizada en el libro de Taleb Nassim de 2007, *El Cisne Negro*. Estos modelos son más consistentes con los días muy malos —o muy buenos— que pueden verse en la realidad en Wall Street.

pendientemente distribuida con m grados de libertad dividida por m . Es decir, sea Z una variable aleatoria normal estándar, sea W una variable aleatoria con distribución chi-cuadrado con m grados de libertad, y sean Z y W independientemente distribuidas. Entonces la variable aleatoria $Z/\sqrt{W/m}$ presenta una distribución t de Student (asimismo denominada **distribución t**) con m grados de libertad. Esta distribución se expresa mediante t_m . En la Tabla 2 del Apéndice se ofrecen algunos percentiles de la distribución t de Student.

La distribución t de Student depende de los grados de libertad m . De ese modo, el percentil 95 de la distribución t_m depende de los grados de libertad m . La distribución t de Student presenta una forma de campana similar a la de la distribución normal, pero cuando m es pequeño (20 o menos), presenta más masa probabilística en las colas —es decir, tiene una forma acampanada «más gruesa» que la normal. Si m es 30 o mayor, la distribución t de Student se aproxima bien mediante la distribución normal estándar y la distribución t_∞ es igual a la distribución normal estándar.

La distribución F

La distribución F con m y n grados de libertad, expresada mediante $F_{m,n}$, se define como la distribución del cociente entre una variable aleatoria chi-cuadrado con m grados de libertad, dividida por m , y una variable aleatoria chi-cuadrado independientemente distribuida con n grados de libertad, dividida por n . Matemáticamente, sea W una variable aleatoria chi-cuadrado con m grados de libertad y sea V una variable aleatoria chi-cuadrado con n grados de libertad, estando W y V independientemente distribuidas. Entonces $\frac{W/m}{V/n}$ presenta una distribución $F_{m,n}$ —es decir, una distribución F con m grados de libertad en el numerador y con n grados de libertad en el denominador.

En estadística y econometría, un caso particular importante de la distribución F surge cuando los grados de libertad del denominador son lo suficientemente elevados como para que la distribución $F_{m,n}$ sea aproximadamente una distribución $F_{m,\infty}$. En este caso límite, la variable aleatoria del denominador, V , es la media de infinitas variables aleatorias chi-cuadrado, y esa media es 1 porque la media de una variable aleatoria normal estándar al cuadrado es 1 (véase Ejercicio 2.24). Por tanto, la distribución $F_{m,\infty}$ es la distribución de una variable aleatoria chi-cuadrado con m grados de libertad, dividida por m : W/m se distribuye $F_{m,\infty}$. Por ejemplo, según la Tabla 4 del Apéndice, el percentil 95 de la distribución $F_{3,\infty}$ es 2,60, que es igual al percentil 95 de la distribución χ^2_3 , 7,81 (de la Tabla 2 del Apéndice), dividido por los grados de libertad, que son 3 ($7,81/3 = 2,60$).

Se recogen los percentiles 90, 95, y 99 de la distribución $F_{m,n}$ en la Tabla 5 del Apéndice para algunos valores de m y n . Por ejemplo, el percentil 95 de la distribución $F_{3,30}$ es 2,92, y el percentil 95 de la distribución $F_{3,90}$ es 2,71. A medida que aumenten los grados de libertad del denominador, n , el percentil 95 de la distribución $F_{3,n}$ tiende al límite de $F_{3,\infty}$ de 2,60.

2.5 Muestreo aleatorio y distribución de la media muestral

Casi todos los procedimientos estadísticos y econométricos utilizados en este libro incluyen medias o medias ponderadas de una muestra de datos. Por tanto, la caracterización de las distribuciones de las medias muestrales es un paso esencial hacia la comprensión del funcionamiento de los procedimientos econométricos.

Esta sección introduce algunos conceptos básicos sobre el muestreo aleatorio y la distribución de medias que se utilizan a lo largo de este libro. Comenzamos tratando el muestreo aleatorio. La acción del muestreo aleatorio —es decir, la extracción aleatoria de una muestra a partir de una población mayor— causa el efecto de hacer de la media muestral una variable aleatoria en sí misma. Al ser la media muestral una variable aleatoria, posee una distribución de probabilidad, que se denomina su distribución muestral. Esta sección concluye con algunas propiedades de la distribución muestral de la media muestral.

Muestreo aleatorio

Muestreo aleatorio simple. Supongamos que nuestra estudiante viajera de la Sección 2.1 aspira a ser estadística de profesión y decide recopilar los tiempos de sus desplazamientos a lo largo de varios días. Ella selecciona esos días aleatoriamente a lo largo de su año académico, y su tiempo de desplazamiento diario posee la función de distribución acumulada de la Figura 2.2a. Por estar estos días seleccionados aleatoriamente, el conocimiento del valor del tiempo de desplazamiento en uno de esos días seleccionados aleatoriamente no proporciona información sobre el tiempo de desplazamiento en otro de los días; es decir, como los días fueron seleccionados aleatoriamente, los valores de los tiempos de desplazamiento de cada uno de los otros días son variables aleatorias independientemente distribuidas.

La situación descrita en el párrafo anterior es un ejemplo del diseño de muestreo más sencillo utilizado en estadística, llamado **muestreo aleatorio simple**, en el que se seleccionan aleatoriamente n objetos a partir de una **población** (la población de días de desplazamiento).

Las n observaciones de la muestra se expresan mediante Y_1, Y_2, \dots, Y_n , donde Y_1 es la primera observación, Y_2 es la segunda observación, etcétera. En el ejemplo de los desplazamientos, Y_1 es el tiempo de desplazamiento del primero de los n días seleccionados y Y_i es el tiempo de desplazamiento del i -ésimo de los días seleccionados aleatoriamente.

Debido a que los miembros de la población incluidos en la muestra fueron seleccionados aleatoriamente, los valores de las observaciones Y_1, Y_2, \dots, Y_n son asimismo aleatorios. Si se seleccionan diferentes miembros de la población, los valores de Y serán distintos. Así, el hecho de seleccionar la muestra aleatoriamente significa que Y_1, Y_2, \dots, Y_n pueden ser tratados como variables aleatorias. Antes del muestreo, Y_1, Y_2, \dots, Y_n pueden tomar muchos valores diferentes; tras haber sido seleccionados, se registra un valor específico para cada observación.

**CONCEPTO
CLAVE**
2.5**Muestro aleatorio simple y variables aleatorias i.i.d.**

En un muestreo aleatorio simple, se seleccionan aleatoriamente n objetos de una población y cada objeto tiene la misma probabilidad de ser seleccionado. El valor de la variable aleatoria Y para el objeto i -ésimo seleccionado aleatoriamente se expresa mediante Y_i . Como cada objeto tiene la misma probabilidad de ser seleccionado y la distribución de Y_i es la misma para todo i , las variables aleatorias Y_1, Y_2, \dots, Y_n son independientes e idénticamente distribuidas (i.i.d.); es decir, la distribución de Y_i es la misma para todo $i = 1, \dots, n$ e Y_i está independientemente distribuida de Y_2, \dots, Y_n , etc.

Extracciones i.i.d. Al ser Y_1, Y_2, \dots, Y_n extracciones aleatorias de una misma población, la distribución marginal de Y_i es la misma para cada $i = 1, \dots, n$; esta distribución marginal es la distribución de Y en la población de la que procede la muestra. Cuando Y_i tiene la misma distribución marginal para $i = 1, \dots, n$, entonces se dice que Y_1, Y_2, \dots, Y_n están **idénticamente distribuidas**.

En muestreo aleatorio simple, conocer el valor de Y_1 no proporciona información acerca de Y_2 , por lo que la distribución condicional de Y_2 dada Y_1 es la misma que la distribución marginal de Y_2 . En otras palabras, en muestreo aleatorio simple, Y_1 se distribuye independientemente de Y_2, \dots, Y_n .

Cuando Y_1, Y_2, \dots, Y_n son extracciones de la misma distribución y están independientemente distribuidas, se dice que son **independientes idénticamente distribuidas** (o **i.i.d.**).

El muestreo aleatorio simple y la selección i.i.d. están recogidos en el Concepto clave 2.5.

La distribución muestral de la media muestral

La **media muestral** o **promedio muestral**, \bar{Y} , de las n observaciones Y_1, Y_2, \dots, Y_n es

$$\bar{Y} = \frac{1}{n} (Y_1 + Y_2 + \dots + Y_n) = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (2.43)$$

Un concepto esencial es que el hecho de seleccionar la muestra de forma aleatoria produce el efecto de hacer que la media muestral \bar{Y} sea una variable aleatoria. Por estar la muestra seleccionada aleatoriamente, el valor de cada Y_i es aleatorio. Como Y_1, Y_2, \dots, Y_n son aleatorias, su media es aleatoria. Si se hubiera seleccionado una muestra diferente, entonces las observaciones y su media muestral habrían sido diferentes: El valor de \bar{Y} es distinto entre distintas muestras extraídas aleatoriamente.

Por ejemplo, supóngase que nuestra estudiante viajera seleccionó cinco días aleatoriamente para recopilar sus tiempos de desplazamiento, y que calculó la media para esos cinco tiempos. Si hubiera elegido cinco días diferentes, habría registrado cinco tiempos diferentes, y por tanto habría calculado un valor diferente de la media muestral.

Por ser \bar{Y} aleatoria, tiene una distribución de probabilidad. La distribución de \bar{Y} se denomina la **distribución muestral** de \bar{Y} porque es la distribución de probabilidad asociada a los posibles valores de \bar{Y} que podrían haberse calculado para diferentes muestras de Y_1, Y_2, \dots, Y_n .

La distribución muestral de las medias y las medias ponderadas interpretan un papel central en estadística y econometría. Comenzamos nuestro estudio de la distribución muestral de \bar{Y} calculando su media y varianza bajo condiciones generales de la distribución poblacional de Y .

Media y varianza de \bar{Y} . Supongamos que las observaciones Y_1, Y_2, \dots, Y_n son i.i.d., y sean μ_Y y σ_Y^2 la media y la varianza de Y_i (por ser las observaciones i.i.d., la media y la varianza son las mismas para todo $i = 1, \dots, n$). Cuando $n = 2$, la media de la suma $Y_1 + Y_2$ se obtiene aplicando la Ecuación (2.28): $E(Y_1 + Y_2) = \mu_Y + \mu_Y = 2\mu_Y$. Por tanto la media de la media muestral es $E[\frac{1}{2}(Y_1 + Y_2)] = \frac{1}{2} \times 2\mu_Y = \mu_Y$. En general,

$$E(\bar{Y}) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \mu_Y. \quad (2.44)$$

La varianza de \bar{Y} se halla aplicando la Ecuación (2.37). Por ejemplo, para $n = 2$, $\text{var}(Y_1 + Y_2) = 2\sigma_Y^2$, por tanto [aplicando la Ecuación (2.31) con $a = b = \frac{1}{2}$ y $\text{cov}(Y_1, Y_2) = 0$], $\text{var}(\bar{Y}) = \frac{1}{2}\sigma_Y^2$. Para cualquier n , por ser Y_1, Y_2, \dots, Y_n son i.i.d., Y_i e Y_j están independientemente distribuidas para $i \neq j$, por lo que $\text{cov}(Y_i, Y_j) = 0$. Por tanto,

$$\begin{aligned}\text{var}(\bar{Y}) &= \text{var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(Y_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j \neq 1}^n \text{cov}(Y_i, Y_j) \\ &= \frac{\sigma_Y^2}{n}.\end{aligned}\tag{2.45}$$

La desviación típica de \bar{Y} es la raíz cuadrada de la varianza, $\sigma_{\bar{Y}}/\sqrt{n}$.

En resumen, la media, la varianza, y la desviación típica de \bar{Y} son

$$E(\bar{Y}) = \mu_Y.\tag{2.46}$$

$$\text{var}(\bar{Y}) = \sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}, \text{ y}\tag{2.47}$$

$$\text{desv. t\'ip. } (\bar{Y}) = \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}.\tag{2.48}$$

Esos resultados son válidos cualquiera que sea la distribución de Y_i ; es decir la distribución de Y_i no necesita tomar una forma específica, como la de la distribución normal, para que se cumplan las Ecuaciones (2.46) a (2.48).

La notación $\sigma_{\bar{Y}}^2$ expresa la varianza de la distribución muestral de la media muestral \bar{Y} . Por el contrario, σ_Y^2 es la varianza de cada Y_i individual, es decir, la varianza de la distribución poblacional de la cual se ha extraído cada observación. De forma similar, $\sigma_{\bar{Y}}$ es la desviación típica de la distribución muestral de \bar{Y} .

Distribución muestral de \bar{Y} cuando Y está distribuida normalmente. Supongamos que Y_1, Y_2, \dots, Y_n son i.i.d. extraídas de la distribución $N(\mu_Y, \sigma_Y^2)$. Como se estableció partiendo de la Ecuación (2.42), la suma de n variables aleatorias normalmente distribuidas está distribuida normalmente. Al ser la media de \bar{Y} , μ_Y y la varianza de \bar{Y} $\sigma_{\bar{Y}}^2/n$, esto significa que si Y_1, \dots, Y_n son i.i.d. extraídas a partir de $N(\mu_Y, \sigma_Y^2)$, entonces \bar{Y} se distribuye $N(\mu_Y, \sigma_{\bar{Y}}^2/n)$.

Diversificación financiera y carteras

El principio de diversificación dice que es posible reducir el riesgo manteniendo pequeñas inversiones en varios activos, en comparación con la colocación de todo su dinero en un único activo. Es decir, no deberían ponerse todos los huevos en la misma cesta.

Las matemáticas de la diversificación se deducen de la Ecuación (2.45). Supongamos que se reparte 1 \$ de manera equitativa entre n activos. Sea Y_i el flujo de pagos generados en 1 año por 1 \$ invertido en el activo i -ésimo. Como se han invertido $1/n$ dólares en cada activo, el flujo real de la cartera tras el trascurso de 1 año es $(Y_1 + Y_2 + \dots + Y_n)/n = \bar{Y}$. Con el fin de mantener la simplicidad, supongamos que cada activo tiene el mismo flujo de pagos esperado, μ_Y , la misma varianza, σ^2 , y la correlación positiva ρ entre los activos es la misma [por lo que $\text{cov}(Y_i, Y_j) = \rho\sigma^2$]. Por tanto el flujo de pagos esperado es $E\bar{Y} = \mu_Y$, y, para n grande, la varianza del

flujo de pagos de toda la cartera es $\text{var}(\bar{Y}) = \rho\sigma^2$ (Ejercicio 2.26). La colocación de todo el dinero en un solo activo o el reparto equitativo entre los n activos presentan el mismo flujo de pagos esperado, pero la diversificación reduce la varianza desde σ^2 a $\rho\sigma^2$.

Las matemáticas de la diversificación han conducido a productos financieros tales como los fondos de inversión colectiva en acciones, en los que el fondo mantiene muchas acciones mientras un individuo posee una participación del fondo, poseyendo así una pequeña cantidad de muchas acciones. Pero la diversificación tiene sus límites: para muchos activos, los pagos se encuentran correlacionados positivamente por lo que la $\text{var}(\bar{Y})$ sigue siendo positiva incluso si n es grande. En el caso de las acciones, el riesgo se reduce manteniendo una cartera, pero esta cartera está sujeta a las fluctuaciones globales impredecibles de las bolsas.

2.6 Aproximación para muestras grandes de las distribuciones muestrales

Las distribuciones muestrales desempeñan un papel central en el desarrollo de los procedimientos estadísticos y econométricos, por lo que es importante conocer, en forma matemática, cuál es la distribución muestral de \bar{Y} . Existen dos métodos para la caracterización de las distribuciones muestrales: el método «exacto» y el método «aproximado».

El método «exacto» conlleva la obtención de una fórmula para la distribución muestral que se cumpla con exactitud para cualquier valor de n . La distribución muestral que describe exactamente la distribución de \bar{Y} para cualquier n se llama la **distribución exacta** o **distribución en muestras finitas** de \bar{Y} . Por ejemplo, si Y está distribuida normalmente y Y_1, \dots, Y_n son i.i.d., entonces (como se vio en la Sección 2.5) la distribución exacta de \bar{Y} es normal con media μ_Y y varianza σ_Y^2/n . Desafortunadamente, si la distribución de Y no es normal, en general, la distribución muestral exacta de \bar{Y} será muy complicada y dependerá de la distribución de Y .

El método «aproximado» utiliza aproximaciones para la distribución muestral que se basan en el hecho de que el tamaño muestral es grande. La aproximación para muestras grandes de la distribución muestral a menudo se denomina **distribución asintótica** —«asintótica» porque la aproximación se hace exacta en el límite, cuando $n \rightarrow \infty$. Como vemos en esta sección estas aproximaciones pueden ser muy precisas incluso si el tamaño muestral es solamente de $n = 30$ observaciones. Como los tamaños muestrales utilizados en la práctica econométrica por lo general ascienden a miles o al menos a cientos, puede considerarse que estas distribuciones asintóticas proporcionan unas muy buenas aproximaciones de la distribución muestral exacta.

Esta sección presenta las dos herramientas clave utilizadas para aproximar distribuciones muestrales cuando el tamaño muestral es grande: la ley de los grandes números y el teorema central del límite. La ley de los grandes números dice que, cuando el tamaño muestral es elevado, \bar{Y} estará cerca de μ_Y con probabilidad muy elevada. El teorema central del límite dice que, cuando el tamaño muestral es elevado, la distribución muestral de la media muestral estandarizada $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$, es aproximadamente normal.

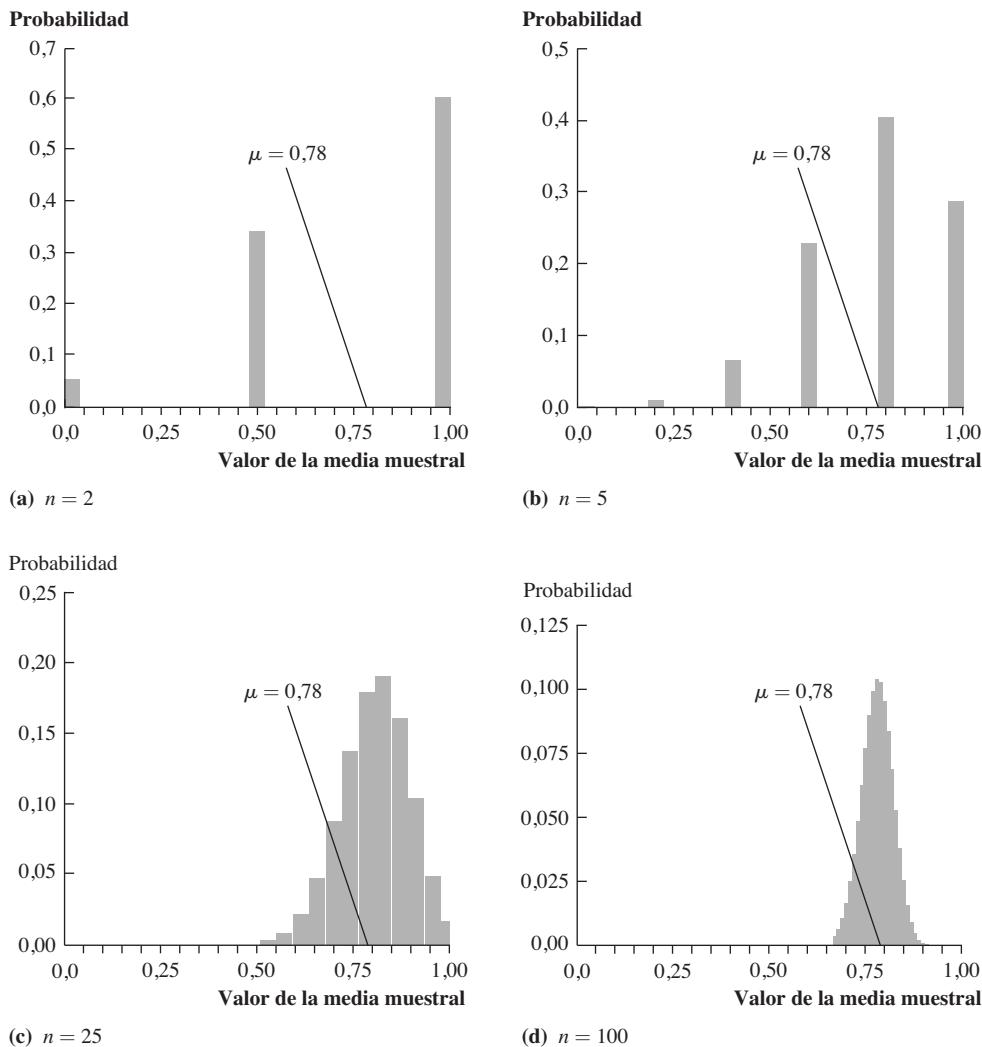
Aunque las distribuciones muestrales exactas son complicadas y dependen de la distribución de Y , las distribuciones asintóticas son sencillas. Además —hay que subrayarlo— la distribución asintótica normal de $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ no depende de la distribución de Y . Esta distribución normal aproximada proporciona simplificaciones enormes y es la base de la teoría de la regresión utilizada a lo largo de este libro.

La ley de los grandes números y la consistencia

La **ley de los grandes números** establece que, bajo condiciones generales, \bar{Y} estará cerca de μ_Y con muy alta probabilidad cuando n es grande. Esto a veces se denomina la «ley de promedios». Cuando un número elevado de variables aleatorias con la misma media se promedian conjuntamente, los valores altos compensan los valores pequeños y su media muestral estará cercana a su media común.

Por ejemplo, considérese una versión simplificada de nuestro experimento de la estudiante viajera en el que ella simplemente registraba si su tiempo de desplazamiento había sido corto (menos de 20 minutos) o largo. Sea Y_i igual a 1 si su desplazamiento fue corto en el i -ésimo día aleatoriamente elegido e igual a 0 si fue largo. Debido a que ella utilizaba muestreo aleatorio simple, Y_1, \dots, Y_n son i.i.d. Por lo tanto, $Y_i, i = 1, \dots, n$ son extracciones i.i.d. de una variable aleatoria de Bernoulli, donde (a partir de la Tabla 2.2) la probabilidad de que $Y_i = 1$ es 0,78. Debido a que la esperanza de una variable aleatoria de Bernoulli es su probabilidad de éxito, $E(Y_i) = \mu_Y = 0,78$. La media muestral \bar{Y} es la proporción de días en su muestra en los cuales sus desplazamientos fueron cortos.

La Figura 2.8 muestra la distribución muestral de \bar{Y} para varios tamaños muestrales n . Cuando $n = 2$ (Figura 2.8a), \bar{Y} puede tomar solamente tres valores: 0, $\frac{1}{2}$, y 1 (ningún desplazamiento fue corto, uno fue corto, o ambos fueron cortos), ninguno de los cuales está particularmente cercano a la verdadera proporción poblacional, 0,78. No obstante, cuando n aumenta (Figuras 2.8b-d), \bar{Y} toma más valores y la distribución muestral se hace cada vez más centrada en μ_Y .

FIGURA 2.8 Distribución muestral de la media muestral de n variables aleatorias de Bernouilli

Las distribuciones son las distribuciones muestrales de \bar{Y} , la media muestral de n variables aleatorias de Bernouilli con $\rho = \Pr(Y_i = 1) = 0,78$ (la probabilidad de un desplazamiento corto es del 78 %). La varianza de la distribución muestral de \bar{Y} disminuye a medida que n se hace grande, por tanto la distribución muestral va estando cada vez más estrechamente concentrada alrededor de su media $\mu = 0,78$ a medida que el tamaño muestral aumenta.

La propiedad de que \bar{Y} esté cerca de μ_Y con probabilidad creciente cuando n aumenta se denomina **convergencia en probabilidad** o, de forma más concisa, **consistencia** (véase Concepto clave 2.6). La ley de los grandes números establece que, bajo ciertas condiciones, \bar{Y} converge en probabilidad a μ_Y o, de manera equivalente, que \bar{Y} es consistente.

Las condiciones sobre la ley de los grandes números que se utilizan en este libro son que $Y_i, i = 1, \dots, n$ son i.i.d. y que la varianza de Y_i , $\sigma^2_{Y_i}$, es finita. La importancia matemática de esas condiciones se aclara en la Sección 17.2, donde se demuestra la ley de los grandes números. Si los datos se seleccionan mediante muestreo aleatorio simple, entonces se cumple el supuesto de i.i.d. El supuesto de varianza finita dice que los valores extremos altos de Y_i —es decir, los atípicos— son improbables y se observan de manera poco frecuente; de otro modo esos valores altos podrían dominar a \bar{Y} y la media muestral resultaría poco fiable. Este supuesto resulta creíble para las aplicaciones y ejemplos de este libro. Por ejemplo, dado que existe un límite superior para el tiempo de desplazamiento de nuestra estudiante (si el tráfico fuera horroroso podría aparcar y caminar), la varianza de la distribución del tiempo de desplazamiento es finita.

CONCEPTO CLAVE
2.6
Convergencia en probabilidad, consistencia y ley de los grandes números

La media muestral \bar{Y} converge en probabilidad a μ_Y (o de forma equivalente, \bar{Y} es consistente) si la probabilidad de que \bar{Y} se encuentre en el rango $\mu_Y - c$ a $\mu_Y + c$ se hace arbitrariamente cercana a 1 cuando n aumenta para cualquier constante $c > 0$. La convergencia en probabilidad de \bar{Y} a μ_Y se expresa mediante $\bar{Y} \xrightarrow{P} \mu_Y$.

La ley de los grandes números establece que si $Y_i, i = 1, \dots, n$ son independientes e idénticamente distribuidas con $E(Y_i) = \mu_Y$ y si los valores atípicos elevados resultan improbables (técticamente, si $\text{var}(Y_i) = \sigma_Y^2 < \infty$), entonces $\bar{Y} \xrightarrow{P} \mu_Y$.

El teorema central del límite

El **teorema central del límite** establece que, bajo ciertas condiciones, la distribución de \bar{Y} se aproxima a una distribución normal cuando n es grande. Recuérdese que la media de \bar{Y} es μ_Y y su varianza es $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$. De acuerdo con el teorema central del límite, cuando n es grande, la distribución de \bar{Y} es aproximadamente $N(\mu_Y, \sigma_{\bar{Y}}^2)$. Como se estudió al final de la Sección 2.5, la distribución de \bar{Y} es *exactamente* $N(\mu_Y, \sigma_{\bar{Y}}^2)$ cuando la muestra se obtiene a partir de una población con distribución normal $N(\mu_Y, \sigma_Y^2)$. El teorema central del límite establece que este mismo resultado es *aproximadamente* cierto cuando n es grande incluso si Y_1, \dots, Y_n no están distribuidas normalmente.

La convergencia de la distribución de \bar{Y} a la aproximación normal con forma de campana puede verse (un poco) en la Figura 2.8. Sin embargo, dado que la distribución se ajusta bastante para un n elevado, esto requiere algo de esfuerzo visual. La forma de la distribución de \bar{Y} sería más fácil de ver si utilizáramos una lente de aumento o dispusiéramos de alguna forma de realizar un *zoom* o de expandir el eje horizontal de la figura.

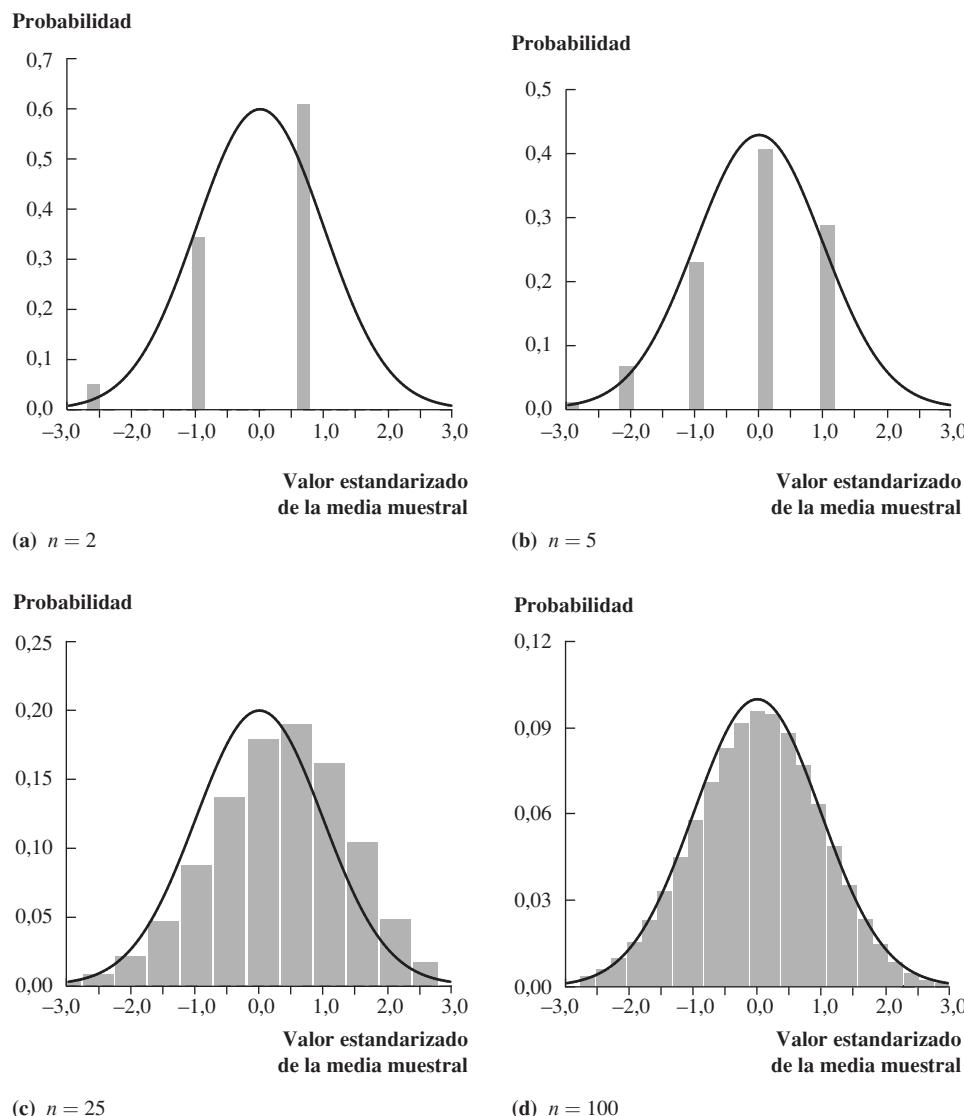
Un modo de hacerlo es estandarizar \bar{Y} restándole su media y dividiéndola por su desviación típica, por lo que tendrá una media de 0 y una varianza de 1. Este proceso conduce a examinar la distribución de la versión estandarizada de \bar{Y} , $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$. De acuerdo con el teorema central del límite, esta distribución se aproximaría adecuadamente mediante una distribución $N(0, 1)$ cuando n es grande.

La distribución de la media estandarizada $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ se representa en la Figura 2.9 para la distribución de la Figura 2.8; las distribuciones de la Figura 2.9 son exactamente las mismas que en la Figura 2.8, excepto por la escala del eje horizontal que se ha cambiado para que la variable estandarizada tenga una media de 0 y una varianza de 1. Después de este cambio de escala, es fácil de ver que, si n es suficientemente grande, la distribución de \bar{Y} se aproxima adecuadamente mediante una distribución normal.

Podríamos preguntarnos, ¿cuánto es «suficientemente grande»? Es decir, ¿cómo de grande debe ser n para que la distribución de \bar{Y} sea aproximadamente normal? La respuesta es «depende». La calidad de la aproximación normal depende de la distribución de Y_i subyacente que conforma la media. Por un lado, si las Y_i están distribuidas normalmente, entonces \bar{Y} está distribuida normalmente de forma exacta para todo n . Por el contrario, cuando las Y_i subyacentes presentan por sí mismas una distribución muy distinta de la normal, entonces esta aproximación puede requerir una $n = 30$ o incluso mayor.

Este hecho se ilustra en la Figura 2.10 para una distribución poblacional, como la mostrada en la Figura 2.10a, que es bastante diferente de la distribución de Bernouilli. Esta distribución posee una cola derecha larga (es «asimétrica» derecha). La distribución muestral de \bar{Y} , tras haber sido centrada y reescalada, se muestra en las Figuras 2.10b-d para $n = 5, 25$, y 100 , respectivamente. Aunque la distribución muestral se aproxime a la forma de campana para $n = 25$, la aproximación normal presenta todavía algunas imperfecciones. Para $n = 100$, sin embargo, la aproximación normal es bastante buena. De hecho, para $n \geq 100$, la aproximación normal de la distribución de \bar{Y} resulta habitualmente muy buena para una extensa variedad de distribuciones poblacionales.

El teorema central del límite es un resultado destacado. Mientras las distribuciones de \bar{Y} con « n pequeño» de los apartados b y c de las Figuras 2.9 y 2.10 son complicadas y bastante diferentes una de otra, las distribuciones con « n grande» de las Figuras 2.9d y 2.10d son sencillas, y asombrosamente, tienen una forma

FIGURA 2.9 Distribución de la media muestral estandarizada de n variables aleatorias de Bernouilli con $p = 0,78$ 

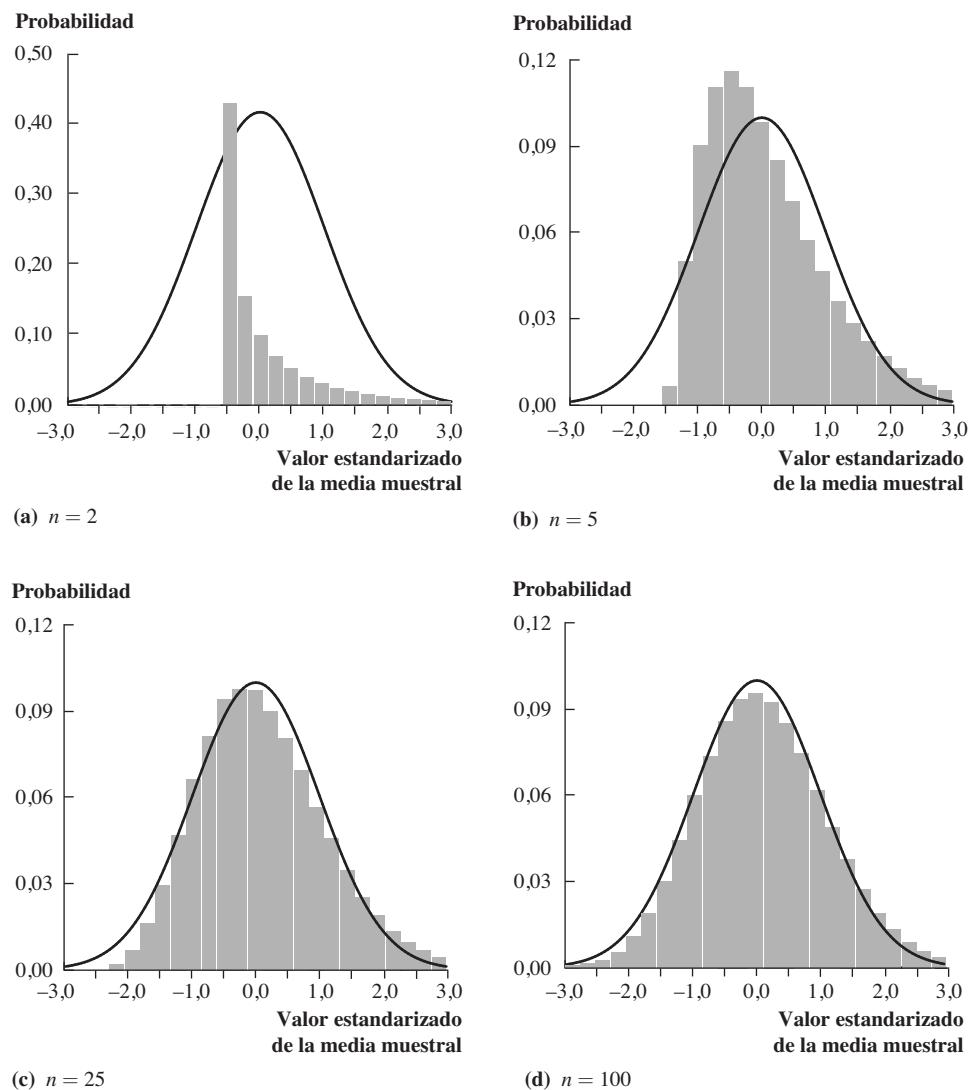
La distribución muestral de \bar{Y} en la Figura 2.8 está representada aquí tras la estandarización de \bar{Y} . Este dibujo centra la distribución de la Figura 2.8 y magnifica la escala del eje horizontal por un factor \sqrt{n} . Cuando el tamaño muestral es grande, las distribuciones muestrales se aproximan cada vez mejor con la distribución normal (la línea gruesa), tal y como predice el teorema central del límite. La distribución muestral ha sido escalada para que la altura de las distribuciones sea aproximadamente la misma en todas las figuras.

CONCEPTO CLAVE
2.7
El teorema central del límite

Supongamos que Y_1, \dots, Y_n son i.i.d. con $E(Y_i) = \mu_Y$ y $\text{var}(Y_i) = \sigma_Y^2$, donde $0 < \sigma_Y^2 < \infty$. A medida que $n \rightarrow \infty$, la distribución $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ (donde $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$) se approxima arbitrariamente bien a la distribución normal estándar.

similar. Debido a que la distribución de \bar{Y} se approxima a la normal cuando n aumenta mucho, se dice que \bar{Y} tiene una **distribución asintóticamente normal**.

La adecuación de la aproximación normal, combinada con su extensa aplicabilidad debido al teorema central del límite, la convierte en un punto crucial para la econometría aplicada moderna. El teorema central del límite está recogido en el Concepto clave 2.7.

FIGURA 2.10 Distribución de la media muestral estandarizada de n extracciones de una distribución asimétrica

Las figuras muestran la distribución muestral de la media muestral estandarizada de n extracciones de una distribución poblacional asimétrica (sesgada) que se muestra en la Figura 2.10a. Cuando n es pequeño ($n = 5$), la distribución muestral, así como la distribución poblacional, es asimétrica. Pero cuando n es grande ($n = 100$), la distribución muestral se aproxima bien mediante una distribución normal estándar (línea gruesa), tal y como predice el teorema central del límite. La distribución normal se ha reescalado para que la altura de las distribuciones sea aproximadamente la misma en todas las figuras.

Resumen

1. Las probabilidades con las que una variable aleatoria toma diferentes valores están recogidas por la función de distribución acumulada, la función de distribución de probabilidad (para las variables aleatorias discretas), y la función de densidad de probabilidad (para las variables aleatorias continuas).
2. El valor esperado de una variable aleatoria Y (denominado asimismo, su media, μ_Y), se expresa mediante $E(Y)$, es su valor medio ponderado por probabilidad. La varianza de Y es $\sigma_Y^2 = E[(Y - \mu_Y)^2]$, y la desviación típica de Y es la raíz cuadrada de su varianza.
3. Las probabilidades conjuntas de dos variables aleatorias X e Y están resumidas por su distribución de probabilidad conjunta. La distribución de probabilidad condicional de Y dado $X = x$ es la distribución de probabilidad de Y , condicionada a que X tome el valor x .

4. Una variable aleatoria normalmente distribuida presenta la densidad de probabilidad con forma de campana de la Figura 2.5. Para calcular una probabilidad asociada a una variable aleatoria normal, antes se estandariza la variable, y después se utiliza la distribución normal estándar acumulada tabulada en la Tabla 1 del Apéndice.
5. El muestreo aleatorio simple genera n observaciones aleatorias Y_1, \dots, Y_n que son independientes e idénticamente distribuidas (i.i.d.).
6. La media muestral \bar{Y} , varía de una muestra obtenida aleatoriamente a otra y por lo tanto es una variable aleatoria con una distribución muestral. Si Y_1, \dots, Y_n son i.i.d., entonces:
 - a) La distribución muestral de \bar{Y} tiene media $\mu_{\bar{Y}}$ y varianza $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$.
 - b) La ley de los grandes números establece que \bar{Y} converge en probabilidad a μ_Y .
 - c) El teorema central del límite establece que la versión estandarizada de \bar{Y} , $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$, presenta una distribución normal estándar [distribución $N(0, 1)$] cuando n es grande.

Términos clave

resultados (12)	varianza condicional (22)
probabilidad (12)	independientemente distribuida (22)
espacio muestral (12)	independiente (22)
suceso (12)	covarianza (23)
variable aleatoria discreta (12)	correlación (23)
variable aleatoria continua (12)	incorrelacionada (23)
distribución de probabilidad (12)	distribución normal (26)
distribución de probabilidad acumulada (12)	distribución normal estándar (26)
función de distribución acumulada (f.d.a.) (13)	estandarizar la variable (26)
variable aleatoria de Bernouilli (13)	distribución normal multivariante (28)
distribución de Bernouilli (13)	distribución normal bivariante (28)
función de densidad de probabilidad (f.d.p.) (14)	distribución chi-cuadrado (28)
función de densidad (14)	distribución t de Student (28)
densidad (14)	distribución t (30)
valor esperado (15)	distribución F (30)
esperanza (15)	muestreo aleatorio simple (31)
media (15)	población (31)
varianza (16)	idénticamente distribuida (32)
desviación típica (16)	independientes e idénticamente
momentos de una distribución (17)	distribuidas (i.i.d.) (32)
asimetría (17)	promedio muestral (32)
curtosis (18)	media muestral (32)
atípico (18)	distribución muestral (32)
leptocútica (19)	distribución exacta (muestras finitas) (34)
momento r -ésimo (19)	distribución asintótica (34)
distribución de probabilidad conjunta (19)	ley de los grandes números (34)
distribución de probabilidad marginal (20)	convergencia en probabilidad (35)
distribución condicional (20)	consistencia (35)
esperanza condicional (21)	teorema central del límite (36)
media condicional (21)	distribución asintóticamente normal (37)
ley de esperanzas iteradas (22)	

Revisión de conceptos

2.1 Los ejemplos de variables aleatorias utilizadas en este capítulo incluyen:

- a) El género de la próxima persona que va a conocer.
- b) El número de veces que se estropea un ordenador.

- c) El tiempo empleado en el desplazamiento a la escuela.
- d) Si el ordenador que le asignan en la biblioteca es nuevo o viejo.
- e) Si llueve o no.

Explique por qué cada una de ellas puede considerarse aleatoria.

- 2.2 Supóngase que las variables aleatorias X e Y son independientes y que sus distribuciones son conocidas. Explique por qué conocer el valor de X no dice nada acerca del valor de Y .
- 2.3 Supóngase que X representa la cantidad de precipitaciones en su ciudad durante un mes determinado e Y representa el número de niños nacidos en Los Angeles durante el mismo mes. ¿Son independientes X e Y ? Explíquelo.
- 2.4 Una clase de econometría tiene 80 estudiantes, y el peso medio de los estudiantes es de 145 lb¹. Se selecciona una muestra aleatoria de cuatro estudiantes de la clase, y se calcula su peso medio. ¿El peso medio de los estudiantes de la muestra será igual a 145 lb? ¿Por qué o por qué no? Utilice este ejemplo para explicar por qué la media muestral \bar{Y} , es una variable aleatoria.
- 2.5 Supóngase que Y_1, \dots, Y_n son variables aleatorias i.i.d. con distribución $N(1, 4)$. Dibuje la densidad de probabilidad de \bar{Y} para $n = 2$. Repítalo para $n = 10$ y $n = 100$. Describa en palabras las diferencias entre las densidades. ¿Cuál es la relación entre su respuesta y la ley de los grandes números?
- 2.6 Supóngase que Y_1, \dots, Y_n son variables aleatorias i.i.d. con la distribución de probabilidad dada en la Figura 2.10a. Si se desea calcular $\Pr(\bar{Y} \leq 0,1)$, ¿sería razonable utilizar la aproximación normal si $n = 5$? ¿Y si fuera $n = 25$ o $n = 100$? Explíquelo.
- 2.7 Y es una variable aleatoria con $\mu_Y = 0$, $\sigma_Y = 1$, asimetría = 0, y curtosis = 100. Dibuje una distribución de probabilidad hipotética de Y . Explique por qué n variables aleatorias extraídas de esta distribución podrían presentar algunos valores atípicos elevados.

Ejercicios

- 2.1 Sea Y el número de «caras» que salen al lanzar dos monedas.
 - a) Obtenga la distribución de probabilidad de Y .
 - b) Obtenga la distribución de probabilidad acumulada de Y .
 - c) Obtenga la media y la varianza de Y .
- 2.2 Utilice la distribución de probabilidad proporcionada en la Tabla 2.2 para calcular:
 - a) $E(Y)$ y $E(X)$
 - b) σ_X^2 y σ_Y^2
 - c) σ_{XY} y $\text{corr}(X, Y)$
- 2.3 Utilizando las variables aleatorias X e Y de la Tabla 2.2, considérense dos nuevas variables aleatorias $W = 3 + 6X$ y $V = 20 - 7Y$. Calcule:
 - (a) $E(W)$ y $E(V)$
 - b) σ_W^2 y σ_V^2
 - c) σ_{WV} y $\text{corr}(W, V)$
- 2.4 Supóngase que X es una variable aleatoria de Bernoulli con $P(X = 1) = p$.
 - a) Demuestre que $E(X^3) = p$.
 - b) Demuestre que $E(X^k) = p$ para $k > 0$.
 - c) Supóngase que $p = 0,3$. Calcule la media, varianza, asimetría y curtosis de X . (Pista: puede resultar útil la utilización de las fórmulas dadas en el Ejercicio 2.21).
- 2.5 En septiembre, la temperatura máxima diaria en Seattle presenta una media de 70 °F² y una desviación típica de 7 °F. ¿Cuál es la media, desviación típica, y varianza en °C?
- 2.6 La tabla siguiente proporciona la distribución de probabilidad conjunta entre situación laboral y titulación universitaria entre los que se encuentran tanto empleados como buscando trabajo (desempleados) dentro de la población en edad de trabajar de EE.UU. en 2008.

¹ N. del T.: Lb = libra (1 libra = 450 g).

² N. del T.: Aproximadamente 21 °C.

Distribución conjunta de situación laboral y titulación de la población mayor de 25 años en EE.UU., 2008

	Desempleados ($Y = 0$)	Empleados ($Y = 1$)	Total
Titulados no universitarios ($X = 0$)	0,037	0,622	0,659
Titulados universitarios ($X = 1$)	0,009	0,332	0,341
Total	0,046	0,954	1,000

- a) Calcule $E(Y)$.
- b) La tasa de desempleo es la proporción de la fuerza laboral que se encuentra desempleada. Demuestre que la tasa de desempleo está dada por $1 - E(Y)$.
- c) Calcule $E(Y|X = 1)$ y $E(Y|X = 0)$.
- d) Calcule la tasa de desempleo para (i) titulados universitarios y (ii) titulados no universitarios.
- e) Un miembro de esta población seleccionado aleatoriamente dice estar desempleado. ¿Cuál es la probabilidad de que este trabajador sea titulado universitario? ¿Y titulado no universitario?
- f) ¿Son independientes los logros educativos y la situación laboral?
- 2.7 En una población dada de dos parejas hombre/mujer asalariadas, los ingresos masculinos presentan una media de 40.000 \$ al año y una desviación típica de 12.000 \$. Los ingresos femeninos presentan una media de 45.000 \$ al año y una desviación típica de 18.000 \$. La correlación entre los ingresos masculinos y femeninos para una pareja es 0,80. Sean C los ingresos combinados de una pareja seleccionada al azar.
- a) ¿Cuál es la media de C ?
- b) ¿Cuál es la covarianza entre los ingresos masculinos y femeninos?
- c) ¿Cuál es la desviación típica de C ?
- d) Convierta las respuestas de la (a) a la (c) de dólares USA a euros (€).
- 2.8 La variable aleatoria Y tiene una media de 1 y una varianza de 4. Sea $Z = \frac{1}{2}(Y - 1)$. Demuestre que $\mu_Z = 0$ y $\sigma_Z^2 = 1$.
- 2.9 Sean X e Y variables aleatorias discretas con la distribución conjunta siguiente:

		Valor de Y					
		14	22	30	40	65	
Valor de X		1	0,02	0,05	0,10	0,03	0,01
		5	0,17	0,15	0,05	0,02	0,01
	8	0,02	0,03	0,15	0,10	0,09	

Es decir, $\Pr(X = 1, Y = 14) = 0,02$, etc.

- a) Calcule la distribución de probabilidad, media, y varianza de Y .
- b) Calcule la distribución de probabilidad, media, y varianza de Y dado $X = 8$.
- c) Calcule la covarianza y la correlación entre X e Y .
- 2.10 Calcule las siguientes probabilidades:
- a) Si Y está distribuida $N(1, 4)$, hallar $\Pr(Y \leq 3)$.
- b) Si Y está distribuida $N(3, 9)$, hallar $\Pr(Y > 0)$.
- c) Si Y está distribuida $N(50, 25)$, hallar $\Pr(40 \leq Y \leq 52)$.
- d) Si Y está distribuida $N(5, 2)$, hallar $\Pr(6 \leq Y \leq 8)$.

2.11 Calcule las siguientes probabilidades:

- Si Y está distribuida χ^2_4 , hallar $\Pr(Y \leq 7,78)$.
- Si Y está distribuida χ^2_{10} , hallar $\Pr(Y > 18,31)$.
- Si Y está distribuida $F_{10, \infty}$ hallar $\Pr(Y > 1,83)$.
- ¿Por qué las respuestas a (b) y (c) son las mismas?
- Si Y está distribuida χ^2_1 , hallar $\Pr(Y \leq 1,0)$. (*Pista:* utilizar la definición de la distribución χ^2_1).

2.12 Calcule las siguientes probabilidades:

- Si Y está distribuida t_{15} , hallar $\Pr(Y > 1,75)$.
- Si Y está distribuida t_{90} , hallar $\Pr(-1,99 \leq Y \leq 1,99)$.
- Si Y está distribuida $N(0, 1)$, hallar $\Pr(-1,99 \leq Y \leq 1,99)$.
- ¿Por qué las respuestas a (b) y (c) son aproximadamente las mismas?
- Si Y está distribuida $F_{7,4}$ hallar $\Pr(Y > 4,12)$.
- Si Y está distribuida $F_{7,120}$ hallar $\Pr(Y > 2,79)$.

2.13 X es una variable aleatoria de Bernouilli con $\Pr(X = 1) = 0,99$, Y está distribuida $N(0, 1)$, W está distribuida $N(0, 100)$, X , y W son independientes. Sea $S = XY + (1 - X)W$. (Es decir, $S = Y$ cuando $X = 1$, y $S = W$ cuando $X = 0$).

- Demuestre que $E(Y^2) = 1$ y $E(W^2) = 100$.
- Demuestre que $E(Y^3) = 0$ y $E(W^3) = 0$. (*Pista:* ¿cuál es la asimetría de una distribución simétrica?).
- Demuestre que $E(Y^4) = 3$ y $E(W^4) = 3 \times 100^2$. (*Pista:* utilizar el hecho de que en una distribución normal la curtosis es 3).
- Obtenga $E(S)$, $E(S^2)$, $E(S^3)$ y $E(S^4)$. (*Pista:* utilizar la ley de esperanzas iteradas condicionado a $X = 0$ y $X = 1$).
- Obtenga la asimetría y la curtosis de S .

2.14 En una población $\mu_Y = 100$ y $\sigma_Y^2 = 43$. Utilice el teorema central del límite para resolver las siguientes cuestiones:

- En una muestra aleatoria de tamaño $n = 100$, hallar $\Pr(\bar{Y} \leq 101)$.
- En una muestra aleatoria de tamaño $n = 165$, hallar $\Pr(\bar{Y} > 98)$.
- En una muestra aleatoria de tamaño $n = 64$, hallar $\Pr(101 \leq \bar{Y} \leq 103)$.

2.15 Supóngase las variables aleatorias Y_i con $i = 1, \dots, n$, cada una de ellas distribuida $N(10, 4)$.

- Calcule $\Pr(9,6 \leq \bar{Y} \leq 10,4)$ cuando (i) $n = 20$, (ii) $n = 100$, y (iii) $n = 1.000$.
- Supóngase que c es un número positivo. Demuestre que $\Pr(10 - c \leq \bar{Y} \leq 10 + c)$ se hace más cercano a 1,0 cuando n crece mucho.
- Utilice su respuesta al apartado (b) para argumentar que \bar{Y} converge en probabilidad a 10.

2.16 Y se distribuye $N(5, 100)$ y se desea calcular $\Pr(Y < 3,6)$. Desafortunadamente, no se dispone de un libro de texto y tampoco acceso a una tabla de probabilidades de la distribución normal como la Tabla 1 del Apéndice. Sin embargo, dispone de su ordenador y de un programa informático capaz de generar observaciones i.i.d. de la distribución $N(5, 100)$. Explique cómo puede utilizarse el ordenador para calcular una aproximación precisa para $\Pr(Y < 3,6)$.

2.17 Y_i con $i = 1, \dots, n$ son variables aleatorias con $p = 0,4$. Sea \bar{Y} la expresión de la media muestral

- Utilice el teorema central del límite para calcular las aproximaciones de
 - $\Pr(\bar{Y} \geq 0,43)$ con $n = 100$.
 - $\Pr(\bar{Y} \leq 0,37)$ con $n = 400$.
- ¿A cuánto debe ascender el tamaño muestral para asegurar que $\Pr(0,39 \leq \bar{Y} \leq 0,41) \geq 0,95$? (Utilice el teorema central del límite para calcular una respuesta aproximada).

2.18 A lo largo del año, la meteorología puede infligir daños a una vivienda. De un año a otro, los daños son aleatorios. Sea Y el valor monetario en dólares de los daños a lo largo de un año dado. Supóngase que el 95 % de los años $Y = 0 \$$, pero el 5 % de los años $Y = 20.000 \$$.

- ¿Cuál es la media y la desviación típica de los daños en un año cualquiera?

- b)** Considérese un «consorcio de seguros» de 100 personas cuyas viviendas se encuentran suficientemente dispersas por lo que, un año cualquiera, los daños en las distintas viviendas, pueden considerarse variables aleatorias independientemente distribuidas. Sea \bar{Y} el promedio de daños de las 100 viviendas en un año. (i) ¿Cuál es el valor esperado del promedio de daños \bar{Y} ? (ii) ¿Cuál es la probabilidad de que \bar{Y} sobrepase los 2.000 \$?
- 2.19** Considérense dos variables aleatorias X e Y . Supóngase que Y toma k valores y_1, \dots, y_k y que X toma l valores x_1, \dots, x_l .
- Demuestre que $\Pr(Y = y_j) = \sum_{i=1}^l \Pr(Y = y_j | X = x_i) \Pr(X = x_i)$. [(Pista: utilizar la definición de $\Pr(Y = y_j | X = x_i)$].
 - Utilice su respuesta al apartado (a) para verificar la Ecuación (2.19).
 - Supóngase que X e Y son independientes. Demuestre que $\sigma_{XY} = 0$ y $\text{corr}(X, Y) = 0$.
- 2.20** Considérense tres variables aleatorias X , Y , y Z . Supóngase que Y toma k valores y_1, \dots, y_k , que X toma l valores x_1, \dots, x_l y que Z toma m valores z_1, \dots, z_m . La probabilidad conjunta de X , Y , Z es $\Pr(X=x, Y=y, Z=z) = \frac{\Pr(Y=y, X=x, Z=z)}{\Pr(X=x, Z=z)}$.
- Explique cómo calcular la probabilidad marginal de que $Y = y$ a partir de la distribución de probabilidad conjunta. [Pista: esto es una generalización de la Ecuación (2.16)].
 - Demuestre que $E(Y) = E[E(Y|X, Z)]$. [Pista: esto es una generalización de las Ecuaciones (2.19) y (2.20)].
- 2.21** X es una variable aleatoria con momentos $E(X)$, $E(X^2)$, $E(X^3)$, etc.
- Demuestre $E(X - \mu)^3 = E(X^3) - 3[E(X^2)][E(X)] + 2[E(X)]^3$.
 - Demuestre $E(X - \mu)^4 = E(X^4) - 4[E(X)][E(X^3)] + 6[E(X)]^2[E(X^2)] - 3[E(X)]^4$.
- 2.22** Supóngase que se dispone de una cantidad de dinero para invertir —por simplicidad 1 \$— y se está planificando colocar una fracción w en un fondo de inversión colectiva en acciones y el resto, en un fondo de inversión colectiva en bonos. Supóngase que 1 \$ invertido en un fondo de inversión en acciones genera una rentabilidad R_s el primer año y que 1 \$ invertido en un fondo de bonos genera una rentabilidad R_b , supóngase que R_s es aleatoria con media 0,08 (8 %) y desviación típica de 0,07, y supóngase que R_b es aleatoria con media 0,05 (5 %) y desviación típica 0,04. La correlación entre R_s y R_b es 0,25. Si se coloca una fracción w del dinero en el fondo de acciones y el resto $1 - w$, en el fondo de bonos, entonces la rentabilidad de la inversión es $R = wR_s + (1 - w)R_b$.
- Supóngase que $w = 0,5$. Calcule la media y la desviación típica de R .
 - Supóngase que $w = 0,75$. Calcule la media y la desviación típica de R .
 - ¿Qué valor de w hace la media de R lo más grande posible? ¿Cuál es la desviación típica de R para este valor de w ?
 - (Más difícil). ¿Cuál es el valor de w que minimiza la desviación típica de R ? (Demuéstrelo gráficamente, con álgebra o mediante cálculo).
- 2.23** Este ejercicio muestra un ejemplo de un par de variables aleatorias X e Y para las que la media condicional de Y dado X depende de X pero $\text{corr}(X, Y) = 0$. Sean X y Z dos variables aleatorias normales estándar independientemente distribuidas, y sea $Y = X^2 + Z$.
- Demuestre que $E(Y|X) = X^2$.
 - Demuestre que $\mu_Y = 1$.
 - Demuestre que $E(XY) = 0$. (Pista: utilizar el hecho de que los momentos impares de una variable aleatoria normal estándar son iguales a cero).
 - Demuestre que $\text{cov}(X, Y) = 0$ y por tanto $\text{corr}(X, Y) = 0$.
- 2.24** Supóngase que Y_i se distribuye $N(0, \sigma^2)$ i.i.d. para $i = 1, \dots, n$.
- Demuestre que $E(Y_i^2/\sigma^2) = 1$.
 - Demuestre que $W = (1/\sigma^2) \sum_{i=1}^n Y_i^2$ se distribuye χ_n^2 .

- c) Demuestre que $E(W) = n$. [Pista: utilice su respuesta al apartado (a)].
- d) Demuestre que $V = Y_1 / \sqrt{\frac{\sum_{i=2}^n Y_i^2}{n-1}}$ se distribuye t_{n-1} .
- 2.25** (Revisión de la notación de sumatorios). Sea x_1, \dots, x_n una secuencia de números, sea y_1, \dots, y_n otra secuencia de números, y sean a, b , y c 3 constantes. Demuestre que
- $\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i$
 - $\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$
 - $\sum_{i=1}^n a = na$
 - $\sum_{i=1}^n (a + bx_i + cy_i)^2 = na^2 + b^2 \sum_{i=1}^n x_i^2 + c^2 \sum_{i=1}^n y_i^2 + 2ab \sum_{i=1}^n x_i + 2ac \sum_{i=1}^n y_i + 2bc \sum_{i=1}^n x_i y_i$
- 2.26** Supóngase que Y_1, \dots, Y_n son variables aleatorias con una media común μ_Y , una varianza común σ_Y^2 , y la misma correlación ρ (por lo que la correlación entre Y_i e Y_j es igual a ρ para todos los pares i y j , donde $i \neq j$).
- Demuestre que la $\text{cov}(Y_i, Y_j) = \rho\sigma_Y^2$ para $i \neq j$.
 - Supóngase que $n = 2$. Demuestre que $E(\bar{Y}) = \mu_Y$ y $\text{var}(\bar{Y}) = \frac{1}{2}\sigma_Y^2 + \frac{1}{2}\rho\sigma_Y^2$.
 - Para $n \geq 2$, demuestre que $E(\bar{Y}) = \mu_Y$ y $\text{var}(\bar{Y}) = \sigma_Y^2/n + [(n-1)/n]\rho\sigma_Y^2$.
 - Para n muy grande, demuestre que $\text{var}(\bar{Y}) \approx \rho\sigma_Y^2$.
- 2.27** X y Z son dos variables aleatorias distribuidas de forma conjunta. Supóngase que se conoce el valor de Z , pero no el valor de X . Sea $\tilde{X} = E(X|Z)$ que expresa una expectativa sobre el valor de X a partir de la información sobre Z , y sea $W = X - \tilde{X}$ la expresión que recoge el error asociado a esa expectativa.
- Demuestre que $E(W) = 0$. (Pista: utilizar la ley de esperanzas iteradas).
 - Demuestre que $E(WZ) = 0$.
 - Sea $\hat{X} = g(Z)$ otra expectativa de X a partir de Z , y $V = X - \hat{X}$ su error. Demuestre que $E(V^2) \geq E(W^2)$. [Pista: sea $h(Z) = g(Z) - E(X|Z)$, por lo que $V = [X - E(X|Z)] - h(Z)$. Obtenga $E(V^2)$].

APÉNDICE

2.1 Obtención de los resultados del Concepto clave 2.3

En este apéndice se obtienen las ecuaciones del Concepto clave 2.3.

La Ecuación (2.29) proviene de la definición de la esperanza.

Para obtener la Ecuación (2.30), se utiliza la definición de la varianza para escribir

$$\text{var}(a + bY) = E\{[a + bY - E(a + bY)]^2\} = E\{[b(Y - \mu_Y)]^2\} = b^2E[(Y - \mu_Y)^2] = b^2\sigma_Y^2.$$

Para obtener la Ecuación (2.31), se utiliza la definición de la varianza para escribir

$$\begin{aligned}\text{var}(aX + bY) &= E\{[(aX + bY) - (a\mu_X + b\mu_Y)]^2\} \\ &= E\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\} \\ &= E[a^2(X - \mu_X)^2] + 2E[ab(X - \mu_X)(Y - \mu_Y)] \\ &\quad + E[b^2(Y - \mu_Y)^2] \\ &= a^2 \text{var}(X) + 2ab \text{cov}(X, Y) + b^2 \text{var}(Y) \\ &= a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2,\end{aligned}\tag{2.49}$$

donde la segunda igualdad se obtiene reagrupando términos, la tercera igualdad se obtiene deshaciendo el cuadrado, y la cuarta igualdad se deduce de la definición de la varianza y la covarianza.

Para obtener la Ecuación (2.32), se escribe

$$E(Y^2) = E\{[(Y - \mu_Y) + \mu_Y]^2\} = E[(Y - \mu_Y)^2] + 2\mu_Y E(Y - \mu_Y) + \mu_Y^2 = \sigma_Y^2 + \mu_Y^2 \text{ por ser } E(Y - \mu_Y) = 0.$$

Para obtener la Ecuación (2.33), se utiliza la definición de la covarianza para escribir

$$\begin{aligned}\text{cov}(a + bX + cV, Y) &= E\{[a + bX + cV - E(a + bX + cV)][Y - \mu_Y]\} \\ &= E\{[b(X - \mu_X) + c(V - \mu_V)][Y - \mu_Y]\} \\ &= E\{[b(X - \mu_X)][Y - \mu_Y]\} + E\{[c(V - \mu_V)][Y - \mu_Y]\} \\ &= b\sigma_{XY} + c\sigma_{VY},\end{aligned}\tag{2.50}$$

que es la Ecuación (2.33).

Para obtener la Ecuación (2.34), se expresa

$$E(XY) = E\{[(X - \mu_X) + \mu_X][(Y - \mu_Y) + \mu_Y]\} = E[(X - \mu_X)(Y - \mu_Y)] + \mu_X E(Y - \mu_Y) + \mu_Y E(X - \mu_X) + \mu_X \mu_Y = \sigma_{XY} + \mu_X \mu_Y.$$

Ahora demostramos la desigualdad de la correlación de la Ecuación (2.35); es decir, $|\text{corr}(X, Y)| \leq 1$. Sea $a = -\sigma_{XY}/\sigma_X^2$ y $b = 1$. Aplicando la Ecuación (2.31), tenemos que

$$\begin{aligned}\text{var}(aX + Y) &= a^2\sigma_X^2 + \sigma_Y^2 + 2a\sigma_{XY} \\ &= (-\sigma_{XY}/\sigma_X^2)^2\sigma_X^2 + \sigma_Y^2 + 2(-\sigma_{XY}/\sigma_X^2)\sigma_{XY} \\ &= \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2.\end{aligned}\tag{2.51}$$

Como $\text{var}(aX + Y)$ es una varianza, no puede ser negativa, por tanto de la última línea de la Ecuación (2.51) debe ocurrir que $\sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2 \geq 0$. Reagrupando esta desigualdad queda

$$\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 \quad (\text{desigualdad de la covarianza}).\tag{2.52}$$

La desigualdad de la covarianza implica que $\sigma_{XY}^2/(\sigma_X^2\sigma_Y^2) \leq 1$ o, de forma equivalente, $|\sigma_{XY}|/(\sigma_X\sigma_Y) \leq 1$, lo que (utilizando la definición de correlación) demuestra la desigualdad de la correlación, $|\text{corr}(XY)| \leq 1$.

3

Repaso de estadística

La estadística es la ciencia que utiliza los datos con el fin de conocer el mundo que nos rodea. Las herramientas estadísticas nos ayudan a resolver preguntas sobre las características de distribuciones desconocidas de las poblaciones de interés. Por ejemplo, ¿cuál es la media de la distribución de los ingresos salariales de los titulados universitarios recientes? ¿Difieren los ingresos salariales medios de las mujeres y de los hombres, y si es así, cuánto?

Estas preguntas se refieren a la distribución de los ingresos salariales en la población de trabajadores. Una forma de resolver estas preguntas sería llevar a cabo un cuestionario exhaustivo sobre la población de trabajadores, midiendo los ingresos salariales de cada trabajador y así conocer la distribución poblacional de los ingresos salariales. En la práctica, sin embargo, un estudio tan exhaustivo resultaría extremadamente caro. La única encuesta completa de población de EE.UU. es el censo decenal. El censo 2000 de EE.UU. costó 10 mil millones de dólares, y el censo de 2010, 15 mil millones o más. El proceso de diseño de los formularios del censo, así como la gestión y realización de las encuestas y la recopilación y el análisis de los datos se realiza durante diez años. A pesar de este extraordinario esfuerzo, muchos miembros de la población son pasados por alto y no están recopilados. Por tanto, es necesario un método diferente, más práctico.

La idea clave de las estadísticas es que se pueden conocer aspectos sobre la distribución de la población mediante la selección de una muestra aleatoria. En lugar de estudiar toda la población de los EE.UU., podemos encuestar, por ejemplo, a 1.000 miembros de la población, seleccionados al azar mediante muestreo aleatorio simple. Mediante métodos estadísticos, se puede utilizar esta muestra para llegar a conclusiones provisionales —obtener inferencias estadísticas— acerca de las características de la población total.

Existen tres tipos de métodos estadísticos que se utilizan en la econometría: la estimación, los contrastes de hipótesis y los intervalos de confianza. La estimación implica el cálculo de «la mejor predicción», un valor numérico acerca de una característica desconocida de una distribución poblacional, como la media, a partir de una muestra de datos. El contraste de hipótesis consiste en la formulación de una hipótesis específica acerca de la población, y a continuación, utilizando la evidencia muestral, decidir si es cierta. Los intervalos de confianza utilizan un conjunto de datos con el fin de estimar un intervalo o rango para una característica desconocida de la población. Las Secciones 3.1, 3.2, y 3.3 repasan la estimación, los contrastes de hipótesis, y los intervalos de confianza en el contexto de la inferencia estadística acerca de una media poblacional desconocida.

En economía, la mayoría de las cuestiones se refieren a relaciones entre dos o más variables o comparaciones entre diferentes poblaciones. Por ejemplo, ¿existe una brecha entre los ingresos salariales medios de hombres y mujeres recién graduados en la universidad? En la Sección 3.4, los métodos para el estudio de la media de una única población de las Secciones 3.4 a 3.3 se generalizan a la comparación entre las medias de diferentes poblaciones. La Sección 3.5 trata sobre cómo pueden ser utilizados los métodos para la comparación de las medias entre dos poblaciones para la estimación de los efectos causales en experimentos. Las Secciones 3.2 a 3.5 se centran en el uso de la distribución normal para la realización de contrastes de hipótesis y para diseñar intervalos de confianza cuando el tamaño muestral es grande. En algunas circunstancias determinadas, los contrastes de hipótesis y los intervalos de confianza pueden estar basados en la distribución *t* de Student en vez de la distribución normal; estas circunstancias determinadas se tratan en la Sección 3.6. El capítulo concluye con un estudio de la correlación muestral y los diagramas de dispersión o nubes de puntos en la Sección 3.7.

3.1 Estimación de la media poblacional

Suponga que se desea conocer el valor medio de Y (es decir, μ_Y) de una población, como por ejemplo la media de los ingresos salariales de las mujeres recién graduadas en la universidad. Un método natural para estimar esta media es calcular la media muestral \bar{Y} a partir de una muestra de n observaciones independientes e idénticamente distribuidas (i.i.d.), Y_1, \dots, Y_n (recuérdese que Y_1, \dots, Y_n son i.i.d. si proceden de muestreo aleatorio simple). Esta sección trata la estimación de μ_Y y las propiedades de \bar{Y} como estimador de μ_Y .

Los estimadores y sus propiedades

Estimadores. La media muestral \bar{Y} es la manera natural de estimar μ_Y , pero no es el único método. Por ejemplo, otra forma de estimar μ_Y es simplemente utilizar la primera observación. Tanto \bar{Y} como Y_1 son funciones de los datos diseñadas para estimar μ_Y ; utilizando la terminología del Concepto clave 3.1, ambos son estimadores de μ_Y . Si se calculan para muestras repetidas, \bar{Y} e Y_1 toman valores diferentes (proporcionan estimaciones diferentes) de una muestra a otra. Así, los estimadores \bar{Y} e Y_1 poseen ambos distribuciones muestrales. Existen, de hecho, muchos estimadores de μ_Y , de los cuales \bar{Y} y Y_1 son dos ejemplos.

Existen muchos estimadores potenciales, así que ¿qué hace que un estimador sea «mejor» que otro? Al tratarse los estimadores de variables aleatorias, esta cuestión puede ser reformulada de una manera más precisa: ¿cuáles son las características deseables de la distribución muestral de un estimador? En general, nos gustaría un estimador que se hiciera tan próximo como fuera posible al verdadero valor desconocido, al menos en algún sentido promedio; en otras palabras, nos gustaría que la distribución muestral de un estimador estuviera tan estrechamente centrada sobre el valor desconocido como fuera posible. Esta observación nos conduce a las tres características específicas deseables de un estimador: insesgadez (ausencia de sesgo), consistencia y eficiencia.

Insesgadez. Supongamos que se evalúa un estimador varias veces sobre muestras repetidas aleatoriamente extraídas. Es sensato esperar que, en promedio, se obtendrá el resultado correcto. Por tanto una propiedad deseable para una estimador es que la media de su distribución muestral sea igual a μ_Y ; si esto ocurre, se dice que el estimador es insesgado.

Con el fin de definir matemáticamente este concepto, sea $\hat{\mu}_Y$ la expresión de un estimador de μ_Y como \bar{Y} o Y_1 . El estimador $\hat{\mu}_Y$ es insesgado si $E(\hat{\mu}_Y) = \mu_Y$, donde $E(\hat{\mu}_Y)$ es la media de la distribución muestral de $\hat{\mu}_Y$; en cualquier otro caso $\hat{\mu}_Y$ es sesgado.

Consistencia. Otra propiedad deseable de un estimador $\hat{\mu}_Y$ es que, cuando el tamaño muestral sea grande, la incertidumbre acerca del valor de μ_Y proveniente de las variaciones aleatorias de la muestra sea muy pequeña. De un modo más preciso, una propiedad deseable de $\hat{\mu}_Y$ es que la probabilidad de que se encuentre dentro de un intervalo pequeño del verdadero valor μ_Y se aproxime a 1 al aumentar la muestra, es decir, que $\hat{\mu}_Y$ sea un estimador consistente de μ_Y (Concepto clave 2.6).

Varianza y eficiencia. Supongamos que se dispone de dos aspirantes a estimadores, $\hat{\mu}_Y$ y $\tilde{\mu}_Y$. ¿Cómo se puede elegir entre ellos? Una manera de hacerlo es elegir el estimador con la distribución muestral más ajustada. Esto sugiere elegir entre $\hat{\mu}_Y$ y $\tilde{\mu}_Y$ seleccionando el estimador con la menor varianza. Si $\hat{\mu}_Y$ presenta una varianza menor que $\tilde{\mu}_Y$, entonces $\hat{\mu}_Y$ se dice que es más eficiente que $\tilde{\mu}_Y$. La terminología «eficiencia» procede de la noción de que si $\hat{\mu}_Y$ presenta una menor varianza, se utiliza la información de los datos de forma más eficiente que con $\tilde{\mu}_Y$.

CONCEPTO
CLAVE

3.1

Estimadores y estimaciones

Un **estimador** es una función de una muestra de datos extraída aleatoriamente a partir de una población. Una **estimación** es el valor numérico de un estimador cuando se calcula efectivamente utilizando los datos de una muestra específica. Un estimador es una variable aleatoria a causa de la aleatoriedad de la selección de la muestra, mientras que una estimación es un número no aleatorio.

**CONCEPTO
CLAVE**
3.2**Sesgo, consistencia y eficiencia**

Sea $\hat{\mu}_Y$ un estimador de μ_Y . Entonces:

- El *sesgo* de $\hat{\mu}_Y$ es $E(\hat{\mu}_Y) - \mu_Y$.
- $\hat{\mu}_Y$ es un *estimador insesgado* de μ_Y si $E(\hat{\mu}_Y) = \mu_Y$.
- $\hat{\mu}_Y$ es un *estimador consistente* de μ_Y si $\hat{\mu}_Y \xrightarrow{P} \mu_Y$.
- Sea $\hat{\mu}_Y$ otro estimador de μ_Y y supongamos que tanto $\hat{\mu}_Y$ como $\tilde{\mu}_Y$ son insesgados. Entonces $\hat{\mu}_Y$ se dice que es más *eficiente* que $\tilde{\mu}_Y$ si $\text{var}(\hat{\mu}_Y) < \text{var}(\tilde{\mu}_Y)$.

Sesgo, consistencia y eficiencia se recogen en el Concepto clave 3.2.

Propiedades de \bar{Y}

¿Qué ocurre con \bar{Y} como estimador de μ_Y cuando es evaluado de acuerdo a los tres criterios de sesgo, consistencia y eficiencia?

Sesgo y consistencia. La distribución muestral de \bar{Y} ha sido ya analizada en las Secciones 2.5 y 2.6. Como se muestra en la Sección 2.5, $E(\bar{Y}) = \mu_Y$, por lo que \bar{Y} es un estimador insesgado de μ_Y . De forma similar, la ley de los grandes números (Concepto clave 2.6) establece que $\bar{Y} \xrightarrow{P} \mu_Y$; es decir, \bar{Y} es consistente.

Eficiencia. ¿Qué puede decirse acerca de la eficiencia de \bar{Y} ? Como la eficiencia implica una comparación de estimadores, es necesario especificar el estimador o estimadores con los cuales se compara \bar{Y} .

Comenzamos comparando la eficiencia de \bar{Y} con el estimador Y_1 . Al ser Y_1, \dots, Y_n i.i.d., la media de la distribución muestral de Y_1 , es $E(Y_1) = \mu_Y$; por tanto Y_1 es un estimador insesgado de μ_Y . Su varianza es $\text{var}(Y_1) = \sigma_Y^2$. De acuerdo con la Sección 2.5, la varianza de \bar{Y} es σ_Y^2/n . Por tanto, para $n \geq 2$, la varianza de \bar{Y} es menor que la varianza de Y_1 ; es decir, \bar{Y} es un estimador más eficiente que Y_1 , por lo que, de acuerdo con el criterio de eficiencia, debería ser utilizado \bar{Y} en lugar de Y_1 . El estimador Y_1 podría ser considerado un estimador peor de forma evidente —¿por qué la molestia de recoger una muestra de n observaciones para quedarse solamente con la primera?— y el concepto de eficiencia proporciona un método formal de demostrar que \bar{Y} es un estimador más deseable que Y_1 .

¿Qué ocurre con un estimador que es peor de modo menos obvio? Consideremos la media ponderada en la que las observaciones se ponderan alternativamente con $\frac{1}{2}$ y $\frac{3}{2}$:

$$\tilde{Y} = \frac{1}{n} \left(\frac{1}{2} Y_1 + \frac{3}{2} Y_2 + \frac{1}{2} Y_3 + \frac{3}{2} Y_4 + \dots + \frac{1}{2} Y_{n-1} + \frac{3}{2} Y_n \right), \quad (3.1)$$

donde el número de observaciones n se supone que es par por conveniencia. La media de \tilde{Y} es μ_Y y su varianza es $\text{var}(\tilde{Y}) = 1,25\sigma_Y^2/n$ (Ejercicio 3.11). Por tanto, \tilde{Y} es insesgado y, al ser $\text{var}(\tilde{Y}) \rightarrow 0$ cuando $n \rightarrow \infty$, \tilde{Y} es consistente. Sin embargo, \tilde{Y} presenta una varianza mayor que \bar{Y} . Por tanto, \bar{Y} es más eficiente que \tilde{Y} .

Los estimadores \bar{Y} , Y_1 , e \tilde{Y} poseen una estructura matemática común: son medias ponderadas de Y_1, \dots, Y_n . Las comparaciones de los dos párrafos anteriores muestran que las medias ponderadas Y_1 e \tilde{Y} tienen

**CONCEPTO
CLAVE**
3.3**Eficiencia de \bar{Y} : \bar{Y} es ELIO**

Sea un estimador de μ_Y que es una media ponderada de Y_1, \dots, Y_n , es decir, $\hat{\mu}_Y = (1/n) \sum_{i=1}^n a_i Y_i$, donde a_1, \dots, a_n son constantes no aleatorias. Si $\hat{\mu}_Y$ es insesgado, entonces $\text{var}(\hat{\mu}_Y) < \text{var}(\bar{Y})$ a menos que $\hat{\mu}_Y = \bar{Y}$. Por tanto, \bar{Y} es el estimador lineal insesgado óptimo (ELIO o BLUE)¹; es decir, \bar{Y} es el estimador más eficiente de μ_Y de entre todos los estimadores insesgados que son medias ponderadas de Y_1, \dots, Y_n .

¹ N. del T.: Best Lineal Unbiased Estimator en inglés.

varianzas mayores que \bar{Y} . De hecho, estas conclusiones reflejan un resultado más general: \bar{Y} es el estimador más eficiente de *todos* los estimadores insesgados que son medias ponderadas de Y_1, \dots, Y_n . Dicho de forma distinta, \bar{Y} es el **Estimador Lineal Insesgado Óptimo (ELIO)**²; es decir, es el estimador más eficiente (óptimo) entre todos los estimadores insesgados que son funciones lineales de Y_1, \dots, Y_n . Este resultado se enumera en el Concepto clave 3.3 y se demuestra en el Capítulo 5.

\bar{Y} es el estimador de mínimos cuadrados de μ_Y . La media muestral \bar{Y} proporciona el mejor ajuste a los datos en el sentido de que el promedio de las diferencias cuadráticas entre las observaciones e \bar{Y} son las menores de entre todos los estimadores posibles.

Considérese el problema de encontrar el estimador m que minimiza

$$\sum_{i=1}^n (Y_i - m)^2, \quad (3.2)$$

que es una medida de la brecha o distancia cuadrática total entre el estimador m y los puntos muestrales. Debido a que m es un estimador de $E(Y)$, se puede pensar en él como una predicción del valor de Y_i , por lo que la distancia $Y_i - m$ puede ser interpretada como un error de predicción. La suma de los cuadrados de las distancias de la Expresión (3.2) puede interpretarse como la suma de los cuadrados de los errores de predicción.

El estimador m que minimiza la suma de las distancias al cuadrado $Y_i - m$ en la Expresión (3.2) se denomina **estimador de mínimos cuadrados**. Puede imaginarse la utilización del método de prueba y error para resolver el problema de mínimos cuadrados: intentar muchos valores para m hasta llegar a convencerse de haber obtenido el valor de m que hace la Expresión (3.2) tan pequeña como resulte posible. Alternativamente, como se explica en el Apéndice 3.2, puede utilizarse álgebra o cálculo para demostrar que $m = \bar{Y}$ minimiza la suma de las distancias al cuadrado de la Expresión (3.2) por lo que \bar{Y} es el estimador de mínimos cuadrados de μ_Y .

La importancia del muestreo aleatorio

Hemos supuesto que Y_1, \dots, Y_n son extracciones i.i.d. como las que hubiésemos obtenido por medio de muestreo aleatorio simple. Este supuesto es importante debido a que el muestreo no aleatorio puede originar que \bar{Y} sea sesgado. Supongamos que, para estimar la tasa de desempleo mensual nacional, una agencia estadística adopta un método de muestreo en el que los entrevistadores interrogan a los adultos en edad de trabajar que estén sentados en un parque de la ciudad a las 10:00 h. del segundo miércoles de mes. Debido a que a esa hora, la mayor parte de los trabajadores se encuentra en el trabajo (¡no sentados en el parque!), la tasa de desempleo estará sobrevalorada en la muestra, y una estimación de la tasa de desempleo basada sobre esa muestra resultaría sesgada. Este sesgo surge porque esta estructura muestral sobrerepresenta o sobrepondera, el número de los miembros de la población que se encuentran en desempleo. Este ejemplo es ficticio, pero el recuadro «¡Landon gana!» proporciona un ejemplo del mundo real acerca del sesgo introducido por el muestreo no totalmente aleatorio.

iLandon gana!

Poco antes de las elecciones presidenciales de 1936 en EE.UU., la *Literary Gazette* publicó un sondeo que indicaba que Alf M. Landon derrotaría al presidente en funciones Franklin D. Roosevelt, por un abrumador 57 % a 43 %. La *Gazette* acertó en que la elección fue abrumadora, pero se equivocó sobre el ganador: ¡Roosevelt ganó por 59 % a 41 %!

¿Cómo pudo la *Gazette* haber cometido un error tan grande? La muestra de la *Gazette* fue seleccionada a partir de los registros telefónicos y archivos del registro de automóviles.

Pero en 1936 muchos hogares no tenían ni automóvil ni teléfono, y aquellos que los tenían habitualmente eran más ricos —y con mayor probabilidad republicanos—. Al no haber sido la encuesta telefónica seleccionada aleatoriamente a partir de la población, si no que más bien subestimaba a los demócratas, el estimador estaba sesgado y la *Gazette* cometió un embarazoso error.

¿Es posible que las encuestas realizadas por medio de internet pudieran presentar un problema de sesgo similar?

² N. del T.: BLUE en inglés.

Es importante diseñar estructuras de diseño muestral de un modo que minimice el sesgo. El Apéndice 3.1 incluye un análisis de lo que la Oficina de Estadísticas Laborales realiza en realidad cuando lleva a cabo la Encuesta Actualizada de Población³ de EE.UU. (CPS en sus siglas en inglés⁴), la encuesta se utiliza para estimar la tasa de desempleo mensual de EE.UU.

3.2 Contrastes de hipótesis sobre la media poblacional

Muchas hipótesis acerca del mundo que nos rodea pueden expresarse en forma de preguntas con respuesta sí/no. ¿La media de los ingresos salariales por hora de los recién titulados en la universidad es de 20 \$ la hora? ¿Son los ingresos salariales medios los mismos para los titulados universitarios de sexo masculino y femenino? Ambas preguntas incluyen hipótesis específicas acerca de la distribución poblacional de los ingresos. El desafío estadístico consiste en responder estas preguntas en base a la evidencia muestral. Esta sección describe los **contrastes de hipótesis** que se refieren a la media muestral (¿Es la media poblacional de ingresos por hora igual a 20 \$?). Los contrastes de hipótesis que se refieren a dos poblaciones (¿Es la media de ingresos la misma para hombres y mujeres?) se recogen en la Sección 3.4.

Hipótesis nula y alternativa

El punto de partida para la contrastación estadística de hipótesis consiste en especificar la hipótesis a contrastar, denominada **hipótesis nula**. El contraste de hipótesis implica la utilización de datos para comparar la hipótesis nula con una segunda hipótesis, denominada **hipótesis alternativa**, que se cumple si la nula no lo hace.

La hipótesis nula consiste en que la media poblacional, $E(Y)$, toma un valor específico, denominado $\mu_{Y,0}$. La hipótesis nula se denomina H_0 y por tanto es

$$H_0: E(Y) = \mu_{Y,0}. \quad (3.3)$$

Por ejemplo, la conjectura de que, en promedio de la población, los titulados universitarios ingresan 20 \$ por hora es una hipótesis nula acerca de la distribución poblacional de los ingresos por hora. Expresado de forma matemática, si Y es el ingreso por hora de un recién graduado en la universidad seleccionado aleatoriamente, entonces la hipótesis nula es que $E(Y) = 20$; es decir, $\mu_{Y,0} = 20$ en la Ecuación (3.3).

La hipótesis alternativa especifica qué es cierto si la hipótesis nula no lo es. La hipótesis alternativa más general es que $E(Y) \neq \mu_{Y,0}$, lo que se denomina **hipótesis alternativa bilateral** porque permite que $E(Y)$ pueda ser tanto menor como mayor que $\mu_{Y,0}$. La alternativa bilateral se escribe como

$$H_1: E(Y) \neq \mu_{Y,0} \quad (\text{alternativa bilateral}). \quad (3.4)$$

La alternativa unilateral es asimismo posible, y se tratará más tarde en esta sección.

El problema al que se enfrenta el estadístico es el de utilizar la evidencia que proporciona una muestra de datos seleccionada aleatoriamente con el fin de decidir si aceptar la hipótesis nula H_0 o rechazarla en favor de la hipótesis alternativa H_1 . Si la hipótesis nula es «aceptada», esto no significa que el estadístico declare que es cierta; sino que es aceptada provisionalmente reconociendo que puede ser rechazada más tarde en base a la evidencia adicional. Por esta razón, el contraste estadístico de hipótesis puede plantearse en términos tanto de rechazar la hipótesis nula como de no hacerlo.

El *p*-valor

En un ejemplo concreto, la media muestral \bar{Y} raramente será exactamente igual al valor que toma la hipótesis especificada $\mu_{Y,0}$. Las diferencias entre \bar{Y} y $\mu_{Y,0}$ pueden aparecer debido a que en realidad la verdadera media no es igual a $\mu_{Y,0}$ (la hipótesis nula es falsa) o debido a que el verdadero valor de la media es

³ U.S. Current Population Survey.

⁴ N. del T.: CPS en inglés.

igual a $\mu_{Y,0}$ (la hipótesis nula es verdadera), pero \bar{Y} es distinto de $\mu_{Y,0}$ debido al muestreo aleatorio. Es imposible distinguir entre dos posibilidades con certeza. Aunque una muestra de datos no pueda proporcionar evidencia concluyente acerca de la hipótesis nula, es posible realizar un cálculo probabilístico que permita contrastar la hipótesis nula de modo que se tenga en cuenta la incertidumbre del muestreo. Este cálculo implica la utilización de datos para obtener el *p*-valor de la hipótesis nula.

El ***p*-valor**, denominado asimismo **probabilidad de significación**, es la probabilidad de obtener un valor del estadístico al menos tan adverso para la hipótesis nula como el calculado en la muestra, suponiendo que la hipótesis nula es cierta. En el caso que nos ocupa, el *p*-valor es la probabilidad de obtener una \bar{Y} al menos tan alejada en lo que respecta a las colas de su distribución, bajo la hipótesis nula, como la media muestral calculada realmente.

Por ejemplo, supongamos que, en la muestra de los recién titulados en la universidad, el salario medio es de 22,64 \$. El *p*-valor es la probabilidad de haber observado un valor de \bar{Y} al menos tan diferente de 20 \$ (la media poblacional bajo la nula) como el valor observado de 22,64 \$ debido solamente a la variación muestral, suponiendo que la hipótesis nula es cierta. Si el *p*-valor es pequeño, digamos el 5 %, entonces resulta muy improbable que, siendo cierta la hipótesis nula, se haya extraído esa muestra; por lo que resulta razonable concluir que la hipótesis nula no es cierta. Por el contrario, si el *p*-valor es grande, digamos un 40 %, entonces es muy probable que la media muestral observada de 22,64 \$ pueda haber sido observada exclusivamente debido a la variación del muestreo aleatorio si la hipótesis nula es cierta; por tanto, la evidencia en contra de la hipótesis nula es débil en sentido probabilístico, y resulta razonable no rechazar la hipótesis nula.

Matemáticamente, para establecer la definición del *p*-valor, sea \bar{Y}^{act} la expresión del valor de la media muestral calculada realmente con los datos en cuestión, y sea \Pr_{H_0} la expresión de la probabilidad calculada bajo la hipótesis nula (es decir, calculada suponiendo que $E(Y_i) = \mu_{Y,0}$). El *p*-valor es

$$p\text{-valor} = \Pr_{H_0}[|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}|]. \quad (3.5)$$

Es decir, el *p*-valor es el área de las colas de la distribución de \bar{Y} bajo la hipótesis nula más allá de $|\bar{Y}^{act} - \mu_{Y,0}|$. Si el *p*-valor es elevado, entonces el valor observado de \bar{Y}^{act} es coherente con la hipótesis nula, pero no lo es si el *p*-valor es pequeño.

Para calcular el *p*-valor, es necesario conocer la distribución muestral de \bar{Y} bajo la hipótesis nula. Como se expuso en la Sección 2.6, cuando el tamaño muestral es pequeño esta distribución es complicada. Sin embargo, de acuerdo con el teorema central del límite, cuando el tamaño muestral es grande la distribución muestral de \bar{Y} se aproxima correctamente mediante una distribución normal. Bajo la hipótesis nula la media de esta distribución normal es $\mu_{Y,0}$, por lo que bajo la hipótesis nula \bar{Y} se distribuye $N(\mu_{Y,0}, \sigma_{\bar{Y}}^2)$, donde $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$. Esta aproximación normal para grandes muestras hace posible calcular el *p*-valor sin necesidad de conocer la distribución poblacional de Y , en tanto en cuanto el tamaño muestral sea elevado. Los detalles del cálculo, sin embargo, dependen de si σ_Y^2 es conocido.

Cálculo del *p*-valor con σ_Y^2 conocido

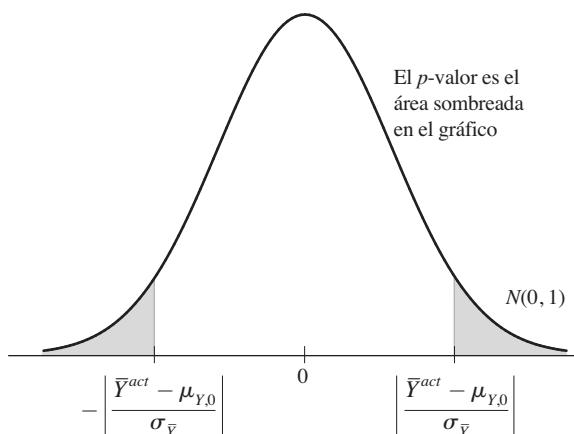
El cálculo de *p*-valor cuando σ_Y^2 es conocido se recoge en la Figura 3.1. Si el tamaño muestral es grande, entonces bajo la hipótesis nula la distribución muestral de \bar{Y} es $N(\mu_{Y,0}, \sigma_{\bar{Y}}^2)$, donde $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$. Por tanto, bajo la hipótesis nula, la versión estandarizada de \bar{Y} , $(\bar{Y} - \mu_{Y,0})/\sigma_{\bar{Y}}$, presenta una distribución normal estándar. El *p*-valor es la probabilidad de obtener un valor de \bar{Y} más alejado de $\mu_{Y,0}$ que \bar{Y}^{act} bajo la hipótesis nula o, de manera equivalente, es la probabilidad de obtener un $(\bar{Y} - \mu_{Y,0})/\sigma_{\bar{Y}}$ mayor que $(\bar{Y}^{act} - \mu_{Y,0})/\sigma_{\bar{Y}}$ en valor absoluto. Esta probabilidad es el área sombreada que se muestra en la Figura 3.1. Expresada matemáticamente, la probabilidad sombreada de las colas en la Figura 3.1 (es decir, el *p*-valor) es

$$p\text{-valor} = \Pr_{H_0}\left(\left|\frac{\bar{Y} - \mu_{Y,0}}{\sigma_{\bar{Y}}}\right| > \left|\frac{\bar{Y}^{act} - \mu_{Y,0}}{\sigma_{\bar{Y}}}\right|\right) = 2\Phi\left(-\left|\frac{\bar{Y}^{act} - \mu_{Y,0}}{\sigma_{\bar{Y}}}\right|\right), \quad (3.6)$$

donde Φ es la función de distribución normal estándar acumulada. Es decir, el *p*-valor es el área de las colas de una distribución normal estándar más allá de $\pm(\bar{Y}^{act} - \mu_{Y,0})/\sigma_{\bar{Y}}$.

FIGURA 3.1 Cálculo del *p*-valor

El *p*-valor es la probabilidad de obtener un valor de \bar{Y} al menos tan distinto de $\mu_{Y,0}$ como \bar{Y}^{act} . En muestras grandes, \bar{Y} se distribuye $N(\mu_{Y,0}, \sigma_{\bar{Y}}^2)$, bajo la hipótesis nula, por lo que $(\bar{Y} - \mu_{Y,0})/\sigma_{\bar{Y}}$ se distribuye $N(0,1)$. Así, el *p*-valor es la probabilidad de la cola sombreada de la distribución normal estándar más allá de $\pm|\bar{Y} - \mu_{Y,0}|/\sigma_{\bar{Y}}$.



La fórmula para el *p*-valor de la Ecuación (3.6) depende de la varianza de la distribución poblacional, σ_Y^2 . En la práctica, habitualmente esta varianza es desconocida. [Una excepción es cuando Y_i es binaria, por lo que su distribución es de Bernouilli, en cuyo caso la varianza está determinada por la hipótesis nula; véase la Ecuación (2.7) y el Ejercicio 3.2]. Debido a que en general σ_Y^2 debe ser estimada antes de poder calcular el *p*-valor, volvemos ahora al problema de la estimación de σ_Y^2 .

La varianza muestral, la desviación típica muestral y el error estándar

La varianza muestral s_Y^2 es un estimador de la varianza poblacional σ_Y^2 , la desviación típica muestral s_Y es un estimador de la desviación típica poblacional σ_Y , y el error estándar de la media muestral \bar{Y} es un estimador de la desviación típica de la distribución muestral de \bar{Y} .

La varianza muestral y la desviación típica. La **varianza muestral**, s_Y^2 , es

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2. \quad (3.7)$$

La **desviación típica muestral**, s_Y , es la raíz cuadrada de la varianza muestral.

La fórmula de la varianza muestral es bastante similar a la fórmula de la varianza poblacional. La varianza poblacional, $E(Y - \mu_Y)^2$, es el valor promedio de $(Y - \mu_Y)^2$, en la distribución poblacional. De forma similar, la varianza muestral es el promedio muestral de $(Y_i - \mu_Y)$, $i = 1, \dots, n$, con dos modificaciones: primera, μ_Y se reemplaza por \bar{Y} , y segunda, el promedio utiliza el divisor $n - 1$ en vez de n .

La razón para la primera modificación —la sustitución de μ_Y por \bar{Y} — es que μ_Y es desconocido y debe ser estimado; el estimador natural de μ_Y es \bar{Y} . La razón para la segunda modificación —dividir por $n - 1$ en vez de n — es que la estimación de μ_Y mediante \bar{Y} introduce un pequeño sesgo a la baja en $(Y_i - \bar{Y})^2$. Concretamente, como se muestra en el Ejercicio 3.18, $E[(Y_i - \bar{Y})^2] = [(n-1)/n]\sigma_Y^2$. Por tanto, $E \sum_{i=1}^n (Y_i - \bar{Y})^2 = nE[(Y_i - \bar{Y})^2] = (n-1)\sigma_Y^2$. Dividiendo por $n - 1$ en la Ecuación (3.7) en vez de por n se corrige el pequeño sesgo a la baja, y como resultado s_Y^2 es insesgado.

CONCEPTO CLAVE

3.4

El error estándar de \bar{Y}

El error estándar de \bar{Y} es un estimador de la desviación típica de \bar{Y} . El error estándar de \bar{Y} se expresa como $ES(\bar{Y})$ o bien $\hat{\sigma}_{\bar{Y}}$. Si Y_1, \dots, Y_n son i.i.d.

$$ES(\bar{Y}) = \hat{\sigma}_{\bar{Y}} = s_Y / \sqrt{n}. \quad (3.8)$$

Dividir por $n - 1$ la Ecuación (3.7) en lugar de n se denomina corrección de los **grados de libertad**: la estimación de la media consume parte de la información —es decir, consume 1 grado de libertad de los datos—, por lo que solo quedan $n - 1$ grados de libertad.

Consistencia de la varianza muestral. La varianza muestral es un estimador consistente de la varianza poblacional:

$$s_Y^2 \rightarrow \sigma_Y^2. \quad (3.9)$$

En otras palabras, la varianza muestral está cercana a la varianza poblacional con elevada probabilidad cuando n es grande.

El resultado de la Ecuación (3.9) se demuestra en el Apéndice 3.3 bajo el supuesto de que Y_1, \dots, Y_n son i.i.d. e Y_i presenta un momento de cuarto orden finito: es decir, $E(Y_i^4) < \infty$. Intuitivamente, la razón por la que s_Y^2 es consistente es que se trata de una media muestral, por lo que s_Y^2 obedece la ley de los grandes números. Pero para que s_Y^2 obedezca la ley de los grandes números del Concepto clave 2.6, $(Y_i - \mu_Y)^2$ debe tener varianza finita, lo que de hecho significa que $E(Y_i^4)$ debe ser finita; en otras palabras, Y_i debe tener un momento de cuarto orden finito.

El error estándar de \bar{Y} . Debido a que la desviación típica de la distribución muestral de \bar{Y} es $\sigma_{\bar{Y}} = \sigma_Y / \sqrt{n}$, la Ecuación (3.9) justifica la utilización de s_Y / \sqrt{n} , como estimador de $\sigma_{\bar{Y}}$. El estimador de $\sigma_{\bar{Y}}$, s_Y / \sqrt{n} , se denomina **error estándar de \bar{Y}** , y se escribe $ES(\bar{Y})$ o bien $\hat{\sigma}_{\bar{Y}}$ (el carácter « $\hat{\cdot}$ ») sobre el símbolo significa que es un estimador de $\sigma_{\bar{Y}}$). El error estándar de \bar{Y} se recoge como Concepto clave 3.4.

Si Y_1, \dots, Y_n son extracciones i.i.d. de una distribución de Bernouilli con probabilidad de éxito p , la fórmula para la varianza de \bar{Y} se simplifica a $p(1 - p)/n$ (véase el Ejercicio 3.2). La fórmula del error estándar asimismo toma una forma sencilla que solo depende de \bar{Y} y n : $ES(\bar{Y}) = \sqrt{\bar{Y}(1 - \bar{Y})/n}$.

Cálculo del *p*-valor con σ_Y desconocido

Al ser s_Y^2 un estimador consistente de σ_Y^2 el *p*-valor puede calcularse reemplazando $\sigma_{\bar{Y}}$ en la Ecuación (3.6) por el error estándar, $ES(\bar{Y}) = \hat{\sigma}_{\bar{Y}}$. Es decir, cuando σ_Y es desconocido e Y_1, \dots, Y_n son i.i.d., el *p*-valor se calcula mediante la fórmula

$$p\text{-valor} = 2\Phi\left(-\left|\frac{\bar{Y}^{act} - \mu_{Y,0}}{SE(\bar{Y})}\right|\right). \quad (3.10)$$

El estadístico *t*

La media muestral estandarizada $(\bar{Y} - \mu_{Y,0})/ES(\bar{Y})$ desempeña un papel central en el contraste de hipótesis y tiene un nombre concreto, **estadístico *t* o *t*-ratio**:

$$t = \frac{\bar{Y} - \mu_{Y,0}}{SE(\bar{Y})}. \quad (3.11)$$

En general, un **estadístico** de contraste es un estadístico utilizado para llevar a cabo un contraste de hipótesis. El estadístico *t* es un ejemplo importante de un contraste de hipótesis.

Distribución para grandes muestras del estadístico *t*. Cuando n es grande, s_Y^2 es cercano a σ_Y^2 con alta probabilidad. Por tanto la distribución del estadístico *t* es aproximadamente la misma que la distribución $(\bar{Y} - \mu_{Y,0})/\sigma_{\bar{Y}}$, que a su vez se aproxima bien mediante la distribución normal estándar cuando n es grande por el teorema central del límite (Concepto clave 2.7). Por tanto, bajo la hipótesis nula,

$$t \text{ se distribuye aproximadamente como } N(0, 1) \text{ para } n \text{ grande.} \quad (3.12)$$

La fórmula para el *p*-valor de la Ecuación (3.10) puede reescribirse en términos del estadístico *t*. Sea t^{act} el valor del estadístico realmente calculado:

$$t^{act} = \frac{\bar{Y}^{act} - \mu_{Y,0}}{ES(\bar{Y})}. \quad (3.13)$$

Por tanto, cuando n es grande, el *p*-valor puede calcularse mediante

$$p\text{-valor} = 2\Phi(-|t^{act}|). \quad (3.14)$$

A modo de ejemplo hipotético, supóngase que se utiliza una muestra de $n = 200$ recién titulados universitarios para contrastar la hipótesis nula de que el salario medio, $E(Y)$, es de 20 \$ la hora. El salario medio muestral es $\bar{Y}^{act} = 22,64$ \$, y la desviación típica muestral es $s_Y = 18,14$ \$. Por tanto el error estándar de \bar{Y} es $s_Y/\sqrt{n} = 18,14/\sqrt{200} = 1,28$. El valor del estadístico *t* es $t^{act} = (22,64 - 20)/1,28 = 2,06$. De la Tabla 1 del Apéndice, el *p*-valor es $2\Phi(-2,06) = 0,039$, o 3,9 %. Es decir, suponiendo que la hipótesis nula es cierta, la probabilidad de obtener una media muestral al menos tan diferente de la nula como la calculada realmente es el 3,9 %.

Contrastes de hipótesis con nivel de significación preestablecido

Cuando se lleva a cabo un contraste de hipótesis estadístico, se pueden cometer dos tipos de errores. Se puede rechazar de forma incorrecta la hipótesis nula cuando es verdadera, o se puede errar al aceptar la hipótesis nula cuando es falsa. Los contrastes de hipótesis pueden llevarse a cabo sin calcular el *p*-valor si se está dispuesto a especificar previamente la probabilidad que se está dispuesto a tolerar de cometer el primer tipo de error —es decir, de rechazar incorrectamente la hipótesis nula cuando es verdadera. Si se elige una probabilidad predeterminada de rechazar la hipótesis nula cuando es verdadera (por ejemplo, el 5 %), entonces se rechazará la hipótesis nula si y solo si el *p*-valor es menor que 0,05. Este método proporciona un tratamiento preferente a la hipótesis nula, pero en muchas situaciones prácticas este tratamiento preferencial resulta apropiado.

Contraste de hipótesis utilizando un nivel de significación fijo. Supóngase que se ha decidido que la hipótesis será rechazada si el *p*-valor es menor que el 5 %. Debido a que el área bajo las colas de la distribución normal estándar más allá de $\pm 1,96$ % es el 5 %, esto proporciona una regla simple:

$$\text{Rechazar } H_0 \text{ si } |t^{act}| > 1,96. \quad (3.15)$$

Es decir, rechazar si el valor absoluto del estadístico *t* calculado en la muestra es mayor que 1,96. Si n es lo suficientemente grande, bajo la hipótesis nula el estadístico *t* presenta una distribución $N(0, 1)$. Por tanto la probabilidad de rechazar erróneamente la hipótesis nula (rechazar la hipótesis nula cuando en realidad es verdadera) es del 5 %.

Este sistema para el contraste de hipótesis estadístico posee una terminología especializada, recogida en el Concepto clave 3.5. El nivel de significación del contraste en la Ecuación (3.15) es el 5 %, el valor crítico de este contraste bilateral es 1,96 y la región de rechazo son los valores del estadístico *t* que quedan fuera de $\pm 1,96$. Si el contraste rechaza al 5 % de nivel de significación, se dice que la media poblacional μ_Y es significativamente distinta de $\mu_{Y,0}$ desde el punto de vista estadístico al 5 % de nivel de significación.

El contraste de hipótesis mediante un nivel de significación preestablecido no necesita el cálculo de *p*-valores. En el ejemplo anterior de contraste de la hipótesis de que los ingresos salariales medios de los recién titulados universitarios es de 20 \$ la hora, el estadístico *t* era 2,06. Este valor supera 1,96, por lo que la hipótesis nula se rechaza al nivel del 5 %. Aunque llevar a cabo este contraste con un 5 % de nivel de significación es fácil, el hecho de ofrecer tan solo el resultado de si la hipótesis nula se rechaza a un nivel de significación preestablecido proporciona menos información que ofrecer el *p*-valor.

¿Qué nivel de significación debería utilizarse en la práctica? En muchos casos, los estadísticos y los económetras utilizan el 5% de nivel de significación. Si se fueran a contrastar muchas hipótesis estadís-

**CONCEPTO
CLAVE**
3.5**La terminología del contraste de hipótesis**

Un contraste de hipótesis estadístico puede ocurrir en dos tipos de errores: un **error tipo I**, en el cual la hipótesis nula se rechaza cuando en realidad es cierta, y un **error tipo II**, en el que la hipótesis nula no es rechazada cuando en realidad es falsa. La probabilidad de rechazo preestablecida de un contraste de hipótesis estadístico cuando la hipótesis nula es cierta —es decir, la probabilidad preestablecida de error tipo I— es el **nivel de significación** del contraste. El **valor crítico** del estadístico de contraste es el valor del estadístico para el cual el contraste exactamente rechaza la hipótesis nula a un nivel de significación dado. El conjunto de valores del estadístico de contraste para los que el contraste rechaza la hipótesis nula es la **región de rechazo**, y los valores del estadístico de contraste para los cuales no se rechaza la hipótesis nula es la **región de aceptación**. La probabilidad de que el contraste rechace efectivamente de forma incorrecta la hipótesis nula cuando es verdadera es el **tamaño del contraste**, y la probabilidad de que el contraste rechace correctamente la hipótesis nula cuando la alternativa es cierta es la **potencia del contraste**.

El *p*-valor es la probabilidad de obtener un estadístico de contraste, debido a la variabilidad del muestreo aleatorio, al menos tan adversa para el valor de la hipótesis nula como el valor realmente observado, suponiendo que la hipótesis nula es cierta. De manera equivalente, el *p*-valor es el menor nivel de significación al que se puede rechazar la hipótesis nula.

ticas al 5 % de nivel, debería rechazarse incorrectamente la nula, en promedio, una vez cada 20 veces. A veces podría resultar adecuado un nivel de significación más conservador. Por ejemplo, los casos legales a veces incluyen evidencia estadística, y la hipótesis nula podría ser que el defendido fuera no culpable; entonces se debería estar bastante seguro de que el rechazo de la nula (conclusión de culpabilidad) no sea solamente el resultado de la variabilidad del muestreo aleatorio. En algunos marcos legales, el nivel de significación utilizado es el 1 %, o incluso el 0,1 % para evitar este tipo de error. De forma similar, si una agencia del gobierno está considerando permitir la venta de un nuevo medicamento, podría resultar adecuado un estándar muy conservador con el fin de que los consumidores puedan estar seguros de que los medicamentos disponibles en el mercado funcionan en realidad.

Ser conservador, en el sentido de utilizar un nivel de significación muy bajo, tiene un coste: cuanto menor es el nivel de significación, mayor es el valor crítico y más difícil se convierte rechazar la nula cuando la nula es falsa. De hecho, lo más conservador es no rechazar nunca la hipótesis nula, pero si este es el punto de vista, entonces nunca se necesitará la evidencia estadística porque ¡nunca se cambiará de opinión! Cuanto menor sea el nivel de significación, menor será la potencia del contraste. Muchas aplicaciones económicas y políticas pueden reclamar menos conservadurismo que el caso legal, por lo que a menudo se considera un 5 % de nivel de significación como una convención razonable.

El Concepto clave 3.6 recoge los contrastes de hipótesis para la media poblacional frente a la alternativa bilateral.

**CONCEPTO
CLAVE**
3.6**El contraste de la hipótesis $E(Y) = \mu_{Y,0}$ frente a la alternativa $E(Y) \neq \mu_{Y,0}$**

1. Calcular el error estándar de \bar{Y} , $ES(\bar{Y})$ [Ecuación (3.8)].
2. Calcular el estadístico t [Ecuación (3.13)].
3. Calcular el *p*-valor [Ecuación (3.14)]. Rechazar la hipótesis al 5 % de nivel de significación si el *p*-valor es menor que 0,05 (de modo equivalente, si $|t^{act}| > 1,96$).

Alternativas unilaterales

En algunas circunstancias, la hipótesis alternativa podría ser que la media supere el valor $\mu_{Y,0}$. Por ejemplo, se espera que la educación ayude en el mercado de trabajo, por lo que la alternativa relevante a la hipótesis nula de que los ingresos son los mismos para los titulados universitarios y para los no titulados no es precisamente que sus ingresos difieran, sino más bien que los titulados ingresan más que los no titulados. Esto se denomina **hipótesis alternativa unilateral** y puede escribirse como

$$H_1: E(Y) > \mu_{Y,0} \quad (\text{alternativa unilateral}). \quad (3.16)$$

El método general para el cálculo de los p -valores y para contrastar hipótesis es el mismo para alternativas unilaterales y para alternativas bilaterales, con la variante de que solamente los grandes valores positivos del estadístico t rechazan la hipótesis nula en lugar de que los valores grandes en valor absoluto. Concretamente, para contrastar la hipótesis unilateral de la Ecuación (3.16), se construye el estadístico t de la Ecuación (3.13). El p -valor es el área bajo la distribución normal estándar a la derecha del estadístico t calculado. Es decir, el p -valor, basado en la aproximación $N(0, 1)$ para la distribución del estadístico t , es

$$p\text{-valor} = \Pr_{H_0}(Z > t^{act}) = 1 - \Phi(t^{act}). \quad (3.17)$$

El valor crítico $N(0, 1)$ para un contraste unilateral con un 5 % de nivel de significación es 1,64. La región de rechazo para este contraste son todos los valores del estadístico t superiores a 1,64.

La hipótesis unilateral de la Ecuación (3.16) se refiere a valores de μ_Y superiores a $\mu_{Y,0}$. Si, en cambio, la hipótesis alternativa es que $E(Y) < \mu_{Y,0}$, entonces es aplicable lo expresado en los párrafos anteriores excepto que cambia el signo; por ejemplo, la región de rechazo al 5 % consiste en los valores del estadístico t menores que -1,64.

3.3 Intervalos de confianza para la media poblacional

Debido al error de muestreo aleatorio, es imposible saber el valor exacto de la media poblacional de Y utilizando solo la información de una muestra. Sin embargo, es posible utilizar los datos de una muestra aleatoria para construir un conjunto de valores que contengan la verdadera media poblacional μ_Y con una cierta probabilidad preestablecida. Este conjunto de valores se denomina **conjunto de confianza**, y la probabilidad preestablecida de que μ_Y esté contenida en este conjunto se denomina **nivel de confianza**. El conjunto de confianza para μ_Y resulta ser todos los posibles valores de la media entre un límite inferior y otro superior, por lo que el conjunto de confianza es un intervalo, denominado **intervalo de confianza**.

He aquí un modo de construir un conjunto de confianza al 95 % para la media poblacional. Comienza por escoger algún valor arbitrario para la media; denominado $\mu_{Y,0}$. Se contrasta la hipótesis nula de que $\mu_Y = \mu_{Y,0}$, frente a la alternativa $\mu_Y \neq \mu_{Y,0}$, mediante el cálculo del estadístico t : si es menor que 1,96, entonces no se rechaza este valor hipotético $\mu_{Y,0}$ al nivel del 5 %, y se anota este valor no rechazado $\mu_{Y,0}$. Ahora se selecciona otro valor arbitrario para $\mu_{Y,0}$ y se contrasta; si no se puede rechazar, se anota este valor en la lista. Se hace una y otra vez; de hecho, se hace para todos los posibles valores de la media poblacional. Continuando con este proceso se obtiene el conjunto de todos los valores de la media poblacional que no pueden rechazarse al 5 % de nivel para un contraste de hipótesis bilateral.

Este listado es útil porque recoge el conjunto de hipótesis que pueden y no pueden rechazarse (al 5 % de nivel) en base a los datos: si alguien acudiese con un número específico en mente, podría decirse si la hipótesis se rechaza o no simplemente buscando ese número en la lista. Un poco de razonamiento inteligente muestra que este conjunto de valores posee una propiedad destacada: la probabilidad de que contenga el verdadero valor de la media poblacional es el 95 %.

El razonamiento inteligente funciona así. Supóngase que el verdadero valor de μ_Y es 21,5 (aunque no se sabe). Entonces \bar{Y} presenta una distribución normal centrada en 21,5 y el estadístico t para el contraste de la hipótesis nula μ_Y presenta una distribución $N(0, 1)$. Por tanto, si n es grande, la probabilidad de rechazar la hipótesis nula $\mu_Y = 21,5$ al 5 % de nivel es el 5 %. Pero como se contrastaron todos los posibles valores

**CONCEPTO
CLAVE**
3.7
Intervalos de confianza para la media poblacional

Un intervalo de confianza bilateral al 95 % para μ_Y es un intervalo construido para que contenga el verdadero valor de μ_Y en el 95 % de todas las posibles muestras aleatorias. Cuando el tamaño muestral n es grande, los intervalos de confianza al 95 %, 90 %, y 99 % para μ_Y son:

- Intervalo de confianza al 95 % para $\mu_Y = [\bar{Y} \pm 1,96ES(\bar{Y})]$.
- Intervalo de confianza al 95 % para $\mu_Y = [\bar{Y} \pm 1,64ES(\bar{Y})]$.
- Intervalo de confianza al 95 % para $\mu_Y = [\bar{Y} \pm 2,58ES(\bar{Y})]$.

de la media poblacional para construir el conjunto, en particular se contrastó el verdadero valor $\mu_Y = 21,5$. En el 95 % de todas las muestras, se aceptará correctamente 21,5; esto significa que en el 95 % de todas las muestras, la lista contendrá el verdadero valor de μ_Y . Por tanto, los valores de la lista constituyen un conjunto de confianza al 95 % para μ_Y .

Este método para construir un conjunto de confianza es poco práctico, para ello se requiere contrastar todos los posibles valores de μ_Y como hipótesis nulas. Afortunadamente, existe un método mucho más fácil. Siguiendo la fórmula para el estadístico t de la Ecuación (3.13), un valor de prueba para $\mu_{Y,0}$ se rechaza al 5 % de nivel si está más alejado que 1,96 errores estándar de \bar{Y} . Por tanto el conjunto de valores de μ_Y que no se rechazan al 5 % de nivel consta de aquellos valores incluidos en $1,96ES(\bar{Y})$ de \bar{Y} . Es decir, un intervalo de confianza al 95 % para μ_Y es $\bar{Y} - 1,96ES(\bar{Y}) \leq \mu_Y \leq \bar{Y} + 1,96ES(\bar{Y})$. El Concepto clave 3.7 recoge este método.

A modo de ejemplo, considérese el problema de construir un intervalo de confianza al 95 % para el promedio de ingresos salariales por hora de los recién titulados universitarios mediante una muestra aleatoria hipotética de 200 titulados universitarios recientes donde $\bar{Y} = 22,64 \$$ y $ES(\bar{Y}) = 1,28 \$$. El intervalo de confianza al 95 % para el promedio de ingresos salariales por hora es $22,64 \pm 1,96 \times 1,28 = 22,64 \pm 2,51 = [20,13 \$, 25,15 \$]$.

Hasta este punto, el análisis se ha centrado en los intervalos de confianza bilaterales. Se podría construir además un intervalo de confianza unilateral como el conjunto de valores de μ_Y que no pueden ser rechazados mediante un contraste de hipótesis unilateral. Aunque los intervalos de confianza unilaterales se aplican en algunas ramas de la estadística, son poco comunes en el análisis econométrico aplicado.

Probabilidades de cobertura. La **probabilidad de cobertura** de un intervalo de confianza para la media poblacional es la probabilidad, calculada sobre todas las posibles muestras aleatorias, de que contenga el verdadero valor de la media poblacional.

3.4 Comparación de medias de diferentes poblaciones

¿Tienen en promedio los mismos ingresos salariales los hombres y mujeres recién titulados en la universidad? Esta pregunta implica tener que comparar las medias de dos distribuciones poblacionales diferentes. Esta sección recoge cómo contrastar hipótesis y cómo construir intervalos de confianza para la diferencia entre las medias de dos poblaciones diferentes.

Contraste de hipótesis para la diferencia entre dos medias

Para ilustrar un **contraste para la diferencia entre dos medias**, sea μ_w el ingreso medio por hora en la población de mujeres recientemente tituladas en la universidad y sea μ_m la media poblacional para los hombres recientemente titulados. Considérese la hipótesis nula de que los ingresos medios para esas dos poblaciones difieren en una cierta cuantía, digamos d_0 . Entonces la hipótesis nula y la hipótesis alternativa bilateral son

$$H_0: \mu_m - \mu_w = d_0 \text{ vs. } H_1: \mu_m - \mu_w \neq d_0. \quad (3.18)$$

La hipótesis nula de que los hombres y mujeres de esas poblaciones presentan los mismos ingresos medios corresponde a H_0 en la Ecuación (3.18) con $d_0 = 0$.

Al ser las medias de esas poblaciones desconocidas deben estimarse a partir de las muestras de los hombres y las mujeres. Supongamos que tenemos muestras de n_m hombres y n_w mujeres seleccionados de forma aleatoria a partir de sus poblaciones. Sea la media muestral de los ingresos anuales \bar{Y}_m para los hombres e \bar{Y}_w para las mujeres. Entonces un estimador de $\mu_m - \mu_w$ es $\bar{Y}_m - \bar{Y}_w$.

Para contrastar la hipótesis nula de que $\mu_m - \mu_w = d_0$ mediante $\bar{Y}_m - \bar{Y}_w$, es necesario conocer la distribución de $\bar{Y}_m - \bar{Y}_w$. Recordemos que \bar{Y}_m está, de acuerdo con el teorema central del límite, distribuida aproximadamente como una $N(\mu_m, \sigma_m^2/n_m)$, donde σ_m^2 es la varianza poblacional de los ingresos de los hombres. De forma similar, \bar{Y}_w está distribuida aproximadamente como $N(\mu_w, \sigma_w^2/n_w)$, donde σ_w^2 es la varianza poblacional de los ingresos de las mujeres. Además, recordemos de la Sección 2.4 que una media ponderada de dos variables aleatorias normales está asimismo distribuida normalmente. Como \bar{Y}_m y \bar{Y}_w están construidas a partir de diferentes muestras obtenidas aleatoriamente, son variables aleatorias independientes. Por tanto, $\bar{Y}_m - \bar{Y}_w$ está distribuida $N[\mu_m - \mu_w, (\sigma_m^2/n_m) + (\sigma_w^2/n_w)]$.

Si σ_m^2 y σ_w^2 son conocidas, entonces esta distribución aproximada normal puede utilizarse para calcular p -valores para el contraste de la hipótesis nula de que $\mu_m - \mu_w = d_0$. En la práctica, sin embargo, estas varianzas poblacionales son habitualmente desconocidas por lo que deben estimarse. Como antes, pueden estimarse mediante las varianzas muestrales s_m^2 y s_w^2 , donde s_m^2 se define como en la Ecuación (3.7), excepto que el estadístico se calcula solamente para los hombres de la muestra, y s_w^2 , se define similarmente para las mujeres. Por tanto el error estándar de $\bar{Y}_m - \bar{Y}_w$ es

$$ES(\bar{Y}_m - \bar{Y}_w) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}. \quad (3.19)$$

Para una versión simplificada de la Ecuación (3.19) cuando Y es una variable aleatoria de Bernouilli véase el Ejercicio 3.15.

El estadístico t para contrastar la hipótesis nula se construye análogamente al estadístico t para contrastar una hipótesis sobre una media poblacional, restando el valor bajo la hipótesis nula de $\mu_m - \mu_w$ a partir del estimador $\bar{Y}_m - \bar{Y}_w$, y dividiendo el resultado por el error estándar de $\bar{Y}_m - \bar{Y}_w$:

$$t = \frac{(\bar{Y}_m - \bar{Y}_w) - d_0}{ES(\bar{Y}_m - \bar{Y}_w)} \quad (\text{estadístico } t \text{ para comparar dos medias}). \quad (3.20)$$

Si tanto n_m como n_w son grandes, entonces el estadístico t presenta una distribución normal estándar.

Debido a que el estadístico t de la Ecuación (3.20) presenta una distribución normal estándar bajo la hipótesis nula cuando n_m y n_w son grandes, el p -valor del contraste bilateral se calcula exactamente igual a como se hizo en el caso de una única población. Es decir, el p -valor se calcula utilizando la Ecuación (3.14).

Para llevar a cabo un contraste con nivel de significación predeterminado, simplemente se calcula el estadístico t de la Ecuación (3.20) y se compara con el valor crítico apropiado. Por ejemplo, la hipótesis nula se rechazaría al 5 % de nivel de significación si el valor absoluto del estadístico t supera 1,96.

Si la alternativa es unilateral en vez de bilateral (es decir, si la alternativa es que $\mu_m - \mu_w > d_0$), entonces el contraste se modifica como se sugirió en la Sección 3.2. El p -valor se calcula mediante la Ecuación (3.17), y un contraste con un nivel de significación del 5 % se rechaza para $t > 1,64$.

Intervalos de confianza para la diferencia entre dos medias poblacionales

El método para la construcción de intervalos de confianza recogido en la Sección 3.3 es extensible a la construcción del intervalo de confianza para la diferencia entre las medias, $d = \mu_m - \mu_w$. Debido a que el valor hipotético d_0 se rechaza al 5 % de nivel de significación si $|t| > 1,96$, d_0 estará en el conjunto de confianza si $|t| \leq 1,96$. Pero $|t| \leq 1,96$ significa que la diferencia estimada, $\bar{Y}_m - \bar{Y}_w$, estará menos alejada

de d_0 que 1,96 errores estándar. Por tanto el intervalo de confianza bilateral al 95 % para d consiste en los valores de d dentro de $\pm 1,96$ errores estándar de $\bar{Y}_m - \bar{Y}_w$:

El intervalo de confianza para $d = \mu_m - \mu_w$, es

$$(\bar{Y}_m - \bar{Y}_w) \pm 1,96ES(\bar{Y}_m - \bar{Y}_w). \quad (3.21)$$

Una vez disponemos de estas fórmulas, el recuadro «La brecha de género en los ingresos salariales de los titulados universitarios en los Estados Unidos» presenta una investigación empírica sobre las diferencias de género en los ingresos salariales de los titulados universitarios de EE.UU.

3.5 Estimación de la diferencia de medias de los efectos causales mediante datos experimentales

Recordemos de la Sección 1.2 que un experimento aleatorizado controlado selecciona aleatoriamente sujetos (individuos o, de forma más general, entidades) de una población de interés, posteriormente se les asigna aleatoriamente o bien a un grupo de tratamiento, que recibe el tratamiento experimental, o bien a un grupo de control, que no recibe el tratamiento. La diferencia entre las medias muestrales del grupo de tratamiento y de control es un estimador de los efectos causales del tratamiento.

Los efectos causales como diferencia de las esperanzas condicionales

El efecto causal de un tratamiento es el efecto previsto sobre los resultados de interés del tratamiento de acuerdo con lo medido en un experimento aleatorizado controlado ideal. Este efecto puede ser expresado como la diferencia de dos esperanzas condicionales. En concreto, **el efecto causal** sobre Y de un nivel de tratamiento x es la diferencia de las esperanzas condicionales, $E(Y|X = x) - E(Y|X = 0)$, donde $E(Y|X = x)$ es el valor esperado de Y para el grupo de tratamiento (que recibe el nivel de tratamiento $X = x$) en un experimento aleatorio controlado ideal y $E(Y|X = 0)$ es el valor esperado de Y para el grupo de control (que recibe el nivel de tratamiento $X = 0$). En el contexto de los experimentos, el efecto causal se denomina asimismo **efecto del tratamiento**. Si solamente existen dos niveles de tratamiento (es decir, si el tratamiento es binario), entonces podemos hacer que $X = 0$ exprese el grupo de control y $X = 1$ refleje el grupo de tratamiento. Si el tratamiento es un tratamiento binario, entonces el efecto causal (es decir, el efecto del tratamiento) es $E(Y|X = 1) - E(Y|X = 0)$ en un experimento aleatorizado controlado ideal.

Estimación de los efectos causales mediante las diferencias de medias

Si el tratamiento en un experimento aleatorizado controlado es binario, entonces el efecto causal puede ser estimado por la diferencia en los resultados medios muestrales entre los grupos de tratamiento y de control. La hipótesis de que el tratamiento es ineficaz es equivalente a la hipótesis de que ambas medias son iguales, lo cual puede contrastarse utilizando el estadístico t para comparar dos medias, dado en la Ecuación (3.20). Un intervalo de confianza al 95 % para la diferencia de medias de los dos grupos es un intervalo de confianza al 95 % para el efecto causal, por lo que puede construirse un intervalo de confianza al 95 % para el efecto causal utilizando la Ecuación (3.21).

Un experimento bien diseñado y llevado a cabo correctamente puede proporcionar una estimación convincente de un efecto causal. Por esta razón, habitualmente se llevan a cabo los experimentos aleatorizados controlados en algunas disciplinas, como la medicina. En economía, sin embargo, los experimentos suelen ser caros, difíciles de administrar, y, en algunos casos, éticamente cuestionables, por lo que siguen siendo muy escasos. Por esta razón, los económetras a veces analizan «experimentos naturales», asimismo llamados cuasi experimentos, en los que algún suceso relacionado con las características del tratamiento o del sujeto tiene el efecto de asignar los diferentes tratamientos a diferentes sujetos, *como si* hubieran sido parte de un experimento aleatorizado controlado. El recuadro, «Una nueva forma de fomentar el ahorro para la jubilación», proporciona un ejemplo de cuasi experimento que da lugar a algunas conclusiones sorprendentes.

La brecha de género en los ingresos salariales de los titulados universitarios en los Estados Unidos

El recuadro del Capítulo 2, «La distribución de ingresos salariales en Estados Unidos en 2008», muestra que, en media, los titulados universitarios varones ganan más que las tituladas mujeres. ¿Cuál es la tendencia reciente de esta «brecha de género» en los ingresos? Las normas sociales y las leyes que rigen la discriminación por género en el lugar de trabajo han cambiado sustancialmente en los Estados Unidos. ¿Es estable la brecha de género en los ingresos salariales de los titulados universitarios o ha disminuido en el tiempo?

La Tabla 3.1 presenta las estimaciones de los ingresos por hora de los graduados universitarios a tiempo completo para los trabajadores entre 25-34 años en los Estados Unidos en 1992, 1996, 2000, 2004 y 2008, utilizando los datos recopilados por el Censo Actualizado de Población. Los ingresos del año 1992, 1996, 2000 y 2004 están ajustados por la inflación, y están expresados en dólares de 2008 por el Índice de Precios al Consumo (IPC)¹. En 2008, los ingresos medios por hora de los 1.838 hombres encuestados fue de 24,98 \$, y la desviación típica de los ingresos para los hombres fue 11,78 \$. El promedio de los ingresos por hora en 2008 de las 1.871 mujeres encuestadas fue 20,87 \$, y la desviación típica de los ingresos fue de 9,66 \$. Por tanto la estimación de la brecha de género en los ingresos para el año 2008 es de 4,11 \$ (24,98 \$ - 20,87 \$), con un error estándar de 0,35 \$ ($= \sqrt{11,78^2/1.838 + 9,66^2/1.871}$). El intervalo de confianza al 95 % para la brecha de género en los ingresos salariales en 2008 es $4,11 \pm 1,96 \times 0,35 = (3,41 \$, 4,80 \$)$.

Los resultados de la Tabla 3.1 sugieren cuatro conclusiones. En primer lugar, la brecha de género es grande. Una brecha de 4,11 \$ por hora podría no parecer mucho, pero en un año asciende hasta 8.220 \$, suponiendo una semana laboral de 40 horas y 50 semanas laborables al año. En segundo lugar, entre 1992 y 2008, la brecha de género estimada aumentó en 0,87 \$ por hora en términos reales, desde 3,22 \$ por hora hasta 4,11 \$ por hora; sin embargo, este aumento no es estadísticamente significativo al 5 % de nivel de significación (Ejercicio

3.17). En tercer lugar, la brecha es grande si se mide en términos porcentuales: de acuerdo con las estimaciones de la Tabla 3.1, en 2008 las mujeres ganaban un 16 % menos por hora que los hombres (4,11 \$/24,98 \$), algo más que la brecha del 14 % vista en el año 1992. En cuarto lugar, la brecha de género es menor para los titulados universitarios jóvenes (el grupo analizado en la Tabla 3.1) que para los graduados universitarios en su conjunto (analizados en la Tabla 2.4): como se indica en la Tabla 2.4, los ingresos medios para el conjunto de las mujeres con educación universitaria trabajando a tiempo completo en 2008 eran de 23,93 \$, mientras que para los hombres esta media fue de 30,97 \$, lo que corresponde a una brecha de género del 23 % [$= (30,97 - 23,93)/30,97$] entre todos los trabajadores a tiempo completo con educación universitaria.

Este análisis empírico documenta que la «brecha de género» en los ingresos salariales por hora es grande y ha sido bastante estable (o tal vez ha aumentado ligeramente) en el pasado reciente. El análisis no nos cuenta, sin embargo, por qué existe esta brecha. ¿Surge por discriminación en el mercado laboral? ¿Refleja las diferencias en las capacidades, la experiencia o la educación entre hombres y mujeres? ¿Refleja las diferencias en la elección de puestos de trabajo? ¿O existe alguna otra causa? Volveremos a estas cuestiones una vez que dispongamos de las herramientas del análisis de regresión múltiple, el tema de la Parte II.

¹ Debido a la inflación, un dólar de 1992 valía más que un dólar de 2008, en el sentido de que un dólar en 1992 podía comprar más bienes y servicios que los que podía comprar un dólar en 2008. Por tanto, los ingresos de 1992 no pueden compararse directamente con los ingresos de 2008 sin haber ajustado antes por la inflación. Un modo de hacer este ajuste es mediante el IPC, una medida del precio de una «cesta de la compra» de bienes y servicios de consumo construida por la Oficina de Estadísticas Laborales. En 16 años desde 1992 hasta 2008, la cesta de bienes y servicios tomada como referencia del IPC que costaba 100 \$ en 1992, costaba 153,40 en 2008. Para hacer comparables los ingresos de 1992 y 2008 en la Tabla 3.1, los salarios de 1992 se han inflactado por la cuantía de la inflación global de los precios, es decir, se han multiplicado los ingresos de 1992 por 1,534 para poderlos expresar en «dólares de 2008».

TABLA 3.1 Tendencias en los ingresos salariales por hora de los trabajadores titulados universitarios, edad 25-35. Desde 1992 hasta 2008, en dólares de 2008

Año	Hombres			Mujeres			Diferencia, hombres vs. mujeres			Intervalo de confianza para d al 95 %
	\bar{Y}_m	s_m	n_m	\bar{Y}_w	s_w	n_w	$\bar{Y}_m - \bar{Y}_w$	$SE(\bar{Y}_m - \bar{Y}_w)$		
1992	23,27	10,17	1.594	20,05	7,87	1.368	3,22**	0,33		2,58-3,88
1996	22,48	10,10	1.379	18,98	7,95	1.230	3,50**	0,35		2,80-4,19
2000	24,88	11,60	1.303	20,74	9,36	1.181	4,14**	0,42		3,32-4,97
2004	25,12	12,01	1.894	21,02	9,36	1.735	4,10**	0,36		3,40-4,80
2008	24,98	11,78	1.838	20,87	9,66	1.871	4,11**	0,35		3,41-4,80

Estas estimaciones están calculadas utilizando los datos de todos los trabajadores a tiempo completo con edades entre 25 y 34 años entrevistados en la Encuesta Actualizada de Población llevada a cabo en marzo del año siguiente (por ejemplo, los datos para 2008 se recogieron en marzo de 2009).

**La diferencia es significativamente distinta de cero al nivel de significación del 1 %.

3.6 Utilización del estadístico t cuando el tamaño muestral es pequeño

En las Secciones 3.2 a 3.5, el estadístico t se utiliza junto con los valores críticos de la distribución normal estándar para el contraste de hipótesis y para la construcción de intervalos de confianza. La utilización de la distribución normal estándar está justificada por el teorema central del límite, que es aplicable cuando el tamaño muestral es grande. Cuando el tamaño muestral es pequeño, la distribución normal estándar puede proporcionar una aproximación pobre para la distribución del estadístico t . Si, sin embargo, la distribución poblacional está por sí misma distribuida normalmente, entonces la distribución exacta (es decir, la distribución para muestras finitas; véase la Sección 2.6) del estadístico t para contrastar la media de una única población es la distribución t de Student con $n - 1$ grados de libertad, y con valores críticos que pueden obtenerse de la distribución t de Student.

El estadístico t y la distribución t de Student

El estadístico t para el contraste de la media. Consideremos el estadístico t utilizado para contrastar la hipótesis de que la media de Y es $\mu_{Y,0}$ utilizando los datos Y_1, \dots, Y_n . La fórmula para este estadístico está dada por la Ecuación (3.10), donde el error estándar de \bar{Y} está dado por la Ecuación (3.8). Sustituyendo la última expresión en la anterior se obtiene la fórmula del estadístico t :

$$t = \frac{\bar{Y} - \mu_{Y,0}}{\sqrt{s_Y^2/n}}, \quad (3.22)$$

donde s_Y^2 está dado en la Ecuación (3.7).

Como se vio en la Sección 3.2, bajo ciertas condiciones generales, el estadístico t presenta una distribución normal estándar si el tamaño muestral es grande y la hipótesis nula es cierta [véase la Ecuación (3.12)]. Aunque la aproximación normal estándar para el estadístico t resulta adecuada para un extenso rango de distribuciones de Y si n es grande, puede resultar poco adecuada si n es pequeño. La distribución exacta del estadístico t depende de la distribución de Y , y puede ser muy complicada. Existe, sin embargo, un caso particular en el que la distribución exacta del estadístico t es relativamente sencilla: si Y está normalmente distribuida, entonces el estadístico t de la Ecuación (3.22) presenta una distribución t de Student con $n - 1$ grados de libertad.

Para verificar este resultado, recordemos de la Sección 2.4 que la distribución t de Student con $n - 1$ grados de libertad se define como la distribución de $Z/\sqrt{W/(n - 1)}$ donde Z es una variable aleatoria con distribución normal estándar, W es una variable aleatoria con una distribución chi-cuadrado con $n - 1$ grados de libertad, y Z y W están independientemente distribuidas. Cuando Y_1, \dots, Y_n son i.i.d. y la distribución poblacional de Y es $N(\mu_Y, \sigma_Y^2)$, el estadístico t puede escribirse como un cociente. Concretamente, sea $Z = (\bar{Y} - \mu_{Y,0})/\sqrt{\sigma_Y^2/n}$ y sea $W = (n - 1)s_Y^2/\sigma_Y^2$; entonces con algo de álgebra⁵ se puede demostrar que el estadístico t de la Ecuación (3.22) puede escribirse como $t = Z/\sqrt{W/(n - 1)}$. Recordemos de la Sección 2.4 que si Y_1, \dots, Y_n son i.i.d. y la distribución poblacional de Y es $N(\mu_Y, \sigma_Y^2)$, entonces la distribución muestral de \bar{Y} es exactamente $N(\mu_Y, \sigma_Y^2/n)$ para todo n ; por tanto, si la hipótesis nula $\mu_Y = \mu_{Y,0}$ es cierta, entonces $Z = (\bar{Y} - \mu_{Y,0})/\sqrt{\sigma_Y^2/n}$ presenta una distribución normal estándar para todo n . Además, $W = (n - 1)s_Y^2/\sigma_Y^2$ presenta una distribución χ_{n-1}^2 para todo n , e \bar{Y} y s_Y^2 están independientemente distribuidas. Esto lleva a que si la distribución poblacional de Y es normal, entonces bajo la hipótesis nula, el estadístico t dado en la Ecuación (3.22) tiene una distribución exacta t de Student con $n - 1$ grados de libertad.

⁵ La expresión deseada se obtiene multiplicando y dividiendo por $\sqrt{\sigma_Y^2}$ y agrupando términos:

$$t = \frac{\bar{Y} - \mu_{Y,0}}{\sqrt{s_Y^2/n}} = \frac{(\bar{Y} - \mu_{Y,0})}{\sqrt{\sigma_Y^2/n}} \div \sqrt{\frac{s_Y^2}{\sigma_Y^2}} = \frac{(\bar{Y} - \mu_{Y,0})}{\sqrt{\sigma_Y^2/n}} \div \sqrt{\frac{(n - 1)s_Y^2/\sigma_Y^2}{n - 1}} = Z \div \sqrt{W/(n - 1)}.$$

Si la distribución poblacional está distribuida normalmente, entonces pueden utilizarse los valores críticos de la distribución t de Student para llevar a cabo contrastes de hipótesis y para construir intervalos de confianza. A modo de ejemplo, considérese el problema hipotético en el que $t^{act} = 2,15$ y $n = 20$, por lo que los grados de libertad son $n - 1 = 19$. A partir de la Tabla 2 del Apéndice, el valor crítico bilateral al 5 % para la distribución t_{19} es 2,09. Como el estadístico t es mayor en valor absoluto que el valor crítico ($2,15 > 2,09$), la hipótesis nula sería rechazada al nivel de significación del 5 % frente a la alternativa bilateral. El intervalo de confianza al 95 % para μ_Y , construido utilizando la distribución t_{19} , sería $\bar{Y} \pm 2,09 ES(\bar{Y})$. Este intervalo de confianza es un poco más ancho que el intervalo de confianza construido utilizando el valor crítico 1,96 de la normal estándar.

El estadístico t para el contraste de diferencia de medias. El estadístico t para el contraste de la diferencia de dos medias, dado en la Ecuación (3.20), no tiene una distribución t de Student, incluso si la distribución poblacional de Y es normal. La distribución t de Student no es aplicable aquí porque el estimador de la varianza utilizado para calcular el error estándar en la Ecuación (3.19) no genera un denominador del estadístico t con distribución chi-cuadrado.

Una versión modificada del estadístico t para la diferencia de medias, basado en una fórmula para el error estándar diferente —la fórmula del error estándar «agrupado»— tiene una distribución exacta t de Student cuando Y está normalmente distribuida; sin embargo, la fórmula del error estándar agrupado es aplicable solamente en el caso particular de que los dos grupos tengan la misma varianza o de que ambos grupos tengan el mismo número de observaciones (Ejercicio 3.21). Adoptamos la notación de la Ecuación (3.19) de forma que los grupos se denominan m y w . El estimador de la varianza agrupada es

$$s_{\text{agrupada}}^2 = \frac{1}{n_m + n_w - 2} \left[\sum_{\substack{i=1 \\ \text{grupo } m}}^{n_m} (Y_i - \bar{Y}_m)^2 + \sum_{\substack{i=1 \\ \text{grupo } w}}^{n_w} (Y_i - \bar{Y}_w)^2 \right], \quad (3.23)$$

donde el primer sumatorio es para las observaciones del grupo m y el segundo sumando es para las observaciones del grupo w . El error estándar agrupado de la diferencia de medias es $ES_{\text{agrupado}}(\bar{Y}_m - \bar{Y}_w) = s_{\text{agrupada}} \times \sqrt{1/n_m + 1/n_w}$ y el estadístico t agrupado se calcula utilizando la Ecuación (3.20), donde el error estándar es el error estándar agrupado, $ES_{\text{agrupado}}(\bar{Y}_m - \bar{Y}_w)$.

Si la distribución poblacional de Y en el grupo m es $N(\mu_m, \sigma_m^2)$, si la distribución poblacional de Y en el grupo w es $N(\mu_w, \sigma_w^2)$, y si para los dos grupos ambas varianzas son iguales (es decir, $\sigma_m^2 = \sigma_w^2$), entonces bajo la hipótesis nula el estadístico t calculado mediante el error estándar agrupado presenta la distribución t de Student con $n_m + n_w - 2$ grados de libertad.

El inconveniente de utilizar el estimador de la varianza agrupada s_{agrupada}^2 es que solo es aplicable si las dos varianzas poblacionales son iguales (suponiendo $n_m \neq n_w$). Si las varianzas poblacionales son diferentes, el estimador de la varianza agrupada es sesgado e inconsistente. Si las varianzas poblacionales son diferentes, pero aun así se utiliza la fórmula de la varianza agrupada, la distribución bajo la nula del estadístico agrupado no es la distribución t de Student, incluso aunque los datos estén distribuidos normalmente; en realidad ni siquiera tiene una distribución normal estándar en muestras grandes. Por tanto, el error estándar agrupado y el estadístico t no deberían utilizarse a menos que se tenga una buena razón para creer que las varianzas poblacionales son iguales.

La utilización de la distribución t de Student en la práctica

Para el problema de contrastar la media de Y , la distribución t de Student es aplicable si la distribución poblacional de Y subyacente es normal. Sin embargo, en las variables económicas, las distribuciones normales son la excepción (por ejemplo, véanse los recuadros del Capítulo 2, «La distribución de los ingresos salariales en Estados Unidos en 2008» y «Un mal día en Wall Street»). Incluso si los datos subyacentes no están normalmente distribuidos, la aproximación normal para la distribución del estadístico t es válida si el tamaño de la muestra es grande. Por lo tanto, la inferencia —los contrastes de hipótesis y los intervalos de confianza— acerca de la media de una distribución debería basarse en la aproximación normal para muestras grandes.

Una nueva forma de fomentar el ahorro para la jubilación

Muchos economistas sostienen que la gente no ahorra lo suficiente para su jubilación. Los métodos convencionales para fomentar el ahorro destinado a la jubilación se centran en los incentivos financieros, pero además ha habido asimismo un aumento del interés en las formas no convencionales para el fomento del ahorro para la jubilación.

En un estudio importante publicado en 2001, Brigitte Madrian y Dennis Shea presentaron uno de estos métodos no convencionales para el fomento del ahorro para la jubilación. Muchas empresas ofrecen planes de ahorro para la jubilación en los que la empresa deposita, en su totalidad o en parte, el ahorro procedente de la nómina de los empleados adscritos. La adscripción a estos planes, llamados planes 401 (k) debido a la sección correspondiente del código fiscal de EE.UU., es siempre opcional. Sin embargo, en algunas empresas, los empleados son inscritos automáticamente en un plan, aunque pueden optar por abandonarlo; en otras empresas, los empleados están inscritos solo si así lo desean. De acuerdo con los modelos económicos convencionales sobre el comportamiento, el método de inscripción —bien sea optar por estar incluido u optar por no estarlo— no importa: El trabajador racional calculará su estrategia óptima y la llevará a cabo. No obstante, Madrian y Shea se cuestionaron ¿podría estar equivocada la teoría económica convencional? ¿Podría el método de inscripción afectar directamente a la tasa de inscripción?

Para medir el efecto del método de inscripción, Madrian y Shea estudiaron una gran empresa que cambió la opción pre-determinada de inscripción en su plan 401 (k) desde la no inscripción o participación a la de participación. Compararon dos grupos de trabajadores: los contratados el año anterior al cambio, que no estaban inscritos de forma automática (pero que podían optar por estarlo), y los contratados en el año siguiente al cambio y por tanto inscritos automáticamente (pero que podían optar por no estarlo). Las condiciones financieras del

plan siguieron siendo las mismas, además Madrian y Shea no encontraron diferencias sistemáticas entre los trabajadores contratados antes y después del cambio. Por lo tanto, desde la perspectiva de un económetra el cambio resulta como un tratamiento asignado al azar y el efecto causal del cambio podría ser estimado por la diferencia de medias entre los dos grupos.

Madrian y Shea hallaron que la norma de inscripción por defecto causaba una gran diferencia: la tasa de inscripción para el grupo «optar por la inscripción» (de control) fue del 37,4 % ($n = 4.249$), mientras que la tasa de inscripción para el grupo «optar por la no inscripción» (de tratamiento) fue del 85,9 %. ($n = 5.801$). La estimación del efecto del tratamiento fue del 48,5 % (= 85,9 % – 37,4 %). Debido a que su muestra era grande, el intervalo de confianza al 95 % (calculado en el Ejercicio 3.15) para el efecto del tratamiento es estrecho, 46,8 % al 50,2 %.

¿Cómo podría importar tanto la elección de la opción pre-determinada por la empresa? Tal vez los trabajadores consideraban esas decisiones financieras demasiado confusas, o tal vez simplemente no querían pensar en envejecer. Ninguna de estas explicaciones resulta económicamente racional —pero ambas son consistentes con las predicciones del incipiente campo de la «economía del comportamiento» y ambas conllevarían la adopción de la inscripción por defecto.

Esta investigación tuvo un efecto práctico importante. En agosto de 2006, el Congreso aprobó la Ley de Protección de las Pensiones que (entre otras cosas) animaba a las empresas a ofrecer planes 401 (k) en los que la inscripción se realizase por defecto. Los resultados económicos de Madrian y Shea y otros ocuparon un lugar destacado en la argumentación de esta legislación.

Para conocer más sobre la economía del comportamiento y el diseño de planes de ahorro para la jubilación, véase Benartzi y Thaler (2007) y Beshears, Choi, Laibson, y Madrian (2008).

Al comparar las dos medias, cualquier razón económica que justifique que dos grupos distintos puedan tener diferentes medias habitualmente implica asimismo que los dos grupos podrían tener varianzas distintas. En consecuencia, la fórmula del error estándar agrupado no resulta adecuada, y la fórmula del error estándar correcta, que permite distintas varianzas entre grupos, es como la de la Ecuación (3.19). Incluso si las distribuciones poblacionales son normales, el estadístico t calculado mediante la fórmula del error estándar de la Ecuación (3.19) no tiene una distribución t de Student. En la práctica, por lo tanto, las inferencias acerca de las diferencias en las medias deberían basarse en la Ecuación (3.19), utilizada conjuntamente con la aproximación normal estándar para muestras grandes.

A pesar de que la distribución t de Student resulta escasamente aplicable en economía, algunos programas de software utilizan la distribución t de Student para calcular los p -valores y los intervalos de confianza. En la práctica, esto no supone un problema, porque la diferencia entre la distribución t de Student y la distribución normal estándar es insignificante si el tamaño muestral es grande. Para $n > 15$, la diferencia en los p -valores calculados utilizando las distribuciones t de Student y la normal estándar, nunca supera 0,002. En la mayoría de las aplicaciones modernas, y en todas las aplicaciones de este libro, el tamaño de las muestras son cientos o miles, suficientemente grandes para que la diferencia entre la distribución t de Student y la distribución normal estándar sea insignificante.

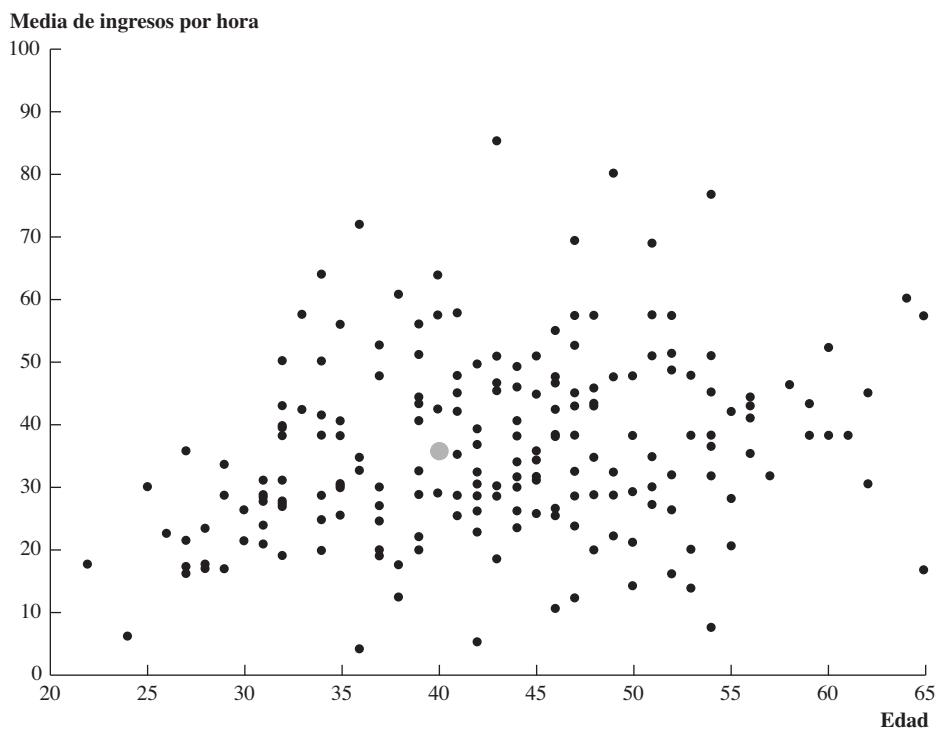
3.7 Diagramas de dispersión, covarianza muestral, y correlación muestral

¿Cuál es la relación entre la edad y los ingresos salariales? Esta pregunta, como muchas otras, relaciona una variable X (edad), con otra, Y (ingresos). Esta sección examina tres maneras de recoger la relación entre variables: el diagrama de dispersión, la covarianza muestral, y el coeficiente de correlación muestral.

Diagramas de dispersión

Un **diagrama de dispersión** es una gráfica de n observaciones sobre X_i e Y_i , en la que cada observación está representada por el punto (X_i, Y_i) . Por ejemplo, la Figura 3.2 es un diagrama de dispersión de la edad (X) y los ingresos salariales por hora (Y) para una muestra de 200 directivos de la industria de la información procedente de la Encuesta Actualizada de Población (CPS) en marzo de 2009. Cada punto de la Figura 3.2 corresponde a un par (X, Y) de una de las observaciones. Por ejemplo, uno de los trabajadores en esta muestra tiene 40 años y gana 35,78 \$ por hora; la edad de este trabajador y sus ingresos están representados por el punto destacado en la Figura 3.2. El diagrama de dispersión muestra una relación positiva entre la edad y los ingresos en esta muestra: los trabajadores de mayor edad tienden a ganar más que los trabajadores más jóvenes. Sin embargo, esta relación no es exacta, y los ingresos podrían no poder predecirse perfectamente usando solamente la edad de una persona.

FIGURA 3.2 Diagrama de dispersión de los ingresos salariales medios por hora vs. edad



Cada punto de la gráfica representa la edad y los ingresos salariales de uno de los 200 trabajadores de la muestra. El punto destacado corresponde a un trabajador de 40 años que ingresa 35,78 \$ por hora. Los datos corresponden a directivos informáticos y directivos de sistemas de información de la CPS de marzo de 2009.

Covarianza muestral y correlación

La covarianza y la correlación se introdujeron en la Sección 2.3 como dos propiedades de la distribución de probabilidad conjunta de las variables aleatorias X e Y . Debido a que la distribución de la población es desconocida, en la práctica no se conocen ni la covarianza ni la correlación poblacionales. La covarianza y la correlación poblacional ES pueden calcularse, no obstante, tomando una muestra aleatoria de n miembros de la población y registrando los datos (X_i, Y_i) , $i = 1, \dots, n$.

La covarianza y correlación muestrales son estimadores de la covarianza y correlación poblacionales. Al igual que los estimadores analizados previamente en este capítulo, se calculan sustituyendo la media poblacional (la esperanza), por la media muestral. La **covarianza muestral**, expresada mediante s_{XY} , es

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}). \quad (3.24)$$

Al igual que la varianza muestral, de la Ecuación (3.24) se calcula dividiendo por $n-1$ en lugar de n ; aquí, otra vez, esta diferencia se deriva del uso de \bar{X} e \bar{Y} para estimar sus respectivas medias poblaciones. Cuando n es grande, la diferencia entre si la división se realiza por n o por $n-1$ es pequeña.

El **coeficiente de correlación muestral**, o **correlación muestral**, se expresa mediante r_{XY} y es la ratio entre la covarianza muestral y las desviaciones típicas muestrales:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}. \quad (3.25)$$

La correlación muestral mide la intensidad de la asociación lineal entre X e Y en una muestra de n observaciones. Como la correlación poblacional, la correlación muestral no tiene unidades de medida y toma valores entre -1 y 1 : $|r_{XY}| \leq 1$.

La correlación muestral es igual a 1 si $X_i = Y_i$ para todo i y es igual a -1 si $X_i = -Y_i$ para todo i . De forma más general, la correlación es ± 1 si el diagrama de dispersión es una línea recta. Si la recta tiene pendiente positiva, entonces existe una relación positiva entre X e Y y la correlación es 1 . Si la recta tiene pendiente negativa, entonces existe una relación negativa y la correlación es -1 . Cuanto más cercano es el diagrama de dispersión a una línea recta, más cercana es la relación a ± 1 . Un coeficiente de correlación elevado no significa necesariamente que la línea tenga una pendiente pronunciada; más bien significa que los puntos del diagrama de dispersión se encuentran muy cerca de una línea recta.

Consistencia de la covarianza y correlación muestrales. Al igual que la varianza muestral, la covarianza muestral es consistente. Es decir,

$$s_{XY} \xrightarrow{P} \sigma_{XY}. \quad (3.26)$$

En otras palabras, la covarianza muestral en muestras grandes se encuentra cercana a la covarianza poblacional con alta probabilidad.

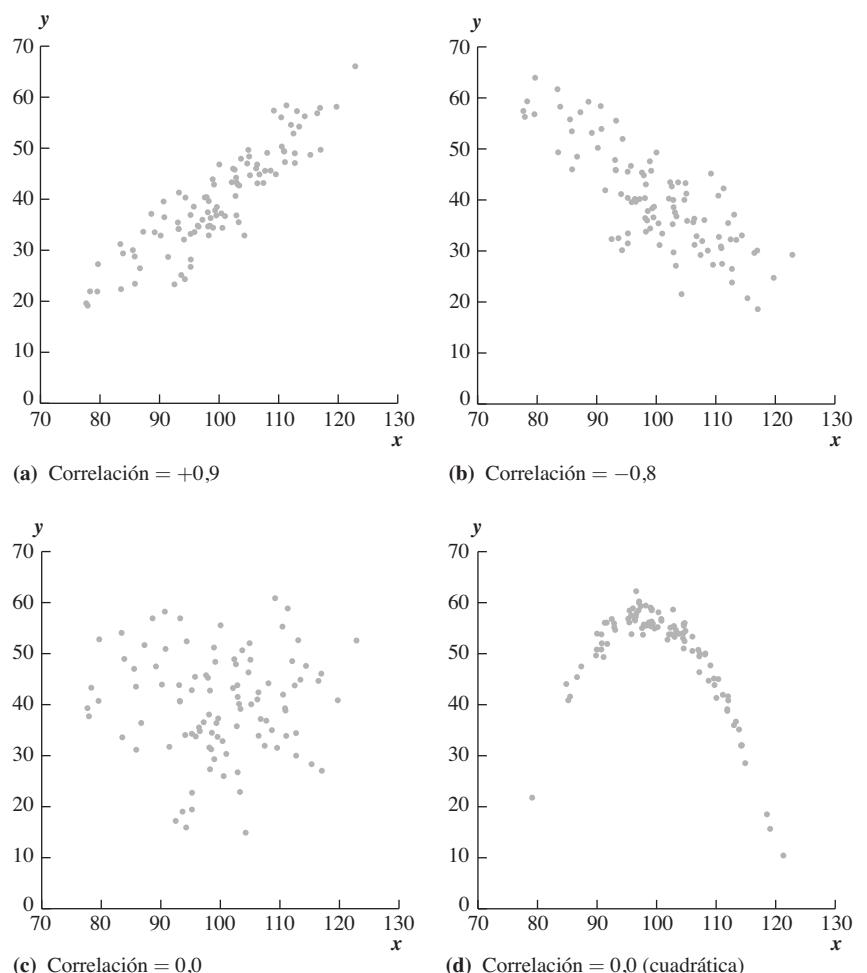
La demostración del resultado de la Ecuación (3.26), bajo el supuesto de que (X_i, Y_i) son i.i.d. y de que X_i e Y_i tienen momentos de cuarto orden finitos es similar a la demostración del Apéndice 3.3 de que la covarianza muestral es consistente y se deja como ejercicio (Ejercicio 3.20).

Debido a que la varianza y la covarianza de la muestra son consistentes, el coeficiente de correlación muestral es consistente, es decir, $r_{XY} \xrightarrow{P} \text{corr}(X_i, Y_i)$.

Ejemplo. A modo de ejemplo, considérense los datos de edad e ingresos de la Figura 3.2. Para estos 200 trabajadores, la desviación típica muestral de la edad es $s_A = 9,07$ años y la desviación típica muestral de los ingresos es $s_E = 14,37$ \$ por hora. La covarianza entre la edad y los ingresos es $s_{AE} = 33,16$ (las

FIGURA 3.3 Diagramas de dispersión para cuatro bases de datos hipotéticas

El diagrama de dispersión de las Figuras 3.3a y 3.3b muestran fuertes relaciones lineales entre X e Y . En la Figura 3.3c, X es independiente de Y y las dos variables están incorrelacionadas. En la Figura 3.3d, las dos variables están asimismo incorrelacionadas incluso aunque estén relacionadas de forma no lineal.



unidades son años \times dólares por hora, y no resultan fácilmente interpretables). Por tanto, el coeficiente de correlación es $r_{AE} = 33,16/(9,07 \times 14,37) = 0,25$ o 25 %. La correlación de 0,25 significa que existe una relación positiva entre edad e ingresos, pero como es evidente en el diagrama de dispersión, esta relación está lejos de ser perfecta.

Para verificar que la correlación no depende de las unidades de medida, supongamos que los ingresos estuvieran expresados en centavos de dólar, en cuyo caso la desviación típica de la muestra de ingresos sería 1.437 ¢ por hora y la covarianza entre edad e ingresos sería 3.316 (las unidades son años \times centavos por hora), entonces la correlación es $3.316/(9,07 \times 1.437) = 0,25$ o 25 %.

La Figura 3.3 proporciona otros ejemplos de diagramas de dispersión y correlación. La Figura 3.3a muestra una fuerte relación lineal positiva entre esas variables, y la correlación muestral es 0,9. La Figura 3.3b muestra una fuerte relación negativa con una correlación muestral de -0,8. La Figura 3.3c muestra un diagrama de dispersión, sin relación evidente, y la correlación muestral es cero. La Figura 3.3d muestra una relación clara: cuando X aumenta, Y inicialmente aumenta pero después disminuye. A pesar de esta clara relación entre X e Y , la correlación muestral es cero; la razón es que, para esos datos, los valores pequeños de Y están asociados tanto a valores grandes de X como a valores pequeños.

Este ejemplo final enfatiza un punto importante; el coeficiente de correlación es una medida acerca de la relación *lineal*. Existe una relación en la Figura 3.3d, pero no es lineal.

Resumen

1. La media muestral, \bar{Y} , es un estimador de la media de la población, μ_Y . Cuando Y_1, \dots, Y_n son i.i.d.
 - a) La distribución muestral de \bar{Y} tiene media $\mu_{\bar{Y}}$ y varianza $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$.
 - b) \bar{Y} es insesgado.
 - c) Por la ley de grandes números, \bar{Y} es consistente.
 - d) Por el teorema central del límite, \bar{Y} tiene una distribución muestral aproximada normal estándar cuando el tamaño de la muestra es grande.
2. El estadístico t se utiliza para contrastar la hipótesis nula de que la media poblacional toma un valor en concreto. Si n es grande, el estadístico t tiene una distribución muestral normal estándar cuando la hipótesis nula es verdadera.
3. El estadístico t se puede utilizar para calcular el p -valor asociado con la hipótesis nula. Un p -valor pequeño evidencia que la hipótesis nula es falsa.
4. Un intervalo de confianza del 95 % para μ_Y es un intervalo construido de manera que contenga el verdadero valor de μ_Y en el 95% de todas las muestras posibles.
5. Los contrastes de hipótesis y los intervalos de confianza para la diferencia de las medias de dos poblaciones son conceptualmente similares a los contrastes e intervalos para la media de una única población.
6. El coeficiente de correlación muestral es un estimador del coeficiente de correlación poblacional y mide la relación lineal entre dos variables —es decir, en qué medida su diagrama de dispersión se approxima razonablemente bien a una línea recta.

Términos clave

estimador (48)	error tipo II (56)
estimación (48)	nivel de significación (56)
sesgo, consistencia y eficiencia (49)	valor crítico (56)
ELIO (Estimador Lineal Insesgado Óptimo) (50)	región de rechazo (56)
estimador de mínimos cuadrados (50)	región de aceptación (56)
contrastos de hipótesis (51)	tamaño de un contraste (56)
hipótesis nula (51)	potencia de un contraste (56)
hipótesis alternativa (51)	hipótesis alternativa unilateral (57)
hipótesis alternativa bilateral (51)	conjunto de confianza (57)
p -valor (probabilidad de significación) (52)	nivel de confianza (57)
varianza muestral (53)	intervalo de confianza (57)
desviación típica muestral (53)	probabilidad de cobertura (58)
grados de libertad (54)	contraste para la diferencia entre dos medias (58)
error estándar de \bar{Y} (54)	efecto causal (60)
estadístico t (t -ratio) (54)	efecto del tratamiento (60)
estadístico de contraste (54)	diagrama de dispersión (65)
error tipo I (56)	covarianza muestral (66)
	coeficiente de correlación muestral (correlación muestral) (66)

Revisión de conceptos

- 3.1 Explique la diferencia entre la media muestral \bar{Y} y la media poblacional.
- 3.2 Explique la diferencia entre un estimador y una estimación. Proponga un ejemplo de cada uno.

3.3 Una distribución poblacional tiene una media de 10 y una varianza de 16. Determine la media y la varianza de \bar{Y} de una muestra i.i.d. de esta población para:

- a) $n = 10$ b) $n = 100$ c) $n = 1.000$

Relacione sus respuestas con la ley de los grandes números.

3.4 ¿Qué papel desempeña el teorema central del límite en la contrastación de hipótesis estadísticas? ¿Y en la construcción de intervalos de confianza?

3.5 ¿Cuál es la diferencia entre hipótesis nula y alternativa? ¿Y entre el tamaño, el nivel de significación, y la potencia? ¿Y entre una hipótesis alternativa unilateral y una hipótesis alternativa bilateral?

3.6 ¿Por qué un intervalo de confianza contiene más información que el resultado de un único contraste de hipótesis?

3.7 Explique por qué el estimador de las diferencias de las medias, aplicado a los datos de un experimento aleatorizado controlado, es un estimador del efecto del tratamiento.

3.8 Dibuje un diagrama de dispersión hipotético para una muestra de tamaño 10 para dos variables aleatorias con una correlación poblacional:

- a) 1,0 b) -1,0 c) 0,9 d) -0,5 e) 0,0

Ejercicios

3.1 En una población, $\mu_Y = 100$ y $\sigma_Y^2 = 43$. Utilice el teorema central del límite central para contestar las siguientes preguntas:

- a) En una muestra aleatoria de tamaño $n = 100$, hallar $\Pr(\bar{Y} < 101)$.
- b) En una muestra aleatoria de tamaño $n = 64$, hallar $\Pr(101 < \bar{Y} < 103)$.
- c) En una muestra aleatoria de tamaño $n = 165$, hallar $\Pr(\bar{Y} > 98)$.

3.2 Sea Y una variable aleatoria de Bernoulli con probabilidad de éxito $\Pr(Y = 1) = p$, sean Y_1, \dots, Y_n extracciones i.i.d. de esta distribución. Sea \hat{p} la proporción de éxitos (1's) en esta muestra.

- a) Demuestre que $\hat{p} = \bar{Y}$.
- b) Demuestre que \hat{p} es un estimador insesgado de p .
- c) Demuestre que $\text{var}(\hat{p}) = p(1 - p)/n$.

3.3 En una encuesta a 400 posibles votantes, 215 respondieron que votarían por el candidato saliente y 185 respondieron que votarían por el aspirante. Sea p la proporción de todos los posibles votantes que prefirieron el candidato saliente en el momento de la encuesta, y sea la proporción de los encuestados que prefirieron al candidato saliente.

- a) Utilice los resultados de la encuesta para estimar p .
- b) Utilice el estimador de la varianza de \hat{p} , $\hat{p}(1 - \hat{p})/n$, para calcular el error estándar del estimador.
- c) ¿Cuál es el p -valor para el contraste $H_0: p = 0,5$ vs. $H_1: p \neq 0,5$?
- d) ¿Cuál es el p -valor para el contraste $H_0: p = 0,5$ vs. $H_1: p > 0,5$?
- e) ¿Por qué difieren los resultados de (c) y (d)?
- f) ¿Proporciona la encuesta evidencia estadísticamente significativa de que el candidato saliente fuera por delante del candidato aspirante en el momento de la encuesta? Explíquelo.

3.4 Utilizando los datos del Ejercicio 3.3:

- a) Construya un intervalo de confianza al 95 % para p .
- b) Construya un intervalo de confianza al 99 % para p .
- c) ¿Por qué es el intervalo de (b) más ancho que el intervalo de (a)?
- d) Sin hacer ningún cálculo adicional, contrasta la hipótesis nula $H_0: p = 0,50$ vs. $H_1: p \neq 0,50$ al nivel de significación del 5 %.

- 3.5** Se lleva a cabo una encuesta sobre los 1.055 votantes registrados, y se solicita a los votantes que elijan entre el candidato A y el candidato B. Sea p la proporción de los votantes de la población que prefieren al candidato A, y sea \hat{p} la proporción de votantes de la muestra que prefieren al candidato A.
- Se está interesado en dos hipótesis rivales $H_0: p = 0,5$ vs. $H_1: p \neq 0,5$. Suponga que se decide rechazar H_0 si $|\hat{p} - 0,5| > 0,02$.
 - ¿Cuál es el tamaño de este contraste?
 - Calcule la potencia de este contraste si $p = 0,53$.
 - En la encuesta, $\hat{p} = 0,54$.
 - Contraste $H_0: p = 0,5$ vs. $H_1: p \neq 0,5$ utilizando un nivel de significación del 5 %.
 - Contraste $H_0: p = 0,5$ vs. $H_1: p > 0,5$ utilizando un nivel de significación del 5 %.
 - Construya un intervalo de confianza al 95 % para p .
 - Construya un intervalo de confianza al 99 % para p .
 - Construya un intervalo de confianza al 50 % para p .
 - Supóngase que la encuesta se realiza 20 veces, a partir de electores seleccionados de forma independiente en cada encuesta. Para cada una de estas 20 encuestas, se construye un intervalo de confianza al 95 % para p .
 - ¿Cuál es la probabilidad de que el verdadero valor de p se encuentre en la totalidad de los 20 intervalos de confianza?
 - ¿Cuántos de estos intervalos de confianza se espera que contengan el verdadero valor de p ?
 - En la jerga de encuestas, el «margen de error» es $1,96 \times ES(\hat{p})$; es decir, la mitad de la longitud del intervalo de confianza al 95 %. Supóngase que se desea diseñar una encuesta que tenga un margen de error de un máximo del 1 %. Es decir, se desea que $\Pr(|\hat{p} - p| > 0,01) \leq 0,05$. ¿Cuánto debe ser n si la encuesta utiliza muestreo aleatorio simple?
- 3.6** Sean Y_1, \dots, Y_n extracciones i.i.d. de una distribución con media μ . Un contraste de $H_0: \mu = 5$ vs. $H_1: \mu \neq 5$ que utiliza el estadístico t habitual arroja un p -valor de 0,03.
- ¿Contiene a $\mu = 5$ el intervalo de confianza al 95 %? Explíquelo.
 - ¿Se puede determinar si $\mu = 6$ se encuentra en el intervalo de confianza al 95 %? Explíquelo.
- 3.7** En una población dada, el 11 % de los posibles votantes son afroamericanos. Una encuesta que utiliza una muestra aleatoria simple de 600 números de teléfonos fijos halla un 8 % de afroamericanos. ¿Existe evidencia de que la encuesta está sesgada? Explíquelo.
- 3.8** Se ofrece una nueva versión del examen SAT a 1.000 estudiantes de secundaria seleccionados al azar. La calificación media en la prueba es 1.110, y la desviación típica de la muestra es 123. Construya un intervalo de confianza al 95 % para la calificación media poblacional para los estudiantes de secundaria.
- 3.9** Supóngase que una planta de fabricación de bombillas produce bombillas con una vida media de 2.000 horas y una desviación típica de 200 horas. Un inventor declara que ha desarrollado un proceso de mejora que permite producir bombillas con una mayor duración media y la misma desviación típica. La gerente de la planta selecciona 100 bombillas al azar producidas mediante ese proceso. Dice que ella creería la afirmación del inventor si la vida media muestral de las bombillas fuese mayor a 2.100 horas; de lo contrario, llegaría a la conclusión de que el nuevo proceso no es mejor que el proceso anterior. Sea μ la media del nuevo proceso. Considérense las hipótesis nula y alternativa $H_0: \mu = 2.000$ vs. $H_1: \mu > 2.000$.
- ¿Cuál es el tamaño del procedimiento de contraste de la gerente de la planta?
 - Supóngase que el nuevo proceso es en realidad mejor y proporciona una vida media a las bombillas de 2.150 horas. ¿Cuál es la potencia del procedimiento de contraste de la gerente de la planta?
 - ¿Qué procedimiento de contraste debería utilizar la gerente de la fábrica si quisiera un tamaño del 5 % para su contraste?

- 3.10** Supóngase que se somete a un nuevo examen estándar a 100 estudiantes de tercer curso seleccionados al azar en Nueva Jersey. La calificación media muestral en la prueba es de 58 puntos, y la desviación típica muestral es de 8 puntos.
- Los autores planean someter a examen a todos los estudiantes de tercer curso de Nueva Jersey. Construya un intervalo de confianza al 95 % para la calificación media de todos los estudiantes de tercer grado de Nueva Jersey.
 - Supóngase que se somete al mismo examen a 200 estudiantes de tercer curso de Iowa seleccionados al azar, y se obtiene una media en la muestra de 62 puntos y una desviación típica de 11 puntos. Construya un intervalo de confianza al 90 % para la diferencia entre las calificaciones medias de Iowa y Nueva Jersey.
 - ¿Se puede concluir con un alto grado de confianza que las medias poblacionales para los estudiantes de Iowa y Nueva Jersey son diferentes? (¿Cuál es el error estándar de la diferencia entre las dos medias muestrales? ¿Cuál es el p -valor del contraste de igualdad de las medias frente a la existencia de alguna diferencia?)
- 3.11** Considérese el estimador \tilde{Y} , definido en la Ecuación (3.1). Demuestre que:
- $E(\tilde{Y}) = \mu_Y$
 - $\text{var}(\tilde{Y}) = 1,25\sigma_Y^2/n$.
- 3.12** Para investigar la posible discriminación por género en una empresa, se seleccionó aleatoriamente una muestra de 100 hombres y 64 mujeres con una similar descripción de su trabajo. Se proporciona a continuación un resumen de los salarios mensuales registrados:

	Salario medio (\bar{Y})	Desviación típica (s_y)	n
Hombres	3.100 \$	200 \$	100
Mujeres	2.900 \$	320 \$	64

- ¿Qué expresan estos datos acerca de las diferencias salariales en la empresa? ¿Proporcionan evidencia estadísticamente significativa de que los salarios medios de hombres y mujeres son diferentes? (Para responder a esta pregunta, establezca antes las hipótesis nula y alternativa; en segundo lugar, calcule estadístico t relevante; en tercer lugar, calcule el p -valor asociado al estadístico t ; y finalmente, utilice el p -valor para responder a la pregunta).
 - ¿Sugieren estos datos que la empresa es culpable de discriminación de género con sus políticas retributivas? Explíquelo.
- 3.13** Los datos sobre las calificaciones en los exámenes de quinto curso (lectura y matemáticas) para los 420 distritos escolares de California arrojan una media $\bar{Y} = 646,2$ y una desviación típica $s_Y = 19,5$.
- Construya un intervalo de confianza al 95 % para la calificación media poblacional.
 - Cuando los distritos se dividieron entre distritos con grupos pequeños (< 20 estudiantes por maestro) y distritos con grupos grandes (≥ 20 estudiantes por maestro), los resultados fueron los siguientes:

Tamaño de la clase	Calificación media (\bar{Y})	Desviación típica (s_y)	n
Pequeño	657,4	19,4	238
Grande	650,0	17,9	182

- ¿Existe evidencia estadísticamente significativa de que los distritos con menor tamaño de las clases tienen un mayor promedio de puntuación en la prueba? Explíquelo.
- 3.14** Se han recogido los valores acerca de la altura en pulgadas (X) y del peso en libras (Y) a partir de una muestra de 300 estudiantes universitarios varones. Los estadísticos de resumen resultantes son: $\bar{X} = 70,5$ pulgadas, $\bar{Y} = 158$ libras, $s_X = 1,8$ pulgadas, $s_Y = 14,2$ libras, $s_{XY} = 21,73$ pulgadas por libras, y $r_{XY} = 0,85$. Convierta estos estadísticos al sistema métrico (metros y kilogramos).

- 3.15** Sean Y_a e Y_b las expresiones de las variables aleatorias de Bernoulli de dos poblaciones diferentes, denominadas a y b . Supongamos que $E(Y_a) = p_a$ y $E(Y_b) = p_b$. Se selecciona aleatoriamente una muestra de tamaño n_a de la población a , con una media muestral \hat{p}_a , y se selecciona una muestra aleatoria de tamaño n_b de la población b , con una media muestral expresada por \hat{p}_b . Supóngase que la muestra de la población a es independiente de la muestra de la población b .
- Demuestre que $E(\hat{p}_a) = p_a$ y $\text{var}(\hat{p}_a) = p_a(1 - p_a)/n_a$. Demuestre que $E(\hat{p}_b) = p_b$ y $\text{var}(\hat{p}_b) = p_b(1 - p_b)/n_b$.
 - Demuestre que $\text{var}(\hat{p}_a - \hat{p}_b) = \frac{p_a(1 - p_a)}{n_a} + \frac{p_b(1 - p_b)}{n_b}$. (*Pista:* recuerde que las muestras son independientes).
 - Supóngase que n_a y n_b son grandes. Demuestre que un intervalo de confianza al 95 % para $p_a - p_b$ está dado por $(\hat{p}_a - \hat{p}_b) \pm 1,96 \sqrt{\frac{\hat{p}_a(1 - \hat{p}_a)}{n_a} + \frac{\hat{p}_b(1 - \hat{p}_b)}{n_b}}$. ¿Cómo se construiría un intervalo de confianza al 90 % para $p_a - p_b$?
 - Léase el recuadro «Una nueva forma de fomentar el ahorro para la jubilación» en la Sección 3.5. Sea la población a el grupo «optar por la no adscripción» (de tratamiento) y la población b el grupo «optar por la adscripción» (de control). Construya un intervalo de confianza al 95 % para el efecto del tratamiento, $p_a - p_b$.
- 3.16** Se sabe que las calificaciones en un examen estandarizado tienen una media de 1.000 para los estudiantes en los Estados Unidos. Se realiza el examen a 453 estudiantes seleccionados al azar en Florida; en esta muestra, la media es de 1.013 y la desviación típica (s) es de 108.
- Construya un intervalo de confianza al 95 % para la calificación media de los estudiantes de Florida.
 - ¿Existe evidencia estadísticamente significativa de que los estudiantes de Florida se comportan de modo diferente a otros estudiantes de Estados Unidos?
 - Se seleccionan aleatoriamente otros 503 estudiantes de Florida. Se les imparte un curso de preparación de 3 horas antes del examen. Su puntuación media en la prueba es 1.019 con una desviación típica de 95.
 - Construya un intervalo de confianza al 95 % para la variación de la calificación media en los exámenes asociadas al curso de preparación.
 - ¿Existe evidencia estadísticamente significativa de que el curso de preparación ayudó?
 - A los 453 estudiantes originales se les imparte el curso de preparación y después se les insta a hacer el examen otra vez. La variación media de la calificación en sus exámenes es de 9 puntos, y la desviación típica de la variación es de 60 puntos.
 - Construya un intervalo de confianza al 95 % para la variación en la media de las calificaciones en los exámenes.
 - ¿Existe evidencia estadísticamente significativa de que los estudiantes realizarán mejor los exámenes en su segundo intento tras recibir el curso de preparación?
 - Los estudiantes pueden tener un mejor desempeño en su segundo intento debido al curso de preparación o bien debido a la experiencia adquirida al hacer el examen en su primer intento. Describa un experimento con el que cuantificar estos dos efectos.
- 3.17** Léase el recuadro de «La brecha de género en los ingresos salariales de los titulados universitarios en Estados Unidos» de la Sección 3.5.
- Construya un intervalo de confianza al 95 % para la variación en los ingresos medios por hora de los hombres entre 1992 y 2008.
 - Construya un intervalo de confianza al 95 % para la variación en los ingresos medios por hora de las mujeres entre 1992 y 2008.

- c) Construya un intervalo de confianza al 95 % para la variación de la brecha de género en los ingresos medios por hora entre 1992 y 2008. (*Pista:* $\bar{Y}_{m,1992} - \bar{Y}_{w,1992}$ es independiente de $\bar{Y}_{m,2008} - \bar{Y}_{w,2008}$).
- 3.18** Este ejercicio demuestra que la varianza muestral es un estimador insesgado de la varianza poblacional si Y_1, \dots, Y_n son i.i.d. con media μ_Y y varianza σ_Y^2 .
- Utilice la Ecuación (2.31) para demostrar que $E[(Y_i - \bar{Y})^2] = \text{var}(Y_i) - 2 \text{cov}(Y_i, \bar{Y}) + \text{var}(\bar{Y})$.
 - Utilice la Ecuación (2.33) para demostrar que $\text{cov}(\bar{Y}, Y_i) = \sigma_Y^2/n$.
 - Utilice los resultados de (a) y (b) para demostrar que $E(s_Y^2) = \sigma_Y^2$.
- 3.19** a) \bar{Y} es un estimador insesgado de μ_Y . ¿ \bar{Y}^2 es un estimador insesgado de μ_Y^2 ?
- b) \bar{Y} es un estimador consistente de μ_Y . ¿ \bar{Y}^2 es un estimador consistente de μ_Y^2 ?
- 3.20** Supóngase que (X_i, Y_i) son i.i.d. con momentos de cuarto orden finitos. Demuestre que la covarianza de la muestra es un estimador consistente de la covarianza de la población, es decir, $s_{XY} \xrightarrow{P} \sigma_{XY}$, donde s_{XY} se define en la Ecuación (3.24). (*Pista:* utilice la estrategia del Apéndice 3.3 y la desigualdad de Cauchy-Schwartz).
- 3.21** Demuestre que el error estándar agrupado $[ES_{\text{agrupado}}(\bar{Y}_m - \bar{Y}_w)]$ dado siguiendo la Ecuación (3.23), es igual al error estándar habitual de la diferencia en las medias de la Ecuación (3.19) cuando el tamaño de los dos grupos es el mismo ($n_m = n_w$).

Ejercicio empírico

- E3.1** En la página web del libro http://www.pearsonhighered.com/stock_watson se encuentra el archivo de datos **CPS92_08** que contiene una versión ampliada de la base de datos utilizada en la Tabla 3.1 de este libro para los años 1992 y 2008. Contiene datos sobre trabajadores a tiempo completo, año completo, de entre 25 y 34 años, con título de educación secundaria o B.A./B.S.⁶ como grado de educación alcanzado más alto. Se ofrece una descripción detallada en el archivo **CPS92_08_Description**, disponible en la página web. Utilice estos datos para responder a las siguientes preguntas.
- Calcule la media muestral de los ingresos medios por hora (AHE) en 1992 y en 2008. Construya un intervalo de confianza al 95 % para las medias poblacionales de AHE en 1992 y 2008, y para la variación entre 1992 y 2008.
 - En 2008, el valor del Índice de Precios al Consumidor (CPI) fue de 215,2. En 1992, el valor de CPI fue 140,3. Repita (a) pero utilizando los AHE medidos en dólares constantes de 2008 (\$ 2008), es decir, ajustando los datos de 1992 por la inflación de los precios que se produjo entre 1992 y 2008.
 - Si se estuviera interesado en la variación del poder adquisitivo de los trabajadores de 1992 a 2008, se utilizarían los resultados de (a) o de (b)? Explíquelo.
 - Utilice los datos de 2008 para construir un intervalo de confianza al 95 % para la media de AHE de los graduados en secundaria. Construya un intervalo de confianza al 95 % para la media de los AHE de los trabajadores con título universitario. Construya un intervalo de confianza al 95 % para la diferencia entre las dos medias.
 - Repita (d) utilizando los datos de 1992 expresados en dólares de 2008.
 - ¿Crecieron los salarios reales (ajustados por inflación) de los graduados en educación secundaria de 1992 hasta 2008? Explíquelo. ¿Aumentaron los salarios reales de los graduados universitarios? ¿Aumentó la brecha entre los ingresos de los titulados universitarios y los de educación secundaria? Explíquelo, utilizando las estimaciones, los intervalos de confianza, y los estadísticos de contraste adecuados.
 - La Tabla 3.1 presenta información sobre la brecha de género para los titulados universitarios. Prepare una tabla similar para los graduados de secundaria utilizando los datos de 1992 y 2008. ¿Existe alguna diferencia notable entre los resultados para los graduados en educación secundaria y los titulados universitarios?

⁶ N. del T.: Licenciado o ingeniero.

APÉNDICE

3.1 La encuesta actualizada de población de EE.UU. (CPS)

Cada mes, la Oficina de Estadísticas Laborales del Departamento de Trabajo de EE.UU. lleva a cabo la Encuesta Actualizada de Población⁷ (CPS), que proporciona datos sobre las características de la fuerza de trabajo de la población, incluyendo el nivel de empleo, el desempleo y los ingresos salariales. Más de 50.000 hogares de EE.UU. son encuestados cada mes. La muestra está escogida mediante selección aleatoria de direcciones a partir de una base de datos de direcciones del censo decenal más reciente, ampliada con los datos de nuevas viviendas construidas después del último censo. El esquema exacto de muestreo aleatorio es bastante complicado (en primer lugar, se escogen aleatoriamente pequeñas áreas geográficas, después son seleccionadas al azar las unidades de vivienda dentro de estas áreas), los detalles pueden encontrarse en el *Manual de Estadísticas Laborales* y en la página web de la Oficina de Estadísticas Laborales (www.bls.gov).

La encuesta llevada a cabo cada mes de marzo es más detallada que el resto de los meses e incluye preguntas sobre los ingresos del año anterior. Los estadísticos de las Tablas 2.4 y 3.1 se calculan utilizando las encuestas de marzo. Los datos CPS de los ingresos de los trabajadores a tiempo completo, definidos para alguien empleado más de 35 horas a la semana durante al menos 48 semanas del año anterior.

APÉNDICE

3.2 Dos pruebas de que \bar{Y} es el estimador de mínimos cuadrados de μ_Y

Este apéndice proporciona dos pruebas —una utiliza el cálculo y la otra no— de que \bar{Y} minimiza la suma de los errores de predicción al cuadrado de la Ecuación (3.2); es decir, que \bar{Y} es el estimador de mínimos cuadrados de $E(Y)$.

Prueba con cálculo

Para minimizar la suma de los errores de predicción al cuadrado, se calcula su derivada y se iguala a cero:

$$\frac{d}{dm} \sum_{i=1}^n (Y_i - m)^2 = -2 \sum_{i=1}^n (Y_i - m) = -2 \sum_{i=1}^n Y_i + 2nm = 0. \quad (3.27)$$

Resolviendo la última ecuación para m se demuestra que $\sum_{i=1}^n (Y_i - m)^2$ se minimiza cuando $m = \bar{Y}$.

Prueba sin cálculo

La estrategia consiste en demostrar que la diferencia entre el estimador de mínimos cuadrados y \bar{Y} debe ser cero, de lo que se deduce que es el estimador de mínimos cuadrados. Sea $d = \bar{Y} - m$, por lo que $m = \bar{Y} - d$. Entonces $(Y_i - m)^2 = (Y_i - [\bar{Y} - d])^2 = ([Y_i - \bar{Y}] + d)^2 = (Y_i - \bar{Y})^2 + 2d(Y_i - \bar{Y}) + d^2$. Por tanto, la suma de los errores de predicción al cuadrado [Ecuación (3.2)] es

$$\sum_{i=1}^n (Y_i - m)^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 + 2d \sum_{i=1}^n (Y_i - \bar{Y}) + nd^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 + nd^2, \quad (3.28)$$

donde la segunda igualdad utiliza el hecho de que $\sum_{i=1}^n (Y_i - \bar{Y}) = 0$. Debido a que los dos términos de la última línea de la Ecuación (3.28) son no negativos y debido a que el primer término no depende de d , $\sum_{i=1}^n (Y_i - m)^2$ se minimiza con una d que hace el segundo término, nd^2 , lo más pequeño posible. Esto se consigue haciendo $d = 0$, por lo que \bar{Y} es el estimador de mínimos cuadrados de $E(Y)$.

⁷ N. del T.: Current Population Survey en su denominación original.

APÉNDICE

3.3 Una prueba de que la varianza muestral es consistente

Se utiliza en este apéndice la ley de los grandes números para demostrar que la varianza muestral s_Y^2 es un estimador consistente de la varianza de la población σ_Y^2 , tal y como se indica en la Ecuación (3.9), si Y_1, \dots, Y_n son i.i.d., y $E(Y_i^4) < \infty$.

En primer lugar, se suma y se resta μ_Y para escribir $(Y_i - \bar{Y})^2 = [(Y_i - \mu_Y) - (\bar{Y} - \mu_Y)]^2 = (Y_i - \mu_Y)^2 - 2(Y_i - \mu_Y)(\bar{Y} - \mu_Y) + (\bar{Y} - \mu_Y)^2$. Sustituyendo esta expresión por $(Y_i - \bar{Y})^2$ en la definición de s_Y^2 [Ecuación (3.7)], tenemos que

$$\begin{aligned} s_Y^2 &= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu_Y)^2 - \frac{2}{n-1} \sum_{i=1}^n (Y_i - \mu_Y)(\bar{Y} - \mu_Y) + \frac{1}{n-1} \sum_{i=1}^n (\bar{Y} - \mu_Y)^2 \\ &= \left(\frac{n}{n-1} \right) \left[\frac{1}{n} \sum_{i=1}^n (Y_i - \mu_Y)^2 \right] - \left(\frac{n}{n-1} \right) (\bar{Y} - \mu_Y)^2 \end{aligned} \quad (3.29)$$

donde la última igualdad se deduce a partir de la definición de \bar{Y} [lo que implica que $\sum_{i=1}^n (Y_i - \mu_Y) = n(\bar{Y} - \mu_Y)$] y reagrupando términos.

La ley de los grandes números puede aplicarse ahora a los dos términos de la última línea de la Ecuación (3.29). Se define $W_i = (Y_i - \mu_Y)^2$. Ahora $E(W_i) = \sigma_Y^2$ (por la definición de la varianza). Debido a que las variables aleatorias Y_1, \dots, Y_n son i.i.d., las variables aleatorias W_1, \dots, W_n son i.i.d. Además, $E(W_i^2) = E[(Y_i - \mu_Y)^4] < \infty$ debido a que, por hipótesis, $E(Y_i^4) < \infty$. Por tanto W_1, \dots, W_n son i.i.d. y $\text{var}(W_i) < \infty$, por lo que \bar{W} cumple las condiciones de la ley de los grandes números del Concepto clave 2.6 y $\bar{W} \xrightarrow{P} E(W_i)$. Pero $\bar{W} = (1/n)\sum_{i=1}^n (Y_i - \mu_Y)^2$ y $E(W_i) = \sigma_Y^2$, por lo que $(1/n)\sum_{i=1}^n (Y_i - \mu_Y)^2 \xrightarrow{P} \sigma_Y^2$. Asimismo, $n/(n-1) \longrightarrow 1$, por lo que el primer término de la Ecuación (3.29) converge en probabilidad a σ_Y^2 . Debido a que $\bar{Y} \xrightarrow{P} \mu_Y$, $(\bar{Y} - \mu_Y)^2 \xrightarrow{P} 0$, por lo que el segundo término converge en probabilidad a cero. Combinando estos resultados se obtiene $s_Y^2 \xrightarrow{P} \sigma_Y^2$.

4 Regresión lineal con regresor único

Un estado impone penas nuevas más duras a los conductores ebrios: ¿cuál es el efecto sobre la mortalidad en accidentes de tráfico? Un distrito escolar reduce el tamaño de sus clases de educación primaria: ¿cuál es su efecto sobre las calificaciones de sus estudiantes en los exámenes estandarizados? Si se completa satisfactoriamente un año adicional de clases en la universidad: ¿cuál es el efecto sobre los futuros ingresos salariales?

Estas tres preguntas versan sobre los efectos desconocidos del cambio en una variable, X (siendo X las sanciones por conducir ebrio, el tamaño de la clase, o los años de escolaridad), sobre otra variable, Y (siendo Y la mortalidad en las carreteras, las calificaciones de los estudiantes, o los ingresos salariales).

En este capítulo se introduce el modelo de regresión lineal que relaciona una variable, X , con otra, Y . Este modelo plantea una relación lineal entre X e Y ; la pendiente de la recta que relaciona X con Y es el efecto de la variación en una unidad de X sobre Y . Al igual que la media de Y es una característica desconocida de la distribución poblacional de Y , la pendiente de la recta que relaciona X e Y es una característica desconocida de la distribución poblacional conjunta de X e Y . El problema estadístico consiste en estimar esta pendiente, es decir, estimar el efecto sobre Y de una variación unitaria en X , utilizando una muestra de datos de estas dos variables.

En este capítulo se describen los métodos de estimación de la pendiente que utilizan una muestra aleatoria de datos de X e Y . Por ejemplo, utilizando los datos del número de alumnos por clase y las calificaciones en los exámenes de los diferentes distritos escolares, se muestra la forma de estimar el efecto esperado sobre los resultados en el examen de la reducción del tamaño de las clases en, pongamos por caso, un estudiante por clase. La pendiente y la constante de la recta que relaciona X con Y pueden estimarse mediante un método denominado mínimos cuadrados ordinarios (MCO).

4.1 El modelo de regresión lineal

La directora de un distrito de escuela primaria debe decidir si contrata más maestros y quiere consejo. Si contrata los maestros, reducirá en dos el número de alumnos por maestro (la ratio estudiantes-maestros). Se enfrenta a un dilema. Los padres quieren clases más pequeñas para que sus hijos puedan recibir una atención más individualizada. Sin embargo, la contratación de más maestros significa gastar más dinero, ¡lo que no es del agrado de los que pagan la cuenta! Por lo tanto ella pregunta: si recorta el tamaño de las clases, ¿cuál será el efecto sobre el rendimiento escolar?

En muchos distritos escolares, el rendimiento escolar se mide mediante exámenes estandarizados, y el estatus laboral o la retribución de algunos administradores pueden depender en parte del desempeño de los estudiantes en esos exámenes. Por lo tanto, afinamos la pregunta de la directora: si se reduce el tamaño medio de las clases en dos estudiantes, ¿cuál será el efecto en su distrito sobre las calificaciones obtenidas en los exámenes estandarizados?

Una respuesta precisa a esta pregunta exige una cuantificación de las variaciones. Si la directora varía el número de alumnos por clase en una cierta cantidad, ¿qué variación esperaría que sucediese sobre las puntuaciones de los exámenes? Podemos expresar esto en forma de relación matemática mediante la letra

griega $\beta_{TamañoClase}$ donde el subíndice *TamañoClase* distingue el efecto de la variación en el tamaño de las clases de otros posibles efectos. Por tanto,

$$\beta_{TamañoClase} = \frac{\text{Variación CalificaciónExamen}}{\text{Variación TamañoClase}} = \frac{\Delta\text{CalificaciónExamen}}{\Delta\text{TamañoClase}} \quad (4.1)$$

donde la letra griega Δ (delta) significa «variación de». Es decir, $\beta_{TamañoClase}$ es la variación en la calificación en el examen que provoca la variación en el tamaño de la clase dividido por la variación en el tamaño de las clases.

Si se tuviera la suficiente fortuna de conocer $\beta_{TamañoClase}$, sería posible decir a la directora que una disminución del tamaño de la clase en un alumno cambiaría las calificaciones de los exámenes de todo el distrito en $\beta_{TamañoClase}$. Además se podría responder a la pregunta real de la directora, que se refiere a la variación del tamaño en dos estudiantes por clase. Para ello, se reorganiza la Ecuación (4.1) por lo que

$$\Delta\text{CalificaciónExamen} = \beta_{TamañoClase} \times \Delta\text{TamañoClase}. \quad (4.2)$$

Supongamos que $\beta_{TamañoClase} = -0,6$. Entonces una reducción del tamaño de las clases en dos alumnos por clase daría lugar a una variación en las calificaciones esperadas de $(-0,6) \times (-2) = 1,2$; es decir, se podría esperar que las calificaciones en los exámenes se incrementaran en 1,2 puntos como consecuencia de la reducción del tamaño de las clases en dos estudiantes por clase.

La Ecuación (4.1) es la definición de la pendiente de una línea recta que relaciona las calificaciones y el tamaño de las clases. Esta recta puede escribirse

$$\text{CalificaciónExamen} = \beta_0 + \beta_{TamañoClase} \times \text{TamañoClase}, \quad (4.3)$$

donde β_0 es la pendiente o intercepto de esta recta y, como antes, $\beta_{TamañoClase}$ es la pendiente. De acuerdo con la Ecuación (4.3), si conocieran β_0 y $\beta_{TamañoClase}$, no solo sería posible determinar la variación asociada a una variación en el tamaño de las clases sobre las calificaciones en el examen para un distrito, sino que además sería posible predecir la calificación media en el examen dado el tamaño de la clase.

Cuando se le propuso a la directora la Ecuación (4.3), ella dijo que algo estaba mal en su formulación. Señalaba que el tamaño de la clase es tan solo uno de tantos aspectos de la educación primaria y que dos distritos con el mismo tamaño de clases podrían tener resultados diferentes en las pruebas debido a muchas razones. Un distrito podría tener mejores maestros o podría utilizar libros de texto mejores. Dos distritos con tamaños de clase, maestros y libros de texto comparables, podrían sin embargo presentar poblaciones de estudiantes muy diferentes; quizás uno de los distritos tenga más inmigrantes (y por tanto menos hablantes nativos de inglés) o familias más ricas. Por último, señalaba que incluso si dos distritos son iguales en todos estos aspectos podrían tener diferentes resultados en los exámenes por motivos esencialmente aleatorios que tienen que ver con el desempeño de cada alumno en el día del examen. Ella tiene razón, por supuesto; por todas estas razones, la Ecuación (4.3) no se cumple con exactitud para todos los distritos. Debe interpretarse en cambio como una definición sobre una relación que se cumple *en promedio* entre la población de los distritos.

Una versión de esta relación lineal que se cumpliera para *cada* distrito debería incorporar esos otros factores que influyen en las calificaciones de los exámenes, incluyendo las características únicas de cada uno de los distritos (por ejemplo, la calidad de sus maestros, las características de sus alumnos, y la fortuna de los estudiantes el día del examen). Un método podría ser hacer una lista con los factores más importantes y posteriormente introducirlos en la Ecuación (4.3) de forma explícita (una idea que retomaremos en el Capítulo 6). Por ahora, sin embargo, simplemente recopilamos todos estos «otros factores» de forma conjunta y expresamos la relación para un distrito determinado como

$$\text{CalificaciónExamen} = \beta_0 + \beta_{TamañoClase} \times \text{TamañoClase} + \text{otros factores}. \quad (4.4)$$

Por tanto, la calificación en el examen para el distrito está expresada en términos de una componente, $\beta_0 + \beta_{TamañoClase} \times \text{TamañoClase}$, que representa el efecto promedio del tamaño de las clases sobre las cali-

ficaciones en la población de distritos escolares y una segunda componente que representa todos los demás factores.

Aunque este análisis se ha centrado en las calificaciones obtenidas y el tamaño de las clases, la idea expresada en la Ecuación (4.4) es mucho más general, por lo que resulta útil para introducir una notación más general. Supongamos que se dispone de una muestra de n distritos. Sea Y_i la media de las calificaciones en los exámenes del distrito i -ésimo, sea X_i el tamaño medio de las clases en el distrito i -ésimo, y sea u_i la expresión de los otros factores que influyen en las calificaciones de los exámenes del distrito i -ésimo. Entonces la Ecuación (4.4) puede escribirse de manera más general como

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad (4.5)$$

para cada distrito (es decir, $i = 1, \dots, n$), donde β_0 es el término constante (intercepto) de esta recta y β_1 es la pendiente. [En notación general, se utiliza β_1 para la pendiente de la ecuación (4.5) en lugar de $\beta_{TamañoClase}$ debido a que esta ecuación está expresada en términos de una variable general X_i].

La ecuación (4.5) es el **modelo de regresión lineal con regresor único**, en el que Y es la **variable dependiente** y X es la **variable independiente o regresor**.

La primera parte de la Ecuación (4.5), $\beta_0 + \beta_1 X_i$ es la **recta de regresión poblacional o función de regresión poblacional**. Esta es la relación entre Y y X que se cumple en promedio para la población. Por tanto, si se conociera el valor de X , de acuerdo con esta recta de regresión poblacional se podría predecir que el valor de la variable dependiente, Y , es $\beta_0 + \beta_1 X$.

El **término constante o independiente o valor en el origen o intercepto**, β_0 y la **pendiente** β_1 son los **coeficientes** de la recta de regresión poblacional, conocidos asimismo como **parámetros** de la recta de regresión poblacional. La pendiente β_1 es la variación en Y asociada a un cambio unitario en X . El intercepto es el valor de la recta de regresión poblacional cuando $X = 0$, es el punto en el que la recta de regresión corta el eje Y . En algunas aplicaciones estadísticas, el intercepto tiene una interpretación económica significativa. En otras aplicaciones, el intercepto no tiene ningún significado en el mundo real; por ejemplo, cuando X es el tamaño de las clases, en sentido estricto el término independiente es el valor esperado de las calificaciones en los exámenes cuando en la clase ¡no hay estudiantes! Cuando el significado en el mundo real del término independiente no tiene sentido, es mejor pensar en su sentido matemático como el coeficiente que determina el nivel de la recta de regresión.

El término u_i de la Ecuación (4.5) es el **término de error**. El término de error recoge todos los factores responsables de la diferencia entre la calificación media en los exámenes del distrito i -ésimo y el valor predicho por la recta de regresión poblacional. Este término de error contiene todos los otros factores que además de X determinan el valor de la variable dependiente, Y , para una observación concreta, i . En el ejemplo del tamaño de las clases, estos otros factores incluyen todas las características específicas del distrito i -ésimo que afectan al rendimiento de sus estudiantes en el examen, incluyendo la calidad del profesorado, las características económicas del estudiante, la suerte, e incluso los errores en la corrección de la prueba.

El modelo de regresión lineal y su terminología se resumen en el Concepto clave 4.1.

La Figura 4.1 recoge el modelo de regresión lineal con un único regresor para siete observaciones hipotéticas de las calificaciones obtenidas en el examen (Y) y el tamaño de las clases (X). La recta de regresión poblacional es la línea recta $\beta_0 + \beta_1 X$. La recta de regresión poblacional tiene pendiente negativa ($\beta_1 < 0$), lo que significa que los distritos con menores ratios de estudiantes por maestro (clases más pequeñas) tienden a tener mejores calificaciones en los exámenes. El término independiente β_0 tiene sentido matemático como el valor en el eje Y del punto de intersección del eje Y y la recta de regresión poblacional, pero, como se mencionó anteriormente, no tiene ningún significado en el mundo real, en este caso.

Debido a los otros factores que determinan el rendimiento en los exámenes, las observaciones hipotéticas de la Figura 4.1 no se encuentran exactamente en la recta de regresión de la población. Por ejemplo, el valor de Y para el distrito #1, Y_1 está por encima de la recta de regresión poblacional. Esto significa que los resultados en las pruebas para el distrito #1 fueron mejores de lo previsto por la recta de regresión poblacional, por lo que el término de error para este distrito, u_1 , es positivo. Por el contrario, Y_2 está por debajo de la recta de regresión poblacional, por lo que las calificaciones en los exámenes de este distrito fueron peores de lo previsto, y $u_2 < 0$.

**CONCEPTO
CLAVE**
4.1**Terminología del modelo de regresión lineal con regresor único**

El modelo de regresión lineal es

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

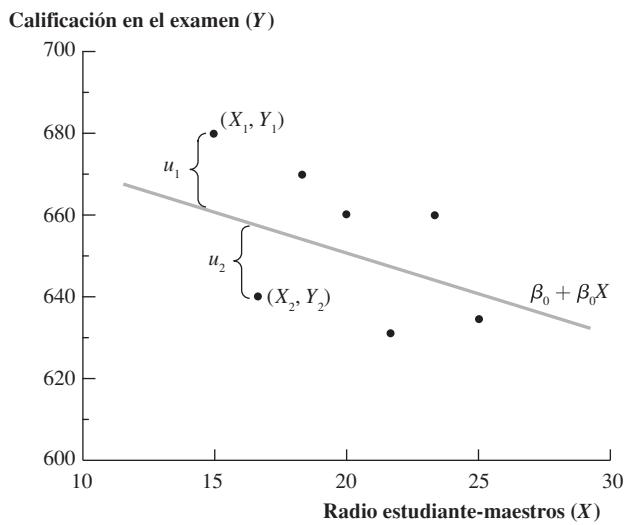
donde

- el subíndice i recorre las observaciones, $i = 1, \dots, n$;
- Y_i es la *variable dependiente*, el *regresando*, o simplemente la *variable de la parte izquierda*;
- X_i es la *variable independiente*, el *regresor*, o simplemente la *variable de la parte derecha*;
- $\beta_0 + \beta_1 X$ es la *recta de regresión poblacional* o *función de regresión poblacional*;
- β_0 es el *intercepto* de la recta de regresión poblacional;
- β_1 es la *pendiente* de la recta;
- u_i es el *término de error*.

FIGURA 4.1

Diagrama de dispersión de las calificaciones en el examen vs. ratio estudiante-maestros (datos hipotéticos)

El diagrama muestra observaciones hipotéticas sobre siete distritos escolares. La recta de regresión poblacional es $\beta_0 + \beta_1 X$. La distancia vertical desde el punto i -ésimo a la recta de regresión poblacional es $Y_i - (\beta_0 + \beta_1 X_i)$, que es el término de error poblacional u_i para la i -ésima observación.



Ahora volvemos al problema de asesorar a la directora: ¿cuál es el efecto esperado sobre resultados en los exámenes de la reducción de la ratio estudiante-maestros en dos estudiantes por maestro? La respuesta es sencilla: el cambio esperado es $(-2) \times \beta_{TamañoClase}$. Pero ¿cuál es el valor de $\beta_{TamañoClase}$?

4.2 Estimación de los coeficientes del modelo de regresión lineal

En una situación práctica, como es el caso del tamaño de las clases y las calificaciones en los exámenes, el término independiente β_0 y la pendiente β_1 de la recta de regresión poblacional son desconocidos. Por tanto, para estimar la pendiente y el término independiente poblacionales desconocidos de la recta de regresión deben utilizarse datos.

Este problema de estimación es similar a otros a los que se enfrenta la estadística. Por ejemplo, supongamos que se desea comparar la media de los ingresos salariales de hombres y mujeres que se han graduado recientemente en la universidad. Aunque el ingreso salarial promedio de la población es desconocido, se

pueden estimar las medias poblacionales utilizando una muestra aleatoria de hombres y mujeres titulados universitarios. Entonces, por ejemplo, el estimador natural de los ingresos medios poblacionales desconocidos para las mujeres, es el ingreso promedio de las mujeres graduadas universitarias de la muestra.

La misma idea es extensible para el modelo de regresión lineal. Se desconoce el valor poblacional de $\beta_{TamañoClase}$, la pendiente de la recta de regresión poblacional desconocida que relaciona X (tamaño de la clase) con Y (calificaciones en los exámenes). Pero así como es posible conocer la media poblacional mediante una muestra de datos extraídos de esa población, es posible conocer la pendiente poblacional de $\beta_{TamañoClase}$, utilizando una muestra de datos.

Los datos que se analizan aquí consisten en calificaciones en los exámenes y tamaño de clases en el año 1999 de 420 distritos escolares de California que proporcionan educación desde preescolar hasta el octavo curso. La calificación del examen es el promedio para todo el distrito de las puntuaciones en lectura y matemáticas de los alumnos de quinto curso. El tamaño de las clases se puede medir de varias formas. El que aquí se utiliza es una de las más generales, el número total de estudiantes del distrito dividido por el número de maestros; es decir, la ratio estudiantes-maestros en el global del distrito. Estos datos se describen con más detalle en el Apéndice 4.1.

La Tabla 4.1 resume las distribuciones de las calificaciones en los exámenes y los tamaños de las clases para esta muestra. El promedio de la ratio estudiantes por maestro es de 19,6 alumnos por maestro, y la desviación típica es de 1,9 alumnos por maestro. El percentil 10 de la distribución de la ratio estudiantes-maestros es de 17,3 (es decir, solo el 10 % de los distritos presentan una ratio estudiantes-maestros por debajo de 17,3), mientras que el distrito en el percentil 90 tiene una ratio estudiantes-maestros de 21,9.

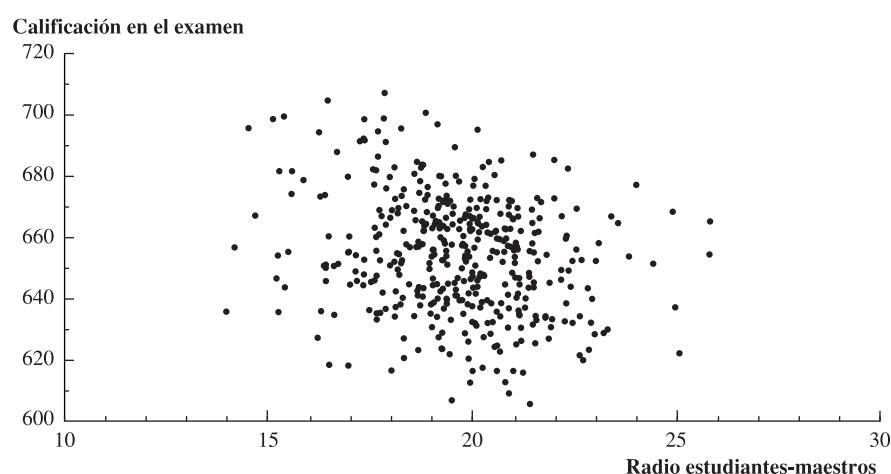
Se muestra un diagrama de dispersión de las 420 observaciones sobre las calificaciones obtenidas y ratios estudiantes-maestros en la Figura 4.2. La correlación muestral es $-0,23$, lo que indica una relación

TABLA 4.1 Resumen de la distribución de las ratios estudiantes-maestros y calificaciones en los exámenes de quinto curso de 420 distritos escolares de California en 1999

	Promedio	Desviación típica	Porcentaje						
			10 %	25 %	40 %	50 % (mediana)	60 %	75 %	90 %
Ratio estudiantes-maestros	19,6	1,9	17,3	18,6	19,3	19,7	20,1	20,9	21,9
Calificación examen	654,2	19,1	630,4	640,0	649,1	654,5	659,4	666,7	679,1

FIGURA 4.2 Diagrama de dispersión de las calificaciones en los exámenes vs. ratio estudiantes-maestros (datos de los distritos escolares de California)

Datos sobre 420 distritos escolares de California. Existe un relación negativa entre la ratio estudiantes-maestros y la calificación en el examen: la correlación muestral es $-0,23$.



negativa débil entre las dos variables. Aunque las clases más grandes de esta muestra tienden a obtener calificaciones más bajas en los exámenes, existen otros factores determinantes de los resultados de las pruebas que explican que las observaciones no se encuentren colocadas exactamente sobre una línea recta.

A pesar de esta baja correlación, si se pudiera trazar de alguna forma una línea recta a través de estos datos, la pendiente de esa recta sería una estimación de $\beta_{TamañoClase}$ basada en esos datos. Una forma de trazar la recta sería con un lápiz y una regla y «a ojo» hacer la mejor recta que se pudiera. Si bien este método es fácil, es poco científico, además de que distintas personas trazarían diferentes rectas estimadas.

Entonces, ¿cómo debe elegirse entre las muchas posibles rectas? Con mucho, el modo más común es el de elegir la recta que proporciona el ajuste de «mínimos cuadrados» a esos datos, es decir, mediante el estimador de mínimos cuadrados ordinarios (MCO).

El estimador de mínimos cuadrados ordinarios

El estimador MCO elige los coeficientes de regresión de tal forma que la recta de regresión estimada se encuentre lo más cercana posible a los datos observados, y la cercanía está medida por la suma de los errores al cuadrado que se cometan con la predicción de Y dado X .

Tal y como se discutió en la Sección 3.1, la media muestral, \bar{Y} , es el estimador de mínimos cuadrados de la media poblacional, $E(Y)$; es decir, \bar{Y} minimiza la suma total de los errores de estimación al cuadrado $\sum_{i=1}^n (Y_i - m)^2$ de entre todos los posibles estimadores m [véase la Expresión (3.2)].

El estimador MCO extiende esta idea al modelo de regresión lineal. Sean b_0 y b_1 algunos de los estimadores de β_0 y β_1 . La recta de regresión basada en esos estimadores es $b_0 + b_1X$ por lo que el valor de Y_i previsto mediante esta recta es $b_0 + b_1X_i$. Por tanto, el error cometido en la predicción de la observación i -ésima es $Y_i - (b_0 + b_1X_i) = Y_i - b_0 - b_1X_i$. La suma de estos errores de predicción al cuadrado para las n observaciones es

$$\sum_{i=1}^n (Y_i - b_0 - b_1X_i)^2. \quad (4.6)$$

La suma de los errores al cuadrado para el modelo de regresión lineal de la Expresión (4.6) es la extensión de la suma de los errores al cuadrado del problema de la estimación de la media en la Expresión (3.2). De hecho, si no hay regresores, b_1 no entra en la Expresión (4.6) y los dos problemas son idénticos, excepto en que la notación es diferente [m en la Expresión (3.2), b_0 en la Expresión (4.6)]. Así como existe un único estimador, \bar{Y} , que minimiza la Expresión (3.2), existe un único par de estimadores de β_0 y β_1 que minimizan la Expresión (4.6).

Los estimadores del término constante y de la pendiente que minimizan la suma de los cuadrados de los errores de la Expresión (4.6) se denominan **estimadores de mínimos cuadrados ordinarios (MCO)** de β_0 y β_1 .

MCO tiene su propia notación especial y terminología. El estimador MCO de β_0 se expresa por $\hat{\beta}_0$, y el estimador de β_1 se expresa como $\hat{\beta}_1$. La **recta de regresión MCO**, asimismo denominada **recta de regresión muestral o función de regresión muestral**, es la línea recta construida utilizando los estimadores MCO: $\hat{\beta}_0 + \hat{\beta}_1X$. El **valor de predicción** de Y_i dado X , está basado en la recta de regresión MCO. El **residuo** de la observación i -ésima es la diferencia entre Y_i y su valor de predicción: $\hat{u} = Y_i - \hat{Y}_i$.

Los estimadores MCO, $\hat{\beta}_0$ y $\hat{\beta}_1$ son los homólogos muestrales de los coeficientes poblacionales β_0 y β_1 . Del mismo modo, la recta de regresión MCO $\hat{\beta}_0 + \hat{\beta}_1X$ es la homóloga muestral de la recta de regresión poblacional $\beta_0 + \beta_1X$, y los residuos MCO \hat{u}_i son los homólogos muestrales de los errores poblacionales u_i .

Se podrían calcular los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ probando diferentes valores de b_0 y b_1 reiteradamente hasta encontrar los que minimizan la suma total de los errores al cuadrado de la Expresión (4.6), son las estimaciones mínimo cuadráticas. No obstante, este método podría resultar bastante tedioso. Afortunadamente, existen fórmulas, obtenidas de la minimización de la Expresión (4.6) mediante la utilización de cálculo, que simplifican la obtención de los estimadores MCO.

Las fórmulas MCO, así como la terminología, se recogen en el Concepto clave 4.2. Esas fórmulas se aplican en casi todos los programas estadísticos y hojas de cálculo. Estas fórmulas se obtienen en el Apéndice 4.2.

**CONCEPTO
CLAVE**

4.2

El estimador MCO, valores estimados y residuos

Los estimadores MCO de la pendiente β_1 y del término independiente (intercepto) β_0 , son

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2} \quad (4.7)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (4.8)$$

El valor de predicción MCO \hat{Y}_i y los residuos \hat{u}_i son

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, i = 1, \dots, n \quad (4.9)$$

$$\hat{u}_i = Y_i - \hat{Y}_i, i = 1, \dots, n. \quad (4.10)$$

El término independiente ($\hat{\beta}_0$), la pendiente ($\hat{\beta}_1$) y los residuos (\hat{u}_i) estimados, se calculan a partir de la muestra de n observaciones de X_i e Y_i , $i = 1, \dots, n$. Son estimadores de los verdaderos término independiente (β_0), pendiente (β_1) y residuos (u_i) poblacionales desconocidos.

Estimaciones MCO de la relación entre calificaciones en los exámenes y ratio estudiantes-maestros

Cuando se utiliza MCO para estimar una recta que relacione la ratio estudiantes-maestros con las calificaciones obtenidas en los exámenes a partir de las 420 observaciones de la Figura 4.2, la pendiente estimada es $-2,28$ y el término independiente estimado es $698,9$. Por tanto, la recta de regresión MCO para esas 420 observaciones es

$$\widehat{\text{CalificaciónExamen}} = 698,9 - 2,28 \times \text{REM}, \quad (4.11)$$

donde $\text{CalificaciónExamen}$ es la calificación media en el examen del distrito y REM es la ratio estudiantes-maestros. El « $\widehat{}$ » sobre la variable $\text{CalificaciónExamen}$ en la Ecuación (4.11) indica que es el valor esperado de acuerdo a la recta de regresión MCO. La Figura 4.3 representa esta recta de regresión MCO superpuesta sobre el diagrama de dispersión de los datos mostrados anteriormente en la Figura 4.2.

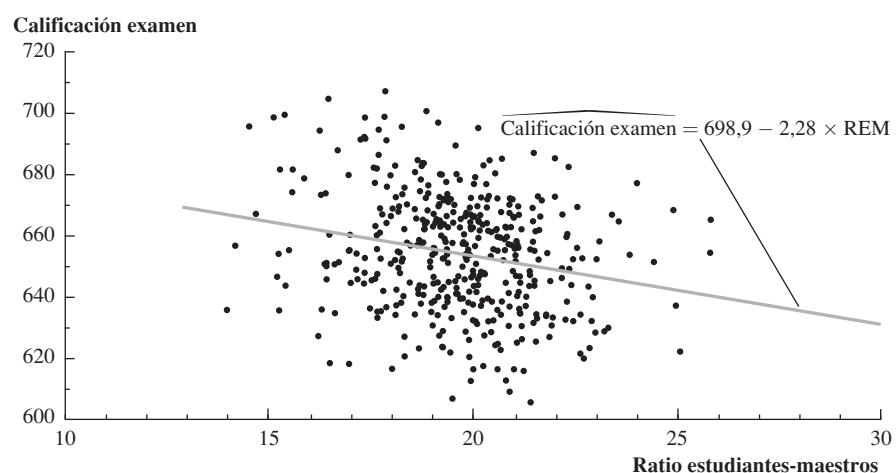
La pendiente de $-2,28$ significa que un aumento en la proporción estudiantes-maestros en un estudiante por clase está, en promedio, asociado a una disminución en las calificaciones del distrito en el examen de $2,28$ puntos. Una disminución de la ratio estudiantes-maestros en dos estudiantes por clase está, en promedio, asociada a un aumento en las calificaciones de $4,56$ puntos [$= -2 \times (-2,28)$]. La pendiente negativa indica que cuantos más estudiantes por maestro (clases más grandes), peor rendimiento en el examen.

Ahora es posible predecir la calificación en el examen para un distrito dado un valor de la ratio estudiantes-maestros. Por ejemplo, para un distrito con 20 estudiantes por maestro, la calificación en la prueba es $698,9 - 2,28 \times 20 = 653,3$. Por supuesto, esta predicción no será cierta de forma exacta a causa de los otros factores que determinan el rendimiento de un distrito. Sin embargo, la recta de regresión proporciona una predicción (la predicción MCO) de cuál será el resultado en los exámenes de este distrito, en base a su ratio estudiantes-maestros, omitiendo esos otros factores.

¿Es esta estimación de la pendiente grande o pequeña? Para responder a esto, volvemos al problema de la directora. Hay que recordar que se está contemplando la posibilidad de contratación de maestros suficientes para permitir una reducción de la proporción estudiantes-maestros en 2. Supongamos que el distrito de la directora se encuentra en la mediana de los distritos de California. A partir de la Tabla 4.1, la mediana de la ratio estudiantes por maestro es de 19,7 y la mediana de las calificaciones en los exámenes

FIGURA 4.3 La Recta de regresión para los datos de California

La recta de regresión estimada muestra una relación negativa entre las calificaciones en el examen y la ratio estudiantes-maestros. Si el tamaño de las clases disminuye en un estudiante, la regresión estimada predice que las calificaciones en el examen aumentarán en 2,28 puntos.



es 654,5. Una reducción de dos alumnos por clase, de 19,7 a 17,7, trasladaría esta ratio estudiantes-maestros del percentil 50 a muy cerca del percentil 10. Este es un gran cambio, y ella tendría que contratar a muchos maestros. ¿Cómo afectaría esto a los resultados en los exámenes?

De acuerdo con la Ecuación (4.11), se prevé que la reducción del número de estudiantes por maestro en 2 incremente las calificaciones en los exámenes en aproximadamente 4,6 puntos; si las calificaciones en los exámenes de su distrito están en la mediana, 654,5, se preverá un aumento hasta 659,1. ¿Es este aumento grande o pequeño? De acuerdo con la Tabla 4.1, este incremento trasladaría a su distrito desde la mediana hasta estar cercano al percentil 60. Por tanto una disminución en el tamaño de las clases que colocara a su distrito cerca del 10 % con las clases más pequeñas trasladaría las clasificaciones en los exámenes desde el percentil 50 al percentil 60. De acuerdo con las estimaciones, al menos, el recorte en la ratio estudiantes-maestros en una gran cuantía (dos estudiantes por maestro) ayudaría y merecería la pena hacerlo en función de su situación presupuestaria, pero no sería una panacea.

¿Qué sucedería si la directora estuviera contemplando un cambio mucho más radical, tal como la reducción de la ratio estudiantes-maestros en 20 alumnos por maestro hasta 5? Desafortunadamente, las estimaciones de la Ecuación (4.11) no resultarían muy útiles para ella. Esta regresión se estimó con los datos de la Figura 4.2, y, como muestra la figura, la ratio estudiantes-maestros más pequeña en estos datos es 14. Estos datos no contienen información acerca de cómo se comportan los distritos con tamaños de clases muy pequeños, por lo que estos datos por sí solos no constituyen una base fiable para predecir el efecto de un cambio radical hasta ese extremo de baja ratio estudiantes-maestros.

¿Por qué utilizar el estimador MCO?

Existen razones tanto teóricas como prácticas para la utilización de los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$. Debido a que MCO es el método dominante utilizado en la práctica, se ha convertido en el idioma común del análisis de regresión para la economía, las finanzas (véase el recuadro «El ‘Beta’ de una acción»), y las ciencias sociales en general. Presentar resultados mediante MCO (o sus variantes estudiadas más adelante en este libro) significa que se está «hablando el mismo idioma» que otros economistas y estadísticos. Las fórmulas MCO están integradas en prácticamente todas las hojas de cálculo y paquetes estadísticos, por lo que MCO es fácil de utilizar.

Los estimadores MCO presentan asimismo propiedades teóricas deseables. Son análogas a las propiedades deseables, estudiadas en la Sección 3.1, de \bar{Y} como estimador de la media poblacional. Bajo los supuestos introducidos en la Sección 4.4, el estimador MCO es insesgado y consistente. El estimador MCO es asimismo eficiente dentro de un cierto tipo de estimadores insesgados; sin embargo, este resultado de eficiencia tiene una limitación importante.

El «beta» de una acción

Una idea fundamental en las finanzas modernas es que un inversor necesita un incentivo financiero para asumir un riesgo. Dicho de otro modo, la rentabilidad esperada¹ de una inversión arriesgada, R , debe ser mayor a la rentabilidad de una inversión segura o libre de riesgo, R_f . Por tanto, la prima de riesgo, $R - R_f$, en una inversión arriesgada, como la tenencia de acciones de una empresa, debe ser positiva.

En principio podría parecer que el riesgo de una acción debería medirse por su varianza. La mayor parte de ese riesgo, sin embargo, puede reducirse mediante la tenencia de otras acciones en una «cartera» —en otras palabras, diversificando la tenencia de activos. Esto significa que la forma correcta de medir el riesgo de una acción no es mediante su *varianza*, sino más bien mediante su *covarianza* con el mercado.

El Modelo de Valoración de Activos Financieros (Capital Asset Pricing Model, CAPM) formaliza esta idea. De acuerdo con el CAPM, la prima de riesgo esperada de un activo es proporcional a la prima de riesgo esperada de una cartera formada por todos los activos disponibles (la «cartera o portafolio de mercado»). Es decir, el CAPM dice que

$$R - R_f = \beta(R_m - R_f) \quad (4.12)$$

donde R_m es la rentabilidad esperada de la cartera de mercado y β es el coeficiente de la regresión poblacional de $R - R_f$ sobre $R_m - R_f$. En la práctica, la rentabilidad libre de riesgo a menudo se considera como el tipo de interés de la deuda pública a corto plazo de EE.UU. De acuerdo con el CAPM, una acción con un $\beta < 1$ tiene un riesgo menor que el de la cartera de mercado y por lo tanto tiene una menor prima esperada de riesgo que la cartera de mercado. Por el contrario, una acción

con un $\beta > 1$ es más arriesgada que la cartera de mercado y por lo tanto presenta una mayor prima de riesgo esperada.

La «beta» de una acción se ha convertido en el caballo de batalla de la industria de la inversión, y se pueden obtener estimaciones de las betas para cientos de acciones de las páginas web de las empresas de inversión. Esas betas por lo general se estiman mediante regresión MCO de la prima de riesgo efectiva de una acción sobre la prima de riesgo efectiva de un índice amplio de mercado.

La siguiente tabla presenta las betas estimadas para siete acciones de EE.UU. Los productores de productos básicos de consumo, de bajo riesgo, como Kellogg, tienen acciones con betas bajas, las acciones con mayor riesgo presentan betas altas.

Compañía	β estimado
Wal-Mart (minorista de descuento)	0,3
Kellogg (cereales para el desayuno)	0,5
Waste Management (residuos)	0,6
Verizon (telecomunicaciones)	0,6
Microsoft (software)	1,0
Best Buy (minorista de equipos electrónicos)	1,3
Bank of America (banco)	2,4

Fuente: SmartMoney.com.

¹ La rentabilidad (o retorno) de una inversión es la variación cambio de su precio más cualquier pago (dividendo) generado por la inversión en términos de porcentaje sobre su precio inicial. Por ejemplo, una acción comprada el 1 de enero por 100 \$ que paga un dividendo a lo largo del año de 2,50 \$ y vendida el 31 de diciembre por 105 \$, tendría una rentabilidad $R = [(105 \$ - 100 \$) + 2,50 \$]/100 \$ = 7,5\%$.

ciencia se mantiene bajo algunas condiciones adicionales particulares, se pospone un estudio más profundo de este resultado hasta la Sección 5.5.

4.3 Medidas de ajuste

Una vez estimada una regresión lineal, es posible preguntarse en qué medida esta regresión lineal describe correctamente los datos. ¿Recoge el regresor mucha o poca proporción de la variación de la variable dependiente? ¿Están las observaciones muy agrupadas alrededor de la recta de regresión o se encuentran dispersas?

El R^2 y el error estándar de la regresión miden la bondad del ajuste de la recta de regresión MCO a los datos. El R^2 oscila entre 0 y 1 y mide la proporción de la varianza de Y_i explicada por X_i . El error estándar de la regresión mide la distancia que habitualmente separa a Y_i de su valor esperado.

El R^2

El **R^2 de la regresión** es la proporción de la varianza muestral de Y_i explicada por (o predicha por) X_i . Las definiciones del valor esperado (o de predicción) y el residuo (véase el Concepto clave 4.2) nos permiten escribir la variable dependiente Y_i como la suma del valor de predicción \hat{Y}_i , más el residuo \hat{u}_i

$$Y_i = \hat{Y}_i + \hat{u}_i. \quad (4.13)$$

En esta notación R^2 es el cociente de la varianza muestral de \hat{Y}_i entre la varianza muestral de Y_i .

Matemáticamente, el R^2 puede escribirse como el cociente entre la suma explicada (de cuadrados) y la suma total (de cuadrados). La **suma explicada (SE)** es la suma de las desviaciones al cuadrado de los valores de predicción de Y_i , \hat{Y}_i , respecto de su media y la **suma total (ST)** es la **suma de los cuadrados de las desviaciones de Y_i** respecto de su media:

$$SE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (4.14)$$

$$ST = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (4.15)$$

La Ecuación (4.14) utiliza el hecho de que la media muestral de los valores de predicción MCO es igual a \bar{Y} (demonstrado en el Apéndice 4.3).

El R^2 es la ratio entre la suma explicada y la suma total

$$R^2 = \frac{SE}{ST} \quad (4.16)$$

Alternativamente, el R^2 se puede expresar en términos del cociente entre la varianza de Y_i no explicada por X_i . La **suma de los cuadrados de los residuos, suma residual, o SR**, es la suma de los residuos MCO al cuadrado:

$$SR = \sum_{i=1}^n \hat{u}_i^2 \quad (4.17)$$

En el Apéndice 4.3, se demuestra que $ST = SE + SR$. Por tanto, el R^2 puede expresarse asimismo como 1 menos el cociente entre la suma de los cuadrados de los residuos y la suma total:

$$R^2 = 1 - \frac{SR}{ST} \quad (4.18)$$

Por último, el R^2 de la regresión de Y sobre un único regresor X es el cuadrado del coeficiente de correlación entre Y y X .

El R^2 toma valores entre 0 y 1. Si $\hat{\beta}_1 = 0$, entonces X_i no explica nada acerca de la variación de Y_i y el valor de predicción de Y_i en base a la regresión es exactamente la media muestral de Y_i . En este caso, la suma explicada es cero y la suma residual es igual a la suma total; por lo que R^2 es cero. Por el contrario, si X_i explica toda la variación de Y_i , entonces $Y_i = \hat{Y}_i$ para todo i y todos los residuos son iguales a cero (es decir, $\hat{u}_i = 0$), de modo que $SE = ST$ y $R^2 = 1$. En general, el R^2 no toma esos valores extremos 0 o 1, sino que se encuentra en algún punto intermedio. Un R^2 cercano a 1 indica que el regresor es un buen predictor de Y_i , mientras que un R^2 cercano a 0 indica que el regresor no es un muy buen predictor de Y_i .

El error estándar de la regresión

El **error estándar de la regresión (ESR)** es un estimador de la desviación típica del error de regresión u_i . Las unidades de u_i e Y_i son las mismas, por lo que el **ESR** es una medida de la dispersión de las observaciones en torno a la recta de regresión, medida en las unidades de la variable dependiente. Por ejemplo, si las unidades de la variable dependiente son dólares, el **ESR** mide en dólares la magnitud de una desviación típica de la recta de regresión, es decir, la magnitud de un error de regresión típico.

Debido a que los errores de regresión u_1, \dots, u_n no son observables, el **ESR** se calcula mediante sus homólogos muestrales, los residuos MCO $\hat{u}_1, \dots, \hat{u}_n$. La fórmula para el **ESR** es

$$ESR = s_{\hat{u}}, \text{ donde } s_{\hat{u}}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{SR}{n-2} \quad (4.19)$$

donde la fórmula de s_u^2 se obtiene del hecho (demonstrado en el Apéndice 4.3) de que la media muestral de los residuos MCO es cero.

La fórmula del ESR en la Ecuación (4.19) es similar a la fórmula de la desviación típica muestral de Y dada en la Ecuación (3.7) en la Sección 3.2, excepto que el $Y_i - \bar{Y}$ en la Ecuación (3.7) se sustituye por \hat{u}_i y el divisor en la Ecuación (3.7) es $n - 1$, mientras que aquí es $n - 2$. La razón de utilizar aquí el divisor $n - 2$ (en lugar de n) es la misma que la razón para usar el divisor $n - 1$ en la Ecuación (3.7): corrige un ligero sesgo a la baja introducido al estimar dos coeficientes de regresión. Esto se denomina corrección por los «grados de libertad» debido a que se estiman dos coeficientes (β_0 y β_1), se pierden «dos grados de libertad», por lo que el divisor en este factor es $n - 2$. (Las matemáticas utilizadas se tratan en la Sección 5.6). Cuando n es grande, la diferencia entre dividir por n , por $n - 1$, o por $n - 2$ es insignificante.

Aplicación a los datos de las calificaciones en los exámenes

La Ecuación (4.11) proporciona la recta de regresión, que se calcula utilizando los datos de las calificaciones en los exámenes de California, que relaciona la calificación en el examen estandarizado (*Calificación-Examen*) con la ratio estudiantes-maestros (*REM*). El R^2 de esta regresión es 0,051, o el 5,1 %, y el *ESR* es 18,6.

El R^2 de 0,051 significa que el regresor *REM* explica el 5,1 % de la varianza de la variable dependiente *CalificaciónExamen*. La Figura 4.3 superpone esta recta de regresión en el diagrama de dispersión de los datos de *CalificaciónExamen* y *REM*. Como muestra el diagrama de dispersión, la ratio estudiantes-maestros explica parte de la variación en las calificaciones en los exámenes, sin embargo la mayor parte de la variación permanece en paradero desconocido.

El ESR de 18,6 significa que la desviación típica de los residuos de la regresión es 18,6, donde las unidades son puntos en el examen estandarizado. Debido a que la desviación típica es una medida de dispersión, el ESR de 18,6 significa que existe una gran dispersión en el diagrama de dispersión de la Figura 4.3 alrededor de la recta de regresión, medida en puntos obtenidos en el examen. Esta gran dispersión significa que las predicciones realizadas acerca de los resultados en las pruebas de examen para ese distrito se equivocarán frecuentemente en una gran cuantía.

¿Qué debería hacerse con este bajo R^2 y este *ESR* elevado? El hecho de que el R^2 de esta regresión sea bajo (y el *ESR* sea grande) no implica, en sí mismo, que esta regresión sea «buena» o «mala». Lo que el R^2 bajo expresa es que otros factores importantes influyen en la calificación de los exámenes. Esos factores podrían incluir las diferencias en el conjunto de los estudiantes entre distritos, las diferencias en la calidad de la escuela no relacionadas con la ratio de estudiantes por maestro, o de la suerte en el examen. El R^2 bajo y el *ESR* elevado no nos dicen cuáles son esos factores, no obstante, indican que la ratio estudiantes-maestros solo explica una pequeña parte de la variación en las calificaciones en los exámenes en estos datos.

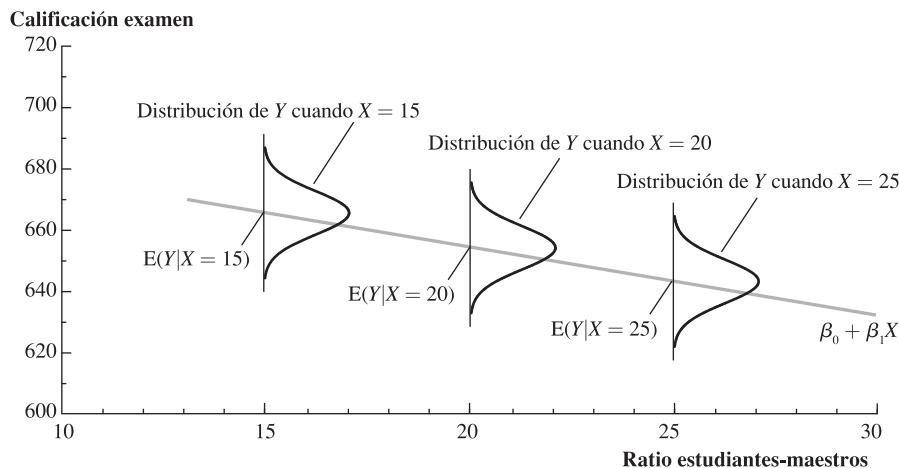
4.4 Los supuestos de mínimos cuadrados

En esta sección se presenta un conjunto de tres supuestos para el modelo de regresión lineal y el esquema de muestreo bajo el cual MCO proporciona un estimador apropiado para los coeficientes de regresión desconocidos, β_0 y β_1 . Inicialmente, estos supuestos pueden resultar abstractos. Sin embargo, tienen interpretaciones naturales, y la comprensión de estos supuestos resulta esencial para comprender cuándo MCO proporcionará —y cuándo no— estimaciones útiles de los coeficientes de regresión.

Supuesto #1: La distribución condicional de u_i dado X_i tiene media igual a cero

El primero de los tres **supuestos de mínimos cuadrados** es que la distribución condicionada de u_i dado X_i tiene una media igual a cero. Este supuesto es una condición formalizada matemáticamente, acerca de los «otros factores» contenidos en u_i , y establece que estos otros factores están incorrelacionados con X_i , en el sentido de que, dado un valor de X_i , la media de la distribución de estos otros factores es cero.

FIGURA 4.4 Las distribuciones condicionales de probabilidad y la recta de regresión poblacional



La figura muestra la probabilidad condicional de las calificaciones en los exámenes para los distritos con tamaño de clases de 15, 20, y 25 estudiantes. La media de la distribución condicional de las calificaciones en los exámenes, dada la ratio estudiantes-maestros, $E(Y|X)$, es la recta de regresión poblacional $\beta_0 + \beta_1 X$. Para un valor dado de X , Y se distribuye en torno a la recta de regresión y el error, $u = Y - (\beta_0 + \beta_1 X)$, tiene una media condicional igual a cero para todos los valores de X .

Este supuesto se ilustra en la Figura 4.4. La regresión poblacional es la relación que se cumple en promedio entre el tamaño de la clase y los resultados en los exámenes en la población, y el término de error u_i representa los otros factores que conducen a que los resultados en los exámenes de un distrito dado difieran de su predicción basada en la recta de regresión poblacional. Como se muestra en la Figura 4.4, a un valor dado de tamaño de las clases, por ejemplo 20 alumnos por clase, esos otros factores a veces conducen a un mejor desempeño de lo previsto ($u_i > 0$) y a veces a uno peor ($u_i < 0$), pero en promedio de población la predicción es correcta. En otras palabras, dado $X_i = 20$, la media de la distribución de u_i es cero. En la Figura 4.4, esto se representa mediante la distribución de u_i que está centrada en la recta de regresión poblacional en $X_i = 20$ y, de forma más general, lo mismo para otros valores x de X_i . Dicho de otra forma, la distribución de u_i condicionada a $X_i = x$ tiene media cero; matemáticamente, $E(u_i|X_i = x) = 0$, o, en una notación algo más sencilla, $E(u_i|X_i) = 0$.

Como se muestra en la Figura 4.4, el supuesto de que $E(u_i|X_i) = 0$ es equivalente a suponer que la recta de regresión poblacional es la media condicional de Y_i dado X_i (se deja como ejercicio una demostración matemática de esto, Ejercicio 4.6).

La media condicional de u en un experimento aleatorizado controlado. En un experimento aleatorizado controlado, los sujetos son asignados al azar al grupo de tratamiento ($X = 1$) o al grupo de control ($X = 0$). La asignación aleatoria se realiza habitualmente mediante un programa informático que no utiliza ninguna información sobre el sujeto, asegurando así que X se distribuye de forma independiente de todas las características personales de los sujetos. La asignación al azar hace a X y a u independientes, lo que a su vez implica que la media condicional de u dado X es cero.

En los datos observacionales, X no se asigna aleatoriamente en un experimento. En su lugar, lo mejor que se puede esperar es que X sea *como si* fuera asignada al azar, en el sentido exacto de que $E(u_i|X_i) = 0$. El hecho de si este supuesto se cumple en una determinada aplicación empírica con datos observacionales requiere de una cuidadosa reflexión y de una valoración, y volveremos sobre esta cuestión en varias ocasiones.

La correlación y la media condicional. Recordemos de la Sección 2.3 que si la media condicional de una variable aleatoria dado otra es cero, entonces las dos variables aleatorias tienen covarianza cero y por lo

tanto no están correlacionadas [Ecuación (2.27)]. Por tanto, el supuesto de media condicional $E(u_i|X_i) = 0$ implica que X_i y u_i están incorrelacionadas, o que $\text{corr}(X_i, u_i) = 0$. Debido a que la correlación es una medida de asociación lineal, esta implicación no se cumple en sentido inverso; incluso aunque X_i y u_i no estén correlacionadas, la media condicional de u_i dado X_i puede ser distinta de cero. Sin embargo, si X_i y u_i están correlacionadas, entonces debe ocurrir que $E(u_i|X_i)$ es distinta de cero. Por lo tanto, suele ser conveniente estudiar el supuesto de media condicional igual a cero en términos de la posible correlación entre X_i y u_i . Si X_i y u_i están correlacionadas, entonces se viola el supuesto de media condicional.

Supuesto #2: $(X_i, Y_i), i = 1, \dots, n$, son independientes e idénticamente distribuidas

El segundo supuesto de mínimos cuadrados consiste en que $(X_i, Y_i), i = 1, \dots, n$, son independientes e idénticamente distribuidas (i.i.d.) para distintas observaciones. Como se estudió en la Sección 2.5 (Concepto clave 2.5), este supuesto es una condición acerca del método de extracción de la muestra. Si las observaciones se extraen mediante muestreo aleatorio simple de una única y gran población, entonces, $(X_i, Y_i), i = 1, \dots, n$ son i.i.d. Por ejemplo, sea X la edad de un trabajador e Y sus ingresos salariales, e imaginemos que se selecciona una persona al azar a partir de la población de trabajadores. Esta persona extraída aleatoriamente tendrá una edad y unos ingresos determinados (es decir, X e Y tomarán unos valores). Si se extrae una muestra de n trabajadores de esta población, entonces $(X_i, Y_i), i = 1, \dots, n$, necesariamente presentan la misma distribución. Si se han seleccionado al azar, se distribuyen además de manera independiente de una observación a otra, es decir, son i.i.d.

El supuesto de i.i.d. es razonable para muchos sistemas de recopilación de datos. Por ejemplo los datos de encuesta de un subgrupo elegido al azar de la población pueden ser generalmente considerados como i.i.d.

No obstante, no todas las técnicas de muestreo proporcionan observaciones i.i.d. sobre (X_i, Y_i) . Un ejemplo de ello es cuando los valores de X no han sido extraídos de una muestra aleatoria de la población, pero aun así se establecen por un investigador como parte de un experimento. Por ejemplo, supongamos que una horticultora quiere estudiar los efectos de diferentes métodos orgánicos de escardar (X) sobre la producción de tomates (Y) y el distinto crecimiento que conlleva sobre las diferentes parcelas de tomate la utilización de diferentes técnicas orgánicas de escardado. Si elige la técnica a usar (el nivel de X) sobre la i -ésima parcela y se aplica la misma técnica a la i -ésima parcela en todas las repeticiones del experimento, entonces el valor de X_i no cambia de una muestra a la otra. Por lo tanto X_i es no aleatoria (aunque el resultado Y_i sea aleatorio), por lo que el esquema muestral no es i.i.d. Los resultados mostrados en este capítulo desarrollados para regresores i.i.d. son igualmente ciertos si los regresores son no aleatorios. Sin embargo, el caso de un regresor no aleatorio es bastante especial. Por ejemplo, los protocolos experimentales modernos habrían asignado para la horticultora el nivel de X para las diferentes parcelas utilizando un generador de números aleatorios, eludiendo así cualquier posible sesgo por parte de la horticultora (que podría utilizar su método favorito para escardar los tomates de la parcela más soleada). Cuando se utiliza este protocolo experimental moderno, el nivel de X es aleatorio y (X_i, Y_i) es i.i.d.

Otro ejemplo de muestreo no i.i.d. es cuando las observaciones se refieren a la misma unidad de observación para distintos períodos de tiempo. Por ejemplo, se pueden tener datos sobre los niveles de inventario (Y) de una empresa y el tipo de interés al que la empresa puede pedir prestado (X), donde estos datos se recogen para una empresa concreta a lo largo del tiempo; por ejemplo, se podrían haber recopilado cuatro veces al año (trimestral) durante 30 años. Este es un ejemplo de datos de series temporales, y una característica crucial de los datos de series temporales es que las observaciones que están cercanas en el tiempo unas de las otras no son independientes, sino que más bien tienden a estar correlacionadas unas con otras; si los tipos de interés ahora son bajos, es probable que sean bajos el trimestre próximo. Este patrón de correlación viola el supuesto de i.i.d. en la parte de la «independencia». Los datos de series temporales introducen un conjunto de complicaciones que se manejan mejor después de desarrollar las herramientas básicas de análisis de regresión.

Supuesto #3: Los datos atípicos elevados son improbables

El tercer supuesto de mínimos cuadrados es que los grandes atípicos, es decir, las observaciones con valores de X_i , Y_i , o de ambos, que están muy lejos de los límites del rango habitual de los datos, son poco probables. Los grandes atípicos pueden hacer que los resultados de la regresión MCO sean engañosos. Esta sensibilidad potencial de MCO a los valores atípicos extremos se ilustra en la Figura 4.5 con datos hipotéticos.

En este libro, el supuesto de que los valores atípicos grandes son improbables se precisa matemáticamente suponiendo que X e Y tienen momentos de cuarto orden que existen y son finitos: $0 < E(X_i^4) < \infty$ y $0 < E(Y_i^4) < \infty$. Otra forma de formular la hipótesis es que X e Y tienen curtosis finita.

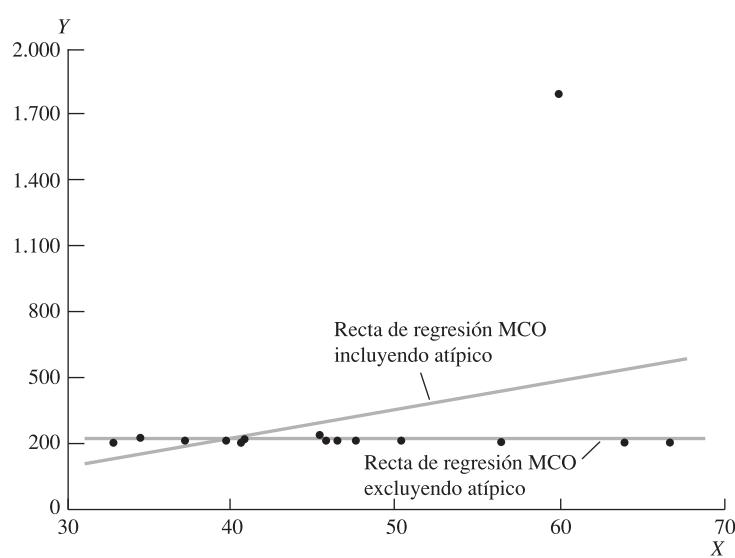
El supuesto de curtosis finita se utiliza en las matemáticas que justifican las aproximaciones para muestras grandes de las distribuciones de los estadísticos de contraste MCO. Nos toparemos con este supuesto en el Capítulo 3 cuando estudiemos la consistencia de la varianza muestral. En concreto, la Ecuación (3.9) establece que la varianza muestral s_Y^2 es un estimador consistente de la varianza poblacional σ_Y^2 ($s_Y^2 \xrightarrow{P} \sigma_Y^2$). Si Y_1, \dots, Y_n son i.i.d. y el momento de cuarto orden de Y_i es finito, entonces se aplica la ley de los grandes números del Concepto clave 2.6 al promedio $\frac{1}{n} \sum_{i=1}^n (Y_i - \mu_Y)^2$, un paso clave de la demostración del Apéndice 3.3 que demuestra que s_Y^2 es consistente.

Una fuente de grandes valores atípicos son los errores en la grabación o registro de los datos, tales como errores tipográficos o el uso incorrecto de diferentes unidades en observación diferentes. Imaginemos la recogida de datos sobre la altura de ocho estudiantes en metros, pero de forma inadvertida, en lugar de eso se registra la altura de un estudiante en centímetros. Una forma de detectar atípicos consiste en representar los datos. Si se decide que un atípico es debido a un error en el registro de los datos, se puede elegir entre corregir el error o, si hacerlo no es posible, sacar el dato de la base de datos.

Dejando de lado los errores de registro de datos, el supuesto de curtosis finita es verosímil en muchas aplicaciones con datos económicos. El tamaño de las clases está limitado por la capacidad física de las aulas; lo mejor que se puede realizar un examen estandarizado es obtener todas las respuestas correctas. Al tener el tamaño de las clases y la puntuación en los exámenes un rango finito, necesariamente tienen curtosis finita. De modo más amplio, las distribuciones utilizadas habitualmente, tales como la distribución normal, tienen momentos de cuarto orden. Adicionalmente, como cuestión matemática, algunas distribuciones tienen momentos de cuarto orden infinitos, que este supuesto excluye. Si se cumple el supuesto de momentos de cuarto orden finitos, entonces es improbable que las inferencias estadísticas que utilizan MCO estén dominadas por unas pocas observaciones.

FIGURA 4.5 La sensibilidad de MCO a los datos atípicos elevados

Esta base de datos hipotética presenta un atípico. La recta de regresión MCO estimada con el atípico presenta una fuerte relación positiva entre X e Y , pero la recta de regresión MCO estimada sin el atípico no muestra relación.



**CONCEPTO
CLAVE**
4.3**Los supuestos de mínimos cuadrados**

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n, \text{ donde}$$

1. el término de error u_i presenta una media condicional igual a cero dado X_i : $E(u_i|X_i) = 0$;
2. (X_i, Y_i) , $i = 1, \dots, n$ son extracciones independientes e idénticamente distribuidas (i.i.d.) de su distribución conjunta; y
3. los valores atípicos grandes son improbables: X_i e Y_i presentan momentos de cuarto orden finitos distintos de cero.

La utilización de los supuestos de mínimos cuadrados

Los tres supuestos de mínimos cuadrados para el modelo de regresión lineal están recogidos en el Concepto clave 4.3. Los supuestos de mínimos cuadrados desempeñan papeles gemelos, y acudiremos a ellos repetidamente en este libro.

El primero de los papeles es matemático: si se cumplen estos supuestos, entonces, como se muestra en la siguiente sección, en muestras grandes los estimadores MCO tienen distribuciones muestrales que son normales. A su vez, esta distribución normal para muestras grandes permite desarrollar métodos para el contraste de hipótesis y la construcción de intervalos de confianza mediante los estimadores MCO.

Su segunda función es la de organizar las circunstancias que plantean dificultades para la regresión MCO. Como veremos, el primer supuesto de mínimos cuadrados es el más importante a tener en cuenta en la práctica. Se discute una razón por la cual el primer supuesto de mínimos cuadrados podría no cumplirse en la práctica en el Capítulo 6, además, se tratan otras razones en la Sección 9.2.

Asimismo, es importante considerar si en una aplicación se cumple el segundo supuesto. Aunque es verosímil que se cumpla en muchas ocasiones para los datos de sección cruzada, el supuesto de independencia no es adecuado para datos de series temporales. Por tanto, los métodos de regresión desarrollados bajo el segundo supuesto requieren alguna modificación para algunas aplicaciones con datos de series temporales.

El tercer supuesto sirve como un recordatorio de que MCO, al igual que la media muestral, puede ser sensible a los valores atípicos de gran tamaño. Si el conjunto de datos contiene valores atípicos grandes, deberían examinarse con cuidado los valores extremos con el fin de asegurarse de que esas observaciones están registradas de forma correcta y pertenecen a la base de datos.

4.5 Distribución muestral de los estimadores MCO

Debido a que los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ se calculan a partir de una muestra seleccionada aleatoriamente, los estimadores en sí mismos son variables aleatorias con una distribución de probabilidad —la distribución muestral— que describe los valores que podrían tomar en las diferentes muestras aleatorias posibles. Esta sección presenta estas distribuciones muestrales.

En muestras pequeñas, estas distribuciones son complicadas, pero en muestras grandes, son aproximadamente normales por el teorema central del límite.

La distribución muestral de los estimadores MCO

Repaso de la distribución muestral de \bar{Y} . Recordemos la discusión sobre la distribución muestral de la media muestral \bar{Y} de las Secciones 2.5 y 2.6, como estimador de la media poblacional desconocida de \bar{Y} , μ_Y . Debido a que \bar{Y} se calcula utilizando una muestra aleatoria, \bar{Y} es una variable aleatoria que toma diferentes valores entre una muestra y otra; la probabilidad de esos valores está recogida por su distribución muestral. Aunque la distribución muestral de \bar{Y} puede ser complicada si el tamaño de la muestra es pequeño, es posible señalar ciertas propiedades que se cumplen para todo n . En particular, la media de la distribución muestral es μ_Y . Es decir, $E(\bar{Y}) = \mu_Y$, por lo que \bar{Y} es un estimador insesgado de μ_Y . Si n es grande, entonces

puede decirse más acerca de la distribución muestral. En particular, el teorema central del límite (Sección 2.6) establece que esta distribución es aproximadamente normal.

La distribución muestral de $\hat{\beta}_0$ y $\hat{\beta}_1$. Estas ideas se transfieren a los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ del término independiente β_0 y de la pendiente β_1 poblacionales y desconocidos de la recta de regresión. Debido a que los estimadores MCO se calculan utilizando una muestra aleatoria, $\hat{\beta}_0$ y $\hat{\beta}_1$ son variables aleatorias que toman valores diferentes de una muestra a otra; la probabilidad de estos diferentes valores se resume en sus distribuciones muestrales.

A pesar de que las distribuciones muestrales de $\hat{\beta}_0$ y $\hat{\beta}_1$, pueden ser complicadas cuando el tamaño muestral es pequeño, es posible hacer ciertas afirmaciones acerca de ellas que se cumplen para todos los n . En particular, la media de las distribuciones muestrales de $\hat{\beta}_0$ y $\hat{\beta}_1$ son β_0 y β_1 . En otras palabras, bajo los supuestos de mínimos cuadrados del Concepto clave 4.3,

$$E(\hat{\beta}_0) = \beta_0 \quad \text{y} \quad E(\hat{\beta}_1) = \beta_1; \quad (4.20)$$

es decir, $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados de β_0 y β_1 . La prueba de que $\hat{\beta}_1$ es insesgado se ofrece en el Apéndice 4.3, y la demostración de que $\hat{\beta}_0$ es insesgado se deja como Ejercicio 4.7.

Si la muestra es suficientemente grande, por el teorema central del límite la distribución muestral de $\hat{\beta}_0$ y $\hat{\beta}_1$ está bien aproximada por la distribución normal bivariante (Sección 2.4). Esto implica que las distribuciones marginales de $\hat{\beta}_0$ y $\hat{\beta}_1$ son normales en muestras grandes.

Este argumento invoca el teorema central del límite. Técnicamente, el teorema central del límite se refiere a la distribución de las medias (como \bar{Y}). Si se examina el numerador de la Ecuación (4.7) para, $\hat{\beta}_1$, se verá que, asimismo, es un tipo de media y no una simple media, como \bar{Y} , sino una media del producto $(Y_i - \bar{Y})(X_i - \bar{X})$. Como se analiza adicionalmente en el Apéndice 4.3, el teorema central del límite es aplicable a esta media por lo que, como la media más sencilla \bar{Y} , se distribuye normalmente en muestras grandes.

La aproximación normal de la distribución de los estimadores MCO en muestras grandes se resume en el Concepto clave 4.4 (el Apéndice 4.3 recoge la deducción de estas fórmulas). Una cuestión relevante en la práctica es en qué medida debe ser n grande para que estas aproximaciones sean fiables. En la Sección 2.6, sugerimos que $n = 100$ es suficientemente grande para que la distribución muestral de \bar{Y} se aproxime bien a una distribución normal, y a veces es suficiente con un n más pequeño. Este criterio se traslada a las medias más complicadas que aparecen en el análisis de regresión. En casi todas las aplicaciones econométricas modernas, $n > 100$, por lo que consideraremos fiables las aproximaciones normales para las distribuciones de los estimadores MCO, a menos que existan buenas razones para pensar lo contrario.

Los resultados del Concepto clave 4.4 implican que los estimadores MCO son consistentes; es decir, cuando el tamaño de la muestra es grande, $\hat{\beta}_0$ y $\hat{\beta}_1$ se acercarán a los verdaderos coeficientes poblacionales.

**CONCEPTO
CLAVE**

4.4

Distribuciones para grandes muestras de $\hat{\beta}_0$ y $\hat{\beta}_1$

Si los supuestos de mínimos cuadrados del Concepto clave 4.3 se cumplen, entonces $\hat{\beta}_0$ y $\hat{\beta}_1$ poseen distribuciones muestrales conjuntas normales para muestras grandes. La distribución normal para muestras grandes de $\hat{\beta}_1$ es $N(\beta_1, \sigma_{\hat{\beta}_1}^2)$, donde la varianza de esta distribución $\sigma_{\hat{\beta}_1}^2$, es

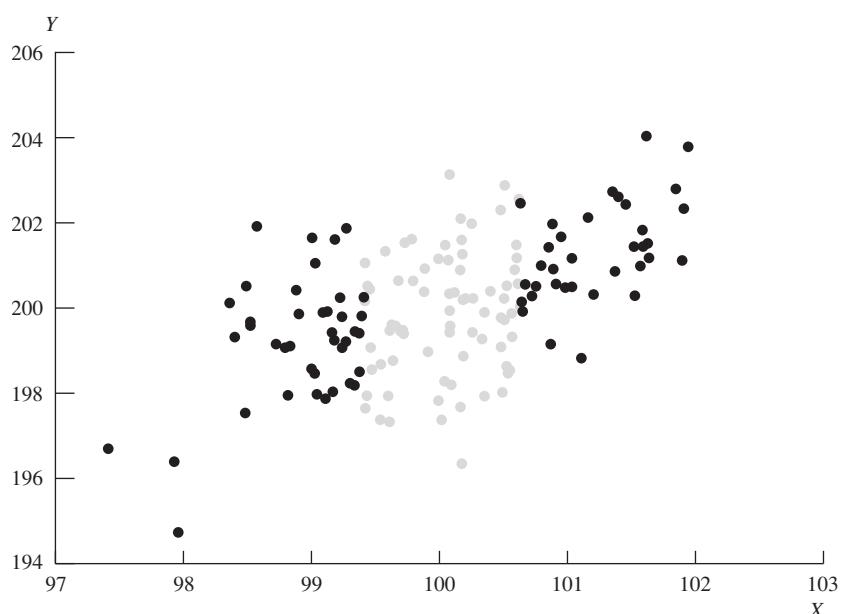
$$\sigma_{\hat{\beta}_1}^2 = \frac{1}{n} \frac{\text{var}[(X_i - \mu_X)u_i]}{[\text{var}(X_i)]^2} \quad (4.21)$$

La distribución normal para muestras grandes de $\hat{\beta}_0$ es $N(\beta_0, \sigma_{\hat{\beta}_0}^2)$, donde

$$\sigma_{\hat{\beta}_0}^2 = \frac{1}{n} \frac{\text{var}(H_i u_i)}{[E(H_i^2)]^2}, \text{ donde } H_i = 1 - \left[\frac{\mu_X}{E(X_i^2)} \right] X_i \quad (4.22)$$

FIGURA 4.6 La varianza de $\hat{\beta}_1$ y la varianza de X

Los puntos grises representan un conjunto de X_i con varianza pequeña. Los círculos negros representan un conjunto de X_i con varianza grande. La recta de regresión puede ser estimada de forma más precisa con los puntos negros que con los puntos grises.



β_0 y β_1 con alta probabilidad. Esto se debe a que las varianzas $\sigma_{\hat{\beta}_0}^2$ y $\sigma_{\hat{\beta}_1}^2$ de los estimadores tienden a cero al aumentar n (n aparece en el denominador de las fórmulas para las varianzas), por lo que la distribución de los estimadores MCO estará fuertemente concentrada alrededor de sus medias β_0 y β_1 , cuando n es grande.

Otra implicación de las distribuciones del Concepto clave 4.4 es que, en general, cuanto mayor es la varianza de X_i , menor es la varianza $\sigma_{\hat{\beta}_1}^2$ de $\hat{\beta}_1$. Matemáticamente, esta implicación surge debido a que la varianza de $\hat{\beta}_1$ en la Ecuación (4.21) es inversamente proporcional al cuadrado de la varianza de X_i ; cuanto mayor es $\text{var}(X_i)$, mayor es el denominador de la Ecuación (4.21) por lo que menor es $\sigma_{\hat{\beta}_1}^2$. Para obtener una mejor idea de por qué esto es así, obsérvese la Figura 4.6, que presenta un diagrama de dispersión de 150 puntos de datos artificiales sobre X e Y . Los puntos de datos indicados con círculos grises son las 75 observaciones más cercanas a \bar{X} . Supongamos que se le pide trazar una recta con la mayor precisión posible a través tanto de los círculos negros como de los círculos grises, ¿cuál elegiría? Sería más fácil dibujar una línea precisa a través de los círculos negros, que tienen una mayor varianza que los círculos grises. Del mismo modo, cuanto mayor sea la varianza de X , más preciso es $\hat{\beta}_1$.

Las distribuciones del Concepto clave 4.4 implican asimismo que cuanto menor es la varianza del término de error u_i , más pequeña es la varianza de $\hat{\beta}_1$. Esto puede verse matemáticamente en la Ecuación (4.21), ya que u_i aparece en el numerador, pero no en el denominador, de $\sigma_{\hat{\beta}_1}^2$: si todos los u_i fueran menores en una escala de un medio pero las X no cambiaron, entonces, $\sigma_{\hat{\beta}_1}$ sería menor en una escala de un medio y $\sigma_{\hat{\beta}_1}^2$ sería menor en una escala de un cuarto (Ejercicio 4.13). Expresado de una forma menos matemática, si los errores son menores (manteniendo las X fijas), entonces los datos presentarán una menor dispersión alrededor de la recta de regresión poblacional, por lo que su pendiente se estimará de manera más precisa.

La aproximación normal para la distribución muestral de $\hat{\beta}_0$ y $\hat{\beta}_1$ es una herramienta poderosa. Con esta aproximación disponible, somos capaces de desarrollar métodos para realizar inferencias sobre los valores poblacionales de los coeficientes de regresión utilizando solamente datos muestrales.

4.6 Conclusión

Este capítulo se ha centrado en la utilización de los mínimos cuadrados ordinarios para la estimación del término independiente y la pendiente de una recta de regresión poblacional mediante una muestra de n

observaciones de una variable dependiente, Y , y un único regresor, X . Existen muchas formas de trazar una línea recta a través de un diagrama de dispersión, pero realizarlo utilizando MCO tiene algunas virtudes. Si se cumplen los supuestos de mínimos cuadrados, entonces los estimadores MCO de la pendiente y del término independiente son insesgados, son consistentes, y presentan una distribución muestral con una varianza que es inversamente proporcional al tamaño de la muestra n . Por otra parte, si n es grande, la distribución muestral del estimador MCO es normal.

Estas propiedades importantes de la distribución muestral del estimador MCO se cumplen bajo los tres supuestos de mínimos cuadrados.

El primer supuesto es que el término de error del modelo de regresión lineal tiene una media condicional igual a cero, dado el regresor X . Este supuesto implica que el estimador MCO es insesgado.

El segundo supuesto es que (X_i, Y_i) son i.i.d., como en el caso en que los datos se recogen por muestreo aleatorio simple. Este supuesto da lugar a la fórmula, que se presenta en el Concepto clave 4.4, para la varianza de la distribución muestral del estimador MCO.

El tercer supuesto es que los valores atípicos grandes son poco probables. Enunciado de manera más formal, X e Y tienen momentos de cuarto orden finitos (curtosis finita). La razón de este supuesto es que MCO pueden ser poco fiables en presencia de valores atípicos extremos. En conjunto, los tres supuestos de mínimos cuadrados implican que el estimador MCO presenta una distribución normal para muestras grandes como se describe en el Concepto clave 4.4.

Los resultados de este capítulo describen la distribución muestral del estimador MCO. Por sí mismos, sin embargo, estos resultados no son suficientes para contrastar una hipótesis acerca del valor de β_1 o para construir un intervalo de confianza para β_1 . Hacerlo requiere un estimador de la desviación típica de la distribución muestral, es decir, el error estándar del estimador MCO. Este paso —pasar de la distribución muestral de $\hat{\beta}_1$ a su error estándar, los contrastes de hipótesis y los intervalos de confianza— se tratan en el capítulo siguiente.

Resumen

1. La recta de regresión poblacional, $\beta_0 + \beta_1 X$, es la media de Y como función del valor de X . La pendiente, β_1 , es la variación esperada en Y asociada con una variación de X en una unidad. El término independiente o intercepto, β_0 , determina el nivel (o altura) de la recta de regresión. El Concepto clave 4.1 resume la terminología del modelo de regresión lineal poblacional.
2. La recta de regresión poblacional puede ser estimada utilizando observaciones muestrales (X_i, Y_i) , $i = 1, \dots, n$ por mínimos cuadrados ordinarios (MCO). Los estimadores MCO del término independiente o intercepto de la regresión y la pendiente se expresan mediante $\hat{\beta}_0$ y $\hat{\beta}_1$.
3. El R^2 y el error estándar de la regresión (ESR) son medidas sobre la cercanía de los valores de Y_i a la recta de regresión estimada. El R^2 toma valores entre 0 y 1, un valor elevado indica que las Y_i están más cerca de la recta. El error estándar de la regresión es un estimador de la desviación típica del error de regresión.
4. Existen tres supuestos básicos para el modelo de regresión lineal: (1) los errores de regresión, u_i , presentan una media condicional a los regresores X_i igual a cero; (2) las observaciones muestrales son extracciones aleatorias i.i.d. de la población; y (3) los valores extremos grandes son poco probables. Si estos supuestos se cumplen, los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son (1) insesgados, (2) consistentes, y (3) están normalmente distribuidos cuando la muestra es grande.

Términos clave

modelo de regresión lineal con un regresor único (79)
variable dependiente (79)
variable independiente (79)
regresor (79)

recta de regresión poblacional (79)
función de regresión poblacional (79)
intercepto poblacional (79)
pendiente poblacional (79)
coeficientes poblacionales (79)

parámetros (79)	residuo (82)
término de error (79)	R^2 de la regresión (85)
estimadores de mínimos cuadrados ordinarios (MCO)(82)	suma explicada de cuadrados (SE) (86)
recta de regresión MCO (82)	suma total de cuadrados (ST) (86)
recta de regresión muestral (82)	suma de los cuadrados de los residuos (SR) (86)
función de regresión muestral (82)	error estándar de la regresión (ESR) (86)
valor de predicción (82)	supuestos de mínimos cuadrados (87)

Revisión de conceptos

- 4.1** Explique la diferencia entre $\hat{\beta}_1$ y β_1 ; entre los residuos \hat{u}_i y el error de la regresión u_i ; y entre el valor de predicción MCO \hat{Y}_i y $E(Y_i|X_i)$.
- 4.2** Para cada supuesto de mínimos cuadrados, proporcione un ejemplo en el cual el supuesto es válido, y después proporcione un ejemplo en el que el supuesto no se cumpla.
- 4.3** Dibuje un diagrama de dispersión con datos hipotéticos de una regresión estimada con $R^2 = 0,9$. Dibuje un diagrama de dispersión con datos hipotéticos de una regresión con $R^2 = 0,5$.

Ejercicios

- 4.1** Suponga que un investigador utiliza datos sobre el tamaño de las clases (TC) y de los promedios de las calificaciones en los exámenes para 100 clases de tercer curso, para estimar la regresión MCO

$$\widehat{\text{CalificaciónExamen}} = 520,4 - 5,82 \times TC, R^2 = 0,08, ESR = 11,5.$$

- a)** Un aula tiene 22 estudiantes. ¿Cuál es la predicción de la regresión para la calificación media en el examen para esa clase?
 - b)** El año pasado, un aula tenía 19 estudiantes, y este año cuenta con 23 alumnos. ¿Cuál es la predicción de la regresión para la variación en la media de las calificaciones en el examen para la clase?
 - c)** La media muestral del tamaño de la clase para 100 aulas es de 21,4. ¿Cuál es la media muestral de las calificaciones en el examen entre las 100 aulas? (Pista: repasar las fórmulas de los estimadores MCO).
 - d)** ¿Cuál es la desviación típica muestral de las calificaciones en los exámenes entre las 100 aulas? (Pista: repasar las fórmulas de R^2 y del ESR).
- 4.2** Supóngase que se selecciona una muestra aleatoria de 200 varones de veinte años de edad de una población y que se registra la altura y el peso de estos hombres. Una regresión del peso sobre la altura da

$$\widehat{\text{Peso}} = -99,41 + 3,94 \times \text{Altura}, R^2 = 0,81, ESR = 10,2,$$

donde $Peso$ se mide en libras¹ y $Altura$ se mide en pulgadas.

- a)** ¿Cuál es la predicción que proporciona la regresión para el peso de alguien que mide 70 pulgadas? ¿y para alguien de 74 pulgadas?
- b)** Un hombre da un estirón tardío y crece 1,5 pulgadas a lo largo de un año. ¿Cuál es la predicción que proporciona la regresión para el aumento de peso de este hombre?
- c)** Supóngase que en lugar de medir el peso y la altura en libras y pulgadas, esas variables se miden en centímetros y kilogramos. ¿Cuáles son las estimaciones de la regresión para esa nueva regresión centímetros-kilogramos? (Proporcione todos los resultados, los coeficientes estimados el R^2 y el ESR).

¹ N. del T.: 1 libra = 0,45 kg; 1 pulgada = 2,54 cm.

- 4.3** Una regresión del promedio de los ingresos salariales semanales (*ISM*, medidos en dólares) sobre la edad (medida en años), utiliza una muestra aleatoria de trabajadores con estudios universitarios a tiempo completo entre 25 y 65 años de edad, y obtiene lo siguiente:

$$\widehat{ISM} = 696,7 + 9,6 \times Edad, R^2 = 0,023, ESR = 624,1$$

- a) Explique qué significan los valores de los coeficientes 696,7 y 9,6.
- b) El error estándar de la regresión (*ESR*) es 624,1. ¿Cuáles son las unidades de medida del *ESR*? (¿Dólares? ¿Años? ¿O el *ESR* no tiene unidades?).
- c) El R^2 de la regresión es 0,023. ¿Cuáles son las unidades de medida de R^2 ? (¿Dólares? Años? ¿O el R^2 no tiene unidades?).
- d) ¿Cuáles son los ingresos salariales pronosticados por la regresión para un trabajador de 25 años de edad? ¿Y para un trabajador de 45 años de edad?
- e) ¿Será fiable la regresión en sus predicciones sobre un trabajador de 99 años de edad? ¿Por qué o por qué no?
- f) Teniendo en cuenta lo que se sabe acerca de la distribución de los ingresos, ¿cree que es posible que la distribución de los errores de la regresión sea normal? (*Pistas*: ¿piensa que la distribución es simétrica o asimétrica?, ¿cuál es el menor valor de los ingresos? y ¿es compatible con una distribución normal?).
- g) El promedio de edad de esta muestra es de 41,6 años. ¿Cuál es el valor medio muestral de *ISM*? (*Pista*: repasar el Concepto clave 4.2).

- 4.4** Lea el recuadro «La ‘beta’ de una acción» en la Sección 4.2.

- a) Supóngase que el valor de β es mayor que 1 para una determinada acción. Demuestre que la varianza de $(R - R_f)$ para esta acción es mayor que la varianza de $(R_m - R_t)$.
 - b) Supóngase que el valor de β es menor que 1 para una determinada acción. ¿Es posible que la varianza de $(R - R_f)$ para esta acción sea mayor que la varianza de $(R_m - R_t)$? (*Pista*: no olvide el error de regresión).
 - c) En un año determinado, la rentabilidad de las letras del Tesoro a 3 meses es del 3,5 % y la rentabilidad de una cartera diversificada de acciones (el S&P 500) es de 7,3 %. Para cada empresa recogida en la tabla del recuadro, utilice el valor estimado de β para estimar la rentabilidad esperada de la acción.
- 4.5** Un profesor decide realizar un experimento para medir el efecto de la presión del tiempo sobre las calificaciones en los exámenes finales. Da a cada uno de los 400 estudiantes de su curso el mismo examen final, pero algunos estudiantes tienen 90 minutos para completar el examen, mientras que otros tienen 120 minutos. A cada estudiante le es asignado al azar uno de los tiempos disponibles para la realización del examen mediante el lanzamiento de una moneda. Sea Y_i el número de puntos obtenidos en el examen por el estudiante i -ésimo. ($0 \leq Y_i \leq 100$), sea X_i la cantidad de tiempo del que dispone el estudiante para completar el examen ($X_i = 90$ o 120), y considérese el modelo de regresión $Y_i = \beta_0 + \beta_1 X_i + u_i$.

- a) Explique qué representa el término u_i . ¿Por qué presentarán diferentes estudiantes diferentes valores de u_i ?
- b) Explique por qué $E(u_i|X_i) = 0$ en este modelo de regresión.
- c) ¿Se cumplen los otros supuestos del Concepto clave 4.3? Explíquelo.
- d) La regresión estimada es $\hat{Y}_i = 49 + 0,24X_i$.
 - I) Calcule la predicción de la regresión estimada para la calificación media de los estudiantes a los que dieron 90 minutos para completar el examen. Repítalo para los de 120 minutos y 150 minutos.
 - II) Calcule la ganancia estimada en la puntuación de un estudiante al que se le dan 10 minutos más en el examen.

- 4.6** Demuestre que el primer supuesto de mínimos cuadrados, $E(u_i|X_i) = 0$, implica que $E(Y_i|X_i) = \beta_0 + \beta_1 X_i$.
- 4.7** Demuestre que $\hat{\beta}_0$ es un estimador insesgado de β_0 . (*Pista:* utilice el hecho de que $\hat{\beta}_1$ es insesgado, que se muestra en el Apéndice 4.3).
- 4.8** Supóngase que todos los supuestos de la regresión del Concepto clave 4.3 se cumplen excepto que el primer supuesto se sustituye por $E(u_i|X_i) = 2$. ¿Qué partes del Concepto clave 4.4 se siguen cumpliendo? ¿Cuáles cambian? ¿Por qué? (¿ $\hat{\beta}_1$ está normalmente distribuido en muestras grandes con la media y la varianza dadas en el Concepto clave 4.4? ¿Qué ocurre con $\hat{\beta}_0$?)
- 4.9** a) Una regresión lineal obtiene $\hat{\beta}_1 = 0$. Demuestre que $R^2 = 0$.
 b) Una regresión lineal obtiene que $R^2 = 0$. ¿Implica esto que $\hat{\beta}_1 = 0$?
- 4.10** Supóngase que $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde (X_i, u_i) son i.i.d., y X_i es una variable aleatoria de Bernoulli con $\Pr(X = 1) = 0,20$. Cuando $X = 1$, u_i es $N(0, 4)$; cuando $X = 0$, u_i es $N(0, 1)$.
 a) Demuestre que los supuestos de regresión del Concepto clave 4.3 se cumplen.
 b) Obtenga una expresión para la varianza en muestras grandes de $\hat{\beta}_1$. [*Pista:* valorar los términos de la Ecuación (4.21)].
- 4.11** Considérese el modelo de regresión $Y_i = \beta_0 + \beta_1 X_i + u_i$.
 a) Suponga que se sabe que $\beta_0 = 0$. Deduzca una fórmula para estimador de mínimos cuadrados de β_1 .
 b) Supóngase que se sabe que $\beta_0 = 4$. Deduzca una fórmula para el estimador de mínimos cuadrados de β_1 .
- 4.12** a) Demuestre que el R^2 de la regresión en la regresión de Y sobre X es el valor al cuadrado de la correlación muestral entre X e Y . Es decir, demuestre que $R^2 = r_{XY}^2$.
 b) Demuestre que el R^2 de la regresión de Y sobre X es el mismo que el R^2 de la regresión de X sobre Y .
 c) Demuestre que $\hat{\beta}_1 = r_{XY}(s_Y/s_X)$, donde r_{XY} es la correlación muestral entre X e Y , y s_Y y s_X son las desviaciones típicas muestrales de X e Y .
- 4.13** Supóngase que $Y_i = \beta_0 + \beta_1 X_i \kappa u_i$, donde κ es una constante distinta de cero y (X_i, Y_i) satisfacen los tres supuestos de mínimos cuadrados. Demuestre que la varianza para grandes muestras de $\hat{\beta}_1$ está dada por $\sigma_{\hat{\beta}_1}^2 = \kappa^2 \frac{1}{n} \frac{\text{var}[(X_i - \mu_X)u_i]}{\text{var}(X_i)^2}$. [*Pista:* esta ecuación es la varianza dada en la Ecuación (4.21) multiplicada por κ^2].
- 4.14** Demuestre que la recta de regresión muestral pasa por el punto (\bar{X}, \bar{Y}) .

Ejercicios empíricos

- E4.1** En la página web del libro http://www.pearsonhighered.com/stock_watson, se encuentra un archivo de datos **CPS08** que contiene una versión ampliada del conjunto de datos utilizado en la Tabla 3.1 para el año 2008. Contiene datos relativos a trabajadores a tiempo completo, para todo el año, de entre 25 y 34 años, titulados en la escuela secundaria o licenciados/ingenieros como grado más alto de educación alcanzado. Se ofrece una descripción detallada en **CPS08_Description**, asimismo disponible en la página web. (Son los mismos datos que en **CPS92_08** pero se limitan al año 2008). En este ejercicio, se investigará la relación entre la edad de un trabajador y sus ingresos salariales. (Por lo general, los trabajadores de más edad tienen más experiencia laboral, lo que conduce a una mayor productividad e ingresos).
 a) Realice una regresión de los ingresos medios por hora (AHE) sobre la edad (Age). ¿Cuál es el término independiente estimado? ¿Cuál es la pendiente estimada? Utilice la regresión estimada para responder a esta pregunta: ¿Cuánto aumentarán los ingresos al aumentar la edad de los trabajadores en un año?
 b) Bob es un trabajador de 26 años de edad. Prediga los ingresos de Bob a partir de la regresión estimada. Alexis es un trabajador de 30 años de edad. Prediga los ingresos de Alexis utilizando la regresión estimada.

- c) ¿Representa la edad una gran proporción de la varianza de los ingresos entre los individuos? Explíquelo.
- E.4.2** En la página web del libro http://www.pearsonhighered.com/stock_watson, se encuentra un archivo de datos **TeachingRatings** que contiene datos sobre las evaluaciones de la asignatura, las características de la asignatura y las características del profesor para 463 cursos de la Universidad de Texas en Austin². Una descripción detallada se ofrece en el archivo **TeachingRatings_Description**, del mismo modo disponible en la página web. Una de las características es un índice de «belleza» del profesor de acuerdo con la clasificación de un jurado de seis jueces. En este ejercicio, se investigará cómo las evaluaciones del curso están relacionados con la belleza del profesor.
- a) Construya un diagrama de dispersión para las evaluaciones medias del curso (*Course_Eval*) sobre la belleza del profesor (*Beauty*). ¿Parece haber una relación entre las variables?
 - b) Realice una regresión de las evaluaciones medias del curso (*Course_Eval*) sobre la belleza del profesor (*Beauty*). ¿Cuál es el término independiente estimado? ¿Cuál es la pendiente estimada? Explique por qué el término independiente estimado es igual a la media muestral de la variable *Course_Eval*. (Pista: ¿cuál es la media muestral de la variable *Beauty*?)
 - c) El Profesor Watson presenta un valor medio para la variable *Beauty*, mientras que el Profesor Stock presenta un valor para *Beauty* de una desviación estándar por encima de la media. Prediga las evaluaciones del curso del Profesor Stock y del Profesor Watson.
 - d) Opine acerca del tamaño de la pendiente de la regresión. ¿El efecto estimado de *Beauty* sobre *Course_Eval* es grande o pequeño? Explique qué entiende por «grande» y por «pequeño».
 - e) ¿Explica *Beauty* una proporción grande de la varianza de las evaluaciones entre los cursos? Explíquelo.
- E4.3** En la página web del libro http://www.pearsonhighered.com/stock_watson, se encuentra el archivo de datos **CollegeDistance** que contiene datos de una muestra aleatoria de alumnos de último año de secundaria entrevistados en 1980 y vueltos a entrevistar en 1986. En este ejercicio, se utilizarán estos datos con el fin de investigar la relación entre el número de años de educación completados por adultos jóvenes y la distancia de la escuela secundaria de cada estudiante a la universidad más cercana. (La proximidad a la universidad reduce el coste de la educación, por lo que los estudiantes que viven más cerca de una universidad deberían, en promedio, completar más años de educación superior). Se recoge una descripción detallada en **CollegeDistance_Description**, asimismo disponible en la página web³.
- a) Realice una regresión de los años completados de educación (*ED*) sobre la distancia a la universidad más cercana (*Dist*), donde *Dist* está medida en decenas de millas. (Por ejemplo, *Dist* = 2 significa que la distancia es de 20 millas.) ¿Cuál es la estimación para el término independiente? ¿Cuál es la pendiente estimada? Utilice la regresión estimada para responder a esta pregunta: ¿cuánto cambia el valor medio de los años de educación completados cuando las universidades se construyen cerca de donde los estudiantes acuden a la escuela secundaria?
 - b) La escuela secundaria de Bob estaba a 20 millas de la universidad más cercana. Prediga los años de educación completados por Bob utilizando la regresión estimada. ¿Cómo cambiarían los pronósticos sobre si Bob hubiera vivido a 10 millas de la universidad más cercana?
 - c) ¿Explica la distancia a la universidad una proporción grande de la varianza de los logros educativos entre los individuos? Explíquelo.
 - d) ¿Cuál es el valor del error estándar de la regresión? ¿Cuáles son las unidades del error estándar (metros, gramos, años, dólares, centavos, u otros)?

² Estos datos fueron proporcionados por el profesor Daniel Hamermesh de la Universidad de Texas en Austin y fueron utilizados en su artículo con Amy Parker, Parker, «Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity», *Economics of Education Review*, agosto de 2005, 24 (4): 369-376.

³ Estos datos fueron proporcionados por la profesora Cecilia Rouse de la Universidad de Princeton y fueron utilizados en su artículo «Democratization or Diversion? The Effect of Community Colleges on Educational Attainment», *Journal of Business and Economic Statistics*, abril 1995, 12(2): 217-224.

- E4.4** En la página web del libro, http://www.pearsonhighered.com/stock_watson, se encuentra el archivo de datos **Growth** que contiene datos sobre las tasas medias de crecimiento desde 1960 a 1995 para 65 países, junto con variables que potencialmente están relacionadas con el crecimiento. Se recoge una descripción detallada en **Growth_Description**, asimismo disponible en la página web. En este ejercicio, se investigará la relación entre el crecimiento y el comercio⁴.
- Construya un diagrama de dispersión de la tasa media de crecimiento anual (*Growth*) sobre la cuota media de participación del comercio (*TradeShare*). ¿Parece que existe relación entre las variables?
 - Un país, Malta, tiene una cuota de participación del comercio mucho mayor que la de otros países. Encuentre a Malta en el diagrama de dispersión. ¿Parece Malta un atípico?
 - Utilizando todas las observaciones, realice una regresión del *Growth* sobre *TradeShare*. ¿Cuál es la pendiente estimada? ¿Cuál es el término independiente estimado? Utilice la regresión para predecir la tasa de crecimiento de un país con una participación del comercio de un 0,5 y con una participación en el comercio igual a 1,0.
 - Estime la misma regresión excluyendo los datos de Malta. Conteste las mismas preguntas que en c.
 - ¿Dónde está Malta? ¿Por qué es la participación del comercio para Malta tan grande? ¿Debería Malta estar incluida o excluida del análisis?

APÉNDICE

4.1 La base de datos de las calificaciones en el examen de California

La base de datos del Programa de Exámenes Estandarizados de California contiene datos sobre el rendimiento en la prueba, las características de la escuela, y las características demográficas de los estudiantes. Los datos aquí utilizados proceden del total de los 420 distritos K-6 y K-8 de California con los datos disponibles para 1999. Las calificaciones en los exámenes son el promedio de las puntuaciones en lectura y matemáticas en el «Stanford 9 Achievement Test», un examen estándar que se realiza a los estudiantes de quinto curso. Las características de la escuela (promedio del distrito) incluyen la matrícula, el número de profesores (medido como «equivalentes a tiempo completo»), el número de ordenadores por aula, y el gasto por estudiante. La ratio estudiantes-maestros que se utiliza aquí es el número de estudiantes del distrito dividido por el número de docentes equivalentes a tiempo completo. Las variables demográficas de los estudiantes, asimismo, están promediadas para todo el distrito. Las variables demográficas incluyen el porcentaje de estudiantes que están en el programa público de asistencia CalWorks (anteriormente AFDC), el porcentaje de estudiantes admitidos para manutención a precio reducido, y el porcentaje de estudiantes que están aprendiendo inglés (es decir, estudiantes para los que el inglés es su segundo idioma). Todos estos datos fueron obtenidos del Departamento de Educación de California (www.cde.ca.gov).

APÉNDICE

4.2 Obtención de los estimadores MCO

Este apéndice utiliza el cálculo para deducir las fórmulas de los estimadores MCO recogidas en el Concepto clave 4.2. Para minimizar la suma de los errores de predicción al cuadrado $\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$ [Ecuación (4.6)], en primer lugar se toman las derivadas parciales con respecto a b_0 y b_1 :

$$\frac{\partial}{\partial b_0} \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2 = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) \text{ y} \quad (4.23)$$

$$\frac{\partial}{\partial b_1} \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2 = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) X_i. \quad (4.24)$$

⁴ Estos datos fueron proporcionados por el profesor Ross Levine, de la Universidad de Brown y fueron utilizados en su artículo con Thorsten Beck y Norman Loayza, «Finance and the Sources of Growth», *Journal of Financial Economics*, 2000, 58: 261-300.

Los estimadores MCO, $\hat{\beta}_0$ y $\hat{\beta}_1$, son los valores de b_0 y b_1 que minimizan $\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$ o, equivalentemente, los valores de b_0 y b_1 para los cuales las derivadas de las Ecuaciones (4.23) y (4.24) son iguales a cero. Por tanto, igualando estas derivadas a cero, agrupando términos y dividiendo por n , demostramos que los estimadores MCO, $\hat{\beta}_0$ y $\hat{\beta}_1$, deben cumplir las dos ecuaciones

$$\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0 \text{ y} \quad (4.25)$$

$$\frac{1}{n} \sum_{i=1}^n X_i Y_i - \hat{\beta}_0 \bar{X} - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^n X_i^2 = 0. \quad (4.26)$$

Resolviendo este par de ecuaciones para $\hat{\beta}_0$ y $\hat{\beta}_1$ queda

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \bar{Y}}{\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.27)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (4.28)$$

Las Ecuaciones (4.27) y (4.28) son las fórmulas para $\hat{\beta}_0$ y $\hat{\beta}_1$ que figuran en el Concepto clave 4.2; la fórmula $\hat{\beta}_1 = s_{YX}/s_X^2$ se obtiene dividiendo el numerador y el denominador de la Ecuación (4.27) por $n - 1$.

APÉNDICE

4.3 Distribución muestral del estimador MCO

En este apéndice se demuestra que el estimador MCO $\hat{\beta}_1$ es insesgado y, en muestras grandes, presenta la distribución muestral normal, dada en el Concepto clave 4.4.

Representación de $\hat{\beta}_1$, en términos de los regresores y los errores

Comenzamos ofreciendo una expresión para $\hat{\beta}_1$, en términos de las variables explicativas y los errores. Debido a que $Y_i = \beta_0 + \beta_1 X_i + u_i$, $Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + u_i - \bar{u}$, por lo que el numerador de la fórmula para $\hat{\beta}_1$ en la Ecuación (4.27) es

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n (X_i - \bar{X})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})] \\ &= \beta_1 \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}). \end{aligned} \quad (4.29)$$

Ahora $\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n (X_i - \bar{X})u_i - \sum_{i=1}^n (X_i - \bar{X})\bar{u} = \sum_{i=1}^n (X_i - \bar{X})u_i$, donde la última igualdad se obtiene a partir de la definición de \bar{X} , lo que implica que $\sum_{i=1}^n (X_i - \bar{X})\bar{u} = [\sum_{i=1}^n X_i - n\bar{X}]\bar{u} = 0$. Sustituyendo $\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n (X_i - \bar{X})u_i$ en la última expresión de la Ecuación (4.29), se obtiene $\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \beta_1 \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (X_i - \bar{X})u_i$. Sustituyendo esta expresión a su vez en la fórmula de $\hat{\beta}_1$ en la Ecuación (4.27) se obtiene

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.30)$$

Prueba de que $\hat{\beta}_1$ es insesgado

La esperanza de $\hat{\beta}_1$ se obtiene tomando la esperanza a ambos lados de la Ecuación (4.30). Por lo tanto,

$$\begin{aligned} E(\hat{\beta}_1) &= \beta_1 + E \left[\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right] \\ &= \beta_1 + E \left[\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) E(u_i | X_1, \dots, X_n)}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right] = \beta_1 \end{aligned} \quad (4.31)$$

donde la segunda igualdad de la ecuación (4.31) se obtiene utilizando la ley de esperanzas iteradas (Sección 2.3). Por el segundo supuesto de mínimos cuadrados, u_i se distribuye independientemente de X para todas las demás observaciones distintas de i , por lo que $E(u_i | X_1, \dots, X_n) = E(u_i | X_i)$. Por el primer supuesto de mínimos cuadrados, sin embargo, $E(u_i | X_i) = 0$. Esto se desprende de que la esperanza condicional entre corchetes de la segunda línea de la Ecuación (4.31) es cero, por lo que $E(\hat{\beta}_1 - \beta_1 | X_1, \dots, X_n) = 0$. De manera equivalente $E(\hat{\beta}_1 | X_1, \dots, X_n) = \beta_1$; es decir, $\hat{\beta}_1$ es condicionalmente insesgado, dados X_1, \dots, X_n . Por la ley de esperanzas iteradas, $E(\hat{\beta}_1 - \beta_1) = E[E(\hat{\beta}_1 - \beta_1 | X_1, \dots, X_n)] = 0$, por lo que $E(\hat{\beta}_1) = \beta_1$; es decir, $\hat{\beta}_1$ es insesgado.

Distribución normal del estimador MCO para grandes muestras

La aproximación normal para muestras grandes de la distribución límite de $\hat{\beta}_1$ (Concepto clave 4.4) se obtiene considerando el comportamiento del último término de la Ecuación (4.30).

Consideremos primero el numerador de este término. Debido a que \bar{X} es consistente, si el tamaño de la muestra es grande, \bar{X} es casi igual a μ_X . Por tanto, para una aproximación ajustada, el término en el numerador de la Ecuación (4.30) es la media muestral \bar{v} , donde $v_i = (X_i - \mu_X)u_i$. Por el primero de los supuestos de mínimos cuadrados, v_i tiene una media igual a cero. Por el segundo supuesto de mínimos cuadrados, v_i es i.i.d. La varianza de v_i es $\sigma_v^2 = \text{var}[(X_i - \mu_X)u_i]$, que, por el tercer supuesto de mínimos cuadrados, es finita y distinta de cero. Por tanto, \bar{v} satisface todos los requisitos del teorema central del límite (Concepto clave 2.7). Por tanto $\bar{v}/\sigma_{\bar{v}}$ está, en muestras grandes, distribuida $N(0, 1)$, donde $\sigma_{\bar{v}}^2 = \sigma_v^2/n$. Por tanto, la distribución de \bar{v} está bien aproximada por la distribución $N(0, \sigma_{\bar{v}}^2/n)$.

Consideremos ahora la expresión del denominador de la Ecuación (4.30); esta es la varianza muestral de X (excepto la división por n en lugar de $n - 1$, que es intrascendente si n es grande). Como se estudió en la Sección 3.2 [Ecuación (3.8)], la varianza muestral es un estimador consistente de la varianza poblacional, por lo que en muestras grandes está arbitrariamente cerca de la varianza poblacional de X .

Combinando estos dos resultados, tenemos que, en muestras grandes, $\hat{\beta}_1 - \beta_1 \cong \bar{v}/\text{var}(X_i)$, de modo que la distribución muestral de $\hat{\beta}_1$ es, en muestras grandes, $N(\beta_1, \sigma_{\hat{\beta}_1}^2)$, donde $\tilde{\sigma}_{\hat{\beta}_1}^2 = \text{var}(\bar{v})/\text{var}(X_i)^2 = \text{var}[(X_i - \mu_X)u_i]/\{n[\text{var}(X_i)]^2\}$, que es la expresión de la Ecuación (4.21).

Algunas propiedades algebraicas adicionales acerca de MCO

Los residuos MCO y los valores estimados satisfacen:

$$\frac{1}{n} \sum_{i=1}^n \hat{u}_i = 0, \quad (4.32)$$

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_i = \bar{Y}, \quad (4.33)$$

$$\sum_{i=1}^n \hat{u}_i X_i = 0 \quad \text{y} \quad s_{\hat{u}X} = 0, \text{ y} \quad (4.34)$$

$$ST = SR + SE \quad (4.35)$$

Las Ecuaciones (4.32) a (4.35) dicen que la media muestral de los residuos MCO es igual a cero; la media muestral de los valores estimados MCO es igual a \bar{Y} ; la covarianza muestral $s_{\hat{u}X}$ entre los residuos MCO y los regresores es igual a cero; y la suma total es la suma residual más la suma explicada [SE, ST, y SR están definidas en las Ecuaciones (4.14), (4.15) y (4.17)].

Para verificar la Ecuación (4.32), debe tenerse en cuenta que la definición de $\hat{\beta}_0$ nos permite escribir los residuos MCO como $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = (Y_i - \bar{Y}) - \hat{\beta}_1(X_i - \bar{X})$; por tanto

$$\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n (Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^n (X_i - \bar{X}).$$

Pero las definiciones de \bar{Y} y \bar{X} implican que $\sum_{i=1}^n (Y_i - \bar{Y}) = 0$ y $\sum_{i=1}^n (X_i - \bar{X}) = 0$, por lo que $\sum_{i=1}^n \hat{u}_i = 0$.

Para verificar la Ecuación (4.33), téngase en cuenta que $Y_i = \hat{Y}_i + \hat{u}_i$ por lo que $\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i + \sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n \hat{Y}_i$, donde la segunda igualdad es consecuencia de la Ecuación (4.32).

Para verificar la Ecuación (4.34), téngase en cuenta que $\sum_{i=1}^n \hat{u}_i = 0$ implica que $\sum_{i=1}^n \hat{u}_i X_i = \sum_{i=1}^n \hat{u}_i (X_i - \bar{X})$, por lo que

$$\begin{aligned} \sum_{i=1}^n \hat{u}_i X_i &= \sum_{i=1}^n [(Y_i - \bar{Y}) - \hat{\beta}_1(X_i - \bar{X})](X_i - \bar{X}) \\ &= \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}) - \hat{\beta}_1 \sum_{i=1}^n (X_i - \bar{X})^2 = 0, \end{aligned} \tag{4.36}$$

donde la última igualdad de la Ecuación (4.36) se obtiene mediante la fórmula de $\hat{\beta}_1$ de la Ecuación (4.27). Este resultado, combinado con el resultado anterior, implica que $s_{\hat{u}X} = 0$.

La Ecuación (4.35) se obtiene de los resultados previos y con algo de álgebra:

$$\begin{aligned} ST &= \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 \\ &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + 2 \sum_{i=1}^n (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) \\ &= SR + SE + 2 \sum_{i=1}^n \hat{u}_i \hat{Y}_i = SR + SE, \end{aligned} \tag{4.37}$$

donde la última igualdad se obtiene a partir de $\sum_{i=1}^n \hat{u}_i \hat{Y}_i = \sum_{i=1}^n \hat{u}_i (\hat{\beta}_0 + \hat{\beta}_1 X_i) = \hat{\beta}_0 \sum_{i=1}^n \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^n \hat{u}_i X_i = 0$ por los resultados anteriores.

Regresión con regresor único: contrastes de hipótesis e intervalos de confianza

Este capítulo continúa el tratamiento de la regresión lineal con un único regresor. En el Capítulo 4 se explicaba cómo el estimador MCO $\hat{\beta}_1$ del coeficiente de la pendiente β_1 difiere de una muestra a otra, es decir, cómo $\hat{\beta}_1$ posee una distribución muestral. En este capítulo, se muestra cómo el conocimiento de esta distribución muestral puede utilizarse para hacer afirmaciones acerca de β_1 que resuman de forma precisa la incertidumbre del muestreo. El punto de partida es el error estándar del estimador MCO, que mide la dispersión de la distribución muestral de $\hat{\beta}_1$. La Sección 5.1 proporciona una expresión para este error estándar (y para el error estándar del estimador MCO del término independiente), a continuación se muestra cómo utilizar $\hat{\beta}_1$ y su error estándar para contrastar hipótesis. La Sección 5.2 explica cómo construir intervalos de confianza para β_1 . La Sección 5.3 recoge el caso especial de un regresor binario.

En las Secciones 5.1 a 5.3 se supone que se cumplen los tres supuestos de mínimos cuadrados del Capítulo 4. Si, además, se cumplen algunas condiciones más fuertes, entonces pueden obtenerse algunos resultados más fuertes a partir de la distribución del estimador MCO. Una de estas condiciones más fuertes consiste en que los errores sean homocedásticos, un concepto introducido en la Sección 5.4. En la Sección 5.5 se formula el teorema de Gauss-Markov, que establece que, bajo ciertas condiciones, MCO es eficiente (tiene la menor varianza) entre una cierta clase de estimadores. La Sección 5.6 analiza la distribución del estimador MCO cuando la distribución poblacional de los errores de regresión es normal.

5.1 Contraste de hipótesis acerca de uno de los coeficientes de regresión

El cliente, la directora, llama con un problema. Tiene un contribuyente enfadado en su oficina, que afirma que el recorte en el tamaño de las clases no ayuda a impulsar los resultados de los exámenes, por lo que reducirlo aún más es un desperdicio de dinero. El tamaño de las clases, clama el contribuyente, no tiene ningún efecto sobre las calificaciones obtenidas.

La reclamación del contribuyente puede ser reformulada en el lenguaje del análisis de regresión. Debido a que el efecto en las calificaciones de la prueba de una variación del tamaño de las clases en una unidad es $\beta_{TamañoClase}$, el contribuyente está afirmando que la recta de regresión poblacional es plana, es decir, la pendiente $\beta_{TamañoClase}$, de la recta de regresión poblacional, es cero. ¿Existe, pregunta la directora, evidencia en su muestra de 420 observaciones sobre los distritos escolares de California de que la pendiente es distinta de cero? ¿Se puede rechazar la hipótesis del contribuyente de que $\beta_{TamañoClase} = 0$, o debería aceptarla, al menos provisionalmente, en espera de nueva evidencia adicional?

Esta sección trata sobre los contrastes de hipótesis acerca de la pendiente β_1 o del intercepto β_0 de la recta de regresión poblacional. Comenzamos con el análisis detallado de los contrastes bilaterales sobre la pendiente β_1 , y posteriormente pasamos a los contrastes de hipótesis unilaterales acerca del término independiente β_0 .

**CONCEPTO
CLAVE**
Forma general del estadístico *t*

En general, el estadístico *t* tiene la forma

5.1

$$t = \frac{\text{estimador} - \text{valor en la hipótesis nula}}{\text{error estándar del estimador}} \quad (5.1)$$

Hipótesis bilaterales acerca de β_1

El método general para el contraste de hipótesis sobre el coeficiente β_1 es el mismo que para el contraste de hipótesis acerca de la media poblacional, por lo que comenzamos con un breve repaso.

Contraste de hipótesis sobre la media poblacional. Recordemos de la Sección 3.2 que la hipótesis nula de que la media de Y toma un valor concreto $\mu_{Y,0}$, puede expresarse como $H_0: E(Y) = \mu_{Y,0}$, y la alternativa bilateral como $H_1: (Y) \neq \mu_{Y,0}$.

El contraste de la hipótesis nula H_0 frente a la alternativa bilateral se realiza siguiendo los tres pasos recogidos en el Concepto clave 3.6. El primero consiste en calcular el error estándar de \bar{Y} , $ES(\bar{Y})$, que es un estimador de la desviación típica de la distribución muestral de \bar{Y} . El segundo paso consiste en calcular el estadístico *t*, que presenta la forma general dada en el Concepto clave 5.1: aplicado aquí, el estadístico *t* es $t = (\bar{Y} - \mu_{Y,0})/ES(\bar{Y})$.

El tercer paso consiste en calcular el *p*-valor, que es el menor nivel de significación con el que la hipótesis nula puede ser rechazada, en base al estadístico de contraste observado en realidad; de forma equivalente, el *p*-valor es la probabilidad de obtener un estadístico, debido a la variabilidad del muestreo aleatorio, al menos tan diferente del valor de la hipótesis nula como el estadístico observado en realidad, suponiendo que la hipótesis nula es cierta (Concepto clave 3.5). Debido a que el estadístico *t* presenta una distribución normal estándar en muestras grandes bajo la hipótesis nula, el *p*-valor para un contraste de hipótesis bilateral es $2\Phi(-|t^{act}|)$, donde t^{act} es el valor del estadístico *t* calculado en realidad y Φ es la distribución normal estándar acumulada de la Tabla 1 del Apéndice. Alternativamente, el tercer paso puede ser sustituido por la mera comparación entre el estadístico *t* con el valor crítico apropiado para el contraste con el nivel de significación deseado. Por ejemplo, un contraste de dos colas (bilateral) con un nivel de significación del 5 % rechazaría la hipótesis nula si $|t^{act}| > 1,96$. En este caso, la media poblacional se dice que es estadística y significativamente distinta del valor de la hipótesis nula con un nivel de significación del 5 %.

Contraste de hipótesis sobre la pendiente β_1 . A nivel teórico, la característica clave que justifica el procedimiento de contraste anterior para la media poblacional es que, en muestras grandes, la distribución muestral de \bar{Y} es aproximadamente normal. Como $\hat{\beta}_1$ también tiene una distribución muestral normal para muestras grandes, las hipótesis acerca del verdadero valor de la pendiente β_1 pueden contrastarse utilizando el mismo método general.

La hipótesis nula y la alternativa deben establecerse con precisión antes de ser contrastadas. La hipótesis del contribuyente enfadado es que $\beta_{TamañoClase} = 0$. En términos más generales, bajo la hipótesis nula, la verdadera pendiente poblacional β_1 toma un valor concreto, $\beta_{1,0}$. Bajo la alternativa bilateral, β_1 , es distinto de $\beta_{1,0}$. Es decir, la **hipótesis nula** y la **hipótesis alternativa bilateral** son

$$H_0: \beta_1 = \beta_{1,0} \text{ vs. } H_1: \beta_1 \neq \beta_{1,0} \quad (\text{alternativa bilateral}). \quad (5.2)$$

Para contrastar la hipótesis nula H_0 , se siguen los mismos tres pasos que para la media poblacional.

El primer paso es calcular el **error estándar de $\hat{\beta}_1$** , $ES(\hat{\beta}_1)$. El error estándar de $\hat{\beta}_1$ es un estimador de $\sigma_{\hat{\beta}_1}$, la desviación típica de la distribución muestral de $\hat{\beta}_1$. De forma concreta,

$$ES(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}^2} \quad (5.3)$$

donde

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} \quad (5.4)$$

El estimador de la varianza de la Ecuación (5.4) se estudia en el Apéndice 5.1. Aunque la fórmula de $\hat{\sigma}_{\hat{\beta}_1}^2$ es complicada, en las aplicaciones, el error estándar se calcula por el *software* de regresión, por lo que su cálculo resulta sencillo en la práctica.

El segundo paso es calcular el **estadístico t** ,

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{ES(\hat{\beta}_1)}. \quad (5.5)$$

El tercer paso consiste en calcular el **p -valor**, la probabilidad de observar un valor de $\hat{\beta}_1$ al menos tan diferente de $\beta_{1,0}$ como el estimado en realidad ($\hat{\beta}_1^{act}$), suponiendo que la hipótesis nula es correcta. De forma matemática,

$$\begin{aligned} p\text{-valor} &= \Pr_{H_0}[|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{act} - \beta_{1,0}|] \\ &= \Pr_{H_0}\left[\left|\frac{\hat{\beta}_1 - \beta_{1,0}}{ES(\hat{\beta}_1)}\right| > \left|\frac{\hat{\beta}_1^{act} - \beta_{1,0}}{ES(\hat{\beta}_1)}\right|\right] \Pr_{H_0}(|t| > |t^{act}|), \end{aligned} \quad (5.6)$$

donde \Pr_{H_0} expresa la probabilidad calculada bajo la hipótesis nula, la segunda igualdad se obtiene dividiendo por $ES(\hat{\beta}_1)$, y t^{act} es el valor del estadístico t calculado en realidad. Debido a que $\hat{\beta}_1$ está distribuido aproximadamente como una normal en muestras grandes, bajo la hipótesis nula el estadístico t se distribuye aproximadamente como una variable aleatoria normal estándar, por lo que en muestras grandes,

$$p\text{-valor} = \Pr(|Z| > |t^{act}|) = 2\Phi(-|t^{act}|). \quad (5.7)$$

Un p -valor de menos del 5 % proporciona evidencia contra la hipótesis nula en el sentido de que, bajo la hipótesis nula, la probabilidad de obtener un valor de $\hat{\beta}_1$ al menos tan alejado del de la hipótesis nula como el observado en la realidad es inferior al 5 %. Si es así, la hipótesis nula se rechaza al nivel de significación del 5 %.

Por otra parte, la hipótesis puede ser contrastada al nivel de significación del 5 % simplemente comparando el valor del estadístico t con $\pm 1,96$, el valor crítico para un contraste bilateral, y rechazando la hipótesis nula al nivel del 5 % si $|t^{act}| > 1,96$.

Estos pasos se resumen en el Concepto clave 5.2.

Presentación de las ecuaciones de regresión y aplicación a las calificaciones en los exámenes. La regresión MCO de las calificaciones en los exámenes sobre la ratio estudiantes-maestros, presentada en la Ecuación (4.11), obtenía $\hat{\beta}_0 = 698,9$ y $\hat{\beta}_1 = -2,28$. Los errores estándar de esas estimaciones son $ES(\hat{\beta}_0) = 10,4$ y $ES(\hat{\beta}_1) = 0,52$.

CONCEPTO CLAVE

5.2

Contraste de la hipótesis $\beta_1 = \beta_{1,0}$ frente a la alternativa $\beta_1 \neq \beta_{1,0}$

1. Calcular el error estándar de $\hat{\beta}_1$, $ES(\hat{\beta}_1)$ [Ecuación (5.3)].
2. Calcular el estadístico t [Ecuación (5.5)].
3. Calcular el p -valor [Ecuación (5.7)]. Rechazar la hipótesis al 5 % de nivel de significación si el p -valor es inferior a 0,05 o, equivalentemente, si $|t^{act}| > 1,96$.

El error estándar y (generalmente) el estadístico t y el p -valor para el contraste de $\beta_1 = 0$ se calculan automáticamente por el *software* de regresión.

Debido a la importancia de los errores estándar, por convención, se incluyen cuando se presentan los coeficientes MCO estimados. Una forma compacta de presentar los errores estándar consiste en colocarlos entre paréntesis debajo de los respectivos coeficientes de la recta de regresión MCO:

$$\text{CalificaciónExamen} = 698,9 - 2,28 \times REM, R^2 = 0,051, ESR = 18,6. \quad (5.8)$$

$$(10,4) \quad (0,52)$$

La Ecuación (5.8) informa asimismo del R^2 de la regresión y del error estándar de la regresión (ESR) de acuerdo con la recta de regresión estimada. Por tanto, la Ecuación (5.8) proporciona la recta de regresión estimada, las estimaciones de la incertidumbre por muestreo sobre la pendiente y el término independiente (los errores estándar), y dos medidas de ajuste de esta recta de regresión (ESR y R^2). Se trata de un formato habitual para la presentación de la ecuación de regresión, y será utilizado a lo largo del libro.

Supongamos que se desea contrastar la hipótesis nula de que la pendiente β_1 es cero en la homóloga poblacional de la Ecuación (5.8) al 5 % de nivel de significación. Para hacerlo, se construye el estadístico t y se compara con 1,96, el valor crítico al 5 % (bilateral) tomado de la distribución normal estándar. El estadístico t se construye mediante la sustitución del valor hipotético de β_1 bajo la hipótesis nula (cero), la pendiente estimada, y su error estándar en la Ecuación (5.8) en la fórmula general de la Ecuación (5.5), el resultado es $t^{act} = (-2,28 - 0)/0,52 = -4,38$. Este estadístico t es mayor (en valor absoluto) al valor crítico bilateral del 5 % que es 1,96, por lo que la hipótesis nula es rechazada en favor de la alternativa bilateral al nivel de significación del 5 %.

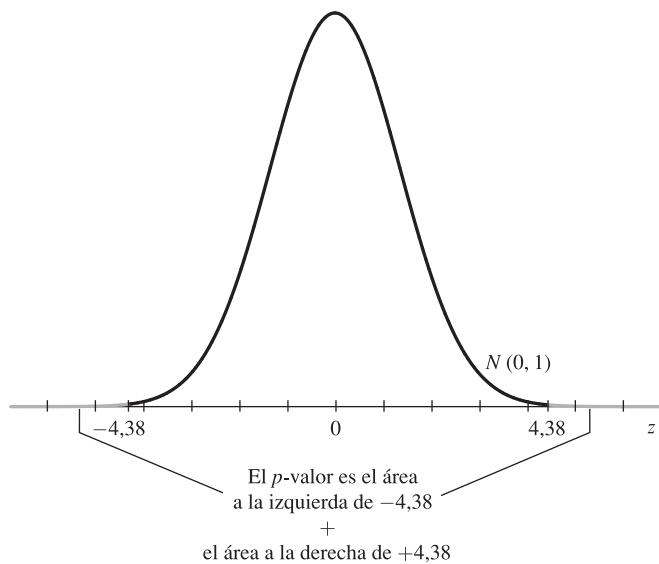
Alternativamente, se puede calcular el p -valor asociado con $t^{act} = -4,38$. Esta probabilidad es el área en las colas de la distribución normal estándar, tal y como se muestra en Figura 5.1. Esta probabilidad es extremadamente pequeña, aproximadamente 0,00001 o 0,001 %. Es decir, si la hipótesis nula $\beta_{TamañoClase} = 0$ es cierta, la probabilidad de obtener un valor de $\hat{\beta}_1$ tan alejado de la nula como el valor obtenido en realidad es extremadamente pequeña, menos del 0,001 %. Debido a que este suceso es tan poco probable, es razonable concluir que la hipótesis nula es falsa.

Hipótesis unilaterales sobre β_1

Hasta ahora se ha centrado el análisis en el contraste de la hipótesis de que $\beta_1 = \beta_{1,0}$ frente a la hipótesis alternativa de que $\beta_1 \neq \beta_{1,0}$. Esto es un contraste de hipótesis bilateral, porque bajo la alternativa β_1 podría

FIGURA 5.1 Cálculo del p -valor de un contraste bilateral cuando $t^{act} = -4,38$

El p -valor de un contraste bilateral es la probabilidad de que $|Z| > |t^{act}|$ donde Z es una variable aleatoria normal estándar y t^{act} es el valor del estadístico t calculado a partir de la muestra. Cuando $t^{act} = -4,38$, el p -valor es tan solo 0,00001.



ser tanto mayor como menor que $\beta_{1,0}$. A veces, no obstante, es conveniente utilizar un contraste de hipótesis unilateral. Por ejemplo, en el problema de las calificaciones en el examen/ratio estudiantes-maestros, mucha gente piensa que clases más pequeñas proporcionan un mejor entorno para el aprendizaje. Bajo esta hipótesis, β_1 es negativo: clases más pequeñas llevan a mejores calificaciones. Podría tener sentido, por tanto, contrastar la hipótesis nula de que $\beta_1 = 0$ (sin efecto) frente a la alternativa unilateral de que $\beta_1 < 0$.

Para un contraste unilateral, la hipótesis nula y la hipótesis alternativa unilateral es

$$H_0: \beta_1 = \beta_{1,0} \text{ vs. } H_1: \beta_1 < \beta_{1,0} \quad (\text{alternativa unilateral}) \quad (5.9)$$

donde $\beta_{1,0}$ es el valor de β_1 bajo la nula (0 en el ejemplo de la ratio estudiantes-maestros) y la alternativa es que β_1 sea menor que $\beta_{1,0}$. Si la alternativa es que β_1 sea mayor que $\beta_{1,0}$ la desigualdad de la Ecuación (5.9) se invierte.

Por ser la hipótesis nula la misma para los contrastes de hipótesis unilaterales y bilaterales, la construcción del estadístico t es la misma. La única diferencia entre los contrastes de hipótesis unilaterales y bilaterales es la interpretación del estadístico t . Para la alternativa unilateral de la Ecuación (5.9), la hipótesis nula se rechaza frente a la alternativa unilateral para valores del estadístico elevados y negativos pero no para valores elevados y positivos. En vez de ser rechazada si $|t^{act}| > 1,96$, la hipótesis se rechaza al 5 % de nivel de significación si $t^{act} < -1,645$.

El p -valor para un contraste unilateral se obtiene a partir de la distribución normal estándar acumulada como

$$p\text{-valor} = \Pr(Z < t^{act}) = \Phi(t^{act}) \quad (5.10)$$

Si la hipótesis alternativa es que β_1 es mayor que $\beta_{1,0}$, las desigualdades de las Ecuaciones (5.9) a (5.10) se invierten, por lo que el p -valor es la probabilidad de la cola derecha, $\Pr(Z > t^{act})$.

¿Cuándo debe utilizarse un contraste unilateral? En la práctica, una hipótesis alternativa unilateral debería ser utilizada solamente cuando exista un claro motivo para ello. Este motivo puede proceder de la teoría económica, la evidencia empírica previa, o de ambas. Sin embargo, aunque en un principio pudiera parecer que la alternativa relevante es unilateral, tras una reflexión, podría no serlo necesariamente. Una nueva formulación de un medicamento sometido recientemente a ensayos clínicos, podría en realidad resultar perjudicial debido a no haberse detectado previamente los efectos secundarios. En el ejemplo del tamaño de las clases, recordemos la broma de graduación de que el secreto del éxito de la universidad consiste en admitir a estudiantes con talento y asegurarse de que los profesores queden fuera de su camino y les dañen lo menos posible. En la práctica, esta ambigüedad conduce a menudo a que los económetras utilicen contrastes bilaterales.

Aplicación a las calificaciones en los exámenes. El estadístico t para el contraste de la ausencia de efecto del tamaño de las clases sobre las calificaciones obtenidas [$\beta_{1,0} = 0$ en la Ecuación (5.9)] es $t^{act} = -4,38$. Este valor es menor que $-2,33$ (el valor crítico para un contraste unilateral con un nivel de significación del 1 %), por lo que la hipótesis nula se rechaza frente a la alternativa unilateral al nivel del 1 %. De hecho, el p -valor es inferior a 0,0006 %. En base a estos datos, se puede rechazar la afirmación del contribuyente enfadado de que la estimación negativa de la pendiente surgió únicamente debido a la variación procedente del muestreo aleatorio con un nivel de significación del 1 %.

Contraste de hipótesis acerca del término independiente β_0

Este análisis se ha centrado en el contraste de hipótesis sobre la pendiente, β_1 . A veces, sin embargo, la hipótesis se refiere al término independiente β_0 . La hipótesis nula acerca del término independiente y la alternativa bilateral es

$$H_0: \beta_0 = \beta_{0,0} \text{ vs. } H_1: \beta_0 \neq \beta_{0,0} \quad (\text{alternativa bilateral}) \quad (5.11)$$

El método general para el contraste de esta hipótesis nula consta de los tres pasos del Concepto clave 5.2 aplicados a β_0 (la fórmula para el error estándar de $\hat{\beta}_0$ se ofrece en el Apéndice 5.1). Si la alternativa es

unilateral, este procedimiento se modifica tal y como se analizó en el apartado anterior para la hipótesis sobre la pendiente.

Los contrastes de hipótesis son útiles si se tiene en mente una hipótesis nula específica (como hacía nuestro contribuyente enfadado). La capacidad de aceptar o rechazar esta hipótesis nula en base a la evidencia estadística es una herramienta poderosa para afrontar la incertidumbre inherente a la utilización de una muestra para conocer la población. Aun así, hay muchas veces en que ninguna hipótesis sencilla sobre un coeficiente de la regresión es dominante, y en su lugar a uno le gustaría conocer un rango de valores del coeficiente que sea consistente con los datos. Lo cual reclama la construcción de un intervalo de confianza.

5.2 Intervalos de confianza para un coeficiente de regresión

Debido a que cualquier estimación estadística de la pendiente β_1 presenta necesariamente incertidumbre debida al muestreo, no es posible determinar el verdadero valor exacto de β_1 a partir de una muestra de datos. Sin embargo, sí es posible utilizar el estimador MCO y su error estándar para construir un intervalo de confianza para la pendiente β_1 o la ordenada en el origen β_0 .

Intervalo de confianza para β_1 . Recordemos que un **intervalo de confianza al 95 % para β_1** tiene dos definiciones equivalentes. En primer lugar, es el conjunto de valores que no pueden rechazarse mediante un contraste de hipótesis bilateral con un nivel de significación del 5 %. En segundo lugar, se trata de un intervalo que presenta una probabilidad del 95 % de contener el verdadero valor de β_1 ; es decir, en el 95 % de las posibles muestras que podrían ser seleccionadas, el intervalo de confianza contendrá el verdadero valor de β_1 . Debido a que este intervalo contiene el valor real en el 95 % de todas las muestras, se dice que tiene un **nivel de confianza del 95 %**.

La razón de que estas dos definiciones sean equivalentes es la siguiente. Un contraste de hipótesis con un nivel de significación del 5 %, por definición, rechazará el verdadero valor de β_1 solamente el 5 % de todas las muestras posibles; es decir, en el 95 % de todas las muestras posibles, el verdadero valor de β_1 no será rechazado. Debido a que el intervalo de confianza del 95 % (según la primera definición) es el conjunto de todos los valores de β_1 que no son rechazados al nivel de significación del 5 %, se deduce que el verdadero valor de β_1 estará contenido en el intervalo de confianza en el 95 % de todas las muestras posibles.

Como en el caso de un intervalo de confianza para la media poblacional (Sección 3.3), en principio, se puede calcular un intervalo de confianza al 95 % mediante el contraste de todos los valores posibles de β_1 (es decir, contrastando la hipótesis nula $\beta_1 = \beta_{1,0}$ para todos los valores de $\beta_{1,0}$) al 5 % de nivel de significación mediante el estadístico t . El intervalo de confianza al 95 % es por tanto la colección de todos los valores de que no se rechazan. Pero calcular el estadístico t para todos los valores de β_1 no terminaría nunca.

Una forma más sencilla de construir el intervalo de confianza es tener en cuenta que el estadístico t rechazaría el valor $\beta_{1,0}$ si $\beta_{1,0}$ está fuera del rango $\hat{\beta}_1 \pm 1,96ES(\hat{\beta}_1)$. Es decir, el intervalo de confianza al 95 % para β_1 es el intervalo $[\hat{\beta}_1 - 1,96ES(\hat{\beta}_1), \hat{\beta}_1 + 1,96ES(\hat{\beta}_1)]$. Este argumento es similar al argumento utilizado para desarrollar un intervalo de confianza para la media muestral.

La construcción de un intervalo de confianza para β_1 se recoge en el Concepto clave 5.3.

CONCEPTO
CLAVE

5.3

Intervalo de confianza para β_1

Un intervalo de confianza bilateral al 95 % para β_1 es un intervalo que contiene al verdadero valor de β_1 con una probabilidad del 95 %: es decir, contiene el verdadero valor de β_1 en el 95 % de todas las posibles muestras extraídas de forma aleatoria. De modo equivalente, es el conjunto de valores de β_1 que no pueden ser rechazados por un contraste de hipótesis bilateral al 5 %. Cuando el tamaño muestral es grande, se construye como

$$\begin{aligned} \text{intervalo de confianza al 95 \% para } \beta_1 &= \\ &= [\hat{\beta}_1 - 1,96ES(\hat{\beta}_1), \hat{\beta}_1 + 1,96ES(\hat{\beta}_1)]. \end{aligned} \quad (5.12)$$

Intervalo de confianza para β_0 . El intervalo de confianza al 95 % para β_0 se construye como en el Concepto clave 5.3, con $\hat{\beta}_0$ y $ES(\hat{\beta}_0)$ sustituyendo a $\hat{\beta}_1$ y $ES(\hat{\beta}_1)$.

Aplicación a las calificaciones en los exámenes. La regresión MCO de las calificaciones en los exámenes sobre la ratio estudiantes-maestros, recogida en la Ecuación (5.8), presentaba $\hat{\beta}_1 = -2,28$ y $ES(\hat{\beta}_1) = 0,52$. El intervalo de confianza bilateral al 95 % para β_1 es $\{-2,28 \pm 1,96 \times 0,52\}$, o $-3,30 \leq \beta_1 \leq -1,26$. El valor $\beta_1 = 0$ no está contenido en este intervalo de confianza, por lo que (como ya sabemos de la Sección 5.1) la hipótesis $\beta_1 = 0$ puede rechazarse al 95 % de nivel de significación.

Intervalos de confianza para la predicción de los efectos de la variación de X . El intervalo de confianza al 95 % para β_1 puede utilizarse para construir un intervalo de confianza al 95 % para la predicción del efecto de una variación general en X .

Consideremos la variación de X en una cantidad dada, Δx . La variación predicha en Y asociada a la variación de X es $\beta_1 \Delta x$. La pendiente poblacional β_1 es desconocida, pero como se puede construir un intervalo de confianza para β_1 , se puede construir un intervalo de confianza para el efecto esperado $\beta_1 \Delta x$. Como uno de los extremos del intervalo de confianza al 95 % para β_1 es $\hat{\beta}_1 - 1,96ES(\hat{\beta}_1)$, el efecto esperado de la variación Δx utilizando esta estimación de β_1 es $[\hat{\beta}_1 - 1,96ES(\hat{\beta}_1)] \times \Delta x$. El otro extremo del intervalo de confianza es $\hat{\beta}_1 + 1,96ES(\hat{\beta}_1)$, y el efecto esperado de la variación utilizando esta estimación es $[\hat{\beta}_1 + 1,96ES(\hat{\beta}_1)] \times \Delta x$. Por tanto un intervalo de confianza al 95 % para el efecto de la variación de x en la cuantía Δx se puede expresar como

$$\begin{aligned} \text{intervalo de confianza al 95 \% para } \beta_1 \Delta x = \\ [\hat{\beta}_1 \Delta x - 1,96ES(\hat{\beta}_1) \times \Delta x, \hat{\beta}_1 \Delta x + 1,96ES(\hat{\beta}_1) \times \Delta x]. \end{aligned} \quad (5.13)$$

Por ejemplo, nuestra hipotética directora está contemplando la reducción de la ratio estudiantes-maestros en 2. Debido a que el intervalo de confianza al 95 % para β_1 es $[-3,30, -1,26]$, el efecto de la reducción de la ratio estudiantes-maestros en 2 podría ser tan grande como $-3,30 \times (-2) = 6,60$ o tan pequeño como $-1,26 \times (-2) = 2,52$. Por tanto la disminución de la ratio estudiantes-maestros en 2 se prevé que aumente las calificaciones en los exámenes entre 2,52 y 6,60 puntos, con un nivel de confianza del 95 %.

5.3 Regresión cuando X es una variable binaria

Hasta ahora la discusión se ha centrado en el caso de que el regresor sea una variable continua. El análisis de regresión también puede ser utilizado cuando el regresor es binario, es decir, cuando solamente toma dos valores, 0 o 1. Por ejemplo, X puede ser el género de un trabajador ($= 1$ si es mujer, $= 0$ si es hombre), si un distrito escolar es urbano o rural ($= 1$ si es urbano, $= 0$ si es rural), o si el tamaño de las clases en el distrito es pequeño o grande ($= 1$ si es pequeño, $= 0$ si es grande). Una variable binaria se denomina asimismo **variable indicador** o a veces **variable ficticia** o **variable dummy**.

Interpretación de los coeficientes de regresión

La mecánica de la regresión con un regresor binario es la misma que si es continuo. Sin embargo, la interpretación de β_1 , es diferente, y resulta que la regresión con una variable binaria es equivalente a realizar un análisis de la diferencia de medias, tal como se describe en la Sección 3.4.

Para ver esto, supongamos que se tiene una variable D_i que es igual a 0 o 1, dependiendo de si la ratio estudiantes-maestros es menor que 20:

$$D_i = \begin{cases} 1 & \text{si la ratio estudiantes-maestros del distrito } i\text{-ésimo es } < 20 \\ 0 & \text{si la ratio estudiantes-maestros del distrito } i\text{-ésimo es } \geq 20. \end{cases} \quad (5.14)$$

El modelo de regresión poblacional con D_i como variable explicativa es

$$Y_i = \beta_0 + \beta_1 D_i + u_i, \quad i = 1, \dots, n. \quad (5.15)$$

Esto es lo mismo que el modelo de regresión con el regresor continuo X_i , salvo que ahora el regresor es la variable binaria D_i . Debido a que D_i no es continua, no resulta útil interpretar β_1 como una pendiente; de hecho, debido a que D_i solo puede tomar dos valores, no existe una «línea recta», por lo que no tiene sentido hablar de una pendiente. Por tanto no nos referiremos a β_1 como una pendiente en la Ecuación (5.15); sino que simplemente nos referiremos a β_1 como el **coeficiente que multiplica a D_i** en esta regresión o, de forma más compacta, el **coeficientes sobre D_i** .

Si en la Ecuación (5.15) β_1 no es una pendiente, ¿qué es? La mejor manera de interpretar β_0 y β_1 en una regresión con variable explicativa binaria consiste en examinar, uno a uno, los dos casos posibles, $D_i = 0$ y $D_i = 1$. Si la ratio estudiantes-maestros es alta, entonces $D_i = 0$ y la Ecuación (5.15) se convierte en

$$Y_i = \beta_0 + u_i \quad (D_i = 0). \quad (5.16)$$

Como $E(u_i|D_i) = 0$, la esperanza condicional de Y_i cuando $D_i = 0$ es $E(Y_i|D_i = 0) = \beta_0$, es decir, β_0 es el valor de la media poblacional de las calificaciones en los exámenes cuando la ratio estudiantes-maestros es alta. Del mismo modo, cuando $D_i = 1$,

$$Y_i = \beta_0 + \beta_1 + u_i \quad (D_i = 1). \quad (5.17)$$

Por tanto, cuando $D_i = 1$, $E(Y_i|D_i = 1) = \beta_0 + \beta_1$; es decir; $\beta_0 + \beta_1$ es el valor de la media poblacional de las calificaciones en los exámenes cuando el cociente estudiantes-maestros es bajo.

Debido a que $\beta_0 + \beta_1$ es la media poblacional de Y_i cuando $D_i = 1$ y β_0 es la media poblacional de Y_i cuando $D_i = 0$, la diferencia $(\beta_0 + \beta_1) - \beta_0 = \beta_1$ es la diferencia entre estas dos medias. En otras palabras, β_1 es la diferencia entre la esperanza condicional de Y_i cuando $D_i = 1$ y cuando $D_i = 0$, o $\beta_1 = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$. En el ejemplo de las calificaciones en los exámenes, β_1 es la diferencia entre la media de las calificaciones en los exámenes para los distritos con baja ratio de estudiantes por maestro y la media de las calificaciones en los exámenes para los distritos con alta ratio de estudiantes por maestro.

Debido a que β_1 es la diferencia de las medias poblacionales, tiene sentido que el estimador MCO de β_1 sea la diferencia entre las medias muestrales de Y_i entre los dos grupos, y, de hecho, este es el caso.

Contrastes de hipótesis e intervalos de confianza. Si las dos medias poblacionales son iguales, entonces β_1 en la Ecuación (5.15) es cero. Por tanto, se puede contrastar la hipótesis nula de que las dos medias poblacionales son iguales frente a la hipótesis alternativa de que son distintas, contrastando la hipótesis nula $\beta_1 = 0$ frente a la alternativa $\beta_1 \neq 0$. Esta hipótesis puede contrastarse mediante el procedimiento descrito en la Sección 5.1. En concreto, la hipótesis nula puede ser rechazada al nivel del 5 % frente a la alternativa bilateral si el estadístico t MCO, $t = \hat{\beta}_1/ES(\hat{\beta}_1)$ es mayor que 1,96 en valor absoluto. Del mismo modo, un intervalo de confianza al 95 % para β_1 , construido como $\hat{\beta}_1 \pm 1,96ES(\hat{\beta}_1)$ como se describe en la Sección 5.2, proporciona un intervalo de confianza al 95 % para la diferencia entre las dos medias poblacionales.

Aplicación a las calificaciones en los exámenes. A modo de ejemplo, una regresión de la calificación en los exámenes con la ratio estudiante-maestro como variable binaria D definida en la Ecuación (5.14) estimada por MCO a partir de las 420 observaciones de la Figura 4.2 obtiene

$$\overline{\text{CalificaciónExamen}} = 650,0 + 7,4D, R^2 = 0,037, ESR = 18,7, \\ (1,3) \quad (1,8) \quad (5.18)$$

donde los errores estándar de las estimaciones MCO de los coeficientes β_0 y β_1 se presentan entre paréntesis bajo las estimaciones MCO. Por tanto la puntuación media en los exámenes para la submuestra con ratios de estudiantes por maestro superiores o iguales a 20 (es decir para los cuales $D = 0$) es 650,0, y la puntuación media en los exámenes para la submuestra con ratios de estudiantes por maestro menores a 20 (por lo que $D = 1$) es $650,0 + 7,4 = 657,4$. La diferencia entre la media muestral de las calificaciones entre los dos grupos es 7,4. Esta es la estimación MCO de β_1 , el coeficiente sobre la variable binaria D , la ratio de estudiantes por maestro.

¿Es estadística y significativamente distinta de cero la diferencia entre las medias poblacionales entre los dos grupos al nivel del 5 %? Para averiguarlo, se formula el estadístico t para β_1 : $t = 7,4/1,8 = 4,04$. Este valor es mayor que 1,96 en valor absoluto, por lo que la hipótesis de que la media poblacional de las calificaciones en las pruebas es la misma entre los distritos con ratio de estudiantes por maestro alta y baja puede ser rechazada al nivel de significación del 5 %.

El estimador MCO y su error estándar pueden utilizarse para construir un intervalo de confianza al 95 % para la verdadera diferencia entre las medias. Es decir, $7,4 \pm 1,96 \times 1,8 = (3,9, 10,9)$. Este intervalo de confianza excluye $\beta_1 = 0$, de modo que (como sabemos por el párrafo anterior) la hipótesis $\beta_1 = 0$ puede ser rechazada al nivel de significación del 5 %.

5.4 Heterocedasticidad y homocedasticidad

El único supuesto realizado sobre la distribución de u_i condicionada a X_i es que tiene una media igual a cero (el primero de los supuestos de mínimos cuadrados). Si, además, la varianza de esta distribución condicional no depende de X_i , entonces se dice que los errores son homocedásticos. Esta sección trata la homocedasticidad, sus implicaciones teóricas, las fórmulas simplificadas de los errores estándar de los estimadores MCO que aparecen si los errores son homocedásticos, así como los riesgos que se corren al utilizar estas fórmulas simplificadas en la práctica.

¿Qué es la heterocedasticidad y la homocedasticidad?

Definiciones de heterocedasticidad y homocedasticidad. El término de error u_i es **homocedástico** si la varianza de la distribución condicional de u_i dado X_i es constante para $i = 1, \dots, n$ y en particular, no depende de X_i . De lo contrario, el término de error es **heterocedástico**.

A modo de ejemplo, volvamos a la Figura 4.4. En ella se muestra la distribución de los errores u_i para algunos valores de x . Debido a que esta distribución se aplica de forma específica a los valores de x señalados, esta es la distribución condicional de u_i dado $X_i = x$. Como se representa en esta figura, todas estas distribuciones condicionadas presentan la misma dispersión; de forma más concreta, la varianza de estas distribuciones es la misma para los distintos valores de x . Es decir, en la Figura 4.4, la varianza condicional de u_i dado $X_i = x$ no depende de x , por lo que los errores que se ilustran en la Figura 4.4 son homocedásticos.

Por contra, la Figura 5.2 ilustra el caso en el que la distribución condicionada de u_i se dispersa a medida que aumenta x . Para valores pequeños de x , la distribución se concentra, pero para valores mayores de x , presenta una dispersión mayor. Por tanto, en la Figura 5.2 la varianza de u_i dado $X_i = x$ aumenta con x , por lo que los errores de la Figura 5.2 son heterocedásticos.

Las definiciones de heterocedasticidad y homocedasticidad se recogen en el Concepto clave 5.4.

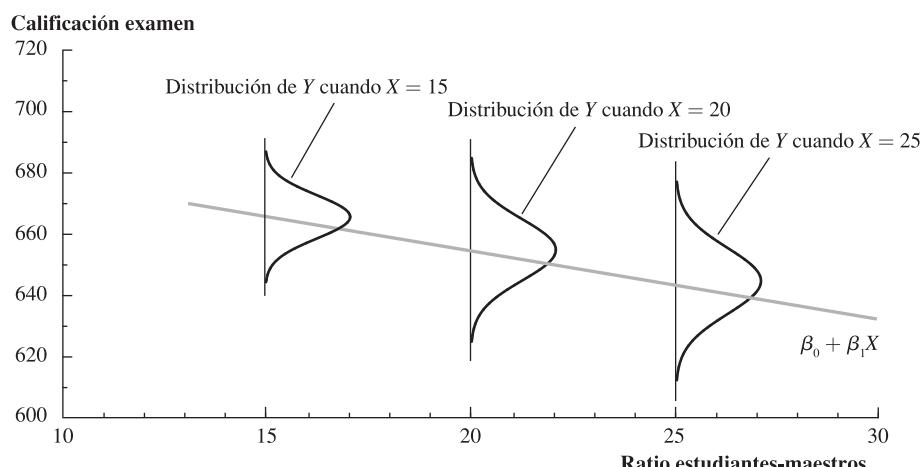
Ejemplo. Estos términos son un trabalenguas, y las definiciones pueden parecer abstractas. Para ayudar a aclararlas con un ejemplo, dejamos el problema de la ratio estudiantes-maestros/calificación en los exámenes, y en su lugar volvemos al ejemplo de los ingresos salariales de los titulados universitarios masculinos frente a los femeninos considerado en el recuadro del Capítulo 3, «La brecha de género en los ingresos salariales de los titulados universitarios en Estados Unidos». Sea $MASCULINO_i$ una variable binaria que es igual a 1 para los titulados universitarios varones e igual a 0 para las tituladas mujeres. El modelo de regresión con variable binaria correspondiente que relaciona los ingresos salariales de un titulado universitario con su género es

$$Ingresos_i = \beta_0 + \beta_1 MASCULINO_i + u_i \quad (5.19)$$

para $i = 1, \dots, n$. Debido a que el regresor es binario, β_1 es la diferencia entre las medias poblacionales de los dos grupos, en este caso, la diferencia entre los ingresos medios entre hombres y mujeres que se graduaron en la universidad.

FIGURA 5.2 Un ejemplo de heterocedasticidad

Al igual que en la Figura 4.4, se muestra la distribución condicional de las calificaciones en los exámenes para tres tamaños de clases diferentes. A diferencia de la Figura 4.4, estas distribuciones se van haciendo más dispersas (tienen mayor varianza) a mayor tamaño de las clases. Debido a que la varianza condicional de u , dado X , $\text{var}(u|X)$, depende de X , u es heterocedástico.


CONCEPTO CLAVE
5.4
Heterocedasticidad y homocedasticidad

El término de error u_i es homocedástico si la varianza de la distribución condicional de u_i dado X_i , $\text{var}(u_i|X_i = x)$, es constante para $i = 1, \dots, n$ y en particular, no depende de x . De lo contrario, el término de error es heterocedástico.

La definición de homocedasticidad establece que la varianza de u_i no depende del regresor. Aquí el regresor es $MASCULINO_i$, por lo que una cuestión importante es la de analizar si la varianza del término de error depende de $MASCULINO_i$. En otras palabras, ¿es igual la varianza del término de error para hombres y mujeres? Si es así, el error es homocedástico; si no, es heterocedástico.

Decidir si la varianza de u_i depende de $MASCULINO_i$ requiere una reflexión detenida acerca de lo que el término de error es en realidad. En este sentido, resulta útil escribir la Ecuación (5.19) como dos ecuaciones distintas, una para los hombres y otra para las mujeres:

$$\text{Ingresos}_i = \beta_0 + u_i \quad (\text{mujeres}) \quad (5.20)$$

$$\text{Ingresos}_i = \beta_0 + \beta_1 + u_i \quad (\text{hombres}) \quad (5.21)$$

Por lo tanto, para las mujeres, u_i es la desviación de los ingresos de la mujer i -ésima respecto de los ingresos poblacionales medios de las mujeres (β_0), y para los hombres, u_i es la desviación de los ingresos del i -ésimo hombre respecto de la media poblacional de los ingresos de los hombres ($\beta_0 + \beta_1$). De ello se desprende que la frase, «la varianza de u_i no depende de $MASCULINO$ », es equivalente a la frase, «la varianza de los ingresos es la misma para hombres y mujeres». En otras palabras, en este ejemplo, el término de error es homocedástico si la varianza de la distribución poblacional de los ingresos es la misma para hombres y mujeres; si estas varianzas son distintas, el término de error es heterocedástico.

Implicaciones matemáticas de la homocedasticidad

Los estimadores MCO siguen siendo insesgados y asintóticamente normales. Debido a que los supuestos de mínimos cuadrados del Concepto clave 4.3 no establecen restricciones sobre la varianza condicional, son aplicables tanto al caso general de heterocedasticidad como al caso particular de homocedasticidad. Por tanto, los estimadores MCO siguen siendo insesgados y consistentes, incluso si los errores son homocedásticos. Además, los estimadores MCO presentan distribuciones muestrales que son normales

en muestras grandes, incluso si los errores son homocedásticos. Tanto si los errores son homocedásticos como si son heterocedásticos, el estimador MCO es insesgado, consistente y asintóticamente normal.

Eficiencia del estimador MCO cuando los errores son homocedásticos. Si se cumplen los supuestos de mínimos cuadrados del Concepto clave 4.3 y los errores son homocedásticos, entonces los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$ son eficientes entre todos los estimadores que son lineales en Y_1, \dots, Y_n y son insesgados, condicionados a X_1, \dots, X_n . Este resultado, denominado teorema de Gauss-Markov, se trata en la Sección 5.5.

Fórmula de la varianza válida con homocedasticidad. Si el término de error es homocedástico, entonces las fórmulas de las varianzas de $\hat{\beta}_0$ y $\hat{\beta}_1$ del Concepto clave 4.4 se simplifican. En consecuencia, si los errores son homocedásticos, entonces existe una fórmula específica que puede utilizarse para los errores estándar de $\hat{\beta}_0$ y $\hat{\beta}_1$. El **error estándar válido con homocedasticidad** de $\hat{\beta}_1$, deducido en el Apéndice 5.1, es $ES(\hat{\beta}_1) = \sqrt{\tilde{\sigma}_{\hat{\beta}_1}^2}$, donde $\tilde{\sigma}_{\hat{\beta}_1}^2$ es el estimador de la varianza de $\hat{\beta}_1$ válido con homocedasticidad

$$\tilde{\sigma}_{\hat{\beta}_1}^2 = \frac{s_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{válido con homocedasticidad}), \quad (5.22)$$

dónde s_u^2 está dado en la Ecuación (4.19). La fórmula del error estándar de $\hat{\beta}_0$ válido con homocedasticidad se ofrece en el Apéndice 5.1. En el caso particular en que X es una variable binaria, el estimador de la varianza de $\hat{\beta}_1$ con homocedasticidad (es decir, el cuadrado del error estándar de $\hat{\beta}_1$ en presencia de homocedasticidad) es la denominada fórmula de la varianza agrupada para la diferencia de medias, dada en la Ecuación (3.23).

Debido a que estas fórmulas alternativas se obtienen para el caso particular en que los errores son homocedásticos y no son aplicables si los errores son heterocedásticos, nos referiremos a ellas como las fórmulas «válidas con homocedasticidad» para la varianza y el error estándar de los estimadores MCO. Tal y como el nombre sugiere, si los errores son heterocedásticos, entonces los errores estándar válidos con homocedasticidad no resultan adecuados. En concreto, si los errores son heterocedásticos, entonces el estadístico t calculado utilizando los errores estándar válidos con homocedasticidad no presenta una distribución normal estándar, incluso en muestras grandes. De hecho, los valores críticos adecuados a utilizar para este estadístico t válido con homocedasticidad dependen de la naturaleza exacta de la heterocedasticidad, por lo que los valores críticos no pueden ser tabulados. Del mismo modo, si los errores son heterocedásticos pero se construye un intervalo de confianza como $\pm 1,96$ veces el error estándar válido con homocedasticidad, en general, la probabilidad de que este intervalo contenga el verdadero valor del coeficiente no es del 95 %, incluso en muestras grandes.

Por el contrario, debido a que la homocedasticidad es un caso particular de heterocedasticidad, los estimadores $\hat{\sigma}_{\hat{\beta}_1}^2$ y $\hat{\sigma}_{\hat{\beta}_0}^2$ de las varianzas de $\hat{\beta}_1$ y $\hat{\beta}_0$ dadas en las Ecuaciones (5.4) y (5.26) dan lugar a inferencias estadísticas válidas tanto si los errores son heterocedásticos como si son homocedásticos. Por tanto, los contrastes de hipótesis e intervalos de confianza basados en esos errores estándar son válidos tanto si los errores son heterocedásticos como si no. Debido a que los errores estándar que hemos utilizado hasta ahora [es decir, los basados en las Ecuaciones (5.4) y (5.26)] conducen a inferencias estadísticas que son válidas tanto si los errores son heterocedásticos como si no lo son, se los denomina **errores estándar heterocedástico-robustos**. Debido a que estas fórmulas fueron propuestas por Eicker (1967), Huber (1967) y White (1980), se les conoce asimismo como errores estándar Eicker-Huber-White.

¿Qué significa esto en la práctica?

¿Es más realista la heterocedasticidad o la homocedasticidad? La respuesta a esta pregunta depende de su aplicación a un caso concreto. Sin embargo, esta cuestión puede ser aclarada volviendo al ejemplo de la brecha de género en los ingresos entre los titulados universitarios. El conocimiento del salario que cobra la gente en el mundo que nos rodea proporciona algunas pistas acerca de cual es el supuesto más

El valor económico de un año de educación: ¿homocedasticidad o heterocedasticidad?

En promedio, los trabajadores con mayor grado de educación alcanzado tienen mayores ingresos salariales que los trabajadores que han alcanzado un menor grado educativo. Pero si los trabajos mejor pagados, principalmente son para los universitarios, podría ocurrir asimismo que la *dispersión* de la distribución de los ingresos salariales sea mayor para los trabajadores con mayor nivel de educación alcanzado. ¿Aumenta su dispersión la distribución de los ingresos a medida que aumenta la educación?

Esta es una pregunta empírica, por lo que para obtener una respuesta es necesario analizar los datos. La Figura 5.3 es un diagrama de dispersión de los ingresos salariales por hora y el número de años de educación para una muestra de 2.989 trabajadores a tiempo completo, entre 29 y 30 años de edad, con entre 6 y 18 años de educación en los Estados Unidos en 2008. Los datos provienen de la Encuesta Actualizada de Población de marzo de 2009 (CPS), que se describe en el Apéndice 3.1.

La Figura 5.3 tiene dos características notables. La primera es que la media de la distribución de los ingresos salariales aumenta con el número de años de educación. Este aumento se resume en la recta de regresión MCO,

$$\widehat{\text{Ingresos}} = -5,38 + 1,76 \text{ AñosEducación},$$

$$(1,05) \quad (0,8)$$

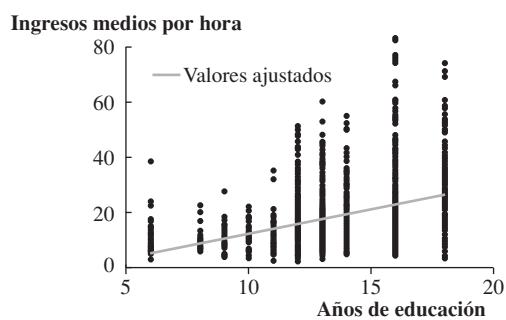
$$R^2 = 0,159, ESR = 9,50. \quad (5.23)$$

Esta recta se representa en la Figura 5.3. El coeficiente de 1,76 de la recta de regresión MCO significa que, en promedio, los ingresos salariales por hora aumentan en 1,76 \$ por cada año adicional de educación. El intervalo de confianza al 95 % para este coeficiente es $1,76 \pm 1,96 \times 0,08$, o de 1,60 a 1,91.

La segunda característica notable de la Figura 5.3 es que la dispersión de la distribución de los ingresos aumenta con los años de educación. Mientras que algunos trabajadores con muchos años de educación completados tienen empleos con baja remuneración, muy pocos trabajadores con niveles de educación bajos ocupan empleos bien remunerados. Esto puede cuantificarse fijándose en la dispersión de los residuos alrededor de la recta de regresión MCO. Para los trabajadores con diez años de educación, la desviación típica de los residuos es 4,34 \$; para los trabajadores con un diploma de escuela secundaria, esta desviación típica es de 7,30 \$; y para los trabajadores con título universitario esta desviación típica se eleva a 12,25 \$. Debido a que estas desviaciones típicas son diferentes para diferentes niveles de educación, la varianza de los residuos en la regresión de la Ecuación (5.23) depende del valor del regresor (los años de educación); en otras palabras, los errores de regresión son heterocedásticos. En términos del mundo real, no todos los titulados universitarios ganan 50 \$ por hora a los 29 años, pero algunos sí, y los trabajadores con tan solo diez años de educación no aspiran a esos puestos de trabajo.

FIGURA 5.3 Diagrama de dispersión de los ingresos salariales por hora y años de educación para edades entre 29 y 30 años en Estados Unidos en 2008

Se representan los ingresos por hora frente a los años de educación para 2,989 trabajadores a tiempo completo, entre 29 y 30 años de edad. La dispersión alrededor de la recta de regresión aumenta con los años de educación, lo que indica que los errores de regresión son heterocedásticos.



sensato. Durante muchos años —y sin tener que extendernos tanto, en la actualidad— las mujeres no se encontraban ocupando los puestos de trabajo mejor pagados: siempre ha habido hombres mal pagados, pero rara vez ha habido mujeres muy bien pagadas. Esto sugiere que la distribución de los ingresos salariales entre las mujeres está más concentrada que para los hombres (véase el recuadro del Capítulo 3, «La brecha de género en los ingresos salariales de los titulados universitarios en Estados Unidos»). En otras palabras, resulta verosímil que la varianza del término de error de la Ecuación (5.20) para las mujeres sea menor que la varianza del término de error de la Ecuación (5.21) para los hombres. Por tanto, la presencia de un

«techo de cristal» para los salarios y los empleos de las mujeres sugiere que el término de error en el modelo de regresión con variable binaria de la Ecuación (5.19) es heterocedástico. A menos que haya razones de peso para lo contrario —y no podemos pensar en ninguna— tiene sentido tratar el término de error en este ejemplo como heterocedástico.

Tal y como ilustra este ejemplo de modelización de los ingresos salariales, la heterocedasticidad se presenta en muchas aplicaciones econométricas. A nivel general, la teoría económica rara vez proporciona razones para creer que los errores son homocedásticos. Por lo tanto, resulta sensato suponer que los errores pueden ser heterocedásticos a menos que existan razones de peso para creer lo contrario.

Implicaciones prácticas. La cuestión principal con relevancia práctica en esta cuestión es si se deben utilizar los errores estándar heterocedástico-robustos o los errores estándar válidos con homocedasticidad. En este sentido, resulta útil pensar en calcular ambos, y posteriormente elegir entre ellos. Si los errores estándar válidos con homocedasticidad y los heterocedástico-robustos son iguales, no se pierde nada al utilizar los errores estándar heterocedástico-robustos; no obstante, si son distintos, deberían utilizarse los más fiables, los que toleran la presencia de heterocedasticidad. Lo más sencillo entonces resulta utilizar siempre los errores estándar heterocedástico-robustos.

Por motivos históricos, muchos programas de *software* proporcionan solamente errores estándar válidos con homocedasticidad como configuración predeterminada, por lo que es necesario que el usuario especifique la opción de errores estándar heterocedástico-robustos. Los detalles sobre cómo llevar a cabo el cálculo de los errores estándar heterocedástico-robustos dependen del paquete de *software* que se utilice.

En todos los ejemplos empíricos de este libro se emplean errores estándar heterocedástico-robustos a menos que explícitamente se establezca lo contrario¹.

*5.5 Fundamentos teóricos de mínimos cuadrados ordinarios

Como se analizó en la Sección 4.5, el estimador MCO es insesgado, es consistente, tiene una varianza que es inversamente proporcional a n , y tiene una distribución muestral normal cuando el tamaño muestral es grande. Además, bajo ciertas condiciones el estimador MCO es más eficiente que algunos otros candidatos a estimadores. En concreto, si se cumplen los supuestos de mínimos cuadrados y si los errores son homocedásticos, el estimador MCO tiene la menor varianza entre todos los estimadores condicionalmente insesgados que son funciones lineales de Y_1, \dots, Y_n . En esta sección se explica y analiza este resultado, que es una consecuencia del teorema de Gauss-Markov. Esta sección concluye con un análisis de los estimadores alternativos que son más eficientes que los MCO cuando no se cumplen las condiciones del teorema de Gauss-Markov.

Estimadores lineales condicionalmente insesgados y teorema de Gauss-Markov

Si los tres supuestos de mínimos cuadrados (Concepto clave 4.3) se cumplen y si el error es homocedástico, entonces el estimador MCO tiene la menor varianza, condicionada a X_1, \dots, X_n , de entre todos los estimadores de la clase de estimadores lineales condicionalmente insesgados. En otras palabras, el estimador MCO es el **Estimador Lineal condicionalmente Insesgado Óptimo**, es decir, es ELIO². Este resultado es una extensión del resultado que se recoge en el Concepto clave 3.3, de que la media muestral \bar{Y} es el estimador más eficiente de la media poblacional de entre la clase de todos los estimadores que son insesgados y son funciones lineales (medias ponderadas) de Y_1, \dots, Y_n .

¹ En el caso de que este libro se utilice junto con otros, podría ser útil tener en cuenta que algunos libros de texto añaden la homocedasticidad a la lista de los supuestos de mínimos cuadrados. Sin embargo, como acabamos de comentar, este supuesto adicional no es necesario para la validez del análisis de regresión MCO, siempre y cuando se utilicen errores estándar heterocedástico-robustos.

* Esta sección es opcional y no se utiliza en los capítulos posteriores.

² *N. del T.*: Best Lineal condicional Unbiased Estimator (BLUE) en su denominación original (BLUE).

**CONCEPTO
CLAVE**

5.5

El teorema de Gauss-Markov para $\hat{\beta}_1$

Si se cumplen los tres supuestos de mínimos cuadrados del Concepto clave 4.3 y si además, los errores son homocedásticos, entonces el estimador MCO $\hat{\beta}_1$ es el Estimador Lineal condicionalmente Insesgado Óptimo (más eficiente) (es **ELIO**)³.

Estimadores lineales condicionalmente insesgados. La clase de los estimadores lineales condicionalmente insesgados está compuesta por todos los estimadores de β_1 que son funciones lineales de Y_1, \dots, Y_n y que son insesgados, condicionado a X_1, \dots, X_n . Es decir, si $\tilde{\beta}_1$ es un estimador lineal, entonces puede escribirse como

$$\tilde{\beta}_1 = \sum_{i=1}^n a_i Y_i \quad (\tilde{\beta}_1 \text{ es lineal}), \quad (5.24)$$

donde las ponderaciones a_1, \dots, a_n pueden depender de X_1, \dots, X_n pero no de Y_1, \dots, Y_n . El estimador $\tilde{\beta}_1$ es condicionalmente insesgado si la media de su distribución muestral condicional, dado X_1, \dots, X_n es β_1 . Es decir, el estimador $\tilde{\beta}_1$ es condicionalmente insesgado si

$$E(\tilde{\beta}_1 | X_1, \dots, X_n) = \beta_1 \quad (\tilde{\beta}_1 \text{ es condicionalmente insesgado}). \quad (5.25)$$

El estimador $\tilde{\beta}_1$ es un estimador condicionalmente insesgado si puede escribirse de la forma de la Ecuación (5.24) (es lineal) y si la Ecuación (5.25) se cumple (es condicionalmente insesgado). En el Apéndice 5.2 se demuestra que el estimador MCO es lineal y condicionalmente insesgado.

El teorema de Gauss-Markov. El **teorema de Gauss-Markov** establece que, bajo un conjunto de condiciones conocidas como las condiciones de Gauss-Markov, el estimador MCO $\hat{\beta}_1$ tiene la menor varianza condicional, dados X_1, \dots, X_n , de todos los estimadores lineales condicionalmente insesgados de β_1 ; es decir, el estimador MCO es ELIO. Las condiciones del teorema de Gauss-Markov, que se establecen en el Anexo 5.2, están implícitas en los tres supuestos de mínimos cuadrados más el supuesto de que los errores son homocedásticos. En consecuencia, si los tres supuestos de mínimos cuadrados se cumplen y los errores son homocedásticos, entonces MCO es ELIO. El teorema de Gauss-Markov se enuncia en el Concepto clave 5.5 y se demuestra en el Apéndice 5.2.

Limitaciones del teorema de Gauss-Markov. El teorema de Gauss-Markov proporciona una justificación teórica al uso de MCO. No obstante, el teorema tiene dos limitaciones importantes. En primer lugar, sus condiciones podrían no cumplirse en la práctica. En particular, si el término de error es heterocedástico, tal y como sucede a menudo en las aplicaciones económicas, el estimador MCO ya no es ELIO. Tal y como se trató en la Sección 5.4, la presencia de heterocedasticidad no representa una amenaza para la inferencia basada en errores estándar heterocedástico-robustos, pero lo que sí significa es que MCO ya no es el estimador eficiente entre los lineales y condicionalmente insesgados. Existe un estimador alternativo a MCO cuando existe heterocedasticidad con forma conocida, llamado estimador de mínimos cuadrados ponderados, que se discute a continuación.

La segunda limitación del teorema de Gauss-Markov es que incluso si las condiciones del teorema se cumplen, existen otros posibles estimadores que no son lineales y condicionalmente insesgados; bajo ciertas condiciones, estos estimadores son más eficientes que MCO.

Estimadores de regresión alternativos a MCO

Bajo ciertas condiciones, algunos estimadores de regresión son más eficientes que MCO.

El estimador de mínimos cuadrados ponderados. Si los errores son heterocedásticos, entonces MCO ya no es ELIO. Si la naturaleza de la heterocedasticidad es conocida, en concreto, si la varianza con-

³ N. Del T.: BLUE es el correspondiente en el original en inglés.

dicional de u_i dado X_i es conocida con un factor constante de proporcionalidad, entonces es posible obtener un estimador que presente una menor varianza que el estimador MCO. Este método, denominado de **mínimos cuadrados ponderados** (MCP), pondera la i -ésima observación por la inversa de la raíz cuadrada de la varianza condicional de u_i dado X_i . Debido a esta ponderación, los errores de esta regresión ponderada son homocedásticos, por lo que MCO, cuando se aplican a los datos ponderados, es ELIO. Aunque impecable desde el punto de vista teórico, el problema práctico de los mínimos cuadrados ponderados es que es necesario conocer cómo la varianza condicional de u_i depende de X_i , algo que raramente se conoce en las aplicaciones económicas. Por tanto, los mínimos cuadrados ponderados se utilizan con mucha menos frecuencia que MCO, y el análisis más detallado se postpone al Capítulo 17.

El estimador de mínima desviación absoluta. Como se trató en la Sección 4.3, el estimador MCO puede ser sensible a los valores atípicos. Si los valores atípicos extremos no son infrecuentes, entonces otros estimadores pueden ser más eficientes que los MCO y pueden llevarse a cabo inferencias que resulten más fiables. Un estimador así es el estimador de mínima desviación absoluta (MDA), en el que los coeficientes de regresión β_0 y β_1 se obtienen resolviendo un problema de minimización como el de la Ecuación (4.6), salvo que se utiliza el valor absoluto del «error» de predicción en lugar de su cuadrado. Es decir, el estimador MDA de β_0 y β_1 son los valores de b_0 y b_1 que minimizan $\sum_{i=1}^n |Y_i - b_0 - b_1 X_i|$. El estimador MDA es menos sensible a los valores atípicos grandes de u que MCO.

En muchas bases de datos económicos, los valores extremos de u son escasos, por lo que el uso de los estimadores MDA, u otros con baja sensibilidad a los valores atípicos, es poco frecuente en las aplicaciones. Por tanto, el tratamiento de la regresión lineal en el resto de este texto se centra exclusivamente en los métodos de mínimos cuadrados.

*5.6 La utilización del estadístico t en regresión para muestras pequeñas

Cuando el tamaño de la muestra es pequeño, la distribución exacta del estadístico t es compleja y depende de la distribución poblacional de los datos que es desconocida. Si, no obstante, los tres supuestos de mínimos cuadrados se cumplen, los errores de regresión son homocedásticos, y además los errores de regresión se distribuyen normalmente, entonces el estimador MCO se distribuye normalmente y el estadístico t válido con homocedasticidad presenta una distribución t de Student. Estos cinco supuestos, los tres supuestos de mínimos cuadrados, que los errores son homocedásticos, y que los errores se distribuyen normalmente, se conocen colectivamente como los **supuestos de la regresión normal homocedástica**.

El estadístico t y la distribución de t de Student

Recordemos de la Sección 2.4 que la distribución t de Student con m grados de libertad se define como la distribución de $Z/\sqrt{W/m}$, donde Z es una variable aleatoria con una distribución normal estándar, W es una variable aleatoria con una distribución chi-cuadrado con m grados de libertad, y Z y W son independientes. Bajo la hipótesis nula, el estadístico t calculado utilizando el error estándar válido con homocedasticidad puede escribirse de esta forma.

El estadístico t válido con homocedasticidad para contrastar $\beta_1 = \beta_{1,0}$ es $\tilde{t} = (\hat{\beta}_1 - \beta_{1,0})/\tilde{\sigma}_{\hat{\beta}_1}$, donde $\tilde{\sigma}_{\hat{\beta}_1}^2$ está definido en la Ecuación (5.22). Bajo los supuestos de la regresión normal homocedástica, Y presenta una distribución normal, condicionada a X_1, \dots, X_n . Como ya se ha analizado en la Sección 5.5, el estimador MCO es una media ponderada de Y_1, \dots, Y_n , donde las ponderaciones dependen de X_1, \dots, X_n [véase la Ecuación (5.32) en el Apéndice 5.2]. Debido a que una media ponderada de variables aleatorias normales independientes se distribuye asimismo normalmente, $\hat{\beta}_1$ tiene una distribución normal condicionada a X_1, \dots, X_n . Por tanto, $(\hat{\beta}_1 - \beta_{1,0})$ presenta una distribución normal bajo la hipótesis nula, condicionada a X_1, \dots, X_n .

* Esta sección es opcional y no se utiliza en los capítulos posteriores.

Además, el estimador (normalizado) de la varianza válido con homocedasticidad tiene una distribución chi-cuadrado con $n - 2$ grados de libertad, dividido por $n - 2$, y $\tilde{\sigma}_{\hat{\beta}_1}^2$ y $\hat{\beta}_1$ están independientemente distribuidos. En consecuencia, el estadístico t válido con la homocedasticidad presenta una distribución t de Student con $n - 2$ grados de libertad.

Este resultado está estrechamente relacionado con un resultado analizado en la Sección 3.5 en el contexto del contraste para la igualdad de las medias de dos muestras. En ese problema, si las dos distribuciones poblaciones son normales con la misma varianza, y si el estadístico t se construye con la fórmula del error estándar agrupado [Ecuación (3.23)], entonces el estadístico t (agrupado) presenta una distribución t de Student. Cuando X es binaria, el error estándar válido con homocedasticidad para $\hat{\beta}_1$ se simplifica hasta la fórmula del error estándar agrupado para la diferencia de medias. De ello se deduce que el resultado de la Sección 3.5 es un caso particular del resultado de que si los supuestos de la regresión homocedástica normal se cumplen, entonces el estadístico t de la regresión válido con homocedasticidad presenta una distribución t de Student (véase el Ejercicio 5.10).

La utilización de la distribución t de Student en la práctica

Si los errores de la regresión son homocedásticos y se distribuyen normalmente y si se utiliza el estadístico t válido con homocedasticidad, entonces los valores críticos que deben tomarse son los de la distribución t de Student (Tabla 2 del Apéndice) en lugar de los de la distribución normal estándar. Debido a que la diferencia entre la distribución t de Student y la distribución normal es insignificante si n es mediano o grande, esta distinción solo es relevante si el tamaño de la muestra es pequeño.

En las aplicaciones econométricas, rara vez existe una razón para creer que los errores sean homocedásticos y estén normalmente distribuidos. Debido a que los tamaños muestrales por lo general son grandes, no obstante, en el proceso de inferencia puede procederse como se describe en las Secciones 5.1 y 5.2, es decir, en primer lugar, calculando los errores estándar heterocedástico-robustos y más tarde utilizar la distribución normal estándar para calcular los p -valores, los contrastes de hipótesis, y los intervalos de confianza.

5.7 Conclusión

Volvamos por un momento al problema con el que comenzaba el Capítulo 4: la directora que estaba considerando la contratación de más maestros para reducir la ratio estudiantes-maestros. ¿Qué hemos aprendido que pueda resultar útil?

Nuestro análisis de regresión, sobre la base de las 420 observaciones del año 1998 de la base de datos de calificaciones en los exámenes de California, mostraba que existía una relación negativa entre la ratio estudiantes-maestros y las calificaciones en los exámenes: los distritos con clases más pequeñas tienen mejores calificaciones en los exámenes. El coeficiente es moderadamente elevado, en un sentido práctico: los distritos con dos alumnos menos por maestro presentan, en media, calificaciones en los exámenes que son 4,6 puntos superiores. Esto se corresponde con el traslado de un distrito desde el percentil 50 de la distribución de las calificaciones en los exámenes hasta aproximadamente el percentil 60.

El coeficiente sobre la ratio estudiantes-maestros es estadística y significativamente distinto de 0 al 5 % de nivel de significación. El coeficiente poblacional podría ser 0, y podríamos simplemente haber estimado nuestro coeficiente negativo debido a la variación del muestreo aleatorio. Sin embargo, la probabilidad de haberlo hecho (y de haber obtenido un estadístico t para β_1 tan alto como lo hicimos nosotros) debido únicamente a la variación aleatoria de las muestras potenciales es muy pequeña, aproximadamente el 0,001 %. Un intervalo de confianza al 95 % para β_1 es $-3,30 \leq \beta_1 \leq -1,26$.

Este resultado representa un avance considerable hacia la respuesta a la pregunta de la directora que todavía sigue siendo una preocupación acuciante. Existe una relación negativa entre la ratio estudiantes-maestros y las calificaciones en los exámenes, pero ¿es esta relación necesariamente la relación causal que la directora necesita para tomar su decisión? Los distritos con menores ratios de estudiantes por maestro obtienen, en promedio, mejores resultados en la prueba. Pero ¿significa esto que la reducción de la ratio estudiantes-maestros, en realidad mejora los resultados?

De hecho, existen razones para temer que tal vez no. La contratación de más profesores, después de todo, cuesta dinero, por lo que los distritos escolares más ricos pueden darse el lujo de clases más pequeñas. Pero los estudiantes de las escuelas más ricas también tienen otras ventajas sobre sus vecinos más pobres, incluyendo mejores instalaciones, libros más nuevos, y maestros mejor pagados. Por otra parte, los estudiantes de las escuelas más ricas suelen provenir de familias más acomodadas y, por lo tanto, tienen otras ventajas que no están directamente relacionadas con su escuela. Por ejemplo, California tiene una comunidad de inmigrantes grande; esos inmigrantes suelen ser más pobres que la población en general, y, en muchos casos, sus hijos no son hablantes nativos de inglés. Por tanto, podría ser que nuestra relación negativa estimada entre las calificaciones en los exámenes y la ratio estudiantes-maestros sea una consecuencia de clases más grandes junto con muchos otros factores que son, de hecho, la causa real de las puntuaciones más bajas en los exámenes.

La existencia de estos otros factores, o «variables omitidas», podría significar que el análisis MCO hasta ahora realizado tiene poco valor para la directora. De hecho, podría resultar engañoso: la varición aislada de la ratio estudiantes-maestros por sí sola no cambiaría estos otros factores que determinan el desempeño del niño en la escuela. Para abordar este problema, necesitamos un método que nos permita aislar el efecto sobre las calificaciones en los exámenes de la variación en la ratio de estudiantes-maestros, manteniendo los otros factores constantes. Ese método es el análisis de regresión múltiple, el tema de los Capítulos 6 y 7.

Resumen

1. El contraste de hipótesis sobre los coeficientes de regresión es análogo al contraste de hipótesis sobre la media poblacional: utiliza el estadístico t para calcular los p -valores y o bien aceptar o bien rechazar la hipótesis nula. Al igual que un intervalo de confianza para la media poblacional, un intervalo de confianza al 95 % para un coeficiente de regresión se calcula como el estimador $\pm 1,96$ veces el error estándar.
2. Cuando X es binaria, el modelo de regresión puede utilizarse para estimar y contrastar la hipótesis sobre la diferencia entre las medias poblacionales del grupo « $X = 0$ » y el grupo « $X = 1$ ».
3. En general, el error u_i es heterocedástico, es decir, la varianza de u_i para un valor dado de X_i , $\text{var}(u_i|X_i = x)$ depende de x . Un caso particular es cuando el error es homocedástico, es decir, $\text{var}(u_i|X_i = x)$ es constante. Los errores estándar válidos con homocedasticidad no dan lugar a inferencias estadísticas válidas si los errores son heterocedásticos, pero sí los errores estándar heterocedástico-robustos.
4. Si se cumplen los tres supuestos de mínimos cuadrados y además los errores de regresión son homocedásticos, entonces, como consecuencia del teorema de Gauss-Markov, el estimador de MCO es ELIO.
5. Si los tres supuestos de mínimos cuadrados se cumplen, y los errores de regresión son homocedásticos, y *si además* los errores de regresión están normalmente distribuidos, entonces el estadístico t MCO, calculado utilizando los errores estándar válidos con homocedasticidad, presenta una distribución t de Student cuando la hipótesis nula es verdadera. La diferencia entre la distribución t de Student y la distribución normal es insignificante si el tamaño de la muestra es mediano o grande.

Términos clave

hipótesis nula (104)	coeficiente sobre D_i (110)
hipótesis alternativa bilateral (104)	heterocedasticidad y homocedasticidad (111)
error estándar de $\hat{\beta}_1$ (104)	errores estándar válidos con homocedasticidad (113)
estadístico t (105)	errores estándar heterocedástico-robustos (113)
p -valor (105)	teorema de Gauss-Markov (116)
intervalo de confianza para β_1 (108)	estimador lineal insesgado óptimo (ELIO) (116)
nivel de confianza (108)	mínimos cuadrados ponderados (117)
variable indicador (109)	supuestos de la regresión normal
variable ficticia o dummy (109)	homocedástica (117)
coeficiente que multiplica a D_i (110)	condiciones de Gauss-Markov (125)

Revisión de conceptos

- 5.1** Describa los procedimientos para calcular el p -valor de un contraste bilateral de $H_0: \mu_Y = 0$ utilizando un conjunto de observaciones i.i.d. $Y_i, i = 1, \dots, n$. Resuma los procedimientos para calcular el p -valor de un contraste bilateral de $H_0: \beta_1 = 0$ en un modelo de regresión utilizando un conjunto de observaciones i.i.d. $(Y_i, X_i) i = 1, \dots, n$.
- 5.2** Explique cómo se puede utilizar un modelo de regresión para estimar la brecha de género en los salarios por hora utilizando los datos sobre los ingresos salariales de hombres y mujeres. ¿Cuáles son las variables dependientes y las variables independientes?
- 5.3** Defina *homocedasticidad* y *heterocedasticidad*. Proporcione un ejemplo empírico hipotético para el que crea que los errores son heterocedásticos y explique el razonamiento.

Ejercicios

- 5.1** Supóngase que un investigador, con datos sobre el tamaño de la clase (TC) y el promedio de las calificaciones en las pruebas para 100 clases de tercer curso, estima la regresión MCO

$$\widehat{\text{CalificaciónExamen}} = 520,4 - 5,82 \times TC, R^2 = 0,08, ESR = 11,5 \\ (20,4) \quad (2,21)$$

- a) Construya un intervalo de confianza al 95 % para β_1 , el coeficiente de la pendiente de la regresión.
 - b) Calcule el p -valor para el contraste bilateral de la hipótesis nula $H_0: \beta_1 = 0$. ¿Rechazaría la hipótesis nula al nivel del 5 %? ¿Y al nivel del 1 %?
 - c) Calcule el p -valor para el contraste bilateral de la hipótesis nula $H_0: \beta_1 = -5,6$. Sin realizar ningún cálculo adicional, determine si $-5,6$ está contenido en el intervalo de confianza al 95 % para β_1 .
 - d) Construya un intervalo de confianza al 99 % para β_0 .
- 5.2** Supóngase que un investigador, con datos salariales sobre 250 trabajadores y 280 trabajadoras seleccionados aleatoriamente, estima la regresión MCO

$$\widehat{\text{Salario}} = 12,52 + 2,12 \times \text{Masculino}, R^2 = 0,06, ESR = 4,2, \\ (0,23) \quad (0,36)$$

donde *Salario* se mide en dólares por hora y *Masculino* es una variable binaria que es igual a 1 si la persona es un varón y 0 si la persona es una mujer. Defina la brecha salarial por género como la diferencia de ingresos salariales medios entre hombres y mujeres.

- a) ¿Cuál es la brecha de género estimada?
- b) ¿Es la brecha de género estimada significativamente distinta de cero? (Calcule el p -valor para el contraste de la hipótesis nula de que no existe brecha de género).
- c) Construya un intervalo de confianza al 95 % para la brecha de género.
- d) En la muestra, ¿cuál es el salario medio de las mujeres? ¿Y de los hombres?
- e) Otro investigador utiliza estos mismos datos pero regresa la variable *Salario* sobre la variable *Femenino*, una variable que es igual a 1 si la persona es una mujer y 0 si la persona es un hombre. ¿Cuáles son las estimaciones de la regresión calculadas a partir de esta regresión?

$$\widehat{\text{Salario}} = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} \times \text{Femenino}, R^2 = \underline{\hspace{2cm}}, ESR = \underline{\hspace{2cm}}.$$

- 5.3** Supóngase que se selecciona una muestra aleatoria de 200 hombres de veinte años de edad de una población y se registran su peso y estatura. Una regresión del peso sobre la altura da como resultado

$$\widehat{\text{Peso}} = -99,41 + 3,94 \times \text{Altura}, R^2 = 0,81, ESR = 10,2, \\ (2,15) \quad (0,31)$$

donde *Peso* se mide en libras y *Altura* se mide en pulgadas. Un hombre da un estirón tardío y crece 1,5 pulgadas en el transcurso de un año. Construya un intervalo de confianza al 99 % para el aumento de peso de esta persona.

- 5.4** Véase el recuadro «El valor económico de un año de educación: homocedasticidad o heterocedasticidad?» de la Sección 5.4. Utilice la regresión proporcionada por la Ecuación (5.23) para responder a lo siguiente.

- Un trabajador seleccionado al azar de 30 años de edad, presenta un nivel de educación de 16 años. ¿Cuál es la esperanza del promedio de los ingresos salariales para ese trabajador?
- Un graduado de secundaria (12 años de educación) está contemplando acudir a un centro universitario de primer ciclo para obtener un título (2 años). ¿Cuánto se espera que aumente la media del salario por hora de este trabajador?
- Un consejero de la escuela secundaria dice a un estudiante que, en promedio, los graduados universitarios ganan 10 \$ por hora más que los graduados en la escuela secundaria. ¿Es esta afirmación congruente con la evidencia que proporciona la regresión? ¿Qué rango de valores es congruente con la evidencia de la regresión?

- 5.5** En la década de 1980, Tennessee llevó a cabo un experimento en el que los estudiantes de guardería fueron asignados aleatoriamente a clases de distintos tamaños, «normal» o «pequeño», realizándose a final del curso unos exámenes o pruebas estandarizadas. (Las clases normales constaban aproximadamente de 24 estudiantes, y las clases pequeñas constaban aproximadamente de 15 estudiantes). Supóngase que, en la población, las pruebas estandarizadas arrojan una puntuación media de 925 puntos y una desviación típica de 75 puntos. Sea *ClasePequeña* la denominación de una variable binaria igual a 1 si el estudiante es asignado a una clase pequeña e igual a 0 en cualquier otro caso. Una regresión de la variable *CalificaciónExamen* sobre *ClasePequeña* proporciona estos resultados:

$$\widehat{\text{CalificaciónExamen}} = 918,0 + 13,9 \times \text{ClasePequeña}, R^2 = 0,01, ESR = 74,6$$

$$(1,6) \quad (2,5)$$

- ¿Mejoran las clases pequeñas los resultados en la prueba? ¿En cuánto? ¿Es grande el efecto? Explíquelo.
 - ¿Es estadísticamente significativo el efecto estimado del tamaño de las clases sobre las calificaciones obtenidas? Realice un contraste al 5 % de nivel.
 - Construya un intervalo de confianza al 99 % para el efecto de *ClasePequeña* sobre las calificaciones en las pruebas.
- 5.6** Respecto a la regresión descrita en el Ejercicio 5.5.
- ¿Cree usted verosímil que los errores de regresión sean homocedásticos? Explíquelo.
 - El $ES(\hat{\beta}_1)$ se calculó mediante la Ecuación (5.3). Supóngase que los errores de regresión fueran homocedásticos: ¿afectaría esto a la validez del intervalo de confianza del Ejercicio 5.5(c)? Explíquelo.
- 5.7** Supóngase que (X_i, Y_i) satisfacen los supuestos del Concepto clave 4.3. Se extrae una muestra aleatoria de tamaño $n = 250$ que arroja los siguientes resultados:

$$\hat{Y} = 5,4 + 3,2X, R^2 = 0,26, ESR = 6,2$$

$$(3,1) \quad (1,5)$$

- Contraste $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ al nivel del 5 %.
- Construya un intervalo de confianza al 95 % para β_1 .
- Supóngase que se averigua que Y_i y X_i son independientes. ¿Le sorprendería? Explíquelo.
- Supóngase que Y_i y X_i son independientes y se extraen muchas muestras de tamaño $n = 250$, se estiman las regresiones, y se responde a (a) y (b). ¿En qué proporción de las muestras se rechazaría la H_0 de (a)? ¿En qué proporción de las muestras estaría incluido el valor $\beta_1 = 0$ en el intervalo de confianza de (b)?

- 5.8** Supóngase que (Y_i, X_i) satisfacen los supuestos del Concepto clave 4.3 y, además, u_i es $N(0, \sigma_u^2)$ y es independiente de X_i . Una muestra de tamaño $n = 30$ da como resultado

$$\hat{Y} = 43,2 + 61,5X, R^2 = 0,54, ESR = 1,52, \\ (10,2) \quad (7,4)$$

donde los números entre paréntesis son los errores estándar válidos con homocedasticidad de los coeficientes de regresión.

- a) Construya un intervalo de confianza al 95 % para β_0 .
- b) Contraste $H_0: \beta_1 = 55$ vs. $H_1: \beta_1 \neq 55$ al nivel del 5 %.
- c) Contraste $H_0: \beta_1 = 55$ vs. $H_1: \beta_1 > 55$ al nivel del 5 %.

- 5.9** Considérese el modelo de regresión

$$Y_i = \beta X_i + u_i,$$

donde u_i y X_i satisfacen los supuestos del Concepto clave 4.3. Sea $\bar{\beta}$ un estimador de β que se construye como $\bar{\beta} = \bar{Y}/\bar{X}$, donde \bar{Y} y \bar{X} son las medias muestrales de Y_i y X_i , respectivamente.

- a) Demuestre que $\bar{\beta}$ es una función lineal de Y_1, Y_2, \dots, Y_n .
 - b) Demuestre que $\bar{\beta}$ es condicionalmente insesgado.
- 5.10** Sea X_i una variable binaria y considérese la regresión $Y_i = \beta_0 + \beta_1 X_i + u_i$. Sea \bar{Y}_0 la media muestral de las observaciones con $X = 0$ e \bar{Y}_1 la media muestral para las observaciones con $X = 1$. Demuestre que $\hat{\beta}_0 = \bar{Y}_0$, $\hat{\beta}_0 + \hat{\beta}_1 = \bar{Y}_1$, y $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$.
- 5.11** Una muestra aleatoria de trabajadores contiene $n_m = 120$ hombres y $n_w = 131$ mujeres. La media muestral de los ingresos salariales semanales de los hombres [$\bar{Y}_m = (1/n_m) \sum_{i=1}^{n_m} Y_{m,i}$] es de 523,10 \$, y la desviación típica muestral [$s_m = \sqrt{\frac{1}{n_m-1} \sum_{i=1}^{n_m} (Y_{m,i} - \bar{Y}_m)^2}$] es de 68,1 \$. Los valores correspondientes para las mujeres son $\bar{Y}_w = 485,10$ \$ y $s_w = 51,10$ \$. Sea *Mujer* una variable indicador que es igual a 1 para las mujeres y 0 para los hombres y supóngase que se utilizan las 251 observaciones en la regresión $Y_i = \beta_0 + \beta_1 Mujer + u_i$. Halle las estimaciones MCO de β_0 y β_1 sus correspondientes errores estándar.
- 5.12** A partir de la Ecuación (4.22), obtenga la varianza de $\hat{\beta}_0$ con homocedasticidad dada en la Ecuación (5.28) del Apéndice 5.1.
- 5.13** Supóngase que (Y_i, X_i) satisfacen los supuestos del Concepto clave 4.3 y, además, u_i es $N(0, \sigma_u^2)$ y es independiente de X_i .
- a) ¿Es $\hat{\beta}_1$ condicionalmente insesgado?
 - b) ¿Es $\hat{\beta}_1$ el estimador lineal condicionalmente insesgado óptimo de β_1 ?
 - c) ¿Cómo cambiaría su respuesta a (a) y a (b) si solamente se supone que (Y_i, X_i) cumplen los supuestos del Concepto clave 4.3 y $\text{var}(u_i | X_i = x)$ es constante?
 - d) ¿Cómo cambiaría su respuesta a (a) y a (b) si solamente se supone que (Y_i, X_i) cumplen los supuestos del Concepto clave 4.3?
- 5.14** Supóngase que $Y_i = \beta X_i + u_i$, donde (u_i, X_i) cumplen las condiciones de Gauss-Markov dadas en la Ecuación (5.31).
- a) Obtenga el estimador de mínimos cuadrados de β y demuestre que es una función lineal de Y_1, \dots, Y_n .
 - b) Demuestre que el estimador es condicionalmente insesgado.
 - c) Obtenga la varianza condicional del estimador.
 - d) Demuestre que el estimador es ELIO.

- 5.15** Un investigador tiene dos muestras independientes de observaciones sobre (Y_i, X_i) . En concreto, supóngase que Y_i representa los ingresos salariales, X_i se refiere a los años de escolaridad, y las muestras independientes son para hombres y mujeres. Escriba la regresión para los hombres como $Y_{m,i} = \beta_{m,0} + \beta_{m,1}X_{m,i} + u_{m,i}$ y la regresión de las mujeres como $Y_{w,j} = \beta_{w,0} + \beta_{w,1}X_{w,j} + u_{w,j}$. Sea $\hat{\beta}_{m,1}$ el estimador MCO construido utilizando la muestra de los hombres, $\hat{\beta}_{w,1}$ es el estimador MCO construido a partir de la muestra de mujeres, y $ES(\hat{\beta}_{m,1})$ y $ES(\hat{\beta}_{w,1})$ indican los correspondientes errores estándar. Demuestre que el error estándar de $\hat{\beta}_{m,1} - \hat{\beta}_{w,1}$ está dado por $ES(\hat{\beta}_{m,1} - \hat{\beta}_{w,1}) = \sqrt{[ES(\hat{\beta}_{m,1})]^2 + [ES(\hat{\beta}_{w,1})]^2}$.

Ejercicios empíricos

- E5.1** Con la base de datos **CPS08** descrita en el Ejercicio empírico E4.1, ejecute una regresión de los ingresos salariales medios por hora (*AHE*) sobre la variable Edad (*Age*) y realice los siguientes ejercicios.
- ¿Es estadísticamente significativo el coeficiente de la pendiente de la regresión estimado? Es decir, se puede rechazar la hipótesis nula $H_0: \beta_1 = 0$ frente a una alternativa bilateral al nivel de significación del 10 %, 5 %, o 1 %? ¿Cuál es el *p*-valor asociado al estadístico *t* del coeficiente?
 - Construya un intervalo de confianza al 95 % para el coeficiente de la pendiente.
 - Repita (a) utilizando solo los datos de los graduados de escuela secundaria.
 - Repita (a) utilizando solo los datos de los graduados universitarios.
 - ¿Es distinto el efecto de la variable *Age* sobre los ingresos salariales para los graduados de secundaria que para los graduados universitarios? Explíquelo. (*Sugerencia:* véase el Ejercicio 5.15).
- E5.2** Con la base de datos **TeachingRatings** descrita en el Ejercicio empírico E4.2, realice una regresión de la variable *Course_Eval* sobre la variable *Beauty*. ¿Es estadísticamente significativo el coeficiente de la pendiente de la regresión estimado? Es decir, se puede rechazar la hipótesis nula $H_0: \beta_1 = 0$ frente a una alternativa bilateral al 10 %, 5 %, o 1 % de nivel de significación? ¿Cuál es el *p*-valor asociado con el estadístico *t* del coeficiente?
- E5.2** Con la base de datos de **CollegeDistance** descrita en el Ejercicio empírico E4.3, realice una regresión de los años de educación (*ED*) sobre la distancia a la universidad más cercana (*Dist*) y realice los siguientes ejercicios.
- ¿Es el coeficiente estimado de la pendiente de la regresión estadísticamente significativo? Es decir, ¿se puede rechazar la hipótesis nula $H_0: \beta_1 = 0$ frente a una alternativa bilateral al nivel de significación del 10 %, 5 %, o 1 %? ¿Cuál es el *p*-valor asociado al estadístico *t* del coeficiente?
 - Construya un intervalo de confianza al 95 % para el coeficiente de la pendiente
 - Realice la regresión utilizando solamente los datos para las mujeres y repita (b).
 - Realice la regresión utilizando solamente los datos para los hombres y repita (b).
 - ¿Es diferente el efecto de la distancia sobre los años completados de educación para los hombres y para las mujeres? (*Sugerencia:* véase el Ejercicio 5.15).

APÉNDICE

5.1 Fórmulas de los errores estándar MCO

Este apéndice analiza las fórmulas para los errores estándar MCO. Se presentan en primer lugar bajo los supuestos de mínimos cuadrados del Concepto clave 4.3, que toleran heterocedasticidad; estos son los errores estándar «heterocedástico-robustos». Posteriormente se obtienen las fórmulas para la varianza de los estimadores MCO y los errores estándar asociados para el caso particular de homocedasticidad.

Errores estándar heterocedástico-robustos

El estimador $\hat{\sigma}_{\hat{\beta}_1}^2$ definido en la Ecuación (5.4) se obtiene mediante la sustitución de las varianzas poblacionales en la Ecuación (4.21) por las varianzas muestrales correspondientes con una modificación. La varianza del numerador de la Ecuación (4.21) se estima mediante $\frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2$ en la que el divisor $n - 2$ (en vez de n) incorpora un ajuste por los grados de libertad para corregir de sesgo a la baja, de forma análoga al ajuste por los grados de libertad utilizado en la definición del ESR de la Sección 4.3. La varianza del denominador se estima mediante $(1/n) \sum_{i=1}^n (X_i - \bar{X})^2$. Sustituyendo $\text{var}[(X_i - \mu_X)u_i]$ y $\text{var}(X_i)$ en la Ecuación (4.21) por esos dos estimadores se obtiene $\hat{\sigma}_{\hat{\beta}_1}^2$ en la Ecuación (5.4). La consistencia de los errores estándar heterocedástico-robustos se trata en la Sección 17.3.

El estimador de la varianza de $\hat{\beta}_0$ es

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n \hat{H}_i^2 \hat{u}_i^2}{\left(\frac{1}{n} \sum_{i=1}^n \hat{H}_i^2 \right)^2} \quad (5.26)$$

donde $\hat{H}_i = 1 - (\bar{X}/\frac{1}{2} \sum_{i=1}^n X_i^2) X_i$. El error estándar de $\hat{\beta}_0$ es $ES(\hat{\beta}_0) = \sqrt{\hat{\sigma}_{\hat{\beta}_0}^2}$. El razonamiento tras el estimador $\hat{\sigma}_{\hat{\beta}_0}^2$ es el mismo que el que está tras $\hat{\sigma}_{\hat{\beta}_1}^2$ y se deriva de la sustitución de las esperanzas poblacionales por las medias muestrales.

Varianzas válidas con homocedasticidad

Con homocedasticidad, la varianza condicional de u_i dado X_i es una constante: $\text{var}(u_i|X_i) = \sigma_u^2$, si los errores son homocedásticos, las fórmulas del Concepto clave 4.4 se simplifican a

$$\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_u^2}{n\sigma_X^2} \text{ y} \quad (5.27)$$

$$\sigma_{\hat{\beta}_0}^2 = \frac{E(X_i^2)}{n\sigma_X^2} \sigma_u^2. \quad (5.28)$$

Para deducir la Ecuación (5.27), se expresa el numerador de la Ecuación (4.21) como

$$\begin{aligned} \text{var}[(X_i - \mu_X)u_i] &= E((X_i - \mu_X)u_i - E[(X_i - \mu_X)u_i])^2 = E\{(X_i - \mu_X)u_i]^2\} = E[(X_i - \mu_X)^2 u_i^2] = \\ &= E[(X_i - \mu_X)^2 \text{var}(u_i|X_i)] \end{aligned}$$

donde la segunda igualdad se obtiene porque $E[(X_i - \mu_X)u_i] = 0$ (por el primer supuesto de mínimos cuadrados) y donde la última igualdad se desprende de la ley de esperanzas iteradas (Sección 2.3). Si u_i es homocedástico, entonces $\text{var}(u_i|X_i) = \sigma_u^2$, por lo que $E[(X_i - \mu_X)^2 \text{var}(u_i|X_i)] = \sigma_u^2 E[(X_i - \mu_X)^2] = \sigma_u^2 \sigma_X^2$. El resultado de la Ecuación (5.27) se obtiene sustituyendo esta expresión en el numerador de la Ecuación (4.21) y simplificando. Mediante cálculos similares se obtiene la Ecuación (5.28).

Errores estándar válidos con homocedasticidad

Los errores estándar válidos con homocedasticidad se obtienen mediante la sustitución de las medias y las varianzas muestrales por las medias y las varianzas poblacionales de las Ecuaciones (5.27) y (5.28), y estimando la varianza de u_i mediante el cuadrado del ESR. Los estimadores de estas varianzas válidos con homocedasticidad son

$$\tilde{\sigma}_{\hat{\beta}_1}^2 = \frac{s_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{válido con homocedasticidad}) \quad (5.29)$$

$$\tilde{\sigma}_{\hat{\beta}_0}^2 = \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) s_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{válido con homocedasticidad}) \quad (5.30)$$

donde s_u^2 viene dado por la Ecuación (4.19). Los errores estándar válidos con homocedasticidad son las raíces cuadradas de $\tilde{\sigma}_{\hat{\beta}_0}^2$ y $\tilde{\sigma}_{\hat{\beta}_1}^2$.

APÉNDICE

5.2 Las condiciones de Gauss-Markov y la demostración del teorema de Gauss-Markov

Tal y como se analizó en la Sección 5.5, el teorema de Gauss-Markov establece que si se cumplen las condiciones de Gauss-Markov, entonces el estimador MCO es el estimador lineal condicionalmente insesgado óptimo (más eficiente) —es ELIO—. Este apéndice comienza estableciendo las condiciones de Gauss-Markov y demostrando que están implícitas en los supuestos de mínimos cuadrados más el de homocedasticidad. A continuación demostraremos que el estimador MCO es un estimador lineal y condicionalmente insesgado. Por último, pasamos a la prueba del teorema.

Las condiciones de Gauss-Markov

Las tres condiciones de Gauss-Markov son

- (i) $E(u_i|X_1, \dots, X_n) = 0$
 - (ii) $\text{var}(u_i|X_1, \dots, X_n) = \sigma_u^2, 0 < \sigma_u^2 < \infty$
 - (iii) $E(u_i u_j|X_1, \dots, X_n) = 0, i \neq j,$
- (5.31)

donde las condiciones se cumplen para $i, j = 1, \dots, n$. Las tres condiciones, respectivamente, establecen que u_i tiene media cero, que u_i tiene varianza constante, y que los errores no están correlacionados entre las diferentes observaciones, todos estos requisitos se cumplen condicionados a todos los (X_1, \dots, X_n) observados.

Las condiciones de Gauss-Markov están implícitas en los tres supuestos de mínimos cuadrados (Concepto clave 4.3), además del supuesto adicional de que los errores son homocedáticos. Debido a que las observaciones son i.i.d. (Supuesto 2), $E(u_i|X_1, \dots, X_n) = E(u_i|X_i)$ y por el Supuesto 1, $E(u_i|X_i) = 0$, por lo que la condición (i) se cumple. Del mismo modo, por el Supuesto 2, $\text{var}(u_i|X_1, \dots, X_n) = \text{var}(u_i|X_i)$, y debido a que los errores se han supuesto homocedásticos, $\text{var}(u_i|X_i) = \sigma_u^2$, que es constante. El Supuesto 3 (momentos de cuarto orden finitos) asegura que, $0 < \sigma_u^2 < \infty$ por lo que la condición (ii) se cumple. Para demostrar que la condición (iii) está implícita en los supuestos de mínimos cuadrados, debe tenerse en cuenta que debido a que $E(u_i u_j|X_1, \dots, X_n) = E(u_i u_j|X_i, X_j)$ debido a que (X_i, Y_i) son i.i.d. por el Supuesto 2. El Supuesto 2 implica asimismo que $E(u_i u_j|X_i, X_j) = E(u_i|X_i)E(u_j|X_j)$ para $i \neq j$; debido a que $E(u_i|X_i) = 0$ para todo i , se deduce que $E(u_i u_j|X_1, \dots, X_n) = 0$ para todo $i \neq j$, por lo que la condición (iii) se cumple. Por lo tanto, el cumplimiento de los supuestos de mínimos cuadrados del Concepto clave 4.3, más la homocedasticidad de los errores, implica el cumplimiento de las condiciones de Gauss-Markov en la Ecuación (5.31).

El estimador MCO $\hat{\beta}_1$ es un estimador lineal condicionalmente insesgado

Para demostrar que $\hat{\beta}_1$ es lineal, hay que tener en cuenta en primer lugar que, debido a que $\sum_{i=1}^n (X_i - \bar{X}) = 0$ (por la definición de \bar{X}), $\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n (X_i - \bar{X})Y_i - \bar{Y} \sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n (X_i - \bar{X})Y_i$. Sustituyendo este resultado en la fórmula de $\hat{\beta}_1$ en la Ecuación (4.7) se obtiene

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{j=1}^n (X_j - \bar{X})^2} = \sum_{i=1}^n \hat{a}_i Y_i, \text{ donde } \hat{a}_i = \frac{(X_i - \bar{X})}{\sum_{j=1}^n (X_j - \bar{X})^2} \quad (5.32)$$

Debido a que las ponderaciones \hat{a}_i , $i = 1, \dots, n$ en la Ecuación (5.32) dependen de X_1, \dots, X_n , pero no de Y_1, \dots, Y_n , el estimador MCO $\hat{\beta}_1$ es un estimador lineal.

Bajo las condiciones de Gauss-Markov, $\hat{\beta}_1$ es condicionalmente insesgado, y la varianza de la distribución condicional de $\hat{\beta}_1$, dados X_1, \dots, X_n , es

$$\text{var}(\hat{\beta}_1 | X_1, \dots, X_n) = \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2}. \quad (5.33)$$

El resultado de que $\hat{\beta}_1$ es condicionalmente insesgado se demostró en el Apéndice 4.3.

Prueba del teorema de Gauss-Markov

Comenzamos obteniendo algunas propiedades válidas para todos los estimadores lineales condicionalmente insesgados, es decir, para todos los estimadores $\tilde{\beta}_1$ que satisfacen las Ecuaciones (5.24) y (5.25). Sustituyendo $Y_i = \beta_0 + \beta_1 X_i + u_i$ en $\tilde{\beta}_1 = \sum_{i=1}^n a_i Y_i$ y agrupando términos, se obtiene que

$$\tilde{\beta}_1 = \beta_0 \left(\sum_{i=1}^n a_i \right) + \beta_1 \left(\sum_{i=1}^n a_i X_i \right) + \sum_{i=1}^n a_i u_i. \quad (5.34)$$

Por la primera condición de Gauss-Markov, $E(\sum_{i=1}^n a_i u_i | X_1, \dots, X_n) = \sum_{i=1}^n a_i E(u_i | X_1, \dots, X_n) = 0$; por tanto, tomando esperanzas condicionales a ambos lados de la Ecuación (5.34) se obtiene $E(\tilde{\beta}_1 | X_1, \dots, X_n) = \beta_0 (\sum_{i=1}^n a_i) + \beta_1 (\sum_{i=1}^n a_i X_i)$. Debido a que $\tilde{\beta}_1$ es condicionalmente insesgado por hipótesis, debe ocurrir que $\beta_0 (\sum_{i=1}^n a_i) + \beta_1 (\sum_{i=1}^n a_i X_i) = \beta_1$, pero para que esta igualdad se cumpla para todos los valores de β_0 y de β_1 debe ocurrir que, para que $\hat{\beta}_1$ sea condicionalmente insesgado

$$\sum_{i=1}^n a_i = 0 \quad \text{y} \quad \sum_{i=1}^n a_i X_i = 1. \quad (5.35)$$

Bajo las condiciones de Gauss-Markov, la varianza de $\tilde{\beta}_1$ condicionada a X_1, \dots, X_n , tiene una forma sencilla. Al sustituir la Ecuación (5.35) en la Ecuación (5.34) se obtiene $\tilde{\beta}_1 - \beta_1 = \sum_{i=1}^n a_i u_i$. Por lo tanto $\text{var}(\tilde{\beta}_1 | X_1, \dots, X_n) = \text{var}(\sum_{i=1}^n a_i u_i | X_1, \dots, X_n) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{cov}(u_i u_j | X_1, \dots, X_n)$; aplicando las condiciones de Gauss-Markov segunda y tercera, los términos cruzados del doble sumatorio desaparecen y la expresión para la varianza condicional se simplifica a

$$\text{var}(\tilde{\beta}_1 | X_1, \dots, X_n) = \sigma_u^2 \sum_{i=1}^n a_i^2. \quad (5.36)$$

Hay que tener en cuenta que las Ecuaciones (5.35) y (5.36) se aplican a $\hat{\beta}_1$ con ponderaciones $a_i = \hat{a}_i$ dados en la Ecuación (5.32).

Ahora vamos a demostrar que las dos restricciones de la Ecuación (5.35) y la expresión de la varianza condicional de la Ecuación (5.36) implican que la varianza condicional de $\tilde{\beta}_1$ es mayor que la varianza condicional de $\hat{\beta}_1$ a menos que sea igual a $\tilde{\beta}_1$. Sea $a_i = \hat{a}_i + d_i$, por lo que $\sum_{i=1}^n a_i^2 = \sum_{i=1}^n (\hat{a}_i + d_i)^2 = \sum_{i=1}^n \hat{a}_i^2 + 2 \sum_{i=1}^n \hat{a}_i d_i + \sum_{i=1}^n d_i^2$.

Utilizando la definición de \hat{a}_i de la Ecuación (5.32), se tiene que

$$\begin{aligned}\sum_{i=1}^n \hat{a}_i d_i &= \frac{\sum_{i=1}^n (X_i - \bar{X}) d_i}{\sum_{j=1}^n (X_j - \bar{X})^2} = \frac{\sum_{i=1}^n d_i X_i - \bar{X} \sum_{i=1}^n d_i}{\sum_{j=1}^n (X_j - \bar{X})^2} \\ &= \frac{\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n \hat{a}_i X_i \right) - \bar{X} \left(\sum_{i=1}^n a_i - \sum_{i=1}^n \hat{a}_i \right)}{\sum_{j=1}^n (X_j - \bar{X})^2} = 0,\end{aligned}$$

donde la penúltima igualdad se deduce de que $d_i = a_i - \hat{a}_i$ y la última igualdad a partir de la Ecuación (5.35) (que se cumple tanto para a_i como para \hat{a}_i). Por tanto $\sigma_u^2 \sum_{i=1}^n a_i^2 = \sigma_u^2 \sum_{i=1}^n \hat{a}_i^2 + \sigma_n^2 \sum_{i=1}^n d_i^2 = \text{var}(\hat{\beta}_1 | X_1, \dots, X_n) + \sigma_n^2 \sum_{i=1}^n d_i^2$; sustituyendo este resultado en la Ecuación (5.36) se obtiene que

$$\text{var}(\tilde{\beta}_1 | X_1, \dots, X_n) - \text{var}(\hat{\beta}_1 | X_1, \dots, X_n) = \sigma_n^2 \sum_{i=1}^n d_i^2. \quad (5.37)$$

Por lo tanto $\tilde{\beta}_1$ tiene una mayor varianza condicional que $\hat{\beta}_1$ si d_i es distinto de cero para cualquier $i = 1, \dots, n$. Pero si $d_i = 0$ para todo i , entonces $a_i = \hat{a}_i$ y $\tilde{\beta}_1 = \hat{\beta}_1$, lo que demuestra que MCO es ELIO.

El teorema de Gauss-Markov si X es no aleatoria

Con un pequeño cambio en la interpretación, el teorema de Gauss-Markov es aplicable asimismo a los regresores no aleatorios; es decir, es aplicable a las variables explicativas que no cambian sus valores en muestras repetidas. En concreto, si el segundo supuesto de mínimos cuadrados se sustituye por el supuesto de que X_1, \dots, X_n son no aleatorias (fijas en muestras repetidas) y u_1, \dots, u_n son i.i.d., entonces lo anteriormente definido, así como la prueba del teorema de Gauss-Markov, son aplicables directamente, salvo que todo lo definido como «condicionado a» resulta innecesario, debido a que X_1, \dots, X_n toman los mismos valores de una muestra a otra.

La media muestral es el estimador lineal eficiente de $E(Y)$

Una consecuencia del teorema de Gauss-Markov es que la media muestral, \bar{Y} , es el estimador lineal más eficiente de $E(Y_i)$ si Y_1, \dots, Y_n son i.i.d. Para comprobarlo, consideremos el caso de la regresión sin una « X », por lo que el único regresor es la variable constante $X_{0i} = 1$. Entonces el estimador MCO $\hat{\beta}_0 = \bar{Y}$. De ello se desprende que, bajo los supuestos de Gauss-Markov \bar{Y} es ELIO. Téngase en cuenta que el requisito de Gauss-Markov de que el error sea homocédastico se satisface automáticamente en este caso porque no hay regresores, por lo que se deduce que \bar{Y} es ELIO si Y_1, \dots, Y_n son i.i.d. Este resultado fue enunciado anteriormente en el Concepto clave 3.3.

Regresión lineal con varios regresores

El Capítulo 5 terminaba con una observación preocupante. Aunque los distritos escolares con una menor ratio de estudiantes-maestros suelen tener mejores calificaciones en los exámenes según la base de datos de California, tal vez los estudiantes de los distritos con grupos pequeños tienen otras ventajas que ayudan a los buenos resultados en los exámenes estandarizados. ¿Podría esto haber dado lugar a resultados engañosos, y si es así, ¿qué se puede hacer?

Los factores omitidos, tales como las características de los estudiantes, pueden, de hecho, occasionar que el estimador de mínimos cuadrados ordinarios (MCO) del efecto del tamaño de las clases sobre las calificaciones en los exámenes pueda inducir a error o, de forma más precisa, sea sesgado. En este capítulo se explica este «sesgo de variable omitida» y se introduce la regresión múltiple, un método que puede eliminar el sesgo de variable omitida. La idea clave de la regresión múltiple es que si se dispone de datos sobre estas variables omitidas, entonces se pueden incluir como regresores adicionales y por tanto se puede calcular el efecto de un regresor (la ratio maestros-alumnos) mientras se mantienen constantes las otras variables (como las características de los estudiantes).

En este capítulo se explica cómo estimar los coeficientes del modelo de regresión lineal múltiple. Muchos aspectos de la regresión múltiple son análogos a los de la regresión con un único regresor, que se trató en los Capítulos 4 y 5. Los coeficientes del modelo de regresión múltiple se pueden estimar a partir de los datos utilizando MCO; los estimadores MCO de regresión múltiple son variables aleatorias porque dependen de los datos de una muestra aleatoria; y en muestras grandes las distribuciones muestrales de los estimadores MCO son aproximadamente normales.

6.1 Sesgo de variable omitida

Al haberse centrado únicamente en la ratio estudiantes-maestros, el análisis empírico de los Capítulos 4 y 5 ignoraba algunos factores que potencialmente podían ser determinantes importantes de las calificaciones en los exámenes, quedando recopilada su influencia en el término de error de la regresión. Estos factores omitidos incluyen las características de la escuela, tales como la calidad de los maestros y el uso del ordenador, y las características de los estudiantes, tales como el entorno familiar. Comenzamos por considerar una característica omitida de los estudiantes que resulta especialmente relevante en California debido a su gran población inmigrante: la prevalencia en el distrito escolar de estudiantes que por no ser hablantes nativos se encuentran todavía aprendiendo inglés.

Al pasar por alto el porcentaje de estudiantes de inglés en el distrito, el estimador de MCO de la pendiente de la regresión de las calificaciones en las pruebas sobre la ratio estudiantes-maestros podría estar sesgado; es decir, la media de la distribución muestral del estimador MCO podría no ser igual al verdadero efecto sobre los resultados en las pruebas de una variación unitaria en la ratio estudiantes-maestros. He aquí el razonamiento. Los estudiantes que todavía están aprendiendo inglés podrían tener un peor desempeño en los exámenes estandarizados que los hablantes nativos de inglés. Si los distritos con clases grandes asimismo tienen muchos estudiantes que aún están aprendiendo inglés, entonces la regresión MCO de las calificaciones en las pruebas sobre la ratio estudiantes-maestros podría encontrar erróneamente una correlación y procurar un coeficiente estimado grande, cuando en realidad el verdadero efecto causal de la reducción del

tamaño de las clases sobre las calificaciones obtenidas es pequeño, e incluso nulo. En consecuencia, basándose en el análisis de los Capítulos 4 y 5, la directora podría contratar suficientes maestros nuevos con el fin de reducir la proporción estudiantes-maestros en 2, pero la mejora esperada de las calificaciones en la prueba podría no materializarse si el verdadero coeficiente fuera pequeño o nulo.

Una ojeada a los datos de California proporciona credibilidad a esta cuestión. La correlación entre la ratio de estudiantes-maestros y el porcentaje de estudiantes de inglés (estudiantes que no son hablantes nativos de inglés y que aún no dominan el idioma) en el distrito es 0,19. Esta pequeña correlación positiva sugiere que los distritos con más estudiantes de inglés suelen tener una mayor proporción de alumnos por maestro (clases más grandes). Si la ratio estudiantes-maestros no estuviera correlacionada con el porcentaje de estudiantes de inglés, entonces estaría justificado pasar por alto el dominio del inglés en la regresión de las calificaciones en las pruebas sobre la ratio estudiantes-maestros. Pero debido a que la proporción estudiantes-maestros y el porcentaje de estudiantes de inglés están correlacionados, es posible que la estimación MCO de los coeficientes de la regresión de las calificaciones en las pruebas sobre la ratio estudiantes-maestros refleje esta influencia.

Definición del sesgo de variable omitida

Si el regresor (la ratio estudiantes-maestros) está correlacionado con una variable que ha sido omitida en el análisis (el porcentaje de estudiantes de inglés) y ésta determina, en parte, la variable dependiente (las calificaciones en las pruebas), el estimador MCO presentará **sesgo de variable omitida**.

El sesgo de variable omitida se produce cuando se cumplen dos condiciones: (1) cuando la variable omitida está correlacionada con los regresores incluidos en la regresión y (2) cuando la variable omitida es un factor determinante de la variable dependiente. Con el fin de ilustrar estas condiciones, consideremos tres ejemplos de variables que se omiten en la regresión de las calificaciones en las pruebas sobre la ratio estudiantes-maestros.

Ejemplo #1: Porcentaje de estudiantes de inglés. Debido a que el porcentaje de estudiantes de inglés está correlacionado con el número de alumnos por maestro, la primera condición para el sesgo de variable omitida se cumple. Resulta verosímil que a los estudiantes que están aprendiendo inglés les vaya peor en las pruebas estandarizadas que a los hablantes nativos de inglés, en cuyo caso el porcentaje de estudiantes de inglés es un factor determinante de las calificaciones en la prueba y la segunda condición para el sesgo de variable omitida se cumple. Por tanto el estimador MCO de la regresión de las calificaciones en los exámenes sobre la proporción estudiantes-maestros podría reflejar incorrectamente la influencia de la variable omitida, el porcentaje de estudiantes de inglés. Es decir, la omisión del porcentaje de estudiantes de inglés puede introducir sesgo de variable omitida.

Ejemplo #2: La hora del día de la prueba. Otra de las variables omitidas en el análisis es la hora del día en la que fue realizada la prueba. Para esta variable omitida, resulta verosímil que no se cumpla la primera condición para el sesgo de variable omitida, pero sí que se cumpla la segunda condición. Por ejemplo, si la hora del día en la que se realiza la prueba varía de un distrito a otro de manera que no esté correlacionada con el tamaño de la clase, entonces la hora del día y el tamaño de las clases no estarían correlacionados y la primera condición no se cumpliría. Por el contrario, la hora del día del examen podría afectar a los resultados (el estado de alerta varía a lo largo de la jornada escolar), por lo que la segunda condición se cumple. Sin embargo, como en este ejemplo la hora del día en la que se realiza la prueba no está correlacionada con la ratio estudiantes-maestros, la ratio estudiantes-maestros no recogería de forma incorrecta el efecto de la «hora del día». Por lo que la omisión de la hora del día en la que se lleva a cabo examen no provoca un sesgo de variable omitida.

Ejemplo #3: Espacio de aparcamiento por alumno. Otra variable que se omite es el espacio para aparcar por alumno (el área de estacionamiento para maestros dividida por el número de estudiantes). Esta variable satisface la primera condición, pero no la segunda para la existencia de sesgo de variable omitida. En concreto, las escuelas con mayor número de docentes por alumno probablemente tendrán más espacio de estacionamiento por maestro, por lo que la primera condición se cumpliría. Sin embargo, bajo el supuesto

**CONCEPTO
CLAVE**
6.1**Sesgo de variable omitida en la regresión con un único regresor**

El sesgo de variable omitida es el sesgo que aparece en el estimador MCO cuando el regresor, X , está correlacionado con una variable omitida. Para que se produzca el sesgo de variable omitida, deben cumplirse dos condiciones:

1. X está correlacionada con la variable omitida.
2. La variable omitida es un determinante de la variable dependiente, Y .

de que el aprendizaje se lleve a cabo en el aula, y no en el aparcamiento, la abundancia de espacio de estacionamiento no tiene efecto directo sobre el aprendizaje; por lo que la segunda condición no se cumple. Debido a que el espacio de estacionamiento por alumno no es un factor determinante para las calificaciones en los exámenes, su omisión del análisis no da lugar a un sesgo de variable omitida.

El sesgo de variable omitida se resume en el Concepto clave 6.1.

El sesgo de variable omitida y el primer supuesto de mínimos cuadrados. El sesgo de variable omitida significa que el primer supuesto de mínimos cuadrados, que $E(u_i|X_i) = 0$, como se enumeró en el Concepto clave 4.3, no se cumple. Para comprobar por qué, hay que recordar que el término de error u_i en el modelo de regresión lineal con un único regresor representa todos los factores, distintos de X_i , que son determinantes de Y_i . Si uno de esos otros factores está correlacionado con X_i , esto significa que el término de error (que contiene a este factor) está correlacionado con X_i . En otras palabras, si una variable omitida es un determinante de Y_i , entonces está en el término de error, y si está correlacionada con X_i , entonces el término de error está correlacionado con X_i . Debido a que u_i y X_i están correlacionados, la media condicional de u_i dado X_i es distinta de cero. Esta correlación por lo tanto, viola el primer supuesto de mínimos cuadrados, y la consecuencia es grave: el estimador MCO es sesgado. Este sesgo no desaparece incluso en muestras muy grandes, y el estimador MCO es inconsistente.

Fórmula del sesgo de variable omitida

La discusión de la sección anterior sobre el sesgo de variable omitida puede resumirse matemáticamente mediante una fórmula para ese sesgo. Sea la correlación entre X_i y u_i $\text{corr}(X_i, u_i) = \rho_{Xu}$. Supongamos que se cumplen los supuestos de mínimos cuadrados segundo y tercero, pero no el primero debido a que ρ_{Xu} es distinto de cero. Entonces el estimador MCO tiene el límite (obtenido en el Anexo 6.1)

$$\hat{\beta}_1 \xrightarrow{P} \beta_1 + \rho_{Xu} \frac{\sigma_u}{\sigma_X}. \quad (6.1)$$

Es decir, a medida que aumenta el tamaño de muestra, $\hat{\beta}_1$ se acerca a $\beta_1 + \rho_{Xu}(\sigma_u/\sigma_X)$ con probabilidad creciente.

La fórmula de la Ecuación (6.1) resume varias de las ideas que se trataron anteriormente sobre el sesgo de variable omitida:

1. El Sesgo de variable omitida es un problema tanto si el tamaño de la muestra es grande como si es pequeño. Debido a que $\hat{\beta}_1$ no converge en probabilidad al verdadero valor β_1 , $\hat{\beta}_1$ es sesgado y es inconsistente; es decir, $\hat{\beta}_1$ no es un estimador consistente de β_1 cuando existe sesgo de variable omitida. El término $\rho_{Xu}(\sigma_u/\sigma_X)$ de la Ecuación (6.1) es el sesgo de $\hat{\beta}_1$, que persiste incluso en muestras grandes.
2. Si este sesgo es grande o pequeño, en la práctica depende de la correlación ρ_{Xu} entre el regresor y el término de error. Cuanto mayor sea $|\rho_{Xu}|$, mayor sesgo.
3. La dirección del sesgo de $\hat{\beta}_1$ depende de si X y u están positiva o negativamente correlacionadas. Por ejemplo, se especuló acerca de si el porcentaje de estudiantes que están aprendiendo inglés tiene un efecto negativo sobre las calificaciones del distrito en el examen (los estudiantes en proceso de aprendizaje del idioma obtienen puntuaciones más bajas), por lo que el porcentaje de estudiantes de inglés entra en el término de error con signo negativo. En nuestros datos, la proporción de estu-

El efecto Mozart: ¿sesgo de variable omitida?

Un estudio publicado en *Nature* en 1993 (Rauscher, Shaw y Ky, 1993) sugería que escuchar música de Mozart durante 10 a 15 minutos podría aumentar temporalmente el cociente intelectual en 8 o 9 puntos. Este estudio fue una gran noticia, y los políticos y los padres vieron una manera fácil de hacer a sus hijos más listos. Durante un tiempo, el estado de Georgia llegó incluso a distribuir CD de música clásica a todos los niños del estado.

¿Cuál es la evidencia del «efecto Mozart»? Una revisión realizada sobre docenas de estudios publicados halló que los estudiantes que asistieron a asignaturas optativas de música o de arte en la escuela secundaria tuvieron, de hecho, mejores calificaciones en los exámenes de inglés y matemáticas que aquellos que no las cursaron¹. Una lectura más detenida de estos estudios, no obstante, sugiere que la verdadera razón del mejor desempeño en los exámenes tenía poco que ver con esas asignaturas. En cambio, los autores de la revisión sugirieron que la correlación entre buenos exámenes y cursar asignaturas de arte o música podía surgir por un gran número de motivos. Por ejemplo, los estudiantes académicamente mejores posiblemente tenían más tiempo para acudir a las asignaturas optativas de música o tenían más interés en hacerlo, o las escuelas con un plan de estudios más completos en música podrían ser en realidad mejores escuelas en todos los ámbitos. En la ter-

minología de la regresión, la relación estimada entre las calificaciones en los exámenes y el hecho de cursar asignaturas optativas de música parece presentar sesgo de variable omitida. Al omitir factores tales como la capacidad innata del estudiante o la calidad general de la escuela, estudiar música parece tener efecto sobre las calificaciones en las pruebas, cuando en realidad no tiene ninguno.

Entonces, ¿existe un efecto Mozart? Una forma de saberlo es llevando a cabo un experimento aleatorizado controlado. (Como se trató en el Capítulo 4, los experimentos aleatorizados controlados eliminan el sesgo de variable omitida mediante la asignación aleatoria de los participantes a grupos de «tratamiento» y de «control»). En conjunto, los abundantes experimentos controlados sobre el efecto Mozart fracasan al demostrar que escuchar la música de Mozart mejora el CI o el desempeño general en el examen. Sin embargo, por razones que no se entienden totalmente, parece ser que escuchar música clásica ayuda temporalmente en un área muy concreta: la papiroflexia y la visualización de formas. Así que la próxima vez que se enrede en un examen de origami, trate de escuchar a la vez un poco de Mozart.

¹ Véase el número de otoño/invierno de 2000 del *Journal of Aesthetic Education*, 34, especialmente el artículo de Ellen Winner y Monica Cooper (pp. 11-76) y el de Lois Hetland (pp. 105-148).

diantes de inglés está *positivamente* correlacionada con la ratio estudiantes-maestros (los distritos con más estudiantes de inglés tienen clases más grandes). Por tanto, la ratio estudiantes-maestros (X) estaría *negativamente* correlacionada con el término de error (u), por lo que $\rho_{Xu} < 0$ y el coeficiente de la proporción estudiantes-maestros $\hat{\beta}_1$ estaría sesgado hacia un número negativo. En otras palabras, tener un porcentaje pequeño de alumnos aprendiendo inglés está asociado tanto con *elevadas* calificaciones en los exámenes como con una *baja* proporción estudiantes-maestros, por lo que una de las razones de que el estimador MCO sugiera que las clases pequeñas mejoran los resultados en los exámenes puede ser que los distritos con clases pequeñas tienen menos estudiantes de inglés.

Solución del sesgo de variable omitida mediante la división de los datos en grupos

¿Qué se puede hacer con respecto al sesgo de variable omitida? Nuestra directora está considerando aumentar el número de profesores de su distrito, pero ella no tiene control sobre la proporción de inmigrantes en su comunidad. Por lo tanto, está interesada en el efecto de la ratio estudiantes-maestros sobre las calificaciones obtenidas, *manteniendo constantes* otros factores, incluyendo el porcentaje de estudiantes aprendiendo inglés. Esta nueva forma de plantear su pregunta sugiere que, en lugar de utilizar los datos de todos los distritos, tal vez deberíamos centrarnos en los distritos con porcentajes comparables de alumnos aprendiendo inglés. Dentro de este subconjunto de distritos, ¿aquellos con clases más pequeñas obtienen mejores resultados en las pruebas estandarizadas?

La Tabla 6.1 presenta la evidencia acerca de la relación entre el tamaño de las clases y las calificaciones en los exámenes dentro de los distritos con porcentajes similares de alumnos aprendiendo inglés. Los distritos se dividen en ocho grupos. En primer lugar, los distritos se dividen en cuatro categorías que se corresponden con los cuartiles de la distribución del porcentaje de estudiantes aprendiendo inglés entre los distritos. En segundo lugar, en cada una de estas cuatro categorías, los distritos se subdividen a su vez en dos grupos, dependiendo de si la ratio estudiantes-maestros es pequeña ($REM < 20$) o grande ($REM \geq 20$).

TABLA 6.1

Diferencias en las calificaciones en los exámenes para los distritos escolares de California con alta y baja ratio de estudiantes por maestro, agrupados por porcentaje de estudiantes de inglés del distrito

	Ratio estudiantes-maestros < 20		Ratio estudiantes-maestros ≥ 20		Diferencia en la calificación en el examen, bajo REM vs alto REM	
	Promedio Calificación Examen	n	Promedio Calificación Examen	n	Diferencia	Estadístico t
Todos los distritos	657,4	238	650,0	182	7,4	4,04
Porcentaje de estudiantes de inglés						
< 1,9 %	664,5	76	665,4	27	-0,9	-0,30
1,9-8,8 %	665,2	64	661,8	44	3,3	1,13
8,8-23,0 %	654,9	54	649,7	50	5,2	1,72
> 23,0 %	636,7	44	634,8	61	1,9	0,68

La primera fila de la Tabla 6.1 presenta la diferencia global en las calificaciones medias en los exámenes entre los distritos con ratios altas y bajas de estudiantes-maestros, es decir, la diferencia en las calificaciones de los exámenes entre estos dos grupos sin desglosarlas por los cuartiles de alumnos aprendiendo inglés. (Recordemos que esta diferencia fue previamente presentada en forma de regresión en la Ecuación (5.18). Como la estimación MCO del coeficiente de D_i en la regresión de la variable *CalificacionExamen* sobre la variable D_i , donde D_i es un regresor binario que es igual a 1 si $REM_i < 20$ y es igual a 0 en caso contrario). A lo largo de la muestra total de 420 distritos, la calificación media en la prueba es 7,4 puntos mayor en los distritos con una ratio baja de estudiantes-maestros que en los que presentan una ratio alta; el estadístico *t* es de 4,04, por lo que la hipótesis nula de que la media de las calificaciones en la prueba es la misma para los dos grupos se rechaza al nivel de significación del 1 %.

Las últimas cuatro filas de la Tabla 6.1 presentan la diferencia en las calificaciones de la prueba entre distritos con alta y baja proporción de alumnos por maestro, desglosada por los cuartiles del porcentaje de estudiantes aprendiendo inglés. Esta evidencia presenta una imagen diferente. De los distritos con menor número de estudiantes aprendiendo inglés (< 1,9 %), la calificación media en la prueba para los 76 con menor proporción de alumnos por maestro es 664,5 y la media para los 27 con proporción alta estudiantes-maestros es 665,4. Por tanto, para los distritos con menos estudiantes aprendiendo inglés, las calificaciones en los exámenes fueron en media 0,9 puntos *más bajas* ¡en los distritos con ratios de estudiantes-maestros bajas! En el segundo cuartil, los distritos con una baja proporción estudiantes-maestros tienen calificaciones en las pruebas con una media de 3,3 puntos por encima de aquellos con alta proporción de estudiantes por maestro; esta brecha fue de 5,2 puntos para el tercer cuartil y de solo 1,9 puntos para el cuartil de los distritos con más alumnos aprendiendo inglés. Una vez que mantenemos constante el porcentaje de alumnos que están aprendiendo inglés, la diferencia en el desempeño entre los distritos con altas y bajas proporciones de alumnos por maestro es tal vez la mitad (o menos) de la estimación global de 7,4 puntos.

Al principio, este hallazgo puede parecer desconcertante. ¿Cómo puede ser el efecto global sobre las calificaciones en los exámenes dos veces el efecto sobre los resultados en los exámenes dentro de cualquier cuartil? La respuesta es que los distritos con mayor cantidad de estudiantes aprendiendo inglés suelen tener *tan* la mayor proporción de estudiantes por maestro *como* las menores calificaciones en los exámenes. La diferencia en la calificación media en los exámenes entre los distritos del cuartil más bajo y más alto del porcentaje de alumnos aprendiendo inglés es grande, de aproximadamente 30 puntos. Los distritos con pocos estudiantes aprendiendo inglés tienden a tener menor proporción de alumnos por maestro: el 74 % (76 de 103) de los distritos del primer cuartil de alumnos aprendiendo inglés tienen clases pequeñas ($REM < 20$), mientras que solo el 42 % (44 de 105) de los distritos del cuartil con mayor cantidad de estu-

diantes aprendiendo inglés tienen clases pequeñas. Por lo tanto, los distritos con más estudiantes aprendiendo inglés tienen tanto calificaciones más bajas en los exámenes como mayores ratios estudiantes-maestros que los otros distritos.

Este análisis refuerza la preocupación de la directora de que el sesgo de variable omitida está presente en la regresión de las calificaciones en las pruebas sobre la ratio estudiantes-maestros. Al fijar la atención en los cuartiles del porcentaje de estudiantes aprendiendo inglés, las diferencias en las calificaciones en los exámenes de la segunda parte de la Tabla 6.1 mejora el análisis sencillo de la diferencia de medias de la primera línea de la Tabla 6.1. Aun así, este análisis no proporciona todavía a la directora una estimación útil de los efectos sobre las calificaciones en los exámenes de la variación en el tamaño de las clases, manteniendo constante la proporción de estudiantes aprendiendo inglés. No obstante, se puede obtener una estimación de este tipo mediante el método de regresión múltiple.

6.2 El modelo de regresión múltiple

El **modelo de regresión múltiple** extiende el modelo de regresión simple (con regresor único) de los Capítulos 4 y 5 para incluir variables adicionales como regresores. Este modelo permite estimar el efecto sobre Y_i de la variación de una variable (X_{1i}) manteniendo constantes el resto de regresores (X_{2i}, X_{3i} , etc.). En el problema del tamaño de las clases, el modelo de regresión múltiple proporciona una manera de aislar el efecto sobre las calificaciones en los exámenes (Y_i) de la ratio estudiantes-maestros (X_{1i}) manteniendo constante el porcentaje de estudiantes del distrito que están aprendiendo inglés (X_{2i}).

La recta de regresión poblacional¹

Supongamos por el momento que solo hay dos variables independientes, X_{1i} y X_{2i} . En el modelo de regresión lineal múltiple, la relación promedio entre estas dos variables independientes y la variable dependiente, Y , está dada por la función lineal

$$E(Y_i|X_{1i} = x_1, X_{2i} = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad (6.2)$$

donde $E(Y_i|X_{1i} = x_1, X_{2i} = x_2)$ es la esperanza condicional de Y_i , dado que $X_{1i} = x_1$ y $X_{2i} = x_2$. Es decir, si el número de alumnos por maestro en el distrito i -ésimo (X_{1i}) es igual a un cierto valor x_1 y el porcentaje de alumnos aprendiendo inglés en el distrito i -ésimo (X_{2i}) es igual a x_2 , entonces el valor esperado de Y_i dados la proporción de alumnos por maestro y el porcentaje de estudiantes aprendiendo inglés viene dado por la Ecuación (6.2).

La Ecuación (6.2) es la **recta de regresión poblacional** o **función de regresión poblacional** en el modelo de regresión múltiple. El coeficiente β_0 es el **intercepto, término independiente** o **término constante**; el coeficiente β_1 es el **coeficiente de la pendiente de X_{1i}** , o más sencillamente, el **coeficiente de X_{1i}** ; y el coeficiente β_2 es el **coeficiente de la pendiente de X_{2i}** , o simplemente, el **coeficiente de X_{2i}** . A una o más variables independientes del modelo de regresión múltiple se les denomina a veces variables de control.

La interpretación del coeficiente β_1 de la Ecuación (6.2) es diferente de lo que lo era cuando X_{1i} era el único regresor: en la Ecuación (6.2), β_1 es el efecto sobre Y de la variación en una unidad de X_1 , **mantiene constante X_2 o teniendo en cuenta X_2 , o controlando por X_2** .

Esta interpretación de β_1 se deriva de la definición según la cual el efecto esperado sobre Y de un cambio en X_1 , ΔX_1 , manteniendo X_2 constante, es la diferencia entre el valor esperado de Y cuando las variables independientes toman los valores $X_1 + \Delta X_1$ y X_2 y el valor esperado de Y cuando las variables independientes toman los valores X_1 y X_2 . Por lo tanto, se escribe la función de regresión poblacional de la Ecu-

¹ N. del T.: Se emplea aquí la denominación «recta» de regresión para mantener la analogía con el modelo de regresión simple. No obstante, es más correcto hablar de «hiperplano» de regresión. Este hiperplano resume la información de los datos muestrales en un hiperplano de dimensión k (número de variables explicativas, aparte del intercepto) de forma análoga al modo en que una recta (dimensión 1) resume la información muestral en el modelo de regresión simple (una variable explicativa además del intercepto).

ción (6.2) como $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ e imaginamos una variación de X_1 en la cuantía ΔX_1 mientras X_2 no cambia, es decir, se mantiene constante X_2 . Debido a que X_1 ha cambiado, Y cambiará en una cierta cantidad, por ejemplo ΔY . Tras esta variación, el nuevo valor de Y , $Y + \Delta Y$, es

$$Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2 \quad (6.3)$$

Se obtiene una ecuación para ΔY en términos de ΔX_1 restando la ecuación $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ a la Ecuación (6.3), por lo que se obtiene $\Delta Y = \beta_1 \Delta X_1$. Es decir,

$$\text{Expresión } \beta_1 = \frac{\Delta Y}{\Delta X_1} \text{ manteniendo } X_2 \text{ constante} \quad (6.4)$$

El coeficiente β_1 es el efecto sobre Y (la esperanza de la variación de Y) de un cambio unitario en X_1 , manteniendo fija X_2 . Otra frase que se utiliza para describir β_1 es el **efecto parcial** sobre Y de X_1 , manteniendo constante X_2 .

La interpretación del término independiente en el modelo de regresión múltiple β_0 es similar a la interpretación del término independiente en el modelo con regresor único: es el valor esperado de Y_i , cuando X_{1i} y X_{2i} son iguales a cero. En pocas palabras, el intercepto β_0 determina el punto del eje Y del que parte la recta de regresión poblacional.

El modelo de regresión múltiple poblacional

La recta de regresión poblacional de la Ecuación (6.2) es la relación entre Y y X_1 y X_2 que se cumple en media en la población. No obstante, al igual que en el caso de la regresión con un único regresor, esta relación no se cumple con exactitud debido a que muchos otros factores influyen en la variable dependiente. Por ejemplo, además de la proporción de estudiantes por maestro y de la proporción de estudiantes que aún están aprendiendo inglés, las calificaciones en los exámenes están influidas por las características de la escuela, otras características de los estudiantes, y la fortuna. Por tanto, es necesario ampliar la función de regresión poblacional de la Ecuación (6.2) con el fin de incorporar estos factores adicionales.

Al igual que en el caso de la regresión con un único regresor, los factores que determinan Y_i además de X_{1i} y X_{2i} se incorporan a la Ecuación (6.2) en forma de un «término de error» u_i . Este término de error es la desviación de una observación concreta (calificaciones en el examen del distrito i -ésimo, en nuestro ejemplo) respecto de la relación poblacional promedio. En consecuencia, se tiene que

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad i = 1, \dots, n, \quad (6.5)$$

donde el subíndice i indica la i -ésima de las n observaciones (distritos) de la muestra.

La ecuación (6.5) es el **modelo de regresión múltiple poblacional** cuando existen dos regresores, X_{1i} y X_{2i} .

En la regresión con regresores binarios, puede ser útil considerar β_0 como el coeficiente de un regresor que es siempre igual a 1; piénsese en β_0 como el coeficiente de X_{0i} siendo $X_{0i} = 1$ para $i = 1, \dots, n$. Por tanto, el modelo de regresión múltiple poblacional de la Ecuación (6.5) puede escribirse alternativamente como

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \text{ donde } X_{0i} = 1, \quad i = 1, \dots, n. \quad (6.6)$$

La variable X_{0i} se denomina a veces **regresor constante**, ya que toma el mismo valor, el valor 1, para todas las observaciones. Del mismo modo, el intercepto β_0 , a veces se denomina **término constante** de la regresión.

Las dos formas de escribir el modelo de regresión poblacional, las Ecuaciones (6.5) y (6.6), son equivalentes.

El análisis hasta el momento se ha centrado en el caso de una sola variable adicional X_2 . En la práctica, no obstante, podrían existir múltiples factores omitidos en el modelo de regresión simple. Por ejemplo, al ignorar las circunstancias económicas de los estudiantes podría aparecer un sesgo de variable omitida, tal

**CONCEPTO
CLAVE**
El modelo de regresión múltiple

El modelo de regresión múltiple es

6.2

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n, \quad (6.7)$$

donde

- Y_i es la i -ésima observación de la variable dependiente; $X_{1i}, X_{2i}, \dots, X_{ki}$ son las i -ésimas observaciones de cada uno de los k regresores; y u_i es el término de error.
- La recta de regresión poblacional es la relación que se cumple entre Y y las X en media en la población:

$$E(Y|X_{1i} = x_1, X_{2i} = x_2, \dots, X_{ki} = x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

- β_1 es el coeficiente de la pendiente de X_1 , β_2 es el coeficiente de X_2 , etc. El coeficiente β_1 es la esperanza de la variación de Y_i que resulta de variar X_{1i} en una unidad, manteniendo constantes X_{2i}, \dots, X_{ki} . Los coeficientes de las otras X se interpretan de forma similar.
- El intercepto β_0 es el valor esperado de Y cuando todas las X son iguales a 0. Se puede interpretar el intercepto como el coeficiente de un regresor, X_{0i} , que es igual a 1 para todo i .

y como sucedía al ignorar la proporción de estudiantes aprendiendo inglés. Este razonamiento conduce a considerar un modelo con tres regresores o, de manera más general, un modelo que incluya k regresores. El modelo de regresión múltiple con k regresores, $X_{1i}, X_{2i}, \dots, X_{ki}$ se encuentra resumido en el Concepto clave 6.2.

Las definiciones de homocedasticidad y heterocedasticidad en el modelo de regresión múltiple son extensiones de sus definiciones para el modelo de regresión con regresor único. El término de error u_i en el modelo de regresión múltiple es **homocedástico** si la varianza de la distribución condicional de u_i dados $X_{1i}, X_{2i}, \dots, X_{ki}$, $\text{var}(u_i|X_{1i}, \dots, X_{ki})$, es constante para $i = 1, \dots, n$ y por tanto no depende de los valores de X_{1i}, \dots, X_{ki} . En cualquier otro caso, el término de error es **heterocedástico**.

El modelo de regresión múltiple mantiene la promesa de proporcionar exactamente lo que la directora quiere saber: el efecto de la variación en la proporción estudiantes-maestros, manteniendo constantes los otros factores que quedan más allá de su control. Estos factores incluyen no solo el porcentaje de estudiantes aprendiendo inglés, sino además otros factores medibles que pueden afectar a los resultados del examen, incluyendo las circunstancias económicas de los estudiantes. No obstante, para que resulte de utilidad práctica a la directora, es necesario que se le proporcionen las estimaciones de los coeficientes poblacionales desconocidos β_0, \dots, β_k , del modelo de regresión poblacional calculados utilizando una muestra de datos. Afortunadamente, esos coeficientes pueden estimarse utilizando mínimos cuadrados ordinarios.

6.3 El estimador MCO en regresión múltiple

En esta Sección se describe cómo los coeficientes del modelo de regresión múltiple puede ser estimados mediante MCO.

El estimador MCO

La Sección 4.2 muestra cómo se pueden estimar los coeficientes del término independiente y de la pendiente del modelo de regresión simple aplicando MCO a una muestra de observaciones de Y y X . La idea clave es que estos coeficientes pueden estimarse minimizando la suma de los cuadrados de los errores de predicción, es decir, seleccionando los estimadores b_0 y b_1 que minimizan $\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$. Los estimadores que lo cumplen son los estimadores MCO, $\hat{\beta}_0$, y $\hat{\beta}_1$.

El método MCO puede asimismo aplicarse para estimar los coeficientes $\beta_0, \beta_1, \dots, \beta_k$ en el modelo de regresión múltiple. Sean b_0, b_1, \dots, b_k los estimadores de $\beta_0, \beta_1, \dots, \beta_k$. El valor de predicción de Y_i calculado mediante esos estimadores, es $b_0 + b_1X_{1i} + \dots + b_kX_{ki}$, y el error al predecir Y_i es $Y_i - (b_0 + b_1X_{1i} + \dots + b_kX_{ki}) = Y_i - b_0 - b_1X_{1i} - \dots - b_kX_{ki}$. La suma de los cuadrados de los errores de predicción para las n observaciones es

$$\sum_{i=1}^n (Y_i - b_0 - b_1X_{1i} - \dots - b_kX_{ki})^2. \quad (6.8)$$

La suma de los cuadrados de los errores para el modelo de regresión lineal de la Expresión (6.8) es la extensión de la suma de los cuadrados de los errores dada en la Ecuación (4.6) para el modelo de regresión lineal con regresor único.

Los estimadores de los coeficientes $\beta_0, \beta_1, \dots, \beta_k$ que minimizan la suma de los errores al cuadrado en la Expresión (6.8) se denominan **estimadores de mínimos cuadrados ordinarios (MCO)** de $\beta_0, \beta_1, \dots, \beta_k$. Los estimadores MCO se expresan mediante $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$.

La terminología de MCO en el modelo lineal de regresión múltiple es la misma que en el modelo de regresión lineal con regresor único. La **recta de regresión MCO** es la línea recta construida utilizando los estimadores MCO: $\hat{\beta}_0 + \hat{\beta}_1X_1 + \dots + \hat{\beta}_kX_k$. El **valor de predicción** de Y_i dados X_{1i}, \dots, X_{ki} , basado en la recta de regresión es $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_1 + \dots + \hat{\beta}_kX_k$. El **residuo MCO** para la i -ésima observación es la diferencia entre Y_i y su valor de predicción MCO; es decir, el residuo MCO es $\hat{u}_i = Y_i - \hat{Y}_i$.

Los estimadores MCO pueden calcularse mediante el método de prueba y error, probando repetidamente diferentes valores para b_0, \dots, b_k hasta llegar a considerar que se ha alcanzado el mínimo de la suma total de la Expresión (6.8). Es mucho más fácil, sin embargo, utilizar fórmulas explícitas para la estimación MCO deducidas mediante cálculo. Las fórmulas para los estimadores MCO en el modelo de regresión múltiple son similares a las del Concepto clave 4.2 para el modelo de regresión simple. Estas fórmulas se encuentran incorporadas dentro del *software* estadístico moderno. En el modelo de regresión múltiple, las fórmulas se expresan y analizan de mejor manera mediante notación matricial, por lo que su representación se pospone a la Sección 18.1.

Las definiciones y la terminología de MCO en la regresión múltiple se resumen en el Concepto clave 6.3.

Aplicación a las calificaciones en los exámenes y la ratio estudiantes-maestros

En la Sección 4.2, se utilizaron MCO para estimar el término independiente y el coeficiente de la pendiente de la regresión que relacionaba los resultados en las pruebas (*CalificaciónExamen*) con la ratio

CONCEPTO CLAVE

6.3

Los estimadores MCO, valores de predicción y residuos en el modelo de regresión múltiple

Los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ son los valores de b_0, b_1, \dots, b_k que minimizan la suma de los errores de predicción al cuadrado $\sum_{i=1}^n (Y_i - b_0 - b_1X_{1i} - \dots - b_kX_{ki})^2$. Los valores de predicción MCO \hat{Y}_i y los residuos \hat{u}_i son

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_{1i} + \dots + \hat{\beta}_kX_{ki}, \quad i = 1, \dots, n \quad (6.9)$$

y

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, \dots, n. \quad (6.10)$$

Los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ y el residuo \hat{u}_i se calculan a partir de una muestra de n observaciones de $(X_{1i}, \dots, X_{ki}, Y_i)$, $i = 1, \dots, n$. Estos son los estimadores de los verdaderos y desconocidos coeficientes poblacionales $\beta_0, \beta_1, \dots, \beta_k$, y del término de error u_i .

estudiantes-maestros (*REM*), utilizando nuestras 420 observaciones acerca de los distritos escolares de California; la recta de regresión MCO estimada, presentada en la Ecuación (4.11), es

$$\widehat{\text{CalificaciónExamen}} = 698,9 - 2,28 \times \text{REM} \quad (6.11)$$

Nuestra preocupación había sido acerca de si esta relación era engañosa debido a que la ratio estudiantes-maestros podría estar recogiendo el efecto de tener muchos alumnos aprendiendo inglés en los distritos con clases más grandes. Es decir, es posible que el estimador MCO esté sujeto al sesgo de variable omitida.

Ahora estamos en condiciones de abordar este problema mediante el uso de MCO para estimar una regresión múltiple en la que la variable dependiente es la calificación en los exámenes (Y_i) y existen dos regresores: la ratio estudiantes-maestros (X_{1i}) y el porcentaje de alumnos aprendiendo inglés en el distrito escolar (X_{2i}) para nuestros 420 distritos ($i = 1, \dots, 420$). La recta de regresión MCO estimada para esta regresión múltiple es

$$\widehat{\text{CalificaciónExamen}} = 686,0 - 1,10 \times \text{REM} - 0,65 \times \text{PctEI} \quad (6.12)$$

donde PctEI es el porcentaje de estudiantes del distrito que están aprendiendo inglés. La estimación MCO del término independiente ($\hat{\beta}_0$) es 686,0, la estimación MCO del coeficiente de la ratio estudiantes-maestros ($\hat{\beta}_1$) es $-1,10$, y la estimación MCO del coeficiente del porcentaje de alumnos que están aprendiendo inglés ($\hat{\beta}_2$), es $-0,65$.

El efecto estimado sobre las calificaciones en las pruebas de una variación en la ratio estudiantes-maestros en la regresión múltiple es aproximadamente la mitad de grande que cuando la ratio estudiantes-maestros es el único regresor: en la ecuación con un solo regresor [Ecuación (6.11)], se estima que una disminución de la *REM* en una unidad aumentará las calificaciones en los exámenes en 2,28 puntos, pero en la ecuación de regresión múltiple [Ecuación (6.12)], se estima que aumentará las calificaciones en los exámenes en tan solo 1,10 puntos. Esta diferencia se debe a que el coeficiente de *REM* en regresión múltiple es el efecto de una variación de *REM*, manteniendo constante (o controlando por) *PctEI*, mientras que en la regresión de un solo regresor, *PctEI* no se mantiene constante.

Estas dos estimaciones se pueden reconciliar concluyendo que existe un sesgo de variable omitida en la estimación del modelo de un solo regresor de la Ecuación (6.11). En la Sección 6.1, vimos que los distritos con un alto porcentaje de alumnos aprendiendo inglés suelen presentar no solo peores calificaciones en los exámenes, sino que asimismo presentan una ratio estudiantes-maestros elevada. Si la proporción de estudiantes aprendiendo inglés se omite en la regresión, se estima que la reducción de la ratio estudiantes-maestros tendrá un mayor efecto sobre las calificaciones en los exámenes de la prueba, pero esta estimación reflejará tanto el efecto de una variación en la ratio estudiantes-maestros como el efecto omitido de tener menos alumnos aprendiendo inglés en el distrito.

Hemos llegado a la misma conclusión de que existe un sesgo de variable omitida en la relación entre las calificaciones en los exámenes de la prueba y la ratio estudiantes-maestros por dos caminos diferentes: el método de la tabla de dividir los datos en grupos (Sección 6.1) y el método de regresión múltiple [Ecuación (6.12)]. De estos dos métodos, el de regresión múltiple presenta dos ventajas importantes. En primer lugar, proporciona una estimación cuantitativa del efecto de la disminución del cociente estudiantes-maestros en una unidad, que es lo que la directora necesitaba para tomar su decisión. En segundo lugar, es extensible a más de dos variables explicativas de modo que la regresión múltiple se puede utilizar para controlar por otros factores medibles diferentes al porcentaje de estudiantes de inglés.

El resto de este capítulo está dedicado a la comprensión y al uso de MCO en el modelo de regresión múltiple. Mucho de lo aprendido acerca del estimador MCO con un regresor único se traslada a la regresión múltiple con pocas o ninguna modificación, por lo que nos centraremos en aquello que es nuevo en la regresión múltiple. Comenzamos por tratar las medidas de ajuste para el modelo de regresión múltiple.

6.4 Medidas de ajuste en regresión múltiple

Tres estadísticos que se utilizan habitualmente en la regresión múltiple son el error estándar de la regresión, el R^2 de la regresión, y el R^2 ajustado (asimismo conocido como \bar{R}^2). Los tres estadísticos miden la bondad de la estimación MCO de la recta de regresión múltiple, es decir, en qué medida la recta describe, o «se ajusta» a los datos.

El error estándar de la regresión (ESR)

El error estándar de la regresión (ESR) estima la desviación típica del término de error u_i . Por tanto, el ESR es una medida de la dispersión de la distribución de Y alrededor de la recta de regresión. En regresión múltiple, el ESR es

$$ESR = s_{\hat{u}} \text{ donde } s_{\hat{u}}^2 = \frac{1}{n - k - 1} \sum_{i=1}^n \hat{u}_i^2 = \frac{SR}{n - k - 1} \quad (6.13)$$

y donde SR es la suma de los cuadrados de los residuos, $SR = \sum_{i=1}^n \hat{u}_i^2$.

La única diferencia entre la definición de la Ecuación (6.13) y la definición del ESR en la Sección 4.3 para el modelo de un único regresor es que aquí el divisor es $n - k - 1$ en vez de $n - 2$. En la Sección 4.3, el divisor $n - 2$ (en vez de n) se ajustaba por el sesgo a la baja que surge por la estimación de los dos coeficientes (la pendiente y el término independiente de la recta de regresión). En este caso, el divisor $n - k - 1$ ajusta por el sesgo a la baja que surge por la estimación de los $k + 1$ coeficientes (los k coeficientes de pendiente además del intercepto). Como en la Sección 4.3, la utilización de $n - k - 1$ en lugar de n se denomina ajuste por los grados de libertad. Si aquí hubiera habido un único regresor, entonces $k = 1$, por lo que el fórmula de la Sección 4.3 es la misma que en la Ecuación (6.13). Cuando n es grande, el efecto del ajuste por los grados de libertad es insignificante.

El R^2

El R^2 de la regresión es la proporción de la varianza muestral de Y_i que está explicada (o predicha) por los regresores. De manera equivalente, el R^2 es 1 menos la proporción de la varianza de Y_i no explicada por las variables explicativas.

La definición matemática de R^2 es la misma que para la regresión con regresor único:

$$R^2 = \frac{SE}{ST} = 1 - \frac{SR}{ST} \quad (6.14)$$

donde la suma explicada (de cuadrados) es $SE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ y la suma total (de cuadrados) es $ST = \sum_{i=1}^n (Y_i - \bar{Y})^2$.

En regresión múltiple, el R^2 aumenta cada vez que se añade un regresor, a menos que el coeficiente estimado del regresor adicional sea exactamente cero. Para comprobarlo, comenzamos con un único regresor y posteriormente añadimos un segundo regresor. Cuando se utiliza MCO para estimar el modelo con ambas variables explicativas, MCO halla aquellos valores de los coeficientes que reduzcan al mínimo la suma de los cuadrados de los residuos. Si resulta que MCO elige un coeficiente para el nuevo regresor que sea exactamente cero, entonces SR será la misma tanto si se incluye la segunda variable en la regresión como si no. Pero si MCO escoge cualquier otro valor distinto de cero, entonces debe ocurrir que ese valor reduzca la SR de la regresión que excluye a este regresor. En la práctica, es extremadamente inusual que un coeficiente estimado sea igual a cero, por lo que en general la SR disminuye al añadirse un nuevo regresor. Pero esto significa que el R^2 en general aumenta (y nunca disminuye) al añadirse un nuevo regresor.

El « R^2 ajustado»

Debido a que R^2 aumenta cuando se añade una variable, un aumento de R^2 no significa que la adición de una variable mejore realmente el ajuste del modelo. En este sentido, R^2 proporciona una estimación exagerada acerca de la bondad con la que la regresión se ajusta a los datos. Una forma de corregir esto es deflactar o reducir el R^2 mediante algún factor, y esto es lo que hace el R^2 ajustado, o bien \bar{R}^2 .

El R^2 ajustado, o \bar{R}^2 , es una versión modificada de R^2 que no necesariamente aumenta al añadirse un nuevo regresor. El \bar{R}^2 es

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \frac{SR}{ST} = 1 - \frac{s_u^2}{s_Y^2} \quad (6.15)$$

La diferencia entre esta fórmula y la segunda definición de R^2 de la Ecuación (6.14) es que la ratio entre la suma de los cuadrados de los residuos y la suma total de cuadrados se multiplica por el factor $(n-1)/(n-k-1)$. Tal y como muestra la segunda expresión de la Ecuación (6.15), esto significa que el R^2 ajustado es 1 menos el cociente entre la varianza muestral de los residuos MCO [con la corrección por los grados de libertad de la Ecuación (6.13)] y la varianza muestral de Y .

Resulta útil saber tres cosas sobre \bar{R}^2 . En primer lugar $(n-1)/(n-k-1)$, es siempre mayor que 1, por lo que \bar{R}^2 siempre es menor que R^2 .

En segundo lugar, la adición de un regresor tiene dos efectos opuestos sobre \bar{R}^2 . Por un lado, la SR disminuye, lo que hace aumentar \bar{R}^2 . Por otro lado, el factor $(n-1)/(n-k-1)$ aumenta. El hecho de que aumente o disminuya, dependerá de cuál de estos dos efectos sea más fuerte.

En tercer lugar, el \bar{R}^2 puede ser negativo. Esto sucede si todos los regresores, considerados de forma conjunta, reducen la suma de los cuadrados de los residuos en una cantidad tan pequeña que la reducción no pueda compensar el factor $(n-1)/(n-k-1)$.

Aplicación a las calificaciones en los exámenes

La Ecuación (6.12) proporciona el hiperplano de regresión estimado para la regresión múltiple que relaciona las calificaciones obtenidas en los exámenes (*CalificaciónExamen*) con la ratio estudiantes-maestros (*REM*) y el porcentaje de estudiantes que estudian inglés (*PctEI*). El R^2 de este hiperplano de regresión es $R^2 = 0,426$, el R^2 ajustado es $\bar{R}^2 = 0,424$, y el error estándar de la regresión es $ESR = 14,5$.

La comparación de estas medidas de ajuste con las de la regresión en la que está excluida la variable *PctEI* [Ecuación (6.11)] muestra que la inclusión de *PctEI* en la regresión aumenta el R^2 desde 0,051 hasta 0,426. Cuando el único regresor es la variable *REM*, solo se explica una pequeña fracción de la variabilidad de la variable *CalificaciónExamen*; sin embargo, cuando se añade *PctEI* a la regresión, se explican más de las dos quintas partes (42,6 %) de la variabilidad de las calificaciones en las pruebas. En este sentido, la inclusión del porcentaje de estudiantes de inglés mejora sustancialmente el ajuste de la regresión. Al ser n grande y al aparecer solamente dos regresores en la Ecuación (6.12), la diferencia entre el R^2 y el $-\bar{R}^2$ ajustado es muy pequeña ($R^2 = 0,426$ versus $\bar{R}^2 = 0,424$).

El *ESR* que corresponde a la regresión que excluye la variable *PctEI* es 18,6; este valor disminuye hasta 14,5 cuando se incluye la variable *PctEI* como segundo regresor. Las unidades del *ESR* son los puntos en el examen. La reducción en el *ESR* nos dice que las predicciones sobre las calificaciones en las pruebas estandarizadas son mucho más precisas si se realizan utilizando la regresión que incluye tanto la variable *REM* como la variable *PctEI*, que cuando se realizan utilizando la regresión con la variable *REM* como regresor único.

Utilización del R^2 y del R^2 ajustado. El \bar{R}^2 es útil, debido a que cuantifica la medida en la que los regresores representan, o explican, la variabilidad de la variable dependiente. Sin embargo, fiarse demasiado de una medida como la \bar{R}^2 (o de R^2) puede resultar una trampa. En las aplicaciones prácticas, «maximizar la \bar{R}^2 » rara vez es la respuesta adecuada a alguna cuestión importante de naturaleza económica o estadística. En cambio, la decisión sobre si incluir una variable en una regresión múltiple debería basarse en si la inclusión de la variable permite estimar mejor el efecto causal de interés. Volveremos a la cuestión de

cómo decidir qué variables incluir —y cuáles excluir— en el Capítulo 7. En primer lugar, no obstante, resulta necesario desarrollar métodos que cuantifiquen la incertidumbre del estimador MCO que proviene del muestreo. El punto de partida para hacerlo es la extensión de los supuestos de mínimos cuadrados del Capítulo 4 para el caso de varias variables explicativas.

6.5 Los supuestos de mínimos cuadrados en regresión múltiple

Existen al menos cuatro supuestos de mínimos cuadrados en el modelo de regresión múltiple. Los tres primeros son los de la Sección 4.3 para el modelo de un solo regresor (Concepto clave 4.3), ampliados de forma que se permitan varios regresores, los cuales se analizan brevemente. El cuarto supuesto es nuevo y se analiza con más detalle.

Supuesto #1: La distribución condicional de u_i dados $X_{1i}, X_{2i}, \dots, X_{ki}$ tiene media igual a cero

El primer supuesto es que la distribución condicional de u_i dados X_{1i}, \dots, X_{ki} tiene una media igual a cero. Este supuesto generaliza el primer supuesto de mínimos cuadrados con regresor único al caso de regresores múltiples. Este supuesto significa que Y_i a veces se encuentra por encima del hiperplano de regresión poblacional y otras veces Y_i se encuentra por debajo del hiperplano de regresión, pero en promedio sobre la población Y_i se encuentra en el hiperplano de regresión poblacional. Por tanto, para cualquier valor de las variables explicativas, el valor esperado de u_i es cero. Como en el caso de la regresión con un único regresor, este es el supuesto clave que hace que los estimadores MCO sean insesgados. Volveremos al sesgo de variable omitida en regresión múltiple en la Sección 7.5.

Supuesto #2: $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i), i = 1, \dots, n$, son i.i.d.

El segundo supuesto es que $(X_{1i}, \dots, X_{ki}, Y_i), i = 1, \dots, n$ son variables aleatorias que están distribuidas de forma independiente e idéntica (i.i.d.). Este supuesto se cumple de forma automática si los datos se han obtenido por muestreo aleatorio simple. Los comentarios sobre este supuesto que aparecen en la Sección 4.3 para un único regresor son aplicables asimismo con múltiples regresores.

Supuesto #3: Los valores atípicos elevados son improbables

El tercer supuesto de mínimos cuadrados es que los valores extremos muy grandes son poco probables, es decir, las observaciones con valores muy alejados del rango habitual de los datos son improbables. Este supuesto sirve como recordatorio de que, como en el caso de regresor único, el estimador MCO de los coeficientes del modelo de regresión múltiple puede ser sensible a los valores atípicos de gran tamaño.

El supuesto de que los valores atípicos grandes son poco probables se formula de un modo matemáticamente preciso suponiendo que X_{1i}, \dots, X_{ki} e Y_i presentan momentos de cuarto orden finitos y distintos de cero: $0 < E(X_{1i}^4) < \infty, \dots, 0 < E(X_{ki}^4) < \infty$ y $0 < E(Y_i^4) < \infty$. Otra forma de formular este supuesto es que la variable dependiente y los regresores presentan curtosis finita. Este supuesto se utiliza para obtener las propiedades de los estadísticos de regresión MCO en muestras grandes.

Supuesto #4: Ausencia de multicolinealidad perfecta

El cuarto supuesto es nuevo para el modelo de regresión múltiple. Descarta una situación incómoda, denominada multicolinealidad perfecta, bajo la que resulta imposible calcular el estimador MCO. Se dice de los regresores que presentan **multicolinealidad perfecta** (o que son perfectamente multicolineales), si uno de los regresores es función lineal perfecta del resto de los regresores. El cuarto supuesto de mínimos cuadrados es que los regresores no son perfectamente multicolineales.

¿Por qué la multicolinealidad perfecta hace que sea imposible calcular el estimador MCO? Supóngase que se desea calcular el coeficiente de la variable *REM* en una regresión de la variable *CalificaciónExamen_i*.

sobre las variables REM_i y $PctEI_i$, pero que se comete un error tipográfico y se escribe accidentalmente la variable REM_i por segunda vez en lugar de la variable $PctEI_i$, es decir, que se regresa la variable $CalificaciónExamen_i$ sobre las variables REM_i y REM_i . Este es un caso de multicolinealidad perfecta, debido a que uno de los regresores (la primera aparición de REM) es una función lineal perfecta de otro de los regresores (la segunda aparición de la variable REM). Dependiendo de cómo trate el paquete de *software* la presencia de multicolinealidad perfecta. Al tratar de estimar esta regresión, el *software* hará una de estas dos cosas: o bien elimina una de las apariciones de la variable REM o se niega a calcular las estimaciones de MCO y da un mensaje de error. La razón matemática para este fallo es que la multicolinealidad perfecta produce un cociente con divisor igual a cero en las fórmulas de MCO.

A nivel intuitivo, la multicolinealidad perfecta constituye un problema debido a que se está pidiendo a la regresión que responda a una pregunta ilógica. En regresión múltiple, el coeficiente de uno de los regresores es el efecto de un cambio en ese regresor, manteniendo constantes el resto de los regresores. En la regresión hipotética de la variable $CalificaciónExamen$ sobre las variables REM y REM , el coeficiente de la primera aparición de la variable REM es el efecto sobre las calificaciones en los exámenes de una variación de la variable REM , manteniendo constante la otra variable REM (que es la misma). Esto no tiene sentido, por lo que MCO no puede estimar este efecto parcial sin sentido.

La solución a la aparición de la multicolinealidad perfecta en esta regresión hipotética consiste simplemente en corregir el error tipográfico y reemplazar una de las apariciones de la variable REM por la variable que originalmente se quería incluir. Este es un ejemplo típico: cuando se presenta multicolinealidad perfecta, a menudo es el reflejo de un error lógico en la elección de los regresores o bien alguna otra característica particular no reconocida previamente del conjunto de datos. En general, la solución a la presencia de multicolinealidad perfecta consiste en modificar las variables explicativas para eliminar la problema.

Se ofrecen otros ejemplos de multicolinealidad perfecta en la Sección 6.7, en la que además se define y analiza la multicolinealidad imperfecta.

Los supuestos de mínimos cuadrados para el modelo de regresión múltiple se resumen en el Concepto clave 6.4.

**CONCEPTO
CLAVE**

6.4

Los supuestos de mínimos cuadrados en el modelo de regresión múltiple

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n,$$

donde

1. u_i tiene media condicional igual a cero dados $X_{1i}, X_{2i}, \dots, X_{ki}$; es decir,

$$E(u_i | X_{1i}, X_{2i}, \dots, X_{ki}) = 0.$$

2. $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i), i = 1, \dots, n$ son extracciones independientes e idénticamente distribuidas (i.i.d.) de su distribución conjunta.
3. Los valores extremos elevados son poco probables: X_{1i}, \dots, X_{ki} e Y_i presentan momentos de cuarto orden finitos y distintos de cero.
4. No existe multicolinealidad perfecta.

6.6 La distribución de los estimadores MCO en regresión múltiple

Debido a que los datos difieren de una muestra a otra, muestras diferentes producen diferentes valores para los estimadores MCO. Esta variación entre las distintas muestras posibles da lugar a la incertidumbre asociada a los estimadores MCO de los coeficientes de regresión poblacionales $\beta_0, \beta_1, \dots, \beta_k$. Al igual que para el caso de la regresión con un único regresor, esta variación se encuentra recogida por la distribución muestral de los estimadores MCO.

Recordemos de la Sección 4.4 que, bajo los supuestos de mínimos cuadrados, los estimadores MCO ($\hat{\beta}_0$ y $\hat{\beta}_1$) son estimadores insesgados y consistentes de los coeficientes desconocidos (β_0 y β_1) en el modelo de regresión lineal con un único regresor. Además, en muestras grandes, la distribución muestral de $\hat{\beta}_0$ y $\hat{\beta}_1$ está bien aproximada mediante una distribución normal bivariante.

Estos resultados se trasladan al análisis de regresión múltiple. Es decir, bajo los supuestos de mínimos cuadrados del Concepto clave 6.4, los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ son estimadores insesgados y consistentes de $\beta_0, \beta_1, \dots, \beta_k$ en el modelo de regresión lineal múltiple. En muestras grandes, la distribución muestral conjunta de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ se aproxima bien mediante una distribución normal multivariante, que es la extensión de la distribución normal bivariante para el caso general de dos o más variables aleatorias normales conjuntas (Sección 2.4).

Aunque el álgebra resulta más complicada cuando existen varios regresores, el teorema central del límite es aplicable a los estimadores MCO en el modelo de regresión múltiple por la misma razón que se aplica a \bar{Y} y a los estimadores MCO cuando existe un único regresor: los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ son promedios de una muestra aleatoria de datos, y si el tamaño de la muestra es suficientemente grande, la distribución muestral de estos promedios se convierte en normal. Debido a que la distribución normal multivariante se maneja matemáticamente mejor utilizando álgebra matricial, las expresiones para la distribución conjunta de los estimadores MCO se tratan más adelante, en el Capítulo 18.

El Concepto clave 6.5 resume el resultado de que, en muestras grandes, la distribución conjunta de los estimadores MCO en la regresión múltiple es aproximadamente la normal. En general, los estimadores MCO están correlacionados; esta correlación se debe a la correlación entre los regresores. La distribución conjunta de los estimadores MCO se discute con más detalle para el caso de dos regresores y con errores homocedásticos en el Apéndice 6.2, y el caso general se analiza en la Sección 18.2.

**CONCEPTO
CLAVE**

6.5

Distribución para muestras grandes de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$

Si se cumplen los supuestos de mínimos cuadrados (Concepto clave 6.4), entonces en muestras grandes los estimadores MCO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ están distribuidos normalmente de forma conjunta y cada $\hat{\beta}_j$ se distribuye $N(\beta_j, \sigma_{\hat{\beta}_j}^2)$, $j = 0, \dots, k$.

6.7 Multicolinealidad

Como se analizó en la Sección 6.5, la multicolinealidad perfecta surge cuando uno de los regresores es una combinación lineal perfecta del resto de los regresores. Esta sección proporciona algunos ejemplos de multicolinealidad perfecta y en ella se explica cómo puede surgir la multicolinealidad, y cómo puede evitarse, en las regresiones con múltiples regresores binarios. La multicolinealidad imperfecta surge cuando uno de los regresores está muy altamente correlacionado, pero no perfectamente correlacionado, con los otros regresores. A diferencia de la multicolinealidad perfecta, la multicolinealidad imperfecta no impide la estimación de la regresión, ni implica un problema lógico en la selección de los regresores. Sin embargo, esto no significa que uno o más coeficientes de la regresión puedan estimarse de forma imprecisa.

Ejemplos de multicolinealidad perfecta

Continuamos la discusión sobre multicolinealidad perfecta de la Sección 6.5 mediante el examen de tres regresiones hipotéticas adicionales. En cada una de ellas, se añade un tercer regresor a la regresión de la variable $CalificaciónExamen_i$ sobre las variables REM_i y $PctEI_i$ de la Ecuación (6.12).

Ejemplo #1: Proporción de estudiantes aprendiendo inglés. Sea $FracEI_i$ la variable que representa la proporción de estudiantes que están aprendiendo inglés en el distrito i -ésimo, y que varía entre los valores 0 y 1. Si la variable $FracEI_i$ se incluyera como un tercer regresor, además de las variables REM_i y $PctEI_i$ los regresores serían perfectamente multicolineales. La razón es que la variable $PctEI$ es el porcentaje de estudiantes que están aprendiendo inglés, por lo que $PctEI_i = 100 \times FracEI_i$ para cada distrito.

Por tanto, uno de los regresores ($PctEI_i$) puede escribirse como una función lineal perfecta de otro regresor ($FraceI_i$).

Debido a esta presencia de multicolinealidad perfecta, es imposible calcular las estimaciones MCO de la regresión de la variable $CalificaciónExamen_i$ sobre las variables REM_i , $PctEI_i$ y $FraceI_i$. A un nivel intuitivo, la estimación MCO falla porque se le está preguntando: ¿Cuál es el efecto de un cambio unitario en el porcentaje de estudiantes aprendiendo inglés, manteniendo constante la proporción de estudiantes que están aprendiendo inglés? Debido a que el porcentaje de estudiantes aprendiendo inglés y la proporción de los estudiantes aprendiendo inglés se mueven juntos en una relación lineal perfecta, esta pregunta no tiene sentido y MCO no puede responderla.

Ejemplo #2: Clases «No muy pequeñas». Sea NMP_i una variable binaria que es igual a 1 si la ratio estudiantes-maestros en el distrito i -ésimo «no es muy pequeña», en concreto, NMP_i es igual a 1 si $REM_i \geq 12$ y es igual a 0 en caso contrario. Esta regresión asimismo muestra multicolinealidad perfecta, pero por una razón más sutil que la regresión del ejemplo anterior. No existen, en realidad, distritos en nuestra base de datos con $REM_i < 12$; como se puede ver en el diagrama de dispersión de la Figura 4.2, el menor valor de la variable REM es 14. Por tanto, $NMP_i = 1$ para todas las observaciones. Recordemos ahora que el modelo de regresión lineal con un término independiente puede considerarse de forma equivalente como un modelo que incluye un regresor, X_{0i} , que es igual a 1 para todo i , como se muestra en la Ecuación (6.6). Por lo tanto se puede escribir $NMP_i = 1 \times X_{0i}$ para todas las observaciones de nuestra base de datos; es decir NMP_i , se puede escribir como una combinación lineal perfecta de las variables explicativas; en concreto, es igual a X_{0i} .

Esto ilustra dos puntos importantes acerca de la multicolinealidad perfecta. En primer lugar, cuando la regresión incluye un término independiente, entonces una de las variables explicativas que puede estar implicada en la multicolinealidad perfecta es el regresor constante X_{0i} . En segundo lugar, la multicolinealidad perfecta, es una propiedad de la base de datos que se maneja. Mientras es posible imaginar un distrito escolar con menos de 12 alumnos por maestro, no existen distritos así en nuestra base de datos, por lo que no pueden analizarse en nuestra regresión.

Ejemplo #3: Porcentaje de angloparlantes. Sea la variable $PctAP_i$ el porcentaje de «angloparlantes» en el distrito i -ésimo, definido como el porcentaje de estudiantes que no están aprendiendo inglés. Una vez más los regresores presentarán multicolinealidad perfecta. Como en el ejemplo anterior, la relación lineal perfecta entre los regresores incluye el regresor constante X_{0i} : para cada distrito, $PctAP_i = 100 \times X_{0i} - PctEI_i$.

Este ejemplo ilustra otro punto: la multicolinealidad perfecta es una característica de todo el conjunto de variables explicativas. Si o bien el término independiente (es decir, el regresor X_{0i}) o bien $PctEI_i$ se excluyen de la regresión, las variables explicativas no serían perfectamente multicolineales.

La trampa de la variable ficticia. Otra posible fuente de multicolinealidad perfecta aparece cuando se utilizan distintas variables binarias, o ficticias, o *dummy*, como regresores. Por ejemplo, supongamos que se han dividido los distritos escolares en tres categorías: rurales, suburbanos y urbanos. Cada distrito se asigna a una (y solo a una) categoría. Sean estas variables binarias la variable $Rural_i$ que es igual a 1 para un distrito rural e igual 0 en caso contrario; la variable $Suburbano_i$; y la variable $Urbano_i$. Si se incluyen las tres variables binarias en la regresión, junto con una constante, los regresores presentarán multicolinealidad perfecta: debido a que cada distrito pertenece a una y solo una categoría, $Rural_i + Suburbano_i + Urbano_i = 1 = X_{i0}$, donde X_{i0} representa el regresor constante introducido en la Ecuación (6.6). Por lo tanto, para estimar la regresión, debe excluirse una de estas cuatro variables, bien sea uno de los indicadores binarios o bien el término constante. Por convención, el término constante se mantiene, en cuyo caso se excluye uno de los indicadores binarios. Por ejemplo, si se excluyera $Rural_i$, entonces el coeficiente de $Suburbano_i$ sería la diferencia promedio entre las calificaciones en los exámenes de los distritos suburbanos y de los distritos rurales, manteniendo constantes las otras variables de la regresión.

En general, si hay G variables binarias, si cada observación corresponde a una y solo una categoría, si hay un término independiente en la regresión, y si todas las variables binarias G se incluyen como variables

explicativas, entonces la regresión va a fallar a causa de la multicolinealidad perfecta. Esta situación se denomina **trampa de la variable ficticia** (*dummy*). El modo habitual de evitar la trampa de la variable ficticia es excluir a una de las variables binarias de la regresión múltiple, por lo que solamente se incluyen $G - 1$ de las G variables binarias como regresores. En este caso, los coeficientes de las variables binarias incluidas representan el efecto incremental de pertenecer a esa categoría, en relación con el caso base representado por la categoría omitida, manteniendo constantes los otros regresores. Alternativamente pueden incluirse todas variables explicativas binarias G si se omite de la regresión el término constante.

Soluciones a la multicolinealidad perfecta. La multicolinealidad perfecta aparece habitualmente cuando se ha cometido un error en la especificación de la regresión. A veces, el error es fácil de detectar (como en el primer ejemplo), pero a veces no lo es (como en el segundo ejemplo). De una forma u otra, el *software* permitirá saber si se está cometiendo un error debido a que el estimador MCO no se puede calcular.

Cuando el *software* detecta que existe multicolinealidad perfecta, es importante modificar la regresión para eliminarlo. Algunos programas resultan poco fiables cuando existe multicolinealidad perfecta, y, como mínimo, se cederá el control sobre la elección de los regresores al ordenador si los regresores presentan multicolinealidad perfecta.

Multicolinealidad imperfecta

A pesar de su similar denominación, la multicolinealidad imperfecta es conceptualmente muy diferente de la multicolinealidad perfecta. La **multicolinealidad imperfecta** significa que dos o más de los regresores están altamente correlacionados en el sentido de que existe una función lineal de los regresores que se encuentra altamente correlacionada con otro regresor. La multicolinealidad imperfecta no plantea ningún problema para la teoría de los estimadores MCO; de hecho, un propósito de MCO es separar las influencias independientes de los distintos regresores cuando estos regresores están potencialmente correlacionados.

Si los regresores presentan multicolinealidad imperfecta, entonces los coeficientes de al menos un regresor individual se estimarán de forma imprecisa. Por ejemplo, considérese la regresión de la variable *CalificaciónExamen* sobre las variables *REM* y *PctEI*. Supongamos que se tuviera que añadir un tercer regresor, el porcentaje de residentes en el distrito que son inmigrantes de primera generación. Los inmigrantes de primera generación frecuentemente hablan inglés como segunda lengua, por lo que las variables *PctEI* y el porcentaje de inmigrantes de primera generación estarán altamente correlacionados: los distritos con muchos inmigrantes recién llegados suelen tener muchos estudiantes que todavía están aprendiendo inglés. Debido a que estas dos variables están altamente correlacionadas, sería difícil utilizar estos datos para estimar el efecto parcial sobre las calificaciones en los exámenes de un aumento en la variable *PctEI*, manteniendo constante el porcentaje de inmigrantes. En otras palabras, la base de datos proporciona poca información sobre lo que ocurre con las calificaciones en los exámenes cuando el porcentaje de estudiantes aprendiendo inglés es bajo, pero la proporción de inmigrantes es elevada, o viceversa. Si se cumplen los supuestos de mínimos cuadrados, entonces el estimador MCO del coeficiente de *PctEI* en esta regresión será insesgado; sin embargo, tendrá una varianza mayor que si los regresores *PctEI* y porcentaje de inmigrantes no estuvieran correlacionados.

El efecto de la multicolinealidad imperfecta sobre la varianza de los estimadores MCO se puede comprobar matemáticamente mediante la inspección de la Ecuación (6.17) en el Apéndice 6.2, que es la varianza de $\hat{\beta}_1$ en una regresión múltiple con dos variables explicativas (X_1 y X_2) para el caso particular de un error homocedástico. En este caso, la varianza de $\hat{\beta}_1$ es inversamente proporcional a $1 - \rho_{X_1, X_2}^2$, donde ρ_{X_1, X_2} es la correlación entre X_1 y X_2 . Cuanto mayor sea la correlación entre las dos variables explicativas, más cercano estará este término a cero y mayor es la varianza de $\hat{\beta}_1$. De manera más general, cuando varios regresores presentan multicolinealidad imperfecta, los coeficientes de uno o más de estos regresores se estimarán de manera imprecisa, es decir, tendrán mayor varianza muestral.

La multicolinealidad perfecta es un problema que a menudo indica la presencia de un error lógico. Por el contrario, la multicolinealidad imperfecta no es necesariamente un error, sino más bien una característica

de MCO, de los datos, y de la pregunta que se está tratando de responder. Si las variables de la regresión son las únicas que se pretenden incluir —las que se han elegido para hacer frente a la posibilidad de sesgo de variable omitida— entonces la multicolinealidad imperfecta implica que será difícil estimar con precisión uno o más de los efectos parciales utilizando los datos disponibles.

6.8 Conclusión

La regresión con regresor único es vulnerable al sesgo de variable omitida: si una variable omitida es un factor determinante de la variable dependiente y está correlacionada con el regresor, entonces el estimador MCO del coeficiente de la pendiente será sesgado y reflejará tanto el efecto del regresor como el efecto de la variable omitida. La regresión múltiple hace posible mitigar el sesgo de variable omitida mediante la inclusión de la variable omitida en la regresión. El coeficiente de un regresor, X_1 , en regresión múltiple es el efecto parcial de un cambio en X_1 , manteniendo constante el resto de variables explicativas incluidas. En el ejemplo de las calificaciones en los exámenes, la inclusión del porcentaje de estudiantes que están aprendiendo inglés como regresor hace posible estimar el efecto sobre las calificaciones obtenidas de un cambio en la ratio estudiantes-maestros, manteniendo constante el porcentaje de estudiantes que están aprendiendo inglés. Hacerlo reduce a la mitad el efecto estimado sobre las calificaciones en los exámenes de un cambio en la ratio estudiantes-maestros.

La teoría estadística de regresión múltiple se basa en la teoría estadística de regresión con un único regresor. Los supuestos de mínimos cuadrados para regresión múltiple son ampliaciones de los tres supuestos de mínimos cuadrados de la regresión con un solo regresor, más un cuarto supuesto que descarta la multicolinealidad perfecta. Debido a que los coeficientes de regresión se estiman utilizando una única muestra, los estimadores MCO presentan una distribución muestral conjunta y, por tanto, presentan incertidumbre debida al muestreo. Esta incertidumbre muestral debe ser cuantificada como parte de un estudio empírico, y las formas de hacerlo en el modelo de regresión múltiple constituyen el tema del siguiente capítulo.

Resumen

1. El sesgo de variable omitida se produce cuando una variable omitida (1) está correlacionada con un regresor ya incluido y (2) es un factor determinante de Y .
2. El modelo de regresión múltiple es un modelo de regresión lineal que incluye varios regresores, X_1, X_2, \dots, X_k . Asociado con cada regresor existe un coeficiente de regresión, $\beta_1, \beta_2, \dots, \beta_k$. El coeficiente β_1 es la variación esperada en Y asociada con una variación unitaria de X_1 , manteniendo constantes el resto de regresores. Los otros coeficientes de la regresión tienen una interpretación análoga.
3. Los coeficientes de regresión múltiple pueden estimarse por MCO. Cuando se cumplen los cuatro supuestos de mínimos cuadrados del Concepto clave 6.4, los estimadores MCO son insesgados, consistentes y están normalmente distribuidos en muestras grandes.
4. La multicolinealidad perfecta, que ocurre cuando uno de los regresores es una combinación lineal exacta del resto, aparece habitualmente por un error en la elección de los regresores incluidos en la regresión múltiple. La solución a la multicolinealidad perfecta requiere cambiar el conjunto de regresores.
5. El error estándar de la regresión, el R^2 y el \bar{R}^2 son medidas de ajuste del modelo de regresión múltiple.

Términos clave

sesgo de variable omitida (130)

coeficiente de la pendiente de X_{1i} (134)

modelo de regresión múltiple (134)

coeficiente de X_{1i} (134)

recta de regresión poblacional (134)

coeficiente de la pendiente de X_{2i} (134)

función de regresión poblacional (134)

coeficiente de X_{2i} (134)

intercepto (término independiente) (134)

manteniendo X_2 constante (134)

controlando por X^2 (134)	recta de regresión MCO (137)
efecto parcial (135)	valor de predicción (137)
modelo de regresión múltiple poblacional (135)	residuo MCO (137)
regresor constante (135)	R^2 (139)
término constante (135)	R^2 ajustado (\bar{R}^2) (140)
homocedástico (136)	multicolinealidad perfecta (141)
heterocedástico (136)	trampa de la variable ficticia (145)
estimadores de mínimos cuadrados ordinarios (MCO) de $\beta_0, \beta_1, \dots, \beta_k$ (137)	multicolinealidad imperfecta (145)

Revisión de conceptos

- 6.1** Una investigadora está interesada en el efecto sobre las calificaciones en los exámenes del uso del ordenador. Utilizando los datos del distrito escolar como los utilizados en este capítulo, regresa la media de las calificaciones en el examen del distrito sobre el número de ordenadores por estudiante. ¿Será $\hat{\beta}_1$ un estimador insesgado del efecto sobre las calificaciones en los exámenes de un aumento en el número de ordenadores por estudiante? ¿Por qué o por qué no? Si se cree que $\hat{\beta}_1$ está sesgado, estará sesgado al alza o a la baja? ¿Por qué?
- 6.2** Una regresión múltiple incluye dos regresores: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. ¿Cuál es la variación esperada en Y si X_1 aumenta en 3 unidades y X_2 no varía? ¿Cuál es la variación esperada en Y si se reduce X_2 en 5 unidades y X_1 no se modifica? ¿Cuál es la variación esperada en Y si X_1 aumenta en 3 unidades y X_2 disminuye en 5 unidades?
- 6.3** Explique por qué dos regresores perfectamente multicolineales no pueden ser incluidos en una regresión lineal múltiple. Proponga dos ejemplos de una pareja de regresores que presenten multicolinealidad perfecta.
- 6.4** Explique por qué es difícil estimar con precisión el efecto parcial de X_1 , manteniendo constante X_2 , si X_1 y X_2 están altamente correlacionados.

Ejercicios

Los primeros cuatro ejercicios se refieren a la tabla de regresiones estimadas de la página 148, calculadas utilizando los datos de 1998 de la Encuesta Actualizada de Población (CPS). La base de datos consta de información sobre 4.000 trabajadores a tiempo completo durante todo el año. El mayor grado educativo alcanzado por cada trabajador es o bien un diploma de escuela secundaria o bien un título de licenciatura. El rango de edades de los trabajadores oscila entre los 25 y los 34 años. La base de datos asimismo contiene información sobre la región del país donde reside la persona, el estado civil y el número de hijos. A los efectos de estos ejercicios, sean

IMH = ingresos medios por hora (en dólares de 1998).

Universidad = variable binaria (1 si titulado en universidad, 0 si titulado escuela secundaria).

Femenino = variable binaria (1 si es mujer, 0 si hombre).

Edad = edad (en años).

Noreste = variable binaria (1 = si Región = Noreste, 0 en caso contrario).

Centro-Oeste = variable binaria (1 = si Región = Centro-Oeste, 0 en caso contrario).

Sur = variable binaria (1 = si Región = Sur, 0 en caso contrario).

Oeste = variable binaria (1 = si Región = Oeste, 0 en caso contrario).

- 6.1** Calcule \bar{R}^2 para cada una de las regresiones.

- 6.2** Utilizando los resultados de la regresión de la columna (1):

- a)** ¿Ganan más los trabajadores con títulos universitarios en promedio que los trabajadores con tan solo grado de secundaria? ¿Cuánto más?
- b)** ¿Ganan los hombres más que las mujeres en promedio? ¿Cuánto más?

6.3 Utilizando los resultados de la regresión de la columna (2):

- a) ¿Es la edad un determinante importante para los ingresos? Explíquelo.
- b) Sally es una mujer titulada universitaria de 29 años de edad. Betsy es una mujer titulada universitaria de 34 años de edad. Prediga los ingresos de Sally y los de Betsy.

6.4 Utilizando los resultados de regresión de la columna (3):

- a) ¿Parece que existen diferencias regionales importantes?
- b) ¿Por qué se ha omitido la variable explicativa Oeste de la regresión? ¿Qué sucedería si se incluyese?

Resultados de las regresiones de los ingresos salariales medios por hora sobre el género y las variables binarias de educación y otras características utilizando datos de 1998 de la encuesta actualizada de población			
Variable dependiente: ingresos salariales medios por hora (AHE).			
Regresor	(1)	(2)	(3)
Universidad (X_1)	5,46	5,48	5,44
Femenino (X_2)	-2,64	-2,62	-2,62
Edad (X_3)		0,29	0,29
Noreste (X_4)			0,69
Centro-Oeste (X_5)			0,60
Sur (X_6)			-0,27
Intercepto	12,69	4,40	3,75
Estadísticos de resumen			
ESR	6,27	6,22	6,21
R^2	0,176	0,190	0,194
\bar{R}^2			
n	4.000	4.000	4.000

- c) Juanita es una mujer titulada universitaria de 28 años de edad de la región Sur. Jennifer es una mujer de 28 años de edad, titulada universitaria de la región Centro-Oeste. Calcule la diferencia esperada entre los ingresos de Juanita y los de Jennifer.

6.5 Los datos fueron recogidos de una muestra aleatoria de 220 viviendas vendidas en una comunidad en el año 2003. Sea la variable *Precio* la que recoge el precio de venta (en miles de \$), sea *DORM* la variable que expresa el número de dormitorios, la variable *Baños* indica el número de cuartos de baño, la variable *CTam* indica el tamaño de la vivienda (en pies cuadrados), la variable *PTam* expresa el tamaño de la parcela (en pies cuadrados), la variable *Edad* expresa la edad de la vivienda (en años), y la variable *Pobre* es una variable binaria que es igual a 1 si el estado general de la casa se puede calificar como «pobre». La estimación de la regresión ofrece estos resultados

$$\widehat{\text{Precio}} = 119,2 + 0,485\text{DORM} + 23,4\text{Baño} + 0,156\text{CTam} + 0,002\text{PTam} + \\ + 0,090\text{Edad} - 48,8\text{Pobre}, \bar{R}^2 = 0,72, \text{ESR} = 41,5$$

- a) Supóngase que un propietario convierte parte de una sala de estar que ya existía en la vivienda en un cuarto de baño. ¿Cuál es el aumento esperado en el valor de la casa?
- b) Supóngase que el propietario añade un nuevo cuarto de baño a su casa, lo que aumenta el tamaño de la vivienda en 100 pies cuadrados. ¿Cuál es el aumento esperado en el valor de la casa?

- c) ¿Cuál es la pérdida de valor si un propietario deja deteriorar su vivienda hasta que las condiciones generales la convierten en «pobre»?
- d) Calcule el R^2 de la regresión.
- 6.6** Un investigador planea estudiar el efecto causal de la policía sobre el crimen a partir de los datos de una muestra aleatoria de condados de los EE.UU. Se plantea realizar una regresión de la tasa de criminalidad del condado sobre el tamaño (per cápita) del cuerpo de policía del condado.
- Explique por qué esta regresión probablemente presente sesgo de variable omitida. ¿Qué variables añadiría a la regresión para controlar por las variables omitidas importantes?
 - Utilice su respuesta en (a) y la expresión del sesgo de variable omitida dado en la Ecuación (6.1) para determinar si la regresión probablemente sobre o subestime el efecto de la policía sobre la tasa de criminalidad. (Es decir, ¿piensa que $\hat{\beta}_1 > \beta_1$ o que $\hat{\beta}_1 < \beta_1$?).
- 6.7** Realice una crítica sobre cada uno de los siguientes planes de investigación propuestos. Su crítica debe explicar cualquier problema de la investigación propuesta y describir cómo el plan de investigación podría ser mejorado. Incluya una discusión sobre cualquier otro dato adicional que necesite, así como sobre las técnicas estadísticas apropiadas para el análisis de los datos.
- Un investigador está interesado en determinar si una gran empresa aeroespacial es culpable de discriminación por género en la fijación de salarios. Para determinar el sesgo potencial, el investigador recoge información sobre los salarios y el género sobre todos los ingenieros de la empresa. El investigador tiene previsto entonces realizar un contraste de «diferencia de medias» para determinar si el salario medio de las mujeres es significativamente menor que el salario medio de los hombres.
 - Un investigador está interesado en determinar si el tiempo pasado en prisión tiene un efecto permanente sobre el salario de una persona. Recoge datos de una muestra aleatoria de personas que han estado fuera de la cárcel durante al menos 15 años. Recoge datos similares sobre una muestra aleatoria de personas que nunca han estado en prisión. La base de datos incluye información sobre el salario actual de cada persona, la educación, la edad, el grupo étnico al que pertenece, el género, y la permanencia (el tiempo que lleva en el puesto de trabajo actual), la ocupación, el estatus sindical, así como si la persona ha sido alguna vez encarcelada. El investigador planea estimar el efecto de la encarcelación sobre los salarios mediante la regresión de los salarios sobre una variable indicadora del encarcelamiento, incluyendo en la regresión los otros determinantes potenciales de los salarios (la educación, la permanencia, el estatus sindical, etc.).
- 6.8** Un estudio reciente encontró que la tasa de mortalidad de las personas que duermen de 6 a 7 horas cada noche es más baja que la tasa de mortalidad de las personas que duermen ocho horas o más. Los 1.1 millones de observaciones utilizadas para este estudio provienen de una encuesta aleatoria de estadounidenses de entre 30 y 102 años de edad. A cada encuestado que respondió a la encuesta se le realizó un seguimiento durante 4 años. La tasa de mortalidad de las personas que dormían siete horas se calculó como la ratio entre el número de muertes durante el lapso del estudio entre las personas que dormían 7 horas sobre el número total de encuestados que respondieron que dormían 7 horas. Este cálculo se repitió para las personas que dormían seis horas, y así sucesivamente. En base en este resumen, ¿recomendaría a los estadounidenses que duermen 9 horas por noche que considerasen la reducción de su sueño a 6 ó 7 horas si quieren prolongar su vida? ¿Por qué o por qué no? Explíquelo.
- 6.9** (Y_i, X_{1i}, X_{2i}) satisfacen los supuestos del Concepto clave 6.4. Se tiene interés en β_1 , el efecto causal de X_1 sobre Y . Supóngase que X_1 y X_2 no están correlacionadas. Se estima β_1 mediante la regresión de Y sobre X_1 (por lo que X_2 no está incluida en la regresión). ¿Este estimador presenta un sesgo de variable omitida? Explíquelo.
- 6.10** (Y_i, X_{1i}, X_{2i}) satisfacen los supuestos del Concepto clave 6.4; además $\text{var}(u_i | X_{1i}, X_{2i}) = 4$, y $\text{var}(X_{1i}) = 6$. Se extrae una muestra aleatoria de tamaño $n = 400$ de la población.
- Supóngase que X_1 y X_2 no están correlacionadas. Calcule la varianza de $\hat{\beta}_1$. [Sugerencia: mire la Ecuación (6.17) del Apéndice 6.2.]

- b) Supóngase que $\text{Corr}(X_1, X_2) = 0,5$. Calcule la varianza de $\hat{\beta}_1$.
- c) Comente las siguientes afirmaciones: «Si X_1 y X_2 están correlacionadas, la varianza de $\hat{\beta}_1$ es mayor de lo que lo sería si X_1 y X_2 no estuvieran correlacionadas. Por tanto, si interesa β_1 , es mejor dejar fuera de la regresión a X_2 si está correlacionada con X_1 ».

6.11 (Precisa cálculo) Considérese el modelo de regresión

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

para $i = 1, \dots, n$. (Téngase en cuenta que existe término constante en la regresión). Siguiendo el análisis utilizado en el Apéndice 4.2:

- a) Especifique la función de mínimos cuadrados que se minimiza por MCO.
- b) Calcular las derivadas parciales de la función objetivo con respecto a β_1 y β_2 .
- c) Supóngase que $\sum_{i=1}^n X_{1i} X_{2i} = 0$. Demuestre que $\hat{\beta}_1 = \sum_{i=1}^n X_{1i} Y_i / \sum_{i=1}^n X_{1i}^2$.
- d) Supóngase que $\sum_{i=1}^n X_{1i} X_{2i} \neq 0$. Obtenga una expresión para $\hat{\beta}_1$ como función de los datos (Y_i, X_{1i}, X_{2i}) , $i = 1, \dots, n$.
- e) Supóngase que el modelo incluye un intercepto:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$
 Demuestre que los estimadores MCO cumplen $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$.
- f) Supóngase como en e) que el modelo contiene un intercepto. Supóngase además que $\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$. Demuestre que $\hat{\beta}_1 = \sum_{i=1}^n (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) / \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2$. ¿Cómo se compara con el estimador MCO de β_1 de la regresión en la que se omite X_2 ?

Ejercicios empíricos

E6.1 Con la base de datos **TeachingRatings** descrita en los Ejercicios empíricos 4.2, lleve a cabo los siguientes ejercicios:

- a) Realice una regresión de la variable *Course_Eval* (recuerde que son las calificaciones en los exámenes) sobre la variable *Beauty* (la variable que mide la belleza del profesor). ¿Cuál es la pendiente estimada?
- b) Realice una regresión de la variable *Course_Eval* sobre la variable *Beauty*, incluyendo algunas variables de control adicionales del tipo de curso y de las características del profesor. En particular, incluya como regresores adicionales las variables *Intro*, *OneCredit*, *Female*, *Minority* y *NNEnglish*. ¿Cuál es el efecto estimado de la variable *Beauty* sobre la variable *Course_Eval*? ¿Presenta la regresión (a) un sesgo de variable omitida importante?
- c) Estime el coeficiente de la variable *Beauty* del modelo de regresión múltiple en (b) mediante el proceso en tres etapas del Apéndice 6.3 (el Teorema de Frisch-Waugh). Verifique que el proceso en tres etapas obtiene el mismo coeficiente estimado para la variable *Beauty* que el obtenido en (b).
- d) El profesor Smith es un hombre negro con un valor de la variable *Beauty* promedio y es angloparlante nativo. Es profesor de una asignatura de tres créditos del curso superior. Prediga la evaluación de la asignatura del profesor Smith.

E6.2 Con la base de datos **CollegeDistance** descrita en el Ejercicio Empírico 4.3, realice los siguientes ejercicios:

- a) Realice una regresión de la variable años de educación completados (ED) sobre la variable de la distancia a la universidad más cercana (*Dist*). ¿Cuál es la pendiente estimada?
- b) Realice una regresión de la variable *ED* sobre la variable *Dist*, pero incluyendo algunos regresores adicionales de control sobre las características del estudiante, la familia del estudiante, y el mercado laboral local. En concreto, incluyendo como regresores adicionales las variables *Bytest*, *Female*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *DadColl*, *Cue80* y *Stwmfg80*. ¿Cuál es el efecto estimado de la variable *Dist* sobre la variable *ED*?

- c) ¿Es sustancialmente distinto el efecto estimado de la variable *Dist* sobre la variable *ED* en la regresión de (b) de la regresión en (a)? En base a esto, ¿parece que la regresión a) presente un sesgo de variable omitida importante?
- d) Compare el ajuste de la regresión de (a) y de (b) utilizando los errores estándar de regresión, R^2 y \bar{R}^2 . ¿Por qué R^2 y \bar{R}^2 son tan similares en la regresión (b)?
- e) El valor del coeficiente de la variable *DadColl* es positivo. ¿Qué mide este coeficiente?
- f) Explique por qué las variables *Cue80* y *Swmfg80* aparecen en la regresión. ¿Cuáles cree que son los signos de sus coeficientes estimados (+ o -)? Interprete la magnitud de estos coeficientes.
- g) Bob es un hombre negro. Su escuela secundaria estaba a 20 millas de la universidad más cercana. Su calificación en la prueba (*Bytest*) fue de 58. Su renta familiar en 1980 fue de 26.000 \$, y su familia poseía una casa. Su madre acudió a la universidad, pero su padre no. La tasa de desempleo en su condado era del 7,5 %, y el promedio del salario por hora manufacturero en su estado era de 9,75 \$. Estime el número de años completados de estudio por Bob utilizando la regresión de (b).
- h) Jim tiene las mismas características que Bob, salvo que su escuela secundaria estaba a 40 millas de la universidad más cercana. Estime los años completados de estudio por Jim utilizando la regresión de (b).
- E6.3** Con la base de datos **Growth** descrita en el Ejercicio Empírico 4.4, pero excluyendo los datos de Malta, realice los siguientes ejercicios:
- Construya una tabla que muestre la media muestral, la desviación típica, y los valores mínimo y máximo de las series *Growth*, *TradeShare*, *YearsSchool*, *Oil*, *Rev_Coups*, *Assassinations*, *RGDP60*. Incluya las unidades apropiadas para todas las series.
 - Realice una regresión de la variable *Growth* sobre las variables *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations* y *RGDP60*. ¿Cuál es el valor del coeficiente de la variable *Rev_Coups*? Interprete el valor de este coeficiente. ¿Es grande o pequeño para el mundo real?
 - Utilice la regresión para predecir la tasa media de crecimiento anual para un país que presenta valores promedios para todas las variables explicativas.
 - Repita c), pero ahora suponiendo que el valor de la variable *TradeShare* para el país es de una desviación típica por encima de la media.
 - ¿Por qué se omite la variable *Oil* de la regresión? ¿Qué pasaría si se incluyese?

APÉNDICE

6.1 Obtención de la ecuación (6.1)

Este apéndice presenta una deducción para la obtención de la fórmula del sesgo de variable omitida de la Ecuación (6.1). La Ecuación (4.30) en el Apéndice 4.3 establece que

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (6.16)$$

Bajo los dos últimos supuestos del Concepto clave 4.3, $(1/n) \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow{p} \sigma_X^2$ y $(1/n) \sum_{i=1}^n (X_i - \bar{X})u_i \xrightarrow{p} \text{cov}(u_i, X_i) = \rho_{Xu}\sigma_u\sigma_X$. Mediante la sustitución de estos límites en la Ecuación (6.16) se obtiene la Ecuación (6.1).

APÉNDICE

6.2 Distribución de los estimadores MCO en presencia de dos regresores y errores homocedásticos

Aunque la fórmula general para la varianza de los estimadores MCO en regresión múltiple es complicada, con dos variables explicativas ($k = 2$) y si los errores son homocedásticos, entonces la fórmula se simplifica lo suficiente como para proporcionar alguna información sobre la distribución de los estimadores MCO.

Debido a que los errores son homocedásticos, la varianza condicional de u_i puede escribirse como $\text{var}(u_i|X_{1i}, X_{2i}) = \sigma_u^2$. Cuando hay dos variables explicativas, X_{1i} y X_{2i} , y el término de error es homocedástico, en muestras grandes la distribución muestral de $\hat{\beta}_1$ es $N(\beta_1, \sigma_{\hat{\beta}_1}^2)$, donde la varianza de esta distribución, $\sigma_{\hat{\beta}_1}^2$, es

$$\sigma_{\hat{\beta}_1}^2 = \frac{1}{n} \left(\frac{1}{1 - \rho_{X_1, X_2}^2} \right) \frac{\sigma_u^2}{\sigma_{X_1}^2} \quad (6.17)$$

donde ρ_{X_1, X_2} es la correlación poblacional entre las dos variables explicativas X_1 y X_2 , y $\sigma_{X_1}^2$ es la varianza poblacional de X_1 .

La varianza $\sigma_{\hat{\beta}_1}^2$ de la distribución muestral de $\hat{\beta}_1$ depende del cuadrado de la correlación entre los regresores. Si X_1 y X_2 están altamente correlacionadas, ya sea positiva o negativamente, entonces ρ_{X_1, X_2}^2 se acerca a 1 y por tanto el término $1 - \rho_{X_1, X_2}^2$ en el denominador de la Ecuación (6.17) es pequeño y la varianza de $\hat{\beta}_1$ es mayor de lo que lo sería si ρ_{X_1, X_2} se acercara a 0.

Otra característica de la distribución normal conjunta para muestras grandes de los estimadores MCO es $\hat{\beta}_1$ y $\hat{\beta}_2$ están, en general, correlacionados. Cuando los errores son homocedásticos, la correlación entre los estimadores MCO $\hat{\beta}_1$ y $\hat{\beta}_2$ es el opuesto (cambia de signo) de la correlación entre los dos regresores:

$$\text{corr}(\hat{\beta}_1, \hat{\beta}_2) = -\rho_{X_1, X_2}. \quad (6.18)$$

APÉNDICE

6.3 El teorema de Frisch-Waugh

El estimador MCO en regresión múltiple se puede calcular mediante una serie de regresiones más cortas. Consideremos el modelo de regresión múltiple de la Ecuación (6.7). El estimador MCO de β_1 se puede calcular en tres etapas:

1. Se regresa X_1 sobre X_2, X_3, \dots, X_k y sea \tilde{X}_1 la expresión de los residuos de esta regresión.
2. Se regresa Y sobre X_2, X_3, \dots, X_k y sea \tilde{Y} la expresión de los residuos de esta regresión.
3. Se regresa \tilde{Y} sobre \tilde{X}_1 ,

donde las regresiones incluyen un término constante (intercepto). El teorema de Frisch-Waugh establece que el coeficiente MCO de la etapa 3 es igual al coeficiente de MCO de X_1 del modelo de regresión múltiple (6.7).

Este resultado proporciona una formulación matemática de la forma en que el coeficiente de regresión múltiple $\hat{\beta}_1$ estima el efecto sobre Y de X_1 , controlando por las otras X : debido a que las dos primeras regresiones (etapas 1 y 2) eliminan de Y y X_1 su variabilidad asociada a las otras X , la tercera regresión estima el efecto sobre Y de X_1 utilizando lo que queda después de eliminar (controlar por) el efecto de las otras X . El teorema de Frisch-Waugh se demuestra en el Ejercicio 18.17.

Este teorema sugiere de qué manera la Ecuación (6.17) se puede deducir a partir de la Ecuación (5.27). Debido a que $\hat{\beta}_1$ es el coeficiente de regresión MCO de la regresión de \tilde{Y} sobre \tilde{X}_1 , la Ecuación (5.27) sugiere que la varianza válida con homocedasticidad de $\hat{\beta}_1$ es $\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_u^2}{n \sigma_{\tilde{X}_1}^2}$, donde $\sigma_{\tilde{X}_1}^2$ es la varianza de \tilde{X}_1 . Debido a que \tilde{X}_1 es el residuo de la regresión de X_1 sobre X_2 (recordemos que la Ecuación (6.17) se refiere al modelo con $k = 2$ regresores), la Ecuación (6.15) implica que $s_{\tilde{X}_1}^2 = (1 - \bar{R}_{X_1 X_2}^2) s_{X_1}^2$, donde $\bar{R}_{X_1 X_2}^2$ es el R^2 ajustado de la regresión de X_1 sobre X_2 . La Ecuación (6.17) se deduce de $s_{\tilde{X}_1}^2 \xrightarrow{P} \sigma_{\tilde{X}_1}^2$, $\bar{R}_{X_1 X_2}^2 \xrightarrow{P} \rho_{X_1 X_2}^2$ y $s_{X_1}^2 \xrightarrow{P} \sigma_{X_1}^2$.

Contrastes de hipótesis e intervalos de confianza en regresión múltiple

Tal y como se analizó en el Capítulo 6, el análisis de regresión múltiple proporciona una manera de mitigar el problema del sesgo de variable omitida mediante la inclusión de regresores adicionales, es decir, controlando por los efectos de estos regresores adicionales. Los coeficientes del modelo de regresión múltiple se pueden estimar por MCO. Al igual que todos los estimadores, el estimador MCO presenta la incertidumbre asociada al muestreo debido a que su valor difiere de una muestra a otra.

Este capítulo presenta los métodos para la cuantificación de la incertidumbre de muestreo del estimador MCO a través de la utilización de errores estándar, contrastes de hipótesis estadísticos e intervalos de confianza. Una nueva posibilidad que aparece en regresión múltiple es una hipótesis que involucra simultáneamente a dos o más coeficientes de regresión. El método general para contrastar esas hipótesis «conjuntas» incluye un nuevo estadístico de contraste, el estadístico F .

La Sección 7.1 amplía los métodos de inferencia estadística de la regresión con regresor único a la regresión múltiple. Las Secciones 7.2 y 7.3 muestran cómo contrastar hipótesis que involucran dos o más coeficientes de regresión. La Sección 7.4 generaliza la noción de intervalos de confianza para un coeficiente a conjuntos de confianza para múltiples coeficientes. La decisión acerca de qué variables incluir en una regresión es un problema práctico importante, por lo que la Sección 7.5 trata los métodos para abordar este problema. En la Sección 7.6, se aplica el análisis de regresión múltiple para la obtención de estimaciones mejoradas del efecto sobre las calificaciones en los exámenes de una reducción en la ratio estudiantes-maestros a partir de la base de datos de California para las calificaciones en los exámenes.

7.1 Contrastes de hipótesis e intervalos de confianza para un único coeficiente

En esta sección se describe cómo calcular el error estándar, cómo contrastar hipótesis y cómo construir intervalos de confianza para un único coeficiente de una ecuación de regresión múltiple

Errores estándar de los estimadores MCO

Hay que recordar que, en el caso de un único regresor, era posible estimar la varianza del estimador MCO mediante la sustitución de las medias muestrales por las esperanzas, lo que conducía al estimador $\hat{\sigma}_{\beta_1}^2$ dado en la Ecuación (5.4). Bajo los supuestos de mínimos cuadrados, la ley de grandes números implica que estas medias muestrales convergen a sus homólogas poblacionales, por lo que, por ejemplo, $\hat{\sigma}_{\beta_1}^2/\sigma_{\beta_1}^2 \xrightarrow{P} 1$. La raíz cuadrada de $\hat{\sigma}_{\beta_1}^2$ es el error estándar de $\hat{\beta}_1$, $ES(\hat{\beta}_1)$, un estimador de la desviación típica de la distribución muestral de $\hat{\beta}_1$.

Todo esto es directamente extensible a la regresión múltiple. El estimador MCO $\hat{\beta}_j$ del j -ésimo coeficiente de regresión tiene una desviación típica, y esta desviación típica se estima mediante su error estándar, $ES(\hat{\beta}_j)$. La fórmula del error estándar es más fácil de definir mediante el uso de matrices (véase Sección 18.2). Lo importante es que, en lo que se refiere a los errores estándar, no hay nada conceptualmente diferente entre los casos de uno o varios regresores. Las ideas clave, la normalidad de los estimadores en muestras grandes y la posibilidad de estimar consistentemente la desviación típica de su distribución muestral, son las mismas, ya sea con uno, dos, o 12 regresores.

Contrastes de hipótesis para un único coeficiente

Supongamos que se desea contrastar la hipótesis de que una variación en la ratio estudiantes-maestros no tiene ningún efecto sobre las calificaciones obtenidas en los exámenes, manteniendo constante el porcentaje de estudiantes en el distrito que están aprendiendo inglés. Esto se corresponde con la hipótesis de que el verdadero coeficiente β_1 de la ratio estudiantes-maestros es igual a cero en la regresión poblacional de las calificaciones en los exámenes sobre REM y PctEI. De forma más general, podría desearse contrastar la hipótesis de que el verdadero coeficiente β_j del j -ésimo regresor toma un valor específico, $\beta_{j,0}$. El valor bajo la hipótesis nula, $\beta_{j,0}$, proviene o bien de la teoría económica o bien, como en el ejemplo del número de estudiantes por maestro, de una decisión tomada en el contexto de un caso o aplicación concreta. Si la hipótesis alternativa es bilateral, entonces las dos hipótesis se pueden expresar matemáticamente como

$$H_0: \beta_j = \beta_{j,0} \text{ vs. } H_1: \beta_j \neq \beta_{j,0} \quad (\text{alternativa bilateral}). \quad (7.1)$$

Por ejemplo, si el primer regresor es REM, entonces la hipótesis nula de que el cambio en la ratio estudiantes-maestros no tiene ningún efecto sobre las calificaciones en los exámenes corresponde con la hipótesis nula de que $\beta_1 = 0$ (por lo que $\beta_{1,0} = 0$). Nuestra tarea es la de contrastar la hipótesis nula H_0 frente a la alternativa H_1 utilizando una muestra de datos.

El Concepto clave 5.2 proporciona un procedimiento para el contraste de esta hipótesis nula en el caso de un único regresor. El primer paso de este procedimiento consiste en calcular el error estándar del coeficiente. El segundo paso es calcular el estadístico t utilizando la fórmula general del Concepto clave 5.1. El tercer paso consiste en calcular el p -valor del contraste, utilizando la distribución normal estándar acumulada de la Tabla 1 del apéndice, o bien, comparar el estadístico t con el valor crítico que corresponde al nivel de significación deseado para el contraste. Los fundamentos teóricos de este procedimiento son que el estimador MCO tiene una distribución normal en muestras grandes que, bajo la hipótesis nula, tiene como media el verdadero valor bajo la hipótesis nula y que la varianza de esta distribución puede estimarse de modo consistente.

Esta base está presente asimismo en regresión múltiple. Tal y como quedó establecido en el Concepto clave 6.5, la distribución muestral de $\hat{\beta}_j$ es aproximadamente normal. Bajo la hipótesis nula, la media de esta distribución es $\beta_{j,0}$. La varianza de esta distribución puede estimarse consistentemente. Por tanto podemos simplemente seguir el mismo procedimiento que para el caso de un solo regresor para contrastar la hipótesis nula de la Ecuación (7.1).

El procedimiento para contrastar una hipótesis sobre un solo coeficiente en regresión múltiple se recoge en el Concepto clave 7.1. El estadístico t efectivamente calculado se expresa en este recuadro como t^{act} . Sin embargo, es habitual referirse a él simplemente como t , por lo que adoptaremos esta notación simplificada para el resto del libro.

CONCEPTO CLAVE

Contraste de la hipótesis $\beta_j = \beta_{j,0}$ frente a la alternativa $\beta_j \neq \beta_{j,0}$

- Calcular el error estándar de $\hat{\beta}_j$, $ES(\hat{\beta}_j)$.

7.1

- Calcular el estadístico t ,

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{ES(\hat{\beta}_j)} \quad (7.2)$$

- Calcular el p -valor

$$p\text{-valor} = 2\Phi(-|t^{act}|) \quad (7.3)$$

donde t^{act} es el valor del estadístico t calculado en realidad. Se rechaza la hipótesis al nivel de significación del 5 % si el p -valor es inferior a 0,05 o, equivalentemente, si $|t^{act}| > 1,96$.

El error estándar y (generalmente) el estadístico t y el p -valor para el contraste de $\beta_j = 0$ se calcula automáticamente mediante el *software* de regresión.

Intervalos de confianza para un único coeficiente

El método para la construcción de un intervalo de confianza en el modelo de regresión múltiple es del mismo modo igual al del modelo de un solo regresor. Este método se recoge en el Concepto clave 7.2.

El método para llevar a cabo un contraste de hipótesis en el Concepto clave 7.1 y el método para construir un intervalo de confianza en el Concepto clave 7.2 se basan en la aproximación normal para muestras grandes de la distribución del estimador MCO, $\hat{\beta}_j$. En consecuencia, hay que tener en cuenta que estos métodos para la cuantificación de la incertidumbre muestral solo están garantizados si se trabaja con muestras grandes.

**CONCEPTO
CLAVE**

7.2

Intervalos de confianza para un único coeficiente en regresión múltiple

Un intervalo de confianza bilateral al 95 % para el coeficiente β_j es un intervalo que contiene el verdadero valor de β_j con una probabilidad del 95 %; es decir, que contiene el verdadero valor de β_j en el 95 % de todas las posibles muestras extraídas al azar. De manera equivalente, es el conjunto de valores de β_j que no puede ser rechazado por un contraste de hipótesis bilateral al 5 %. Cuando el tamaño de la muestra es grande, el intervalo de confianza al 95 % es

$$\text{intervalo de confianza al 95 \% para } \beta_j = [\hat{\beta}_j - 1,96\text{ES}(\hat{\beta}_j), \hat{\beta}_j + 1,96\text{ES}(\hat{\beta}_j)]. \quad (7.4)$$

Un intervalo de confianza al 90 % se obtiene mediante la sustitución de 1,96 en la Ecuación (7.4) por 1,64.

Aplicación a las calificaciones y la ratio estudiantes-maestros

¿Podemos rechazar la hipótesis nula de que un cambio en la ratio estudiantes-maestros no tiene ningún efecto sobre las calificaciones obtenidas, una vez que se controla por el porcentaje de estudiantes del distrito que están aprendiendo inglés? ¿Cuál es el intervalo de confianza al 95 % para el efecto sobre las calificaciones en los exámenes de un cambio en la ratio estudiantes-maestros, controlando por el porcentaje de estudiantes aprendiendo inglés? Ahora somos capaces de averiguarlo. La regresión de las calificaciones en los exámenes sobre REM y $PctEI$, estimada por MCO, está dada en la Ecuación (6.12), y se retoma aquí con los errores estándar entre paréntesis debajo de los coeficientes:

$$\begin{array}{l} \widehat{\text{CalificaciónExamen}} = 686,0 - 1,10 \times REM - 0,650 \times PctEI. \\ \quad (8,7) \quad (0,43) \quad (0,031) \end{array} \quad (7.5)$$

Para contrastar la hipótesis de que el verdadero coeficiente de REM es 0, primero tenemos que calcular el estadístico t en la Ecuación (7.2). Debido a que la hipótesis nula dice que el verdadero valor de este coeficiente es igual a cero, el estadístico t es $t = (-1,10 - 0)/0,43 = -2,54$. El p -valor asociado es $2\Phi(-2,54) = 1,1\%$; es decir, el menor nivel de significación al que se puede rechazar la hipótesis nula es del 1,1 %. Debido a que el p -valor es menor que 5 %, la hipótesis nula puede ser rechazada al nivel de significación del 5 % (pero no al nivel de significación del 1 %).

Un intervalo de confianza al 95 % para el coeficiente poblacional de REM es $-1,10 \pm 1,96 \times 0,43 = (-1,95, -0,26)$; es decir, podemos estar seguros al 95 % de que el verdadero valor del coeficiente está entre $-1,95$ y $-0,26$. Interpretado en el contexto que resulta de interés para la directora de disminuir la ratio alumnos-maestros en 2, el intervalo de confianza al 95 % para el efecto sobre los resultados en los exámenes de esta reducción es $(-1,95 \times 2, -0,26 \times 2) = (-3,90, -0,52)$.

Adición de gastos por alumno a la ecuación. El análisis de la regresión múltiple de la Ecuación (7.5) ha convencido a la directora de que, hasta el momento y en la base a la evidencia, la reducción del tamaño de las clases mejorará los resultados en la prueba de su distrito. Ahora, sin embargo, cambia su pregunta y realiza una pregunta más matizada. Si va a contratar a más maestros, pagará más por esos maes-

tros ya sea a través de recortes en el resto del presupuesto (a costa de nuevos ordenadores, reducción de mantenimiento, etc.) o mediante un aumento del presupuesto, que no será del agrado de los contribuyentes. ¿Cuál, pregunta, es el efecto en las calificaciones de los exámenes de la reducción de la ratio estudiantes-maestros, manteniendo constantes los gastos por alumno (y el porcentaje de estudiantes aprendiendo inglés)?

Esta pregunta puede abordarse mediante la estimación de una regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros, el gasto total por alumno y el porcentaje de estudiantes aprendiendo Inglés. La recta (hiperplano) de regresión MCO es

$$\widehat{\text{CalificaciónExamen}} = 649,6 - 0,29 \times REM + 3,87 \times Gasto - 0,656 \times PctEI, \quad (7.6)$$

(15,5)	(0,48)	(1,59)	(0,032)
--------	--------	--------	---------

donde *Gasto* es el gasto total anual por alumno en el distrito en miles de dólares.

El resultado es sorprendente. Manteniendo constantes los gastos por alumno y el porcentaje de estudiantes aprendiendo inglés, se estima que un cambio en la ratio estudiantes-maestros tiene un efecto muy pequeño sobre los resultados en los exámenes: el coeficiente estimado de *REM* es $-1,10$ en la Ecuación (7.5), pero, tras añadir la variable *Gasto* como regresor en la Ecuación (7.6), es de solamente $-0,29$. Por otra parte, el estadístico *t* para contrastar que el verdadero valor del coeficiente es cero es ahora $t = (-0,29 - 0)/0,48 = 0,60$, por lo que la hipótesis de que el valor poblacional de este coeficiente es en realidad cero no puede ser rechazada incluso al nivel de significación del 10 % ($| -0,60 | < 1,645$). Por lo tanto la Ecuación (7.6) no proporciona evidencia acerca de que la contratación de más maestros mejore los resultados en los exámenes, si los gastos totales por alumno se mantienen constantes.

Una interpretación de la regresión de la Ecuación (7.6) es que, en estos datos para California, los administradores de las escuelas asignan sus presupuestos de manera eficiente. Supongamos, contrafactualmente, que el coeficiente de *REM* en la Ecuación (7.6) fuera negativo y grande. Si es así, los distritos escolares podrían elevar sus calificaciones en las pruebas, simplemente mediante la disminución de los fondos para otros fines (libros de texto, tecnología, deportes, etc.) y la transferencia de los fondos para contratar más maestros, reduciría el tamaño de las clases, mientras se mantendrían constates los gastos. Sin embargo, el coeficiente pequeño y estadísticamente no significativo de *REM* en la Ecuación (7.6) indica que esta transferencia tendría efecto escaso sobre las calificaciones obtenidas. Dicho de otra manera, los distritos ya asignan sus fondos de manera eficiente.

Tenga en cuenta que el error estándar de *REM* aumentaba cuando se añadía la variable *Gasto*, desde 0,43 en la Ecuación (7.5) hasta 0,48 en la Ecuación (7.6). Esto ilustra el hecho general, presentado en la Sección 6.7 en el contexto de multicolinealidad imperfecta, esta correlación entre los regresores (la correlación entre *REM* y *Gasto*) puede hacer que los estimadores MCO sean menos precisos.

¿Qué pasa con el contribuyente enfadado? Él afirma que los valores poblacionales *tanto* del coeficiente de la ratio alumnos-maestros (β_1) *como* del coeficiente del gasto por alumno (β_2) son iguales a cero; es decir, su hipótesis es que $\beta_1 = 0$ y $\beta_2 = 0$. Aunque pudiera parecer que se puede rechazar esta hipótesis porque el estadístico *t* para contrastar $\beta_2 = 0$ en la Ecuación (7.6) es $t = 3,87/1,59 = 2,43$, este razonamiento es erróneo. La hipótesis del contribuyente es una hipótesis conjunta, y para contrastarla es necesaria una nueva herramienta, el estadístico *F*.

7.2 Contraste de hipótesis conjuntas

En esta sección se describe cómo formular hipótesis conjuntas sobre coeficientes sobre regresión múltiple y cómo contrastarlas utilizando un estadístico *F*.

Contraste de hipótesis acerca de dos o más coeficientes

Hipótesis nula conjunta. Consideremos la regresión de la Ecuación (7.6) de las calificaciones en los exámenes sobre la ratio alumnos-maestros, el gasto por alumno y el porcentaje de estudiantes aprendiendo inglés. Nuestro contribuyente enfadado plantea la hipótesis de que ni la ratio de los estudiantes por maestro

ni los gastos por alumno tienen efecto alguno sobre las calificaciones obtenidas en los exámenes, una vez controlado por el porcentaje de alumnos que están aprendiendo inglés. Debido a que la variable *REM* es el primer regresor de la Ecuación (7.6) y la variable *Gasto*, es el segundo, podemos expresar esta hipótesis matemáticamente como

$$H_0: \beta_1 = 0 \quad \text{y} \quad \beta_2 = 0 \quad \text{vs.} \quad H_1: \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0 \quad (7.7)$$

La hipótesis de que tanto el coeficiente de la ratio estudiantes-maestros (β_1) como el coeficiente de gastos por alumno (β_2) son iguales a cero es un ejemplo de hipótesis conjunta sobre los coeficientes del modelo de regresión múltiple. En este caso, la hipótesis nula restringe el valor de dos de los coeficientes, por lo que a efectos de terminología podemos decir que la hipótesis nula de la Ecuación (7.7) impone dos **restricciones** sobre el modelo de regresión múltiple: $\beta_1 = 0$ y $\beta_2 = 0$.

En general, una **hipótesis conjunta** es una hipótesis que impone dos o más restricciones sobre los coeficientes de regresión. Se consideran la hipótesis nula conjunta y la alternativa de la forma

$$\begin{aligned} H_0: \beta_j &= \beta_{j,0}, \beta_m &= \beta_{m,0}, \dots, \text{para un total de } q \text{ restricciones, vs.} \\ H_1: &\text{una o más de las } q \text{ restricciones bajo } H_0 \text{ no se cumple,} \end{aligned} \quad (7.8)$$

donde β_j, β_m, \dots , se refieren a diferentes coeficientes de regresión $\beta_{j,0}, \beta_{m,0}, \dots$, se refieren a los valores de estos coeficientes bajo la hipótesis nula. La hipótesis nula de la Ecuación (7.7) es un ejemplo de la Ecuación (7.8). Otro ejemplo es que, en una regresión con $k = 6$ variables explicativas, la hipótesis nula es que los coeficientes de los regresores 2.^o, 4.^o, y 5.^o sean iguales a cero; es decir $\beta_2 = 0, \beta_4 = 0$, y $\beta_5 = 0$ por lo que existen $q = 3$ restricciones. En general, bajo la hipótesis nula H_0 existen q restricciones de este tipo.

Si alguna (o más de una) de las igualdades bajo la hipótesis nula H_0 en la Ecuación (7.8) es falsa, entonces la hipótesis nula conjunta en sí misma es falsa. Por tanto la hipótesis alternativa es que al menos una de las igualdades de la hipótesis nula no se cumple.

¿Por qué no se pueden contrastar los coeficientes individuales de uno en uno? A pesar de que parece que debería ser posible contrastar una hipótesis conjunta mediante el estadístico t habitual para contrastar las restricciones una por una, el siguiente cálculo demuestra que este procedimiento no es fiable. En concreto, supongamos que interesa contrastar la hipótesis nula de la Ecuación (7.6) de que $\beta_1 = 0$ y que $\beta_2 = 0$. Sea t_1 el estadístico para contrastar la hipótesis nula de que $\beta_1 = 0$ y sea el estadístico t_2 el del contraste de la hipótesis de que $\beta_2 = 0$. ¿Qué sucede cuando se utiliza el procedimiento de contraste «una a una»: ¿se rechaza la hipótesis nula conjunta, si o bien t_1 o bien t_2 son mayores que 1,96 en valor absoluto?

Debido a que esta pregunta implica a las dos variables aleatorias t_1 y t_2 , para responderla se requiere la caracterización de la distribución muestral conjunta de t_1 y t_2 . Como se ha mencionado en la Sección 6.6, en muestras grandes $\hat{\beta}_1$ y $\hat{\beta}_2$ tienen una distribución normal conjunta, por lo que bajo la hipótesis nula conjunta los estadísticos t , t_1 y t_2 presentan una distribución normal bivariante, donde cada estadístico t tiene una media igual a 0 y una varianza igual a 1.

Consideremos primero el caso particular en el que los estadísticos t no están correlacionados y por lo tanto son independientes. ¿Cuál es el tamaño del procedimiento de contraste «una a una»?; es decir, ¿cuál es la probabilidad de que se rechace la hipótesis nula cuando es verdadera? ¡Más del 5%! En este caso particular se puede calcular la probabilidad de rechazo de este método con exactitud. La hipótesis nula no se rechaza solamente si tanto $|t_1| \leq 1,96$ como $|t_2| \leq 1,96$. Debido a que los estadísticos t son independientes, $\Pr(|t_1| \leq 1,96 \text{ y } |t_2| \leq 1,96) = \Pr(|t_1| \leq 1,96) \times \Pr(|t_2| \leq 1,96) = 0,95^2 = 0,9025 = 90,25\%$. Por lo que la probabilidad de rechazar la hipótesis nula cuando es verdadera es $1 - 0,95^2 = 9,75\%$. Este método «una a una» rechaza la hipótesis nula con demasiada frecuencia, debido a que se le dan demasiadas oportunidades: si no se rechaza mediante el primer estadístico t , se intenta otra vez mediante el segundo.

Si los regresores están correlacionados, la situación es aún más complicada. El tamaño del procedimiento «una a una» depende del valor de la correlación entre los regresores. Debido a que el procedimiento de contraste «una a una» tiene el tamaño erróneo, es decir, su tasa de rechazo bajo la hipótesis nula no es igual al nivel de significación deseado, se necesita un nuevo método.

Un método consiste en modificar el procedimiento «una a una» para lo que se utilizan diferentes valores críticos que aseguran que su tamaño sea igual a su nivel de significación. Este método denominado método de Bonferroni, se describe en el Apéndice 7.1 La ventaja del método de Bonferroni es que es aplicable de forma muy general. Su desventaja es que puede tener baja potencia: con frecuencia no rechaza la hipótesis nula cuando en realidad la hipótesis alternativa es verdadera.

Afortunadamente, existe otro método para contrastar hipótesis conjuntas con mayor potencia, particularmente cuando los regresores se encuentran altamente correlacionados. Este método se basa en el estadístico F .

El estadístico F

El **estadístico F** se utiliza para contrastar hipótesis conjuntas sobre los coeficientes de regresión. Las fórmulas para el estadístico F están integradas en los paquetes informáticos de regresión modernos. Tratamos primero el caso de dos restricciones, y a continuación se trata el caso general de q restricciones.

El estadístico F con $q = 2$ restricciones. Cuando la hipótesis nula conjunta tiene las dos restricciones de que $\beta_1 = 0$ y $\beta_2 = 0$, el estadístico F combina los dos estadísticos t , t_1 y t_2 , mediante la fórmula

$$F = \frac{1}{2} \left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1, t_2} t_1 t_2}{1 - \hat{\rho}_{t_1, t_2}^2} \right) \quad (7.9)$$

donde $\hat{\rho}_{t_1, t_2}$ es un estimador de la correlación entre los dos estadísticos t .

Para entender el estadístico F de la Ecuación (7.9), supongamos primero que sabemos que los estadísticos t no están correlacionados, por lo que se pueden eliminar los términos que incluyen a $\hat{\rho}_{t_1, t_2}$. Si es así, la Ecuación (7.9) se simplifica y $F = \frac{1}{2}(t_1^2 + t_2^2)$; es decir, el estadístico F es el promedio de los cuadrados de los estadísticos t . Bajo la hipótesis nula, t_1 y t_2 son variables aleatorias normales estándar independientes (debido a que los estadísticos t no están correlacionados por hipótesis), por lo que bajo la hipótesis nula F tiene una distribución $F_{2, \infty}$ (Sección 2.4). Bajo la hipótesis alternativa de que o bien β_1 es distinto de cero o bien β_2 es distinto de cero (o ambas), entonces o bien t_1^2 , o bien t_2^2 será grande (o ambos) lo que conlleva que el contraste rechace la hipótesis nula.

En general, los estadísticos t están correlacionados, y la fórmula para el estadístico F de la Ecuación (7.9) se ajusta a esta correlación. Este ajuste se realiza para que, bajo la hipótesis nula, el estadístico F presente una distribución $F_{2, \infty}$ en muestras grandes tanto si los estadísticos t están correlacionados como si no lo están.

El estadístico F con q restricciones. La fórmula para el estadístico F heterocedástico-robusto para contrastar las q restricciones de la hipótesis nula conjunta de la Ecuación (7.8) se ofrece en la Sección 18.3. Esta fórmula está incorporada en el *software* de regresión, lo que hace que el estadístico F sea fácil de calcular en la práctica.

Bajo la hipótesis nula, el estadístico F tiene una distribución muestral que, en muestras grandes, está dada por la distribución $F_{q, \infty}$. Es decir, en muestras grandes, bajo la hipótesis nula

$$\text{el estadístico } F \text{ se distribuye } F_{q, \infty}. \quad (7.10)$$

Por tanto, los valores críticos para el estadístico F pueden obtenerse a partir las tablas de la distribución $F_{q, \infty}$ de la Tabla 4 del Apéndice para el valor apropiado de q y el nivel de significación deseado.

Cálculo del estadístico F heterocedástico-robusto mediante software estadístico. Si el estadístico F se calcula utilizando la fórmula general heterocedástico-robusta, su distribución en muestras grandes bajo la hipótesis nula es $F_{q, \infty}$ independientemente de si los errores son homocedásticos o heterocedásticos. Como se analizó en la Sección 5.4, por razones históricas, la mayor parte de los paquetes de *software* estadístico calculan por defecto los errores estándar válidos con homocedasticidad. En consecuencia, en algunos paquetes de *software* debe seleccionarse una opción de «Robusto» para que el estadístico F se calcule utilizando los errores estándar heterocedástico-robustos (y, de forma más general, una estimación de

la «matriz de covarianzas» heterocedástico-robusta). La versión del estadístico F válida con homocedasticidad se analiza al final de esta sección.

Cálculo del p -valor mediante el estadístico F . El p -valor del estadístico F puede calcularse utilizando la aproximación de su distribución para muestras grandes $F_{q,\infty}$. Sea F^{act} la expresión del valor del estadístico F calculado en realidad. Debido a que el estadístico F tiene una distribución para grandes muestras $F_{q,\infty}$ bajo la hipótesis nula, el p -valor es

$$p\text{-valor} = \Pr[F_{q,\infty} > F^{act}]. \quad (7.11)$$

El p -valor de la Ecuación (7.11) puede ser evaluado utilizando una tabla de la distribución $F_{q,\infty}$ (o, alternativamente, una tabla de la distribución χ_q^2 , porque una variable aleatoria distribuida χ_q^2 es q veces una variable aleatoria distribuida $F_{q,\infty}$). De forma alternativa, el p -valor se puede evaluar utilizando un ordenador, ya que las fórmulas para las distribuciones acumuladas chi-cuadrado y F han sido incorporadas a la mayoría de los paquetes estadísticos modernos.

El estadístico F «general». El estadístico F «general» contrasta la hipótesis conjunta de que *todos* los coeficientes de las pendientes son cero. Es decir, la hipótesis nula y la hipótesis alternativa son

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0 \text{ vs. } H_1: \beta_j \neq 0 \text{ para al menos una } j, j = 1, \dots, k. \quad (7.12)$$

Bajo esta hipótesis nula, ninguna de las variables explicativas explica nada de la variabilidad de Y_i , aunque el término independiente (que bajo la hipótesis nula es la media de Y_i) puede ser distinto de cero. La hipótesis nula de la Ecuación (7.12) es un caso particular de la hipótesis nula general de la Ecuación (7.8), y el estadístico F general de la regresión es el estadístico F calculado para la hipótesis nula de la Ecuación (7.12). En muestras grandes, el estadístico F general de la regresión presenta una distribución $F_{k,\infty}$ cuando la hipótesis nula es cierta.

El estadístico F cuando $q = 1$. Cuando $q = 1$, el estadístico F contrasta una única restricción. Entonces la hipótesis nula conjunta se reduce a la hipótesis nula sobre un solo coeficiente de regresión y el estadístico F es el estadístico t .

Aplicación a las calificaciones y la ratio estudiantes-maestros

Estamos ahora en condiciones de contrastar la hipótesis nula de que *ambos* coeficientes, el de la ratio estudiantes-maestros *además* del de los gastos por alumno son iguales a cero, frente a la alternativa de que al menos uno de los coeficientes es distinto de cero, controlando por el porcentaje de estudiantes que están aprendiendo inglés en el distrito.

Para contrastar esta hipótesis, tenemos que calcular el estadístico F heterocedástico-robusto del contraste de que $\beta_1 = 0$ y $\beta_2 = 0$ utilizando la regresión de la variable *CalificaciónExamen* sobre las variables *REM*, *Gasto*, y *PctEI* presentada en la Ecuación (7.6). Este estadístico F es 5,43. Bajo la hipótesis nula, en muestras grandes, este estadístico tiene una distribución $F_{2,\infty}$. El valor crítico del 5 % de la distribución $F_{2,\infty}$ es de 3,00 (Tabla 4 del Apéndice), y el valor crítico del 1 % es 4,61. El valor del estadístico F calculado a partir de los datos, 5,43, es mayor que 4,61, por lo que la hipótesis nula se rechaza al nivel del 1 %. Resulta muy poco probable que se hubiera extraído una muestra que hubiera dado lugar a un estadístico F con un valor tan elevado como 5,43 si la hipótesis nula fuera cierta realmente (el p -valor es 0,005). En base a la evidencia de la Ecuación (7.6) tal como se recoge en este estadístico F , se puede rechazar la hipótesis del contribuyente de que *ni* la ratio estudiantes-maestros, *ni* los gastos por alumno tienen ningún efecto sobre las calificaciones obtenidas en los exámenes (manteniendo constante el porcentaje de estudiantes que están aprendiendo inglés).

El estadístico F válido con homocedasticidad

Una manera de replantear la cuestión abordada por el estadístico F es preguntarse si relajando las q restricciones que constituyen la hipótesis nula se mejora el ajuste de la regresión lo suficiente como para

que resulte improbable que esta mejora sea simplemente el resultado de la variabilidad que proviene del muestreo aleatorio si la hipótesis nula es cierta. Esta reformulación sugiere que existe un vínculo entre el estadístico F y el R^2 de la regresión: un estadístico F elevado debería, al parecer, estar asociado con un aumento sustancial en el R^2 . De hecho, si el error u_i es homocedástico, esta intuición tiene una expresión matemática exacta. En concreto, si el término de error es homocedástico, el estadístico F puede expresarse en términos de mejora en el ajuste de la regresión, medida ya sea por la disminución de la suma de los cuadrados de los residuos o bien por el aumento del R^2 de la regresión. El estadístico F resultante se conoce como el estadístico F válido con homocedasticidad, porque solamente es válido si el término de error es homocedástico. Por el contrario, el estadístico F heterocedástico-robusto calculado mediante la fórmula de la Sección 18.3 es válido tanto si el término de error es homocedástico como si es heterocedástico. A pesar de esta limitación significativa del estadístico F válido con homocedasticidad, su sencilla fórmula arroja luz sobre lo que está haciendo el estadístico F . Además, esa sencilla fórmula puede ser calculada utilizando los resultados estándar de la regresión, tal y como podrían ser presentados en una tabla, que incluya los R^2 de las regresiones, pero no los estadísticos F .

El estadístico F válido con homocedasticidad se calcula utilizando una fórmula sencilla basada en la suma de los cuadrados de los residuos de dos regresiones. En la primera regresión, denominada **regresión restringida**, se impone el cumplimiento de la hipótesis nula. Cuando la hipótesis nula es del tipo de la Ecuación (7.8), en la que todos los valores de la hipótesis son cero, la regresión restringida es la regresión en la que estos coeficientes se hacen iguales a cero; es decir, los regresores relevantes se excluyen de la regresión. En la segunda regresión, denominada **regresión sin restringir**, la hipótesis alternativa se considera cierta. Si la suma de los cuadrados de los residuos es lo suficientemente más pequeña en la regresión sin restringir, libre, que en la regresión restringida, entonces el contraste rechaza la hipótesis nula.

El estadístico F válido con homocedasticidad está dado por la fórmula

$$F = \frac{(SR_{\text{restringida}} - SR_{\text{sin restringir}})/q}{SR_{\text{sin restringir}}/(n - k_{\text{sin restringir}} - 1)} \quad (7.13)$$

donde $SR_{\text{restringida}}$ es la suma de los cuadrados de los residuos de la regresión restringida, $SR_{\text{sin restringir}}$ es la suma de los cuadrados de los residuos de la regresión sin restringir, q es el número de restricciones bajo la hipótesis nula, y $k_{\text{sin restringir}}$ es el número de regresores de la regresión sin restringir. Una fórmula alternativa equivalente para el estadístico F válido con homocedasticidad está basada en el R^2 de las dos regresiones:

$$F = \frac{(R_{\text{sin restringir}}^2 - R_{\text{restringida}}^2)/q}{(1 - R_{\text{sin restringir}}^2)/(n - k_{\text{sin restringir}} - 1)} \quad (7.14)$$

Si los errores son homocedásticos, entonces la diferencia entre el estadístico F válido con homocedasticidad calculado utilizando la Ecuación (7.13) o (7.14) y el estadístico F heterocedástico robusto se desvanece cuando el tamaño de la muestra, n , aumenta. Por tanto, si los errores son homocedásticos, la distribución muestral del estadístico F válido con homocedasticidad bajo la hipótesis nula es $F_{q,\infty}$, para muestras grandes.

Estas fórmulas son fáciles de calcular y tienen una interpretación intuitiva en términos de bondad de ajuste a los datos de las regresiones restringida y no restringida. Desafortunadamente, las fórmulas son aplicables solamente si los errores son homocedásticos. Debido a que la homocedasticidad es un caso particular con el que no se puede contar en las aplicaciones con datos económicos, o de forma más general en las bases de datos con las que habitualmente nos encontramos en ciencias sociales, en la práctica el estadístico F válido con homocedasticidad no es un sustituto satisfactorio del estadístico F heterocedástico-robusto.

Utilización del estadístico F válido con homocedasticidad cuando n es pequeño. Si los errores son homocedásticos y se distribuyen normales i.i.d., entonces el estadístico F válido con homocedasticidad definido en las Ecuaciones (7.13) y (7.14) presenta una distribución $F_{q,n-k_{\text{sin restringir}}-1}$ bajo la hipótesis nula. Los valores críticos de esta distribución, que dependen tanto de q , como de $(n - k_{\text{sin restringir}} - 1)$, figuran en la Tabla 5 del Apéndice. Como se analizó en la Sección 2.4, la distribución $F_{q,n-k_{\text{sin restringir}}-1}$ converge

a la distribución $F_{q,\infty}$ cuando n aumenta; para muestras grandes, las diferencias entre las dos distribuciones son insignificantes. Para muestras pequeñas, sin embargo, los dos conjuntos de valores críticos son diferentes.

Aplicación a las calificaciones y ratio estudiantes-maestros. Para contrastar la hipótesis nula de que los coeficientes poblacionales de las variables *REM* y *Gasto* son iguales a 0, controlando por la variable *PctEI*, debe calcularse la *SR* (o el R^2) para las regresiones restringida y no restringida. La regresión sin restringir tiene como regresores a las variables *REM*, *Gasto*, y *PctEI*, y se ofrece en la Ecuación (7.6); su R^2 es 0,4366; es decir, $R^2_{\text{sin restringir}} = 0,4366$. La regresión restringida impone la hipótesis nula conjunta de que los verdaderos coeficientes de las variables *REM* y *Gasto* son iguales a cero; es decir, bajo la hipótesis nula, las variables *REM* y *Gasto* no entran en la regresión poblacional, aunque sí la variable *PctEI* (la hipótesis nula no restringe el coeficiente de la variable *PctEI*). La regresión restringida, que estimada por MCO, es

$$\widehat{\text{CalificaciónExamen}} = 664,7 - 0,671 \times \text{PctEI}, R^2 = 0,4149, \quad (7.15)$$

$$(1,0) \quad (0,032)$$

por lo que $R^2_{\text{restringido}} = 0,4149$. El número de restricciones es $q = 2$, el número de observaciones es $n = 420$, y el número de variables explicativas en la regresión sin restricciones es $k = 3$. El estadístico *F* válido con homocedasticidad, calculado mediante la Ecuación (7.14), es

$$F = \frac{(0,4366 - 0,4149)/2}{(1 - 0,4366)/(42 - 3 - 1)} = 8,01$$

Debido a que 8,01 es mayor que el valor crítico al 1 % de 4,61, la hipótesis se rechaza al nivel del 1 % mediante el contraste válido con homocedasticidad.

Este ejemplo ilustra las ventajas y desventajas del estadístico *F* válido con homocedasticidad. Su ventaja es que se puede calcular con una calculadora. Su desventaja es que los valores de los estadísticos *F* válido con homocedasticidad y heterocedástico-robusto pueden ser muy diferentes: el estadístico *F* heterocedástico-robusto para contrastar esta hipótesis conjunta es de 5,43, muy diferente del valor menos fiable válido con homocedasticidad de 8,01.

7.3 Contraste de una sola restricción sobre varios coeficientes

A veces, la teoría económica sugiere una única restricción que involucra a dos o más coeficientes de la regresión. Por ejemplo, la teoría podría sugerir una hipótesis nula de la forma $\beta_1 = \beta_2$; es decir, los efectos del primer y segundo regresores son los mismos. En este caso, la tarea consiste en contrastar esta hipótesis nula frente a la alternativa de que los dos coeficientes son diferentes:

$$H_0: \beta_1 = \beta_2 \text{ vs. } H_1: \beta_1 \neq \beta_2. \quad (7.16)$$

Esta hipótesis nula consta de una sola restricción, por lo que $q = 1$, pero esta restricción implica a varios coeficientes (β_1 y β_2). Es necesario modificar los métodos presentados hasta ahora para contrastar esta hipótesis. Existen dos métodos, cuál resulta más fácil depende del *software* empleado.

Método #1: Contrastar la restricción directamente. Algunos paquetes estadísticos disponen de un comando específico diseñado para contrastar las restricciones como la de la Ecuación (7.16) y el resultado es un estadístico *F* que, debido a que $q = 1$, presenta una distribución $F_{1,\infty}$ bajo la hipótesis nula. (Recordemos de la Sección 2.4 que el cuadrado de una variable aleatoria normal estándar tiene una distribución $F_{1,\infty}$, por lo que el percentil 95 % de la distribución $F_{1,\infty}$ es $1,96^2 = 3,84$).

Método #2: Transformar la regresión. Si el paquete estadístico no puede contrastar la restricción directamente, la hipótesis de la Ecuación (7.16) puede contrastarse mediante un truco de acuerdo con el cual la ecuación de regresión original se reescribe de modo que la restricción de la Ecuación (7.16) se convierta

en una restricción sobre un único coeficiente de la regresión. Para ser concretos, supongamos que solo hay dos variables explicativas en la regresión, X_{1i} y X_{2i} , por lo que la regresión poblacional tiene la forma

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i. \quad (7.17)$$

Aquí está el truco: restando y sumando $\beta_2 X_{1i}$, tenemos que $\beta_1 X_{1i} + \beta_2 X_{2i} = \beta_1 X_{1i} - \beta_2 X_{1i} + \beta_2 X_{1i} + \beta_2 X_{2i} = (\beta_1 - \beta_2)X_{1i} + \beta_2(X_{1i} + X_{2i}) = \gamma_1 X_{1i} + \beta_2 W_i$ donde $\gamma_1 = \beta_1 - \beta_2$ y $W_i = X_{1i} + X_{2i}$. Por tanto, la regresión poblacional de la Ecuación (7.17) puede reescribirse como

$$Y_i = \beta_0 + \gamma_1 X_{1i} + \beta_2 W_i + u_i. \quad (7.18)$$

Debido a que el coeficiente γ_1 en esta ecuación es $\gamma_1 = \beta_1 - \beta_2$, bajo la hipótesis nula en la Ecuación (7.16), $\gamma_1 = 0$, mientras bajo la alternativa, $\gamma_1 \neq 0$. Por lo tanto, transformando la Ecuación (7.17) en la Ecuación (7.18), hemos convertido una restricción sobre dos coeficientes de regresión en una restricción sobre un único coeficiente de regresión.

Debido a que la restricción incluye ahora a un único coeficiente γ_1 , la hipótesis nula de la Ecuación (7.16) puede contrastarse mediante el método del estadístico t de la Sección 7.1. En la práctica, esto se realiza en primer lugar construyendo el nuevo regresor W_i como la suma de los dos regresores originales, y estimando después la regresión de Y_i sobre X_{1i} y W_i . Puede calcularse ahora un intervalo de confianza al 95 % para la diferencia de los coeficientes $\beta_1 - \beta_2$ como $\hat{\gamma}_1 \pm 1,96ES(\hat{\gamma}_1)$.

Este método puede ampliarse a otras restricciones sobre las ecuaciones de regresión usando el mismo truco (véase el Ejercicio 7.9).

Los dos métodos (Métodos #1 y #2) son equivalentes, en el sentido de que el estadístico F del primer método es igual al cuadrado del estadístico t del segundo método.

Ampliación a $q > 1$. En general, es posible tener q restricciones bajo la hipótesis nula en las que algunas o todas estas restricciones implican a varios coeficientes. El estadístico F de la Sección 7.2 es extensible a este tipo de hipótesis conjuntas. El estadístico F puede calcularse por cualquiera de los dos métodos que acabamos de mencionar para $q = 1$. La mejor manera de hacer esto en la práctica depende del *software* de regresión que en concreto se utilice.

7.4 Conjuntos de confianza para varios coeficientes

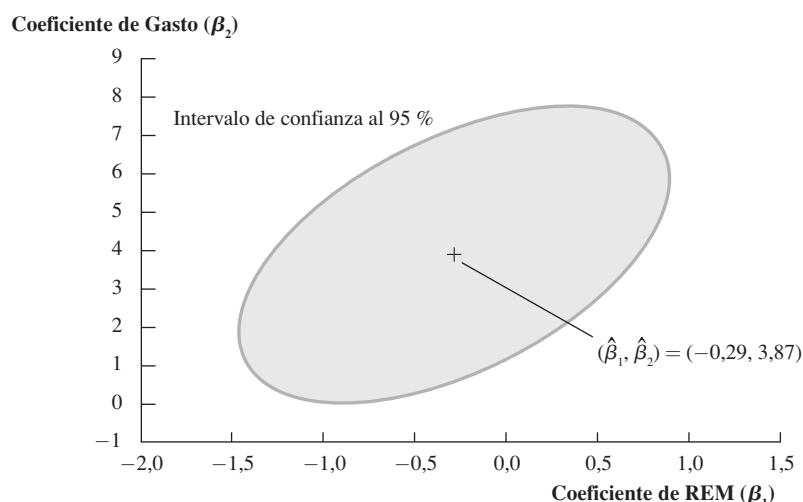
En esta sección se explica cómo construir un conjunto de confianza para dos o más coeficientes de regresión. El método es conceptualmente similar al método de la Sección 7.1 para construir un conjunto de confianza para un único coeficiente utilizando el estadístico t , excepto por que el conjunto de confianza para varios coeficientes se basa en el estadístico F .

Un **conjunto de confianza al 95 %** para dos o más coeficientes es un conjunto que contiene el verdadero valor poblacional de estos coeficientes en el 95 % de las muestras extraídas al azar. Por tanto, un conjunto de confianza es la generalización a dos o más coeficientes de un intervalo de confianza para un único coeficiente.

Recordemos que un intervalo de confianza del 95 % se calcula mediante la búsqueda de un conjunto de valores de los coeficientes que no se rechazan mediante un estadístico t con un nivel de significación del 5 %. Este método puede ser extendido al caso de varios coeficientes. Para concretar esto, supongamos que estamos interesados en construir un conjunto de confianza para dos coeficientes, β_1 y β_2 . En la Sección 7.2 se mostraba cómo utilizar el estadístico F para contrastar la hipótesis nula conjunta de que $\beta_1 = \beta_{1,0}$ y $\beta_2 = \beta_{2,0}$. Supongamos que se estuvieran contrastando todos los valores posibles de $\beta_{1,0}$ y $\beta_{2,0}$ al nivel del 5 %. Para cada par de candidatos $(\beta_{1,0}, \beta_{2,0})$, se calcula el estadístico F y se rechazan si se supera el valor crítico al 5 % de 3,00. Debido a que el contraste tiene un nivel de significación del 5 %, los verdaderos valores poblacionales de β_1 y β_2 no serán rechazados en el 95 % de todas las muestras. Por tanto, el conjunto de valores no rechazados al nivel de confianza del 5 % por este estadístico F constituye un conjunto de confianza al 95 % para β_1 y β_2 .

FIGURA 7.1 Conjunto de confianza para los coeficientes de REM y Gasto de la ecuación

El conjunto de confianza al 95 % para los coeficientes de REM (β_1) y Gasto (β_2) es una elipse. La elipse contiene los pares de valores de β_1 y β_2 que no pueden ser rechazados con el estadístico F al nivel de significación del 5 %.



Aunque este método de probar todos los valores posibles de $\beta_{1,0}$ y $\beta_{2,0}$ funciona en teoría, en la práctica resulta mucho más sencillo utilizar una fórmula explícita para el conjunto de confianza. Esta fórmula del conjunto de confianza para un número arbitrario de coeficientes se basa en la fórmula para el estadístico F . Cuando hay dos coeficientes, los conjuntos de confianza resultantes son elipses.

A modo de ejemplo, la Figura 7.1 muestra un conjunto de confianza al 95 % (elipse de confianza) para los coeficientes de la ratio estudiantes-maestros y del gasto por alumno, manteniendo constante el porcentaje de estudiantes que están aprendiendo inglés, en base a la regresión estimada en la Ecuación (7.6). Esta elipse no incluye el punto (0,0). Esto significa que la hipótesis nula de que estos dos coeficientes son simultáneamente iguales a cero se rechaza utilizando el estadístico F al nivel de significación del 5 %, lo que ya sabíamos de la Sección 7.2. La elipse de confianza es una salchicha gruesa con la parte larga de la salchicha orientada en la dirección abajo-izquierda/arriba-derecha. La razón de esta orientación es que la correlación estimada entre $\hat{\beta}_1$ y $\hat{\beta}_2$ es positiva, lo que a su vez se debe a que la correlación entre los regresores *REM* y *Gasto* es negativa (las escuelas que gastan más por alumno suelen tener menos alumnos por maestro)¹.

7.5 Especificación del modelo en regresión múltiple

La tarea de determinar qué variables incluir en regresión múltiple —es decir, el problema de elegir una especificación para la regresión— puede ser bastante difícil, y no existe ninguna regla sencilla que sea aplicable a todas las situaciones. Pero no hay que desesperar, ya que se dispone de algunas directrices útiles. El punto de partida a la hora de elegir una especificación para la regresión consiste en tener en cuenta las posibles fuentes del sesgo de variable omitida. Es importante confiar en el conocimiento experto del problema empírico y centrarse en la obtención de una estimación insesgada del efecto causal de interés; y no basarse únicamente en las medidas estadísticas de ajuste, tales como el R^2 o el \bar{R}^2 .

¹ N. del T.: Si bien la representación gráfica de las variables *REM* y *Gasto* arrojaría un diagrama de dispersión con una nube de puntos orientada arriba-izquierda/abajo-derecha (por la correlación negativa que existe entre ambas). Los parámetros β_1 y β_2 se refieren a la influencia de estas variables sobre la variable de interés (las calificaciones en los exámenes) esta influencia está positivamente correlacionada, es decir, en este caso cuando aumente la influencia de una de ellas, aumenta la influencia de la otra sobre la variable de las calificaciones. Esa es la interpretación de la correlación positiva entre β_1 y β_2 .

Sesgo de variable omitida en regresión múltiple

Los estimadores MCO de los coeficientes en regresión múltiple presentarán sesgo de variable omitida si un factor determinante de Y_i omitido está correlacionado con al menos uno de los regresores. Por ejemplo, los estudiantes de familias acomodadas suelen tener más oportunidades de aprendizaje fuera del aula (material de lectura en el hogar, viajes, visitas a museos, etc.) que los estudiantes de familias menos acomodadas, lo que podría conducir a mejores resultados en los exámenes de la prueba. Además, si el distrito es rico, entonces las escuelas tenderán a tener mayores presupuestos y menores ratios de estudiantes por maestro. Si es así, la disponibilidad de oportunidades de aprendizaje externo y la ratio estudiantes-maestros estarían correlacionadas negativamente, y la estimación MCO de los coeficientes de la ratio estudiantes-maestros podría captar el efecto de las oportunidades de aprendizaje externo, incluso después de controlar por el porcentaje de los estudiantes que están aprendiendo inglés. En resumen, la omisión de las oportunidades de aprendizaje externo (y otras variables relacionadas con las circunstancias económicas de los alumnos) podría conducir al sesgo de variable omitida en la regresión de las calificaciones en los exámenes sobre el ratio estudiantes-maestros y el porcentaje de alumnos aprendiendo inglés.

Las condiciones generales para el sesgo de variable omitida en regresión múltiple son similares a las de un solo regresor: si una variable omitida es un factor determinante de Y_i y si está relacionada con al menos una de las variables explicativas, el estimador MCO de al menos uno de los coeficientes tendrá sesgo de variable omitida. Las dos condiciones para el sesgo de variable omitida en regresión múltiple se resumen en el Concepto clave 7.3.

A nivel matemático, si se cumplen las dos condiciones para el sesgo de variable omitida, entonces al menos uno de los regresores está correlacionado con el término de error. Esto significa que la esperanza condicional de u_i dados X_{1i}, \dots, X_{ki} es distinta de cero, por lo que se viola el primer supuesto de mínimos cuadrados. Como consecuencia, el sesgo de variable omitida persiste incluso si el tamaño de la muestra es grande; es decir, la presencia de sesgo de variable omitida implica que los Estimadores MCO son inconsistentes.

CONCEPTO CLAVE

7.3

Sesgo de variable omitida en regresión múltiple

El sesgo de variable omitida es el sesgo en el estimador MCO que aparece cuando uno o más regresores incluidos están correlacionados con una variable omitida. Para que surja el sesgo de variable omitida deben cumplirse dos cosas:

1. Al menos uno de los regresores incluidos debe estar correlacionado con la variable omitida.
2. La variable omitida debe ser un factor determinante de la variable dependiente, Y .

El papel de las variables de control en regresión múltiple

Hasta ahora, hemos distinguido de forma implícita entre un regresor para el que se desea estimar un efecto causal —es decir, una variable de interés— y las variables de control. Se analiza ahora esta distinción con más detalle.

Una **variable de control** no es el objeto de interés del estudio; sino que es un regresor incluido para mantener constantes los factores que, si se descuidan, podrían llevar a que la estimación del efecto causal de interés presente sesgo de variable omitida. Los supuestos de mínimos cuadrados de la regresión múltiple (Sección 6.5) consideran los regresores simétricamente. En este apartado, se presenta una alternativa a los supuestos de mínimos cuadrados en la que la distinción entre una variable de interés y una variable de control es explícita. Si se cumple este supuesto alternativo, el estimador de MCO del efecto de interés es insesgado, pero los coeficientes MCO de las variables de control serán, en general, sesgados y no tendrán una interpretación causal.

Por ejemplo, consideremos el sesgo de variable omitida potencial derivado de la omisión de las oportunidades de aprendizaje externo en la regresión de las calificaciones en los exámenes. Aunque las «oportuni-

dades de aprendizaje externo» son un concepto amplio que resulta difícil de medir, esas oportunidades están correlacionadas con las características económicas de los estudiantes, que sí pueden medirse. Por lo tanto puede incluirse una medida de las circunstancias económicas en una regresión de las calificaciones en los exámenes para controlar por los determinantes de los resultados en los exámenes que están relacionados con la renta y que han sido omitidos de la regresión, como son las oportunidades de aprendizaje externo. Con ese fin, se amplía la regresión de las calificaciones en los exámenes sobre las variables *REM* y *PctEI* con el porcentaje de estudiantes que perciben una subvención para el comedor escolar o les resulta gratuito (*PctCom*). Debido a que los estudiantes cumplen los requisitos para entrar en este programa si su ingreso familiar es menor a un cierto umbral (aproximadamente un 150 % del umbral de pobreza), la variable *PctCom* mide la proporción de los niños económicamente desfavorecidos del distrito. La regresión estimada es

$$\widehat{\text{CalificaciónExamen}} = 700,2 - 1,00 \times \text{REM} - 0,122 \times \text{PctEI} - 0,547 \times \text{PctCom}. \quad (7.19)$$

(5,6)	(0,27)	(0,033)	(0,024)
-------	--------	---------	---------

La inclusión de la variable de control *PctCom* no cambia sustancialmente ninguna conclusión acerca del efecto del tamaño de las clases: el coeficiente de *REM* solo cambia ligeramente desde su valor de $-1,10$ en la Ecuación (7.5) hasta $-1,00$ en la Ecuación (7.19), y sigue siendo estadísticamente significativo al nivel del 1 %.

¿Qué se hace con el coeficiente de *PctCom* en la Ecuación (7.19)? Este coeficiente es muy grande: la diferencia en las calificaciones de los exámenes entre un distrito con *PctCom* = 0 % y uno con *PctCom* = 50 % se estima que es de 27,4 puntos [$= 0,547 \times (50 - 0)$], aproximadamente la diferencia entre los percentiles 75 y 25 de las calificaciones en los exámenes de la Tabla 4.1. ¿Tiene este coeficiente una interpretación causal? Supongamos que tras haber visto la Ecuación (7.19) la directora propuso eliminar el programa de comedor a precio reducido, por lo que, en su distrito, *PctCom* se reduciría inmediatamente a cero. ¿La eliminación del programa de ayudas al comedor aumentaría las calificaciones en los exámenes de su distrito? El sentido común sugiere que la respuesta es no; de hecho, dejar con hambre a algunos estudiantes, eliminando el programa de comedor a precio reducido podría tener el efecto contrario. ¿Pero tiene sentido considerar el coeficiente de la variable de interés *REM* como causal, aunque no así el coeficiente de la variable de control *PctCom*?

La distinción entre variables de interés y variables de control puede ser establecida de forma matemáticamente precisa reemplazando el primer supuesto de mínimos cuadrados del Concepto clave 6.4 —es decir, el supuesto de esperanza condicional igual a cero— por un supuesto denominado independencia de la media condicional. Consideremos una regresión con dos variables, en la cual X_{1i} es la variable de interés y X_{2i} es la variable de control. La **independencia en media condicional** requiere que la esperanza condicional de u_i dados X_{1i} y X_{2i} no dependa de (sea independiente de) X_{1i} , aunque pueda depender de X_{2i} . Es decir

$$E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i}) \quad (\text{independencia en media condicional}). \quad (7.20)$$

Como se muestra en el Apéndice 7.2, bajo el supuesto de independencia en media condicional de la Ecuación (7.20), el coeficiente de X_{1i} tiene una interpretación causal, pero no el coeficiente de X_{2i} .

La idea de la independencia en media condicional es que una vez que se controla X_{2i} , X_{1i} puede ser tratada como si estuviera asignada al azar, en el sentido de que la media condicional del término de error ya no depende de X_{1i} . La inclusión de X_{2i} como variable de control hace que X_{1i} no esté correlacionada con el término de error por lo que MCO puede estimar el efecto causal sobre Y_{1i} de un cambio en X_{1i} . La variable de control, sin embargo, sigue estando correlacionada con el término de error, por lo que el coeficiente de la variable de control está sujeto al sesgo de variable omitida y no tiene una interpretación causal.

La terminología de las variables de control puede resultar confusa. La variable de control X_{2i} se incluye debido a que tiene en cuenta (controla) los factores omitidos que afectan a Y_i y están correlacionados con X_{1i} y debido a que podría (aunque no necesariamente) tener un efecto causal por sí misma. Por tanto, el coeficiente de X_{1i} es el efecto sobre Y_i de X_{1i} , utilizando la variable de control X_{2i} tanto para mantener constante el efecto directo de X_{2i} como para controlar por los factores correlacionados con X_{2i} . Debido a que esta

terminología es complicada, lo habitual es simplemente decir que el coeficiente de X_{1i} es el efecto sobre Y_i , controlando por X_{2i} . Cuando se utiliza una variable de control, se controla tanto por su propio efecto causal directo (si existe) como por el efecto de los factores omitidos correlacionados, con el objetivo de garantizar que se cumpla la independencia en media condicional.

En el ejemplo del tamaño de las clases, la variable *PctCom* puede estar correlacionada con factores tales como las oportunidades de aprendizaje fuera de la escuela, que se encuentran en el término de error; de hecho, es *por causa* de esta correlación por lo que la variable *PctCom* es una variable de control útil. Esta correlación entre *PctCom* y el término de error significa que el coeficiente estimado para *PctCom* no tiene una interpretación causal. Lo que el supuesto de independencia en media condicional requiere es que, dadas las variables de control de la regresión (*PctEI* y *PctCom*), la media del término de error no dependa de la ratio estudiantes-maestros. Dicho de otro modo, la independencia en media condicional dice que entre las escuelas con los mismos valores de las variables *PctEI* y *PctCom*, es «como si» se asignase aleatoriamente el tamaño de las clases: la inclusión de *PctEI* y *PctCom* en la regresión controla por los factores omitidos de manera que *REM* no está correlacionada con el término de error. Si es así, el coeficiente de la ratio estudiantes-maestros tiene una interpretación causal, incluso aunque el coeficiente de *PctCom* no la tenga: para la directora que pelea por incrementar las calificaciones en las pruebas, no hay comedor gratis.

La especificación del modelo en teoría y en la práctica

En teoría, cuando se dispone de datos sobre la variable omitida, la solución para el sesgo de variable omitida es incluir la variable omitida en la regresión. En la práctica, sin embargo, la decisión de incluir una variable en particular puede ser difícil y requiere una valoración.

Nuestro sistema para el problema del sesgo potencial de variable omitida es doble. En primer lugar, debería elegirse un conjunto central o conjunto base de variables explicativas mediante una combinación de una opinión experta, la teoría económica, y el conocimiento de cómo fueron recogidos los datos; la regresión que utiliza este conjunto base de regresores se denomina a veces **especificación base**. Esta especificación base debería contener las variables de interés principal y las variables de control sugeridas por la opinión experta y la teoría económica. Sin embargo, la opinión fundamentada y la teoría económica rara vez resultan decisivas, y con frecuencia las variables sugeridas por la teoría económica no son las únicas de las que se disponen datos. Por lo tanto, el siguiente paso es desarrollar una lista de **especificaciones alternativas** candidatas, es decir, conjuntos alternativos de regresores. Si las estimaciones de los coeficientes de interés son numéricamente similares entre las especificaciones alternativas, esto proporciona evidencia de que las estimaciones de la especificación base son fiables. Si, por otro lado, las estimaciones de los coeficientes de interés varían sustancialmente entre las distintas especificaciones, esto a menudo proporciona evidencia de que la especificación original presenta sesgo de variable omitida. Proporcionamos más detalles sobre este enfoque para la especificación del modelo en la Sección 9.2 tras estudiar algunas herramientas para la especificación de las regresiones.

Interpretación del R^2 y del \bar{R}^2 ajustado en la práctica

Un R^2 o un \bar{R}^2 cercanos a 1 significa que los regresores son buenos al predecir los valores de la variable dependiente en la muestra, y un R^2 o \bar{R}^2 cercanos a 0 significa que no lo son. Esto hace que estos estadísticos sean resúmenes útiles de la capacidad predictiva de la regresión. Sin embargo, resulta fácil leer más en ellos de lo que realmente merecen.

Existen cuatro posibles dificultades para protegerse cuando se utilizan R^2 o \bar{R}^2 :

1. **Un aumento en el R^2 o en el \bar{R}^2 no significa necesariamente que una variable que se haya añadido sea estadísticamente significativa.** El R^2 aumenta cada vez que se agrega un regresor, tanto si es estadísticamente significativo como si no lo es. El \bar{R}^2 no siempre aumenta, pero si lo hace, no significa necesariamente que el coeficiente del regresor añadido sea estadísticamente significativo. Para determinar si una variable añadida es estadísticamente significativa, es necesario realizar un contraste de hipótesis utilizando el estadístico t .

2. **Un R^2 o un \bar{R}^2 elevados no significa que los regresores sean una verdadera causa de la variable dependiente.** Imaginemos la regresión de las calificaciones en los exámenes sobre el espacio de estacionamiento por alumno. El espacio de estacionamiento está correlacionado con la ratio estudiantes-maestros, con si la escuela está en las afueras o en el centro de la ciudad, y posiblemente con la renta del distrito; todas ellas están correlacionadas con los resultados en los exámenes. Por tanto, la regresión de las calificaciones en los exámenes sobre el área de estacionamiento por alumno podría tener un R^2 o un \bar{R}^2 elevados, aunque la relación no sea causal (intente decirle a la directora que la manera de aumentar las calificaciones en los exámenes consiste en aumentar el espacio de estacionamiento!).
3. **Un elevado R^2 o \bar{R}^2 no quiere decir que no haya sesgo de variable omitida.** Recordemos el análisis de la Sección 6.1, referido al sesgo de variable omitida en la regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros. El R^2 nunca apareció, porque no desempeñaba ningún papel lógico en esta discusión. El sesgo de variable omitida puede aparecer en una regresión con un R^2 bajo, moderado o elevado. De forma inversa, un R^2 bajo no implica que necesariamente exista sesgo de variable omitida.
4. **Un R^2 o un \bar{R}^2 elevados no significa necesariamente que se disponga del conjunto de variables explicativas más adecuado, ni un R^2 o un \bar{R}^2 bajos necesariamente significa que se disponga de un conjunto inadecuado de regresores.** La cuestión acerca de lo que constituye el conjunto de variables explicativas adecuado en regresión múltiple es difícil, y volveremos a ello a lo largo de este libro de texto. Las decisiones acerca de las variables explicativas deben sopesar los problemas de sesgo de variable omitida, la disponibilidad de los datos, la calidad de los datos y, sobre todo, la teoría económica y la naturaleza de las cuestiones de fondo que están siendo abordadas. Ninguna de estas preguntas puede responderse de forma simple por la obtención de un alto (o bajo) R^2 o \bar{R}^2 en la egresión.

Estos puntos se encuentran resumidos en el Concepto clave 7.4.

**CONCEPTO
CLAVE**

7.4

R^2 y \bar{R}^2 qué nos dicen y qué no

El R^2 y el \bar{R}^2 nos dicen si los regresores son buenos para predecir, o «explicar» los valores de la variable dependiente en la muestra de datos disponible. Si el R^2 (o el \bar{R}^2) está cerca de 1, entonces los regresores proporcionan buenas predicciones sobre la variable dependiente en esa muestra, en el sentido de que la varianza de los residuos MCO es pequeña comparada con la varianza de la variable dependiente. Si el R^2 (o el \bar{R}^2) está cercano a 0, es cierto todo lo contrario.

El R^2 y el \bar{R}^2 no nos dicen si:

1. Una variable incluida es estadísticamente significativa.
2. Los regresores son la verdadera causa de los movimientos de la variable dependiente.
3. Existe un sesgo de variable omitida.
4. Se ha elegido el conjunto más adecuado de regresores.

7.6 Análisis de la base de datos de las calificaciones en los exámenes

Esta sección presenta un análisis acerca del efecto sobre las calificaciones en los exámenes de la ratio estudiantes-maestros utilizando la base de datos de California. Nuestro objetivo primordial es el de proporcionar un ejemplo en el que se utilice el análisis de regresión múltiple para mitigar el sesgo de variable omitida. Nuestro segundo objetivo es demostrar cómo utilizar una tabla para resumir los resultados de la regresión.

Discusión de las especificaciones base y alternativas. Este análisis se centra en la estimación del efecto sobre los resultados en los exámenes de un cambio en la ratio estudiantes-maestros, manteniendo constantes las características de los estudiantes que la superintendente no puede controlar. Existen muchos factores que potencialmente pueden afectar al promedio de las calificaciones en los exámenes de un distrito. Algunos de estos factores están correlacionados con la ratio estudiantes-maestros, por lo que su omisión de la regresión genera un sesgo de variable omitida. Debido a que estos factores, tales como las oportunidades de aprendizaje fuera del aula, no son directamente medibles, se incluyen otras variables de control que están correlacionadas con esos factores omitidos. Si las variables de control son adecuadas en el sentido de que se cumple el supuesto de independencia en media condicional, entonces el coeficiente de la ratio estudiantes-maestros es el efecto de un cambio en la ratio estudiantes-maestros, manteniendo constantes esos otros factores.

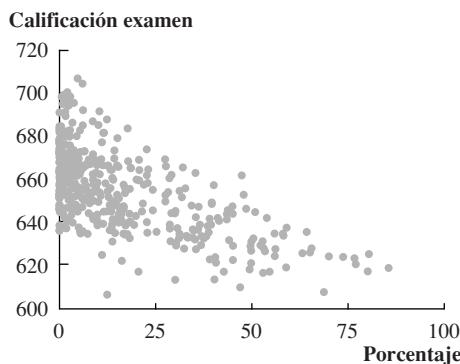
Aquí consideraremos tres variables que controlan por las características de los estudiantes que podrían afectar a las calificaciones en las pruebas: la proporción de estudiantes que están aprendiendo inglés, el porcentaje de estudiantes que cumplen los requisitos para percibir un subsidio o gratuidad en el comedor escolar, y una nueva variable, el porcentaje de estudiantes del distrito cuyas familias cumplen los requisitos para optar a un programa de asistencia económica en California. Los requisitos de selección para estar incluido en este programa de asistencia económica dependen en parte de los ingresos familiares, con un umbral más bajo (más estricto) que el programa de comedores subsidiados. Las dos últimas variables, por tanto, son diferentes medidas de la proporción de niños económicamente desfavorecidos del distrito (su coeficiente de correlación es 0,74). La teoría y la opinión experta no nos dicen cuál de estas dos variables emplear para controlar por los factores determinantes de las calificaciones en los exámenes relacionados con las características económicas. Para la especificación base utilizamos el porcentaje de cumplimiento de las condiciones para el programa de comedor subsidiado, pero consideramos asimismo una especificación alternativa que utiliza la proporción de quienes cumplen los requisitos para participar en el programa de asistencia económica.

Los diagramas de dispersión de las calificaciones en los exámenes y de estas variables se presentan en la Figura 7.2. Cada una de estas variables muestra una correlación negativa con las calificaciones en los exámenes. La correlación entre las calificaciones en los exámenes y el porcentaje de estudiantes que están aprendiendo inglés es $-0,64$; entre las calificaciones en los exámenes y el porcentaje de quienes cumplen los requisitos para participar en el programa de comedor subsidiado es $-0,87$; y entre las calificaciones en los exámenes y el porcentaje de quienes cumplen los requisitos para entrar en el programa de ayuda económica es $-0,63$.

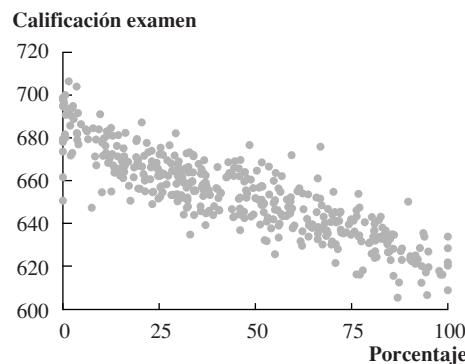
¿Qué escala debe utilizarse para los regresores? Un problema práctico que surge en el análisis de regresión es qué escala se debe utilizar para los regresores. En la Figura 7.2, las unidades de las variables son porcentajes, por lo que el máximo rango posible de los datos está entre 0 y 100. Alternativamente, podrían haberse definido estas variables como una *proporción decimal* en lugar de un porcentaje, por ejemplo, la variable *PctEI* podría reemplazarse por la proporción de estudiantes aprendiendo inglés, *FracEI* ($= PctEI/100$), que podría oscilar entre 0 y 1 en lugar de entre 0 y 100. En términos más generales, en el análisis de regresión por lo general debe tomarse alguna decisión acerca de la escala tanto de la variable dependiente como de las independientes. ¿Cómo elegir entonces la escala, o las unidades, de las variables?

La respuesta general a la cuestión de la elección de la escala de las variables es hacer que los resultados de la regresión sean fáciles de leer y de interpretar. En la aplicación a las calificaciones en los exámenes de la prueba, la unidad natural para la variable dependiente es la puntuación en la prueba en sí misma. En la regresión de la variable *CalificaciónExámenes* sobre las variables *REM* y *PctEI* presentadas en la Ecuación (7.5), el coeficiente de *PctEI* es $-0,650$. Si en cambio el regresor hubiera sido *FracEI*, la regresión habría tenido un R^2 y un *ESR* idénticos; sin embargo, el coeficiente de *FracEI* habría sido $-65,0$. En la especificación con *PctEI*, el coeficiente es la variación prevista en las calificaciones de la prueba de un aumento en 1 punto porcentual en los alumnos que están aprendiendo inglés, manteniendo *REM* constante; en la especificación con *FracEI*, el coeficiente es la variación prevista en las calificaciones de los exámenes de un aumento en 1 en la proporción de estudiantes aprendiendo inglés, es decir, un aumento de 100 puntos porcentuales, manteniendo constante *REM*. Aunque estas dos especificaciones son matemáticamente equivalentes, a los efectos de la interpretación la que contiene *PctEI* nos parece, a nosotros, más natural.

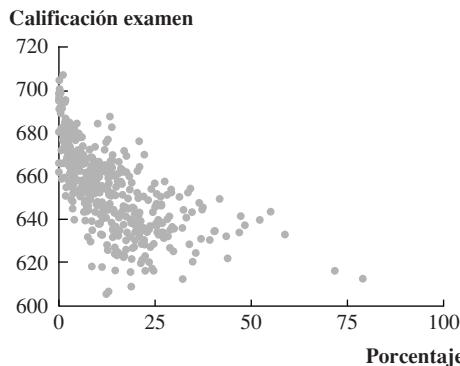
FIGURA 7.2 Diagramas de dispersión de las calificaciones en los exámenes vs. tres características de los estudiantes



(a) Porcentaje de estudiantes aprendiendo inglés



(b) Porcentaje de cumplimiento de los requisitos para comedor subvencionado



(c) Porcentaje de cumplimiento de los requisitos para el programa de ayuda económica

Los diagramas de dispersión muestran una relación negativa entre las calificaciones en los exámenes y (a) el porcentaje de estudiantes aprendiendo inglés (correlación = $-0,64$), (b) el porcentaje de estudiantes que cumplen los requisitos para el programa de comedor subvencionado (correlación = $-0,87$); y (c) el porcentaje de estudiantes que cumplen los requisitos para el programa de ayuda económica (correlación = $-0,63$).

Asimismo otro aspecto a considerar al decidir acerca de la escala es elegir las unidades de los regresores a fin de que los coeficientes de regresión resultantes resulten de fácil lectura. Por ejemplo, si un regresor se mide en dólares y tiene un coeficiente de $0,00000356$, resulta más fácil de leer si el regresor se convierte a millones de dólares y el coeficiente se presenta como $3,56$.

Presentación tabular de los resultados. Ahora nos enfrentamos con un problema de comunicación. ¿Cuál es la mejor manera de mostrar los resultados de varias regresiones múltiples que contienen diferentes subconjuntos de posibles variables explicativas? Hasta ahora, hemos presentado los resultados de regresión escribiendo las ecuaciones de regresión estimadas, como en las Ecuaciones (7.6) y (7.19). Esto funciona bien cuando solo hay unas pocas variables explicativas y solo unas pocas ecuaciones, pero con más variables explicativas y ecuaciones, este método de presentación puede resultar confuso. Un modo mejor de presentar los resultados de varias regresiones es mediante una tabla.

La Tabla 7.1 presenta el resumen de los resultados de las regresiones de las calificaciones en los exámenes sobre varios grupos de regresores. Cada una de las columnas presenta una regresión por separado. Cada regresión tiene la misma variable dependiente, las calificaciones en los exámenes. Las entradas en las cinco primeras filas son los coeficientes de regresión estimados, con sus errores estándar debajo de ellos entre paréntesis. Los asteriscos indican si los estadísticos t , que contrastan la hipótesis de que el coeficiente

TABLA 7.1 Resultados de las regresiones de las calificaciones en los exámenes sobre la ratio estudiantes-maestros y variables de control de las características de los estudiantes utilizando los distritos de escuela primaria de California

Variable dependiente: calificación en el examen media del distrito	(1)	(2)	(3)	(4)	(5)
Regresor					
Ratio estudiantes-maestros (X_1)	-2,28** (0,52)	-1,10* (0,43)	-1,00** (0,27)	-1,31** (0,34)	-1,01** (0,27)
Porcentaje de alumnos aprendiendo inglés (X_2)		-0,650** (0,031)	-0,122** (0,033)	-0,488** (0,030)	-0,130** (0,036)
Porcentaje de cumplimiento de requisitos para comedor subvencionado (X_3)			-0,547** (0,024)		-0,529** (0,038)
Porcentaje de cumplimiento de requisitos para ayudas económicas públicas (X_4)				-0,790** (0,068)	0,048 (0,059)
Intercepto	698,9** (10,4)	686,0** (8,7)	700,2** (5,6)	698,0** (6,9)	700,4** (5,5)
Estadísticos de resumen					
ESR	18,58	14,46	9,08	11,65	9,08
\bar{R}^2	0,049	0,424	0,773	0,626	0,773
n	420	420	420	420	420

Estas regresiones fueron estimadas utilizando los datos de los distritos escolares de California descritos en el Apéndice 4.1. Los errores estándar heterocedástico-robustos se presentan entre paréntesis debajo de los coeficientes. Los coeficientes individuales son estadísticamente significativos al nivel de significación del 5 %* o al nivel del 1 %** utilizando un contraste bilateral.

correspondiente es igual a cero, es significativo al nivel del 5 % (un asterisco) o al nivel del 1 % (dos asteriscos). Las tres últimas filas contienen los estadísticos de resumen para la regresión (el error estándar de la regresión, ESR, y el R^2 ajustado, \bar{R}^2) y el tamaño de la muestra (que es el mismo para todas las regresiones, 420 observaciones).

Toda la información que hemos presentado hasta ahora en formato de ecuación aparece en forma de una columna en esta tabla. Por ejemplo, consideremos la regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros, sin variables de control. En forma de ecuación, esta regresión es

$$\widehat{\text{CalificaciónExamen}} = 698,9 - 2,28 \times \text{REM}, \bar{R}^2 = 0,049, \text{ESR} = 18,58, n = 420. \quad (7.21)$$

Toda esta información aparece en la columna (1) de la Tabla 7.1. El coeficiente estimado de la ratio estudiantes-maestros (-2,28) aparece en la primera fila de entradas numéricas, y su error estándar (0,52) aparece entre paréntesis justo debajo del coeficiente estimado. El término independiente (698,9) y su error estándar (10,4) se ofrecen en la fila con la etiqueta «Intercepto». (A veces, esta fila se etiqueta como «constante» debido a que, como se trató en la Sección 6.2, el intercepto se puede interpretar como el coeficiente de un regresor que siempre es igual a 1). Del mismo modo, el \bar{R}^2 (0,049), el ESR (18,58), y el tamaño muestral n (420) aparecen en las filas finales. Las entradas en blanco en las filas de los otros regresores indican que esos regresores no están incluidos en esta regresión.

Aunque en la tabla no se muestren los estadísticos t , se pueden calcular a partir de la información proporcionada; por ejemplo, el estadístico t para el contraste de la hipótesis de que el coeficiente de la ratio estudiantes-maestros en la columna (1) es igual a cero es $-2,28/0,52 = -4,38$. Esta hipótesis se rechaza al nivel del 1 %, lo que se indica en la tabla mediante el doble asterisco al lado del coeficiente estimado.

Las regresiones que incluyen las variables de control que miden las características de los estudiantes se presentan en las columnas (2) a (5). La columna (2), que presenta la regresión de las calificaciones en los

exámenes sobre la ratio estudiantes-maestros y sobre el porcentaje de alumnos aprendiendo inglés, se estableció anteriormente, como la Ecuación (7.5).

La columna (3) presenta la especificación base, en la que los regresores son la ratio estudiantes-maestros y dos variables de control, el porcentaje de alumnos aprendiendo inglés y el porcentaje de estudiantes que cumplen los requisitos para comedor subvencionado.

Las columnas (4) y (5) presentan especificaciones alternativas que analizan el efecto de los cambios en la forma de medir la situación económica de los estudiantes. En la columna (4) se incluye como regresor el porcentaje de estudiantes dentro del programa de ayuda económica, y en la columna (5) se incluyen ambas variables sobre la situación económica.

Discusión de los resultados empíricos. Estos resultados sugieren tres conclusiones:

1. El control de estas características de los estudiantes reduce el efecto de la ratio estudiantes-maestros sobre las calificaciones en los exámenes en aproximadamente la mitad. Este efecto estimado no es muy sensible a qué variables de control específicas se incluyen en la regresión. En todos los casos el coeficiente de la ratio estudiantes-maestros sigue siendo estadísticamente significativo al nivel del 5 %. En las cuatro especificaciones con variables de control, regresiones (2) a (5), la reducción de la ratio estudiantes-maestros en un alumno por maestro se estima que aumente el promedio de las puntuaciones en las pruebas en aproximadamente 1 punto, manteniendo constantes las características de los estudiantes.
2. Las variables de las características de los estudiantes son predictores potentes de las calificaciones en los exámenes. La ratio estudiantes-maestros explica por sí misma solamente una pequeña proporción de la variabilidad de las calificaciones en los exámenes: El \bar{R}^2 en la columna (1) es 0,049. El \bar{R}^2 se eleva, sin embargo, cuando se añaden las variables de las características de los estudiantes. Por ejemplo, el \bar{R}^2 de la especificación base, la regresión (3), es 0,773. Los signos de los coeficientes de las variables demográficas de los estudiantes son acordes con los patrones vistos en la Figura 7.2: los distritos con muchos estudiantes que están aprendiendo inglés y los distritos con muchos niños pobres obtienen peores resultados en los exámenes.
3. Las variables de control no son siempre individualmente significativas: en la especificación (5), la hipótesis de que el coeficiente del porcentaje que cumple los requisitos para el programa de asistencia económica es igual a cero, no se rechaza al nivel del 5 % (el estadístico t es $-0,82$). Debido a que la adición de esta variable de control a la especificación base (3) tiene un efecto insignificante en el coeficiente estimado para la ratio estudiantes-maestros y su error estándar, y debido a que el coeficiente de esta variable de control no es significativo en la especificación (5), esta variable adicional de control es redundante, al menos a los efectos de este análisis.

7.7 Conclusión

El Capítulo 6 comenzó con una preocupación: en la regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros, las características de los estudiantes omitidas que influyen en las calificaciones en los exámenes podrían estar correlacionadas con el número de alumnos por maestro en el distrito, y, de ser así, la ratio estudiantes-maestros en el distrito podría recoger el efecto sobre las calificaciones en los exámenes de estas características omitidas de los alumnos. Por tanto el estimador MCO podría presentar sesgo de variable omitida. Para mitigar este sesgo potencial de variable omitida, se amplía la regresión mediante la inclusión de las variables que controlan por las distintas características de los estudiantes (el porcentaje de estudiantes aprendiendo inglés y dos medidas de la situación económica de los estudiantes). Al hacerlo, se reduce el efecto estimado del cambio unitario en la ratio estudiantes-maestros a la mitad, aunque sigue siendo posible rechazar la hipótesis nula de que el efecto poblacional sobre las calificaciones obtenidas en el examen, manteniendo constantes esas variables de control, es igual a cero al 5 % de nivel de significación. Debido a que eliminan el sesgo de variable omitida que surge por estas características de los estudiantes, estas estimaciones de regresión múltiple, los contrastes de hipótesis, y los intervalos de confianza

resultan mucho más útiles para asesorar a la directora que las estimaciones con un solo regresor de los Capítulos 4 y 5.

El análisis realizado en este capítulo y en el anterior ha supuesto que la función de regresión poblacional es lineal en las variables explicativas, es decir, que la esperanza condicionada de Y_i dados los regresores es una línea recta. No existe, sin embargo, ninguna razón en particular para pensar que esto sea así. De hecho, el efecto de la reducción de la ratio estudiantes-maestros podría ser muy diferente en los distritos con clases grandes que en los distritos que ya cuentan con clases pequeñas. Si es así, la función de regresión poblacional no es lineal en las X , sino que es una función no lineal de las X . Para ampliar el análisis a las funciones de regresión que son no lineales en las X , no obstante, necesitamos las herramientas que se exponen en el capítulo siguiente.

Resumen

1. Los contrastes de hipótesis y los intervalos de confianza para un único coeficiente de regresión se llevan a cabo utilizando esencialmente los mismos procedimientos que se utilizaron en el modelo de regresión lineal con una única variable del Capítulo 5. Por ejemplo, un intervalo de confianza al 95 % para β_1 viene determinado por $\hat{\beta}_1 \pm 1,96ES(\hat{\beta}_1)$.
2. Las hipótesis que incluyen más de una restricción sobre los coeficientes se denominan hipótesis conjuntas. Una hipótesis conjunta puede contrastarse mediante un estadístico F .
3. La especificación de la regresión se realiza determinando, en primer lugar, una especificación base seleccionada con el fin de afrontar la preocupación por la presencia de sesgo de variable omitida. La especificación base puede modificarse incluyendo regresores adicionales que evitan asimismo otras posibles fuentes de sesgo de variable omitida. Si simplemente se elige la especificación con el R^2 más alto, puede ocurrir que no se estime el efecto causal de interés.

Términos clave

restricciones (157)	conjunto de confianza al 95 % (162)
hipótesis conjunta (157)	variable de control (164)
estadístico F (158)	independencia en media condicional (165)
regresión restringida (160)	especificación base (166)
regresión sin restringir (160)	especificaciones alternativas (166)
estadístico F válido con homocedasticidad (160)	contraste de Bonferroni (177)

Revisión de conceptos

- 7.1 Explique cómo se contrasta la hipótesis de que $\beta_1 = 0$ en el modelo de regresión múltiple $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Explique cómo se contrastaría la hipótesis nula de que $\beta_2 = 0$. Explique cómo se contrastaría la hipótesis conjunta de que $\beta_1 = 0$ y $\beta_2 = 0$. ¿Por qué el resultado de los dos primeros contrastes no implica el resultado de este último?
- 7.2 Proponga un ejemplo de una regresión de la que podría decirse que presenta un alto valor de R^2 , pero proporciona estimadores sesgados e inconsistentes de los coeficientes de regresión. Explique por qué es probable que R^2 sea elevado. Explique por qué los estimadores MCO serían sesgados e inconsistentes.

Ejercicios

Los primeros seis ejercicios tratan sobre la tabla de regresiones estimadas de la página 174, calculadas utilizando los datos de 1998 de la base de datos CPS. La base de datos consta de información sobre 4.000 trabajadores a tiempo completo durante todo el año. El nivel educativo más elevado alcanzado por cada trabajador es o bien un diploma de escuela secundaria o bien un título universitario. El rango

de edades de los trabajadores oscila entre los 25 y los 34 años. La base de datos contiene asimismo información sobre la región del país donde la persona reside, el estado civil y el número de hijos. En estos ejercicios sean,

IMH = ingresos medios por hora (en dólares de 1998)

Universidad = variable binaria (1 si titulado en la universidad, 0 si titulado en escuela secundaria)

Femenino = variable binaria (1 si es mujer, 0 si hombre)

Edad = edad (en años)

Noreste = variable binaria (1 = si la Región = Noreste, 0 en caso contrario)

Coeste = variable binaria (1 = si la Región = Centro-Oeste, 0 en caso contrario)

Sur = variable binaria (1 = si la Región = Sur, 0 en caso contrario)

Oeste = variable binaria (1 = si la Región = Oeste, 0 en caso contrario)

7.1 Añadir «*» (5 %) y «**» (1 %) a la tabla para indicar la significación estadística de los coeficientes.

7.2 Utilizando los resultados de la regresión de la columna (1):

- a) ¿Es estadísticamente significativa la diferencia estimada por esta regresión entre los ingresos salariales de los graduados universitarios y los graduados en enseñanza secundaria al nivel del 5 %? Construya un intervalo de confianza al 95 % para esta diferencia.
- b) ¿Es estadísticamente significativa la diferencia estimada por esta regresión entre los ingresos salariales de hombres y mujeres al nivel del 5 %? Construya un intervalo de confianza al 95 % para esta diferencia.

7.3 Utilizando los resultados de la regresión de la columna (2):

- a) ¿Es la edad un factor importante de los ingresos salariales? Utilice un contraste estadístico apropiado y/o un intervalo de confianza para explicar la respuesta.
- b) Sally es una graduada universitaria de 29 años de edad. Betsy es una mujer de 34 años de edad graduada en la universidad. Construya un intervalo de confianza al 95 % para la diferencia esperada entre sus ingresos salariales.

7.4 Utilizando los resultados de la regresión de la columna (3) (véase la tabla de la página siguiente):

- a) ¿Parece que existan diferencias regionales importantes? Utilice un contraste de hipótesis adecuado para explicar su respuesta.
- b) Juanita es una mujer de 28 años graduada universitaria de la región Sur. Molly es una mujer graduada universitaria de 28 años de la región Oeste. Jennifer es una mujer graduada universitaria de la región Centro-Oeste.
 - i) Construya un intervalo de confianza al 95 % para la diferencia en los ingresos esperados de Juanita y de Molly
 - ii) Explique cómo se construiría un intervalo de confianza al 95 % para la diferencia entre los ingresos esperados entre Juanita y Jennifer (*Pista:* ¿Qué ocurriría si se incluyera la variable Oeste y se excluyera de la regresión la variable Coeste?).

7.5 La regresión mostrada en la columna (2) se estima de nuevo, esta vez utilizando los datos de 1992 (4.000 observaciones seleccionadas al azar de la CPS de marzo de 1993, convertidas a dólares de 1998 utilizando el índice de precios al consumo). Los resultados son

$$\widehat{IMH} = 0,77 + 5,29\text{Universidad} - 2,59\text{Femenino} + 0,40\text{Edad}, ESR = 5,85, \bar{R}^2 = 0,21.$$

(0,98)	(0,20)	(0,18)	(0,03)
--------	--------	--------	--------

Comparando esta regresión con la regresión para 1998 que se muestra en la columna (2), ¿hubo un cambio estadísticamente significativo en el coeficiente de la variable Universidad?

7.6 Comente la siguiente afirmación: «En todas las regresiones, el coeficiente de la variable Femenino es negativo, grande y estadísticamente significativo. Esto proporciona una potente evidencia estadística acerca de la existencia de discriminación por género en el mercado laboral de EE.UU.».

Resultados de la regresión de los ingresos medios por hora sobre el género y variables binarias educativas y otras características utilizando los datos de 1998 de la encuesta actualizada de población (CPS)			
Variable dependiente: ingresos medios por hora (IMH).			
Regresor	(1)	(2)	(3)
Universidad (X_1)	5,46 (0,21)	5,48 (0,21)	5,44 (0,21)
Femenino (X_2)	-2,64 (0,20)	-2,62 (0,20)	-2,62 (0,20)
Edad (X_3)		0,29 (0,04)	0,29 (0,04)
Noreste (X_4)			0,69 (0,30)
Oeste (X_5)			0,60 (0,28)
Sur (X_6)			-0,27 (0,26)
Intercepto	12,69 (0,14)	4,40 (1,05)	3,75 (1,06)
Estadísticos de resumen y contrastes conjuntos			
Estadístico F para los efectos regionales = 0			6,10
ESR	6,27	6,22	6,21
R^2	0,176	0,190	0,194
n	4.000	4.000	4.000

7.7 La pregunta 6.5 presentaba la siguiente regresión (a la que se han añadido los errores estándar):

$$\widehat{\text{Precio}} = 119,2 + 0,485\text{DORM} + 23,4\text{Baño} + 0,156\text{CTam} + 0,002\text{PTam} + 0,090\text{Edad} - \\ (23,9) \quad (2,61) \quad (8,94) \quad (0,011) \quad (0,00048) \quad (0,311) \\ - 48,8\text{Pobre}, \bar{R}^2 = 0,72, \text{ESR} = 41,5 \\ (10,5)$$

- a) ¿Es el coeficiente de la variable DORM estadística y significativamente distinta de cero?
 - b) Por lo general las viviendas de cinco dormitorios se venden por una cuantía mucho mayor que las viviendas de dos dormitorios. ¿Es esto compatible con la respuesta en (a) y en general con la regresión?
 - c) Un propietario de una vivienda compra un solar adyacente de 2.000 pies cuadrados. Construya un intervalo de confianza al 99 % para la variación en el valor de su casa.
 - d) La superficie del solar se mide en pies cuadrados. ¿Cree que podría resultar más apropiada otra escala? ¿Por qué o por qué no?
 - e) El estadístico F una vez omitidas las variables DORM y Edad de la regresión es $F = 0,08$. ¿Son los coeficientes de las variables DORM y Edad estadísticamente distintos de cero a un nivel del 10 %?
- 7.8 En relación con la Tabla 7.1 del texto:
- a) Construya el R^2 de cada una de las regresiones.
 - b) Plantee el estadístico F válido con homocedasticidad para contrastar $\beta_3 = \beta_4 = 0$ en la regresión mostrada en la columna (5). ¿Es el estadístico significativo al nivel del 5 %?

- c) Contraste $\beta_3 = \beta_4 = 0$ en la regresión que se muestra en la columna (5) mediante el contraste de Bonferroni que se trata en el Apéndice 7.1.
- d) Construya un intervalo de confianza al 99 % para β_1 en la regresión de la columna (5).
- 7.9** Considérese el modelo de regresión $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Utilice el Método #2 de la Sección 7.3 para transformar la regresión de modo que se pueda utilizar un estadístico t para contrastar
- $\beta_1 = \beta_2$;
 - $\beta_1 + a\beta_2 = 0$, donde a es una constante;
 - $\beta_1 + \beta_2 = 1$. (*Pista:* debe redefinirse la variable dependiente de la regresión).
- 7.10** Las Ecuaciones (7.13) y (7.14) muestran dos fórmulas para el estadístico F válido con homocedasticidad. Demuestre que las dos fórmulas son equivalentes.
- 7.11** Un distrito escolar lleva a cabo un experimento para estimar el efecto del tamaño de las clases sobre las calificaciones obtenidas en los exámenes de segundo curso. El distrito asigna al 50 % de sus estudiantes de primer curso del año anterior a clases de segundo curso pequeñas (18 estudiantes por aula) y el otro 50 % a clases de tamaño normal (21 alumnos por aula). A los estudiantes nuevos del distrito se les trata de manera diferente: el 20 % son asignados al azar a las clases pequeñas y un 80 % a las clases de tamaño normal. Al final de curso para los alumnos de segundo curso, se somete a cada estudiante a un examen estandarizado. Sea Y_i la calificación obtenida en el examen por el estudiante i -ésimo, sea X_{1i} una variable binaria que es igual a 1 si el estudiante es asignado a una clase pequeña, y X_{2i} es una variable binaria que toma el valor 1 si el estudiante es de nuevo ingreso. Sea β_1 la expresión para el efecto causal sobre las calificaciones en el examen de la reducción del tamaño de las clases desde un tamaño normal a un tamaño pequeño.
- Considérese la regresión $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$. ¿Cree usted que $E(u_i|X_{1i}) = 0$? ¿Es el estimador MCO de β_1 insesgado y consistente? Explíquelo.
 - Considérese la regresión $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. ¿Cree usted que $E(u_i|X_{1i}, X_{2i})$ depende de X_1 ? Es el estimador MCO de β_1 insesgado y consistente? Explíquelo. ¿Cree usted que $E(u_i|X_{1i}, X_{2i})$ depende de X_2 ? ¿Proporcionará el estimador MCO de β_2 una estimación insesgada y consistente del efecto causal del cambio a una nueva escuela (es decir, de ser un estudiante de nuevo ingreso)? Explíquelo.

Ejercicios empíricos

- E7.1** Utilícese la base de datos **CPS08** descrita en el Ejercicio empírico 4.1 para responder a las siguientes cuestiones.
- Realice una regresión de los ingresos medios por hora (variable *AHE* en la base de datos) sobre la variable de edad (*Age*). ¿Cuál es el intercepto estimado? ¿Cuál es la pendiente estimada?
 - Realice una regresión de la variable *AHE* sobre la variable *Age*, la variable género (*Female*), y la variable educación (*Bachelor*). ¿Cuál es el efecto estimado de la variable *Age* sobre los ingresos? Construya un intervalo de confianza al 95 % para el coeficiente de la variable *Age* en la regresión.
 - ¿Son los resultados de la regresión de (b) sustancialmente diferentes de los resultados de (a) con respecto a los efectos de la variable *Age* sobre la variable *AHE*? ¿Parece que existe en la regresión (a) sesgo de variable omitida?
 - Bob es un trabajador de sexo masculino de 26 años de edad, con un diploma de educación secundaria. Prediga los ingresos de Bob utilizando la regresión estimada en (b). Alexis es una mujer trabajadora de 30 años de edad, con título universitario. Prediga los ingresos de Alexis utilizando la regresión (b).
 - Compare el ajuste de la regresión de (a) y (b) utilizando los errores estándar de la regresión, el R^2 y el \bar{R}^2 . ¿Por qué R^2 y \bar{R}^2 son tan similares en la regresión (b)?
 - ¿Son el género y la educación factores determinantes de los ingresos? Contraste la hipótesis nula de que la variable *Female* puede eliminarse de la regresión. Contraste la hipótesis nula de que la

variable *Bachelor* puede eliminarse de la regresión. Contraste la hipótesis de que tanto la variable *Female* como la variable *Bachelor* se pueden eliminar de la regresión.

- g) Una regresión presenta sesgo de variable omitida cuando se cumplen dos condiciones. ¿Cuáles son estas dos condiciones? ¿Parece que estas condiciones se cumplan aquí?
- E7.2** Utilizando la base de datos **TeachingRatings** descrita en el Ejercicio empírico 4.2, realice los siguientes ejercicios.
- Realice una regresión de la variable *Course_Eval* sobre la variable que mide la belleza del profesor (*Beauty*). Construya un intervalo de confianza al 95 % para el efecto de la variable *Beauty* sobre la variable *Course_Eval*.
 - Considérense las distintas variables de control de la base de datos. ¿Cuáles cree que deben ser incluidas en la regresión? Utilizando una tabla como la Tabla 7.1, examine la robustez del intervalo de confianza construido en (a). ¿Cuál es un intervalo de confianza al 95 % razonable para el efecto de la variable *Beauty* sobre la variable *Course_Eval*?
- E7.3** Utilizando la base de datos **CollegeDistance** descrita en el Ejercicio empírico 4.3 conteste a las siguientes preguntas.
- Un grupo de defensa de la educación sostiene que, en promedio, los años universitarios completados por una persona se incrementarían en aproximadamente 0.15 años si la distancia a la universidad más cercana se redujera en 20 millas. Realice una regresión de los años de universidad completados (ED) sobre la distancia a la universidad más cercana (Dist). ¿Es consistente la afirmación de los grupos de defensa de la educación con la estimación de la regresión? Explíquelo.
 - Otros factores afectan asimismo a los años de formación que una persona completa. ¿Cambia el efecto estimado de la distancia sobre los años de universidad completados si se controla por estos factores? Para responder a esta pregunta, construya una tabla como la Tabla 7.1. Incluya una especificación sencilla [construida en (a)], una especificación base (que incluya un conjunto de variables de control importantes), y algunas modificaciones a la especificación base. Analice cómo el efecto estimado de la variable *Dist* sobre la variable *ED* varía entre las distintas especificaciones.
 - Se ha argumentado que, controlando por otros factores, los negros y los hispanos completan más cursos universitarios que los blancos. ¿Es este resultado congruente con las regresiones planteadas en la parte (b)?
- E7.4** Con la base de datos **Growth** descrita en el Ejercicio empírico 4.4, pero excluyendo los datos de Malta, lleve a cabo los siguientes ejercicios.
- Realice una regresión de la variable *Growth* sobre las variables *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations*, y *RGDP60*. Construya un intervalo de confianza al 95 % para el coeficiente de la variable *TradeShare*. ¿Es estadísticamente significativo el coeficiente al nivel del 5 %?
 - Compruebe si, como grupo, las variables *YearsSchool*, *Rev_Coups*, *Assessinations*, y *RGDP60* pueden ser omitidas de la regresión. ¿Cuál es el *p*-valor del estadístico *F*?

APÉNDICE

7.1 El contraste de hipótesis conjunta de Bonferroni

El método de la Sección 7.2 es la mejor manera de contrastar una hipótesis conjunta en regresión múltiple. Sin embargo, si el autor de un estudio presenta los resultados de regresión, pero no realiza la contrastación de una restricción conjunta en la que estemos interesados y tampoco se dispone de los datos originales, entonces no sería posible calcular el estadístico F de la Sección 7.2. En este apéndice se describe una forma de contrastar una hipótesis conjunta que puede utilizarse cuando solo se dispone de una tabla de resultados de regresión. Este método es un caso particular de un método de contraste muy general basado en la desigualdad de Bonferroni.

El método de Bonferroni es un contraste de hipótesis conjunta basado en el estadístico t para una hipótesis individual; es decir, el contraste de Bonferroni es el contraste del estadístico t una a una de la Sección 7.2 realizado correctamente. El **contraste de Bonferroni** para la hipótesis nula conjunta $\beta_1 = \beta_{1,0}$ y $\beta_2 = \beta_{2,0}$ basado en el valor crítico $c > 0$ utiliza la siguiente regla:

$$\begin{aligned} \text{Aceptar si } |t_1| \leq c \quad \text{y} \quad \text{si } |t_2| \leq c; \text{ rechazar en otro caso} \\ (\text{contraste del estadístico } t \text{ una a una de Bonferroni}) \end{aligned} \quad (7.22)$$

donde t_1 y t_2 son los estadísticos t para el contraste de las restricciones sobre β_1 y β_2 , respectivamente.

El truco consiste en elegir el valor crítico c de manera tal que la probabilidad de que el contraste una a una rechace cuando la hipótesis nula es cierta no sea mayor que nivel de significación deseado, digamos del 5 %. Esto se hace mediante la utilización de la desigualdad de Bonferroni para elegir el valor crítico c para permitir tanto el hecho de que se están contrastando dos restricciones como cualquier posible correlación entre t_1 y t_2 .

La desigualdad de Bonferroni

La desigualdad de Bonferroni es un resultado básico de teoría de la probabilidad. Sean A y B sucesos. Sea $A \cap B$ el suceso «tanto A como B » (la intersección de A y B), y sea $A \cup B$ el suceso « A o B o ambos» (la unión de A y B). Entonces $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$. Debido a que $\Pr(A \cap B) \geq 0$, se deduce que $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$. Esta desigualdad a su vez implica que $1 - \Pr(A \cup B) \geq 1 - [\Pr(A) + \Pr(B)]$. Sean A^c y B^c los complementarios de A y B , es decir, los sucesos «no A » y «no B ». Como el complementario de $A \cup B$ es $A^c \cap B^c$, $1 - \Pr(A \cup B) = \Pr(A^c \cap B^c)$, lo que da lugar a la desigualdad de Bonferroni, $\Pr(A^c \cap B^c) \geq 1 - [\Pr(A) + \Pr(B)]$.

Sea ahora A el suceso $|t_1| > c$ y B el suceso $|t_2| > c$. Entonces la desigualdad $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$ da lugar a

$$\Pr(|t_1| > c \text{ o } |t_2| > c, \text{ o ambas}) \leq \Pr(|t_1| > c) + \Pr(|t_2| > c) \quad (7.23)$$

Contrastes de Bonferroni

Debido a que el suceso « $|t_1| > c$ o $|t_2| > c$ o los dos» es la región de rechazo del contraste una a una, la Ecuación (7.23) conduce a un valor crítico válido para el contraste una a una. Bajo la hipótesis nula en muestras grandes, $\Pr(|t_1| > c) = \Pr(|t_2| > c) = \Pr(|Z| > c)$. Por tanto la Ecuación (7.23) implica que, en muestras grandes, la probabilidad de que el contraste una a una rechace la nula es

$$\Pr_{H_0}(\text{contraste una a una rechace}) \leq 2\Pr(|Z| > c) \quad (7.24)$$

La desigualdad de la Ecuación (7.24) proporciona una manera de elegir un valor crítico c por lo que la probabilidad de rechazo bajo la hipótesis nula iguala el nivel de significación deseado. El método de Bonferroni se puede extender a más de dos coeficientes; si existen q restricciones bajo la hipótesis nula, el factor 2 en el lado derecho de la Ecuación (7.24) se reemplaza por q .

La Tabla 7.2 presenta los valores críticos c para el contraste de Bonferroni una a una para varios niveles de significación y $q = 2, 3$, y 4 . Por ejemplo, supongamos que el nivel de significación deseado es el 5 % y $q = 2$. De acuerdo con la Tabla 7.2, el valor crítico c es 2,241. Este valor crítico es el percentil 1,25 de la distribución normal estándar, por lo que $\Pr(|Z| > 2,241) = 2,5\%$. Por tanto la Ecuación (7.24) nos dice que, en muestras grandes, el contraste una a una de la Ecuación (7.22) rechazaría al menos el 5 % de las veces bajo la hipótesis nula.

Los valores críticos de la Tabla 7.2 son mayores que los valores críticos para contrastar una única restricción. Por ejemplo, con $q = 2$, el contraste una a una rechaza si al menos un estadístico t supera 2,241 en valor absoluto. Este valor

TABLA 7.2

Valores críticos c de Bonferroni para el contraste de hipótesis conjunta del estadístico t una a una

Número de restricciones (q)	Nivel de significación		
	10 %	5 %	1 %
2	1,960	2,241	2,807
3	2,128	2,394	2,935
4	2,241	2,498	3,023

crítico es mayor que 1,96, ya está correctamente corregido por el hecho de que, mirando los dos estadísticos t , se obtiene una segunda oportunidad de rechazar la hipótesis nula conjunta, como se analizó en la Sección 7.2.

Si el estadístico t individual se basa en los errores estándar heterocedástico-robustos, entonces el contraste de Bonferroni es válido tanto si existe heterocedasticidad como si no, pero si los estadísticos t se basan en los errores estándar válidos con homocedasticidad, el contraste de Bonferroni es válido solamente bajo homocedasticidad.

Aplicación a las calificaciones en los exámenes

Los estadísticos t para el contraste de la hipótesis nula conjunta de que los verdaderos coeficientes del tamaño de las clases y los gastos por alumno de la Ecuación (7.6) son, respectivamente, $t_1 = -0,60$ y $t_2 = 2,43$. Aunque $|t_1| < 2,241$, debido a que $|t_2| > 2,241$, se puede rechazar la hipótesis nula conjunta al 5 % de nivel de significación mediante el contraste de Bonferroni. Sin embargo, tanto t_1 como t_2 son menores a 2,807 en valor absoluto, por lo que no se puede rechazar la hipótesis nula conjunta al 1 % de nivel de significación con el contraste de Bonferroni. Por el contrario, utilizando el estadístico F de la Sección 7.2, hubiese sido posible rechazar esta hipótesis al 1 % de nivel de significación.

APÉNDICE

7.2 Independencia en media condicional

En este apéndice se demuestra que, bajo el supuesto de independencia en media condicional que se introdujo en la Sección 7.5 [Ecuación (7.20)], el estimador MCO del coeficiente es insesgado para la variable de interés, pero no para la variable de control.

Consideremos una regresión con dos regresores, $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Si $E(u_i | X_{1i}, X_{2i}) = 0$, como sería el caso si X_{1i} y X_{2i} se asignaran aleatoriamente en un experimento, entonces los estimadores MCO $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores insesgados de los efectos causales β_1 y β_2 .

Ahora supongamos que X_{1i} es la variable de interés y que X_{2i} es una variable de control que está correlacionada con factores omitidos del término de error. Aunque el supuesto de media condicional igual a cero no se cumpla, supongamos que se cumple la independencia en media condicional, por lo que $E(u_i | X_{1i}, X_{2i}) = E(u_i | X_{2i})$. Por conveniencia, supongamos además que $E(u_i | X_{2i})$ es lineal en X_{2i} , por lo que $E(u_i | X_{2i}) = \gamma_0 + \gamma_2 X_{2i}$, donde γ_0 y γ_2 son constantes (este supuesto de linealidad se discute más adelante). Se define v_i de modo que sea la diferencia entre u_i y la esperanza condicional de u_i dados X_{1i} y X_{2i} —es decir, $v_i = u_i - E(u_i | X_{1i}, X_{2i})$ —de modo que v_i tiene una media condicional igual a cero: $E(v_i | X_{1i}, X_{2i}) = E[u_i - E(u_i | X_{1i}, X_{2i}) | X_{1i}, X_{2i}] = E(u_i | X_{1i}, X_{2i}) - E(u_i | X_{1i}, X_{2i}) = 0$. Por lo tanto,

$$\begin{aligned}
 Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \\
 &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + E(u_i | X_{1i}, X_{2i}) + v_i \quad (\text{utilizando la definición de } v_i) \\
 &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + E(u_i | X_{2i}) + v_i \quad (\text{utilizando la independencia en media condicional}) \\
 &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (\gamma_0 + \gamma_2 X_{2i}) + v_i \quad [\text{utilizando la linealidad de } E(u_i | X_{2i})] \\
 &= (\beta_0 + \gamma_0) + \beta_1 X_{1i} + (\beta_2 + \gamma_2) X_{2i} + v_i \quad (\text{agrupando términos}) \\
 &= \delta_0 + \beta_1 X_{1i} + \delta_2 X_{2i} + v_i,
 \end{aligned} \tag{7.25}$$

donde $\delta_0 = \beta_0 + \gamma_0$ y $\delta_2 = \beta_2 + \gamma_2$.

El error v_i de la Ecuación (7.25) presenta una media condicional igual a cero; es decir, $E(v_i|X_{1i}, X_{2i}) = 0$. Por lo tanto, el primer supuesto de mínimos cuadrados para la regresión múltiple es aplicable a la última línea de la Ecuación (7.25), y, si los otros supuestos de la regresión múltiple asimismo se cumplen, entonces la regresión MCO de Y_i sobre una constante, X_{1i} , y X_{2i} dará lugar a estimadores insesgados y consistentes de δ_0 , β_1 y δ_2 . Por tanto, el estimador MCO del coeficiente de X_{1i} es insesgado para el efecto causal β_1 . Sin embargo, el estimador MCO del coeficiente de X_{2i} no es insesgado de β_2 , y en su lugar se estima la suma del efecto causal β_2 y del coeficiente γ_2 que aparece a partir de la correlación de la variable de control X_{2i} con el término de error original u_i .

La obtención de la Ecuación (7.25) funciona para cualquier valor de β_2 , incluyendo el cero. Una variable X_{2i} es una variable de control útil si se cumple la independencia en media condicional; no necesita tener efecto causal directo sobre Y_i .

La cuarta línea de la Ecuación (7.25) utiliza el supuesto de que $E(u_i|X_{2i})$ es lineal en X_{2i} . Como se analizó en la Sección 2.4, esto será cierto si u_i y X_{2i} están distribuidas normalmente de forma conjunta. El supuesto de linealidad puede relajarse mediante los métodos que se discuten en el Capítulo 8. El Ejercicio 18.9 opera a través de los pasos de la Ecuación (7.25) para esperanzas condicionales no lineales, variables de interés múltiples, y variables de control múltiples.

En términos del ejemplo de la Sección 7.5 [la regresión de la Ecuación (7.19)], si X_{2i} es *PctCom*, entonces β_2 es el efecto causal del programa de comedor subsidiado (β_2 es positivo si los beneficios nutricionales del programa mejoran las calificaciones en los exámenes), γ_2 es negativo porque *PctCom* está correlacionada de forma negativa con (controlando por) las ventajas de aprendizaje omitidas que mejoran los resultados en los exámenes de la prueba, y $\delta_2 = \beta_2 + \gamma_2$ sería negativo si la contribución del sesgo de variable omitida a través de γ_2 sobreponderara el efecto causal positivo β_2 .

Para entender mejor el supuesto de independencia en media condicional, volvamos al concepto de experimento aleatorizado controlado ideal. Como se analizó en la Sección 4.4, si X_{1i} se asignara aleatoriamente, entonces en una regresión de Y_i sobre X_{1i} , se cumpliría el supuesto de media condicional igual a cero. Sin embargo, si X_{1i} se asignara aleatoriamente, condicionada a otra variable X_{2i} , entonces se cumpliría el supuesto de independencia en media condicional, pero si X_{2i} estuviera correlacionada con u_i , el supuesto de media condicional igual a cero no. Por ejemplo, consideremos un experimento para estudiar el efecto en las calificaciones de econometría de las tareas obligatorias frente a las voluntarias. Entre los de la especialidad en economía ($X_{2i} = 1$), al 75 % se les asigna al grupo de tratamiento (tarea obligatoria: $X_{1i} = 1$), mientras que entre los que no pertenecen a la especialidad en economía ($X_{2i} = 0$), solo el 25 % es asignado al grupo de tratamiento. Debido a que el tratamiento se asigna aleatoriamente a los de la especialidad en economía y a los que no pertenecen a la especialidad, u_i es independiente de X_{1i} , dado X_{2i} , por lo que en particular $E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i})$. Si la elección de la especialidad está relacionada con otras características (como la formación previa en matemáticas) que determinan el rendimiento en una asignatura de econometría, entonces $E(u_i|X_{2i}) \neq 0$, y la regresión de la nota en el examen final (Y_i) sobre solamente la variable X_{1i} estará sujeta a sesgo de variable omitida (X_{1i} está correlacionada con la especialidad y por lo tanto con otros determinantes de la nota omitidos). Al incluir la especialidad (X_{2i}) en la regresión se elimina este sesgo de variable omitida (el tratamiento se asigna aleatoriamente, dada la especialidad), haciendo que el estimador MCO del coeficiente de X_{1i} sea un estimador insesgado del efecto causal sobre las calificaciones en econometría de las tareas encomendadas. Sin embargo, el estimador MCO del coeficiente de la especialidad no es insesgado para el efecto causal de cambiarse a economía, ya que la especialidad no se asigna aleatoriamente y está correlacionada con otros factores omitidos que podrían no cambiar (como la formación previa en matemáticas) si un estudiante cambia la especialidad.

8 Funciones de regresión no lineales

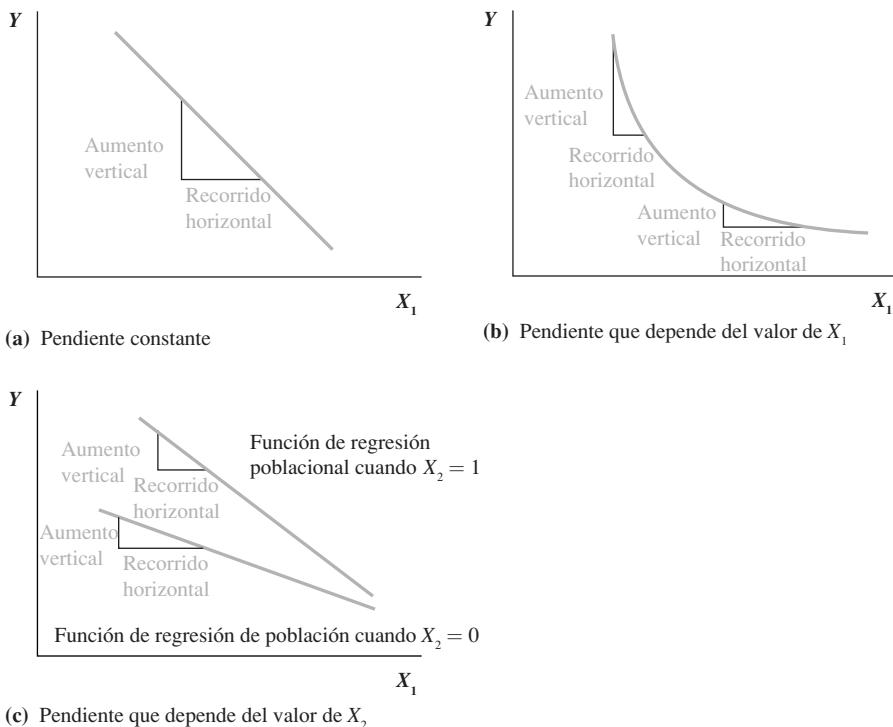
En los Capítulos 4 a 7, se suponía que la función de regresión poblacional era lineal. En otras palabras, la pendiente de la función de regresión poblacional era constante. Por lo que el efecto sobre Y de un cambio unitario en X no depende del valor de X . ¿Pero qué pasa si el efecto sobre Y de un cambio en X depende del valor de una o más de las variables independientes? Si es así, la función de regresión poblacional es no lineal.

En este capítulo se desarrollan dos grupos de métodos para la detección y modelización de funciones de regresión poblacionales no lineales. Los métodos del primer grupo son útiles cuando el efecto sobre Y de un cambio en una variable independiente X_1 , depende del valor de X_1 en sí misma. Por ejemplo, la reducción del tamaño de las clases en un estudiante por maestro podría tener un efecto mayor si las clases fueran ya aceptablemente pequeñas y manejables que si fueran tan grandes que el maestro pudiera hacer poco más que mantener a la clase controlada. Si es así, las calificaciones en los exámenes (Y) es una función no lineal de la ratio estudiantes-maestros (X_1), donde esta función es más inclinada o pronunciada cuando X_1 es pequeño. Se muestra un ejemplo de una función de regresión no lineal con esta característica en la Figura 8.1. Mientras que la función de regresión poblacional lineal de la Figura 8.1a tiene una pendiente constante, la función de regresión poblacional no lineal de la Figura 8.1b tiene una pendiente más pronunciada cuando X_1 es pequeño que cuando es grande. Este primer grupo de métodos se presenta en la Sección 8.2.

Los métodos del segundo grupo resultan útiles cuando el efecto sobre Y de un cambio en X_1 depende del valor de otra variable independiente, digamos X_2 . Por ejemplo, los estudiantes que están aprendiendo inglés podrían beneficiarse en especial de tener una atención más personalizada; si es así, el efecto sobre las calificaciones en los exámenes de la reducción de la ratio estudiantes-maestros será mayor en los distritos con muchos estudiantes que están aprendiendo inglés que en los distritos con pocos estudiantes que están aprendiendo inglés. En este ejemplo, el efecto sobre las calificaciones obtenidas (Y) de una reducción en la ratio estudiantes-maestros (X_1) depende del porcentaje de alumnos aprendiendo inglés en el distrito (X_2). Como se muestra en la Figura 8.1c, la pendiente de este tipo de función de regresión poblacional depende del valor de X_2 . Este segundo grupo de métodos se presenta en la Sección 8.3.

En los modelos de las Secciones 8.2 y 8.3, la función de regresión poblacional es una función no lineal de las variables independientes; es decir, la esperanza condicional $E(Y_i|X_{1i}, \dots, X_{ki})$ es una función no lineal de una o más de las X . A pesar de que son no lineales en las X , estos modelos son funciones lineales de los coeficientes desconocidos (o parámetros) del modelo de regresión poblacional, y por tanto son versiones del modelo de regresión múltiple de los Capítulos 6 y 7. Por tanto, los parámetros desconocidos de estas funciones de regresión no lineales pueden estimarse y contrastarse utilizando MCO y los métodos de los Capítulos 6 y 7.

En las Secciones 8.1 y 8.2 se introducen las funciones de regresión no lineales en el contexto de la regresión con una única variable independiente, y en la Sección 8.3 se amplían a dos variables independientes. Por simplicidad, se han omitido variables de control adicionales en los ejemplos empíricos de las Secciones 8.1 a 8.3. En la práctica, sin embargo, es importante analizar las funciones de regresión no lineales en los modelos que tienen en cuenta los factores omitidos incluyendo asimismo variables de control. En la Sección 8.5, se combinan las funciones de regresión no lineales y variables de control adicionales al echar un vistazo de cerca a las posibles no linealidades de la relación entre las calificaciones en los exámenes y la ratio estudiantes-maestros, manteniendo constantes las características de los estudiantes. En algunas aplicaciones la función de regresión es una función no lineal de las X y de los parámetros. Si es así, los parámetros no pueden estimarse por MCO, pero pueden estimarse por mínimos cuadrados no lineales. En el Apéndice 8.1 se proporcionan ejemplos de estas funciones y se describe el estimador de mínimos cuadrados no lineales.

FIGURA 8.1 Funciones de regresión poblacional con diferentes pendientes

En la Figura 8.1a, la función de regresión poblacional presenta una pendiente constante. En la Figura 8.1b, la pendiente de la función de regresión poblacional depende del valor de X_1 . En la Figura 8.1c, la pendiente de la función de regresión poblacional depende del valor de X_2 .

8.1 Estrategia general para la modelización de funciones de regresión no lineales

Esta sección proporciona una estrategia general para la modelización de funciones de regresión poblacionales no lineales. En esta estrategia, los modelos no lineales son generalizaciones del modelo de regresión múltiple y, por tanto, pueden ser estimados y contrastados mediante las herramientas de los Capítulos 6 y 7. En primer lugar, sin embargo, volvemos a los datos de las calificaciones en los exámenes de California y consideramos la relación entre las calificaciones en el examen y la renta del distrito.

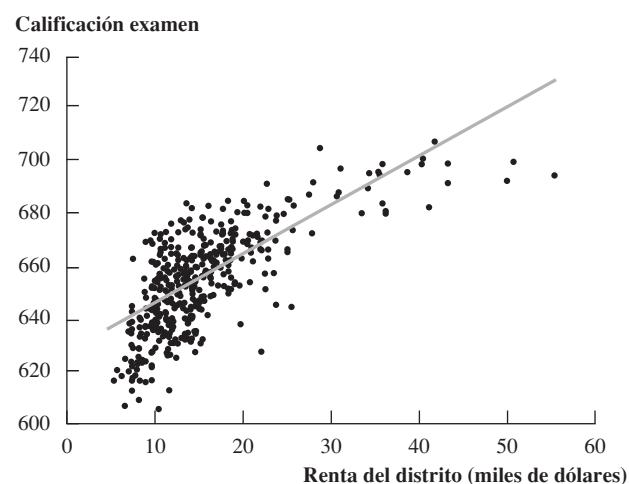
Calificaciones y renta del distrito

En el Capítulo 7, vimos que las circunstancias económicas de los estudiantes son un factor importante para explicar el desempeño en los exámenes estandarizados. Este análisis utilizaba dos variables sobre la situación económica (el porcentaje de estudiantes que cumplían los requisitos para el subsidio de comedor y el porcentaje de familias del distrito que cumplían los requisitos de inclusión en el programa de asistencia económica) para medir la proporción de estudiantes del distrito provenientes de familias pobres. Una medida diferente, más amplia de las circunstancias económicas, es la renta media anual per cápita del distrito escolar («renta del distrito»). La base de datos de California incluye la renta del distrito en miles de dólares de 1998. La muestra contiene un amplio rango de niveles de renta: para los 420 distritos de la muestra, la renta mediana del distrito es de 13,7 (es decir, 13.700 \$ por persona), y va desde 5,3 (5.300 \$ por persona) a 55,3 (55.300 \$ por persona).

FIGURA 8.2

Diagrama de dispersión de las variables calificación examen vs. renta del distrito con una función de regresión lineal

Existe una correlación positiva entre las calificaciones en la prueba estandarizada y la renta del distrito (correlación = 0,71), pero la recta de regresión lineal MCO no describe adecuadamente la relación entre estas variables.



La Figura 8.2 muestra un diagrama de dispersión de las calificaciones en los exámenes de quinto curso sobre la renta del distrito para la base de datos de California, junto con la recta de regresión MCO que relaciona estas dos variables. Los resultados obtenidos en los exámenes y la renta media están fuerte y positivamente correlacionados, con un coeficiente de correlación de 0,71; los estudiantes de los distritos ricos obtienen mejores resultados en los exámenes que los estudiantes de los distritos pobres. Pero esta dispersión presenta una peculiaridad: la mayoría de los puntos están por debajo de la recta MCO cuando la renta es o muy baja (menos de 10.000 \$) o muy alta (más de 40.000 \$), pero están por encima de la recta cuando la renta está entre 15.000 \$ y 30.000 \$. Parece que existe una cierta curvatura en la relación entre las calificaciones en los exámenes y la renta que no se capta con la regresión lineal.

En definitiva, parece que la relación entre la renta del distrito y las calificaciones en los exámenes de la prueba estandarizada no es un línea recta. Más bien es no lineal. Una función no lineal es una función con una pendiente que no es constante: la función $f(X)$ es lineal si la pendiente de $f(X)$ es la misma para todos los valores de X , pero si la pendiente depende del valor de X , entonces $f(X)$ es no lineal.

Si una línea recta no es una descripción adecuada de la relación entre la renta del distrito y las calificaciones en los exámenes, ¿cuál lo es? Imaginemos que se dibuja una curva que se ajusta a los puntos de la Figura 8.2. Esta curva aumentaría su pendiente para los valores bajos de la renta del distrito y luego se aplanaaría al aumentar la renta del distrito. Una forma de aproximar de manera matemática esta curva es la modelización de la relación como una función cuadrática. Es decir, podríamos modelizar las calificaciones en los exámenes como una función de la renta y del cuadrado de la renta.

Un modelo de regresión poblacional cuadrática que relacione las calificaciones en los exámenes y la renta puede expresarse matemáticamente como

$$\text{CalificaciónExamen}_i = \beta_0 + \beta_1 \text{Renta}_i + \beta_2 \text{Renta}_i^2 + u_i, \quad (8.1)$$

donde β_0 , β_1 , y β_2 son coeficientes, Renta_i es la renta del distrito i -ésimo, Renta_i^2 es el cuadrado de la renta del distrito i -ésimo, y u_i es un término de error que, como es habitual, representa todos los otros factores que determinan las calificaciones en los exámenes. La Ecuación (8.1) se denomina **modelo de regresión cuadrática** porque la función de regresión poblacional, $E(\text{CalificaciónExamen}_i | \text{Renta}_i) = \beta_0 + \beta_1 \text{Renta}_i + \beta_2 \text{Renta}_i^2$ es una función cuadrática de la variable independiente, Renta .

Si se conociesen los coeficientes poblacionales β_0 , β_1 y β_2 de la Ecuación (8.1), se podría predecir la calificación en el examen de un distrito en base a su renta media. Sin embargo, estos coeficientes poblacionales son desconocidos, y por lo tanto, deben estimarse a partir de una muestra de datos.

Al principio, podría parecer difícil encontrar los coeficientes de la función cuadrática que mejor se ajusten a los datos de la Figura 8.2. No obstante, si se compara la Ecuación (8.1) con el modelo de regresión

múltiple del Concepto clave 6.2, se verá que la Ecuación (8.1) es en realidad una versión del modelo de regresión múltiple con dos variables explicativas: el primer regresor es *Renta*, y el segundo regresor es *Renta*². Mecánicamente, se puede crear este segundo regresor generando una nueva variable que sea igual al cuadrado de *Renta*, por ejemplo, como una columna adicional en una hoja de cálculo. Por tanto, tras definir las variables explicativas como *Renta* y *Renta*², el modelo no lineal de la Ecuación (8.1) ¡es simplemente un modelo de regresión múltiple con dos variables explicativas!

Debido a que el modelo de regresión cuadrática es una variante de la regresión múltiple, sus coeficientes poblacionales desconocidos se pueden estimar y contrastar mediante los métodos MCO descritos en los Capítulos 6 y 7. De la estimación de los coeficientes de la Ecuación (8.1) por MCO para las 420 observaciones de la Figura 8.2 se obtiene

$$\widehat{\text{CalificaciónExamen}} = 607,3 + 3,85\text{Renta} - 0,0423\text{Renta}^2, \bar{R}^2 = 0,554, \quad (8.2)$$

(2,9) (0,27) (0,0048)

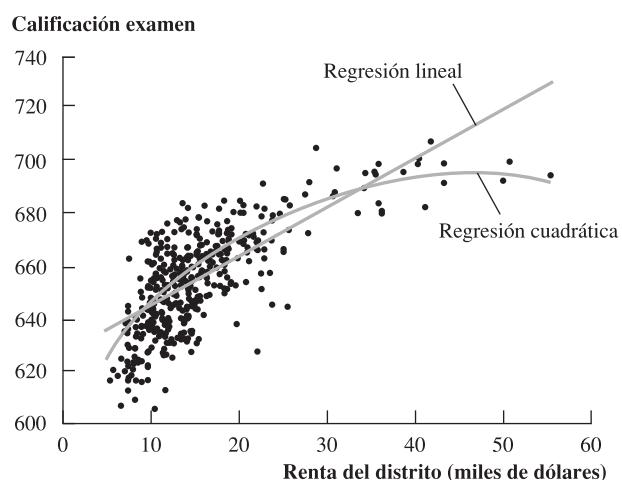
donde (como es habitual) los errores estándar de los coeficientes estimados se presentan entre paréntesis. La función de regresión estimada de la Ecuación (8.2) está representada en la Figura 8.3, superpuesta sobre el diagrama de dispersión de los datos. La función cuadrática capta la curvatura del diagrama de dispersión: es más inclinada para los valores bajos de la renta del distrito pero se aplana al aumentar la renta del distrito. En resumen, la función de regresión cuadrática parece ajustarse a los datos mejor que la lineal.

Se puede ir un paso más allá de esta comparación visual y contrastar de modo formal la hipótesis de que la relación entre la renta y las calificaciones en los exámenes es lineal frente a la alternativa de que no es lineal. Si la relación es lineal, la función de regresión está especificada correctamente mediante la Ecuación (8.1), salvo porque el regresor *Renta*² no está presente; es decir, si la relación es lineal, la Ecuación (8.1) se cumple con $\beta_2 = 0$. Por tanto, se puede contrastar la hipótesis nula de que la función de regresión poblacional es lineal frente a la alternativa de que es cuadrática contrastando la hipótesis nula de que $\beta_2 = 0$ contra la alternativa de que $\beta_2 \neq 0$.

Debido a que la Ecuación (8.1) es solo una variante del modelo de regresión múltiple, la hipótesis nula de que $\beta_2 = 0$ puede contrastarse mediante el planteamiento del estadístico *t* para esta hipótesis. Este estadístico *t* es $t = (\hat{\beta}_2 - 0)/ES(\hat{\beta}_2)$, que por la Ecuación (8.2) es $t = -0,0423/0,0048 = -8,81$. En valor absoluto, supera el valor crítico del 5 % para este contraste (que es 1,96). De hecho, el *p*-valor para el estadístico *t* es inferior a 0,01 %, por lo que puede rechazarse la hipótesis de que $\beta_2 = 0$ a todos los niveles de significación convencionales. Por lo tanto este contraste de hipótesis formal respalda la inspección informal realizada sobre las Figuras 8.2 y 8.3: el modelo cuadrático se ajusta mejor a los datos que el modelo lineal.

FIGURA 8.3 Diagrama de dispersión de las calificaciones en los exámenes sobre la renta del distrito, con función de regresión lineal y cuadrática

La función de regresión MCO cuadrática se ajusta mejor a los datos que la función de regresión MCO lineal.



El efecto sobre Y de un cambio en X con especificaciones no lineales

Dejando de lado por un momento el ejemplo de las calificaciones en los exámenes, se considera un problema general. Se quiere conocer cómo se espera que cambie la variable dependiente Y si la variable independiente X_1 cambia en la cantidad ΔX_1 , manteniendo constantes las otras variables independientes X_2, \dots, X_k . Si la función de regresión poblacional es lineal, este efecto es fácil de calcular: como se muestra en la Ecuación (6.4), el cambio esperado en Y es $\Delta Y = \beta_1 \Delta X_1$, donde β_1 es el coeficiente de la regresión poblacional que multiplica a X_1 . Sin embargo, cuando la función de regresión es no lineal, el cambio esperado en Y es más complicado de calcular porque puede depender de los valores de las variables independientes.

Fórmula general para una función de regresión poblacional no lineal¹. Los modelos de regresión poblacional no lineales considerados en este capítulo son de la forma

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}) + u_i, \quad i = 1, \dots, n, \quad (8.3)$$

donde $f(X_{1i}, X_{2i}, \dots, X_{ki})$ es la **función de regresión no lineal** poblacional, una función posiblemente no lineal de las variables independientes $X_{1i}, X_{2i}, \dots, X_{ki}$ y u_i es el término de error. Por ejemplo, en el modelo de regresión cuadrática de la Ecuación (8.1), está presente una sola variable independiente, por lo que X_1 es *Renta* y la función de regresión poblacional es $f(Renta_i) = \beta_0 + \beta_1 Renta_i + \beta_2 Renta_i^2$.

Debido a que la función de regresión poblacional es la esperanza condicional de Y_i dados $X_{1i}, X_{2i}, \dots, X_{ki}$, en la Ecuación (8.3) se permite la posibilidad de que esta esperanza condicional sea una función no lineal de $X_{1i}, X_{2i}, \dots, X_{ki}$; es decir, $E(Y_i|X_{1i}, X_{2i}, \dots, X_{ki}) = f(X_{1i}, X_{2i}, \dots, X_{ki})$, donde f puede ser una función no lineal. Si la función de regresión poblacional es lineal, entonces $f(X_{1i}, X_{2i}, \dots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$, y la Ecuación (8.3) se convierte en el modelo de regresión lineal del Concepto clave 6.2. Sin embargo, la Ecuación (8.3) permite además una función de regresión no lineal.

CONCEPTO CLAVE

8.1

El efecto esperado en Y de un cambio en X_1 en el modelo de regresión no lineal (8.3)

La variación esperada en Y , ΔY , asociada con una variación en X_1 , ΔX_1 , manteniendo constantes X_2, \dots, X_k , es la diferencia entre el valor de la función de regresión poblacional antes y después de la variación de X_1 , manteniendo constantes X_2, \dots, X_k . Es decir, la variación esperada en Y es la diferencia:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \dots, X_k) - f(X_1, X_2, \dots, X_k). \quad (8.4)$$

El estimador de esta diferencia poblacional desconocida es la diferencia entre los valores esperados para estos dos casos. Sea $\hat{f} = (X_1, X_2, \dots, X_k)$ el valor esperado de Y basado en el estimador \hat{f} de la función de regresión poblacional. Entonces la variación esperada en Y es

$$\Delta \bar{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \dots, X_k) - \hat{f}(X_1, X_2, \dots, X_k). \quad (8.5)$$

El efecto sobre Y de un cambio en X_1 . Tal y como se analizó en la Sección 6.2, el efecto sobre Y de un cambio en X_1 , ΔX_1 , manteniendo constantes X_2, \dots, X_k , es la diferencia en el valor esperado de Y cuando las variables independientes toman los valores $X_1 + \Delta X_1, X_2, \dots, X_k$ y el valor esperado de Y cuando las variables independientes adoptan los valores X_1, X_2, \dots, X_k . La diferencia entre estos dos valores esperados, ΔY , es lo que sucede en Y en promedio de la población cuando se produce un cambio en X_1 en una cuantía

¹ El término *regresión no lineal* se aplica a dos familias de modelos conceptualmente diferentes. En la primera familia, la función de regresión poblacional es una función no lineal de las X , pero es una función lineal de los parámetros desconocidos (los β). En la segunda familia, la función de regresión poblacional es una función no lineal de los parámetros desconocidos y puede, o no, ser una función no lineal de las X . Los modelos en la parte principal de este capítulo se ubican todos en la primera familia. El Apéndice 8.1 recoge los modelos de la segunda familia.

ΔX_1 , manteniendo constantes las demás variables, X_2, \dots, X_k . En el modelo de regresión no lineal de la Ecuación (8.3), este efecto sobre Y es $\Delta Y = f(X_1 + \Delta X_1, X_2, \dots, X_k) - f(X_1, X_2, \dots, X_k)$.

Debido a que la función de regresión f es desconocida, el efecto poblacional en Y de un cambio en X_1 es asimismo desconocido. Para estimar el efecto poblacional, en primer lugar se estima la función de regresión poblacional. En general, se denomina a esta función estimada como \hat{f} ; un ejemplo de función estimada de esta forma es la función de regresión cuadrática estimada en la Ecuación (8.2). El efecto estimado sobre Y (expresado mediante $\Delta \hat{Y}$) del cambio en X_1 es la diferencia entre el valor esperado o previsto de Y cuando las variables independientes toman los valores $X_1 + \Delta X_1, X_2, \dots, X_k$ y el valor esperado de Y cuando las variables independientes toman los valores X_1, X_2, \dots, X_k .

El método para calcular el efecto esperado sobre Y de un cambio en X_1 se recoge en el Concepto clave 8.1. El método del Concepto clave 8.1 siempre funciona, tanto si ΔX_1 es grande como si es pequeño, y tanto si los regresores son continuos como si son discretos. En el Apéndice 8.2 se muestra la forma de evaluar la pendiente mediante el cálculo para el caso particular de un único regresor continuo con ΔX_1 pequeño.

Aplicación a las calificaciones en los exámenes y la renta. ¿Cuál es la variación esperada en las calificaciones de los exámenes asociada a una variación de la renta del distrito de 1.000 \$, en base a la estimación de la función de regresión cuadrática de la Ecuación (8.2)? Debido a que la función de regresión es de segundo grado, este efecto depende de la renta inicial del distrito. Por lo tanto, se consideran dos casos: un aumento en la renta del distrito de 10 a 11 (es decir, desde 10.000 \$ per cápita hasta 11.000 \$) y un aumento de la renta del distrito de 40 a 41.

Para calcular $\Delta \hat{Y}$ asociada con el cambio en la renta de 10 a 11, se puede aplicar la fórmula general de la Ecuación (8.5) para el modelo de regresión cuadrática. Haciéndolo se obtiene

$$\Delta \hat{Y} = (\hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2), \quad (8.6)$$

donde $\hat{\beta}_0, \hat{\beta}_1$ y $\hat{\beta}_2$ son los estimadores MCO.

El término dentro del primer paréntesis de la Ecuación (8.6) es el valor esperado de Y cuando la variable *Renta* = 11, y el término dentro del segundo paréntesis es el valor esperado de Y cuando la variable *Renta* = 10. Estos valores esperados se calculan a partir de las estimaciones MCO de los coeficientes de la Ecuación (8.2). De acuerdo con esto, cuando la variable *Renta* = 10, el valor esperado de las calificaciones en los exámenes es $607,3 + 3,85 \times 10 - 0,0423 \times 10^2 = 641,57$. Cuando la variable *Renta* = 11, el valor esperado es $607,3 + 3,85 \times 11 - 0,0423 \times 11^2 = 644,53$. La diferencia entre estos dos valores esperados es $\Delta \hat{Y} = 644,53 - 641,57 = 2,96$ puntos; es decir, la diferencia esperada en las calificaciones de los exámenes entre un distrito con un promedio de renta de 11.000 \$ y uno con un promedio de renta de 10.000 \$ es de 2,96 puntos.

En el segundo caso, cuando la renta varía desde el nivel de los 40.000 \$ a los 41.000 \$, la diferencia en los valores esperados de la Ecuación (8.6) es $\Delta \hat{Y} = (607,3 + 3,85 \times 41 - 0,0423 \times 41^2) - (607,3 + 3,85 \times 40 - 0,0423 \times 40^2) = 694,04 - 693,62 = 0,42$ puntos. Por tanto, un cambio en la renta en una cuantía de 1.000 \$ está asociado con una mayor variación esperada en las calificaciones de los exámenes si la renta inicial es de 10.000 \$ que si es de 40.000 \$ (la variación esperada es de 2,96 puntos frente a los 0,42 puntos). Dicho de otro modo, la pendiente de la función de regresión cuadrática estimada en la Figura 8.3 es más pronunciada para los valores bajos de la variable *Renta* (como son los 10.000 \$) que para los valores más altos de la variable *Renta* (como los 40.000 \$).

Errores estándar de los efectos estimados. El estimador del efecto sobre Y de un cambio en X_1 depende del estimador de la función de regresión poblacional, \hat{f} , que varía de una muestra a otra. Por tanto, el efecto estimado contiene un error de muestreo. Una forma de cuantificar la incertidumbre en el muestreo asociada al efecto estimado es calcular un intervalo de confianza para el verdadero efecto poblacional. Para hacerlo, es necesario calcular el error estándar de $\Delta \hat{Y}$ en la Ecuación (8.5).

Resulta fácil calcular un error estándar para $\Delta \hat{Y}$ cuando la función de regresión es lineal. El efecto estimado de un cambio en X_1 es $\hat{\beta}_1 \Delta X_1$, por lo que un intervalo de confianza al 95 % para la variación estimada es $\hat{\beta}_1 \Delta X_1 \pm 1,96 \text{ ES}(\hat{\beta}_1) \Delta X_1$.

En los modelos de regresión no lineales de este capítulo, el error estándar de $\Delta \hat{Y}$ puede calcularse utilizando las herramientas introducidas en la Sección 7.3 para el contraste de una sola restricción que incluye varios coeficientes. Para ilustrar este método, consideremos la variación esperada en las calificaciones de los exámenes asociada a un cambio en la renta de 10 a 11 en la Ecuación (8.6), que es $\Delta \hat{Y} = \hat{\beta}_1 \times (11 - 10) + \hat{\beta}_2 \times (11^2 - 10^2) = \hat{\beta}_1 + 21\hat{\beta}_2$. Por tanto, el error estándar de la variación estimada es

$$ES(\Delta \hat{Y}) = ES(\hat{\beta}_1 + 21\hat{\beta}_2). \quad (8.7)$$

Por lo tanto, si se puede calcular el error estándar de $\hat{\beta}_1 + 21\hat{\beta}_2$ entonces se habrá calculado el error estándar de $\Delta \hat{Y}$. Existen dos métodos para hacerlo utilizando el *software* de regresión habitual, que se corresponden con los dos métodos de la Sección 7.3 para contrastar una restricción única con varios coeficientes.

El primer método es el de utilizar el «Método #1» de la Sección 7.3, que consiste en calcular el estadístico F para contrastar la hipótesis de que $\beta_1 + 21\beta_2 = 0$. El error estándar de $\Delta \hat{Y}$ está dado entonces por²

$$ES(\Delta \hat{Y}) = \frac{|\Delta \bar{Y}|}{\sqrt{F}}. \quad (8.8)$$

Cuando se aplica a la regresión cuadrática de la Ecuación (8.2), el estadístico F para el contraste de la hipótesis de que $\beta_1 + 21\beta_2 = 0$ es $F = 299,94$. Debido a que, $\Delta \hat{Y} = 2,96$, aplicando la Ecuación (8.8) se obtiene que $ES(\Delta \hat{Y}) = 2,96/\sqrt{299,94} = 0,17$. Por tanto, un intervalo de confianza al 95 % para la variación en el valor esperado de Y es $2,96 \pm 1,96 \times 0,17$, o bien $(2,63, 3,29)$.

El segundo método consiste en utilizar el «Método #2» de la Sección 7.3, lo que implica la transformación de las variables explicativas de modo que, en la regresión transformada, uno de los coeficientes sea $\beta_1 + 21\beta_2$. Esta transformación se deja como ejercicio (Ejercicio 8.9).

Comentario sobre la interpretación de los coeficientes en especificaciones no lineales. En el modelo de regresión múltiple de los Capítulos 6 y 7, los coeficientes de regresión tenían una interpretación natural. Por ejemplo, β_1 es la variación esperada en Y asociada con un cambio en X_1 , manteniendo los otros regresores constantes. Pero, como se ha visto, este no suele ser el caso habitual en un modelo no lineal. Es decir, no resulta muy útil pensar en el β_1 de la Ecuación (8.1) como el efecto de la variación en la renta del distrito, manteniendo constante el cuadrado de la renta del distrito. En los modelos no lineales la función de regresión se interpreta mejor mediante su representación gráfica y mediante el cálculo del efecto esperado sobre Y de la variación de una o más variables independientes.

Metodología general para la modelización no lineal mediante regresión múltiple

El método general para modelizar las funciones de regresión no lineales adoptado en este capítulo consta de cinco elementos:

1. **Identificación de una posible relación no lineal.** Lo mejor es hacer uso de la teoría económica y de lo que se conoce acerca de lo que el caso concreto trabajado sugiere acerca de una posible relación no lineal. Incluso antes de mirar los datos, preguntarse si la pendiente de la función de regresión que relaciona Y y X , pudiera razonablemente depender del valor de X o de otra variable independiente. ¿Por qué podrían existir esas relaciones de dependencia no lineales? ¿Qué formas no lineales sugieren? Por ejemplo, pensando en la dinámica de un aula con alumnos de 11 años de edad, puede parecer que la reducción del tamaño de la clase de 18 estudiantes a 17 podría tener un mayor efecto que la reducción de 30 a 29.

² La Ecuación (8.8) se obtiene teniendo en cuenta que el estadístico F es el cuadrado del estadístico t para el contraste de esta hipótesis, es decir, $F = t^2 = [(\hat{\beta}_1 + 21\hat{\beta}_2)/ES(\hat{\beta}_1 + 21\hat{\beta}_2)]^2 = [\Delta \hat{Y}/ES(\Delta \hat{Y})]^2$, y resolviendo para $ES(\Delta \hat{Y})$.

2. **Especificación de una función no lineal y estimación de sus parámetros por MCO.** Las Secciones 8.2 y 8.3 contienen varias funciones de regresión no lineales que puede ser estimadas por MCO. Tras haber trabajado estas secciones, se comprenderán las características de cada una de estas funciones.
3. **Determinación de si el modelo no lineal mejora el modelo lineal.** ¡No solo por pensar que una función de regresión puede ser no lineal, significa que realmente lo sea! Debe determinarse empíricamente si el modelo no lineal elegido es el adecuado. La mayoría de las veces pueden utilizarse los estadísticos t y los estadísticos F para contrastar la hipótesis nula de que la función de regresión poblacional es lineal frente a la alternativa de que es no lineal.
4. **Representación de la función de regresión no lineal estimada.** ¿Describe bien los datos la función de regresión estimada? Un vistazo a las Figuras 8.2 y 8.3 sugiere que el modelo cuadrático se ajusta mejor a los datos que el modelo lineal.
5. **Estimación del efecto sobre Y de un cambio en X .** El último paso consiste en utilizar la regresión estimada para calcular el efecto sobre Y de una variación en uno o más regresores X utilizando el método del Concepto clave 8.1.

8.2 Funciones no lineales de una sola variable independiente

En esta sección se presentan dos métodos para la modelización de una función de regresión no lineal. Por simplicidad, estos métodos se desarrollan para una función de regresión no lineal que incluye una sola variable independiente, X . No obstante, tal y como se vio en la Sección 8.5, estos modelos pueden ser modificados de forma que incluyan varias variables independientes.

El primer método presentado en esta sección es la regresión polinomial, una extensión de la regresión cuadrática utilizada en la última sección para modelizar la relación entre las calificaciones en los exámenes y la renta. El segundo método utiliza los logaritmos de X , de Y , o de ambos. A pesar de que estos métodos se presentan por separado, pueden utilizarse combinados.

El Apéndice 8.2 proporciona un tratamiento de los modelos de esta sección basado en el cálculo.

Polinomios

Una forma de especificar una función de regresión no lineal es mediante un polinomio en X . En general, sea r la mayor potencia de X incluida en la regresión. **El modelo de regresión polinomial de grado r** es

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i. \quad (8.9)$$

Cuando $r = 2$, la Ecuación (8.9) es el modelo de regresión cuadrática analizado en la Sección 8.1. Si $r = 3$ entonces la mayor potencia de X incluida es X^3 , a la Ecuación (8.9) se le denomina **modelo de regresión cúbica**.

El modelo de regresión polinomial es similar al modelo de regresión múltiple del Capítulo 6, excepto que en el Capítulo 6 los regresores eran variables independientes distintas, mientras que aquí los regresores son potencias de la misma variable independiente, X ; es decir, las variables explicativas son X , X^2 , X^3 , etc. De esta manera las técnicas para la estimación e inferencia desarrolladas para la regresión múltiple pueden aplicarse aquí. En particular, los coeficientes desconocidos β_0 , β_1 , ..., β_r de la Ecuación (8.9) pueden estimarse mediante regresión MCO de Y_i sobre X_i , X_i^2 , ..., X_i^r .

Contraste de la hipótesis nula de que la función de regresión poblacional es lineal. Si la función de regresión poblacional es lineal, entonces el término cuadrático y los de grado más alto no entran en la función de regresión poblacional. De acuerdo con esto, la hipótesis nula (H_0) de que la regresión es lineal y la alternativa (H_1) de que es un polinomio de grado r se corresponden con

$$H_0: \beta_2 = 0, \beta_3 = 0, \dots, \beta_r = 0 \text{ vs. } H_1: \text{al menos una } \beta_j \neq 0, j = 2, \dots, r. \quad (8.10)$$

La hipótesis nula de que la función de regresión poblacional es lineal se puede contrastar frente a la alternativa de que es un polinomio de grado r contrastando H_0 frente H_1 en la Ecuación (8.10). Debido a que H_0 es una hipótesis nula conjunta con $q = r - 1$ restricciones sobre los coeficientes del modelo de regresión poblacional polinomial, se puede contrastar mediante el estadístico F tal y como se describía en la Sección 7.2.

¿Qué grado de polinomio debería utilizarse? Es decir, ¿cuántas potencias de X deberían ser incluidas en una regresión polinómica? La respuesta es un equilibrio entre flexibilidad y precisión estadística. El aumento en el grado r introduce una mayor flexibilidad en la función de regresión y permite un ajuste hacia más formas; un polinomio de grado r puede tener un máximo de $r - 1$ curvas (es decir, puntos de inflexión) en su gráfica. Pero el aumento de r supone agregar más variables explicativas, lo que puede reducir la precisión de los coeficientes estimados.

Por lo tanto, la respuesta a la pregunta de cuántos términos deberían incluirse es que deben incluirse los suficientes como para poder modelizar la función de regresión no lineal de manera adecuada, pero no más. Desgraciadamente, esta respuesta no es muy útil en la práctica!

Un método práctico para determinar el grado del polinomio consiste en plantear si los coeficientes de la Ecuación (8.9) asociados con los mayores valores de r son iguales a cero. Si es así, entonces estos términos se pueden quitar de la regresión. Este procedimiento, se denomina contraste de hipótesis secuencial porque las hipótesis individuales se contrastan secuencialmente, tal y como queda resumido en los siguientes pasos:

1. Elegir un valor máximo para r y calcular el polinomio de regresión para este r .
2. Utilizar el estadístico t para contrastar la hipótesis de que el coeficiente de X^r [β_r en la Ecuación (8.9)] es igual a cero. Si se rechaza esta hipótesis, entonces X^r pertenece a la regresión, por lo que se utiliza el polinomio de grado r .
3. Si no se rechaza $\beta_r = 0$ en el paso 2, eliminar X^r de la regresión y estimar una regresión polinomial de grado $r - 1$. Contrastar si el coeficiente de X^{r-1} es cero. Si se rechaza, utilizar el polinomio de grado $r - 1$.
4. Si no se rechaza $\beta_{r-1} = 0$ en el paso 3, continuar con este procedimiento hasta que el coeficiente de la mayor potencia del polinomio sea estadísticamente significativo.

A esta receta le falta un ingrediente: el grado inicial r del polinomio. En muchas aplicaciones con datos económicos, las funciones no lineales son suaves, es decir, no presentan saltos bruscos o «picos». Si es así, entonces resulta apropiado seleccionar un máximo grado para el polinomio que sea pequeño, como 2, 3 o 4, es decir, comenzando con $r = 2$ o 3 o 4 en el paso 1.

Aplicación a la renta del distrito y las calificaciones en los exámenes. La función de regresión cúbica estimada que relaciona la renta del distrito con las calificaciones es

$$\widehat{\text{CalificaciónExamen}} = 600,1 + 5,02\text{Renta} - 0,096\text{Renta}^2 + 0,00069\text{Renta}^3, \bar{R}^2 = 0,555. \quad (8.11)$$

(5,1)	(0,71)	(0,029)	(0,00035)
-------	--------	---------	-----------

El estadístico t de la variable Renta^3 es 1,97, por lo que la hipótesis nula de que la función de regresión es cuadrática se rechaza frente a la alternativa de que es cúbica al nivel del 5 %. Por otra parte, el estadístico F para contrastar la hipótesis nula conjunta de que los coeficientes de las variables Renta^2 y Renta^3 son ambos iguales a cero es 37,7, con un p -valor menor que 0,01 %, por lo que la hipótesis nula de que la función de regresión es lineal se rechaza frente a la alternativa de que es o bien cuadrática o bien cúbica.

Interpretación de los coeficientes en los modelos de regresión polinomial. Los coeficientes de las regresiones polinomiales no tienen interpretación sencilla. La mejor forma de interpretar las regresiones polinomiales es mediante la representación de la función de regresión estimada y calculando el efecto estimado sobre Y asociado a una variación de X para uno o más valores de X .

Logaritmos

Otra forma de especificar una función de regresión no lineal es mediante los logaritmos naturales de Y y/o X . Los logaritmos convierten las variaciones en las variables en cambios porcentuales, y muchas relaciones tienen una expresión natural en términos de porcentajes. He aquí algunos ejemplos:

- Un recuadro del Capítulo 3, «La brecha de género en los ingresos salariales de los titulados universitarios en Estados Unidos», examinaba la brecha salarial entre hombres y mujeres graduados universitarios. En ese debate, la brecha salarial se medía en términos de dólares. Sin embargo, es más fácil comparar las diferencias salariales entre profesiones y en el tiempo cuando estas se expresan en términos porcentuales.
- En la Sección 8.1 se encontró que la renta del distrito y las calificaciones en los exámenes estaban relacionadas de forma no lineal. ¿Sería lineal esta relación en términos de cambios porcentuales? Es decir, ¿podría ser que una variación en la renta del distrito de un 1 % —en lugar de en una cuantía de 1.000 \$— esté asociada con una variación en las calificaciones de los exámenes que sea aproximadamente constante para los diferentes valores de renta?
- En el análisis económico de la demanda del consumidor, a menudo se supone que un aumento en el precio de un 1 % conduce a una cierta disminución en el *porcentaje* de la cantidad demandada. La disminución porcentual de la demanda que resulta de un aumento en el precio de un 1 % se denomina **elasticidad**-precio.

Las especificaciones de regresión que utilizan logaritmos naturales permiten que los modelos de regresión puedan estimar las relaciones porcentuales de este tipo. Antes de presentar las especificaciones, se revisan las funciones exponenciales y los logaritmos naturales.

La función exponencial y el logaritmo natural. La función exponencial y su inversa, el logaritmo natural, desempeñan un papel importante en la modelización de las funciones de regresión no lineales. La **función exponencial** de x es e^x (es decir, e elevado a la potencia x), donde e es la constante 2,71828...; la función exponencial se expresa asimismo como $\exp(x)$. El **logaritmo natural** es la inversa de la función exponencial; es decir, el logaritmo natural es la función para la cual $x = \ln(e^y)$ o, equivalentemente, $y = \ln[\exp(x)]$. La base del logaritmo natural es e . Aunque existen logaritmos en otras bases, tales como la base 10, en este libro solo se consideran los logaritmos en base e —es decir, el logaritmo natural—, por lo que cuando se utiliza el término logaritmo siempre significa «logaritmo natural».

La función logarítmica, $y = \ln(x)$, se representa en la Figura 8.4. Debe tenerse en cuenta que la función logarítmica está definida solamente para valores positivos de x . La función logarítmica presenta una pendiente inclinada al principio y luego se aplana (aunque la función continúa aumentando). La pendiente de la función logarítmica, $\ln(x)$, es $1/x$.

La función logarítmica presenta las siguientes propiedades útiles:

$$\ln(1/x) = -\ln(x); \quad (8.12)$$

$$\ln(ax) = \ln(a) + \ln(x); \quad (8.13)$$

$$\ln(x/a) = \ln(x) - \ln(a); \quad (8.14)$$

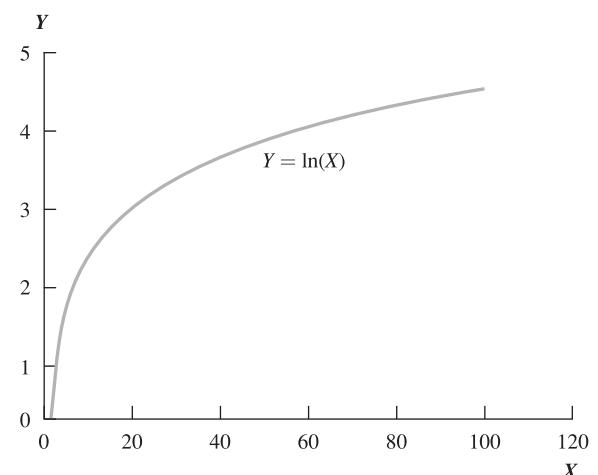
$$\ln(x^a) = a \ln(x). \quad (8.15)$$

Logaritmos y porcentajes. El vínculo entre el logaritmo y los porcentajes se basa en un hecho clave: Cuando Δx es pequeño, la diferencia entre el logaritmo de $x + \Delta x$ y el logaritmo de x es aproximadamente $\Delta x/x$, la variación porcentual en x dividida por 100. Es decir,

$$\ln(x + \Delta x) - \ln(x) \cong \frac{\Delta x}{x} \left(\text{si } \frac{\Delta x}{x} \text{ es pequeño} \right) \quad (8.16)$$

FIGURA 8.4 La función logarítmica, $Y = \ln(X)$

La función logarítmica $Y = \ln(X)$ es más inclinada para valores bajos que para valores altos de X , solamente está definida para $X > 0$, y su pendiente es $1/X$.



donde « \cong » significa «aproximadamente igual a». La deducción de esta aproximación se basa en el cálculo, pero se demuestra fácilmente probando algunos valores de x y Δx . Por ejemplo, cuando $x = 100$ y $\Delta x = 1$, entonces $\Delta x/x = 1/100 = 0,01$ (o el 1 %), al mismo tiempo $\ln(x + \Delta x) - \ln(x) = \ln(101) - \ln(100) = 0,00995$ (o el 0,995 %). Por lo tanto $\Delta x/x$ (que es 0,01) está muy cerca de $\ln(x + \Delta x) - \ln(x)$ (que es 0,00995). Cuando $\Delta x = 5$, $\Delta x/x = 5/100 = 0,05$, mientras que $\ln(x + \Delta x) - \ln(x) = \ln(105) - \ln(100) = 0,04879$.

Los tres modelos de regresión logarítmica. Existen tres casos distintos en los que pueden utilizarse logaritmos: cuando se transforma X tomando sus logaritmos pero no Y , cuando se transforma Y tomando su logaritmo pero no X , y cuando tanto Y como X se transforman en sus logaritmos. La interpretación de los coeficientes de la regresión es diferente en cada caso. Se analizan sucesivamente los tres casos.

Caso I: X está en logaritmos, Y no. En este caso, el modelo de regresión es

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i, \quad i = 1, \dots, n. \quad (8.17)$$

Debido a que Y no está expresada en logaritmos, pero X sí lo está, a veces se lo conoce como **modelo lineal-log**.

En el modelo lineal-log, una variación del 1 % en X está asociada con un cambio en Y de $0,01\beta_1$. Para comprobarlo, consideremos la diferencia en la función de regresión poblacional entre los valores de X que se diferencian en la cantidad ΔX : es decir $[\beta_0 + \beta_1 \ln(X + \Delta X)] - [\beta_0 + \beta_1 \ln(X)] = \beta_1 [\ln(X + \Delta X) - \ln(X)] \cong \beta_1 (\Delta X/X)$, donde para el último paso se utiliza la aproximación de la Ecuación (8.16). Si X cambia en un 1 %, entonces $\Delta X/X = 0,01$; por lo que en este modelo un cambio de un 1 % en X está asociado con un cambio en Y de $0,01\beta_1$.

La única diferencia entre el modelo de regresión de la Ecuación (8.17) y el modelo de regresión del Capítulo 4 con un único regresor es que la variable de la parte derecha ahora es el logaritmo de X en lugar del propio X . Para estimar los coeficientes β_0 y β_1 y de la Ecuación (8.17), primero se calcula una nueva variable, $\ln(X)$, que fácilmente se obtiene mediante una hoja de cálculo o mediante programas estadísticos. A continuación, β_0 y β_1 pueden ser estimados mediante la regresión MCO de la variable Y_i sobre $\ln(X)$, las hipótesis sobre β_1 se pueden contrastar utilizando el estadístico t , y puede construirse un intervalo de confianza al 95 % para β_1 de la forma $\beta_1 \pm 1,96ES(\hat{\beta}_1)$.

A modo de ejemplo, volvamos a la relación entre la renta del distrito y las calificaciones en los exámenes. En lugar de la especificación cuadrática, se puede utilizar la especificación lineal-log de la Ecuación (8.17). De la estimación de la regresión por MCO se obtiene

$$\widehat{\text{CalificaciónExamen}} = 557,8 + 36,2 \ln(\text{Renta}), \quad \bar{R}^2 = 0,561. \quad (8.18)$$

(3,8)	(1,40)
-------	--------

De acuerdo con la Ecuación (8.18), un aumento del 1 % en la renta está asociado con un aumento en las calificaciones de $0,01 \times 36,42 = 0,36$ puntos.

Para estimar el efecto sobre Y de un cambio en la X en sus unidades originales de miles de dólares (no en logaritmos), se puede utilizar el método del Concepto clave 8.1. Por ejemplo, ¿cuál es la diferencia prevista en las calificaciones de los exámenes en los distritos con un promedio de renta de 10.000 \$ frente a los de 11.000 \$? El valor estimado de ΔY es la diferencia entre los valores de predicción: $\Delta \hat{Y} = [557,8 + 36,42 \ln(11)] - [557,8 + 36,42 \ln(10)] = 36,42 \times [\ln(11) - \ln(10)] = 3,47$. Del mismo modo, la diferencia estimada entre un distrito con un promedio de renta de 40.000 \$ y un distrito con un promedio de renta de 41.000 \$ es $36,42 [\ln(41) - \ln(40)] = 0,90$. Por lo tanto, al igual que la especificación cuadrática, esta regresión predice que un aumento de 1.000 \$ en la renta tiene un efecto mayor sobre las calificaciones en los exámenes en los distritos pobres que en los distritos ricos.

La función de regresión lineal-log estimada de la Ecuación (8.18) se representa en la Figura 8.5. Debido a que el regresor de la Ecuación (8.18) es el logaritmo natural de la renta en vez de la renta, la función de regresión estimada no es una línea recta. Al igual que la función de regresión cuadrática de la Figura 8.3, al principio es inclinada, pero luego se aplana para niveles de renta más altos.

Caso II: Y está en logaritmos, X no. En este caso, el modelo de regresión es

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i. \quad (8.19)$$

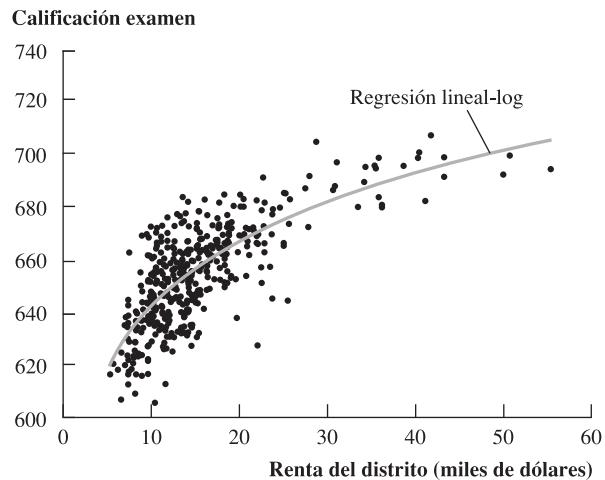
Al estar Y expresada en logaritmos, pero no X , se conoce como el **modelo log-lineal**.

En el modelo log-lineal, un cambio unitario en X ($\Delta X = 1$) está asociado con un cambio de $100 \times \beta_1$ % en Y . Para comprobarlo, se comparan los valores esperados de $\ln(Y)$ para los valores de X que se diferencian en ΔX . El valor esperado de $\ln(Y)$ dado X es $\ln(Y) = \beta_0 + \beta_1 X$. Cuando X es $X + \Delta X$ el valor esperado viene dado por $\ln(Y + \Delta Y) = \beta_0 + \beta_1(X + \Delta X)$. Por tanto, la diferencia entre estos valores esperados es $\ln(Y + \Delta Y) - \ln(Y) = [\beta_0 + \beta_1(X + \Delta X)] - [\beta_0 + \beta_1 X] = \beta_1 \Delta X$. De la aproximación de la Ecuación (8.16), sin embargo, si $\beta_1 \Delta X$ es pequeño, entonces $\ln(Y + \Delta Y) - \ln(Y) \cong \Delta Y/Y$. Por tanto, $\Delta Y/Y \cong \beta_1 \Delta X$. Si $\Delta X = 1$ es decir, X cambia en una unidad, entonces $\Delta Y/Y$ cambia en β_1 . Traducido a porcentajes, un cambio unitario en X está asociado con un cambio en Y de $100 \times \beta_1$ %.

A modo de ejemplo, volvamos al ejemplo empírico de la Sección 3.7, la relación entre la edad y los ingresos salariales de los graduados universitarios. Muchos de los contratos de empleo especifican que, por cada año de servicio adicional, un trabajador obtiene un cierto porcentaje de aumento en su salario. Esta relación porcentual sugiere una estimación de la especificación log-lineal de la Ecuación (8.19) de manera que cada año de edad adicional (X) está asociado, en media de la población, con algún incremento porcen-

FIGURA 8.5 La función de regresión lineal-log

La función de regresión lineal-log estimada $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \ln(X)$ capta la mayor parte de la relación no lineal entre las calificaciones en los exámenes y la renta del distrito.



tual constante de los ingresos salariales (Y). Calculando en primer lugar la nueva variable dependiente, $\ln(\text{Ingresos}_i)$, los coeficientes desconocidos β_0 y β_1 pueden ser estimados mediante regresión MCO de la variable $\ln(\text{Ingresos}_i)$ sobre la variable Edad_i . Cuando se estiman utilizando las 14.407 observaciones sobre los graduados universitarios de la Encuesta Actualizada de Población para marzo de 2009 (los datos se describen en el Apéndice 3.1), esta relación es

$$\widehat{\ln(\text{Ingresos})} = 2,805 + 0,0087\text{Edad}, \bar{R}^2 = 0,027. \quad (8.20)$$

(0,018) (0,0004)

De acuerdo con esta regresión, se prevé que los ingresos aumenten en un 0,87 % [($100 \times 0,0087$) %] por cada año adicional de edad.

Caso III: Tanto X como Y están en logaritmos. En este caso, el modelo de regresión es

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i. \quad (8.21)$$

Debido a que tanto Y como X se especifican en logaritmos, se conoce como **modelo log-log**.

En el modelo log-log, una variación del 1 % en X está asociada con una variación en Y de un β_1 %. Por tanto, en esta especificación β_1 es la elasticidad de Y con respecto a X . Para comprobarlo, se aplica de nuevo el Concepto clave 8.1; por lo que $\ln(Y + \Delta Y) - \ln(Y) = [\beta_0 + \beta_1 \ln(X + \Delta X)] - [\beta_0 + \beta_1 \ln(X)] = = \beta_1[\ln(X + \Delta X) - \ln(X)]$. Aplicando la aproximación de la Ecuación (8.16) en ambos lados de esta ecuación se obtiene

$$\frac{\Delta Y}{Y} \cong \beta_1 \frac{\Delta X}{X} \quad o$$

$$\beta_1 = \frac{\Delta Y/Y}{\Delta X/X} = \frac{100 \times (\Delta Y/Y)}{100 \times (\Delta X/X)} = \frac{\text{Variación porcentual en } Y}{\text{Variación porcentual en } X} \quad (8.22)$$

Por tanto, en la especificación log-log, β_1 es la ratio entre el cambio porcentual en Y asociado con el cambio porcentual en X y ese mismo cambio en X . Si el cambio porcentual en X es del 1 % (es decir, si $\Delta X = 0,01X$), entonces β_1 es el cambio porcentual en Y asociado con el cambio en X del 1 %. Es decir, β_1 es la elasticidad de Y con respecto a X .

A modo de ejemplo, volvemos a la relación entre la renta y las calificaciones. Cuando se especifica esta relación de este modo, los coeficientes desconocidos se estiman mediante una regresión del logaritmo de las calificaciones en los exámenes sobre el logaritmo de la renta. La ecuación estimada resultante es

$$\widehat{\ln(\text{CalificaciónExamen})} = 6,336 + 0,554 \ln(\text{Renta}), \bar{R}^2 = 0,557. \quad (8.23)$$

(0,006) (0,0021)

De acuerdo con esta función de regresión estimada, se espera que un aumento del 1 % en la renta se corresponda con un aumento del 0,0554 % en las calificaciones en los exámenes.

La función de regresión log-log estimada de la Ecuación (8.23) se representa en Figura 8.6. Debido a que Y está en logaritmos, el eje vertical de la Figura 8.6 mide el logaritmo de las calificaciones en los exámenes y el diagrama de dispersión es del logaritmo de las calificaciones en los exámenes versus la renta del distrito. A efectos comparativos, la Figura 8.6 muestra asimismo la estimación de la función de regresión con una especificación log-lineal, que es

$$\widehat{\text{CalificaciónExamen}} = 6,439 + 0,00284(\text{Renta}), \bar{R}^2 = 0,497. \quad (8.24)$$

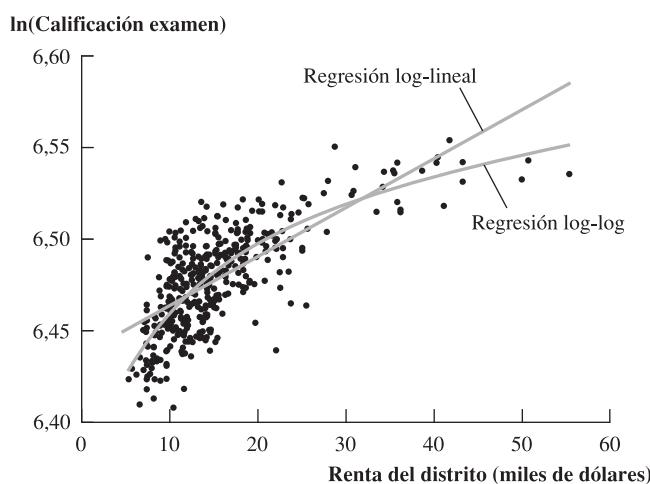
(0,003) (0,0018)

Debido a que el eje vertical está expresado en logaritmos, la función de regresión de la Ecuación (8.24) es la línea recta de la Figura 8.6.

Como se puede ver en la Figura 8.6, la especificación log-log se ajusta un poco mejor que la especificación log-lineal. Esto es congruente con el \bar{R}^2 más alto que presenta la regresión log-log (0,557)

FIGURA 8.6 Las funciones de regresión log-lineal y log-log

En la función de regresión log-lineal, $\ln(Y)$ es una función lineal de X . En la función de regresión log-log, $\ln(Y)$ es una función lineal de $\ln(X)$.



que el de la regresión log-lineal (0,497). Aún así, la especificación log-log no se ajusta especialmente bien a los datos: en los valores más bajos de la renta, la mayor parte de las observaciones se sitúan por debajo de la curva log-log, mientras que para un nivel medio de renta, la mayoría de las observaciones se sitúan por encima de la función de regresión estimada.

Los tres modelos de regresión logarítmica se resumen en el Concepto clave 8.2.

Una dificultad al comparar las especificaciones logarítmicas. ¿Cuál de los modelos logarítmicos de regresión se ajusta mejor a los datos? Tal y como se vio al analizar las Ecuaciones (8.23) y (8.24), el \bar{R}^2 puede utilizarse para comparar los modelos log-lineal y log-log; y ocurre que, el modelo log-log presenta el \bar{R}^2 más alto. Del mismo modo, el \bar{R}^2 puede utilizarse para comparar la regresión lineal-log de la Ecuación (8.18) con la regresión lineal de Y sobre X . En la regresión de las calificaciones en los exámenes y la renta, la regresión lineal-log presenta un \bar{R}^2 de 0,561, mientras que la regresión lineal tiene un \bar{R}^2 de 0,508, por lo que el modelo lineal-log se ajusta mejor a los datos.

CONCEPTO CLAVE

8.2

Logaritmos en la regresión: tres casos

Los logaritmos pueden utilizarse para transformar la variable dependiente Y , una variable independiente X , o ambas (pero la variable que se transforme debe ser positiva). La siguiente tabla resume estos tres casos, así como la interpretación del coeficiente de regresión β_1 . En cada caso, se puede estimar β_1 mediante la aplicación de MCO tras haber tomado logaritmos de la variable dependiente y/o las independientes.

Caso	Especificación de la regresión	Interpretación de β_1
I	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$	Un 1 % de cambio en X está asociado con un cambio en Y de $0,01\beta_1$.
II	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$	Un cambio en X de una unidad ($\Delta X = 1$) está asociado con un cambio en Y de $100\beta_1$.
III	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$	Un 1 % de cambio en X está asociado con un cambio en Y de β_1 %, por lo que β_1 es la elasticidad de Y respecto a X .

¿Cómo se pueden comparar el modelo lineal-log y el modelo log-log? Desafortunadamente, el \bar{R}^2 no puede ser utilizado para comparar estas dos regresiones debido a que sus variables dependientes son diferentes [en uno es Y_i , y en el otro es $\ln(Y_i)$]. Recordemos que el \bar{R}^2 mide la proporción de la varianza de la variable dependiente que está explicada por los regresores. Debido a que las variables dependientes de los modelos log-log y lineal-log son diferentes, no tiene sentido comparar sus respectivas \bar{R}^2 .

Debido a este problema, lo mejor que se puede hacer en cada caso particular es decidir si tiene sentido especificar Y en logaritmos, de acuerdo con la teoría económica y según el propio conocimiento previo del problema en cuestión, así como el de otros expertos. Por ejemplo, en economía laboral, generalmente se modelizan los salarios utilizando logaritmos debido a que las comparaciones de salarios, los aumentos salariales contractuales, etc., son a menudo tratados de un modo más natural en términos porcentuales. En la modelización de las calificaciones de los exámenes, parece natural (en cualquier caso, nos lo parece a nosotros) tratar los resultados del examen en términos de puntos obtenidos en los exámenes de la prueba, en lugar de en términos de porcentajes de aumento en las calificaciones, por lo que nos centramos en los modelos en los que la variable dependiente es la calificación en los exámenes de la prueba, en lugar de su logaritmo.

Cálculo de los valores de predicción de Y cuando Y está expresada en logaritmos³. Si la variable dependiente Y se ha transformado tomando logaritmos, puede utilizarse la regresión estimada para calcular directamente el valor de predicción de $\ln(Y)$. Sin embargo, es un poco más difícil de calcular el valor esperado de Y en sí mismo.

Para comprobarlo, consideremos el modelo de regresión log-lineal de la Ecuación (8.19) y lo reescribimos de forma que quede especificado en términos de Y en vez de $\ln(Y)$. Para ello, se toma la función exponencial a ambos lados de la Ecuación (8.19), el resultado es

$$Y_i = \exp(\beta_0 + \beta_1 X_i + u_i) = e^{\beta_0 + \beta_1 X_i} e^{u_i}. \quad (8.25)$$

El valor esperado de Y_i dado X_i es $E(Y_i|X_i) = E(e^{\beta_0 + \beta_1 X_i} e^{u_i}|X_i) = e^{\beta_0 + \beta_1 X_i} E(e^{u_i}|X_i)$. El problema es que incluso aunque $E(u_i|X_i) = 0$, $E(e^{u_i}|X_i) \neq 1$. Por tanto, el valor de predicción adecuado para Y_i no se obtiene simplemente tomando la función exponencial de $\hat{\beta}_0 + \hat{\beta}_1 X_i$, es decir, haciendo $\hat{Y}_i = e^{\hat{\beta}_0 + \hat{\beta}_1 X_i}$: este valor teórico está sesgado debido al factor omitido $E(e^{u_i}|X_i)$.

Una solución a este problema consiste en estimar el factor $E(e^{u_i}|X_i)$ y utilizar esta estimación para calcular el valor esperado de Y . El Ejercicio 17.12 trabaja con distintas maneras de estimar $E(e^{u_i}|X_i)$, pero esto se hace complicado, sobre todo si u_i es heterocedástico, por lo que no continuaremos más allá.

Otra solución, que es el método utilizado en este libro, consiste en calcular los valores esperados del logaritmo de Y , pero sin transformarlos a sus unidades originales. En la práctica, esto suele ser aceptable cuando la variable dependiente se especifica como un logaritmo, a menudo es más natural utilizar solamente la especificación logarítmica (y las interpretaciones porcentuales asociadas) a lo largo del análisis.

Modelos polinomiales y logarítmicos para calificaciones y renta del distrito

En la práctica, la teoría económica o el juicio experto podrían sugerir utilizar una forma funcional determinada, pero en realidad la verdadera forma de la función de regresión poblacional es desconocida. En la práctica, por tanto, el ajuste de una función no lineal implica decidir qué método o combinación de métodos funciona mejor. A modo de ejemplo, se comparan los modelos logarítmicos y polinomiales para la relación entre la renta del distrito y las calificaciones en los exámenes.

Especificaciones polinomiales. Se han considerado dos especificaciones polinomiales utilizando las potencias de la variable *Renta*, la cuadrática [Ecuación (8.2)] y la cúbica [(8.11)]. Debido a que el coeficiente de la variable *Renta*³ en la Ecuación (8.11) era significativo al nivel del 5 %, la especificación cúbica proporcionaba una mejora sobre la cuadrática, por lo que se seleccionó el modelo cúbico como especificación polinomial preferida.

³ Este material es más avanzado y puede saltarse sin pérdida de continuidad.

Especificaciones logarítmicas. La especificación logarítmica de la Ecuación (8.18) parecía proporcionar un buen ajuste a los datos, pero esto no se contrastó de manera formal. Una manera de hacerlo es generalizándola con potencias superiores del logaritmo de la renta. Si estos términos adicionales no son estadísticamente distintos de cero, entonces se puede concluir que la especificación de la Ecuación (8.18) es adecuada en el sentido de que no puede ser rechazada frente a la función polinomial de los logaritmos. En consecuencia, la regresión cúbica estimada (especificada en potencias del logaritmo de la renta) es

$$\text{Calificación Examen} = 486,1 + 113,4 \ln(\text{Renta}) - 26,9[\ln(\text{Renta})]^2 + 3,06[\ln(\text{Renta})]^3, \bar{R}^2 = 0,560. \quad (8.26)$$

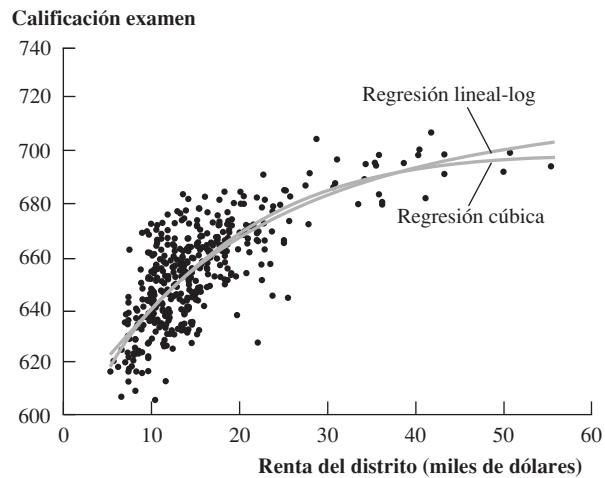
(79,4)	(87,9)	(31,7)	(3,74)
--------	--------	--------	--------

El estadístico t del coeficiente del término cúbico es 0,818, por lo que la hipótesis nula de que el verdadero coeficiente es igual a cero, no se rechaza al nivel del 10 %. El estadístico F para contrastar la hipótesis conjunta de que los verdaderos coeficientes de los términos cuadrático y cúbico son ambos iguales a cero es 0,44, con un p -valor de 0,64, por lo que esta hipótesis nula conjunta no se rechaza al nivel del 10 %. Por tanto, el modelo logarítmico cúbico de la Ecuación (8.26) no proporciona una mejora estadísticamente significativa sobre el modelo de la Ecuación (8.18), que es lineal en el logaritmo de la renta.

Comparación de las especificaciones cúbica y lineal-log. En la Figura 8.7 se representan las funciones de regresión estimadas a partir de la especificación cúbica de la Ecuación (8.11) y de la especificación lineal-log de la Ecuación (8.18). Las dos funciones de regresión estimadas son muy similares. Una herramienta estadística para comparar estas especificaciones es el \bar{R}^2 . El \bar{R}^2 de la regresión logarítmica es 0,561, y el de la regresión cúbica es 0,555. Debido a que la especificación logarítmica presenta una ligera ventaja en términos de \bar{R}^2 y debido a que esta especificación no necesita polinomios de mayor grado sobre los logaritmos de la renta para ajustarse a estos datos, se adopta la especificación logarítmica de la Ecuación (8.18).

FIGURA 8.7 Las funciones de regresión lineal-log y cúbica

La función de regresión cúbica estimada [Ecuación (8.11)] y la función de regresión lineal-log estimada [Ecuación (8.18)] son prácticamente idénticas en este ejemplo.



8.3 Interacciones entre variables independientes

En la introducción de este capítulo nos preguntábamos si la reducción de la ratio estudiantes-maestros podría tener un efecto mayor sobre las calificaciones obtenidas en los distritos donde muchos estudiantes todavía estén aprendiendo inglés, que en aquellos con pocos estudiantes que todavía estén aprendiendo inglés. Esto podría ocurrir si, por ejemplo, los estudiantes que están aprendiendo inglés se beneficiasen de forma distinta de la enseñanza personalizada o en grupos pequeños. Si es así, la presencia de muchos es-

tudiantes aprendiendo inglés en un distrito podría interactuar con la ratio estudiantes-maestros, de tal de manera que el efecto sobre las calificaciones en los exámenes de una variación en la ratio estudiantes-maestros dependiera de la proporción de estudiantes de inglés.

En esta sección se explica cómo incorporar dichas interacciones entre dos variables independientes al modelo de regresión múltiple. La posible interacción entre la ratio estudiantes-maestros y la proporción de estudiantes aprendiendo inglés es un ejemplo de una situación más general, en la que el efecto sobre Y de un cambio en una variable independiente depende del valor de otra variable independiente. Se consideran tres casos: cuando las dos variables independientes son binarias, cuando una es binaria y la otra es continua, y cuando ambas son continuas.

Interacciones entre dos variables binarias

Consideremos la regresión poblacional del logaritmo de los ingresos salariales [Y_i , donde $Y_i = \ln(\text{Ingresos}_i)$] sobre dos variables binarias: si un trabajador tiene un título universitario (D_{1i} , donde $D_{1i} = 1$ si la i -ésima persona es graduada universitaria) y el género del trabajador (D_{2i} , donde $D_{2i} = 1$ si la i -ésima persona es de sexo femenino). La regresión lineal poblacional de Y_i sobre estas dos variables binarias es

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i. \quad (8.27)$$

En este modelo de regresión, β_1 es el efecto de tener un título universitario, manteniendo constante el género, y β_2 es el efecto sobre el logaritmo de los ingresos de ser mujer, manteniendo constante el nivel de educación alcanzado.

La especificación de la Ecuación (8.27) presenta una limitación importante: en esta especificación, el efecto de tener un título universitario, manteniendo constante el género, es el mismo para hombres y mujeres. No existe, sin embargo, ninguna razón para creer que esto deba ser así. Expresado matemáticamente, el efecto sobre Y_i de D_{1i} , manteniendo constante D_{2i} , podría depender del valor de D_{2i} . En otras palabras, podría haber una interacción entre tener un título universitario y el género, de manera que el valor en el mercado de trabajo de un grado universitario podría ser diferente para hombres y mujeres.

Aunque la especificación de la Ecuación (8.27) no permite esta interacción entre tener un título universitario y el género, resulta sencillo modificar la especificación, por lo que así se hace, mediante la introducción de otro regresor, el producto de los dos variables binarias, $D_{1i} \times D_{2i}$. La regresión resultante es

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i. \quad (8.28)$$

El nuevo regresor, el producto $D_{1i} \times D_{2i}$, se denomina **término de interacción o regresor de interacción**, y el modelo de regresión poblacional de la Ecuación (8.28) se denomina **modelo de regresión con interacciones** de variables binarias.

El término de interacción de la Ecuación (8.28) permite el efecto poblacional sobre el logaritmo de los ingresos salariales (Y_i) de tener un título universitario (el cambio de D_{1i} de $D_{1i} = 0$ a $D_{1i} = 1$) dependiendo del género (D_{2i}). Para demostrar esto matemáticamente, se calcula el efecto poblacional de un cambio en D_{1i} utilizando el método general establecido en el Concepto clave 8.1. El primer paso consiste en calcular la esperanza condicional de Y_i para $D_{1i} = 0$, dado un valor de D_{2i} ; es decir $E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_0 + \beta_1 \times 0 + \beta_2 \times d_2 + \beta_3 \times (0 - d_2) = \beta_0 + \beta_2 d_2$, donde se utiliza el supuesto de media condicional igual a cero, $E(u_i | D_{1i}, D_{2i}) = 0$. El siguiente paso es calcular la esperanza condicional de Y_i tras el cambio —es decir, para $D_{1i} = 1$ — dado el mismo valor de D_{2i} ; es decir $E(Y_i | D_{1i} = 1, D_{2i} = d_2) = \beta_0 + \beta_1 \times 1 + \beta_2 \times d_2 + \beta_3 \times (1 - d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2$. El efecto de este cambio es la diferencia de los valores esperados [es decir, la diferencia de la Ecuación (8.4)], que es

$$E(Y_i | D_{1i} = 1, D_{2i} = d_2) - E(Y_i | D_{1i} = 0, D_{2i} = d_2) = \beta_1 + \beta_3 d_2. \quad (8.29)$$

Por lo tanto, en la especificación con interacción de variables binarias de la Ecuación (8.28), el efecto de obtener un título universitario (un cambio unitario en D_{1i}) depende del género de la persona [el valor de D_{2i} , que es d_2 en la Ecuación (8.29)]. Si la persona es de sexo masculino ($d_2 = 0$), el efecto de la obtención de

**CONCEPTO
CLAVE**
8.3
Un método para la interpretación de los coeficientes en regresiones con variables binarias

En primer lugar se calcula el valor esperado de Y para cada uno de los casos posibles descritos por el conjunto de variables binarias. A continuación se comparan estos valores esperados. Cada coeficiente puede expresarse entonces como un valor esperado o como la diferencia entre dos o más valores esperados.

un título universitario es β_1 , pero si la persona es de sexo femenino ($d_2 = 1$), el efecto es $\beta_1 + \beta_3$. El coeficiente β_3 del término de interacción es la diferencia entre el efecto de la obtención de un título universitario para las mujeres frente a los hombres.

Aunque este ejemplo se formuló en términos de la variable logaritmo de los ingresos, la variable tener un título universitario, y la variable de género, la cuestión es de carácter general. La regresión con interacción entre variables binarias permite que el efecto del cambio en una de las variables independientes binarias pueda depender del valor de la otra variable binaria.

El método utilizado aquí para interpretar los coeficientes es, en efecto, trabajar con cada posible combinación de las variables binarias. Este método, que es aplicable a todas las regresiones con variables binarias, se resume en el Concepto clave 8.3.

Aplicación a la ratio estudiantes-maestros y porcentaje de estudiantes aprendiendo inglés. Sea $REMA_i$ una variable binaria que vale 1 si la ratio estudiantes-maestros es de 20 o mayor (alto) y es igual a 0 en cualquier otro caso, y sea EIA_i una variable binaria que es igual a 1 si el porcentaje de estudiantes aprendiendo inglés es del 10 % o más (alto), y es igual a 0 en caso contrario. La regresión de interacción de las calificaciones en las pruebas sobre las variables $REMA_i$ y EIA_i es

$$\widehat{\text{CalificaciónExamen}} = 664,1 - 1,9 \text{ } REMA - 18,2 \text{ } EIA - 3,5(\text{REMA} \times \text{EIA}), \bar{R}^2 = 0,290. \quad (8.30)$$

(1,4)	(1,9)	(2,3)	(3,1)
-------	-------	-------	-------

El efecto esperado del traslado desde un distrito con una ratio de alumnos por maestro baja a uno con una elevada proporción de alumnos por maestro, manteniendo constante el hecho de si el porcentaje de estudiantes que todavía están aprendiendo inglés es alto o bajo, viene dado por la Ecuación (8.29), en la que los coeficientes estimados sustituyen a los coeficientes poblacionales. De acuerdo con las estimaciones de la Ecuación (8.30), este efecto por tanto es $-1,9 - 3,5EIA$. Es decir, si la proporción de estudiantes aprendiendo inglés es baja ($EIA = 0$), entonces el efecto sobre las calificaciones en los exámenes de trasladarse desde un distrito con $REMA = 0$ a uno con $REMA = 1$ es que las calificaciones en los exámenes se reduzcan en 1,9 puntos. Si la proporción de estudiantes aprendiendo inglés es alta, se estima que las calificaciones en los exámenes disminuyan en $1,9 + 3,5 = 5,4$ puntos.

La regresión estimada en la Ecuación (8.30), asimismo, puede utilizarse para estimar la media de las calificaciones en los exámenes para cada una de las cuatro combinaciones posibles de las variables binarias.

Esto se obtiene utilizando el procedimiento del Concepto clave 8.3. Por lo tanto, la media muestral de las calificaciones en los exámenes de los distritos con ratios estudiantes-maestros bajas ($REMA_i = 0$) y de baja proporción de estudiantes aprendiendo inglés ($EIA_i = 0$) es 664,1. Para los distritos con $REMA_i = 1$ (ratio de estudiantes maestros alta) y $EIA_i = 0$ (proporción de estudiantes aprendiendo inglés baja), la media muestral es 662,2 ($= 664,1 - 1,9$). Si $REMA_i = 0$ y $EIA_i = 1$, la media muestral es 645,9 ($= 664,1 - 18,2$), y cuando $REMA_i = 1$ y $EIA_i = 1$, la media muestral es 640,5 ($= 664,1 - 1,9 - 18,2 - 3,5$).

Interacciones entre una variable continua y una variable binaria

Consideremos ahora la regresión poblacional del logaritmo de los ingresos [$Y_i = \ln(\text{Ingresos}_i)$] sobre una variable continua, los años de experiencia laboral de una persona (X_i), y una variable binaria, si el trabajador tiene un título universitario (D_i , donde $D_i = 1$ si la i -ésima persona es graduada universitaria).

Como se muestra en la Figura 8.8, la recta de regresión poblacional que relaciona Y y la variable continua X puede depender de la variable binaria D de tres formas diferentes.

En la Figura 8.8(a), las dos rectas de regresión difieren tan solo en sus correspondientes interceptos. El modelo de regresión poblacional correspondiente es

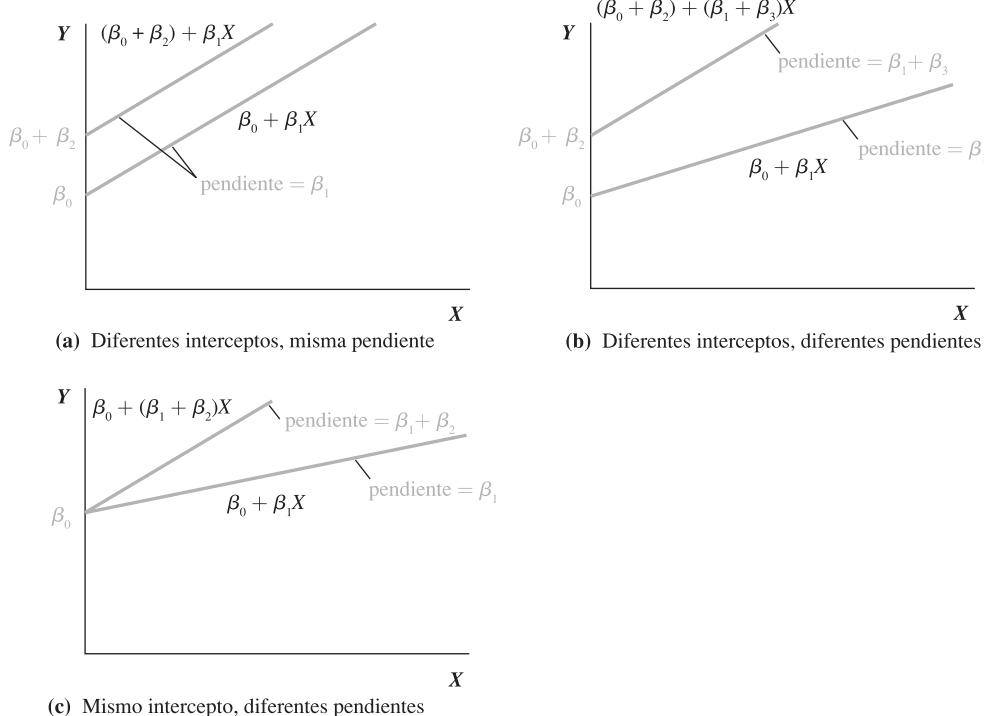
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i. \quad (8.31)$$

Este es el modelo habitual de regresión múltiple con una función de regresión poblacional que es lineal en X_i y D_i . Cuando $D_i = 0$, la función de regresión poblacional es $\beta_0 + \beta_1 X_i$. Por lo que el intercepto es β_0 y la pendiente es β_1 . Cuando $D_i = 1$ la función de regresión poblacional es $\beta_0 + \beta_1 X_i + \beta_2$, por lo que la pendiente sigue siendo β_1 , pero el intercepto es $\beta_0 + \beta_2$. Por lo tanto β_2 es la diferencia entre los interceptos de las dos rectas de regresión, como se muestra en la Figura 8.8(a). Expresado en términos del ejemplo de los ingresos salariales, β_1 es el efecto sobre el logaritmo de los ingresos de un año adicional de experiencia laboral, manteniendo constante el estatus de titulación universitaria, y β_2 es el efecto de un título universitario sobre el logaritmo de los ingresos, manteniendo los años de experiencia laboral constantes. En esta especificación, el efecto de un año adicional de experiencia laboral es el mismo para los graduados universitarios y los no graduados; es decir, las dos rectas de la Figura 8.8(a) tienen la misma pendiente.

En la Figura 8.8(b), las dos rectas tienen diferentes pendientes e interceptos. Las pendientes diferentes permiten que el efecto de un año adicional de trabajo sea distinto para los graduados universitarios y para los no graduados. Para permitir diferentes pendientes, se añade un término de interacción a la Ecuación (8.31):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i. \quad (8.32)$$

FIGURA 8.8 Funciones de regresión con variables binarias y continuas



Las Interacciones entre variables binarias y variables continuas pueden dar lugar a tres funciones de regresión poblacionales diferentes: (a) $\beta_0 + \beta_1 X + \beta_2 D$ permite diferentes interceptos pero presenta la misma pendiente, (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D)$ permite diferentes interceptos y diferentes pendientes, y (c) $\beta_0 + \beta_1 X + \beta_2 (X \times D)$ presenta el mismo intercepto pero permite pendientes diferentes.

donde $X_i \times D_i$ es una nueva variable, el producto de X_i y D_i . Para interpretar los coeficientes de esta regresión, se aplica el procedimiento del Concepto clave 8.3. Al hacerlo se demuestra que, si $D_i = 0$, la función de regresión poblacional es $\beta_0 + \beta_1 X_i$, mientras que si $D_i = 1$, la función de regresión poblacional es $(\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i$. Por tanto, esta especificación permite dos funciones de regresión poblacionales diferentes que relacionan Y_i y X_i , dependiendo del valor de D_i , como se muestra en la Figura 8.8(b). La diferencia entre los dos interceptos es β_2 y la diferencia entre las dos pendientes es β_3 . En el ejemplo de los ingresos salariales, β_1 es el efecto de un año adicional de experiencia laboral para los no graduados ($D_i = 0$) y $\beta_1 + \beta_3$ es este efecto para los graduados, por lo que β_3 es la *diferencia* en el efecto de un año adicional de experiencia laboral para los graduados universitarios frente a los no graduados.

**CONCEPTO
CLAVE**

8.4

Interacciones entre variables binarias y continuas

Mediante el uso del término de interacción $X_i \times D_i$, la recta de regresión poblacional que relaciona Y_i con la variable continua X_i puede tener una pendiente que dependa de la variable binaria D_i . Existen tres posibilidades:

1. Interceptos diferentes, misma pendiente [Figura 8.8(a)]:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$

2. Diferentes interceptos y pendientes [Figura 8.8(b)]:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3(X_i \times D_i) + u_i$$

3. Mismo intercepto, pendientes diferentes [Figura 8.8(c)]:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2(X_i \times D_i) + u_i$$

Una tercera posibilidad, que se muestra en la Figura 8.8(c), es que las dos rectas tengan diferentes pendientes, pero el mismo intercepto. El modelo de regresión con interacciones para este caso es

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2(X_i \times D_i) + u_i. \quad (8.33)$$

Los coeficientes de esta especificación pueden interpretarse asimismo mediante el Concepto clave 8.3. En los términos del ejemplo de los ingresos salariales, esta especificación permite que los efectos sobre el logaritmo de los ingresos de la experiencia sean diferentes entre los titulados universitarios y los no graduados, pero exige que la esperanza del logaritmo de los ingresos sea la misma para ambos grupos cuando no existe experiencia previa. Dicho de otro modo, esta especificación se corresponde con una media poblacional para el nivel inicial de salarios igual para los titulados universitarios y los no titulados. Esto no tiene mucho sentido en este caso, y en la práctica esta especificación se utiliza con menos frecuencia que la Ecuación (8.32), que permite diferentes interceptos y pendientes.

Las tres especificaciones de las Ecuaciones (8.31), (8.32) y (8.33) son versiones del modelo de regresión múltiple del Capítulo 6, y una vez que se ha creado una nueva variable $X_i \times D_i$, los coeficientes de todos ellos pueden estimarse mediante MCO.

Los tres modelos de regresión que incluyen una variable independiente binaria y una variable independiente continua se resumen en el Concepto clave 8.4.

Aplicación a la ratio estudiantes maestros y porcentaje de estudiantes aprendiendo inglés. ¿Depende el efecto de la disminución de la ratio estudiantes-maestros sobre las calificaciones de si el porcentaje de estudiantes que aún están aprendiendo inglés es alto o bajo? Una forma de responder a esta pregunta es mediante una especificación que permita dos rectas de regresión diferentes, en función de si hay un alto o un bajo porcentaje de estudiantes aprendiendo inglés. Esto se logra utilizando diferentes especificaciones para el intercepto y para la pendiente:

$$\overbrace{\text{CalificaciónExamen}} = 682,2 - 0,97 \text{REM} + 5,6 \text{EIA} - 1,28(\text{REM} \times \text{EIA}), \bar{R}^2 = 0,305. \quad (8.34)$$

(11,9)	(0,59)	(19,5)	(0,97)
--------	--------	--------	--------

donde la variable binaria EIA_i es igual a 1 si el porcentaje de estudiantes del distrito que están aprendiendo inglés es superior al 10 % y es igual a 0 en caso contrario.

Para los distritos con una proporción baja de estudiantes aprendiendo inglés ($EIA_i = 0$), la recta de regresión estimada es $682,2 - 0,97REM_i$. Para los distritos con una proporción elevada de estudiantes aprendiendo inglés ($EIA_i = 1$), la recta de regresión estimada es $682,2 + 5,6 - 0,97REM_i - 1,28REM_i = 687,8 - 2,25REM_i$. De acuerdo con estas estimaciones, se prevé que la reducción de la ratio de estudiantes por maestro en uno aumente las calificaciones en los exámenes en 0,97 puntos en los distritos con proporción baja de alumnos aprendiendo inglés, pero en 2,25 puntos en los distritos con proporción alta de estudiantes aprendiendo inglés. La diferencia entre estos dos efectos, 1,28 puntos, es el coeficiente del término de interacción de la Ecuación (8.34).

El modelo de regresión con interacciones de la Ecuación (8.34) permite estimar los efectos de las intervenciones políticas más matizadas que la reducción del tamaño de las clases en todos los ámbitos considerada anteriormente. Por ejemplo, supongamos que el Estado considera una política de reducción de la ratio estudiantes-maestros en una cuantía de 2 en los distritos con una alta proporción de alumnos aprendiendo inglés ($EIA_i = 1$), pero deja el tamaño de las clases en otros distritos igual. La aplicación del método del Concepto Clave 8.1 a las Ecuaciones (8.32) y (8.34) muestra que la estimación del efecto de esta reducción para los distritos en los que $EIA = 1$ es $-2(\hat{\beta}_1 + \hat{\beta}_3) = 4,50$. El error estándar de este efecto estimado es $ES(-2\hat{\beta}_1 - 2\hat{\beta}_3) = 1,53$, que puede calcularse mediante la Ecuación (8.8) y los métodos de la Sección 7.3.

La regresión MCO de la Ecuación (8.34) se puede utilizar para contrastar algunas hipótesis sobre la recta de regresión poblacional. En primer lugar, la hipótesis de que las dos rectas son en realidad la misma puede contrastarse mediante el cálculo del estadístico F para el contraste de la hipótesis conjunta de que el coeficiente de EIA_i y el coeficiente del término de interacción $REM_i \times EIA_i$ son iguales a cero. Este estadístico F es 89,9, que es significativo al nivel del 1 %.

En segundo lugar, la hipótesis de que las dos rectas tienen la misma pendiente se puede contrastar mediante el contraste de que el coeficiente del término de interacción es cero. El estadístico t , $-1,28/0,97 = -1,32$, es menor que 1,645 en valor absoluto, por lo que la hipótesis nula de que las dos rectas tienen la misma pendiente no puede rechazarse mediante un contraste bilateral al 10 % de nivel de significación.

En tercer lugar, la hipótesis de que las dos rectas tienen el mismo intercepto se corresponde con la restricción de que el coeficiente poblacional de EIA es cero. El estadístico t para el contraste de esta restricción es $t = 5,6/19,5 = 0,29$, por lo que la hipótesis de que las rectas tienen el mismo intercepto no puede rechazarse al nivel de 5 %.

Estos tres contrastes arrojan resultados aparentemente contradictorios: el contraste conjunto con el estadístico F rechaza la hipótesis conjunta de que la pendiente y el intercepto son iguales, pero los contrastes de hipótesis individuales que utilizan el estadístico t no lo rechazan. La razón es que las variables explicativas, EIA y $REM \times EIA$, están altamente correlacionadas. Esto da lugar a errores estándar elevados para los coeficientes individuales. A pesar de que resulta imposible decir cuál de los coeficientes es distinto de cero, existe una fuerte evidencia en contra de la hipótesis de que *ambos* son iguales a cero.

Finalmente, la hipótesis de que la ratio estudiantes-maestros no entra en esta especificación, se puede contrastar mediante el cálculo del estadístico F para la hipótesis conjunta de que los coeficientes de REM y del término de interacción son ambos iguales a cero. Este estadístico F es 5,64, que tiene un p -valor de 0,004. Por tanto, los coeficientes de la ratio estudiantes-maestros son estadísticamente significativos al nivel de significación del 1 %.

Interacciones entre dos variables continuas

Supongamos ahora que ambas variables independientes (X_{1i} y X_{2i}) son continuas. Un ejemplo de ello es cuando Y_i es el logaritmo de los ingresos salariales del trabajador i -ésimo, X_{1i} son sus años de experiencia laboral, y X_{2i} es el número de años que él o ella fueron a la escuela. Si la función de regresión poblacional es lineal, el efecto sobre los salarios de un año adicional de experiencia no depende del número de años de educación, o, equivalentemente, el efecto de un año adicional de educación no depende del número de años

La rentabilidad de la educación y la brecha de género

Además de placeres intelectuales, la educación reporta beneficios económicos. Como muestran los recuadros de los Capítulos 3 y 5, los trabajadores con mayor nivel educativo alcanzado tienden a ganar más que sus homólogos con menos educación. No obstante, el análisis en esos recuadros quedaba incompleto, al menos por tres razones. En primer lugar, no tenía en cuenta otros factores determinantes de los ingresos salariales que podrían estar correlacionados con los logros educativos, por lo que el estimador MCO del coeficiente de la educación podría presentar sesgo de variable omitida. En segundo lugar, la forma funcional que se utilizaba en el Capítulo 5 —una relación lineal sencilla—, implica una variación en los ingresos en una cuantía constante de dólares por cada año adicional de educación, mientras que se podría sospechar que la variación en los ingresos salariales en dólares es en realidad mayor para los niveles superiores de educación. En tercer lugar, el recuadro del Capítulo 5 ignoraba las diferencias por género en los ingresos salariales que resaltaba el recuadro del Capítulo 3.

Todas estas limitaciones pueden ser abordadas mediante un análisis de regresión múltiple que tenga en cuenta los determinantes de los ingresos que, si se omiten, podrían causar un sesgo de variable omitida y que utilice una forma funcional lineal para relacionar la educación y los ingresos salariales. La Tabla 8.1 resume las regresiones estimadas utilizando los datos sobre trabajadores a tiempo completo, de edad comprendida entre 30 y 64 años, de la Encuesta Actualizada de Población (los datos de la CPS se describen en el Apéndice 3.1). La variable dependiente es el logaritmo del salario por hora, por lo que un año más de educación se asocia con un aumento porcentual constante (no un aumento en dólares) de los ingresos salariales.

La Tabla 8.1 presenta cuatro resultados destacados. En primer lugar, la omisión del género en la regresión (1) no produce un sesgo de variable omitida importante: a pesar de que el género entra en la regresión (2) de manera significativa y con un coeficiente elevado, el género y los años de educación no están correlacionados; es decir, en promedio, los hombres y las mujeres tienen prácticamente los mismos niveles de educación. En segundo lugar, la rentabilidad de la educación es eco-

nómica, estadística y significativamente diferente para los hombres y las mujeres: en la regresión (3), el estadístico t para el contraste de la hipótesis de que es igual es 7,02 ($=0,0121/0,0017$). En tercer lugar, la regresión (4) tiene en cuenta (controla por) la región del país en la que reside el individuo, abordando por tanto el sesgo potencial de variable omitida que podría surgir si los años de educación difiriesen sistemáticamente en función de la región. Controlar por la región da lugar a una pequeña diferencia en los coeficientes estimados para los términos de educación, en relación a los presentados para la regresión (3). En cuarto lugar, la regresión (4) controla por la experiencia potencial del trabajador, medida por los años transcurridos desde la finalización de los estudios. Los coeficientes estimados implican un valor marginal decreciente para cada año de experiencia potencial.

La rentabilidad económica de la educación estimada en la regresión (4) es 10,32 % para cada año de educación para los hombres y 11,66 % ($=0,1032 + 0,0134$, en porcentaje) para las mujeres. Debido a que las funciones de regresión para los hombres y las mujeres tienen diferentes pendientes, la brecha de género depende de los años de educación. Para 12 años de educación, la brecha de género se estima en 29,0 % ($=0,0134 + 12 - 0,451$, en porcentaje); para 16 años de educación, la brecha de género es menor en términos porcentuales, un 23,7 %.

Estas estimaciones del rendimiento de la educación y la brecha de género todavía presentan limitaciones, entre ellas la posibilidad de que otras variables omitidas, en particular, la capacidad innata del trabajador, y los posibles problemas asociados con la forma en que se miden las variables en la CPS. Sin embargo, las estimaciones de la Tabla 8.1 son congruentes con las obtenidas por los economistas que prestan especial atención a estas limitaciones. Una encuesta, llevada a cabo por el económetra David Card (1999) sobre docenas de estudios empíricos, concluye que las mejores estimaciones de los economistas laborales para la rentabilidad de la educación, en general, la sitúan entre el 8 % y el 11 %, y que la rentabilidad depende de la calidad de la educación. Para aprender más sobre la rentabilidad económica de la educación, véase Card (1999).

de experiencia laboral. En realidad, sin embargo, podría existir una interacción entre estas dos variables, de modo que el efecto sobre los salarios de un año adicional de experiencia dependa del número de años de educación. Esta interacción puede modelizarse ampliando el modelo de regresión lineal con un término de interacción que es el producto de X_{1i} y X_{2i} :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i. \quad (8.35)$$

El término de interacción permite que el efecto de un cambio unitario en X_1 dependa de X_2 . Para comprobarlo, se aplica el método general para calcular los efectos en los modelos de regresión no lineal del Concepto clave 8.1. La diferencia en la Ecuación (8.4), calculada para la función de regresión con interac-

TABLA 8.1 La rentabilidad de la educación y la brecha de género: resultados de regresión para los Estados Unidos en 2008				
Regresor	(1)	(2)	(3)	(4)
Años de educación	0,1035** (0,0009)	0,1050** (0,0009)	0,1001** (0,0011)	0,1032** (0,0012)
Mujer		-0,263** (0,004)	-0,432** (0,024)	-0,451** (0,024)
Mujer × Años de educación			0,0121** (0,0017)	0,0134** (0,0017)
Experiencia potencial				0,0143** (0,0012)
Experiencia potencial ²				-0,000211** (0,000023)
Centro-Oeste				-0,095** (0,006)
Sur				-0,092** (0,006)
Oeste				-0,023** (0,007)
Intercepto	1,533** (0,012)	1,629** (0,012)	1,697** (0,016)	1,503** (0,023)
\bar{R}^2	0,208	0,258	0,258	0,267
Los datos son de la Encuesta Actualizada de Población de marzo de 2009 (“Current Population Survey”, véase el Apéndice 3.1). El tamaño de la muestra es $n = 52,790$ observaciones para cada regresión. Mujer es una variable indicador que es igual a 1 para las mujeres y 0 para los hombres. Centro-Oeste, Sur, y Oeste son variables indicadoras que expresan la región de los Estados Unidos en la que residen los trabajadores: por ejemplo, la variable Centro-Oeste es igual a 1 si el trabajador vive en la región centro oeste y es igual a 0 en caso contrario (se omite la región Nordeste). Los errores estándar se presentan entre paréntesis debajo de los coeficientes estimados. Los coeficientes individuales son estadísticamente significativos al 5%* o al 1%** de nivel de significación.				

CONCEPTO CLAVE
8.5

Interacciones en regresión múltiple

El término de interacción entre las dos variables independientes X_1 y X_2 es su producto $X_1 \times X_2$. La inclusión de este término permite que el efecto sobre Y de un cambio en X_1 dependa del valor de X_2 y, de forma inversa, permite que el efecto de un cambio en X_2 pueda depender del valor de X_1 .

El coeficiente de $X_1 \times X_2$ es el efecto de un aumento unitario en X_1 y en X_2 , mucho más allá de la suma de los efectos individuales de un aumento unitario producido solamente en X_1 y un aumento unitario producido solamente en X_2 . Esto es cierto tanto si X_1 y/o X_2 son continuas como si son binarias.

ciones de la Ecuación (8.35), es $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1$ [Ejercicio 8.10(a)]. Por tanto, el efecto sobre Y de un cambio en X_1 , manteniendo constante X_2 , es

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2 \quad (8.36)$$

que depende de X_2 . Por ejemplo, en el caso de los ingresos salariales, si β_3 es positivo, entonces el efecto sobre el logaritmo de los ingresos salariales de un año adicional de experiencia es mayor, en la cantidad β_3 , para cada año adicional de educación que tiene el trabajador.

Un cálculo similar muestra que el efecto sobre Y de un cambio ΔX_2 en X_2 , manteniendo constante X_1 , es $\Delta Y/\Delta X_2 = (\beta_2 + \beta_3 X_1)$.

Colocando juntos estos dos efectos se muestra que el coeficiente β_3 del término de interacción es el efecto de un aumento unitario en X_1 y en X_2 , mucho más allá de los efectos de un cambio unitario solamente en X_1 y un aumento unitario en X_2 en solitario. Es decir, si X_1 cambia en ΔX_1 y X_2 cambia en ΔX_2 , entonces el cambio esperado en Y es $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$ [Ejercicio 8.10(c)]. El primer término es el efecto que proviene del cambio en X_1 manteniendo constante X_2 ; el segundo término es el efecto que proviene del cambio en X_2 manteniendo constante X_1 ; y el último término $\beta_3 \Delta X_1 \Delta X_2$, es el efecto extra del cambio tanto en X_1 como en X_2 .

Las interacciones entre las dos variables están recogidas como Concepto clave 8.5.

Cuando las interacciones se combinan con transformaciones logarítmicas, pueden utilizarse para estimar las elasticidades precio cuando las elasticidades precio dependen de las características del bien (véase como ejemplo el recuadro «La demanda de revistas de economía» de la página siguiente).

Aplicación para la ratio estudiantes-maestros y el porcentaje de estudiantes aprendiendo inglés. Los ejemplos anteriores consideran las interacciones entre la ratio estudiantes-maestros y la variable binaria que indica si el porcentaje de estudiantes aprendiendo inglés es grande o pequeño. Un modo diferente de estudiar esta interacción es examinar la interacción entre la ratio estudiantes-maestros y la variable continua, porcentaje de estudiantes aprendiendo inglés ($PctEI$). La regresión con interacción estimada es

$$\widehat{\text{CalificaciónExamen}} = 686,3 - 1,12REM - 0,67PctEI + 0,0012(REM \times PctEI), \bar{R}^2 = 0,422. \quad (8.37)$$

(11,8)	(0,59)	(0,37)	(0,019)
--------	--------	--------	---------

Cuando el porcentaje de estudiantes aprendiendo inglés se encuentra en la mediana ($PctEI = 8,85$), la pendiente de la recta que relaciona las calificaciones en los exámenes y la ratio estudiantes-maestros se estima que es $-1,11 (= -1,12 + 0,0012 \times 8,85)$. Cuando el porcentaje de estudiantes aprendiendo inglés se encuentra en el percentil 75 ($PctEI = 23,0$), se estima que esta recta es más plana, con una pendiente de $-1,09 (= -1,12 + 0,0012 \times 23,0)$. Es decir, para un distrito con un 8,85 % de estudiantes aprendiendo inglés, el efecto estimado de la reducción en una unidad de la ratio estudiantes-maestros es el de aumentar las calificaciones en los exámenes en 1,11 puntos, pero para un distrito con un 23,0 % de estudiantes aprendiendo inglés, la reducción de la ratio estudiantes-maestros en una unidad se prevé que aumente las calificaciones en los exámenes en solamente 1,09 puntos. Sin embargo, la diferencia entre estos efectos estimados no es estadísticamente significativa: el estadístico t para contrastar si el coeficiente del término de interacción es igual a cero, es $t = 0,0012/0,019 = 0,06$, que no es significativo al nivel del 10 %.

Para mantener centrado el análisis en los modelos no lineales, las especificaciones de las Secciones 8.1 a 8.3 excluyen variables de control adicionales tales como la situación económica de los estudiantes. En consecuencia, estos resultados podría decirse que son objeto de sesgo de variable omitida. Para extraer conclusiones de fondo sobre el efecto en las calificaciones de la reducción de la ratio estudiantes-maestros, estas especificaciones no lineales deben ampliarse con variables de control, y es el ejercicio al que ahora se vuelve el análisis.

8.4 Efectos no lineales sobre las calificaciones de la ratio estudiantes-maestros

Esta sección aborda tres preguntas específicas acerca de las calificaciones en los exámenes y la ratio estudiantes-maestros. En primer lugar, después de tener en cuenta las diferencias en las características económicas de los diferentes distritos, ¿depende el efecto sobre las calificaciones en los exámenes de la reducción de la ratio estudiantes-maestros de la proporción de los estudiantes que todavía están aprendiendo inglés? En segundo lugar, ¿depende este efecto del valor de la ratio estudiantes-maestros? Tercero, y más importante, después de tener en cuenta los factores económicos y las no linealidades, ¿cuál es el efecto estimado sobre las calificaciones de la reducción de la ratio estudiantes-maestros en dos alumnos por maestro, que la directora del Capítulo 4 se propone hacer?

La demanda de revistas de economía

Los economistas profesionales siguen la más reciente investigación en sus áreas de especialización. La mayor parte de la investigación en economía aparece publicada en primer lugar en las revistas de economía, por lo que los economistas —o sus bibliotecas— se suscriben a las revistas de economía.

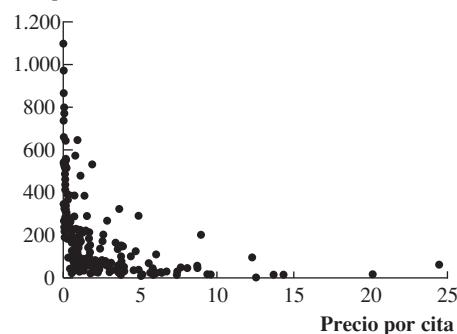
¿Cuál es la elasticidad de la demanda por parte de las bibliotecas para las revistas de economía? Para averiguarlo, se analiza la relación entre el número de suscripciones a una revista de las bibliotecas de EE.UU. (Y_t) y el precio de la suscripción de la biblioteca a la revista a partir de los datos para el año 2000 sobre 180 revistas de economía. Debido a que el producto de una revista no es el papel en el que se imprime, sino más bien las ideas que contiene, su precio lógicamente no se mide en dólares por página sino en dólares por idea. Aunque no se pueden medir las «ideas» directamente, una buena medida indirecta es el número de veces que los artículos de una revista son citados posteriormente por otros investigadores. En consecuencia, se miden los precios como el «precio

por cita» en la revista. El rango de precios es enorme, desde 1/2 centavo por cita (para la *American Economic Review*) hasta 20 centavos por cita o más. Algunas revistas son caras por cita debido a que tienen pocas citas; otras, debido a que el precio de suscripción anual para la biblioteca es muy elevado. En 2010, una suscripción para una biblioteca para la edición impresa del *Journal of Econometrics* costaba 3.264 \$, en comparación con los 455 dólares por una suscripción conjunta a las siete revistas de la American Economics Association que incluye la *American Economic Review*!

Debido a que lo que interesa es la estimación de las elasticidades, se utiliza una especificación log-log (Concepto clave 8.2). Los diagramas de dispersión de las Figuras 8.9(a) y 8.9(b) proporcionan apoyo empírico a esta transformación. Debido a que algunas de las revistas más antiguas y prestigiosas son las más baratas por cita, una regresión del logaritmo de la cantidad sobre el logaritmo del precio podría presentar sesgo de variable omitida. Las regresiones por tanto, incluyen

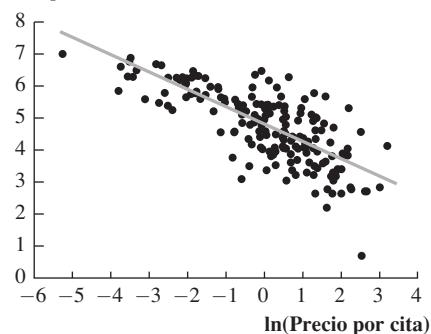
FIGURA 8.9 Suscripciones de bibliotecas y precio de las revistas de economía

Suscripciones



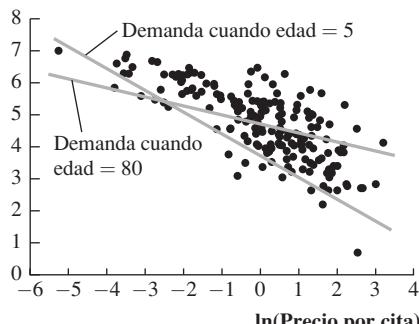
(a) Suscripciones y precio por cita

ln(Suscripciones)



(b) ln(Suscripciones) y ln(Precio por cita)

ln(Suscripciones)



(c) ln(Suscripciones) y ln(Precio por cita)

Existe una relación no lineal inversa entre el número de suscripciones por parte de las bibliotecas de EE.UU. (cantidad) y el precio para la biblioteca por cita (precio), como se muestra en la Figura 8.9a para 180 revistas de economía en 2000. Pero como se aprecia en la Figura 8.9b, la relación entre el logaritmo de la cantidad y el logaritmo del precio parece ser aproximadamente lineal. La Figura 8.9c muestra que la demanda es más elástica para las revistas jóvenes (Edad = 5) que para las revistas antiguas (Edad = 80).

(continúa)

dos variables de control: el logaritmo de la edad y el logaritmo del número de artículos anuales de la revista.

Los resultados de la regresión se resumen en la Tabla 8.2. Los resultados proporcionan las siguientes conclusiones (¡compruebe si puede encontrar la base de estas conclusiones en la Tabla!).

1. La demanda es menos elástica para las revistas más antiguas que para las más nuevas
2. La evidencia apoya una relación lineal, en vez de cúbica, para la función del logaritmo de los precios.
3. La demanda es mayor para las revistas con más artículos, manteniendo los precios y la edad constantes.

Entonces, ¿cuál es la elasticidad de la demanda de las revistas de economía? Depende de la edad de la revista. Las curvas de demanda para una revista de 80 años de antigüedad, y para una recién llegada de 5 años de antigüedad, se encuentran

superpuestas en el diagrama de dispersión de la Figura 8.9(c); la elasticidad de la demanda de la revista más antigua es $-0,28$ ($ES = 0,06$), mientras que para la revista más joven es $-0,67$ ($ES = 0,08$).

Esta demanda es muy inelástica: la demanda es muy poco sensible a los precios, especialmente para las revistas más antiguas. Para las bibliotecas, contar con las investigaciones más recientes es una necesidad, no un lujo. A modo de comparación, los expertos estiman que la elasticidad de la demanda de cigarrillos está en el rango de $-0,3$ a $-0,5$. Las revistas de economía son, al parecer, tan adictivas como el tabaco, ¡pero mucho mejores para la salud!¹.

¹ Estos datos fueron proporcionados amablemente por el profesor Theodore Bergstrom del Departamento de Economía de la Universidad de California, Santa Barbara. Si está interesado en saber más sobre la economía de las revistas de economía, véase Bergstrom (2001).

TABLA 8.2 Estimaciones de la demanda de las revistas de economía

Regresor	(1)	(2)	(3)	(4)
$\ln(\text{Precio por cita})$	$-0,533^{**}$ (0,034)	$-0,408^{**}$ (0,044)	$-0,961^{**}$ (0,160)	$-0,899^{**}$ (0,145)
$[\ln(\text{Precio por cita})]^2$			$0,017$ (0,025)	
$[\ln(\text{Precio por cita})]^3$			$0,0037$ (0,0055)	
$\ln(\text{Edad})$		$0,424^{**}$ (0,119)	$0,373^{**}$ (0,118)	$0,374^{**}$ (0,118)
$\ln(\text{Edad}) \times \ln(\text{Precio por cita})$			$0,156^{**}$ (0,052)	$0,141^{**}$ (0,040)
$\ln(\text{Artículos} \div 1.000.000)$		$0,206^{*}$ (0,098)	$0,235^{*}$ (0,098)	$0,229^{*}$ (0,096)
Intercepto	$4,77^{**}$ (0,055)	$3,21^{**}$ (0,38)	$3,41^{**}$ (0,38)	$3,43^{**}$ (0,38)
Estadísticos F y estadísticos de resumen				
Estadístico F de contraste de los coeficientes de los términos cuadráticos y cúbicos (<i>p</i> -valor)			$0,25$ (0,779)	
ESR	0,750	0,705	0,691	0,688
\bar{R}^2	0,555	0,607	0,622	0,626

El estadístico F contrasta la hipótesis de que los coeficientes de $[\ln(\text{Precio por cita})]^2$ y $[\ln(\text{Precio por cita})]^3$ son ambos iguales a cero. Los errores estándar están entre paréntesis bajo los coeficientes, y los *p*-valores se indican entre paréntesis debajo de los estadísticos F. Los coeficientes individuales son estadísticamente significativos al nivel del 5%* o al nivel del 1%**.

Se responde a estas preguntas considerando especificaciones de regresión no lineales del tipo descrito en las Secciones 8.2 y 8.3, ampliadas para incluir dos medidas sobre la situación económica de los estudiantes: el porcentaje de estudiantes que cumplen los requisitos para percibir un subsidio de comedor y el logaritmo de la renta promedio del distrito. El logaritmo de la renta se utiliza debido a que el análisis empírico de la

Sección 8.2 sugiere que esta especificación capta la relación no lineal entre las calificaciones en los exámenes y la renta. Como en la Sección 7.6, no se incluyen los gastos por alumno como regresor, y al no hacerlo, se considera el efecto de la disminución de la ratio estudiantes-maestros, permitiendo que aumenten los gastos por alumno (es decir, no se mantienen constantes los gastos por alumno).

Discusión de los resultados de la regresión

Los resultados de la regresión MCO se resumen en la Tabla 8.3. Las columnas etiquetadas de (1) a (7) presentan por separado las regresiones. Las entradas de la tabla son los coeficientes, los errores estándar,

TABLA 8.3 Modelos de regresión no lineales para las calificaciones en los exámenes

Regresor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ratio estudiantes-maestros (<i>REM</i>)	-1,00** (0,27)	-0,73** (0,26)	-0,97 (0,59)	-0,53 (0,34)	64,33** (24,86)	83,70** (28,50)	65,29** (25,26)
<i>REM</i> ²					-3,42** (1,25)	-4,38** (1,44)	-3,47** (1,27)
<i>REM</i> ³					0,059** (0,021)	0,075** (0,024)	0,060** (0,021)
% aprendiendo inglés	-0,122** (0,033)	-0,176** (0,034)					-0,166** (0,034)
% aprendiendo inglés ≥ 10%? (<i>EIA</i> , Binaria)			5,64 (19,51)	5,50 (9,80)	-5,47** (1,03)	816,1* (327,7)	
<i>EIA</i> × <i>REM</i>			-1,28 (0,97)	-0,58 (0,50)		-123,3* (50,2)	
<i>EIA</i> × <i>REM</i> ²					6,12* (2,54)		
<i>EIA</i> × <i>REM</i> ³					-0,101* (0,043)		
% Cumplimiento requisitos comedor subvencionado	-0,547** (0,024)	-0,398** (0,033)		-0,411** (0,029)	-0,420** (0,029)	-0,418** (0,029)	-0,402** (0,033)
Renta media distrito (logaritmo)		11,57** (1,81)		12,12** (1,80)	11,75** (1,78)	11,80** (1,78)	11,51** (1,81)
Intercepto	700,2** (5,6)	658,6** (8,6)	682,2** (11,9)	653,6** (9,9)	252,0 (163,6)	122,3 (185,5)	244,8 (165,7)
Estadísticos <i>F</i> y <i>p</i>-valores para hipótesis conjuntas							
(a) Todas las variables <i>REM</i> e interacciones = 0			5,64 (0,004)	5,92 (0,003)	6,31 (< 0,001)	4,96 (< 0,001)	5,91 (0,001)
(b) <i>REM</i> ² , <i>REM</i> ³ = 0					6,17 (< 0,001)	5,81 (0,003)	5,96 (0,003)
(c) <i>EIA</i> × <i>REM</i> , <i>EIA</i> × <i>REM</i> ² , <i>EIA</i> × <i>REM</i> ³ = 0						2,69 (0,046)	
<i>ESR</i>	9,08	8,64	15,88	8,63	8,56	8,55	8,57
<i>R</i> ²	0,773	0,794	0,305	0,795	0,798	0,799	0,798

Estas regresiones se estimaron utilizando los datos sobre los distritos escolares K-8 de California, descritos en el Apéndice 4.1. Los errores estándar se presentan entre paréntesis debajo de los coeficientes, y los *p*-valores se presentan entre paréntesis debajo de los estadísticos *F*. Los coeficientes individuales son estadísticamente significativos al 5%* o al 1%** de nivel de significación.

algunos estadísticos F y sus p -valores, y los estadísticos de resumen, tal como se indica en la descripción de cada fila.

La primera columna de los resultados de regresión, etiquetada como regresión (1) en la tabla, es la regresión (3) de la Tabla 7.1 repetida aquí por conveniencia. Esta regresión no controla por la renta, por lo que lo primero que se hace es comprobar si los resultados cambian sustancialmente cuando se incluye el logaritmo de la renta como una variable económica de control adicional. Los resultados se ofrecen en la regresión (2) de la Tabla 8.3. El logaritmo de la renta es estadísticamente significativo al nivel del 1 % y el coeficiente de la ratio estudiantes-maestros se convierte en algo más cercano a cero, pasando desde $-1,00$ hasta $-0,73$, aunque sigue siendo estadísticamente significativo al nivel del 1 %. La variación en el coeficiente de REM es lo suficientemente grande entre las regresiones (1) y (2) como para justificar la inclusión del logaritmo de la renta en las regresiones restantes como elemento que evite el sesgo de variable omitida.

La regresión (3) de la Tabla 8.3 es la regresión con interacciones de la Ecuación (8.34) con la variable binaria de porcentaje alto o bajo de estudiantes aprendiendo inglés, pero sin variables económicas de control. Cuando se añaden las variables económicas de control (porcentaje de cumplimiento de los requisitos para el comedor subsidiado y logaritmo de la renta) [regresión (4) de la tabla], los coeficientes cambian, pero en ningún caso el coeficiente del término de interacción es significativo al nivel del 5 %. Sobre la base de la evidencia en la regresión (4), la hipótesis de que el efecto de REM es el mismo para los distritos con porcentajes de alumnos aprendiendo inglés bajos y altos no puede rechazarse al nivel del 5 % (el estadístico t es $t = -0,58/0,50 = -1,16$).

La regresión (5) examina si el efecto de la variación en la ratio estudiantes-maestros depende del valor de la ratio estudiantes-maestros mediante la inclusión de una especificación cúbica en REM, además de las otras variables de control de la regresión (4) [el término de interacción, $EIA \times REM$ se eliminó porque no era significativo en la regresión (4) al 10 % de nivel]. Las estimaciones de la regresión (5) son congruentes con el hecho de que la proporción de alumnos por maestro tenga un efecto no lineal. La hipótesis nula de que la relación es lineal se rechaza al nivel de significación del 1 % frente a la alternativa de que es cúbica (el estadístico F para el contraste de la hipótesis de que los verdaderos coeficientes de REM^2 y REM^3 son iguales a cero es 6,17, con un p -valor de $<0,001$).

La regresión (6) además, examina si el efecto de la ratio estudiantes-maestros depende no solo del valor de la ratio estudiantes-maestros, sino además de la proporción de estudiantes aprendiendo inglés. Mediante la inclusión de las interacciones entre EIA y REM , REM^2 y REM^3 , se puede comprobar si las (posiblemente cúbicas) funciones de regresión poblacionales que relacionan las calificaciones en los exámenes y REM son diferentes para los porcentajes altos y bajos de estudiantes aprendiendo inglés. Para ello, se contrasta la restricción de que los coeficientes de los tres términos de interacción son cero. El estadístico F resultante es 2,69, que tiene un p -valor de 0,046 y por lo tanto es significativo al 5 %, pero no al 1 % de nivel. Esto proporciona alguna evidencia de que las funciones de regresión son diferentes para los distritos con porcentajes de alumnos aprendiendo inglés altos y bajos; sin embargo, la comparación de las regresiones (6) y (4) deja claro que estas diferencias están asociadas con los términos cuadráticos y cúbicos.

La regresión (7) es una modificación de la regresión (5), en la que se utiliza la variable continua $PctEI$ en lugar de la variable binaria EIA para controlar por el porcentaje alumnos aprendiendo inglés en el distrito. Los coeficientes de los otros regresores no cambian sustancialmente cuando se realiza esta modificación, lo que indica que los resultados de la regresión (5) no son sensibles a la medida del porcentaje de estudiantes aprendiendo inglés que se utilice en la regresión.

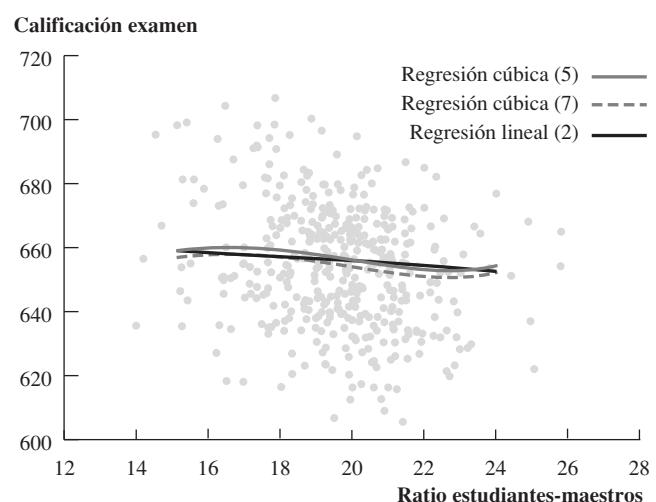
En todas las especificaciones, la hipótesis de que la ratio estudiantes-maestros no entra en la regresión se rechaza al 1 % de nivel.

Las especificaciones no lineales de la Tabla 8.3 se interpretan en su mayoría fácilmente de forma gráfica. La Figura 8.10 representa gráficamente las funciones de regresión estimadas que relacionan las calificaciones y la ratio estudiantes-maestros para la especificación lineal (2) y la especificación cúbica (5) y (7), junto con un diagrama de dispersión de los datos⁴. Estas funciones de regresión estimadas muestran los

⁴ Para cada curva, el valor de predicción se calculó igualando cada variable independiente, distinta a REM , al valor de su media muestral y calculando el valor de predicción multiplicando esos valores fijados de las variables independientes por sus respectivos coeficientes, estimados a partir de la Tabla 8.3. Esto se realizó para distintos valores de REM , y el gráfico de los valores de predicción ajustados resultantes es la función de regresión estimada que relaciona las calificaciones en los exámenes y la variable REM , manteniendo constantes las otras variables en sus medias muestrales.

FIGURA 8.10 Tres funciones de regresión que relacionan las calificaciones y la ratio estudiantes-maestros

La regresión cúbica de las columnas (5) y (7) de la Tabla 8.3 son prácticamente idénticas. Indican una pequeña cantidad de no linealidad en la relación entre las calificaciones y la ratio estudiantes-maestros



valores de predicción de las calificaciones en los exámenes como una función de la ratio estudiantes-maestros, manteniendo fijos otros valores de las variables independientes de la regresión. Las funciones de regresión estimadas están cercanas unas de otras, aunque la regresión cúbica se aplana para valores altos de la ratio estudiantes-maestros.

La regresión (6) indica una diferencia estadísticamente significativa en las funciones de regresión cúbicas que relacionan las calificaciones y *REM*, dependiendo de si el porcentaje de estudiantes que están aprendiendo inglés en el distrito es alto o bajo. La Figura 8.11 muestra gráficamente estas dos funciones de regresión estimadas por lo que puede apreciarse si esta diferencia aparte de ser estadísticamente significativa, tiene importancia en sentido práctico. Como muestra la Figura 8.11, para los valores de la ratio estudiantes-maestros entre 17 y 23 —un rango que incluye el 88 % de las observaciones—, las dos funciones están separadas por aproximadamente diez puntos, pero por lo demás son bastante similares; es decir, para los valores de la variable *REM* entre 17 y 23, los distritos con un menor porcentaje de estudiantes aprendiendo inglés lo hacen mejor, manteniendo constante la ratio estudiantes-maestros, pero el efecto de una variación en la ratio estudiantes-maestros es esencialmente el mismo para ambos grupos. Las dos funciones de regresión son distintas para valores de la variable ratio estudiantes-maestros por debajo de 16,5, pero debe tenerse cuidado en no leer más en esto de lo que no esté justificado. Los distritos con valores de la variable $REM < 16,5$ constituyen solamente el 6 % de las observaciones, por lo que las diferencias entre las funciones de regresión no lineales están reflejando las diferencias en esos muy pocos distritos con ratios estudiantes-maestros muy bajas. Por tanto, en base a la Figura 8.11, se puede concluir que el efecto sobre las calificaciones en los exámenes de una variación en la ratio estudiantes-maestros no depende del porcentaje de estudiantes que todavía están aprendiendo inglés para el rango de ratios estudiantes-maestros para los que se dispone de la mayoría de datos.

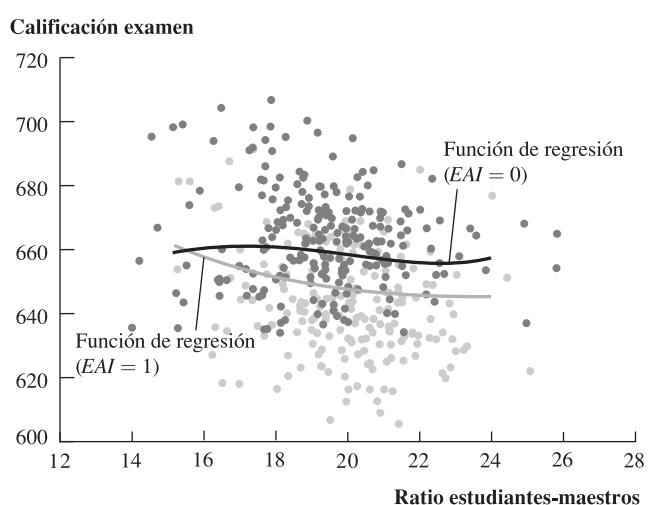
Resumen de resultados

Esos resultados permiten responder a las tres preguntas que se planteaban al principio de esta sección.

En primer lugar, tras tener en cuenta las circunstancias económicas, el hecho de si existen muchos o pocos estudiantes aprendiendo inglés en el distrito no ejerce una influencia sustancial en el efecto sobre las calificaciones de una variación en la ratio estudiantes-maestros. En las especificaciones lineales, no existe evidencia estadísticamente significativa de esta diferencia. La especificación cúbica de la regresión (6) proporciona evidencia estadísticamente significativa (al 5 % de nivel) de que las funciones de regresión son

FIGURA 8.11 Funciones de regresión para los distritos con porcentajes de estudiantes de inglés alto y bajo

Los distritos con porcentajes bajos de estudiantes aprendiendo inglés ($EIA = 0$) están representados mediante puntos gris oscuro y los distritos con $EIA = 1$ se representan mediante los puntos gris claro. La función de regresión cúbica para $EIA = 1$ de la regresión (6) de la Tabla 8.3 se encuentra aproximadamente 10 puntos por debajo de la función de regresión cúbica para $EIA = 0$ para $17 \leq REM \leq 23$, pero por otro lado las dos pendientes tienen forma y pendiente similares en ese rango. Las pendientes de las funciones de regresión difieren en gran medida para valores muy altos y bajos de la variable REM , para los cuales existen pocas observaciones.



diferentes para los distritos con porcentajes de estudiantes aprendiendo inglés altos y bajos; no obstante, como muestra la Figura 8.11, las funciones de regresión estimadas presentan pendientes similares en el rango de la variable ratio estudiantes-maestros para la que se dispone de la mayor parte de los datos.

En segundo lugar, tras controlar por las circunstancias económicas, existe evidencia de un efecto no lineal sobre las calificaciones en los exámenes de la ratio estudiantes-maestros. Este efecto es estadísticamente significativo al 1 % de nivel (los coeficientes de las variables REM^2 y REM^3 son siempre significativos al 1 % de nivel).

En tercer lugar, podemos volver la vista al problema de la directora que abría el Capítulo 4. Ella quería saber el efecto sobre las calificaciones de la reducción de la ratio estudiantes-maestros en dos estudiantes por maestro. En la especificación lineal (2) este efecto no depende de la ratio estudiantes-maestros en sí mismo, y el efecto estimado de esta reducción es la mejora de las calificaciones en 1,46 ($= -0,73 - 2$) puntos. En las especificaciones no lineales, este efecto depende del valor de la ratio estudiantes-maestros. Si su distrito en realidad presenta una ratio estudiantes-maestros de 20 y ella contempla recortarla hasta 18, entonces en base a la regresión (5) el efecto estimado de esta reducción es el aumento de las calificaciones de los exámenes en 3,00 puntos, mientras en base a la regresión (7) esta estimación es 2,93. Si su distrito presenta realmente una ratio de estudiantes-maestros de 22 y ella está considerando reducirla hasta 20, entonces en base a la regresión (5) el efecto estimado de esta reducción es la mejora en las calificaciones en 1,93 puntos, mientras que en base a la regresión (7) la estimación es de 1,90. La estimaciones que provienen de las especificaciones no lineales sugieren que el recorte en la ratio estudiantes-maestros presenta un mayor efecto si esta ratio ya es pequeña.

8.5 Conclusión

En este capítulo se han presentado algunos métodos para modelizar las funciones de regresión no lineales. Debido a que estos modelos son variantes del modelo de regresión múltiple, los coeficientes desconocidos pueden estimarse por MCO, y las hipótesis acerca de sus valores pueden contrastarse utilizando los estadísticos t y F , como se describió en el Capítulo 7. En estos modelos, el efecto esperado sobre Y de un cambio en una de las variables independientes, X_1 , manteniendo constantes las otras variables independientes X_2, \dots, X_k , depende en general de los valores de X_1, X_2, \dots, X_k .

Existen muchos modelos diferentes en este capítulo, y no puede culparse a nadie de quedarse un poco desconcertado acerca de cuál usar en un caso concreto. ¿Cómo podrían analizarse las posibles no linealida-

des en la práctica? La Sección 8.1 proporcionaba un método general para un análisis de este tipo, pero este método requiere tomar decisiones y realizar valoraciones durante el proceso. Sería conveniente que hubiera una única receta que pudiera seguirse y que funcionara siempre, en todos los casos, pero en la práctica el análisis de datos resulta rara vez sencillo.

El paso sencillo y más importante en la especificación de funciones lineales es «utilizar la cabeza». Antes de mirar los datos, ¿puede imaginarse alguna razón, basada en la teoría económica o el juicio experto, por la cual la pendiente de la regresión poblacional podría depender del valor de esta, u otra, variable independiente? Si es así, ¿qué tipo de dependencia podría esperarse? Y, más importante, ¿qué no linealidades (si hubiera) podrían tener implicaciones importantes para los principales aspectos abordados en el estudio? Responder a estas preguntas cuidadosamente centrará el análisis. En el caso de las calificaciones en los exámenes, por ejemplo, ese razonamiento lleva a investigar si contratar más profesores podría tener un mayor efecto en los distritos con un mayor porcentaje de estudiantes que todavía están aprendiendo inglés, quizás porque estos estudiantes podrían beneficiarse de forma diferenciada de una atención más personalizada. Precisando esta cuestión, hemos sido capaces de encontrar una respuesta precisa: tras tener en cuenta las circunstancias económicas de los estudiantes, no encontramos evidencia estadísticamente significativa de esta interacción.

Resumen

1. En una regresión no lineal, la pendiente de la función de regresión poblacional depende del valor de una o más variables independientes.
2. El efecto sobre Y de un cambio en la(s) variable(s) independiente(s) puede calcularse evaluando la función de regresión para dos valores de la(s) variable(s) independiente(s). El procedimiento se resume en el Concepto clave 8.1.
3. Una regresión polinomial incluye las potencias de X como regresores. Una regresión cuadrática incluye X y X^2 , y una regresión cúbica incluye X , X^2 y X^3 .
4. Las variaciones pequeñas en logaritmos pueden interpretarse como variaciones porcentuales o proporcionales en una variable. Las regresiones que incluyen logaritmos se utilizan para estimar variaciones proporcionales y elasticidades.
5. El producto de dos variables se denomina término de interacción. Cuando los términos de interacción se incluyen como regresores, permiten que la pendiente de regresión de una variable dependa del valor de otra variable.

Términos clave

modelo de regresión cuadrático (183)
 función de regresión no lineal (185)
 modelo de regresión polinomial (188)
 modelo de regresión cúbica (188)
 elasticidad (190)
 función exponencial (190)
 logaritmo natural (190)
 modelo lineal-log (191)

modelo log-lineal (192)
 modelo log-log (193)
 término de interacción (197)
 regresor de interacción (197)
 modelo de regresión con interacción (197)
 mínimos cuadrados no lineales (220)
 estimadores de mínimos cuadrados no lineales (220)

Revisión de conceptos

- 8.1** Represente una función de regresión que sea creciente (tenga pendiente positiva) y sea inclinada para valores pequeños de X , pero menos inclinada para valores altos de X . Explique cómo se especificaría una regresión no lineal para modelizar esta forma. ¿Puede imaginar una relación económica con una forma semejante?

- 8.2.** Una función de producción «Coob-Douglas» relaciona la producción (Q) con los factores de producción, capital (K), trabajo (L), y materias primas (M), y se utiliza un término de error u en la ecuación $Q = \lambda K^{\beta_1} L^{\beta_2} M^{\beta_3} e^u$, donde, λ , β_1 , β_2 y β_3 son los parámetros de producción. Supóngase que se dispone de datos sobre la producción y los factores de producción a partir de una muestra aleatoria de empresas con la misma función de producción Coob-Douglas. ¿Cómo utilizaría el análisis de regresión para estimar los parámetros de producción?
- 8.3.** Una función de «demanda de dinero» estándar como las que utilizan los macroeconomistas tiene la forma $\ln(m) = \beta_0 + \beta_1 \ln(PIB) + \beta_2 R$, donde m es la cantidad de saldos reales, PIB es el valor del producto interior bruto (real), y R es el tipo de interés nominal medido en porcentaje anual. Supóngase que $\beta_1 = 1,0$ y $\beta_2 = -0,02$. ¿Qué pasará con el valor de m si el PIB aumenta en un 2 %? ¿Qué pasará con m si el tipo de interés aumenta del 4 % al 5 %?
- 8.4.** Se ha estimado un modelo de regresión lineal que relaciona Y con X . El profesor dice: «Creo que la relación entre Y y X es no lineal». Explique cómo se contrastaría la pertinencia de la regresión lineal.
- 8.5.** Supóngase que en el Ejercicio 8.2 se pensó que el valor de β_2 no era constante, sino que aumentaba cuando K aumentaba. ¿Cómo se utilizaría un término de interacción para captar este efecto?

Ejercicios

- 8.1** Las ventas de una empresa son de 196 millones de dólares en 2009 y aumentaron hasta los 198 millones de dólares en 2010.
- Calcule el porcentaje de aumento en las ventas mediante la fórmula habitual $100 \times \frac{(Ventas_{2010} - Ventas_{2009})}{Ventas_{2009}}$. Compare este valor con la aproximación $100 \times [\ln(Ventas_{2010}) - \ln(Ventas_{2009})]$.
 - Repita (a) suponiendo que $Ventas_{2010} = 205$, $Ventas_{2010} = 250$ y $Ventas_{2010} = 500$.
 - ¿Cuál es la bondad de la aproximación, cuando la variación es pequeña? ¿La calidad de la aproximación se deteriora cuando aumenta el cambio porcentual?
- 8.2.** Supóngase que un investigador recoge datos sobre las viviendas que se han vendido en un determinado vecindario durante el año pasado y obtiene los resultados de regresión de la tabla que aparece en la página siguiente.
- Utilizando los resultados de la columna (1), ¿cuál es la variación esperada en el precio de una vivienda si se construye un anexo a ella de 500 pies cuadrados? Construya un intervalo de confianza al 95 % para la variación porcentual en el precio.
 - Comparando las columnas (1) y (2), ¿es mejor utilizar la variable *Tamaño* o la variable $\ln(Tamaño)$ para explicar el precio de la vivienda?
 - Utilizando la columna (2), ¿cuál es el efecto estimado de tener piscina sobre el precio? (Asegúrese de obtener las unidades de forma correcta). Construya un intervalo de confianza al 95 % para este efecto.
 - La regresión de la columna (3) añade la variable de número de dormitorios a la regresión. ¿Cuál es la cuantía del efecto estimado de tener un dormitorio adicional? ¿Es el efecto estadísticamente significativo? ¿Por qué cree que el efecto estimado es tan pequeño? (Pista: ¿qué otras variables se están manteniendo constantes?)
 - ¿Es el término cuadrático $\ln(Tamaño)^2$ importante?
 - Utilice la regresión de la columna (5) para calcular la variación esperada en el precio cuando se añade una piscina a una casa sin vista. Repita el ejercicio para una vivienda con una buena vista. ¿Existe una gran diferencia? ¿Es la diferencia estadísticamente significativa?

Resultados de regresión para el Ejercicio 8.2					
Variable dependiente: $\ln(\text{Precio})$	(1)	(2)	(3)	(4)	(5)
<i>Tamaño</i>	0,00042 (0,000038)				
$\ln(\text{Tamaño})$		0,69 (0,054)	0,68 (0,087)	0,57 (2,03)	0,69 (0,055)
$\ln(\text{Tamaño})^2$				0,0078 (0,14)	
<i>Dormitorios</i>			0,0036 (0,037)		
<i>Piscina</i>	0,082 (0,032)	0,071 (0,034)	0,071 (0,034)	0,071 (0,036)	0,071 (0,035)
<i>Vista</i>	0,037 (0,029)	0,027 (0,028)	0,026 (0,026)	0,027 (0,029)	0,027 (0,030)
<i>Piscina</i> × <i>Vista</i>					0,0022 (0,10)
<i>Estado</i>	0,13 (0,045)	0,12 (0,035)	0,12 (0,035)	0,12 (0,036)	0,12 (0,035)
Intercepto	10,97 (0,069)	6,60 (0,39)	6,63 (0,53)	7,02 (7,50)	6,60 (0,40)
Estadísticos de resumen					
<i>ESR</i>	0,102	0,098	0,099	0,099	0,099
<i>R</i> ²	0,72	0,74	0,73	0,73	0,73
Definición de las variables: <i>Precio</i> = precio de venta (\$); <i>Tamaño</i> = tamaño de la vivienda (en pies cuadrados); <i>Dormitorios</i> = número de habitaciones; <i>Piscina</i> = variable binaria (1 si la casa tiene piscina, 0 en caso contrario); <i>Vista</i> = variable binaria (1 si la casa tiene una bonita vista, 0 en caso contrario); <i>Estado</i> = variable binaria (1 si el agente inmobiliario informa de que la casa está en excelentes condiciones, 0 en caso contrario)					

- 8.3. Tras leer el análisis de este capítulo sobre las calificaciones en los exámenes y el tamaño de las clases, un educador comenta: «en mi experiencia, el desempeño de los estudiantes depende del tamaño de la clase, pero no del modo que expresan estas regresiones. Por el contrario, los estudiantes lo hacen bien si el tamaño de las clases es de menos de 20 estudiantes y les va muy mal cuando el tamaño de la clase es mayor que 25. No existen ventajas debidas a la reducción del tamaño de las clases por debajo de 20 estudiantes, la relación es constante en la región intermedia entre 20 y 25 estudiantes, y no hay pérdida al aumentar el tamaño de las clases si es ya de más de 25 estudiantes». El educador describe un «efecto umbral», en el que el rendimiento es constante para los tamaños de clase inferiores a 20, luego salta y es constante para los tamaños de clase entre 20 y 25, y luego salta de nuevo para los tamaños de clase superiores a 25. Para modelizar estos efectos umbral, se definen las variables binarias

$$\text{REMpequeño} = 1 \text{ si } \text{REM} < 20 \text{ y } \text{REMpequeño} = 0 \text{ en otro caso;}$$

$$\text{REMmediano} = 1 \text{ si } 20 \leq \text{REM} \leq 25 \text{ y } \text{REMmediano} = 0 \text{ en otro caso; y}$$

$$\text{REMgrande} = 1 \text{ si } \text{REM} > 25 \text{ y } \text{REMgrande} = 0 \text{ en otro caso.}$$

- a) Considérese la regresión. $\text{CalificaciónExamen}_i = \beta_0 + \beta_1 \text{REMpequeño}_i + \beta_2 \text{REMgrande}_i + u_i$. Represente la función de regresión que relaciona la variable *CalificaciónExamen* con *REM* para los valores hipotéticos de los coeficientes de regresión que son compatibles con la afirmación del educador.

- b)** Un investigador intenta estimar la regresión $CalificaciónExamen_i = \beta_0 + \beta_1 REMpequeño_i + \beta_2 REMmediano_i + \beta_3 REMgrande_i + u_i$ y se encuentra con que su ordenador falla. ¿Por qué?

8.4. Lea el recuadro «La rentabilidad de la educación y la brecha de género» de la Sección 8.3.

- a)** Considérese un hombre con 16 años de educación y 2 años de experiencia, que proviene de un estado occidental. Utilice los resultados de la columna (4) de la Tabla 8.1 y el método del Concepto clave 8.1 para estimar la variación esperada en el logaritmo de los ingresos medios por hora (IMH) asociada a un año adicional de experiencia.
- b)** Repita (a) suponiendo 10 años de experiencia.
- c)** Explique por qué las respuestas a (a) y a (b) son diferentes.
- d)** ¿Es estadísticamente significativa la diferencia en las respuestas a (a) y a (b) al nivel del 5 %? Explíquelo.
- e)** ¿Cambiarían las respuestas desde (a) hasta (d) si la persona fuera una mujer? ¿Y si la persona fuera de la región sur? Explíquelo.
- f)** ¿Cómo cambiaría la regresión si se sospechara que el efecto de la experiencia sobre los ingresos salariales fuera diferente para los hombres y para las mujeres?

8.5. Lea el recuadro «La demanda de revistas de economía» de la sección 8.3.

- a)** El recuadro llega a tres conclusiones. Mirando los resultados de la tabla, ¿cuál es la base para cada una de estas conclusiones?
- b)** Utilizando los resultados de la regresión (4), el recuadro muestra que la elasticidad de la demanda para una revista de 80 años de antigüedad es $-0,28$.
- I) ¿Cómo se determina este valor a partir de la regresión estimada?
 - II) El recuadro muestra que el error estándar de la elasticidad estimada es 0,06. ¿Cómo se calcula este error estándar? (Sugerencia: véase el apartado «Errores estándar de los efectos estimados» de la página 186.)
- c)** Suponga que la variable *Artículos* hubiera sido dividida por 1.000 en lugar de por 1.000.000. ¿Cómo cambiarían los resultados de la columna (4)?

8.6. Con respecto a la Tabla 8.3.

- a)** Un investigador sospecha que el efecto del % de cumplimiento de requisitos para el subsidio de comedor tiene un efecto no lineal sobre las calificaciones obtenidas. En particular, piensa que el aumento de esta variable de un 10 % a un 20 % tiene efecto escaso sobre las calificaciones obtenidas, pero una variación desde un 50 % hasta un 60 % tiene un efecto mucho mayor.
- I) Describa una especificación no lineal que pueda ser utilizada para modelizar esta forma de no linealidad.
 - II) ¿Cómo se podría contrastar si la creencia del investigador es mejor que la especificación lineal de la columna (7) de la Tabla 8.3?
- b)** Un investigador sospecha que el efecto de la renta sobre las calificaciones en los exámenes en la prueba es diferente en los distritos con clases pequeñas que en los distritos con clases grandes.
- I) Describa una especificación no lineal que pueda utilizarse para modelizar esta forma de no linealidad.
 - II) ¿Cómo contrastaría si la creencia del investigador es mejor que la especificación lineal de la columna (7) en la Tabla 8.3?

8.7. Este problema se inspira en un estudio sobre la «brecha de género» en los ingresos salariales de la parte más alta de los puestos de trabajo corporativos [Bertrand y Hallock (2001)]. El estudio compara las retribuciones totales de los altos ejecutivos para un gran número de sociedades anónimas de EE.UU. en la década de 1990. (Cada año, estas sociedades anónimas deben informar de los niveles de las retribuciones totales de sus cinco principales ejecutivos).

- a) Sea *Femenino* una variable indicador que es igual a 1 para las mujeres y 0 para los hombres. Una regresión del logaritmo de los ingresos salariales sobre la variable *Femenino* arroja estos resultados

$$\widehat{\ln(\text{Ingresos})} = 6,48 - 0,44\text{Femenino}, ESR = 2,65. \\ (0,01) \quad (0,05)$$

- I) El coeficiente estimado para Femenino es $-0,44$. Explique qué significa este valor.
 - II) El ESR es $2,65$. Explique qué significa este valor
 - III) ¿Esta regresión sugiere que las mujeres que ocupan puestos de alta dirección ganan menos que los altos directivos hombres? Explíquelo.
 - IV) ¿Sugiere esta regresión que existe discriminación de género? Explíquelo.
- b) Se añaden dos nuevas variables a la regresión, el valor de mercado de la empresa (una medida del tamaño de la empresa, en millones de dólares) y la rentabilidad de las acciones (una medida del funcionamiento de la empresa, en puntos porcentuales):

$$\widehat{\ln(\text{Ingresos})} = 3,86 - 0,28\text{Femenino} + 0,37 \ln(\text{ValorMercado}) + 0,004\text{Rentabilidad}, \\ (0,03) \quad (0,04) \quad (0,004) \quad (0,003) \\ n = 46,670, R^2 = 0,345.$$

- I) El coeficiente de $\ln(\text{ValorMercado})$ es $0,37$. Explique cuál es el significado de este valor.
- II) El coeficiente de la variable *Femenino* es ahora $-0,28$. Explique por qué ha cambiado respecto de la regresión de (a).
- c) ¿Son las grandes empresas más propensas a tener altas ejecutivas mujeres que las empresas pequeñas? Explíquelo.

- 8.8.** X es una variable continua que toma valores entre 5 y 100. Z es una variable binaria. Represente las funciones de regresión siguientes (con valores de X entre 5 y 100 en el eje horizontal y los valores de \hat{Y} en el eje vertical):

- a) $\hat{Y} = 2,0 + 3,0 \times \ln(X)$.
- b) $\hat{Y} = 2,0 - 3,0 \times \ln(X)$.
- c) I) $\hat{Y} = 2,0 + 3,0 \times \ln(X) + 4,0Z$, con $Z = 1$.
II) Igual que (I), pero con $Z = 0$.
- d) I) $\hat{Y} = 2,0 + 3,0 \times \ln(X) + 4,0Z - 1,0 \times Z \times \ln(X)$, con $Z = 1$.
II) Igual que (I), pero con $Z = 0$.
- e) $\hat{Y} = 1,0 + 125,0X - 0,01X^2$.

- 8.9.** Explique cómo se utilizaría el «Método #2» de la Sección 7.3 para calcular el intervalo de confianza tratado a continuación en la Ecuación (8.8). [Sugerencia: para ello es necesario estimar una nueva regresión con una definición diferente de los regresores y de la variable dependiente. Véase el Ejercicio (7.9)].

- 8.10.** Considérese el modelo de regresión $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$. Utilice el Concepto clave 8.1 para demostrar:
- a) $\Delta Y / \Delta X_1 = \beta_1 + \beta_3 X_2$ (efecto de la variación en X_1 manteniendo constante X_2).
 - b) $\Delta Y / \Delta X_2 = \beta_2 + \beta_3 X_1$ (efecto de la variación en X_2 manteniendo constante X_1).
 - c) Si X_1 cambia en ΔX_1 y X_2 cambia en ΔX_2 , entonces $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$.

- 8.11.** Deduzca las expresiones para las elasticidades obtenidas en el Apéndice 8.2 para los modelos lineal y log-log. (Sugerencia: para el modelo log-log suponer que u y X son independientes, como se hacía en el Apéndice 8.2 para el modelo log-lineal).

- 8.12.** El análisis de la Ecuación (8.28) interpreta el coeficiente de interacción de variables binarias utilizando el supuesto de media condicional igual a cero. Este ejercicio demuestra que la interpretación es aplicable asimismo bajo independencia en media condicional. Considérese el experimento hipotético del Ejercicio 7.11.
- Supóngase que se estima la regresión $Y_i = \gamma_0 + \gamma_1 X_{1i} + u_i$ utilizando solamente los datos de los estudiantes que regresan. Demuestre que γ_1 es el efecto del tamaño de la clase para los estudiantes que regresan, es decir, que $\gamma_1 = E(Y_i|X_{1i} = 1, X_{2i} = 0) - E(Y_i|X_{1i} = 0, X_{2i} = 0)$. Explique por qué $\hat{\gamma}_1$ es un estimador insesgado de γ_1 .
 - Supóngase que se estima la regresión $Y_i = \delta_0 + \delta_1 X_{1i} + u_i$, utilizando solamente los datos de los nuevos estudiantes. Demuestre que δ_1 es el efecto del tamaño de las clases para los nuevos estudiantes, es decir, que $\delta_1 = E(Y_i|X_{1i} = 1, X_{2i} = 1) - E(Y_i|X_{1i} = 0, X_{2i} = 1)$. Explique por qué $\hat{\delta}_1$ es un estimador insesgado de δ_1 .
 - Considérese la regresión tanto para los estudiantes que regresan como para los nuevos, $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3(X_{1i} \times X_{2i}) + u_i$. Utilice el supuesto de independencia en media condicional, $E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i})$ para demostrar que $\beta_1 = \gamma_1$, $\beta_1 + \beta_3 = \delta_1$ y $\beta_3 = \delta_1 - \gamma_1$ (la diferencia en los efectos del tamaño de las clases).
 - Supóngase que se estima la regresión con interacciones del apartado (c) utilizando los datos combinados y que $E(u_i|X_{1i}, X_{2i}) = E(u_i|X_{2i})$. Demuestre que $\hat{\beta}_1$ y $\hat{\beta}_3$ son insesgados, pero que $\hat{\beta}_2$ es en general sesgado.

Ejercicios empíricos

- E8.1** Utilícese la base de datos **CPS08** descrita en el Ejercicio Empírico 4.1 para responder a las siguientes cuestiones.

- Realice una regresión de los ingresos medios por hora (variable *AHE* en la base de datos) sobre la edad (variable *Age*), el género (variable *Female*), y la educación (variable *Bachelor*). Si la variable *Age* aumenta de 25 a 26, ¿cuánto se espera que cambien los ingresos? Si la variable *Age* aumenta desde 33 hasta 34, ¿cuánto se espera que cambien los ingresos?
- Realice una regresión del logaritmo de los ingresos medios por hora, $\ln(AHE)$, sobre las variables *Age*, *Female* y *Bachelor*. Si la variable *Age* aumenta de 25 a 26, ¿cuánto se espera que cambien los ingresos? Si la variable *Age* aumenta desde 33 hasta 34, ¿cuánto se espera que cambien los ingresos?
- Realice una regresión del logaritmo de los ingresos medios por hora, $\ln(AHE)$, sobre $\ln(\text{Age})$, *Female* y *Bachelor*. Si la variable *Age* aumenta de 25 a 26, ¿cuánto se espera que cambien los ingresos? Si *Age* aumenta desde 33 hasta 34, ¿cuánto se espera que cambien los ingresos?
- Realice una regresión del logaritmo de los ingresos medios por hora, $\ln(AHE)$, sobre las variables *Age*, Age^2 , *Female* y *Bachelor*. Si la variable *Age* aumenta de 25 a 26, ¿cuánto se espera que cambien los ingresos? Si *Age* aumenta desde 33 hasta 34, ¿cuánto se espera que cambien los ingresos?
- ¿Prefiere la regresión de (c) a la regresión de (b)? Explíquelo.
- ¿Prefiere la regresión de (d) a la regresión de (b)? Explíquelo.
- ¿Prefiere la regresión de (d) a la regresión de (c)? Explíquelo.
- Represente gráficamente la relación entre las variables *Age* y $\ln(AHE)$ a partir de (b), (c) y (d) para los varones con un diploma de escuela secundaria. Describa las similitudes y las diferencias entre las funciones de regresión estimadas. ¿Cambiaría su respuesta si se representase gráficamente la función de regresión para las mujeres con título universitario?
- Realice una regresión del $\ln(AHE)$, sobre las variables *Age*, Age^2 , *Female*, *Bachelor* y el término de interacción *Female* \times *Bachelor*. ¿Cuál es el coeficiente del término de interacción? Alexis es una mujer de 30 años de edad con un título de licenciatura. ¿Qué predice la regresión para el

valor de $\ln(AHE)$ que le corresponde? Jane es una mujer de 30 años con un diploma de secundaria. ¿Qué predice la regresión para el valor de $\ln(AHE)$ que le corresponde? ¿Cuál es la predicción para la diferencia entre los ingresos de Alexis y los de Jane? Bob es un hombre de 30 años de edad con un grado de licenciatura. ¿Qué predice la regresión para el valor de $\ln(AHE)$ que le corresponde? Jim es un hombre de 30 años de edad, con un diploma de escuela secundaria. ¿Qué predice la regresión para el valor de $\ln(AHE)$ que le corresponde? ¿Cuál es la predicción para la diferencia entre los ingresos de Bob y de Jim?

- j) ¿Es el efecto de la variable *Age* sobre los ingresos salariales diferente para los hombres y para las mujeres? Especifique y estime una regresión que pueda utilizarse para responder a esta pregunta.
- k) ¿Es el efecto de la variable *Age* sobre los ingresos salariales distinto para los graduados de secundaria que para los graduados en la universidad? Especifique y estime una regresión que pueda utilizarse para responder a esta pregunta.
- l) Después de realizar todas estas regresiones (y cualquier otra que desee), resuma el efecto de la variable *Age* sobre los ingresos salariales de los trabajadores jóvenes.

E8.2. Utilizando la base de datos **TeachingRatings** descrita en el Ejercicio empírico 4.2, realice los siguientes ejercicios.

- a) Estime una regresión de la variable *Course_Eval* sobre las variables *Beauty*, *Intro*, *OneCredit*, *Female*, *Minority* y *NNEnglish*.
- b) Añadiendo las variables *Age* y *Age*² a la regresión. ¿Existe evidencia de que la variable *Age* tenga un efecto no lineal sobre la variable *Course_Eval*? ¿Existe evidencia de que la variable *Age* tenga algún efecto sobre la variable *Course_Eval*?
- c) Modifique la regresión en (a) para que el efecto de la variable *Beauty* sobre la variable *Course_Eval* sea diferente para hombres y mujeres. ¿Es la diferencia entre hombres y mujeres en el efecto de la variable *Beauty* estadísticamente significativa?
- d) El profesor Smith es un hombre. Se realiza una intervención de cirugía estética que aumenta su índice de belleza partiendo de una desviación estándar por debajo de la media hasta una desviación estándar por encima de la media. ¿Cuál el valor de la variable *Beauty* que le correspondía antes de la cirugía? ¿Y después de la cirugía? Utilizando la regresión de (c), construya un intervalo de confianza al 95 % para el aumento de la evaluación de su asignatura.
- e) Repita (d) para la profesora Jones, que es una mujer.

E8.3. Utilice la base de datos **CollegeDistance** descrita en el Ejercicio empírico 4.3 para contestar las siguientes preguntas.

- a) Realice una regresión de la variable *ED* sobre las variables *Dist*, *Female*, *Bytest*, *Tuition*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *DadColl*, *MomColl*, *Cue80* y *Stwmfg80*. Si la variable *Dist* aumenta de 2 a 3 (es decir, de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si la variable *Dist* aumenta desde 6 hasta 7 (es decir, desde 60 hasta 70 millas), ¿cuánto se espera que cambien los años de educación?
- b) Realice una regresión del $\ln(ED)$ sobre las variables *Dist*, *Female*, *Bytest*, *Tuition*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *DadColl*, *MomColl*, *Cue80* y *Stwmfg80*. Si la variable *Dist* aumenta de 2 a 3 (de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si la variable *Dist* aumenta de 6 a 7 (es decir, de 60 a 70 millas), ¿cuánto se espera que cambien los años de educación?
- c) Realice una regresión de la variable *ED* sobre las variables *Dist*, *Dist*², *Female*, *Bytest*, *Tuition*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *DadColl*, *MomColl*, *Cue80* y *Stwmfg80*. Si la variable *Dist* aumenta de 2 a 3 (de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si *Dist* aumenta de 6 a 7 (es decir, de 60 a 70 millas), ¿cuánto se espera que cambien los años de educación?
- d) ¿Prefiere la regresión de (c) a la regresión de (a)? Explíquelo.

- e) Considere el caso de una mujer hispana con $Tuition = 950 \$$, $Bytest = 58$, $Incomehi = 0$, $Ownhome = 0$, $DadColl = 1$, $MomColl = 1$, $Cue80 = 7,1$ y $Stwmfg = 10,06$.
- Represente gráficamente la relación de regresión entre las variables ED y $Dist$ de los apartados (a) y (c) para la variable $Dist$ en el rango de 0 a 10 (de 0 a 100 millas). Describa las similitudes y diferencias entre las funciones de regresión estimadas. ¿Cambiaría su respuesta si se representara gráficamente la función de regresión para un varón blanco con las mismas características?
 - ¿De qué manera se comporta la función de regresión (c) para $Dist > 10$? ¿Cuántas observaciones presentan un valor para la variable $Dist > 10$?
- f) Añada el término de interacción $DadColl \times MomColl$ a la regresión de (c). ¿Qué mide el coeficiente del término de interacción?
- g) Mary, Jane, Alexis y Bonnie tienen los mismos valores para las variables $Dist$, $Bytest$, $Tuition$, $Female$, $Black$, $Hispanic$, $Fincome$, $Ownhome$, $Cue80$ y $Stwmfg80$. Ninguno de los padres de Mary fue a la universidad. El padre de Jane asistió a la universidad, pero su madre no lo hizo. La madre de Alexis asistió a la universidad, pero su padre no lo hizo. Tanto el padre como la madre de Bonnie fueron a la universidad. Utilizando las regresiones de (f).
- ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Jane y Mary?
 - ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Alexis y Mary?
 - ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Bonnie y Mary?
- h) ¿Existe alguna evidencia de que el efecto de la variable $Dist$ sobre la variable ED dependa de la renta de la familia?
- i) Después de realizar todas estas regresiones (y cualquier otra que desee), resuma el efecto de la variable $Dist$ sobre los años de educación.
- E8.4.** Con la base de datos **Growth** descrita en el Ejercicio empírico 4.4, excluyendo los datos para Malta, realice las cinco siguientes regresiones: la variable $Growth$ sobre (1) las variables $TradeShare$ y $YearsSchool$; (2) las variables $TradeShare$ y $\ln(YearsSchool)$; (3) las variables $TradeShare$, $\ln(YearsSchool)$, Rev_Coups , $Assassinations$ y $\ln(RGDP60)$; (4) las variables $TradeShare$, $\ln(YearsSchool)$, Rev_Coups , $Assassinations$, $\ln(RGDP60)$ y $TradeShare \ln(YearsSchool)$; y (5) las variables $TradeShare$, $TradeShare^2$, $TradeShare^3$, $\ln(YearsSchool)$, Rev_Coups , $Assassinations$ y $\ln(RGDP60)$.
- Construya un diagrama de dispersión para las variables $Growth$ y $YearsSchool$. ¿La relación tiene aspecto de ser lineal o no lineal? Explíquelo. Utilice el gráfico para explicar por qué la regresión (2) se ajusta mejor que la regresión (1).
 - En 1960, un país contemplaba una política educativa que aumentara la media de años de escolarización de 4 a 6 años. A partir de la regresión (1) prediga el aumento en la variable $Growth$. Utilice la regresión (2) para predecir el incremento en $Growth$.
 - Contraste si los coeficientes de las variables $Assassinations$ y Rev_Coups son iguales a cero utilizando la regresión (3).
 - A partir de la regresión (4), existe evidencia de que el efecto de la variable $TradeShare$ sobre la variable $Growth$ depende del nivel de la educación del país?
 - A partir de la regresión (5), existe evidencia de una relación no lineal entre las variables $TradeShare$ y $Growth$?
 - En 1960, un país contemplaba una política comercial que aumentara el valor medio de la variable $TradeShare$ desde 0,5 hasta 1. Utilice la regresión (3) para predecir el incremento en la variable $Growth$. Utilice la regresión (5) para predecir el aumento en la variable $Growth$.

APÉNDICE

8.1 Funciones de regresión que son no lineales en los parámetros

Las funciones de regresión no lineales consideradas en las Secciones 8.2 y 8.3 son funciones no lineales de las X , pero son funciones lineales de los parámetros desconocidos. Debido a que son lineales en los parámetros desconocidos, los parámetros pueden ser estimados por MCO después de definir regresores nuevos que son transformaciones lineales de las X originales. Esta familia de funciones de regresión no lineales es amplia y fácil de usar. En algunas aplicaciones, sin embargo, el razonamiento económico lleva a que las funciones de regresión sean no lineales en los parámetros. A pesar de que estas funciones de regresión no pueden ser estimadas por MCO, pueden estimarse mediante una generalización de MCO denominada mínimos cuadrados no lineales.

Funciones que son no lineales en los parámetros

Comenzamos con dos ejemplos de funciones que no son lineales en los parámetros. Posteriormente se proporciona una formulación general.

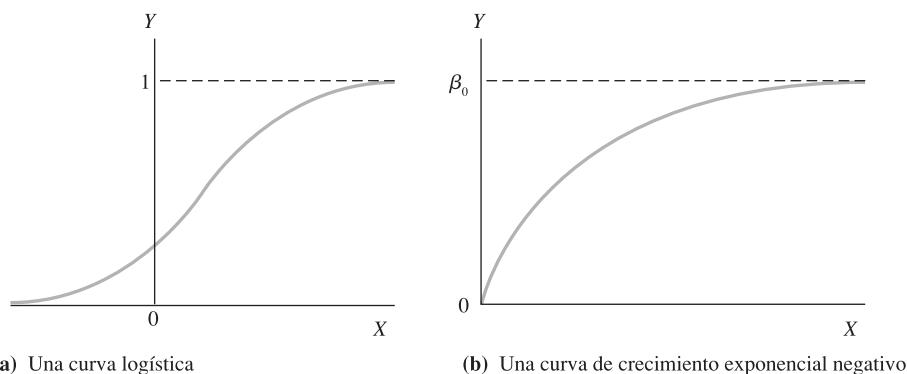
Curva logística. Supongamos que se está estudiando la penetración en el mercado de una tecnología, como es la adopción de un *software* de gestión para bases de datos en diferentes sectores económicos. La variable dependiente es la proporción de empresas en el sector que han adoptado el *software*, una única variable independiente X describe una característica del sector, y se dispone de datos sobre n sectores. La variable dependiente se encuentra entre 0 (las que no lo adoptan) y 1 (las que lo adoptan al 100 %). Debido a que un modelo de regresión lineal podría generar valores de predicción menores que 0 o mayores que 1, tiene sentido utilizar en su lugar una función que genera valores de predicción entre 0 y 1.

La función logística aumenta suavemente a partir de un mínimo de 0 hasta un máximo de 1. El modelo de regresión logística con una única X es

$$Y_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i)}} + u_i \quad (8.38)$$

La función logística con una única X se representa gráficamente en la Figura 8.12a. Como puede observarse en el gráfico, la función logística tiene una forma de «S» alargada. Para valores pequeños de X , el valor de la función está cercano a 0, y la pendiente es plana; la curva es más pronunciada para valores medios de X ; y para valores grandes de X , la función se aproxima a 1 y la pendiente es plana otra vez.

FIGURA 8.12 Dos funciones que son no lineales en los parámetros



La parte (a) representa gráficamente la función logística de la Ecuación (8.38), cuyos valores de predicción se encuentran entre 0 y 1. La parte (b) representa gráficamente la función de crecimiento exponencial negativo de la Ecuación (8.39), que tiene una pendiente que siempre es positiva y disminuye a medida que X aumenta, y una asíntota en β_0 cuando X tiende a infinito.

Crecimiento exponencial negativo. Las funciones utilizadas en la Sección 8.2 para modelizar la relación entre las calificaciones en los exámenes y la renta presentan algunas deficiencias. Por ejemplo, los modelos polinomiales pueden dar lugar a una pendiente negativa para algunos valores de la renta, lo que es inverosímil. La especificación logarítmica presenta una pendiente positiva para todos los valores de la renta; sin embargo, si los valores de la renta se hacen muy grandes, los valores de predicción crecen sin límite, por lo que para algunos valores de la renta el valor de predicción para un distrito puede que exceda la máxima puntuación posible en el examen.

El modelo de crecimiento exponencial negativo proporciona una especificación no lineal que presenta una pendiente positiva para todos los valores de la renta, presenta una pendiente que es mayor para los valores bajos de la renta y disminuye a medida que aumenta la renta, y tiene un límite superior (es decir, una asíntota cuando la renta aumenta hasta el infinito). El modelo de regresión de crecimiento exponencial negativo es

$$Y_i = \beta_0 [1 - e^{-\beta_1(X_i - \beta_2)}] + u_i. \quad (8.39)$$

La función de crecimiento exponencial negativo se representa gráficamente en la Figura 8.12b. La pendiente es pronunciada para los valores bajos de X , pero cuando X aumenta alcanza una asíntota en β_0 .

Funciones generales que son no lineales en los parámetros. Los modelos de regresión logística y de crecimiento exponencial negativo son casos particulares del modelo general de regresión no lineal

$$Y_i = f(X_{1i}, \dots, X_{ki}; \beta_0, \dots, \beta_m) + u_i, \quad (8.40)$$

en el que existen k variables independientes y $m + 1$ parámetros, β_0, \dots, β_m . En los modelos de las Secciones 8.2 y 8.3, las X entraban en esta función de forma no lineal, pero los parámetros entraban de forma lineal. En los ejemplos de este apéndice, los parámetros entran del mismo modo de forma no lineal. Si los parámetros son conocidos, entonces los efectos previstos se pueden calcular utilizando el método descrito en la Sección 8.1. En los casos prácticos concretos, sin embargo, los parámetros son desconocidos y deben estimarse a partir de los datos. Los parámetros que entran en forma no lineal no pueden ser estimados por MCO, pero se pueden estimar por mínimos cuadrados no lineales.

Estimación por mínimos cuadrados no lineales

Mínimos cuadrados no lineales es un método general para estimar los parámetros desconocidos de una función de regresión cuando estos parámetros entran en la función de regresión poblacional no linealmente.

Recordemos el análisis de la Sección 5.3 sobre el estimador MCO de los coeficientes del modelo lineal de regresión múltiple. El estimador MCO minimiza la suma de los errores de predicción al cuadrado de la Ecuación (5.8), $\sum_{i=1}^n [Y_i - (b_0 + b_1X_{1i} + \dots + b_kX_{ki})]^2$. En principio, el estimador MCO puede calcularse probando muchos valores para b_0, \dots, b_k y resolviendo para los valores que minimizan la suma de los errores al cuadrado.

Este mismo método puede utilizarse para estimar los parámetros del modelo general de regresión no lineal de la Ecuación (8.40). Debido a que la función de regresión es no lineal en los coeficientes, este método se denomina **mínimos cuadrados no lineales**. Para un conjunto de valores de prueba de los parámetros b_0, b_1, \dots, b_m se construye la suma de los errores de predicción al cuadrado:

$$\sum_{i=1}^n [Y_i - f(X_{1i}, \dots, X_{ki}, b_1, \dots, b_m)]^2. \quad (8.41)$$

Los **estimadores de mínimos cuadrados no lineales** de $\beta_0, \beta_1, \dots, \beta_m$ son los valores de b_0, b_1, \dots, b_m que minimizan la suma de los errores de predicción al cuadrado de la Ecuación (8.41).

En regresión lineal, una fórmula relativamente sencilla expresa el estimador MCO como una función de los datos. Desafortunadamente, no existe tal fórmula general para mínimos cuadrados no lineales, por lo que el estimador de mínimos cuadrados no lineales debe hallarse por métodos numéricos mediante un ordenador. El *software* de regresión incorpora algoritmos para la resolución del problema de minimización de mínimos cuadrados no lineales, lo que simplifica en la práctica la tarea de calcular el estimador de mínimos cuadrados no lineales.

Bajo condiciones generales sobre la función f , y las X , el estimador de mínimos cuadrados no lineales comparte dos propiedades clave con el estimador MCO del modelo de regresión lineal: es consistente, y está normalmente distribuido en muestras grandes. En el *software* de regresión que incluye la estimación de mínimos cuadrados no lineales, la salida normalmente informa de los errores estándar de los parámetros estimados. Como consecuencia, la inferencia que se refiere a los parámetros puede llevarse a cabo de la forma habitual; en particular, los estadísticos t se pueden construir mediante el método general del Concepto clave 5.1, y puede construirse un intervalo de confianza al 95 % como el coeficiente estimado, más o menos 1,96 errores estándar. Al igual que en la regresión lineal, el término de error en el

modelo de regresión no lineal puede ser heterocedástico, por lo que deben utilizarse los errores estándar heterocedástico-robustos.

Aplicación a la relación calificaciones-renta

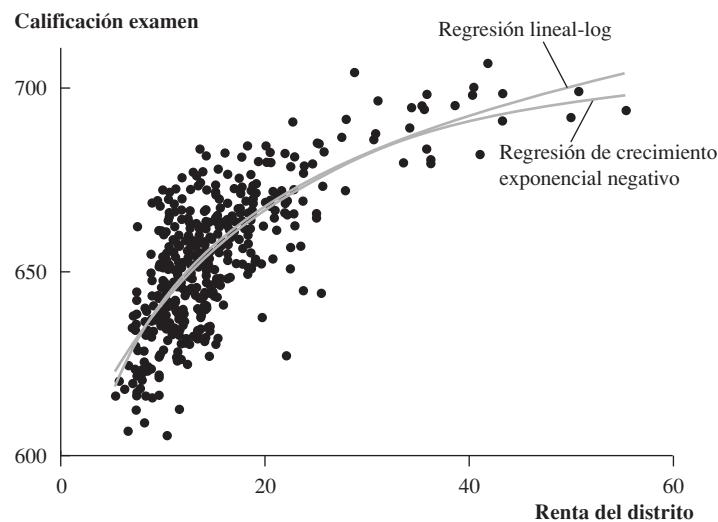
Un modelo de crecimiento exponencial negativo, se ajusta a la renta del distrito (X) y las calificaciones en los exámenes (Y), y presenta las características deseables de una pendiente que siempre es positiva [si β_1 en la Ecuación (8.39) es positivo], y una asíntota en β_0 cuando la renta tiende a infinito. El resultado de estimar β_0 , β_1 y β_2 en la Ecuación (8.39) con los datos de las calificaciones de los exámenes de California se obtiene $\hat{\beta}_0 = 703,2$ (error estándar heterocedástico-robusto = 4,44), $\hat{\beta}_1 = 0,0552$ ($ES = 0,0068$), y $\hat{\beta}_2 = -34,0$ ($ES = 4,48$). Por tanto, la función de regresión no lineal estimada (con los errores estándar presentados debajo de las estimaciones de los parámetros) es

$$\widehat{\text{CalificaciónExamen}} = 703,2[1 - e^{-0,0552(\text{Renta} + 34,0)}]. \quad (8.42)$$

(4,44) (0,0068) (4,48)

Esta función de regresión estimada se representa gráficamente en la Figura 8.13, junto con la función de regresión logarítmica y un diagrama de dispersión de los datos. Las dos especificaciones son, en este caso, muy similares. Una diferencia es que la curva de crecimiento exponencial negativo se aplana en los niveles de renta más altos, lo que es consistente con la presencia de una asíntota.

FIGURA 8.13 Las funciones de regresión de crecimiento exponencial negativo y lineal-log



La función de regresión de crecimiento exponencial negativo [Ecuación (8.42)] y la función de regresión lineal-log [Ecuación (8.18)], captan ambas la relación no lineal entre las calificaciones y la renta del distrito. Una de las diferencias entre las dos funciones es que el modelo de crecimiento exponencial negativo tiene una asíntota cuando la variable Renta aumenta hasta el infinito, pero no así la función de regresión lineal-log.

APÉNDICE

8.2 Pendientes y elasticidades de funciones de regresión no lineales

En este apéndice se utiliza cálculo para evaluar las pendientes y las elasticidades de las funciones de regresión no lineales con regresores continuos. Nos centramos en el caso de la Sección 8.2, en el que hay una sola X . Este enfoque se extiende a múltiples X , mediante derivadas parciales.

Consideremos el modelo de regresión no lineal $Y_i = f(X_i) + u_i$, con $E(u_i|X_i) = 0$. La pendiente de la función de regresión poblacional, $f(X)$ evaluada en el punto $X = x$, es la derivada de f , es decir, $df(X)/dX|_{X=x}$. Para la función de regresión polinomial de la Ecuación (8.9), $f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_r X^r$ y $dX^a/dX = aX^{a-1}$ para cualquier constante a , por lo que $df(X)/dX|_{X=x} = \beta_1 + 2\beta_2 x + \dots + r\beta_r x^{r-1}$. La pendiente estimada en x es $\hat{df}(X)/dX|_{X=x} = \hat{\beta}_1 + 2\hat{\beta}_2 x + \dots + r\hat{\beta}_r x^{r-1}$. El error estándar de la pendiente estimada es $SE(\hat{\beta}_1 + 2\hat{\beta}_2 x + \dots + r\hat{\beta}_r x^{r-1})$; para un valor dado de x , este es el error estándar de la suma ponderada de los coeficientes de regresión, que puede calcularse utilizando los métodos de la Sección 7.3 y la Ecuación (8.8).

La elasticidad de Y con respecto a X es el cambio porcentual en Y para un determinado cambio porcentual en X . Formalmente, esta definición se aplica en el límite en el que el porcentaje de cambio en X tiende a cero, por lo que la pendiente que aparece en la definición de la Ecuación (8.22) se sustituye por la derivada y la elasticidad es

$$\text{elasticidad de } Y \text{ con respecto a } X = \frac{dY}{dX} \times \frac{X}{Y} = \frac{d \ln Y}{d \ln X}$$

En un modelo de regresión, Y depende tanto de la X como del término de error u . Debido a que u es aleatorio, es habitual evaluar la elasticidad no como el cambio porcentual en Y , sino en el componente de predicción de Y , es decir, el cambio porcentual en $E(Y|X)$. De acuerdo con esto, la elasticidad de $E(Y|X)$ con respecto a X es

$$\frac{dE(Y|X)}{dX} \times \frac{X}{E(Y|X)} = \frac{d \ln E(Y|X)}{d \ln X}$$

Las elasticidades para el modelo lineal y para los tres modelos logarítmicos resumidos en el Concepto clave 8.2 se ofrecen en la tabla siguiente.

Caso	Modelo de regresión poblacional	Elasticidad de $E(Y X)$ con respecto a X
lineal	$Y = \beta_0 + \beta_1 X + u$	$\frac{\beta_1 X}{\beta_0 + \beta_1 X}$
lineal-log	$Y = \beta_0 + \beta_1 \ln(X) + u$	$\frac{\beta_1}{\beta_0 + \beta_1 \ln(X)}$
log-lineal	$\ln(Y) = \beta_0 + \beta_1 X + u$	$\beta_1 X$
log-log	$\ln(Y) = \beta_0 + \beta_1 \ln(X) + u$	β_1

La especificación log-log presenta una elasticidad constante, pero para las otras tres especificaciones la elasticidad depende de X .

Ahora se obtienen las expresiones para los modelos lineal-log y log-lineal. Para el modelo lineal-log, $E(Y|X) = \beta_0 + \beta_1 \ln(X)$. Como $d \ln(X)/dX = 1/X$, aplicando la regla de la cadena se obtiene $dE(Y|X)/dX = \beta_1/X$. Por tanto la elasticidad es $dE(Y|X)/dX \times X/E(Y|X) = (\beta_1/X) \times X/[\beta_0 + \beta_1 \ln(X)] = \beta_1/[\beta_0 + \beta_1 \ln(X)]$, como aparece en la tabla. Para el modelo log-lineal, es habitual realizar el supuesto adicional de que u y X están distribuidas independientemente, por lo que la expresión para $E(Y|X)$ obtenida siguiendo la Ecuación (8.25) se convierte en $E(Y|X) = ce^{\beta_0 + \beta_1 X}$ donde $c = E(e^u)$ es una constante que no depende de X por el supuesto adicional de que u y X son independientes. Por tanto $dE(Y|X)/dX = ce^{\beta_0 + \beta_1 X}\beta_1$ y la elasticidad es $dE(Y|X)/dX \times X/E(Y|X) = ce^{\beta_0 + \beta_1 X}\beta_1 \times X/(ce^{\beta_0 + \beta_1 X}) = \beta_1 X$. La obtención de estas expresiones para los modelos lineal y log-log se dejan como Ejercicio 8.11.

9 Evaluación de estudios basados en regresión múltiple

Los cinco capítulos anteriores explican cómo utilizar la regresión múltiple para analizar la relación entre las variables de una base de datos. En este capítulo, volvemos atrás y nos preguntamos ¿qué hace que un estudio que utiliza regresión múltiple sea o no sea fiable?. Nos centramos en los estudios estadísticos cuyo objetivo es estimar el efecto causal de un cambio en alguna variable independiente, como el tamaño de la clase, sobre una variable dependiente, como es la calificación en los exámenes. En estos estudios, ¿cuándo proporcionará la regresión múltiple un estimador útil del efecto causal?, y, lo que es igual de importante, ¿cuándo no lo hará?

Para responder a estas cuestiones, este capítulo presenta un marco para evaluar estudios estadísticos en general, tanto si utilizan análisis de regresión como si no lo hacen. Este marco de análisis descansa en los conceptos de validez interna y externa. Un estudio es válido internamente si sus inferencias estadísticas acerca de los efectos causales son válidas para la población y el escenario estudiados; es válido externamente si sus inferencias pueden generalizarse a otras poblaciones y escenarios. En las Secciones 9.1 y 9.2 se tratan la validez interna y externa, se enumeran una serie de amenazas a la validez interna y externa, y se habla de cómo identificar estas amenazas en la práctica. El estudio en las Secciones 9.1 y 9.2 se centra en la estimación de los efectos causales a partir de datos observacionales. La Sección 9.3 trata acerca de una utilización diferente de los modelos de regresión, la predicción, y proporciona una introducción a las amenazas para la validez de las predicciones realizadas mediante los modelos de regresión.

Como ilustración del marco de análisis de la validez interna y externa, en la Sección 9.4 se evalúa la validez interna y externa del estudio acerca del efecto sobre las calificaciones en los exámenes del recorte de la ratio estudiantes-maestros presentado en los Capítulos 4 a 8.

9.1 Validez interna y externa

Los conceptos de validez interna y externa, definidos en el Concepto clave 9.1 proporcionan el marco para evaluar si un estudio estadístico o estadístico resulta útil para responder a una cuestión específica de interés.

La validez interna y externa distinguen entre la población y el escenario estudiados y la población y el escenario para los cuales se generalizan los resultados. La **población estudiada** es la población de individuos —personas, empresas, distritos escolares, etcétera— de los cuales se extrae la muestra. La población para la cual los resultados se generalizan, o **población de interés**, es la población de entidades individuales para la que se van a aplicar las inferencias causales del estudio. Por ejemplo, un director de una escuela secundaria (cursos 9 a 12) podría desear generalizar nuestros hallazgos sobre el tamaño de las clases y las calificaciones en los exámenes de los distritos de escuela primaria de California (la población estudiada) a la población de escuelas secundarias (la población de interés).

Con «escenario», nos referimos al entorno institucional, legal, social y económico. Por ejemplo, sería importante saber si los hallazgos de un experimento de laboratorio para la evaluación de los métodos de cultivo de tomates orgánicos podrían generalizarse al terreno, es decir, si los métodos orgánicos que funcionan en un entorno de laboratorio funcionan asimismo en un escenario del mundo real. Se proporcionan otros ejemplos de diferencias en poblaciones y escenarios más adelante en esta sección.

**CONCEPTO
CLAVE**
9.1**Validez interna y externa**

Se dice que un análisis estadístico tiene **validez interna** si las inferencias estadísticas acerca de los efectos causales son válidas para la población que está siendo estudiada.

Se dice que el análisis tiene **validez externa** si sus inferencias y conclusiones se pueden generalizar a partir de la población y el escenario estudiados para otras poblaciones y entornos.

Amenazas a la validez interna

La validez interna tiene dos componentes. En primer lugar, el estimador del efecto causal debe ser insesgado y consistente. Por ejemplo, si $\hat{\beta}_{REM}$ es el estimador MCO del efecto sobre las calificaciones en los exámenes de una variación unitaria en la ratio estudiantes-maestros en una determinada regresión, entonces $\hat{\beta}_{REM}$ debe ser un estimador insesgado y consistente del verdadero efecto causal poblacional de una variación en la ratio estudiantes-maestros, β_{REM} .

En segundo lugar, los contrastes de hipótesis deben tener el nivel de significación deseado (la tasa de rechazo efectiva del contraste bajo la hipótesis nula debe ser igual al nivel de significación deseado), y los intervalos de confianza deben tener el nivel de confianza deseado. Por ejemplo, si un intervalo de confianza se construye como $\hat{\beta}_{REM} \pm 1,96ES(\hat{\beta}_{REM})$, este intervalo de confianza debe contener el verdadero efecto causal poblacional con una probabilidad del 95 % en muestras repetidas.

En el análisis de regresión, los efectos causales se estiman mediante las funciones de regresión estimadas y los contrastes de hipótesis se realizan utilizando los coeficientes de regresión estimados y sus errores estándar. De acuerdo con esto, en un estudio basado en la regresión MCO, los requisitos para la validez interna son que el estimador MCO sea insesgado y consistente, y que los errores estándar se calculen de una manera que haga que los intervalos de confianza presenten el nivel de confianza deseado. Por diferentes razones estos requisitos podrían no cumplirse, y estas razones constituyen amenazas a la validez interna. Estas amenazas conducen a incumplimientos de uno o más de los supuestos de mínimos cuadrados del Concepto clave 6.4. Por ejemplo, una amenaza que ya ha sido tratada en detalle es el sesgo de variable omitida; que conduce a la correlación entre una o más variables explicativas y el término de error, lo cual viola el primero de los supuestos de mínimos cuadrados. Si se dispone de los datos sobre la variable omitida o sobre una variable de control adecuada, entonces esta amenaza se puede evitar mediante la inclusión de esta variable como regresor adicional.

La Sección 9.2 ofrece una descripción detallada de las distintas amenazas a la validez interna en el análisis de regresión múltiple y sugiere cómo mitigarlas.

Amenazas a la validez externa

Las posibles amenazas a la validez externa surgen de las diferencias entre la población y el escenario estudiado y la población y el escenario de interés.

Diferencias entre poblaciones. Las diferencias entre la población estudiada y la población de interés pueden suponer una amenaza a la validez externa. Por ejemplo, los estudios de laboratorio sobre los efectos tóxicos de productos químicos suelen utilizar poblaciones de animales como los ratones (la población estudiada), pero los resultados se utilizan para redactar las normas de salud y seguridad para poblaciones humanas (la población de interés). Si ratones y hombres se diferencian lo suficiente como para que este hecho constituya una amenaza a la validez externa de estos estudios, es un asunto de debate.

De manera más general, el verdadero efecto causal puede no ser el mismo en la población estudiada y en la población de interés. Esto podría deberse a que la población fue elegida de una manera que la hace diferente de la población de interés, por las diferencias en las características de la población, las diferencias geográficas, o bien debido a que el estudio no está actualizado.

Diferencias en el escenario. Incluso aunque la población estudiada y la población de interés sean la misma, tal vez no sea posible generalizar los resultados del estudio si los escenarios son distintos. Por ejem-

plo, un estudio acerca del efecto sobre las juergas universitarias de una campaña publicitaria contra el alcohol podría no ser generalizable a otro grupo idéntico de estudiantes universitarios, si las sanciones legales por consumo de alcohol son diferentes entre ambas universidades. En este caso, el escenario legal en el que se lleva a cabo el estudio difiere del escenario legal en que sus resultados se aplican.

De manera más general, los ejemplos de las diferencias en los escenarios incluyen diferencias en el entorno institucional (universidades públicas frente a universidades religiosas), diferencias en las leyes (las diferencias en las sanciones legales), o diferencias en el entorno físico (las borracheras de botellón en el sur de California frente a las de Fairbanks, Alaska).

Aplicación a las calificaciones y ratio estudiantes-maestros. Los Capítulos 7 y 8 presentaban mejoras sobre las calificaciones de los exámenes estadísticamente significativas, pero sustancialmente pequeñas, de la reducción de la ratio estudiantes-maestros. Este análisis se basaba en los resultados de las pruebas para los distritos escolares de California. Supongamos por un momento que estos resultados son internamente válidos. ¿Para qué otras poblaciones y escenarios de interés podrían generalizarse estos resultados?

Cuanto más cercanos a la población y al escenario del estudio se encuentren la población y escenario de interés, más fuertes serán las razones para la validez externa. Por ejemplo, los estudiantes universitarios, y la formación universitaria son muy diferentes a los estudiantes y la formación de la escuela primaria, por lo que resulta poco verosímil que el efecto de la reducción del tamaño de las clases estimado a partir de los datos de los distritos escolares de educación primaria de California sea generalizable a las universidades. Por otro lado, los estudiantes, el currículo y la organización en las escuelas primarias son muy similares a lo ancho de los Estados Unidos, por lo que resulta verosímil que los resultados de California pudieran generalizarse al desempeño en los exámenes estandarizados en otros distritos de educación primaria de EE.UU.

Cómo evaluar la validez externa de un estudio. La validez externa debe ser juzgada mediante el conocimiento específico de las poblaciones y los escenarios estudiados y los de interés. Las diferencias importantes entre ellos pondrán en tela de juicio la validez externa del estudio.

A veces existen dos o más estudios sobre poblaciones diferentes, pero relacionadas. Si es así, la validez externa de ambos estudios se puede comprobar mediante la comparación de sus resultados. Por ejemplo, en la Sección 9.4 se analizan los datos sobre las calificaciones en los exámenes y el tamaño de las clases en los distritos de escuela primaria de Massachusetts y se comparan los resultados de Massachusetts y California. En general, las conclusiones similares en dos o más estudios impulsan las razones para la validez externa, mientras que las diferencias en sus resultados que no resulten fácilmente explicables ponen en duda su validez externa¹.

Cómo diseñar un estudio externamente válido. Dado que las amenazas a la validez externa se derivan de la falta de comparabilidad entre las poblaciones y los escenarios, estas amenazas se minimizan de un modo mejor en las etapas iniciales de un estudio, antes de que se obtengan los datos. El diseño de un estudio queda fuera del alcance de este libro, por lo que se remite al lector interesado al trabajo de Shadish, Cook y Campbell (2002).

9.2 Amenazas a la validez interna del análisis de regresión múltiple

Los estudios basados en el análisis de regresión son internamente válidos si los coeficientes de regresión estimados son insesgados y consistentes, y si sus errores estándar proporcionan intervalos de confianza con el nivel de confianza deseado. Esta sección inspecciona cinco razones por las que el estimador MCO de los coeficientes de regresión múltiple puede ser sesgado incluso en muestras grandes: variables omitidas,

¹ La comparación de muchos estudios relacionados acerca del mismo tema se denomina meta-análisis. Por ejemplo, el debate del recuadro de «El efecto Mozart: ¿sesgo de variable omitida?» del Capítulo 6 se basa en un meta-análisis. Realizar un meta-análisis acerca de muchos estudios presenta sus propios desafíos. ¿Cómo separar los buenos estudios de los malos? ¿Cómo comparar los estudios, cuando las variables dependientes son diferentes? ¿Deben ponderarse más los estudios con muestras más grandes? El estudio sobre el meta-análisis y sus retos queda más allá del alcance de este libro de texto. El lector interesado puede consultar el trabajo de Hedges y Olkin (1985) y Cooper y Hedges (1994).

errores de especificación de la forma funcional de la función de regresión, medición imprecisa de las variables independientes («errores en las variables»), selección muestral, y causalidad simultánea. Las cinco fuentes de sesgo surgen debido a que el regresor está correlacionado con el término de error de la regresión poblacional, violando por tanto el primer supuesto de mínimos cuadrados del Concepto clave 6.4. Para cada una de ellas, se discute qué puede hacerse para reducir este sesgo. La sección concluye con un análisis sobre las circunstancias que conducen a los errores estándar inconsistentes y sobre qué puede hacerse al respecto.

Sesgo de variable omitida

Recordemos que el sesgo de variable omitida se produce cuando se omite de la regresión una variable que determina Y y que además está correlacionada con uno o con más de los regresores incluidos en esa misma regresión. Este sesgo persiste incluso en muestras grandes, por lo que el estimador MCO es inconsistente. La mejor manera de minimizar el sesgo de variable omitida depende de si están disponibles o no las variables que permiten tener en cuenta adecuadamente (o controlar por) la variable omitida potencial.

Soluciones para el sesgo de variable omitida cuando la variable es observable o bien existen variables de control adecuadas. Si se dispone de datos sobre la variable omitida, entonces se puede incluir esa variable en una regresión múltiple, solucionando por tanto el problema. Por otra parte, si se dispone de datos sobre una o más variables de control, y si esas variables de control son adecuadas en el sentido de que conducen a la independencia en media condicional [Ecuación (7.20)], entonces la inclusión de las variables de control elimina el posible sesgo en el coeficiente de la variable de interés.

La adición de una variable a una regresión presenta tanto costes como beneficios. Por un lado, la omisión de la variable podría ocasionar sesgo de variable omitida. Por otro lado, la inclusión de la variable cuando no corresponda (es decir, cuando su coeficiente de regresión poblacional sea cero) reducirá la precisión de los estimadores de los otros coeficientes de regresión. En otras palabras, la decisión acerca de incluir una variable implica un equilibrio entre el sesgo y la varianza del coeficiente de interés. En la práctica, existen cuatro pasos que pueden ayudar a decidir si se incluye una variable o un conjunto de variables en una regresión.

El primer paso consiste en identificar el coeficiente o coeficientes de interés clave en la regresión. En las regresiones de las calificaciones en los exámenes, este es el coeficiente de la ratio estudiantes-maestros, debido a que la pregunta originalmente planteada se refiere al efecto sobre las calificaciones de la reducción de la ratio estudiantes-maestros.

El segundo paso consiste en preguntarse: ¿cuáles son las fuentes más probables de un sesgo importante de variable omitida en esta regresión? Responder a esta pregunta requiere la aplicación de la teoría económica y de un conocimiento experto, y debería realizarse antes de llevar a cabo efectivamente cualquier regresión; debido a que este paso se lleva a cabo antes de analizar los datos, se denomina razonamiento *a priori* («antes del hecho»). En el ejemplo de las calificaciones en los exámenes, este paso implica la identificación de aquellos factores que son determinantes de las calificaciones en los exámenes y que, si se ignoran, podrían sesgar el estimador del efecto del tamaño de las clases. Los resultados de este paso son una especificación base para la regresión, que es el punto de partida para el análisis empírico de regresión, y una lista de otras variables «cuestionables» que podrían ayudar a mitigar los posibles sesgos de variables omitidas.

El tercer paso consiste en aumentar la especificación base con las variables de control adicionales cuestionables identificadas en el segundo paso. Si los coeficientes de las variables de control adicionales son estadísticamente significativos o si los coeficientes de interés estimados cambian sensiblemente cuando se incluyen las variables adicionales, entonces deben permanecer en la especificación y debe modificarse la especificación base. Si no, estas variables pueden ser excluidas de la regresión.

El cuarto paso consiste en presentar un resumen preciso de los resultados en forma de tabla. Esto proporciona «información completa» a un potencial escéptico, de forma tal que pueda obtener sus propias conclusiones. Las Tablas 7.1 y 8.3 son ejemplos de esta estrategia. Por ejemplo, en la Tabla 8.3, podría haberse presentado solamente la regresión de la columna (7), debido a que esta regresión resume los efectos

**CONCEPTO
CLAVE**
9.2
Sesgo de variable omitida: ¿deberían incluirse más variables en la regresión?

Si se incluye otra variable en la regresión múltiple, se eliminará la posibilidad del sesgo de variable omitida que pueda surgir al excluir esa variable, pero la varianza de los estimadores de los coeficientes de interés puede aumentar. Se ofrecen aquí algunas pautas que pueden ayudar a decidir si se debe incluir una variable adicional:

1. Ser específico acerca del coeficiente o coeficientes de interés.
2. Utilizar un razonamiento *a priori* para identificar las fuentes potenciales más importantes de sesgo de variable omitida, lo que lleva a una especificación base y a algunas variables «cuestionables».
3. Contrastar si otras variables de control «cuestionables» tienen un coeficiente distinto de cero.
4. Proporcionar tablas que representen los resultados «de divulgación completa» para que otros puedan ver el efecto de la inclusión de las variables cuestionables sobre el(s) coeficiente(s) de interés. ¿Cambian los resultados si se incluye una variable de control cuestionable?

relevantes y las no linealidades de las otras regresiones de la tabla. La presentación de las otras regresiones, sin embargo, permite al lector escéptico sacar sus propias conclusiones.

Estos pasos se resumen en el Concepto clave 9.2.

Soluciones al sesgo de variable omitida cuando no se dispone de variables de control adecuadas. La adición de una variable omitida a una regresión no es una opción si no se dispone de datos sobre esta variable y si no existen variables de control adecuadas. No obstante, existen otras tres maneras de resolver el sesgo de variable omitida. Cada una de estas tres soluciones evita el sesgo de variable omitida mediante la utilización de diferentes tipos de datos.

La primera solución consiste en utilizar los datos en los que se observa la misma unidad observacional en diferentes momentos del tiempo. Por ejemplo, las calificaciones en los exámenes y los datos relacionados pueden recogerse para los mismos distritos en 1995 y nuevamente en 2000. Los datos de este tipo se denominan datos panel. Tal y como se explica en el Capítulo 10, los datos de panel permiten que sea posible tener en cuenta (controlar por) las variables omitidas observables, siempre y cuando estas variables omitidas no cambien con el tiempo.

La segunda solución consiste en utilizar la regresión de variables instrumentales. Este método se basa en una nueva variable, denominada variable instrumental. La regresión de variables instrumentales se trata en el Capítulo 12.

La tercera solución consiste en utilizar un diseño de estudio en el que el efecto de interés (por ejemplo, el efecto de reducir el tamaño de las clases sobre el rendimiento de los estudiantes) se estudie mediante un experimento aleatorizado controlado. Los experimentos aleatorizados controlados se tratan en el Capítulo 13.

Error de especificación de la forma funcional de la función de regresión

Si la verdadera función de regresión poblacional es no lineal, pero la regresión estimada es lineal, entonces el **error de especificación de la forma funcional** provoca que el estimador MCO sea sesgado. Este sesgo es un tipo de sesgo de variable omitida, en el que las variables omitidas son los términos que reflejan los aspectos no lineales que faltan en la función de regresión.

Por ejemplo, si la función de regresión poblacional es un polinomio de segundo grado, una regresión que no incluya el cuadrado de la variable independiente presentará sesgo de variable omitida. El sesgo debido al error de especificación de la forma funcional se resume en el Concepto clave 9.3.

**CONCEPTO
CLAVE**

9.3

Error de especificación de la forma funcional

El error de especificación de la forma funcional aparece cuando la forma funcional de la función de regresión estimada difiere de la forma funcional de la función de regresión poblacional. Si la forma funcional se especifica de forma errónea, entonces el estimador del efecto parcial de un cambio en una de las variables será, en general, sesgado. El error de especificación de la forma funcional a menudo puede detectarse mediante la representación gráfica de los datos y de la función de regresión, y puede corregirse mediante la utilización de una forma funcional diferente.

Soluciones para el error de especificación de la forma funcional. Cuando la variable dependiente es continua (como las calificaciones en los exámenes), este problema de no linealidad potencial puede resolverse mediante los métodos del Capítulo 8. Sin embargo, si la variable dependiente es discreta o binaria (por ejemplo, Y_i es igual a 1 si la i -ésima persona asistió a la universidad y es igual a 0 en caso contrario), las cosas son más complicadas. La regresión con variable dependiente discreta se trata en el Capítulo 11.

Sesgo de errores de medida y por errores en las variables

Supongamos que en nuestra regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros se hubieran mezclado sin querer los datos de modo que terminamos regresando las calificaciones de los alumnos de quinto curso sobre la ratio estudiantes-maestros para los estudiantes de décimo curso de ese distrito. Aunque las ratios estudiantes-maestros para los estudiantes de primaria y de décimo curso puedan estar correlacionadas, no son iguales, por lo que esta confusión podría conducir a un sesgo en el coeficiente estimado. Este es un ejemplo de **sesgo por errores en las variables** debido a que su origen es un error de medición en las variables independientes. Este sesgo persiste incluso en muestras muy grandes, por lo que el estimador MCO es inconsistente si existe error de medición.

Existen muchas fuentes posibles de error de medición. Si los datos son recogidos a través de una encuesta, el encuestado puede dar una respuesta equivocada. Por ejemplo, una pregunta de la Encuesta Actualizada de Población se refiere a los ingresos del año anterior. Un encuestado podría no saber sus ingresos de forma exacta o bien podría declarar erróneamente la cantidad por otras razones. Si en su lugar los datos se obtienen de los registros administrativos informatizados, podría haber habido errores tipográficos, cuando se introdujeron los datos.

Para comprobar que los errores en las variables pueden dar lugar a la existencia de correlación entre el regresor y el término de error, supongamos que existe un único regresor X_i (por ejemplo, los ingresos en realidad percibidos) pero que X_i está medido de forma imprecisa mediante \tilde{X}_i (los ingresos declarados por los encuestados). Debido a que se observa \tilde{X}_i , pero no X_i , la ecuación de regresión estimada en realidad es la que se basa en \tilde{X}_i . Escrito en términos de la variable medida de forma imprecisa \tilde{X}_i , la ecuación de regresión poblacional $Y_i = \beta_0 + \beta_1 X_i + u_i$ es

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 \tilde{X}_i + [\beta_1(X_i - \tilde{X}_i) + u_i] \\ &= \beta_0 + \beta_1 \tilde{X}_i + v_i, \end{aligned} \tag{9.1}$$

donde $v_i = \beta_1(X_i - \tilde{X}_i) + u_i$. Por lo tanto la ecuación de regresión poblacional expresada en términos de \tilde{X}_i presenta un término de error que contiene el error de medición que es la diferencia entre \tilde{X}_i y X_i . Si esta diferencia está correlacionada con el valor medido \tilde{X}_i , entonces el regresor \tilde{X}_i estará correlacionado con el término de error y $\hat{\beta}_1$ será sesgado e inconsistente.

La cuantía exacta así como el sentido del sesgo de $\hat{\beta}_1$ dependerá de la correlación entre \tilde{X}_i y el error de medición, $\tilde{X}_i - X_i$. Esta correlación depende, a su vez, de la naturaleza específica del error de medición.

Por ejemplo, supongamos que el valor medido \tilde{X}_i , es igual al verdadero valor, no medible, X_i , más un componente puramente aleatorio, w_i , que tiene media igual a cero y varianza σ_w^2 . Debido a que el error es puramente aleatorio, podríamos suponer que w_i no está correlacionado con X_i ni con el error de regresión u_i . Este supuesto

constituye el **modelo clásico de error de medición** en el que $\tilde{X}_i = X_i + w_i$, en el que $\text{corr}(w_i, X_i) = 0$ y $\text{corr}(w_i, u_i) = 0$. En el modelo clásico de error de medición, un poco de álgebra² muestra que $\hat{\beta}_1$ tiene el límite de probabilidad

$$\hat{\beta}_1 \xrightarrow{P} \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2} \beta_1 \quad (9.2)$$

Es decir, si el error de medición tiene el efecto de añadir simplemente un elemento aleatorio al valor real de la variable independiente, entonces $\hat{\beta}_1$ es inconsistente. Debido a que el cociente $\frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2}$ es menor que 1, $\hat{\beta}_1$ estará sesgado hacia 0, incluso en muestras grandes. En el caso extremo de que el error de medición sea tan grande que en esencia no quede información sobre X_i , el cociente de las varianzas de la última expresión de la Ecuación (9.2) es 0 y $\hat{\beta}_1$ converge en probabilidad a 0. En el otro extremo, cuando no existe error de medición, $\sigma_w^2 = 0$ por lo que $\hat{\beta}_1 \xrightarrow{P} \beta_1$.

Un modelo diferente para el error de medición supone que el encuestado realiza su mejor estimación acerca del verdadero valor. En este modelo de «mejor conjetura» la respuesta \tilde{X}_i se modeliza como la media condicional de X_i , dada la información disponible para el encuestado. Debido a que \tilde{X}_i es la mejor conjetura, el error de medición $\tilde{X}_i - X_i$ no está correlacionado con la respuesta \tilde{X}_i (si el error de medición estuviera correlacionado con \tilde{X}_i , entonces eso sería una información útil para predecir X_i , en cuyo caso \tilde{X}_i no sería la mejor estimación de X_i). Es decir, $E[(\tilde{X}_i - X_i)\tilde{X}_i] = 0$, y si la información del encuestado no está correlacionada con el término de error u_i , entonces \tilde{X}_i no está correlacionada con el término de error v_i . Entonces en este modelo de error de medición de «mejor conjetura», $\hat{\beta}_1$ es consistente, pero debido a que $\text{var}(v_i) > \text{var}(u_i)$, la varianza de $\hat{\beta}_1$ es mayor de lo que lo sería en ausencia del error de medición. El modelo de error de medición de «mejor conjetura» se analiza más adelante en el Ejercicio 9.12.

Los problemas creados por el error de medición pueden ser aún más complicados si existe una declaración inexacta intencionada. Por ejemplo, supongamos que los encuestados deben facilitar la renta declarada en su declaración del impuesto sobre la renta, pero intencionalmente declaran una base imponible por debajo de su verdadera renta al no incluir los pagos en efectivo. Si, por ejemplo, todos los encuestados declarasen solamente el 90 % de los ingresos, entonces $\tilde{X}_i = 0,90X_i$, y $\hat{\beta}_1$ estará sesgado en más de un 10 %.

Aunque el resultado de la Ecuación (9.2) es específico para el error de medición clásico, ilustra la proposición más general de que si la variable independiente está medida de manera imprecisa, entonces el estimador MCO es sesgado, incluso en muestras grandes. El sesgo de errores en las variables se resume en el Concepto clave 9.4.

Error de medición en Y. El efecto del error de medición en Y es diferente del error de medición en X . Si Y presenta un error de medición clásico, entonces este error de medición aumenta la varianza de la regresión y de $\hat{\beta}_1$, pero no induce sesgo en $\hat{\beta}_1$. Para comprobarlo, supongamos que la medida de Y_i es \tilde{Y}_i que es igual a la verdadera Y_i más un error de medición aleatorio w_i . Entonces, el modelo de regresión estimado es $\tilde{Y}_i = \beta_0 + \beta_1 X_i + v_i$, donde $v_i = w_i + u_i$. Si w_i es verdaderamente aleatorio, entonces w_i y X_i se distribuyen

CONCEPTO CLAVE

9.4

Sesgo por errores en las variables

El sesgo por errores en las variables en el estimador MCO se produce cuando una variable independiente se mide de forma imprecisa. Este sesgo depende de la naturaleza del error de medida y persiste incluso si el tamaño de la muestra es grande. Si la variable medida es igual al valor real, más un error de medición con media igual a cero, que está independientemente distribuido, entonces el estimador MCO en una regresión con una sola variable en su parte derecha está sesgado hacia cero, y su límite de probabilidad está dado por la Ecuación (9.2).

² Bajo este supuesto de error de medición, $v_i = \beta_1(X_i - \tilde{X}_i) + u_i = -\beta_1 w_i + u_i$, $\text{cov}(X_i, u_i) = 0$, y $\text{cov}(\tilde{X}_i, w_i) = \text{cov}(X_i + w_i, w_i) = \sigma_w^2$, por lo que $\text{cov}(\tilde{X}_i, v_i) = -\beta_1 \text{cov}(\tilde{X}_i, w_i) + \text{cov}(\tilde{X}_i, u_i) = -\beta_1 \sigma_w^2$. Por lo tanto, a partir de la Ecuación (6.1), $\hat{\beta}_1 \xrightarrow{P} \beta_1 - \beta_1 \sigma_w^2 / \sigma_X^2$. Ahora $\sigma_X^2 = \sigma_X^2 + \sigma_w^2$, por lo que $\hat{\beta}_1 \xrightarrow{P} \beta_1 - \beta_1 \sigma_w^2 / (\sigma_X^2 + \sigma_w^2) = [\sigma_X^2 / (\sigma_X^2 + \sigma_w^2)]\beta_1$.

de forma independiente, por lo que $E(w_i|X_i) = 0$, en cuyo caso, $E(v_i|X_i) = 0$, por lo que $\hat{\beta}_1$ es insesgado. Sin embargo, debido a que $\text{var}(v_i) > \text{var}(u_i)$, la varianza de $\hat{\beta}_1$ es mayor de lo que lo sería sin error de medición. En el ejemplo de las calificaciones en los exámenes/tamaño de las clases, supongamos que las calificaciones en los exámenes presentan errores de calificación puramente aleatorios que son independientes de los regresores; entonces el modelo de error de medición clásico de este párrafo es aplicable a \tilde{Y}_i , y $\hat{\beta}_1$ es insesgado. De forma más general, el error de medición en Y que presenta una media condicional igual a cero dados los regresores no provocará sesgo en los coeficientes MCO.

Soluciones para el sesgo por errores en las variables. La mejor manera de resolver el problema de los errores en las variables consiste en obtener una medida precisa de X . No obstante, si esto es imposible, se pueden utilizar métodos econométricos para mitigar el sesgo de errores en las variables.

Uno de estos métodos es la regresión de variables instrumentales. Se basa en disponer de otra variable (la variable «instrumental») que está correlacionada con el valor real X_i , pero que no está correlacionada con el error de medición. Este método se estudia en el Capítulo 12.

Un segundo método consiste en desarrollar un modelo matemático para el error de medición y, si es posible, utilizar las fórmulas resultantes para ajustar las estimaciones. Por ejemplo, si una investigadora cree que es aplicable el modelo clásico del error de medición y si conoce o puede estimar el cociente σ_w^2/σ_X^2 , entonces ella puede utilizar la Ecuación (9.2) para calcular un estimador de β_1 que corrige el sesgo a la baja. Debido a que este método requiere conocimientos especializados acerca de la naturaleza del error de medición, los detalles suelen ser específicos para una determinada base de datos, así como sus problemas de medición, y no se continuará este método en este libro.

Datos perdidos y selección muestral

Los datos perdidos son una característica común de las bases de datos económicos. El hecho de que los datos perdidos representen una amenaza a la validez interna o no depende de la razón por la cual esos datos se consideran perdidos. Se consideran tres casos: cuando los datos faltan de forma totalmente aleatoria, cuando la pérdida de datos se basa en X , y cuando los datos faltan debido a un proceso de selección que está relacionado con Y además de depender de X .

Cuando los datos perdidos son totalmente aleatorios, es decir, por causas aleatorias no relacionadas con los valores de X o de Y , el efecto es el de una reducción del tamaño de la muestra, pero no se introduce sesgo. Por ejemplo, supongamos que se lleva a cabo un muestreo aleatorio simple de 100 compañeros de clase, y se pierden la mitad de los datos al azar. Sería como si nunca se hubiese encuestado a esos alumnos. Qedaría una muestra aleatoria simple de 50 compañeros de clase, por lo que el haber perdido aleatoriamente los datos no introduce sesgo.

Cuando la pérdida de los datos se basa en el valor de un regresor, el efecto es asimismo el de la reducción del tamaño de la muestra, pero no se introduce un sesgo. Por ejemplo, en el ejemplo del tamaño de las clases/ratio estudiantes-maestros, supongamos que se han utilizado solamente los distritos para los que la relación estudiantes-maestros es superior a 20. Aunque no sea posible extraer conclusiones sobre lo que sucede cuando $REM \leq 20$, esto no introduciría sesgo en el análisis del efecto del tamaño de las clases para los distritos con $REM > 20$.

En contraste con los dos primeros casos, si los datos se perdieron debido a un proceso de selección que está relacionado con el valor de la variable dependiente (Y), además de depender de los regresores (X), entonces este proceso de selección puede introducir correlación entre el término de error y los regresores. El sesgo resultante en el estimador MCO se denomina **sesgo de selección muestral**. Se ofreció un ejemplo de sesgo de selección muestral en las votaciones en el recuadro «¡Landon gana!» de la Sección 3.1. En este ejemplo, el método de selección de la muestra (selección aleatoria de números de teléfono de los propietarios de automóviles) está relacionado con la variable dependiente (personas que lo apoyaron para la presidencia en 1936), debido a que en 1936 los propietarios de automóviles con teléfono eran más propensos a ser republicanos. El problema de selección muestral podría ser clasificado tanto como una consecuencia del muestreo no aleatorio como un problema de datos perdidos. En el ejemplo de las votaciones de 1936, la muestra era una muestra aleatoria de propietarios de automóviles con teléfono, no una muestra aleatoria de

**CONCEPTO
CLAVE**

9.5

Sesgo de selección muestral

El sesgo de selección muestral se presenta cuando el proceso de selección influye en la disponibilidad de los datos y el proceso está relacionado con la variable dependiente, además de depender de los regresores. La selección muestral induce correlación entre uno o más regresores y el término de error, lo que da lugar al sesgo y a la inconsistencia el estimador MCO.

los votantes. Por otra parte, este ejemplo puede ser tratado como un problema de datos perdidos, imaginando una muestra aleatoria de votantes, pero en la que se han perdido los datos de los que no tienen automóviles y teléfono. El mecanismo por el cual se pierden los datos está relacionado con la variable dependiente, lo que lleva a un sesgo de selección muestral.

El recuadro «¿Los fondos de inversión baten al mercado?» proporciona un ejemplo de sesgo de selección muestral en economía financiera. El sesgo de selección muestral está resumido en el Concepto clave 9.5³.

Soluciones al sesgo de selección. Los métodos que se han tratado hasta ahora no pueden eliminar el sesgo de selección muestral. Estos métodos para la estimación de modelos con selección muestral quedan fuera del alcance de este libro. Estos métodos se basan en las técnicas que se introducen en el Capítulo 11, en el que se ofrecen más referencias.

¿Los fondos de inversión baten al mercado?

Los fondos de inversión son vehículos de inversión que manejan una cartera de acciones. Comprando participaciones de un fondo de inversión colectiva, un pequeño inversor puede mantener una cartera ampliamente diversificada, sin la molestia y el gasto (costes de transacción) de comprar y vender acciones de sociedades anónimas individuales. Algunos fondos de inversión simplemente siguen la trayectoria del mercado (por ejemplo, manteniendo acciones de las empresas que forman parte del S & P 500), mientras que otros están gestionados activamente por profesionales a tiempo completo, cuyo trabajo es hacer que el fondo obtenga un rendimiento mejor que el conjunto del mercado y que los fondos de la competencia. ¿Pero estos fondos tan activamente gestionados logran este objetivo? ¿Baten los fondos de inversión de forma sistemática a otros fondos y al mercado?

Una forma de responder a estas preguntas consiste en comparar las rentabilidades futuras de los fondos de inversión que obtuvieron un alto rendimiento en el último año con las rentabilidades futuras de otros fondos y del conjunto del mercado. Al realizar estas comparaciones, los economistas financieros saben que es importante seleccionar la muestra de fondos de inversión con cuidado. Sin embargo, esta tarea no es tan sencilla como parece. Algunas bases de datos incluyen datos históricos sobre los fondos que se encuentran disponibles para comprar en la actualidad, pero este método significa que los perros —los fondos de inversión que peor se han comportado— se han omitido de la base de datos debido a que salieron del negocio o se fusionaron con otros fondos. Por esta razón,

un estudio que utilice datos sobre el comportamiento histórico de los fondos que se encuentran disponibles en la actualidad está sujeto al sesgo de selección muestral: la muestra se selecciona en base al valor de la variable dependiente, la rentabilidad, debido a que los fondos con los rendimientos más bajos se han eliminado. La rentabilidad media de todos los fondos (incluyendo los difuntos) para un periodo de diez años será menor que la rentabilidad media de los fondos que todavía existan al final de esos diez años, por lo que un estudio que incluya solamente a estos últimos fondos sobreestimará su comportamiento. Los economistas financieros se refieren a este sesgo de selección mediante la expresión «sesgo de supervivencia», debido a que solo los mejores fondos sobreviven para estar en la base de datos.

Cuando los especialistas en econometría financiera corren el sesgo de supervivencia incorporando los datos de los fondos difuntos, los resultados no pintan un retrato favorecedor de los administradores de fondos. Corregido el sesgo de supervivencia, la evidencia econométrica indica que los fondos de inversión activamente gestionados no superan al mercado en media y que el buen comportamiento pasado no predice un buen comportamiento en el futuro. Para una lectura más amplia sobre los fondos de inversión y el sesgo de supervivencia, véase Malkiel (2003, Capítulo 11) y Carhart (1997). El problema del sesgo de supervivencia asimismo se plantea en las evaluaciones del comportamiento de los fondos de cobertura; para una lectura más amplia, véase Aggarwal y Jorion (2010).

³ El Ejercicio 18.16 proporciona un tratamiento matemático para los tres casos de datos perdidos tratados aquí.

Causalidad simultánea

Hasta ahora, hemos supuesto que la causalidad va desde las variables explicativas hacia la variable dependiente (X causa a Y). Pero ¿y si la causalidad funciona del mismo modo desde la variable dependiente hacia una o más variables explicativas (Y causa a X)? Si es así, la causalidad va «hacia atrás», del mismo modo que hacia adelante; es decir, existe **causalidad simultánea**. Si existe causalidad simultánea, una regresión MCO recogerá ambos efectos, por lo que el estimador MCO será sesgado e inconsistente.

Por ejemplo, nuestro estudio de las calificaciones en los exámenes estaba centrado en el efecto sobre las calificaciones en los exámenes de la reducción de la ratio estudiantes-maestros, por lo que la causalidad se presume que parte desde la ratio estudiantes-maestros y va hacia las calificaciones. Supongamos, sin embargo, que una iniciativa del gobierno subsidia la contratación de maestros en los distritos escolares con peores puntuaciones. Si es así, la causalidad funcionaría en ambos sentidos: por las razones educativas habituales será razonable pensar que las ratios bajas de estudiantes por maestro conduzcan a calificaciones altas en los exámenes, pero debido al programa del gobierno las calificaciones bajas llevarían asimismo a bajas ratios de estudiantes por maestro.

La causalidad simultánea conduce a la correlación entre el regresor y el término de error. En el ejemplo de las calificaciones en los exámenes, supongamos que existe un factor omitido que lleva a bajas calificaciones en los exámenes; debido al programa del gobierno, este factor que produce las bajas calificaciones a su vez conduce a una ratio de estudiantes-maestros baja. Por lo tanto un término de error negativo en la regresión poblacional de las calificaciones en los exámenes sobre la ratio estudiantes-maestros reduce las calificaciones en los exámenes, pero debido al programa del gobierno conduce asimismo a una disminución de la ratio estudiantes-maestros. En otras palabras, la ratio estudiantes-maestros está correlacionada positivamente con el término de error en la regresión poblacional. Esto a su vez conduce a un sesgo de causalidad simultánea y a la inconsistencia del estimador MCO.

Esta correlación entre el término de error y el regresor se puede precisar matemáticamente mediante la introducción de una ecuación adicional que describa el vínculo causal inverso. Para mayor comodidad, se consideran solamente las dos variables X e Y , y se ignoran los otros posibles regresores. De acuerdo con esto, existen dos ecuaciones, una en la que X causa a Y y una en la que Y causa a X :

$$Y_i = \beta_0 + \beta_1 X_i + u_i \text{ y} \quad (9.3)$$

$$X_i = \gamma_0 + \gamma_1 Y_i + v_i. \quad (9.4)$$

La Ecuación (9.3) es la habitual en la que β_1 es el efecto sobre Y de una variación en X , donde u representa los otros factores. La Ecuación (9.4) representa el efecto causal inverso de Y sobre X . En el problema de las calificaciones en los exámenes, la Ecuación (9.3) representa el efecto educativo del tamaño de las clases sobre las calificaciones obtenidas, mientras que la Ecuación (9.4) representa el efecto causal inverso de las calificaciones en los exámenes sobre el tamaño de las clases inducido por el programa del gobierno.

La causalidad simultánea conduce a la correlación entre X_i y el término de error u_i en la Ecuación (9.3). Para comprobarlo, imaginemos que u_i es negativo, lo que hace disminuir Y_i . Sin embargo, este menor valor de Y_i afecta al valor de X_i a través de la segunda de estas ecuaciones, y si γ_1 es positivo, un valor bajo de Y_i llevará a un valor bajo de X_i . Por tanto, si γ_1 es positivo, X_i y u_i estarán positivamente correlacionados⁴.

Debido a que esto puede expresarse matemáticamente mediante dos ecuaciones simultáneas, el sesgo de causalidad simultánea a veces se denomina **sesgo de ecuaciones simultáneas**. El sesgo por causalidad simultánea se resume en el Concepto clave 9.6.

Soluciones al sesgo de causalidad simultánea. Existen dos maneras de mitigar el sesgo de causalidad simultánea. Una de ellas es mediante la regresión por variables instrumentales, el tema del Capítulo 12. La segunda consiste en diseñar y llevar a cabo un experimento aleatorizado controlado en el que se anule el canal de la causalidad inversa, y los experimentos de este tipo se tratan en el Capítulo 13.

⁴ Para demostrarlo matemáticamente, téngase en cuenta que la Ecuación (9.4) implica que $\text{cov}(X_i, u_i) = \text{cov}(\gamma_0 + \gamma_1 Y_i + v_i, u_i) = \gamma_1 \text{cov}(Y_i, u_i) + \text{cov}(v_i, u_i)$. Suponiendo que $\text{cov}(v_i, u_i) = 0$, por la Ecuación (9.3) esto a su vez implica que $\text{cov}(X_i, u_i) = \gamma_1 \text{cov}(Y_i, u_i) = \gamma_1 \text{cov}(\beta_0 + \beta_1 X_i + u_i, u_i) = \gamma_1 \beta_1 \text{cov}(X_i, u_i) + \gamma_1 \sigma_u^2$. Resolviendo para $\text{cov}(X_i, u_i)$ se obtiene entonces el resultado $\text{cov}(X_i, u_i) = \gamma_1 \sigma_u^2 / (1 - \gamma_1 \beta_1)$.

**CONCEPTO
CLAVE**

9.6

Sesgo por causalidad simultánea

El sesgo por causalidad simultánea, asimismo denominado sesgo de ecuaciones simultáneas, aparece en una regresión de Y sobre X , cuando, además del vínculo causal de interés que va desde X hacia Y , existe un vínculo causal desde Y hacia X . Esta causalidad inversa provoca que X esté correlacionado con el término de error en la regresión poblacional de interés.

Origen de la inconsistencia de los errores estándar MCO

Los errores estándar inconsistentes representan una amenaza diferente para la validez interna. Incluso aunque el estimador MCO sea consistente y la muestra sea grande, la inconsistencia de los errores estándar origina que los contrastes de hipótesis presenten un tamaño distinto del nivel de significación deseado, así como que los intervalos de confianza al «95 %» no incluyan al verdadero valor en el 95 % de las muestras repetidas.

Existen dos razones principales para la inconsistencia de los errores estándar: un tratamiento no adecuado de la heterocedasticidad y la correlación del término de error entre observaciones.

Heterocedasticidad. Como se trató en la Sección 5.4, debido a razones históricas, algunos paquetes informáticos de regresión solamente presentan errores estándar válidos con homocedasticidad. Si, no obstante, el error de la regresión es heterocedástico, estos errores estándar no constituyen una base fiable para los contrastes de hipótesis y los intervalos de confianza. La solución a este problema consiste en utilizar los errores estándar heterocedástico-robustos y construir los estadísticos F utilizando un estimador de la varianza heterocedástico-robusto. Los errores estándar heterocedástico-robustos están disponibles como opción en los paquetes de software modernos.

Correlación del término de error entre observaciones. En algunos casos, el error de regresión poblacional puede estar correlacionado entre observaciones. Esto no sucederá si los datos se extraen de una población mediante muestreo aleatorio, debido a que la aleatoriedad en el proceso de muestreo asegura que los errores estén distribuidos de forma independiente entre una observación y la siguiente. A veces, sin embargo, el muestreo tan solo es aleatorio parcialmente. La circunstancia más común es cuando los datos son observaciones repetidas del mismo individuo en el tiempo, tal como el mismo distrito escolar en diferentes años. Si las variables omitidas que forman parte del error de regresión son persistentes (como las características demográficas del distrito), entonces se induce correlación «serial» en el error de regresión a lo largo del tiempo. La correlación serial en el término de error puede aparecer en los datos de panel (datos sobre varios distritos para varios años) y en los datos de series temporales (datos sobre un solo distrito durante varios años).

Otra situación en la que el término de error puede estar correlacionado entre las distintas observaciones es cuando el muestreo está basado en una unidad geográfica. Si existen variables omitidas que reflejan las influencias geográficas, estas variables podrían dar lugar a la correlación entre los errores de regresión para observaciones adyacentes.

La correlación del error de regresión entre las distintas observaciones no hace que el estimador MCO sea sesgado o inconsistente, pero viola el segundo supuesto de mínimos cuadrados del Concepto clave 6.4. La consecuencia es que los errores estándar MCO, tanto los válidos con homocedasticidad como los heterocedástico-robustos, son incorrectos en el sentido de que no dan lugar a intervalos de confianza con el nivel de confianza deseado.

En muchos casos, este problema se puede solucionar mediante el uso de una fórmula alternativa para los errores estándar. En el Capítulo 10 se proporcionan fórmulas para el cálculo de errores estándar que son robustos tanto en presencia de heterocedasticidad como de correlación serial en regresión con datos de panel, y en el Capítulo 15, para la regresión con datos de series temporales.

En el Concepto clave 9.7 se resumen las amenazas a la validez interna de un estudio de regresión múltiple.

**CONCEPTO
CLAVE**
9.7
Amenazas a la validez interna de un estudio de regresión múltiple

Existen cinco amenazas principales a la validez interna de un estudio de regresión múltiple:

1. Variables omitidas.
2. Error de especificación de la forma funcional.
3. Errores en las variables (errores de medición en las variables explicativas).
4. Selección muestral.
5. Causalidad simultánea.

Cada uno de ellos, si está presente, se traduce en el incumplimiento del primer supuesto de mínimos cuadrados, $E(u_i|X_{1i}, \dots, X_{ki}) \neq 0$, lo que a su vez significa que el estimador MCO es sesgado e inconsistente.

El cálculo incorrecto de los errores estándar representa asimismo una amenaza a la validez interna. Los errores estándar válidos con homocedasticidad no son válidos en presencia de heterocedasticidad. Si las variables no son independientes entre distintas observaciones, lo cual puede ocurrir en datos de panel y en datos de series temporales, entonces se necesita un nuevo ajuste en la fórmula de los errores estándar a fin de obtener errores estándar válidos.

La aplicación de esta lista de amenazas a un estudio de regresión múltiple constituye un método sistemático de evaluar la validez interna del estudio.

9.3 Validez interna y externa cuando la regresión se utiliza para predicción

Hasta ahora, el estudio del análisis de regresión múltiple se ha centrado en la estimación de los efectos causales. No obstante, los modelos de regresión pueden utilizarse para otros fines entre los que se incluye la predicción. Cuando los modelos de regresión se utilizan para predicción, la preocupación acerca de la validez externa es muy importante, pero la preocupación acerca de la estimación insesgada de los efectos causales no lo es.

Utilización de modelos de regresión para predicción

El Capítulo 4 empezaba examinando el problema de una directora escolar que quería saber cuánto podrían aumentar las calificaciones en los exámenes si se redujera el tamaño de las clases de su distrito escolar; es decir, la directora quería conocer el efecto causal sobre las calificaciones en los exámenes de una variación en el tamaño de las clases. De acuerdo con esto, los Capítulos 4 a 8 se centraron en la utilización del análisis de regresión para estimar los efectos causales a partir de datos observacionales.

Se considera ahora un problema diferente. Un padre que se muda a un área metropolitana planea elegir dónde vivir basándose en parte en la calidad de las escuelas locales. Al padre le gustaría saber cómo se comportan los diferentes distritos escolares a la hora de llevar a cabo las pruebas estandarizadas. Supongamos, no obstante, que los datos sobre las calificaciones en los exámenes de la prueba no están disponibles (tal vez son confidenciales) pero sí lo están los datos sobre el tamaño de las clases. En esta situación, el padre debe adivinar el desempeño en la realización del examen estandarizado en los diferentes distritos basándose en una cantidad limitada de información. Es decir, el problema del padre consiste en predecir las calificaciones medias en los exámenes de un distrito dado en base a la información relacionada con las calificaciones en los exámenes, en particular, el tamaño de las clases.

¿Cómo puede realizar el padre este pronóstico? Recordemos la regresión de las calificaciones en los exámenes sobre la ratio estudiantes-maestros (REM) del Capítulo 4:

$$\widehat{\text{CalificaciónExamen}} = 698,9 - 2,28 \times \text{REM}. \quad (9.5)$$

Llegamos a la conclusión de que esta regresión no resulta útil para la directora: el estimador MCO de la pendiente es sesgado, debido a la omisión de variables tales como la composición del cuerpo de los estudiantes y las otras oportunidades de aprendizaje fuera de la escuela para los estudiantes.

Sin embargo, la Ecuación (9.5) podría ser útil para el padre que está tratando de elegir una casa. Sin duda, el tamaño de la clase no es el único factor determinante del rendimiento en los exámenes de la prueba, pero desde la perspectiva del padre lo que importa es si es un predictor fiable del desempeño (calificaciones) en la prueba. Al padre interesado en la predicción de las calificaciones en los exámenes no le importa si el coeficiente de la Ecuación (9.5) estima el efecto causal sobre las calificaciones en los exámenes del tamaño de la clase. Por el contrario, el padre simplemente desea que la regresión explique gran parte de la variación de las calificaciones entre los distintos distritos y que sea estable, es decir, que sea aplicable a los distritos a los que el padre está considerando mudarse. Aunque el sesgo de variable omitida hace que la Ecuación (9.5) no tenga valor a la hora de responder a la cuestión de la causalidad, todavía puede resultar útil para fines de pronóstico.

De manera más general, los modelos de regresión pueden originar previsiones fiables, aunque sus coeficientes no tengan una interpretación causal. Este reconocimiento se encuentra detrás de la utilización de la mayoría de los modelos de regresión con fines predictivos.

Evaluación de la validez de los modelos de regresión para predicción

Debido a que el problema de la directora y el problema del padre son conceptualmente muy diferentes, los requisitos para la validez de la regresión son diferentes para sus respectivos problemas. Para obtener estimaciones fiables de los efectos causales, se deben afrontar las amenazas a la validez interna que se recogen en el Concepto clave 9.7.

Por el contrario, si se pretende obtener un pronóstico fiable, la regresión estimada debe poseer un buen poder explicativo, sus coeficientes deben ser estimados con precisión, y debe ser estable en el sentido de que la regresión estimada sobre un conjunto de datos pueda ser fiable para efectuar predicciones a partir de otros datos. Cuando se utiliza un modelo de regresión para predecir, una preocupación primordial es que el modelo sea externamente válido en el sentido de que sea estable y cuantitativamente aplicable a las circunstancias bajo las cuales se realiza la predicción. En la Parte IV, volvemos al problema de evaluar la validez de un modelo de regresión para predecir los valores futuros de los datos de series temporales.

9.4 Ejemplo: calificaciones y tamaño de las clases

El marco de validez interna y externa ayuda a tener una perspectiva crítica acerca de lo que se ha aprendido, y lo que no, en nuestro análisis de los datos sobre las calificaciones en los exámenes de California.

Validez externa

La cuestión de si el análisis para California puede generalizarse, es decir, si es externamente válido, depende de la población y el escenario para los que se realiza la generalización. En este caso, se considera si los resultados pueden generalizarse al desempeño en otros exámenes estandarizados de otros distritos de escuelas primarias públicas en los Estados Unidos.

En la Sección 9.1 se señalaba que el hecho de disponer de más de un estudio sobre el mismo tema ofrece una oportunidad para evaluar la validez externa de ambos estudios mediante la comparación de sus resultados. En el caso de las calificaciones en los exámenes estandarizados y el tamaño de las clases, se dispone en realidad de otras bases de datos comparables. En esta sección, se analiza una base de datos diferente, basada en los resultados en los exámenes estandarizados para los alumnos de cuarto curso en 220 distritos de escuelas públicas de Massachusetts en 1998. Tanto en Massachusetts como en California los exámenes proporcionan unas medidas amplias acerca de los conocimientos y habilidades académicas de los estudiantes, si bien los detalles difieren. Asimismo, la organización de la educación en el aula es muy similar en los dos estados a nivel de escuela primaria (como lo es en la mayoría de los distritos escolares de educación prima-

ria de EE.UU.), aunque los aspectos de financiación de las escuelas primarias y plan de estudios difieran. Por tanto, la obtención de resultados similares acerca del efecto de la ratio estudiantes-maestros sobre el desempeño en la prueba con los datos de California y Massachusetts constituiría una prueba de la validez externa de los hallazgos para California. Por el contrario, la obtención de resultados diferentes para los dos estados plantearía preguntas acerca de la validez interna o externa de al menos uno de los estudios.

Comparación de los datos de California y Massachusetts. Al igual que los datos de California, los datos de Massachusetts son para el nivel de distrito escolar. Las definiciones de las variables en la base de datos de Massachusetts son las mismas que en la base de datos de California, o prácticamente las mismas. Se ofrece más información sobre la base de datos de Massachusetts, incluyendo las definiciones de las variables, en el Apéndice 9.1.

La Tabla 9.1 presenta los estadísticos de resumen para las muestras de California y Massachusetts. La puntuación promedio es más alta en Massachusetts, pero el examen es diferente, por lo que una comparación directa de los resultados no resulta apropiada. El promedio de la proporción de alumnos por maestro es mayor en California (19,6 frente a 17,3). La renta media por distrito es un 20 % mayor en Massachusetts, pero la desviación típica de la renta es mayor en California; es decir, hay una mayor dispersión en las rentas medias de los distritos en California que en Massachusetts. El porcentaje promedio de estudiantes que aún están aprendiendo inglés, así como el porcentaje promedio de estudiantes que reciben subvenciones para comedor, son ambos mucho más altos en California que en los distritos de Massachusetts.

Calificaciones y renta media del distrito. Por motivos de espacio, no se presentan los diagramas de dispersión para todos los datos de Massachusetts. Debido a que era el tema central del Capítulo 8, no obstante, resulta interesante examinar la relación entre las calificaciones en los exámenes y la renta media del distrito en Massachusetts. Este diagrama de dispersión se presenta en la Figura 9.1. El patrón general de este diagrama de dispersión es similar al de la Figura 8.2 para los datos de California: la relación entre la renta y las calificaciones parece ser más inclinada en los valores de renta bajos y se hace más plana en los valores altos. Evidentemente, la regresión lineal representada en el gráfico no recoge esta no linealidad aparente. Las funciones de regresión cúbica y logarítmica se representan asimismo en la Figura 9.1. La función de regresión cúbica presenta un R^2 ligeramente superior al de la especificación logarítmica (0,486 frente a 0,455). La comparación de las Figuras 8.7 y 9.1 muestra que el patrón general de no linealidad hallado en la renta de California y los datos de las calificaciones en los exámenes está presente asimismo en los datos de Massachusetts. Las formas funcionales precisas que mejor describen esta no linealidad difieren; no obstante, la especificación cúbica ajusta mejor en Massachusetts, pero la especificación lineal-log ajusta mejor en California.

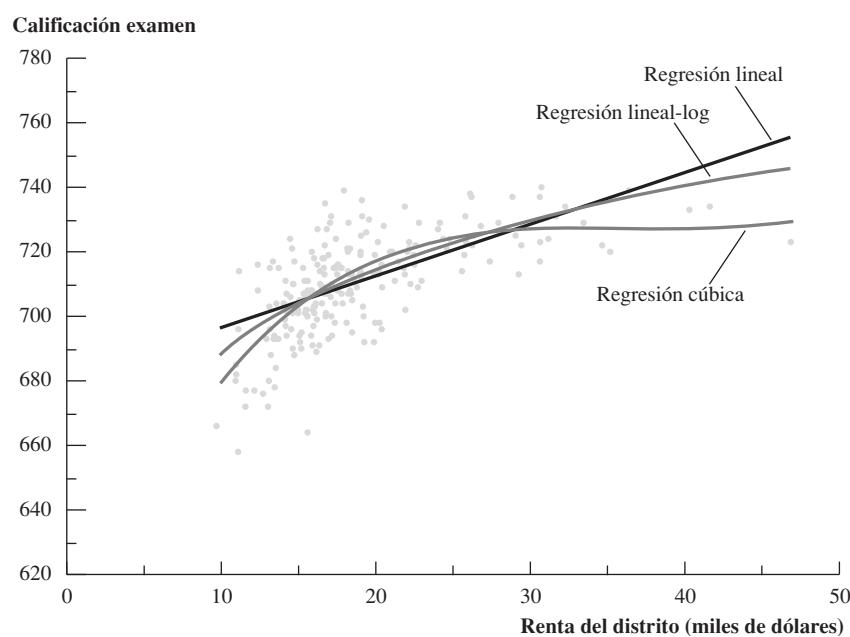
Resultados de regresión múltiple. Los resultados de regresión para los datos de Massachusetts se presentan en la Tabla 9.2. La primera regresión, presentada en la columna (1) de la tabla, solo tiene como

TABLA 9.1 Estadísticos de resumen para los datos de las calificaciones en California y Massachusetts

	California		Massachusetts	
	Media	Desviación típica	Media	Desviación típica
Calificaciones examen	654,1	19,1	709,8	15,1
Ratio estudiantes-maestros	19,6	1,9	17,3	2,3
% estudiantes aprendiendo inglés	15,8 %	18,3 %	1,1 %	2,9 %
% receptores subsidio comedor	44,7 %	27,1 %	15,3 %	15,1 %
Renta media distrito (\$)	15.317 \$	7.226 \$	18.747 \$	5.808 \$
Número de observaciones	420		220	
Año	1999		1998	

FIGURA 9.1 Calificaciones vs. renta para los datos de Massachusetts

La función de regresión lineal estimada no capta la relación no lineal entre la renta y las calificaciones en los datos de Massachusetts. Las funciones de regresión estimadas lineal-log y cúbica son similares para las rentas del distrito entre 13.000 \$ y 30.000 \$, el rango que contiene la mayoría de las observaciones.



regresor la ratio estudiantes-maestros. La pendiente es negativa ($-1,72$), y la hipótesis de que el coeficiente es igual a cero puede ser rechazada al nivel de significación del 1 % ($t = -1,72/0,50 = -3,44$).

Las columnas restantes presentan los resultados de la inclusión de variables adicionales que controlan por las características de los estudiantes y de la inclusión de no linealidades en la función de regresión estimada. Controlar por el porcentaje de estudiantes aprendiendo Inglés, el porcentaje de estudiantes que cumplen los requisitos para la gratuidad del comedor, y la renta media del distrito reduce el coeficiente estimado para la ratio estudiantes-maestros en un 60 %, pasando de $-1,72$ en la regresión (1) a $-0,69$ en la regresión (2) y $-0,64$ en la regresión (3).

Comparando los \bar{R}^2 de las regresiones (2) y (3) se concluye que la especificación cúbica (3) proporciona un modelo mejor acerca de la relación entre las calificaciones en los exámenes y la renta que la especificación logarítmica (2), incluso manteniendo constante la ratio estudiantes-maestros. No existe evidencia estadísticamente significativa de una relación no lineal entre las calificaciones en los exámenes y la ratio estudiantes-maestros: el estadístico F de la regresión (4) para el contraste de que los coeficientes poblacionales de REM^2 y REM^3 son iguales a cero presenta un p -valor de 0,641. Del mismo modo, no existe evidencia de que una reducción en la ratio estudiantes-maestros tenga un efecto diferente en los distritos con muchos estudiantes que están aprendiendo inglés que en los que tienen unos pocos [el estadístico t para $EIA \times REM$ en la regresión (5) es $0,80/0,56 = 1,43$]. Finalmente, la regresión (6) muestra que el coeficiente estimado de la ratio estudiantes-maestros no cambia sustancialmente cuando el porcentaje de estudiantes que están aprendiendo inglés [que no es significativo en la regresión (3)] se excluye. En resumen, los resultados de la regresión (3) no son sensibles a los cambios en la forma funcional y la especificación considerada en las regresiones (4) a (6) de la Tabla 9.2. Por tanto, se adopta la regresión (3) como la estimación base del efecto en las calificaciones en los exámenes de un cambio en la ratio estudiantes-maestros, sobre la base de los datos de Massachusetts.

Comparación de los resultados de Massachusetts y California. Para los datos de California, se obtiene lo siguiente:

1. La adición de variables que controlan por las características de los estudiantes reduce el coeficiente de la ratio estudiantes-maestros desde $-2,28$ [regresión (1), Tabla 7.1] hasta $-0,73$ [regresión (2), Tabla 8.3], una reducción del 68 %.

TABLA 9.2 Estimaciones de regresión múltiple del ratio estudiantes-maestros y calificaciones en los exámenes: datos de Massachusetts

Variable dependiente: media combinada de las calificaciones en el examen de inglés, matemáticas, y ciencias del distrito escolar, cuarto curso; 220 observaciones.

Regresor	(1)	(2)	(3)	(4)	(5)	(6)
Ratio estudiantes-maestros (REM)	-1,72** (0,50)	-0,69* (0,27)	-0,64* (0,27)	12,4 (14,0)	-1,02** (0,37)	-0,67* (0,27)
REM^2				-0,680 (0,737)		
REM^3				0,011 (0,013)		
% estudiantes aprendiendo inglés	-0,411 (0,306)	-0,437 (0,303)	-0,434 (0,300)			
% estudiantes aprendiendo inglés > mediana? (Binaria, EIA)					-12,6 (9,8)	
EIA × REM					0,80 (0,56)	
% cumplimiento requisitos gratuidad comedor	-0,521** (0,077)	-0,582** (0,097)	-0,587** (0,104)	-0,709** (0,091)	-0,653** (0,72)	
Renta distrito (logaritmo)	16,53** (3,15)					
Renta distrito		-3,07 (2,35)	-3,38 (2,49)	-3,87* (2,49)	-3,22 (2,31)	
Renta distrito ²		0,164 (0,085)	0,174 (0,089)	0,184* (0,090)	0,165 (0,085)	
Renta distrito ³		-0,0022* (0,0010)	-0,0023* (0,0010)	-0,0023* (0,0010)	-0,0022* (0,0010)	
Intercepto	739,6** (8,6)	682,4** (11,5)	744,0** (21,3)	665,5** (81,3)	759,9** (23,2)	747,4** (20,3)
Todas las variables REM e interacciones = 0				2,86 (0,038)	4,01 (0,020)	
$REM^2, REM^3 = 0$				0,45 (0,641)		
$Renta^2, Renta^3$			7,74 (< 0,001)	7,75 (< 0,001)	5,85 (0,003)	6,55 (0,002)
$EIA, EIA \times REM$					1,58 (0,208)	
ESR	14,64	8,69	8,61	8,63	8,62	8,64
\bar{R}_2	0,063	0,670	0,676	0,675	0,675	0,674

Estas regresiones fueron estimadas utilizando los datos de los distritos de escuela primaria de Massachusetts descritos en el Apéndice 9.1. Los errores estándar se presentan entre paréntesis debajo de los coeficientes, y los *p*-valores se indican entre paréntesis debajo de los estadísticos F. Los coeficientes individuales son estadísticamente significativos al nivel del 5 % * o al nivel del 1 % **.

2. La hipótesis de que el verdadero coeficiente de la ratio estudiantes-maestros es igual a cero se rechaza al nivel de significación del 1 %, incluso tras la adición de variables que controlan por las características de los estudiantes y las características económicas del distrito.
3. El efecto de reducir la ratio estudiantes-maestros no depende de manera importante del porcentaje de estudiantes aprendiendo inglés en el distrito.
4. Existe cierta evidencia de que la relación entre las calificaciones en los exámenes y la ratio estudiantes-maestros es no lineal.

¿Se puede obtener lo mismo para Massachusetts? De acuerdo con los resultados de (1), (2) y (3), la respuesta es sí. La inclusión de variables de control adicionales reduce el coeficiente de la ratio estudiantes-maestros desde $-1,72$ [regresión (1), Tabla 9.2] hasta $-0,69$ [regresión (2), Tabla 9.2], una reducción del 60 %. Los coeficientes de la ratio estudiantes-maestros siguen siendo significativos después de agregar las variables de control. Estos coeficientes solamente son significativos al nivel del 5 % con los datos de Massachusetts, mientras que son significativos al nivel del 1 % con los datos de California. Sin embargo, hay casi el doble de observaciones en los datos de California, por lo que no es de extrañar que las estimaciones para California sean más precisas. Al igual que en los datos de California, no existe evidencia estadísticamente significativa en los datos de Massachusetts acerca de una interacción entre la ratio estudiantes-maestros y la variable binaria que indica un alto porcentaje de estudiantes aprendiendo inglés en el distrito.

Sin embargo, los resultados de (4) no se mantienen para los datos de Massachusetts: la hipótesis de que la relación entre la ratio estudiantes-maestros y las calificaciones en los exámenes es lineal no puede rechazarse al 5 % de nivel de significación, cuando se contrasta frente a una especificación cúbica.

Debido a que los dos exámenes estandarizados son diferentes, los coeficientes no se pueden comparar directamente: un punto en los exámenes de Massachusetts no es igual a un punto en los exámenes de California. Si, no obstante, las calificaciones se expresan en las mismas unidades, entonces los efectos estimados del tamaño de las clases se pueden comparar. Una forma de hacerlo es transformando las calificaciones en el examen mediante su estandarización: restando la media muestral y dividiendo por la desviación típica, por lo que presentan una media de 0 y una varianza de 1. Los coeficientes de las pendientes de la regresión con las calificaciones transformadas son iguales a los coeficientes de las pendientes de la regresión original divididos por la desviación típica del examen. Por tanto, el coeficiente de la ratio estudiantes-maestros, dividido por la desviación típica de las calificaciones en los exámenes se pueden comparar entre los dos conjuntos de datos.

Esta comparación se lleva a cabo en la Tabla 9.3. La primera columna muestra las estimaciones MCO del coeficiente de la ratio estudiantes-maestros en una regresión con el porcentaje de estudiantes aprendiendo inglés, el porcentaje de estudiantes que cumplen los requisitos para la gratuidad del comedor, y la renta media del distrito como variables de control incluidas. La segunda columna presenta la desviación típica de las calificaciones entre distritos. Las dos últimas columnas presentan el efecto estimado sobre las calificaciones en los exámenes de la reducción de la ratio estudiantes-maestros en dos estudiantes por maestro (la propuesta de nuestra directora), en primer lugar en las unidades de puntuación en el examen y en segundo lugar en unidades de desviaciones típicas. Para la especificación lineal, el coeficiente MCO de la estimación con los datos de California es $-0,73$, por lo que la reducción de la ratio estudiantes-maestros en dos se estima que aumente las calificaciones del distrito en el examen en $-0,73 \times (-2) = 1,46$ puntos. Debido a que la desviación típica de las calificaciones en los exámenes es de 19,1 puntos, esto se corresponde con $1,46/19,1 = 0,076$ desviaciones típicas de la distribución de las calificaciones en los exámenes de los distintos distritos. El error estándar de esta estimación es $0,26 \times 2/19,1 = 0,027$. Los efectos estimados para los modelos no lineales y sus errores estándar se calcularon mediante el método que se describe en la Sección 8.1.

Basándonos en el modelo lineal con los datos de California, una reducción en dos estudiantes por maestro se estima que aumente las calificaciones en los exámenes en 0,076 unidades de desviación típica, con un error estándar de 0,027. Los modelos no lineales para los datos de California sugieren un efecto algo mayor, con el efecto específico que es función de la ratio inicial de estudiantes-maestros. De acuerdo con los datos de Massachusetts, este efecto estimado es de 0,085 unidades de desviación típica, con un error estándar de 0,036.

TABLA 9.3

Ratios estudiantes-maestros y calificaciones: comparación de las estimaciones de California y Massachusetts

		Efecto estimado de dos estudiantes menos por maestro, en unidades de:		
	Estimación MCO $\hat{\beta}_{STR}$	Desviación típica de las calificaciones entre distritos	Puntos en el examen	Desviaciones típicas
California				
Lineal: Tabla 8.3(2)	–0,73 (0,26)	19,1	1,46 (0,52)	0,076 (0,027)
Cúbica: Tabla 8.3(7) Reducción REM de 20 a 18	—	19,1	2,93 (0,70)	0,153 (0,037)
Cúbica: Tabla 8.3(7) Reducción REM de 22 a 20	—	19,1	1,90 (0,69)	0,099 (0,036)
Massachusetts				
Lineal: Tabla 9.2(3)	–0,64 (0,27)	15,1	1,28 (0,54)	0,085 (0,036)
Los errores estándar están entre paréntesis.				

Estas estimaciones son esencialmente las mismas. La disminución de la ratio estudiantes-maestros se prevé que aumente las calificaciones en los exámenes, pero la mejora prevista es pequeña. En los datos de California, por ejemplo, la diferencia en las calificaciones entre el distrito mediano y el distrito en el percentil 75 es de 12,2 puntos de calificación en la prueba (Tabla 4.1), o de 0,64 (= 12,2/19,1) desviaciones típicas. El efecto estimado mediante el modelo lineal es un poco más de una décima parte de este tamaño; en otras palabras, de acuerdo con esta estimación, la disminución del ratio estudiantes-maestros en una cuantía de dos trasladaría a un distrito tan solo una décima parte del camino en la dirección desde la media hasta el percentil 75 de la distribución de las calificaciones en los exámenes entre los distritos. La reducción de la ratio estudiantes-maestros en una cuantía de dos supone un gran cambio para un distrito, pero los beneficios estimados como se muestra en la Tabla 9.3, aun siendo distintos de cero, son pequeños.

Este análisis de los datos de Massachusetts sugiere que los resultados de California son externamente válidos, al menos cuando se generalizan a los distritos escolares de primaria de otra parte de los Estados Unidos.

Validez interna

La similitud de los resultados entre California y Massachusetts no garantiza su validez *interna*. La Sección 9.2 enumera cinco posibles amenazas a la validez interna que podrían inducir un sesgo en la estimación del efecto sobre las calificaciones del tamaño de las clases. A continuación, nos volvemos a centrar en estas amenazas.

VARIABLES OMITIDAS. Las regresiones múltiples presentadas en este capítulo y los anteriores controlan por una característica de los estudiantes (el porcentaje de estudiantes aprendiendo inglés), una característica económica familiar (el porcentaje de estudiantes que perciben una subvención de comedor), y una medida más amplia de la capacidad económica del distrito (renta media del distrito).

Si estas variables de control son adecuadas, entonces en lo que respecta al análisis de regresión, es como si la ratio estudiantes-maestros se asignara aleatoriamente entre los distritos con los mismos valores de estas variables de control, en cuyo caso se cumpliría el supuesto de independencia en media condicional. Sin embargo, todavía podrían existir algunos factores omitidos para los que estas tres variables pudieran no

resultar las variables de control adecuadas. Por ejemplo, si la ratio estudiantes-maestros estuviera correlacionada con la calidad del profesorado, incluso entre los distritos con la misma proporción de inmigrantes y las mismas características socioeconómicas (tal vez porque los mejores profesores se sienten atraídos por las escuelas con menor ratio estudiantes-maestros) y si la calidad del profesorado afectara a las calificaciones en los exámenes, entonces la omisión de la variable que mida la calidad del profesorado podría sesgar el coeficiente de la ratio estudiantes-maestros. Del mismo modo, entre los distritos con las mismas características socioeconómicas, los distritos con menores ratios estudiantes-maestros podrían estar habitados por familias que están más comprometidas con la mejora del aprendizaje de sus hijos en el hogar. Tales factores omitidos podrían llevar a un sesgo de variable omitida.

Una forma de eliminar el sesgo de variable omitida, al menos en teoría, es llevar a cabo un experimento. Por ejemplo, los estudiantes podrían ser asignados aleatoriamente a clases de diferentes tamaños, y podría compararse más tarde su desempeño en la realización de los exámenes estandarizados. Un estudio así se llevó a cabo en realidad en Tennessee, y se examina en el Capítulo 13.

Forma funcional. Este análisis y el del Capítulo 8 exploran varias formas funcionales. Hallamos que algunas de las posibles no linealidades que fueron investigadas no resultaban estadísticamente significativas, mientras que otras que sí lo eran no alteraban de forma sustancial la estimación del efecto de la reducción de la ratio estudiantes-maestros. Aunque podría llevarse a cabo adicionalmente el análisis de otras formas funcionales, esto sugería que no era probable que las principales conclusiones de estos estudios fueran sensibles a la utilización de diferentes especificaciones no lineales para la regresión.

Errores en las variables. El promedio de la ratio estudiantes por maestro en el distrito es una medida amplia y potencialmente inexacta del tamaño de las clases. Por ejemplo, debido a que los estudiantes se trasladan desde y hacia los diferentes distritos, la ratio estudiantes-maestros podría no expresar de forma precisa el tamaño real de las clases disfrutado por los estudiantes que realizan la prueba, lo que a su vez podría provocar una estimación del efecto del tamaño de las clases sesgado hacia cero. Otra variable con un error de medición potencial es la renta media del distrito. Estos datos se tomaron del censo de 1990, mientras que el resto de datos corresponden a 1998 (Massachusetts) o a 1999 (California). Si la composición económica del distrito cambió sustancialmente durante la década de 1990, esto podría ser una medida imprecisa de la renta media del distrito en la realidad.

Selección. Los datos de California y de Massachusetts cubren todos los distritos escolares de educación primaria pública de estos estados que satisfacen restricciones mínimas de tamaño, por lo que no hay razón para creer que exista aquí un problema de selección muestral.

Causalidad simultánea. La causalidad simultánea se presentaría si el desempeño en los exámenes estandarizados afectara a la ratio estudiantes-maestros. Esto podría ocurrir, por ejemplo, si hubiese un mecanismo burocrático o político que aumentara la financiación de las escuelas o de los distritos con peores resultados, lo que a su vez diera lugar a la contratación de más maestros. En Massachusetts, no existía tal mecanismo de igualación de la financiación de las escuelas durante el tiempo de realización de estos exámenes. En California, una serie de casos judiciales condujo a cierta igualación de la financiación, pero esta redistribución de los fondos no estaba basada en el éxito escolar. Por tanto ni en Massachusetts, ni en California la causalidad simultánea parece ser un problema.

Heterocedasticidad y correlación del término de error entre observaciones. Todos los resultados presentados aquí y en los capítulos anteriores utilizan errores estándar heterocedástico-robustos, por lo que la heterocedasticidad no amenaza la validez interna. Sin embargo, la correlación del término de error entre observaciones, podría poner en peligro la consistencia de los errores estándar debido a que no fue utilizado un muestreo aleatorio simple (la muestra consta de todos los distritos escolares de primaria del estado). Aunque existen fórmulas alternativas para los errores estándar que se podrían aplicar en este caso, los detalles son complicados y especializados y, por tanto, se dejan para los textos más avanzados.

Debate e implicaciones

La similitud entre los resultados de Massachusetts y California sugiere que estos estudios son externamente válidos, en el sentido de que los principales resultados se pueden generalizar al desempeño en las pruebas estandarizadas de otros distritos de escuelas primarias de los Estados Unidos.

Algunas de las amenazas potenciales más importantes para la validez interna se han evitado controlando por las características de los estudiantes, la situación económica familiar y la capacidad económica del distrito, y mediante la verificación de la presencia de no linealidades en la función de regresión. Sin embargo algunas amenazas potenciales a la validez interna permanecen. Uno de los principales candidatos es el sesgo de variable omitida, que tal vez aparezca debido a que las variables de control no captan otras características de los distritos escolares o las oportunidades de aprendizaje extraescolar.

De acuerdo con los datos tanto de California como de Massachusetts, somos capaces de responder a la pregunta de la directora de la Sección 4.1: tras tener en cuenta la situación económica familiar, las características de los estudiantes y la renta del distrito, y después de modelizar las no linealidades de la función de regresión, la disminución en dos alumnos de la ratio estudiantes-maestros se prevé que aumente las calificaciones en las pruebas en aproximadamente 0,08 desviaciones típicas de la distribución de las calificaciones en los exámenes entre distritos. Este efecto es estadísticamente significativo, pero es bastante pequeño. Este pequeño efecto estimado está en línea con los resultados de los numerosos estudios que han investigado los efectos sobre las calificaciones de las reducciones del tamaño de las clases⁵.

La superintendente puede ahora utilizar esta estimación como ayuda para decidir si reducir el tamaño de las clases. Al tomar esta decisión, tendrá que sopesar los costes de la reducción propuesta frente a los beneficios. Los costes incluyen los salarios de los maestros y los gastos de las aulas adicionales. Los beneficios incluyen un mejor rendimiento académico, que aquí se ha medido mediante el rendimiento en las pruebas estandarizadas, pero existen otros beneficios potenciales que no se han estudiado, como la reducción de las tasas de abandono escolar y las mejoras en los ingresos salariales futuros. Este efecto estimado de la propuesta sobre el rendimiento en las pruebas estandarizadas es un *input* importante para el cálculo de costes y beneficios.

9.5 Conclusión

Los conceptos de validez interna y externa proporcionan un marco para evaluar lo que se ha aprendido de un estudio econométrico.

Un estudio basado en la regresión múltiple es internamente válido si los coeficientes estimados son insesgados y consistentes, y si los errores estándar son consistentes. Las amenazas a la validez interna de un estudio así incluyen las variables omitidas, los errores de especificación de la forma funcional (no linealidades), la medición imprecisa de las variables independientes (errores en las variables), la selección muestral y la causalidad simultánea. Cada uno de estos factores introduce correlación entre el regresor y el término de error, lo que a su vez hace a los estimadores MCO sesgados e inconsistentes. Si los errores están correlacionados entre las observaciones, lo que puede ocurrir con datos de series temporales, o si son heterocedásticos pero los errores estándar se calculan utilizando la fórmula válida con homocedasticidad, entonces la validez interna se ve comprometida debido a que los errores estándar serán inconsistentes. Estos últimos problemas pueden tratarse mediante un cálculo adecuado de los errores estándar.

Un estudio que utiliza el análisis de regresión, al igual que cualquier estudio estadístico, es externamente válido si sus resultados pueden generalizarse más allá de la población y el escenario estudiados. A veces puede ayudar la comparación de dos o más estudios sobre el mismo tema. Sin embargo, tanto si existen como si no existen dos o más estudios de este tipo, la evaluación de la validez externa requiere efectuar valoraciones acerca de las similitudes entre la población y el escenario estudiados y la población y el escenario para los que los resultados se están generalizando.

⁵ Si se está interesado en saber más acerca de la relación entre el tamaño de las clases y las calificaciones, véase el análisis de Ehrenberg *et al.* (2001a, 2001b).

Las dos partes siguientes de este libro desarrollan distintos modos de afrontar las amenazas a la validez interna que no pueden mitigarse solamente mediante el análisis de regresión múltiple. La Parte III amplia el modelo de regresión múltiple, de forma que permita mitigar las cinco posibles fuentes de sesgo del estimador MCO; la Parte III analiza asimismo un método diferente para la obtención de la validez interna, los experimentos aleatorizados controlados. En la Parte IV se desarrollan métodos de análisis para los datos de series temporales y para la utilización de datos de series temporales en la estimación de los denominados efectos causales dinámicos, que son los efectos causales que varían a lo largo del tiempo.

Resumen

1. Los estudios estadísticos se evalúan mediante la resolución de la cuestión acerca de si el análisis es interna y externamente válido. Un estudio es internamente válido si las inferencias estadísticas acerca de los efectos causales son válidas para la población que está siendo estudiada. Un estudio es externamente válido si sus inferencias y conclusiones pueden generalizarse a partir de la población y el escenario estudiados a otras poblaciones y escenarios.
2. En la estimación de los efectos causales de la regresión, existen dos tipos de amenazas a la validez interna. En primer lugar, los estimadores MCO serán sesgados e inconsistentes si los regresores y los términos de error están correlacionados. En segundo lugar, los intervalos de confianza y los contrastes de hipótesis no son válidos cuando los errores estándar son incorrectos.
3. Los regresores y los términos de error pueden estar correlacionados cuando existen variables omitidas, se utiliza una forma funcional incorrecta, una o más variables explicativas están medidas erróneamente, la muestra se extrae de la población de forma no aleatoria, o existe causalidad simultánea entre las variables explicativas y las variables dependientes.
4. Los errores estándar son incorrectos cuando los errores son heterocedásticos y el *software* utiliza los errores estándar válidos con homocedasticidad, o cuando el término de error está correlacionado entre las diferentes observaciones.
5. Cuando los modelos de regresión se utilizan exclusivamente para predicción, no es necesario que los coeficientes de regresión sean estimadores insesgados de los efectos causales. Resulta crucial, sin embargo, que el modelo de regresión sea externamente válido para la predicción en el caso concreto estudiado.

Términos clave

población estudiada (223)
 población de interés (223)
 validez interna (224)
 validez externa (224)
 error de especificación de la forma funcional (227)

sesgo por errores en las variables (228)
 modelo clásico de error de medición (229)
 sesgo de selección muestral (230)
 causalidad simultánea (232)
 sesgo de ecuaciones simultáneas (232)

Revisión de conceptos

- 9.1 ¿Cuál es la diferencia entre la validez interna y externa? ¿Y entre la población estudiada y la población de interés?
- 9.2 El Concepto clave 9.2 describe el problema de la selección de variables en términos de un equilibrio en la disyuntiva entre sesgo y varianza. ¿En qué consiste esta disyuntiva? ¿Por qué podría disminuir el sesgo la inclusión de un regresor adicional? ¿Aumenta la varianza?
- 9.3 Las variables económicas a menudo se miden con un error. ¿Significa esto que el análisis de regresión no es fiable? Explíquelo.

- 9.4** Supóngase que un estado ofrece exámenes estandarizados voluntarios a todos sus alumnos de tercer curso y que estos datos se utilizaron en un estudio acerca de la influencia del tamaño de las clases en el rendimiento estudiantil. Explique cómo la selección muestral podría invalidar los resultados.
- 9.5** Un investigador estima el efecto sobre las tasas de criminalidad de los gastos policiales mediante datos a nivel urbano. Explique cómo podría invalidar los resultados la presencia de causalidad simultánea.
- 9.6** Un investigador estima una regresión mediante dos paquetes de *software* diferentes. El primero utiliza la fórmula de los errores estándar válidos con homocedasticidad. El segundo utiliza la fórmula de los errores estándar heterocedástico-robustos. Los errores estándar son muy diferentes. ¿Cuál debería utilizar el investigador? ¿Por qué?

Ejercicios

- 9.1** Supóngase que acabamos de leer un cuidadoso estudio estadístico acerca de los efectos de la publicidad en la demanda de cigarrillos. Utilizando los datos de Nueva York de la década de 1970, el estudio concluye que la publicidad en los autobuses y el metro resultó más efectiva que la publicidad impresa. Utilice el concepto de validez externa para determinar si estos resultados probablemente podrían aplicarse a Boston en la década de 1970; Los Ángeles en la década de 1970, y Nueva York en 2010.
- 9.2** Considérese el modelo de regresión con una única variable $Y_i = \beta_0 + \beta_1 X_i + u_i$ y supóngase que se satisface el supuesto del Concepto clave 4.3. Supóngase que Y_i está medida con un error, por lo que los datos son $\tilde{Y}_i = Y_i + w_i$, donde w_i es el error de medida que es i.i.d. e independiente de Y_i y X_i . Considérese la regresión poblacional $\tilde{Y}_i = \beta_0 + \beta_1 X_i + v_i$, donde v_i es el término de error de la regresión con la variable dependiente erróneamente medida, \tilde{Y}_i .
- Demuestre que $v_i = u_i + w_i$.
 - Demuestre que la regresión $\tilde{Y}_i = \beta_0 + \beta_1 X_i + v_i$ cumple los supuestos del Concepto clave 4.3. (Suponga que w_i es independiente de Y_j y X_j para todos los valores de i y j que tienen un momento de cuarto orden finito).
 - ¿Son los estimadores MCO consistentes?
 - ¿Pueden construirse intervalos de confianza de la forma habitual?
 - Evalúe estas afirmaciones: «El error de medición en las X es un grave problema. El error de medición en Y no lo es».
- 9.3** Los economistas laborales descubrieron un resultado empírico desconcertante por medio de un estudio sobre los determinantes de los ingresos salariales de las mujeres. Utilizando mujeres empleadas aleatoriamente seleccionadas, realizaron una regresión de los ingresos salariales sobre el número de hijos de las mujeres y un conjunto de variables de control (edad, educación, ocupación, etc.). Hallaron que las mujeres con más hijos tenían salarios más altos, teniendo en cuenta estos otros factores. Explique cómo la selección muestral podría ser la causa de este resultado. (Sugerencia: tenga en cuenta que las mujeres que no trabajan fuera del hogar no se encuentran en la muestra). [Este enigma empírico motivó la investigación de James Heckman sobre la selección muestral que le llevó al Premio Nobel de Economía en 2000. Véase Heckman (1974)].
- 9.4** A partir de las regresiones que se muestran en la columna (2) de la Tabla 8.3 y la columna (2) de la Tabla 9.2, construya una tabla como la Tabla 9.3 para comparar los efectos estimados de un aumento del 10 % en la renta del distrito sobre las calificaciones obtenidas en los exámenes de California y Massachusetts.
- 9.5** La demanda de un bien está dada por $Q = \beta_0 + \beta_1 P + u$, donde Q expresa la cantidad, P expresa el precio, y u expresa otros factores distintos del precio que determinan la demanda. La oferta del bien está dada por $Q = \gamma_0 + \gamma_1 P + v$, donde v expresa los factores distintos del precio que influyen en la oferta. Supóngase que tanto u como v tienen una media igual a cero, tienen varianzas σ_u^2 y σ_v^2 , y están mutuamente incorrelacionadas.

- a) Resuelva las dos ecuaciones simultáneas para mostrar cómo Q y P dependen de u y de v .
 b) Obtenga las medias de P y Q .
 c) Obtenga la varianza de P , la varianza de Q , y la covarianza entre Q y P .
 d) Se obtiene una muestra aleatoria de observaciones de (Q_i, P_i) , y se regresa Q_i sobre P_i (es decir, Q_i es el regresando y P_i es el regresor). Supóngase que la muestra es muy grande.
- I) Utilice las respuestas a (b) y a (c) para obtener los valores de los coeficientes de la regresión.
 [Sugerencia: utilizar las Ecuaciones (4.7) y (4.8)].
- II) Un investigador utiliza la pendiente de esta regresión como una estimación de la pendiente de la función de demanda (β_1). ¿Es la pendiente estimada demasiado grande o demasiado pequeña? (Sugerencia: recuerde que la pendiente de las curvas de demanda es negativa y las curvas de oferta tienen pendiente positiva).
- 9.6** Supóngase que $n = 100$, las observaciones i.i.d. sobre (Y_i, X_i) dan lugar a los siguientes resultados de regresión:

$$\hat{Y} = 32,1 + 66,8, ESR = 15,1, R^2 = 0,81 \\ (15,1) \quad (12,2)$$

Otro investigador está interesado en la misma regresión, pero comete un error al introducir los datos en su programa de regresión: introduce cada observación dos veces, por lo que cuenta con 200 observaciones (con la observación 1 introducida dos veces, la observación 2 introducida dos veces, y así sucesivamente).

- a) Mediante estas 200 observaciones, ¿qué resultados arroja su programa de regresión? (Sugerencia: exprese los valores «incorrectos», de las medias, varianzas y covarianzas muestrales de Y y X en función de los valores «correctos». Utilícelos para determinar los estadísticos de regresión)

$$\hat{Y} = \frac{\text{_____}}{\text{_____}} + \frac{\text{_____}}{\text{_____}} X, ESR = \text{_____}, R^2 = \text{_____}$$

- b) ¿Qué requisitos para la validez interna se violan (si hubiera alguno)?

- 9.7** ¿Son las siguientes afirmaciones verdaderas o falsas? Explique su respuesta.
- a) «Una regresión de mínimos cuadrados ordinarios de Y sobre X será internamente inconsistente si X está correlacionada con el término de error».
- b) «Cada una de las cinco principales amenazas a la validez interna implica que X está correlacionada con el término de error».
- 9.8** ¿Sería útil la regresión de la Ecuación (9.5) para predecir las calificaciones en los exámenes de las pruebas de un distrito escolar en Massachusetts? ¿Por qué o por qué no?
- 9.9** Considérese la regresión lineal de la variable *CalificaciónExamen* sobre la variable *Renta* que muestra la Figura 8.2 y la regresión no lineal de la Ecuación (8.18). ¿Podría cualquiera de estas regresiones proporcionar una estimación fiable del efecto de la renta sobre las calificaciones en los exámenes? ¿Cualquiera de estas regresiones proporcionaría un método fiable para la predicción de las calificaciones en los exámenes? Explíquelo.
- 9.10** Lea el recuadro «La rentabilidad de la educación y la brecha de género» de la Sección 8.3. Debata acerca de la validez interna y externa de los efectos estimados de la educación sobre los ingresos salariales.
- 9.11** Léase el recuadro «La demanda de revistas de economía» de la Sección 8.3. Discuta la validez interna y externa del efecto estimado del precio por cita sobre las suscripciones.
- 9.12** Considérese el modelo de regresión de una sola variable $Y_i = \beta_0 + \beta_1 X_i + u_i$ y supongamos que se satisfacen los supuestos de mínimos cuadrados del Concepto clave 4.3. Falta el regresor X_i , pero están disponibles los datos sobre una variable relacionada Z_i , y el valor de X_i se calcula mediante $\bar{X}_i = E(X_i|Z_i)$. Sea $w_i = \bar{X}_i - X_i$.

- a) Demuestre que \tilde{X}_i es el estimador del mínimo error cuadrático medio de X_i utilizando Z_i . Es decir, sea $\hat{X}_i = g(Z_i)$ alguna otra conjetura sobre X_i basada en Z_i y demuestre que $\text{var}(\hat{X}_i - X_i) \geq \text{var}(\tilde{X}_i - X_i)$. (Sugerencia: Ejercicio de revisión 2.27).
- b) Demuestre que $E(w_i|\tilde{X}_i) = 0$.
- c) Supóngase que $E(u_i|Z_i) = 0$ y que se utiliza \tilde{X}_i como regresor en lugar de X_i . Demuestre que $\hat{\beta}_1$ es consistente. ¿Es $\hat{\beta}_0$ consistente?
- 9.13** Supóngase que el modelo de regresión $Y_i = \beta_0 + \beta_1 X_i + u_i$ cumple los supuestos de mínimos cuadrados del Concepto clave 4.3 de la Sección 4.4. Usted y un amigo toman una muestra aleatoria de 300 observaciones de Y y X .
- a) Su amigo le informa de que, sin darse cuenta, ha mezclado las observaciones de X para el 20 % de la muestra. Para estas observaciones mezcladas, el valor de X no se corresponde con el X_i de la observación i -ésima, sino con el valor de X para alguna otra observación. En la notación de la Sección 9.2, el valor medido del regresor, \tilde{X}_i , es igual a X_i para el 80 % de las observaciones, pero es igual a un X_i seleccionado aleatoriamente para el 20 % de las observaciones restantes. Usted realiza una regresión de Y_i sobre \tilde{X}_i . Demuestre que $E(\hat{\beta}_1) = 0,8\beta_1$.
- b) Explique cómo se puede construir una estimación insesgada de β_1 utilizando el estimador MCO de (a).
- c) Suponga ahora que su amigo le dice que las X estaban mezcladas para las primeras 60 observaciones, pero que las 240 observaciones restantes son correctas. Estima β_1 mediante la regresión de Y sobre X utilizando solamente las 240 observaciones correctamente medidas. ¿Es este estimador de β_1 mejor que el estimador que se propone en (b)? Explíquelo.

Ejercicios empíricos

- E9.1** Utilícese la base de datos **CPS08** descrita en el Ejercicio empírico 4.1 para responder a las siguientes preguntas.
- a) Discuta la validez interna de las regresiones que se utilizaron para responder al Ejercicio empírico 8.1(l). Incluya una discusión acerca del posible sesgo de variable omitida, error de especificación de la forma funcional de la regresión, errores en las variables, selección muestral, causalidad simultánea, y la inconsistencia de los errores estándar MCO.
- b) La base de datos **CPS92_08** descrita en el Ejercicio empírico 3.1 incluye datos de 2008 y 1992. Utilice estos datos para investigar la validez externa (temporal) de las conclusiones que se alcanzaron en el Ejercicio empírico 8.1(l). [Nota: recuerde que se debe ajustar por la inflación, como se explica en el Ejercicio empírico 3.1(b).]
- E9.2** Un comité sobre la mejora de la docencia de grado en su universidad necesita su ayuda antes de informar al decano. El comité busca su consejo, como un experto económetra, acerca de si la universidad debería tener en cuenta la apariencia física a la hora de la contratación del profesorado. (Esto es legal siempre y cuando no se mire la raza, la religión, la edad y el género). No tiene tiempo para recoger sus propios datos, por lo que debe basar sus recomendaciones en el análisis de la base de datos **TeachingRatings** descrita en el Ejercicio empírico 4.2, que ha servido como base para algunos ejercicios empíricos en la Parte II del libro de texto. Basándose en su análisis de estos datos, ¿cuál es su consejo? Justifique su consejo sobre la base de una evaluación cuidadosa y completa de la validez interna y externa de las regresiones que se llevaron a cabo para responder a los ejercicios empíricos mediante los datos en los capítulos anteriores.
- E9.3** Utilícese la base de datos **CollegeDistance** descrita en el Ejercicio empírico 4.3 para contestar las siguientes preguntas.
- a) Discuta la validez interna de las regresiones que se utilizaron para la respuesta al Ejercicio empírico 8.3(l). Incluya una discusión sobre el posible sesgo de variable omitida, error de especificación de la forma funcional de la regresión, errores en las variables, selección muestral, causalidad simultánea e inconsistencia de los errores estándar MCO.

- b)** La base de datos **CollegeDistance** excluía a los estudiantes de los estados del oeste; los datos de estos estudiantes se incluyen en la base de datos **CollegeDistanceWest**. Utilice estos datos para investigar la validez (geográfica) externa de las conclusiones que se alcanzaron en el Ejercicio empírico 8.3(I).

APÉNDICE

9.1 Los datos de las calificaciones en las pruebas de educación primaria de Massachusetts

Los datos de Massachusetts son las medias de distrito para los distritos de escuelas públicas primarias en 1998. La calificación en la prueba se toma del Sistema de Evaluación Global de Massachusetts (MCAS, en sus siglas originales⁶), prueba exigida a todos los estudiantes de cuarto curso de las escuelas públicas de Massachusetts en la primavera de 1998. La prueba está patrocinada por el Departamento de Educación de Massachusetts y es obligatoria para todas las escuelas públicas. Los datos analizados aquí consisten en la puntuación global total, que es la suma de las puntuaciones en los apartados de inglés, matemáticas y ciencias en las pruebas.

Los datos sobre la ratio estudiantes-maestros, el porcentaje de estudiantes que perciben un subsidio para el comedor, y el porcentaje de estudiantes que aún están aprendiendo inglés son los promedios de cada distrito escolar de primaria para el año escolar 1997-1998 y se obtuvieron del Departamento de Educación de Massachusetts. Los datos sobre la renta media del distrito se obtuvieron del censo de EE.UU. de 1990.

⁶ Massachusetts Comprehensive Assessment System.

10 Regresión con datos de panel

La regresión múltiple es una herramienta poderosa que permite tener en cuenta el efecto de variables de las cuales disponemos de datos. No obstante, si no se dispone de datos para algunas de las variables, no pueden incluirse en la regresión y los estimadores MCO de los coeficientes de regresión podrían presentar sesgo de variable omitida.

En este capítulo se describe un método para tener en cuenta algunos tipos de variables omitidas que de hecho no se observan. Este método requiere un tipo específico de datos, denominados datos de panel, en los que cada unidad observada, o entidad individual, se observa para dos o más períodos de tiempo. Al estudiar los *cambios* en la variable dependiente en el tiempo, es posible eliminar el efecto de variables omitidas que difieren entre los distintos individuos, pero que son constantes en el tiempo.

La aplicación empírica de este capítulo se refiere a la conducción en estado de ebriedad: ¿Cuáles son los efectos de los impuestos sobre el alcohol y las leyes sobre la conducción bajo los efectos del alcohol sobre los accidentes de tráfico? Abordamos esta cuestión a partir de los datos sobre la mortalidad en accidentes de tráfico, impuestos sobre el alcohol, leyes que penalizan la conducción bajo los efectos del alcohol, y de otras variables relacionadas para los 48 estados contiguos de EE.UU. para cada uno de los siete años que van desde 1982 hasta 1988. Este conjunto de datos de panel permite tener en cuenta las variables no observables que difieren de un estado a otro, tales como las actitudes culturales predominantes hacia la bebida y la conducción, pero que no cambian en el tiempo. Asimismo permite tener en cuenta las variables que varían en el tiempo, como las mejoras en seguridad de los coches nuevos, pero que no varían entre los distintos estados.

En la Sección 10.1 se describe la estructura de los datos de panel y se introduce el conjunto de datos sobre la conducción bajo los efectos del alcohol. La regresión de efectos fijos, la principal herramienta para el análisis de regresión con datos de panel, es una extensión de la regresión múltiple que explota los datos de panel para tener en cuenta las variables que difieren entre los distintos individuos o entidades individuales, pero que son constantes en el tiempo. La regresión de efectos fijos se introduce en las Secciones 10.2 y 10.3, en primer lugar para el caso de dos períodos de tiempo y más tarde para varios períodos de tiempo. En la Sección 10.4, se amplían estos métodos con el fin de incorporar los denominados efectos fijos temporales, que tienen en cuenta las variables no observables que son constantes entre distintas entidades individuales pero que cambian en el tiempo. La Sección 10.5 analiza los supuestos de la regresión con datos de panel y los errores estándar de la regresión con datos de panel. En la Sección 10.6, se utilizan estos métodos para estudiar el efecto de los impuestos sobre el alcohol y las leyes de conducción bajo los efectos del alcohol sobre la mortalidad en accidentes de tráfico.

10.1 Datos de panel

Recordemos de la Sección 1.3 que los **datos de panel** (denominados asimismo datos longitudinales) se refieren a los datos para n entidades individuales distintas observadas en T diferentes períodos. Los datos estatales sobre la mortalidad en accidentes de tráfico estudiados en este capítulo son datos de panel. Estos datos son para $n = 48$ entidades individuales (estados), donde cada entidad individual se observa en $T = 7$ períodos de tiempo (cada uno de los años 1982, ..., 1988), para un total de $7 \times 48 = 336$ observaciones.

Cuando se describen datos de sección cruzada resulta útil utilizar un subíndice para expresar el individuo o entidad individual; por ejemplo, Y_i se refiere a la variable Y para la i -ésima entidad individual. Cuando

**CONCEPTO
CLAVE**
10.1
Notación para datos de panel

Los datos de panel consisten en observaciones sobre las mismas n entidades individuales para dos o más períodos de tiempo T , tal y como se ilustra en la Tabla 1.3. Si el conjunto de datos consta de las observaciones sobre las variables X e Y , entonces los datos se expresan como

$$(X_{1i}, Y_{1i}), i = 1, \dots, n \text{ y } t = 1, \dots, T, \quad (10.1)$$

donde el primer subíndice, i , se refiere a la entidad individual que está siendo observada y el segundo subíndice, t , se refiere al periodo en el que se observa.

se describen los datos de panel, se necesita una notación adicional para realizar un seguimiento tanto de la entidad individual como del periodo de tiempo. Esto se hace mediante dos subíndices en lugar de uno: el primero, i , se refiere a la entidad individual, y el segundo, t , se refiere al periodo de tiempo de la observación. Por lo tanto Y_{it} expresa la variable Y observada para la i -ésima de las n entidades individuales en el t -ésimo de los T periodos. Esta notación se resume en el Concepto clave 10.1.

Existen algunos términos adicionales asociados con los datos de panel para indicar si existen algunas observaciones perdidas. Un **panel equilibrado** dispone de todas sus observaciones; es decir, las variables observadas para cada entidad individual y para cada periodo de tiempo. Un panel al que le faltan algunos datos perdidos para al menos un periodo de tiempo o para al menos una entidad individual se denomina **panel incompleto**. La base de datos sobre mortalidad en accidentes de tráfico tiene datos para los 48 estados contiguos EE.UU. para el total de los siete años, por lo que es equilibrado. Si, por el contrario, faltaran algunos datos (por ejemplo, si no se dispusiera de los datos sobre la mortalidad para algunos estados para 1983), entonces el conjunto de datos sería incompleto. Los métodos que se presentan en este capítulo están descritos para un panel equilibrado; sin embargo, todos estos métodos se pueden utilizar con un panel incompleto, aunque la forma exacta de hacerlo en la práctica depende del *software* de regresión que se utilice.

Ejemplo: mortalidad en accidentes de tráfico e impuestos sobre el alcohol

Hay aproximadamente 40.000 muertes en accidentes de tráfico en carretera cada año en Estados Unidos. Aproximadamente en una cuarta parte de los accidentes mortales está involucrado un conductor que había bebido, y esta proporción se eleva durante los períodos de consumo máximo. Un estudio (Levitt y Porter, 2001) calculaba que hasta el 25 % de los conductores en carretera entre la 1 h. AM y las 3 h. AM habían bebido, y que un conductor que está ebrio a efectos legales presenta una probabilidad al menos 13 veces mayor de causar un accidente fatal que un conductor que no ha estado bebiendo.

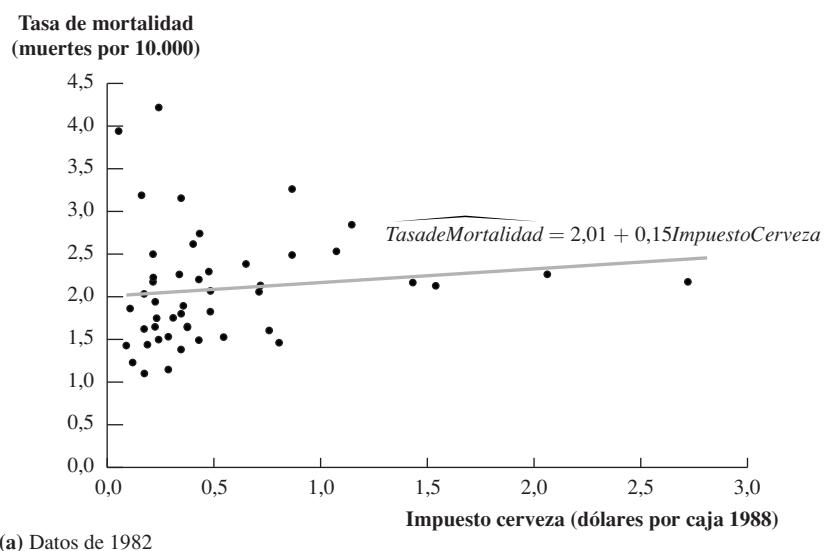
En este capítulo, se estudia la eficacia de varias políticas gubernamentales diseñadas para disuadir a los conductores ebrios que existen en la realidad, sobre la reducción de las muertes en accidentes de tráfico. La base de datos de panel se compone de variables relacionadas con las muertes en accidentes de tráfico y el alcohol, incluyendo el número de muertes en accidentes de tráfico al año en cada uno de los estados, el tipo de leyes sobre conducción en estado de ebriedad en cada estado en cada año, y el impuesto sobre la cerveza en cada uno de los estados. La medida de las muertes en accidentes de tráfico que se utiliza es la tasa de mortalidad, que es el número de muertes anuales en accidentes de tráfico por cada 10.000 personas de la población de cada estado. La medida de los impuestos sobre el alcohol que se utiliza es el impuesto «real» sobre una caja de cervezas, que es el impuesto sobre la cerveza, expresado en dólares de 1988 ajustado por la inflación¹. Los datos se describen con más detalle en el Apéndice 10.1.

La Figura 10.1a es un diagrama de dispersión de los datos para 1982 sobre dos de estas variables, la tasa de mortalidad y el impuesto real sobre una caja de cervezas. Un punto en este diagrama de dispersión representa la tasa de mortalidad en 1982 y el impuesto real sobre la cerveza en 1982 para un estado determinado.

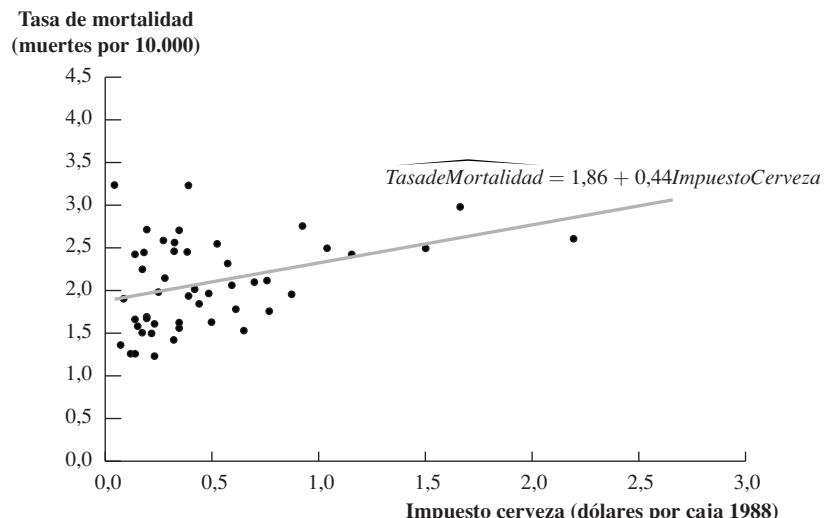
¹ Para hacer comparables los impuestos en el tiempo, se expresan en «dólares de 1988» utilizando el Índice de Precios al Consumidor (IPC). Por ejemplo, debido a la inflación, un impuesto de 1 \$ en 1982 corresponde a un impuesto de 1,23 \$ expresado en dólares de 1988.

FIGURA 10.1 La tasa de mortalidad en accidentes de tráfico y el impuesto sobre la cerveza

El panel (a) es un diagrama de dispersión de las tasas de mortalidad y de los impuestos reales sobre una caja de cervezas (en dólares de 1988) para 48 estados en 1982. El Panel (b) muestra los datos para 1988. Ambos gráficos muestran una relación positiva entre la tasa de mortalidad y el impuesto real sobre la cerveza.



(a) Datos de 1982



(b) Datos de 1988

La recta de regresión MCO obtenida mediante la regresión de la tasa de mortalidad sobre el impuesto real sobre la cerveza asimismo está representada en el gráfico; la recta estimada de regresión es

$$\widehat{\text{TasadeMortalidad}} = 2,01 + 0,15\text{ImpuestoCerveza} \quad (\text{datos de 1982}). \quad (10.2)$$

$(0,15) \quad (0,13)$

El coeficiente del impuesto real sobre la cerveza es positivo, pero no estadísticamente significativo al nivel del 10 %.

Debido a que se dispone de datos para más de un año, es posible reexaminar esta relación para otro año. Esto se hace en la Figura 10.1b, que es el mismo diagrama de dispersión que antes, excepto que utiliza los datos para 1988. La recta de regresión MCO para estos datos es

$$\widehat{\text{TasadeMortalidad}} = 1,86 + 0,44\text{ImpuestoCerveza} \quad (\text{datos de 1988}). \quad (10.3)$$

$(0,11) \quad (0,13)$

A diferencia de la regresión que utiliza los datos para 1982, el coeficiente del impuesto real sobre la cerveza es estadísticamente significativo al nivel del 1 % (el estadístico t es 3,43). Curiosamente, los coeficientes estimados para los datos de 1982 y 1988 son *positivos*: en sentido literal, mayores impuestos reales sobre la cerveza están asociados *con más*, no con menos muertes en accidentes de tráfico.

¿Debería concluirse que un aumento en el impuesto sobre la cerveza colleva más muertes en accidentes de tráfico? No necesariamente, debido a que estas regresiones podrían presentar un sesgo de variable omitida importante. Existen muchos factores que afectan a la tasa de mortalidad, que incluyen la calidad de los automóviles que se conducen en el estado, si las carreteras del lugar en cuestión se encuentran en buen estado, si la mayor parte de la conducción es rural o urbana, la densidad de coches en la carretera, y si está socialmente aceptado beber y conducir. Cualquiera de estos factores puede estar correlacionado con los impuestos sobre el alcohol, y si lo está, va a conllevar un sesgo de variable omitida. Uno de los modos de abordar estas fuentes potenciales de sesgo de variable omitida sería recoger los datos sobre todas estas variables y agregarlos a las regresiones anuales de sección cruzada de las Ecuaciones (10.2) y (10.3). Desafortunadamente, algunas de estas variables, tales como la aceptación cultural de beber y conducir, podrían resultar muy difíciles o incluso imposibles de medir.

Sin embargo, si estos factores se mantienen constantes en el tiempo para un estado determinado, es posible otra vía. Debido a que disponemos de datos de panel, podemos en efecto mantener estos factores constantes, incluso aunque no fuese posible medirlos. Para ello, se utiliza la regresión MCO con efectos fijos.

10.2 Datos de panel con dos períodos temporales: comparaciones «antes y después»

Si se han obtenido los datos para cada estado para $T = 2$ períodos de tiempo, es posible comparar los valores de la variable dependiente en el segundo periodo con los valores en el primer periodo. Debido a que los cambios están centrados en la variable dependiente, esta comparación «antes y después» o de «diferencias» en efecto mantiene constantes los factores no observables que difieren de un estado a otro pero no cambian en el tiempo dentro de un estado.

Sea Z_i una variable que determina la tasa de mortalidad en el estado i -ésimo, que no cambia en el tiempo (por lo que se omite el subíndice t). Por ejemplo, Z_i podría ser la actitud cultural local hacia beber y conducir, la cual cambia lentamente y por lo tanto podría ser considerada como constante entre 1982 y 1988. En consecuencia, la regresión lineal poblacional que relaciona Z_i y el impuesto real sobre la cerveza con la tasa de mortalidad es

$$\widehat{\text{Tasa de Mortalidad}}_{it} = \beta_0 + \beta_1 \text{ImpuestoCerveza}_{it} + \beta_2 Z_i + u_{it} \quad (10.4)$$

donde u_{it} es el término de error y $i = 1, \dots, n$ y $t = 1, \dots, T$.

Debido a que Z_i no cambia en el tiempo, en el modelo de regresión de la Ecuación (10.4) no provocará ningún *cambio* en la tasa de mortalidad entre 1982 y 1988. Por tanto, en este modelo de regresión, la influencia de Z_i puede eliminarse mediante el análisis de la variación de la tasa de mortalidad entre los dos períodos. Para comprobarlo matemáticamente, se considera la Ecuación (10.4) para cada uno de los dos años 1982 y 1988:

$$\text{Tasa de Mortalidad}_{i1982} = \beta_0 + \beta_1 \text{ImpuestoCerveza}_{i1982} + \beta_2 Z_i + u_{i1982}, \quad (10.5)$$

$$\text{Tasa de Mortalidad}_{i1988} = \beta_0 + \beta_1 \text{ImpuestoCerveza}_{i1988} + \beta_2 Z_i + u_{i1988}. \quad (10.6)$$

Restando la Ecuación (10.5) de la Ecuación (10.6) se elimina el efecto de Z_i :

$$\begin{aligned} & \text{Tasa de Mortalidad}_{i1988} - \text{Tasa de Mortalidad}_{i1982} = \\ & = \beta_1 (\text{ImpuestoCerveza}_{i1988} - \text{ImpuestoCerveza}_{i1982}) + u_{i1988} - u_{i1982}. \end{aligned} \quad (10.7)$$

Esta especificación tiene una interpretación intuitiva. Las actitudes culturales hacia la bebida y la conducción afectan al nivel de conducción en estado de ebriedad y por lo tanto a la tasa de mortalidad en un estado. No obstante, si éstas no cambiaron entre 1982 y 1988, entonces no produjeron ningún *cambio* sobre el número de muertes en el estado. Por el contrario, cualquier variación temporal en la mortalidad por accidentes de tráfico debe haber surgido de otras fuentes. En la Ecuación (10.7), las otras fuentes son los cambios en el impuesto sobre la cerveza y los cambios en el término de error (que captura los cambios en otros factores que determinan las muertes por accidentes de tráfico).

Especificando la regresión en términos de variaciones en la Ecuación (10.7), se elimina el efecto de las variables no observables Z_i que son constantes en el tiempo. En otras palabras, el análisis de las variaciones en Y y en X tiene el efecto de tener en cuenta las variables que son constantes con el tiempo, eliminando así esta fuente de sesgo de variable omitida.

La Figura 10.2 representa un diagrama de dispersión del *cambio* en la tasa de mortalidad entre los años 1982 y 1988 sobre el *cambio* en el impuesto real sobre la cerveza entre 1982 y 1988 para los 48 estados de nuestro conjunto de datos. Un punto de la Figura 10.2 representa el cambio en la tasa de mortalidad y el cambio en el impuesto real sobre la cerveza entre 1982 y 1988 para un estado determinado. La recta de regresión MCO, estimada utilizando estos datos y que se representa en el gráfico, es

$$\begin{aligned} \overline{\text{TasadeMortalidad}_{1988} - \text{TasadeMortalidad}_{1982}} &= \\ &= -0,72 - 1,04(\text{ImpuestoCerveza}_{1988} - \text{ImpuestoCerveza}_{1982}). \end{aligned} \quad (10.8)$$

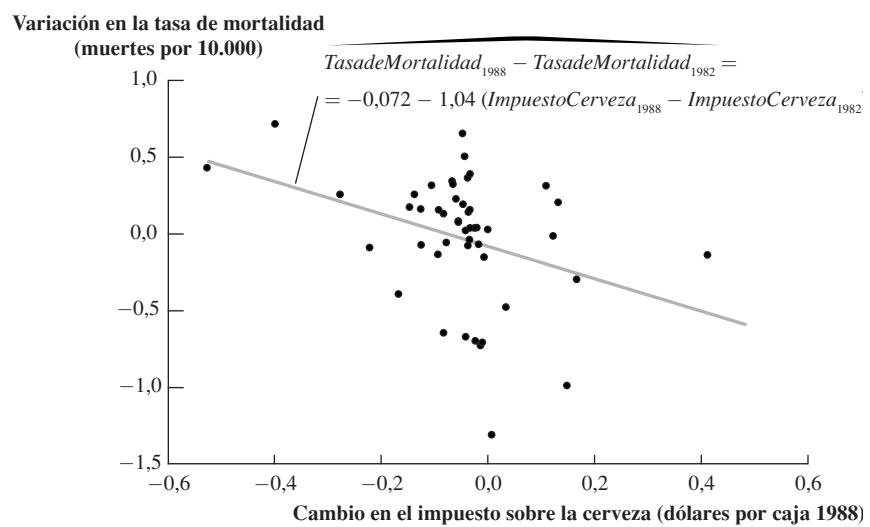
(0,065) (0,36)

Incluyendo un intercepto en la Ecuación (10.8) se permite la posibilidad de que el cambio medio en la tasa de mortalidad, en ausencia de un cambio en el impuesto real sobre la cerveza, sea distinto de cero. Por ejemplo, un intercepto negativo ($-0,072$) podría reflejar las mejoras en la seguridad de los vehículos desde 1982 hasta 1988, que redujeron la tasa media de mortalidad.

En contraposición a los resultados de la regresión de sección cruzada, el efecto estimado de una variación en el impuesto real sobre la cerveza es negativo, de acuerdo con lo que predice la teoría económica. La hipótesis de que el coeficiente poblacional de la pendiente es igual a cero se rechaza al nivel de significación del 5 %. De acuerdo con este coeficiente estimado, un incremento en el impuesto real sobre la cerveza en 1 \$ por caja reduce la tasa de mortalidad por accidentes de tráfico en 1,04 muertes por cada 10.000 personas. Este efecto estimado es muy grande: la tasa de mortalidad promedio en estos datos es de aproxi-

FIGURA 10.2 Cambios o variaciones de la tasa de mortalidad y del impuesto sobre la cerveza, 1982-1988

Este es un diagrama de dispersión del *cambio* de variación en la tasa de mortalidad por accidentes de tráfico y del *cambio* en el impuesto real sobre la cerveza entre 1982 y 1988 para 48 estados. Existe una relación negativa entre los cambios en la tasa de mortalidad y los cambios en el impuesto sobre la cerveza.



madamente 2 (es decir, dos muertos por año cada 10.000 miembros de la población), por lo que la estimación sugiere que las muertes por accidentes de tráfico pueden reducirse a la mitad simplemente aumentando el impuesto real de la cerveza en 1 \$ por caja.

Al examinar las variaciones de la tasa de mortalidad en el tiempo, la regresión de la Ecuación (10.8) controla por los factores fijos, tales como las actitudes culturales hacia la bebida y la conducción. Pero existen muchos factores que influyen en la seguridad del tráfico, y si cambian a lo largo tiempo y están correlacionados con el impuesto real sobre la cerveza, entonces su omisión dará lugar a un sesgo de variable omitida. En la Sección 10.5, se realiza un análisis más cuidadoso que tiene en cuenta varios factores como este, por lo que por ahora, lo mejor es abstenerse de extraer cualquier conclusión preliminar importante acerca del efecto del impuesto real sobre la cerveza en la mortalidad por accidentes de tráfico.

Este análisis «antes y después» funciona cuando los datos se observan para dos años diferentes. Nuestro conjunto de datos, sin embargo, contiene observaciones para siete años diferentes, y parece una tontería descartar estos datos adicionales que son potencialmente útiles. Sin embargo, los métodos «antes y después» no son aplicables directamente cuando $T > 2$. Para analizar todas las observaciones de nuestro conjunto de datos de panel, se utiliza el método de regresión de efectos fijos.

10.3 Regresión de efectos fijos

La regresión de efectos fijos es un método que permite tener en cuenta las variables omitidas en datos de panel cuando las variables omitidas varían entre las distintas entidades individuales (estados), pero no cambian en el tiempo. A diferencia de las comparaciones «antes y después» de la Sección 10.2, la regresión de efectos fijos puede utilizarse cuando hay dos o más observaciones temporales para cada entidad individual.

El modelo de regresión de efectos fijos presenta n interceptos diferentes, uno para cada entidad individual. Estos interceptos pueden representarse mediante un conjunto de variables binarias (o indicadores). Estas variables binarias absorben las influencias de todas las variables omitidas que difieren de una entidad individual a otra, pero son constantes en el tiempo.

El modelo de regresión de efectos fijos

Consideremos el modelo de regresión de la Ecuación (10.4) con la variable dependiente (*TasadeMortalidad*) y el regresor observado (*ImpuestoCerveza*) que vienen expresados por Y_{it} y X_{it} , respectivamente:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}. \quad (10.9)$$

donde Z_i es una variable no observable, que varía de un estado a otro, pero que no cambia en el tiempo (por ejemplo, Z_i representa las actitudes culturales hacia la bebida y la conducción). Se pretende estimar β_1 , el efecto sobre Y de X manteniendo constantes las características no observables del estado Z .

Debido a que Z_i varía de un estado a otro, pero es constante en el tiempo, se puede interpretar que el modelo de regresión poblacional de la Ecuación (10.9) contiene n interceptos, uno para cada estado. En concreto, sea $\alpha_i = \beta_0 + \beta_2 Z_i$. Entonces la Ecuación (10.9) se convierte en

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}. \quad (10.10)$$

La Ecuación (10.10) es el **modelo de regresión de efectos fijos**, en el que $\alpha_1, \alpha_2, \dots, \alpha_n$ se tratan como interceptos desconocidos a estimar, uno para cada estado. La interpretación de α_i como un intercepto específico para cada estado en la Ecuación (10.10) proviene de considerar la recta de regresión poblacional para el estado i -ésimo; esta recta de regresión poblacional es $\alpha_i + \beta_1 X_{it}$. El coeficiente de la pendiente de la recta de regresión poblacional, β_1 , es el mismo para todos los estados, pero el intercepto de la recta de regresión poblacional varía de un estado a otro.

Debido a que el intercepto α_i de la Ecuación (10.10) puede considerarse como el «efecto» de estar en la entidad individual i (en la aplicación actual, las entidades individuales son los estados), los térmi-

nos $\alpha_1, \alpha_2, \dots, \alpha_n$ se conocen como **efectos fijos individuales**. La variación en los efectos fijos individuales proviene de las variables omitidas que, al igual que Z_i en la Ecuación (10.9), varían entre las distintas entidades individuales pero no en el tiempo.

Los interceptos específicos de cada estado en el modelo de regresión de efectos fijos pueden asimismo expresarse utilizando variables binarias que expresen los estados individuales. En la Sección 8.3 se consideraba el caso en que las observaciones pertenecían a uno de los dos grupos y la recta de regresión poblacional presentaba la misma pendiente para ambos grupos, pero interceptos diferentes (véase la Figura 8.8a). Esta recta de regresión poblacional se expresaba matemáticamente utilizando una única variable binaria que indicaba uno de los grupos (caso #1 del Concepto clave 8.4). Si tuviéramos solamente dos estados en nuestra base de datos, ese modelo de regresión de variable binaria sería aplicable aquí. No obstante, debido a que tenemos más de dos estados, se necesitan variables binarias adicionales para captar todos los interceptos específicos de cada estado en la Ecuación (10.10).

Para desarrollar el modelo de regresión de efectos fijos mediante variables binarias, sea $D1_i$ una variable binaria que es igual a 1 cuando $i = 1$ y es igual a 0 en caso contrario, sea $D2_i$ igual a 1 cuando $i = 2$ y es igual a 0 en caso contrario, y así sucesivamente. No pueden incluirse las n variables binarias además de un intercepto común porque si se hace, los regresores serán perfectamente multicolineales (esta es la «trampa de la variable ficticia» de la Sección 6.7), por lo que se omite de forma arbitraria la variable binaria $D1_i$, para el primer grupo. De este modo, el modelo de regresión de efectos fijos de la Ecuación (10.10) puede escribirse de forma equivalente como

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, \quad (10.11)$$

donde $\beta_0, \beta_1, \gamma_2, \dots, \gamma_n$ son los coeficientes desconocidos a estimar. Para obtener la relación entre los coeficientes de la Ecuación (10.11) y los interceptos de la Ecuación (10.10), se comparan las rectas de regresión poblacionales para cada estado entre las dos ecuaciones. En la Ecuación (10.11), la ecuación de regresión poblacional para el primer estado es $\beta_0 + \beta_1 X_{it}$, por lo que $\alpha_1 = \beta_0$. Para los estados segundo y restantes, esta ecuación de regresión es $\beta_0 + \beta_1 X_{it} + \gamma_i$, por lo que $\alpha_i = \beta_0 + \gamma_i$ para $i \geq 2$.

Por tanto, existen dos formas equivalentes de escribir el modelo de regresión de efectos fijos, las Ecuaciones (10.10) y (10.11). En la Ecuación (10.10), el modelo está expresado en términos de n interceptos específicos de cada estado. En la Ecuación (10.11), el modelo de regresión de efectos fijos tiene una intercepto común y $n - 1$ regresores binarios. En ambas formulaciones, el coeficiente de pendiente de X es el mismo de un estado a otro. El intercepto específico para cada estado en la Ecuación (10.10) y los regresores binarios de la Ecuación (10.11) tienen el mismo origen: la variable no observada Z_i que varía entre los estados, pero no en el tiempo.

Extensión a varias X . Si existen otros determinantes observados de Y que estén correlacionados con X y que cambien en el tiempo, entonces deberían del mismo modo estar incluidos en la regresión para evitar el sesgo de variable omitida. Al hacerlo se obtiene el modelo de regresión de efectos fijos con múltiples variables explicativas, que se recoge en el Concepto clave 10.2.

Estimación e inferencia

En principio, la especificación con variables binarias del modelo de regresión de efectos fijos [Ecuación (10.13)] se puede estimar mediante MCO. Esta regresión, sin embargo, tiene $k + n$ regresores (las k X , las $n - 1$ variables binarias, y el intercepto), por lo que en la práctica esta regresión MCO es tediosa o, en algunos paquetes de *software*, es imposible de llevar a cabo si el número de entidades individuales es muy grande. Por lo tanto, el *software* econométrico, tiene rutinas especiales para la estimación MCO de los modelos de regresión de efectos fijos. Estas rutinas especiales son equivalentes a utilizar MCO en la regresión de variables binarias completa, pero son más rápidas debido a que emplean algunas simplificaciones matemáticas que obtienen mediante el álgebra de la regresión de efectos fijos.

El algoritmo MCO «en desviaciones respecto de su media». El *software* de regresión habitualmente calcula el estimador MCO de efectos fijos en dos etapas. En la primera etapa, se le resta a cada

**CONCEPTO
CLAVE**

El modelo de regresión de efectos fijos

El modelo de regresión de efectos fijos es

10.2

$$Y_{it} = \beta_1 X_{1,it} + \dots + \beta_k X_{k,it} + \alpha_i + u_{it}, \quad (10.12)$$

donde $i = 1, \dots, n$; $t = 1, \dots, T$; $X_{1,it}$ es el valor del primer regresor para la entidad individual i en el periodo de tiempo t , $X_{2,it}$ es el valor del segundo regresor, y así sucesivamente; y $\alpha_1, \dots, \alpha_n$ son los interceptos específicos de cada entidad individual.

De manera equivalente, el modelo de regresión de efectos fijos puede expresarse en términos de un intercepto común, las X , y $n - 1$ variables binarias que representan a todas las entidades individuales excepto a una:

$$\begin{aligned} Y_{it} = & \beta_0 + \beta_1 X_{1,it} + \dots + \beta_k X_{k,it} + \gamma_2 D_{2i} \\ & + \gamma_3 D_{3i} + \dots + \gamma_n D_{ni} + u_{it}, \end{aligned} \quad (10.13)$$

donde $D_{2i} = 1$ si $i = 2$ y $D_{2i} = 0$ en caso contrario, y así sucesivamente.

variable la media específica de su entidad individual. En el segundo paso, se estima la regresión utilizando las variables «en desviaciones respecto de su media». En concreto, consideremos el caso de un solo regresor en la versión del modelo de efectos fijos de la Ecuación (10.10) y se toma la media en ambos lados de la Ecuación (10.10); entonces $\bar{Y}_i = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_i$, donde $\bar{Y}_i = (1/T) \sum_{t=1}^T Y_{it}$, y \bar{X}_i y \bar{u}_i se definen de manera similar. Por lo tanto la Ecuación (10.10) implica que $Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i)$. Sea $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$ y $\tilde{u}_{it} = u_{it} - \bar{u}_i$; en consecuencia,

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}. \quad (10.14)$$

De este modo β_1 puede estimarse mediante la regresión MCO de las variables «en desviaciones respecto de su media» \tilde{Y}_{it} sobre \tilde{X}_{it} . De hecho, este estimador es idéntico al estimador MCO de β_1 obtenido a partir de la estimación del modelo de efectos fijos de la Ecuación (10.11) con $n - 1$ variables binarias (Ejercicio 18.6).

La regresión (de las diferencias) «antes y después» versus la especificación con variables binarias. Aunque la Ecuación (10.11) con sus variables binarias parezca muy diferente al modelo de regresión «antes y después» de la Ecuación (10.7), en el caso particular en el que $T = 2$ el estimador MCO de β_1 de la especificación de variable binaria y el de la especificación «antes y después» son idénticos si se excluye el intercepto de la especificaciones «antes y después». Por tanto, cuando $T = 2$ existen tres formas de estimar β_1 por MCO: la especificación «antes y después» de la Ecuación (10.7) (sin intercepto), la especificación con variable binaria de la Ecuación (10.11), y la especificación «en desviaciones respecto de su media» de la Ecuación (10.14). Estos tres métodos son equivalentes; es decir, dan lugar a estimaciones MCO de β_1 idénticas (Ejercicio 10.11).

La distribución muestral, los errores estándar, y la inferencia estadística. En regresión múltiple con datos de sección cruzada, si se cumplen los cuatro supuestos de mínimos cuadrados del Concepto clave 6.4, la distribución muestral del estimador MCO es normal en muestras grandes. La varianza de esta distribución muestral puede estimarse a partir de los datos, y la raíz cuadrada de este estimador de la varianza —es decir, el error estándar— puede utilizarse para el contraste de hipótesis utilizando un estadístico t y para construir intervalos de confianza.

De forma similar, en regresión múltiple con datos de panel, si se cumplen un conjunto de supuestos —denominados supuestos de la regresión de efectos fijos—, entonces la distribución muestral del estimador MCO de efectos fijos es normal en muestras grandes, la varianza de esta distribución puede estimarse a partir de los datos, la raíz cuadrada del estimador es el error estándar, y el error estándar puede utilizarse para construir estadísticos t e intervalos de confianza. Dado el error estándar, la inferencia estadística —los contrastes de hipótesis (incluyendo las hipótesis conjuntas utilizando los estadísticos F) y la construcción de

intervalos de confianza— se realiza de forma exactamente igual que en regresión múltiple con datos de sección cruzada.

Los supuestos de la regresión de efectos fijos y los errores estándar de la regresión de efectos fijos se analizan en la Sección 10.5.

Aplicación a la mortalidad en accidentes de tráfico

La estimación MCO de la recta de regresión de efectos fijos que relaciona el impuesto real sobre la cerveza con la tasa de mortalidad, en base a los datos de los 7 años (336 observaciones), es

$$\widehat{\text{Tasa de Mortalidad}} = -0,66 \text{ImpuestoCerveza} + \text{EfectosFijosdelEstado} \quad (10.15)$$

(0,29)

donde, como es habitual, los interceptos fijos de cada estado estimados no figuran para ahorrar espacio y porque no resultan de gran interés en esta aplicación.

Al igual que la especificación de «diferencias» de la Ecuación (10.8), el coeficiente estimado en la regresión de efectos fijos de la Ecuación (10.15) es negativo, por lo que, según lo previsto por la teoría económica, un impuesto real sobre la cerveza más alto está asociado con menos muertes por accidentes de tráfico, lo cual es lo contrario de lo que encontramos en las primeras regresiones de sección cruzada de las Ecuaciones (10.2) y (10.3). Las dos regresiones no son idénticas, porque la regresión de «diferencias» de la Ecuación (10.8) utiliza solamente los datos para 1982 y 1988 (en concreto, la diferencia entre esos dos años), mientras que la regresión de efectos fijos de la Ecuación (10.15) utiliza los datos para los 7 años. Debido a las observaciones adicionales, el error estándar es menor en la Ecuación (10.15) que en la Ecuación (10.8).

La inclusión de los efectos fijos de cada estado en la regresión de la tasa de mortalidad permite evitar el sesgo de variable omitida derivado de la omisión de factores tales como las actitudes culturales hacia la bebida y la conducción, que varían entre los estados pero que son constantes en el tiempo dentro de un estado. Sin embargo, un escéptico podría sospechar que otros factores podrían llevar a un sesgo de variable omitida. Por ejemplo, durante este periodo los coches podrían haber ido haciéndose cada vez más seguros y sus ocupantes haber utilizado cada vez más los cinturones de seguridad; si el impuesto real sobre la cerveza aumentó en promedio durante la década de 1980, entonces la variable *ImpuestoCerveza* podría estar captando el efecto de la mejora general de la seguridad en el automóvil. Sin embargo, si las mejoras en seguridad evolucionaron en el tiempo, pero fueron las mismas para todos los estados, entonces se puede eliminar su influencia mediante la inclusión de efectos fijos.

10.4 Regresión con efectos fijos temporales

Así como los efectos fijos individuales permiten tener en cuenta las variables que permanecen constantes en el tiempo pero difieren entre las distintas entidades individuales, los efectos fijos temporales permiten tener en cuenta las variables que son constantes entre las entidades individuales, pero que evolucionan en el tiempo.

Debido a que las mejoras de seguridad en los coches nuevos se introducen a nivel nacional, sirven para reducir los accidentes de tráfico en todos los estados. Por lo tanto, resulta verosímil pensar en la seguridad del automóvil como una variable omitida que cambia en el tiempo, pero tiene el mismo valor para todos los estados. La regresión poblacional de la Ecuación (10.9) puede modificarse para hacer explícito el efecto de la seguridad del automóvil, que se expresa mediante S_t :

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \beta_3 S_t + u_{it}, \quad (10.16)$$

donde S_t no es observable y donde el único subíndice t hace hincapié en que la seguridad cambia en el tiempo, pero es constante en todos los estados. Debido a que $\beta_3 S_t$ representa las variables que determinan Y_{it} , si S_t está correlacionada con X_{it} , entonces la omisión de S_t de la regresión conduce a un sesgo de variable omitida.

Solamente efectos temporales

Por el momento, supongamos que las variables Z_i no están presentes por lo que el término $\beta_2 Z_i$ se puede quitar de la Ecuación (10.16), aunque el término $\beta_3 S_t$ permanece. El objetivo es estimar β_1 , teniendo en cuenta S_t .

Aunque S_t no sea observable, su influencia puede eliminarse debido a que varía en el tiempo, pero no entre los estados, del mismo modo que es posible eliminar el efecto de Z_i , que varía entre los estados, pero no en el tiempo. En el modelo de efectos fijos individuales, la presencia de Z_i lleva al modelo de regresión de efectos fijos de la Ecuación (10.10), en que cada estado tiene su propio intercepto (o efecto fijo). Del mismo modo, debido a que S_t varía en el tiempo, pero no entre estados, la presencia de S_t lleva a un modelo de regresión en el que cada periodo de tiempo tiene su propio intercepto.

El **modelo de regresión de efectos fijos temporales** con un único regresor X es

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}. \quad (10.17)$$

Este modelo presenta un intercepto diferente, λ_t , para cada periodo de tiempo. El intercepto λ_t de la Ecuación (10.17) se puede considerar como el «efecto» sobre Y del año t (o, de forma más general, del periodo de tiempo t), por lo que los términos $\lambda_1, \dots, \lambda_T$ se conocen como **efectos fijos temporales**. La variación de los efectos fijos temporales proviene de las variables omitidas que, como S_t en la Ecuación (10.16), varían en el tiempo, pero no entre las entidades individuales.

Al igual que el modelo de regresión de efectos fijos individuales puede representarse mediante $n - 1$ indicadores binarios, asimismo, por tanto, el modelo de regresión de efectos fijos temporales puede representarse mediante $T - 1$ indicadores binarios:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \dots + \delta_T BT_t + u_{it}, \quad (10.18)$$

donde $\delta_2, \dots, \delta_T$ son coeficientes desconocidos y donde $B2_t = 1$ si $t = 2$ y $B2_t = 0$ en caso contrario, y así sucesivamente. Al igual que en el modelo de regresión de efectos fijos de la Ecuación (10.11), en esta versión del modelo de efectos fijos temporales se incluye el intercepto, y se omite la primera variable binaria ($B1_t$) para evitar la multicolinealidad perfecta.

Cuando existen más regresores observados « X », entonces estos regresores aparecen asimismo en las Ecuaciones (10.17) y (10.18).

En la regresión de las muertes por accidentes de tráfico, la especificación de los efectos fijos temporales permite eliminar el sesgo derivado de las variables omitidas como los estándares de seguridad introducidos a nivel nacional que cambian en el tiempo, pero son los mismos en todos los estados para un año determinado.

Efectos fijos individuales y temporales

Si algunas variables omitidas son constantes en el tiempo pero varían entre los estados (como por ejemplo, las normas culturales), mientras que otras son constantes entre los estados pero varían en el tiempo (como por ejemplo, los estándares nacionales de seguridad), entonces resulta apropiado incluir efectos tanto individuales (para los estados) como temporales.

El modelo combinado de **regresión de efectos fijos individuales y temporales** es

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}, \quad (10.19)$$

donde α_i es el efecto fijo individual y λ_t es el efecto fijo temporal. Este modelo puede representarse de manera equivalente mediante $n - 1$ indicadores binarios de la entidad individual y $T - 1$ indicadores binarios temporales, junto con un intercepto:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \dots + \gamma_n Dn_i + \delta_2 B2_t + \dots + \delta_T BT_t + u_{it}, \quad (10.20)$$

donde $\beta_0, \beta_1, \gamma_2, \dots, \gamma_n$ y $\delta_2, \dots, \delta_T$ son coeficientes desconocidos.

Cuando existen regresores observables adicionales « X », estos aparecen asimismo en las Ecuaciones (10.19) y (10.20).

El modelo de regresión combinado de efectos fijos individuales del estado y temporales elimina el sesgo de variables omitidas que aparece tanto por las variables no observables que son constantes en el tiempo como por las variables no observables que son constantes entre los estados.

Estimación. El modelo de efectos fijos temporales y el modelo de efectos fijos individuales y temporales son ambos variantes del modelo de regresión múltiple. Por tanto sus coeficientes pueden estimarse por MCO incluyendo las variables binarias temporales adicionales. Alternativamente, en un panel equilibrado los coeficientes de las X se pueden calcular expresando en primer lugar la Y y las X en términos de desviaciones respecto de sus medias individuales y temporales y posteriormente estimando la ecuación de regresión múltiple de Y en desviaciones sobre las X en desviaciones.

Este algoritmo, que se lleva a cabo habitualmente por el *software* de regresión, elimina la necesidad de construir el conjunto completo de indicadores binarios que aparecen en la Ecuación (10.20). Un método equivalente consiste en expresar Y , las X , y los indicadores temporales en términos de desviaciones respecto de sus medias individuales (pero no temporales) y estimar los $k + T$ coeficientes por regresión múltiple de la Y en desviaciones sobre las X en desviaciones y los indicadores temporales en desviaciones. Finalmente, si $T = 2$ la regresión de efectos fijos individuales y temporales puede estimarse mediante el método «antes y después» de la Sección 10.2, incluyendo el intercepto en la regresión. Por tanto, la regresión «antes y después» presentada en la Ecuación (10.8), en la que el cambio en la variable *TasadeMortalidad* desde 1982 hasta 1988 se regresa sobre el cambio de la variable *ImpuestoCerveza* entre 1982 y 1988 incluyendo un intercepto, proporciona la misma estimación del coeficiente de la pendiente que la regresión MCO de la variable *TasadeMortalidad* sobre la variable *ImpuestoCerveza*, que incluye los efectos fijos individuales y temporales, estimada utilizando los datos de los dos años, 1982 y 1988.

Aplicación a la mortalidad en accidentes de tráfico. Añadiendo los efectos temporales a la regresión de efectos fijos individuales (del estado), los resultados de la estimación MCO de la recta de regresión son:

$$\widehat{\text{TasadeMortalidad}} = -0,64\text{ImpuestoCerveza} + \text{EfectosFijosdelEstado} + \text{EfectosFijosTemporales}. \quad (10.21)$$

(0,36)

Esta especificación incluye el impuesto sobre la cerveza, 47 variables binarias del estado (efectos fijos del estado), 6 variables binarias cada una para un año (efectos fijos temporales), y un intercepto, por lo que esta regresión ¡tiene en realidad $1 + 47 + 6 + 1 = 55$ variables en el lado derecho! Los coeficientes de las variables binarias temporales y del estado y el intercepto no se muestran debido a que no se consideran de gran interés.

La inclusión de los efectos temporales tiene poco impacto sobre el coeficiente del impuesto real sobre la cerveza [comparando las Ecuaciones (10.15) y (10.21)]. A pesar de que este coeficiente se estima de forma menos precisa cuando se incluyen los efectos temporales, sigue siendo significativo al 10 %, pero no al 5 %, de nivel de significación ($t = -0,64/0,36 = -1,78$).

Esta relación estimada entre el impuesto real sobre la cerveza y la mortalidad en accidentes de tráfico es inmune al sesgo de variable omitida de las variables que son constantes tanto en el tiempo como entre los estados. Sin embargo, muchos factores determinantes de las muertes en accidentes de tráfico no entran en esta categoría, por lo que esta especificación podría estar sujeta todavía a sesgo de variable omitida. De acuerdo con esto, en la Sección 10.6 se lleva a cabo una investigación empírica más completa del efecto del impuesto sobre la cerveza y de las leyes dirigidas directamente a eliminar la conducción en estado de ebriedad, teniendo en cuenta varios factores. Antes de pasar a ese estudio, en primer lugar se analizan los supuestos subyacentes en la regresión de datos de panel y la construcción de los errores estándar de los estimadores de los efectos fijos.

10.5 Los supuestos de la regresión de efectos fijos y los errores estándar de la regresión de efectos fijos

En datos de panel, el error de regresión puede estar correlacionado en el tiempo dentro de una entidad individual. Como ocurría en el caso de presencia de heterocedasticidad, esta correlación no introduce un sesgo en el estimador de efectos fijos, pero afecta a la varianza del estimador de efectos fijos y, por lo tanto, afecta a cómo se calculan los errores estándar. Los errores estándar de las regresiones de efectos fijos presentadas en este capítulo son los llamados errores estándar agrupados, que son robustos tanto en presencia de heterocedasticidad como de correlación en el tiempo dentro de una entidad individual. Cuando hay muchas entidades individuales (cuando n es grande), los contrastes de hipótesis y los intervalos de confianza se pueden calcular utilizando los valores críticos habituales para muestras grandes normal y F .

En esta sección se describen los errores estándar agrupados. Comenzamos con los supuestos de la regresión de efectos fijos, que amplían los supuestos de la regresión de mínimos cuadrados a los datos de panel; bajo estos supuestos, el estimador de efectos fijos presenta una distribución asintóticamente normal cuando n es grande. Para mantener la notación tan simple como sea posible, esta sección se centra en el modelo de regresión de efectos fijos individuales de la Sección 10.3, en la que no hay efectos temporales.

Los supuestos de la regresión de efectos fijos

Los cuatro supuestos de la regresión de efectos fijos se resumen en el Concepto clave 10.3. Estos supuestos extienden los cuatro supuestos de mínimos cuadrados, establecidos para los datos de sección cruzada en el Concepto clave 6.4, a los datos de panel.

El primer supuesto es que el término de error presenta una media condicional igual a 0, dados los T valores de X para cada entidad individual. Este supuesto desempeña el mismo papel que el primer supuesto de mínimos cuadrados para los datos de sección cruzada del Concepto clave 6.4 e implica que no existe sesgo de variable omitida. El requisito de que la media condicional de u_{it} no dependa de *ninguno* de los valores de X para esa entidad individual —pasados, presentes o futuros— añade una sutileza importante más allá del primer supuesto de mínimos cuadrados para datos de sección cruzada. Este supuesto se viola si el u_{it} actual está correlacionado con los valores pasados, presentes o futuros de X .

El segundo supuesto es que las variables para una entidad individual se distribuyen de forma idéntica a, pero independientemente de, las variables de otra entidad individual; es decir, las variables son i.i.d. entre entidades individuales para $i = 1, \dots, n$. Al igual que el segundo supuesto de mínimos cuadrados del Concepto clave 6.4, el segundo supuesto de la regresión de efectos fijos se cumple si las entidades individuales son seleccionadas mediante muestreo aleatorio simple a partir de la población.

**CONCEPTO
CLAVE**

10.3

Los supuestos de la regresión de efectos fijos

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

donde

1. u_{it} presenta media condicional igual a cero: $E(u_{it} | X_{i1}, X_{i2}, \dots, X_{iT}, \alpha_i) = 0$.
2. $(X_{i1}, X_{i2}, \dots, X_{iT}, u_{i1}, u_{i2}, \dots, u_{iT})$, $i = 1, \dots, n$ son i.i.d. extraídas a partir de su distribución conjunta.
3. Los datos atípicos elevados son improbables: (X_{it}, u_{it}) tienen momentos de cuarto orden finitos.
4. No existe multicolinealidad perfecta.

Para regresores múltiples, X_{it} debería reemplazarse por la lista completa $X_{1,it}, X_{2,it}, \dots, X_{k,it}$.

El tercer y cuarto supuestos de la regresión de efectos fijos son análogos al tercer y cuarto supuestos de mínimos cuadrados para datos de sección cruzada del Concepto clave 6.4.

Bajo los supuestos de mínimos cuadrados para datos de panel del Concepto clave 10.3, el estimador de efectos fijos es consistente y tiene una distribución normal cuando n es grande. Los detalles se analizan en el Apéndice 10.2.

Una diferencia importante entre los supuestos para datos de panel del Concepto clave 10.3 y los supuestos para datos de sección cruzada del Concepto clave 6.4 es el Supuesto 2. El homólogo para sección cruzada del Supuesto 2 sostiene que cada observación es independiente, lo que surge en virtud de un muestreo aleatorio simple. Por el contrario, el Supuesto 2 para datos de panel sostiene que las variables son independientes entre las distintas entidades individuales, pero no impone ninguna restricción de ese tipo dentro de una entidad individual. Por ejemplo, el Supuesto 2 permite que X_{it} esté correlacionada en el tiempo dentro de una entidad individual.

Si X_{it} está correlacionada con X_{is} para diferentes valores de s y t , es decir, si X_{it} está correlacionada en el tiempo para una determinada entidad individual, entonces se dice que X_{it} está **autocorrelacionada** (correlacionada consigo misma, en diferentes períodos) o **serialmente correlacionada**. La correlación es una característica habitual en los datos de series temporales: lo que sucede un año tiende a estar correlacionado con lo que pasa el año siguiente. En el ejemplo de las muertes en accidentes de tráfico, X_{it} , el impuesto sobre la cerveza en el estado i en el año t , está autocorrelacionado: la mayoría de las veces, el poder legislativo no cambia los impuestos sobre la cerveza, por lo que si es alto un año en relación a su valor promedio para el estado i , tenderá asimismo a ser alto el año siguiente. Del mismo modo, es posible pensar en razones por las que u_{it} estaría autocorrelacionado. Hay que recordar que u_{it} se compone de los factores que varían en el tiempo que son determinantes de Y_{it} pero que no están incluidos como regresores, y algunos de estos factores omitidos podrían estar autocorrelacionados. Por ejemplo, una recesión en la economía local podría generar despidos y disminuir los desplazamientos por carretera, reduciendo así los accidentes de tráfico durante 2 o más años. Del mismo modo, un proyecto de mejora de una carretera principal podría reducir los accidentes de tráfico no solo en el año de realización, sino asimismo en los años siguientes. Tales factores omitidos que persisten durante varios años, dan lugar a errores de regresión autocorrelacionados. No todos los factores omitidos darán lugar a autocorrelación en u_{it} ; por ejemplo, algunos inviernos las condiciones de conducción probablemente afecten a las muertes por accidente, pero si las condiciones meteorológicas invernales para un estado determinado se distribuyen de forma independiente de un año a otro, entonces esta componente del término de error estaría incorrelacionada serialmente. En general, sin embargo, en tanto en cuanto algunos factores omitidos estén autocorrelacionados, entonces u_{it} estará autocorrelacionado.

Errores estándar de la regresión de efectos fijos

Si los errores de la regresión están autocorrelacionados, entonces la fórmula habitual para los errores estándar heterocedástico-robustos para la regresión en sección cruzada [Ecuaciones (5.3) y (5.4)] no es válida. Una forma de comprobarlo es mediante una analogía respecto de heterocedasticidad. En una regresión con datos de sección cruzada, si los errores son heterocedásticos, entonces (como se trató en la Sección 5.4), los errores estándar válidos con homocedasticidad no son válidos debido a que se obtuvieron bajo el falso supuesto de homocedasticidad. Del mismo modo, si los errores están autocorrelacionados, entonces los errores estándar habituales no serán válidos debido a que se obtuvieron bajo el falso supuesto de ausencia de correlación serial.

Los errores estándar que son válidos si u_{it} es potencialmente heterocedástico y está potencialmente correlacionado en el tiempo dentro de una entidad individual se conocen como **errores estándar consistentes a heterocedasticidad y autocorrelación (HAC)**. Los errores estándar que se utilizan en este capítulo son uno de los tipos de errores estándar HAC, los **errores estándar agrupados**. El término *agrupado* se debe a que estos errores estándar permiten la presencia en los errores de regresión de una correlación arbitraria dentro de un conglomerado, agrupación, o «cluster», pero se supone que los errores de regresión no están correlacionados entre los grupos. En el contexto de datos de panel, cada grupo está compuesto por una entidad individual. Por lo que los errores estándar agrupados permiten heterocedasticidad y una autocorrela-

ción arbitraria dentro de una entidad individual, pero consideran que los errores no están correlacionados entre las distintas entidades individuales. Es decir, que los errores estándar agrupados permiten heterocedasticidad y autocorrelación de una manera que sea compatible con el segundo supuesto de regresión de efectos fijos del Concepto clave 10.3.

Al igual que los errores estándar heterocedástico-robustos de la regresión con datos de sección cruzada, los errores estándar agrupados son válidos exista o no exista heterocedasticidad, autocorrelación, o ambas cosas. Si el número de entidades individuales n es grande, la inferencia realizada mediante los errores estándar agrupados puede realizarse con los valores críticos habituales para muestras grandes de la normal para los estadísticos t y de la distribución $F_{q,\infty}$ para los estadísticos F para el contraste de q restricciones.

En la práctica, puede haber una gran diferencia entre los errores estándar agrupados y los errores estándar que no permiten autocorrelación en u_{it} . Por ejemplo, el error estándar habitual (datos de sección cruzada) heterocedástico-robusto para el coeficiente de la variable *ImpuestoCerveza* en la Ecuación (10.21) es 0,25, sustancialmente menor que el error estándar agrupado, 0,36, y los estadísticos t respectivos para contrastar que $\beta_1 = 0$ son $-2,51$ y $-1,78$. La razón por la que se muestra el error estándar agrupado es que permite la correlación serial de u_{it} dentro de una entidad individual, mientras que el error estándar habitual heterocedástico-robusto no. La fórmula de los errores estándar agrupados se muestra en el Apéndice 10.2.

10.6 Las leyes sobre conducción bajo los efectos del alcohol y la mortalidad por accidentes de tráfico

Los impuestos sobre el alcohol constituyen tan solo una de las maneras de disuadir de beber y conducir. Los estados difieren en sus castigos por conducción en estado de ebriedad, y un Estado que tome medidas contra los conductores ebrios podría hacerlo tanto endureciendo las leyes de conducción, como aumentando los impuestos. Si es así, la omisión de estas leyes podría producir un sesgo de variable omitida en el estimador MCO del efecto del impuesto real a la cerveza sobre las muertes en accidentes de tráfico, incluso en las regresiones con efectos fijos individuales y temporales. Además, dado que el uso del vehículo depende en parte de si los conductores tienen un puesto de trabajo y debido a que los cambios en los impuestos pueden reflejar las condiciones económicas (un déficit presupuestario estatal puede conllevar aumentos en los impuestos), la omisión de las condiciones económicas del estado puede dar lugar asimismo a un sesgo de variable omitida. En esta sección, por lo tanto, se amplía el análisis anterior sobre los accidentes de tráfico a fin de incluir otras leyes sobre conducción y las condiciones económicas.

Los resultados se resumen en la Tabla 10.1. El formato de la tabla es el mismo que el de las tablas de resultados de regresión de los Capítulos 7 a 9: cada columna presenta una regresión diferente, y cada fila presenta un coeficiente estimado y el error estándar, el estadístico F y el p -valor, u otra información sobre la regresión.

La columna (1) de la Tabla 10.1 presenta los resultados de la regresión MCO de la tasa de mortalidad sobre el impuesto real sobre la cerveza, sin efectos fijos individuales para el estado ni temporales. Al igual que en las regresiones de sección cruzada para 1982 y 1988 [Ecuaciones (10.2) y (10.3)], el coeficiente del impuesto real sobre la cerveza es *positivo* (0,36): de acuerdo con esta estimación, ¡el aumento de los impuestos sobre la cerveza *aumenta* las muertes en accidentes de tráfico! Sin embargo, la regresión de la columna (2) [presentada previamente como Ecuación (10.15)], que incluye efectos fijos individuales del estado, sugiere que el coeficiente positivo en la regresión (1) es el resultado del sesgo de variable omitida (el coeficiente del impuesto real sobre la cerveza es $-0,66$). El R^2 de la regresión salta desde 0,091 hasta 0,889 cuando se incluyen efectos fijos; evidentemente, los efectos fijos individuales del estado explican una gran parte de la variación en los datos.

Poco cambia cuando se agregan los efectos temporales, como se indica en la columna (3) [presentada previamente como Ecuación (10.21)], excepto que el coeficiente del impuesto sobre la cerveza se estima ahora con menos precisión. Los resultados de las columnas (1) a (3) son acordes con que los factores fijos omitidos —factores históricos y culturales, condiciones generales de las carreteras, densidad de población, actitudes hacia el alcohol y la conducción, etc.— son determinantes importantes de la variación en las muertes por accidentes de tráfico entre los distintos estados.

TABLA 10.1 Análisis de regresión del efecto de las leyes sobre conducción en estado de embriaguez sobre las muertes en accidentes de tráfico

Variable dependiente: tasa de mortalidad por accidentes de tráfico (muertes por 10.000)							
Regresor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Impuesto cerveza	0,36** (0,05)	-0,66* (0,29)	-0,64 ⁺ (0,36)	-0,45 (0,30)	-0,69* (0,35)	-0,46 (0,31)	-0,93** (0,34)
Edad mínima para beber 18				0,028 (0,070)	-0,010 (0,083)		0,037 (0,102)
Edad mínima para beber 19				-0,018 (0,050)	-0,076 (0,068)		-0,065 (0,099)
Edad mínima para beber 20				0,032 (0,051)	-0,100 ⁺ (0,056)		-0,113 (0,125)
Edad mínima para beber						-0,002 (0,021)	
¿Condena preceptiva de cárcel o servicios comunitarios?				0,038 (0,103)	0,085 (0,112)	0,039 (0,103)	0,089 (0,164)
Media de millas del vehículo por conductor				0,008 (0,007)	0,017 (0,011)	0,009 (0,007)	0,124 (0,049)
Tasa de desempleo				-0,063** (0,013)		-0,063** (0,013)	-0,091** (0,021)
Renta real per cápita (logaritmo)				1,82** (0,64)		1,79** (0,64)	1,00 (0,68)
Años	1982-88	1982-88	1982-88	1982-88	1982-88	1982-88	Solamente 1982 y 1988
¿Efectos individuales del estado?	no	sí	sí	sí	sí	sí	sí
¿Efectos temporales?	no	no	sí	sí	sí	sí	sí
¿Errores estándar agrupados?	no	sí	sí	sí	sí	sí	sí
Estadísticos F y p-valores para el contraste de exclusión de grupos de variables							
Efectos temporales = 0		4,22 (0,002)	10,12 (< 0,001)	3,48 (0,006)	10,28 (< 0,001)	37,49 (< 0,001)	
Coeficientes edad mínima para beber = 0			0,35 (0,786)	1,41 (0,253)		0,42 (0,738)	
Tasa de desempleo, renta per cápita = 0			29,62 (< 0,001)		31,96 (< 0,001)	25,20 (< 0,001)	
\bar{R}_2	0,091	0,889	0,891	0,926	0,893	0,926	0,899

Estas regresiones se estimaron utilizando datos de panel sobre 48 estados de EE.UU. Las regresiones (1) a (6) utilizan datos para todos los años desde 1982 a 1988, y la regresión (7) utiliza datos solamente de 1982 y 1988. El conjunto de datos se describe en el Apéndice 10.1. Los errores estándar se muestran entre paréntesis bajo los coeficientes, y los *p*-valores se muestran entre paréntesis debajo de los estadísticos *F*. El coeficiente individual es estadísticamente significativo al ^{+10 %}, ^{*5 %}, o ^{**1 %} de nivel de significación.

Las siguientes cuatro regresiones de la Tabla 10.1 incluyen otros posibles factores determinantes de las tasas de mortalidad junto con efectos individuales del estado y temporales. La especificación base, que se presenta en la columna (4), incluye las variables relacionadas con las leyes sobre conducción en estado de ebriedad, más las variables que tienen en cuenta la cantidad de conducción y las condiciones económicas generales del estado. Las primeras variables legales son la edad mínima legal para consumir alcohol, representadas mediante tres variables binarias de la edad mínima legal para el consumo de alcohol de 18, 19 y 20 años (por lo que el grupo que se omite es el de una edad mínima legal para beber de 21 años o más). La otra variable legal es el castigo asociado a la primera condena por conducir bajo la influencia del alcohol, ya sea condena preceptiva de cárcel o servicio comunitario obligatorio (se omite el grupo de castigo menos severo). Las tres medidas sobre la cantidad de conducción y condiciones económicas son la media de millas recorridas por cada conductor, la tasa de desempleo, y el logaritmo de la renta personal real (en dólares de 1998) per cápita (la utilización del logaritmo de la renta permite que el coeficiente se pueda interpretar en términos de variación porcentual de la renta; véase la Sección 8.2). La última regresión de la Tabla 10.1 sigue el método «antes y después» de la Sección 10.2 y solo utiliza los datos de 1982 y 1988; por lo que la regresión (7) amplía la regresión de la Ecuación (10.8) para incluir las variables explicativas adicionales.

La regresión de la columna (4) arroja cuatro resultados interesantes.

1. La inclusión de las variables adicionales reduce (en términos absolutos) el efecto estimado del impuesto sobre la cerveza desde el $-0,64$ de la columna (3) al $-0,45$ de la columna (4). Una forma de evaluar la magnitud de este coeficiente es imaginar un estado con un impuesto real promedio sobre la cerveza que duplique este impuesto; debido a que el impuesto real medio sobre la cerveza en estos datos es de aproximadamente $0,50$ \$ por caja (en dólares de 1988), esto implica un aumento del impuesto de $0,50$ \$ por caja. El efecto estimado de un aumento de $0,50$ \$ en el impuesto sobre la cerveza es el de la disminución en la tasa de mortalidad esperada en $0,45 \times 0,50 = 0,23$ muertes por cada 10.000 habitantes. Este efecto estimado es grande: debido a que la tasa de mortalidad media es de 2 por cada 10.000 habitantes, una reducción de $0,23$ corresponde a la reducción de las muertes en accidentes de tráfico en casi un octavo. Dicho esto, la estimación es bastante imprecisa: debido a que el error estándar de este coeficiente es $0,30$, el intervalo de confianza al 95 % para este efecto es $-0,45 \times 0,50 \pm 1,96 \times 0,30 \times 0,50 = (-0,52, 0,07)$. Este amplio intervalo de confianza al 95 % incluye el valor cero, por lo que la hipótesis de que el impuesto sobre la cerveza no tiene ningún efecto, no puede rechazarse al nivel de significación del 5 %.
2. La edad mínima legal para consumir alcohol se estima que tiene un pequeño efecto sobre los accidentes de tráfico de forma precisa. De acuerdo con la regresión de la columna (4), el intervalo de confianza al 95 % para el aumento de la tasa de mortalidad de un estado con una edad mínima legal para beber de 18 años, en relación a una edad mínima de 21 años, es $(-0,11, 0,17)$. La hipótesis conjunta de que los coeficientes de las variables de edad mínima legal para beber son iguales a cero no puede ser rechazada al nivel de significación del 10 %: el estadístico F para el contraste de la hipótesis conjunta de que los tres coeficientes son iguales a cero es $0,35$, con un p -valor de $0,786$.
3. Se estima que el coeficiente de la variable de castigo por el primer delito es pequeño y no es significativamente diferente de cero al nivel de significación del 10 %.
4. Las variables económicas tienen un poder explicativo considerable sobre el número de víctimas mortales por accidentes de tráfico. Las altas tasas de desempleo están asociadas con menos muertes: se calcula que un aumento de la tasa de desempleo en un punto porcentual reduce las víctimas de accidentes de tráfico en $0,063$ muertes por 10.000. Del mismo modo, los valores altos de la renta real per cápita están asociados con un número elevado de muertes: el coeficiente es $1,82$, por lo que un aumento del 1 % en la real per cápita está asociado con un aumento en las víctimas de accidentes de tráfico de $0,0182$ muertes por cada 10.000 habitantes (véase el Caso I del Concepto clave 8.2 para la interpretación de este coeficiente). De acuerdo con estas estimaciones, unas buenas condiciones económicas están asociadas con un mayor número de muertes, tal vez debido al aumento de la densidad de tráfico cuando la tasa de desempleo es baja, o al mayor consumo de alcohol cuando la renta es

alta. Las dos variables económicas son conjuntamente significativas al nivel de significación del 0,1 % (el estadístico F es 29,62).

Las columnas (5) a (7) de la Tabla 10.1 presentan las regresiones que comprueban la sensibilidad de estas conclusiones a los cambios en la especificación base. La regresión de la columna (5) suprime las variables que tienen en cuenta las condiciones económicas. El resultado es un aumento en el efecto estimado del impuesto real sobre la cerveza, que pasa a ser significativo al nivel del 5 %, pero ningún cambio apreciable en los otros coeficientes. La sensibilidad del coeficiente estimado del impuesto sobre la cerveza a la inclusión de las variables económicas, en combinación con la significatividad estadística de los coeficientes de las variables de la columna (4), indica que las variables económicas deben permanecer en la especificación base. La regresión de la columna (6) muestra que los resultados de la columna (4) no son sensibles a los cambios en la forma funcional cuando los tres indicadores de la edad mínima para empezar a beber se reemplazan por la existencia de regulación sobre edad mínima para beber en sí misma. Cuando se estiman los coeficientes utilizando los cambios en las variables de 1982 a 1988 [columna (7)], como en la Sección 10.2, los resultados de la columna (4) permanecen en gran parte sin cambios, excepto que el coeficiente del impuesto sobre la cerveza es mayor y es significativo al nivel del 1 %.

La fortaleza de este análisis es que la inclusión de los efectos fijos individuales del estado y temporales mitiga la amenaza de sesgo de variable omitida que surge de las variables no observables que o bien no cambian con el tiempo (como las actitudes culturales hacia la bebida y la conducción) o bien no cambian entre los estados (como las innovaciones de seguridad). Como siempre, no obstante, es importante considerar las posibles amenazas a la validez. Una fuente potencial de sesgo de variable omitida es que la medida de los impuestos sobre alcohol que se utiliza aquí, el impuesto real sobre la cerveza, podría arrastrar a otros impuestos sobre el alcohol, lo que sugiere una interpretación de los resultados de un modo más amplio que lo que corresponde solamente a la cerveza. Una posibilidad sutil es que los aumentos en el impuesto real sobre la cerveza podrían estar asociados con campañas de educación pública. Si es así, los cambios en el impuesto real sobre la cerveza podrían recoger el efecto de una campaña más amplia para reducir la conducción en estado de ebriedad.

En conjunto, estos resultados presentan una imagen provocadora de las medidas de control de la conducción bajo los efectos del alcohol y la mortalidad en accidentes de tráfico. De acuerdo con estas estimaciones, ni los castigos más severos, ni el aumento de la edad mínima legal para consumir alcohol tienen efectos importantes sobre la mortalidad. Por el contrario, existe evidencia de que el aumento de los impuestos sobre el alcohol, medidos mediante el impuesto real sobre la cerveza, reduce las muertes por accidentes de tráfico, presumiblemente a través de la reducción del consumo de alcohol. La imprecisión de la estimación del coeficiente del impuesto sobre la cerveza significa, no obstante, que deberíamos ser cautelosos acerca de sacar conclusiones políticas de este análisis y que es necesaria una investigación adicional².

10.7 Conclusión

En este capítulo se muestra cómo las distintas observaciones en el tiempo de una misma entidad individual pueden ser utilizadas a fin de tener en cuenta las variables omitidas no observables que difieren entre las distintas entidades, pero que son constantes en el tiempo. La idea clave es que si las variables no observables no cambian en el tiempo, entonces cualquier cambio en la variable dependiente debe estar ocasionado por otras influencias distintas a esas características fijas. Si las actitudes culturales hacia la bebida y la conducción no cambian de manera apreciable durante 7 años dentro de un estado, entonces la explicación a los cambios en la tasa de mortalidad por accidentes de tráfico a lo largo esos 7 años debe ser otra.

² Para un análisis adicional de estos datos, véase Ruhm (1996). Un reciente meta-análisis acerca de 112 estudios sobre el efecto de los precios del alcohol y los impuestos sobre el consumo hallaba elasticidades de $-0,46$ para la cerveza, $-0,69$ para el vino, y $-0,80$ para los licores, y concluía que los impuestos sobre el alcohol tienen efectos importantes sobre la reducción del consumo, en relación a otros programas [Wagenaar, Salois y Komro (2009)]. Para obtener más información acerca de la conducción bajo los efectos del alcohol, y sobre la economía del alcohol en general, véase asimismo Cook y Moore (2000), Chaloupka, Grossman y Saffer (2002), Young y Bielinska-Kwapisz (2006), y Dang (2008).

Para explotar esta idea, se necesitan datos en los que se observe la misma entidad individual en dos o más períodos de tiempo; es decir, se necesitan datos de panel. Con datos de panel, el modelo de regresión múltiple de la Parte II se puede ampliar a fin de poder incluir un conjunto completo de variables binarias para las entidades individuales; este es el modelo de regresión de efectos fijos, que puede ser estimado por MCO. Un rasgo del modelo de regresión de efectos fijos es que incluye los efectos fijos temporales que tienen en cuenta las variables no observables que cambian en el tiempo, pero que son constantes entre las distintas entidades. Tanto los efectos fijos individuales de las entidades como los temporales pueden incluirse en la regresión a fin de tener en cuenta las variables que varían entre las distintas entidades pero que son constantes en el tiempo y las variables que varían en el tiempo pero que son constantes entre las distintas entidades.

A pesar de esas virtudes, la regresión de efectos fijos individuales de cada entidad y temporales no puede tener en cuenta las variables omitidas que varían *tanto* entre las entidades individuales *como* en el tiempo. Y, obviamente, los métodos para datos de panel requieren datos de panel, que a menudo no están disponibles. Por tanto permanece la necesidad de un método que pueda eliminar la influencia de las variables omitidas no observables cuando los métodos para datos de panel no pueden hacer el trabajo. Un método potente y general para hacerlo es la regresión de variables instrumentales, el tema del Capítulo 12.

Resumen

1. Los datos de panel consisten en observaciones sobre varias (n) entidades individuales —estados, empresas, personas, etc.— en los que se observa cada entidad individual en dos o más períodos de tiempo (T).
2. La regresión con efectos fijos individuales tiene en cuenta las variables no observables que difieren de una entidad a otra, pero permanecen constantes en el tiempo.
3. Cuando existen dos períodos de tiempo, la regresión de efectos fijos puede estimarse mediante una regresión «antes y después» de la variación de Y del primer período al segundo sobre la variación en X correspondiente.
4. La regresión de efectos fijos individuales se puede estimar mediante la inclusión de variables binarias para $n - 1$ entidades individuales, además de las variables independientes observables (las X) y un intercepto.
5. Los efectos fijos temporales tienen en cuenta las variables no observables que son iguales entre las distintas entidades individuales, pero que varían en el tiempo.
6. Una regresión con efectos fijos temporales e individuales se puede estimar mediante la inclusión de variables binarias para $n - 1$ entidades y de variables binarias para $T - 1$ períodos de tiempo, además de las X y un intercepto.
7. En datos de panel, las variables están habitualmente autocorrelacionadas, es decir, correlacionadas en el tiempo dentro de una entidad individual. Se necesitan errores estándar que permitan tanto esta autocorrelación como la potencial heterocedasticidad, y una manera de hacerlo es utilizando los errores estándar agrupados.

Términos clave

datos de panel (249)

panel equilibrado (250)

panel incompleto (250)

modelo de regresión de efectos fijos (254)

efectos fijos individuales (255)

modelo de regresión de efectos fijos temporales (258)

efectos fijos temporales (258)

modelo de regresión de efectos fijos individuales y temporales (258)

autocorrelacionada (261)

serialmente correlacionada (261)

errores estándar consistentes a heterocedasticidad y autocorrelación (HAC) (261)

errores estándar agrupados (261)

Revisión de conceptos

- 10.1** ¿Por qué es necesario el uso de dos subíndices, i y t , para describir los datos del panel? ¿A qué se refiere i ? ¿A qué se refiere t ?
- 10.2.** Un investigador utiliza un conjunto de datos de panel sobre $n = 1.000$ trabajadores sobre $T = 10$ años (entre 2001 y 2010) que contiene los ingresos salariales, el género, la educación, y la edad de los trabajadores. El investigador está interesado en el efecto de la educación sobre los ingresos salariales. Proponga algunos ejemplos de variables no observables específicas de cada persona que estén correlacionadas con la educación y los ingresos. ¿Puede pensar en ejemplos de variables específicas temporales que pudieran estar correlacionadas con la educación y los ingresos? ¿Cómo podrían tenerse en cuenta estas variables específicas para cada persona y periodo temporal en una regresión de datos de panel?
- 10.3** ¿Podría utilizarse la regresión propuesta como respuesta a la Pregunta 10.2 para estimar el efecto del género sobre los ingresos salariales de una persona? ¿Puede utilizarse esta regresión para estimar el efecto de la tasa nacional de desempleo sobre los ingresos de una persona? Explíquelo.
- 10.4** En el contexto de la regresión sugerida en la Pregunta 10.2, explique por qué el error de regresión para un individuo determinado puede presentar correlación serial.

Ejercicios

- 10.1** Este ejercicio se refiere al panel de datos sobre conducción bajo los efectos del alcohol resumido en la Tabla 10.1.
- Nueva Jersey tiene una población de 8,1 millones de personas. Supongamos que Nueva Jersey hubiera aumentado el impuesto sobre una caja de cerveza por un importe de 1 \$ (en dólares de 1988). Utilice los resultados de la columna (4) para predecir el número de vidas que se salvarían durante el año siguiente. Construya un intervalo de confianza al 95 % para la respuesta.
 - La edad mínima legal para beber en Nueva Jersey es de 21 años. Supongamos que Nueva Jersey bajara su edad legal para beber a 18 años. Utilice los resultados de la columna (4) para predecir el cambio en el número de muertes en accidentes de tráfico para el año siguiente. Construya un intervalo de confianza al 95 % para la respuesta.
 - Supóngase que la renta real per cápita en Nueva Jersey se incrementara en un 1 % durante el próximo año. Utilice los resultados de la columna (4) para predecir el cambio en el número de muertes por accidentes de tráfico del año siguiente. Construya un intervalo de confianza al 90 % para la respuesta.
 - ¿Deberían incluirse efectos temporales en la regresión? ¿Por qué o por qué no?
 - Un investigador se pregunta acerca de si la tasa de desempleo tiene un efecto diferente sobre las muertes por accidentes de tráfico en los estados occidentales que en el resto de estados. ¿Cómo se contrastaría esta hipótesis? (Sea específico acerca de la especificación de la regresión y sobre el contraste estadístico que utilizaría).
- 10.2** Considérese la versión con variable binaria del modelo de regresión de efectos fijos de la Ecuación (10.11), excepto por un regresor adicional, $D1_i$; es decir, sea
- $$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_1 D1_i + \gamma_2 D2 + \cdots + \gamma_n Dn_i + u_{it}.$$
- Supongamos que $n = 3$. Demuestre que los regresores binarios y el regresor «constante» presentan multicolinealidad exacta; es decir, exprese una de las variables $D1_i$, $D2_i$, $D3_i$ y $X_{0,it}$ como una función lineal exacta del resto, donde $X_{0,it} = 1$ para todo i, t .
 - Demuestre el resultado de (a) para una n general.
 - ¿Qué ocurrirá si se intentase estimar los coeficientes de la regresión por MCO?
- 10.3** En la Sección 9.2 se proporciona una lista de cinco posibles amenazas a la validez interna de un análisis de regresión. Aplique esta lista al análisis empírico de la Sección 10.6 y extraiga así las conclusiones pertinentes acerca de su validez interna.

- 10.4** Utilizando la regresión de la Ecuación (10.11), ¿cuál es la pendiente y el intercepto de
- La entidad individual 1 en el periodo de tiempo 1?
 - La entidad individual 1 en el periodo de tiempo 3?
 - La entidad individual 3 en el periodo de tiempo 1?
 - La entidad individual 3 en el periodo de tiempo 3?
- 10.5** Considérese el modelo con un solo regresor $Y_{it} = \beta_0 + \beta_1 X_{1,it} + \alpha_i + \lambda_t + u_{it}$. Este modelo puede expresarse asimismo como
- $$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \delta_2 B2_t + \dots + \delta_T BT_t + \gamma_2 D2_i + \dots + \gamma_n Dn_i + u_{it},$$
- donde $B2_t = 1$ si $t = 2$ y 0 en cualquier otro caso, $D2_i = 1$ si $i = 2$ y 0 en cualquier otro caso, etcétera. ¿Cómo son los coeficientes ($\beta_0, \delta_2, \dots, \delta_T, \gamma_2, \dots, \gamma_n$) en relación a los coeficientes ($\alpha_1, \dots, \alpha_n, \lambda_1, \dots, \lambda_T$)?
- 10.6** ¿Los supuestos de la regresión de efectos fijos del Concepto clave 10.3 implican que $\text{cov}(\tilde{v}_{it}, \tilde{v}_{is}) = 0$ para $t \neq s$ en la Ecuación (10.28)? Explíquelo.
- 10.7** Un investigador cree que las muertes por accidentes de tráfico se incrementan cuando las carreteras están heladas y por tanto los estados donde más nieva tendrán más muertes que otros estados. Comente los siguientes métodos diseñados para estimar el efecto de la nieve sobre las muertes:
- El investigador recoge datos sobre las precipitaciones en forma de nieve para cada estado y añade este regresor (MediadeNieve_i) a las regresiones que figuran en Tabla 10.1.
 - El investigador recoge datos sobre las nevadas en cada estado para cada año de la muestra ($Nieve_{it}$) y añade este regresor a las regresiones.
- 10.8** Considérense las observaciones (Y_{it}, X_{it}) del modelo lineal de datos de panel
- $$Y_{it} = X_{it}\beta_1 + \alpha_i + \lambda_i t + u_{it},$$
- donde $t = 1, \dots, T; i = 1, \dots, N$; y $\alpha_i + \lambda_i t$ es una tendencia temporal inobservable específica para cada individuo. ¿Cómo se estimaría β_1 ?
- 10.9**
- En el modelo de regresión de efectos fijos, ¿se estiman consistentemente los efectos fijos individuales α_i , si $n \rightarrow \infty$ con T fijo? (Sugerencia: analizar el modelo sin X : $Y_{it} = \alpha_i + u_{it}$).
 - Si n es grande (por ejemplo, $n = 2.000$), pero T es pequeño (por ejemplo, $T = 4$), ¿piensa que los valores estimados de α_i presentan aproximadamente una distribución normal? ¿Por qué o por qué no? (Sugerencia: analizar el modelo $Y_{it} = \alpha_i + u_{it}$.)
- 10.10** En un estudio sobre el efecto sobre los ingresos salariales de la educación utilizando datos de panel sobre los ingresos anuales de un gran número de trabajadores, un investigador realiza una regresión de los ingresos salariales en un año dado sobre la edad, la educación, la condición sindical, y los ingresos salariales del trabajador durante el año anterior mediante una regresión de efectos fijos. ¿Proporcionará esta regresión estimaciones fiables sobre los efectos de las variables explicativas (edad, educación, condición sindical e ingresos salariales del año anterior) sobre los ingresos salariales? Explíquelo. (Sugerencia: compruebe la regresión de efectos fijos de la Sección 10.5).
- 10.11** Sea $\hat{\beta}_1^D$ la expresión del estimador en desviaciones respecto de la media individual dado en la Ecuación (10.22), y sea $\hat{\beta}_1^{AD}$ la expresión del estimador «antes y después» sin intercepto, de modo que $\hat{\beta}_1^{AD} = [\sum_{i=1}^n (X_{i2} - X_{i1})(Y_{i2} - Y_{i1}) / \sum_{i=1}^n (X_{i2} - X_{i1})^2]$. Demuestre que, si $T = 2$, $\hat{\beta}_1^D = \hat{\beta}_1^{AD}$ [Sugerencia: utilice la definición de \tilde{X}_{it} antes de la Ecuación (10.22) para demostrar que $\tilde{X}_{i1} = -\frac{1}{2}(X_{i2} - X_{i1})$ y $\tilde{X}_{i2} = \frac{1}{2}(X_{i2} - X_{i1})$].

Ejercicios empíricos

- E10.1** Algunos estados de EE.UU. han promulgado leyes que permiten a los ciudadanos llevar armas. Estas leyes son conocidas como leyes de «emisión obligatoria», debido a que obligan a las autoridades locales a emitir un permiso para llevar armas a todos los solicitantes que sean ciudadanos,

sean mentalmente competentes y no hayan sido condenados por un delito grave (algunos estados imponen algunas restricciones adicionales). Sus defensores sostienen que si más personas llevan armas, el crimen se reducirá debido a que los criminales serán disuadidos de atacar a otras personas. Sus opositores argumentan que el crimen aumentará debido al uso accidental o espontáneo de las armas. En este ejercicio, se analiza el efecto de las leyes sobre la tenencia de armas sobre los crímenes violentos. En la página web del libro de texto http://www.pearsonhighered.com/stock_watson se encuentra el archivo de datos **Guns** que contiene un panel equilibrado de datos sobre 50 estados de EE.UU., más el Distrito de Columbia para los años 1977 a 1999³. Se ofrece una descripción detallada en el archivo **Guns_Description**, disponible en la página web.

- a) Estime (1) una regresión de la variable $\ln(vio)$ sobre la variable *shall* y (2) una regresión de la variable $\ln(vio)$ sobre las variables *shall*, *incarc_rate*, *density*, *avginc*, *pop*, *pb1064*, *pw1064* y *pm1029*.
 - I) Interprete el coeficiente de la variable *shall* en la regresión (2). ¿Es esta estimación grande o pequeña en el sentido del «mundo real»?
 - II) Al añadir las variables de control en la regresión (2) ¿cambia el efecto estimado de la ley de emisión obligatoria en la regresión (1), medido por la significación estadística? ¿Y medido por el significatividad en el mundo real del coeficiente estimado?
 - III) Sugiera una variable que cambie entre los estados, pero que probablemente cambie poco, o nada en absoluto, en el tiempo y que pudiera causar sesgo de variable omitida en la regresión (2).
- b) ¿Cambian los resultados cuando se agregan los efectos fijos individuales de cada estado? Si es así, ¿qué conjunto de resultados de regresión es más creíble y por qué?
- c) ¿Cambian los resultados cuando se añaden los efectos fijos temporales? Si es así, ¿qué conjunto de resultados de regresión es más creíble y por qué?
- d) Repita el análisis utilizando las variables $\ln(rob)$ y $\ln(mur)$ en lugar de la variable $\ln(vio)$.
- e) En su opinión, ¿cuáles son las amenazas más importantes a la validez interna de este análisis de regresión que perduran?
- f) En base a su análisis, ¿qué conclusiones sacaría sobre los efectos de las leyes de tenencia de armas sobre los índices de criminalidad?

E10.2 Los accidentes de tráfico son la principal causa de muerte de los estadounidenses entre los 5 y los 32 años de edad. Mediante distintas políticas de gasto, el gobierno federal ha alentado a los estados a instituir normativas de obligatoriedad de uso del cinturón de seguridad para reducir el número de muertes y lesiones graves. En este ejercicio se investigará la eficacia de estas leyes para el aumento del uso del cinturón de seguridad y la reducción de víctimas mortales. En la página web del libro de texto http://www.pearsonhighered.com/stock_watson se encuentra el archivo de datos **Seatbelts** que contiene un panel de datos sobre 50 estados de EE.UU., además del Distrito de Columbia para los años 1983 a 1997⁴. Se ofrece una descripción detallada en el archivo **Seatbelts_Description**, disponible en la página web.

- a) Estime el efecto del uso del cinturón de seguridad sobre las muertes mediante la regresión de la variable *FatalityRate* sobre las variables *sb_useage*, *speed65*, *speed70*, *ba08*, *drinkage21*, $\ln(income)$, y *age*. ¿La regresión estimada sugiere que un mayor uso del cinturón de seguridad reduce las muertes?
- b) ¿Cambian los resultados cuando se añaden los efectos fijos individuales de cada estado? Proporcione una explicación intuitiva de por qué los resultados cambian.

³ Estos datos fueron proporcionados por el profesor John Donahue de la Universidad de Stanford y se utilizaron en su artículo con Ian Ayres, «Shooting Down the ‘More Guns Less Crime’ Hypothesis», *Stanford Law Review*, 2003, 55: 1193-1312.

⁴ Estos datos fueron proporcionados por el profesor Liran Einav de la Universidad de Stanford y se utilizaron en su artículo con Alma Cohen, «The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities», *The Review of Economics and Statistics*, 2003, 85(4): 828-843.

- c) ¿Cambian los resultados cuando se agregan efectos fijos temporales más los efectos fijos individuales de cada estado?
- d) ¿Qué especificación de la regresión —(a), (b) o (c)— es más fiable? Explique por qué.
- e) Utilizando los resultados de (c), analice el tamaño del coeficiente de la variable *sb_useage*. ¿Es grande? ¿Es pequeño? ¿Cuántas vidas se salvarían si el uso del cinturón de seguridad aumentara del 52 % al 90 %?
- f) Existen dos maneras de aplicar las leyes de obligatoriedad del uso del cinturón de seguridad: la «básica» mediante la cual la aplicación significa que un oficial de policía puede detener un coche y multar al conductor si el oficial observa que un ocupante no lleva puesto el cinturón de seguridad, y la «secundaria», mediante la cual la aplicación significa que un oficial de policía puede poner una multa si un ocupante no lleva puesto el cinturón de seguridad, pero debe existir otra razón para poder detener el coche. En el conjunto de datos, *primary* es una variable binaria para la aplicación básica y *secondary* es una variable binaria para la aplicación secundaria. Realice una regresión de la variable *sb_useage* sobre las variables *primary*, *secondary*, *speed65*, *speed70*, *ba08*, *drinkage21*, *ln(income)*, y *age*, incluyendo efectos fijos individuales del estado y temporales en la regresión. ¿La aplicación básica lleva a un mayor uso del cinturón de seguridad? ¿Qué pasa con la aplicación secundaria?
- g) En 2000, Nueva Jersey cambió de la aplicación secundaria a la aplicación básica. Estime el número de vidas salvadas por año, de la realización de este cambio.

APÉNDICE

10.1 La base de datos estatales sobre mortalidad en accidentes de tráfico

Los datos corresponden a los 48 estados contiguos de los EE.UU. (excluyendo Alaska y Hawaii), son datos anuales desde 1982 hasta 1988. La tasa de mortalidad de tráfico es el número de muertes en accidentes de tráfico en un determinado estado en un año dado, por cada 10.000 personas que viven en ese estado durante ese año. Los datos de mortalidad en accidentes de tráfico se obtuvieron del Sistema de Notificación de Accidentes Mortales del Departamento de Transporte de EE.UU. El impuesto sobre la cerveza (el impuesto sobre una caja de cerveza) se obtuvo del Almanaque de Cerveceros del Instituto de la Cerveza. Las variables de edad legal mínima para beber alcohol de la Tabla 10.1 son variables binarias que indican si la edad mínima legal para beber alcohol es de 18, 19 o 20 años. La variable binaria de castigo en la Tabla 10.1 describe los requisitos mínimos de la sentencia dentro de un estado para una primera condena por conducir ebrio: esta variable es igual a 1 si el estado exige tiempo en prisión o servicios a la comunidad y es igual a 0 en otro caso (una pena menor). Los datos sobre el total de las millas recorridas anualmente por el vehículo por estado se obtuvieron del Departamento de Transporte. La renta personal se obtuvo de la Oficina de Análisis Económico de EE.UU., y la tasa de desempleo se obtuvo de la Oficina de Estadísticas Laborales de EE.UU. Estos datos fueron proporcionados amablemente por el profesor Christopher J. Ruhm del Departamento de Economía de la Universidad de Carolina del Norte.

APÉNDICE

10.2 Errores estándar de la regresión de efectos fijos

Este apéndice proporciona las fórmulas de los errores estándar de la regresión de efectos fijos con un solo regresor. Estas fórmulas se extienden al caso de múltiples variables explicativas en el Ejercicio 18.15.

La distribución asintótica del estimador de efectos fijos con n grande

El estimador de efectos fijos. El estimador de efectos fijos de β_1 es el estimador MCO obtenido mediante la regresión en desviaciones respecto de la media individual de la Ecuación (10.14), en el que se regresa \tilde{Y}_{it} sobre \tilde{X}_{it} ,

donde $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$, $\bar{Y}_i = T^{-1} \sum_{t=1}^T Y_{it}$ y $\bar{X}_i = T^{-1} \sum_{t=1}^T X_{it}$. La fórmula para el estimador MCO se obtiene mediante la sustitución de $X_i - \bar{X}$ por \tilde{X}_{it} e $Y_i - \bar{Y}$ por \tilde{Y}_{it} en la Ecuación (4.7) y mediante la sustitución del sumatorio simple de la Ecuación (4.7) por dos sumatorios, uno sobre las entidades individuales ($i = 1, \dots, n$) y otro sobre los períodos de tiempo ($t = 1, \dots, T$)⁵, por lo que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \tilde{Y}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^2} \quad (10.22)$$

La obtención de la distribución muestral de $\hat{\beta}_1$ es paralela a la deducción del Apéndice 4.3 para la distribución muestral del estimador MCO con datos de sección cruzada. En primer lugar, se sustituye $\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$ [Ecuación (10.14)] en el numerador de la Ecuación (10.22) para obtener la homóloga para datos de panel de la Ecuación (4.30):

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \tilde{u}_{it}}{\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^2} \quad (10.23)$$

A continuación, reorganizando esta expresión y multiplicando ambos lados por \sqrt{nT} se obtiene

$$\sqrt{nT}(\hat{\beta}_1 - \beta_1) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \eta_i}}{\hat{Q}_{\tilde{X}}}, \text{ donde } \eta_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \tilde{X}_{it} \tilde{u}_{it}} \text{ y } \hat{Q}_{\tilde{X}} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^2. \quad (10.24)$$

El factor de escala de la Ecuación (10.24), nT , es el número total de observaciones.

Distribución y errores estándar cuando n es grande. En la mayoría de las aplicaciones de datos de panel n es mucho mayor que T , lo que motiva la aproximación de las distribuciones muestrales cuando $n \rightarrow \infty$, manteniendo T fijo. Bajo los supuestos de la regresión de efectos fijos del Concepto clave 10.3, $\hat{Q}_{\tilde{X}} \xrightarrow{P} Q_{\tilde{X}} = ET^{-1} \sum_{t=1}^T \tilde{X}_{it}^2$ cuando $n \rightarrow \infty$. Además, η_i es i.i.d. para $i = 1, \dots, n$ (por el Supuesto 2) con media cero (por el Supuesto 1) y varianza σ_η^2 (que es finita por el Supuesto 3), por lo que por el teorema central del límite $\sqrt{1/n} \sum_{i=1}^n \eta_i \xrightarrow{d} N(0, \sigma_\eta^2)$. Se desprende de la Ecuación (10.24) que

$$\sqrt{nT}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} -N\left(0, \frac{\sigma_\eta^2}{Q_{\tilde{X}}^2}\right) \quad (10.25)$$

De acuerdo con la Ecuación (10.25), la varianza de la distribución para muestras grandes de $\hat{\beta}_1$ es

$$\text{var}(\hat{\beta}_1) = \frac{1}{nT} \frac{\sigma_\eta^2}{Q_{\tilde{X}}^2} \quad (10.26)$$

La fórmula de los errores estándar agrupados sustituye los momentos poblacionales en la Ecuación (10.26) por sus homólogos muestrales:

$$ES(\hat{\beta}_1) = \sqrt{\frac{1}{nT} \frac{s_\eta^2}{\hat{Q}_{\tilde{X}}^2}} \quad (10.27)$$

$$\text{donde } s_\eta^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{\eta}_i - \bar{\eta})^2 = \frac{1}{n-1} \sum_{i=1}^n \hat{\eta}_i^2 \text{ (errores estándar agrupados)}$$

⁵ El doble sumatorio es la extensión a dos subíndices de un sumatorio sencillo:

$$\begin{aligned} \sum_{i=1}^n \sum_{t=1}^T X_{it} &= \sum_{i=1}^n \left(\sum_{t=1}^T x_{it} \right) \\ &= \sum_{i=1}^n (X_{i1} + X_{i2} + \dots + X_{iT}) \\ &= (X_{11} + X_{12} + \dots + X_{1T}) + (X_{21} + X_{22} + \dots + X_{2T}) + \dots + (X_{n1} + X_{n2} + \dots + X_{nT}). \end{aligned}$$

donde $\hat{\eta}_i = \sqrt{1/T} \sum_{t=1}^T \tilde{X}_{it} \hat{u}_{it}$ es el homólogo muestral de η_i [$\hat{\eta}_i$ es η_i en la Ecuación (10.24), con \tilde{u}_{it} reemplazado por el residuo \hat{u}_{it} de la regresión de efectos fijos], y $\bar{\hat{\eta}} = (1/n) \sum_{i=1}^n \hat{\eta}_i$. La última igualdad de la Ecuación (10.27) se debe a que $\bar{\hat{\eta}} = 0$, lo que a su vez se desprende de que los residuos y los regresores no están correlacionados [Ecuación (4.34)]. Téngase en cuenta que s_{η}^2 es la varianza muestral de $\hat{\eta}_i$ [véase la Ecuación (3.7)].

El estimador s_{η}^2 es un estimador consistente de σ_{η}^2 cuando $n \rightarrow \infty$ incluso en presencia de heterocedasticidad o autocorrelación (Ejercicio 18.15); por tanto el error estándar agrupado de la Ecuación (10.27) es consistente a heterocedasticidad y autocorrelación. Debido a que el error estándar agrupado es consistente, el estadístico t del contraste de $\beta_1 = \beta_{1,0}$ tiene una distribución normal estándar bajo la hipótesis nula cuando $n \rightarrow \infty$.

Todos los resultados anteriores son aplicables con regresores múltiples. Además, si n es grande, el estadístico F para el contraste de q restricciones (calculado mediante la fórmula de la varianza agrupada) tiene su distribución asintótica habitual $F_{q,\infty}$.

¿Por qué no es válido el estimador habitual heterocedástico-robusto del Capítulo 5 para datos de panel? Existen dos razones. La razón más importante es que el estimador heterocedástico-robusto del Capítulo 5 no permite la correlación serial dentro de un grupo (cluster). Recordemos que, para dos variables aleatorias U y V , $\text{var}(U + V) = \text{var}(U) + \text{var}(V) + 2\text{cov}(U, V)$. La varianza η_i en la Ecuación (10.24) puede escribirse como la suma de las varianzas más las covarianzas. Sea $\tilde{v}_{it} = \tilde{X}_{it} \tilde{u}_{it}$; entonces

$$\begin{aligned}\text{var}(\eta_i) &= \text{var}\left(\sqrt{\frac{1}{T} \sum_{t=1}^T \tilde{v}_{it}}\right) = \frac{1}{T} \text{var}(\tilde{v}_{i1} + \tilde{v}_{i2} + \dots + \tilde{v}_{iT}) \\ &= \frac{1}{T} [\text{var}(\tilde{v}_{i1}) + \text{var}(\tilde{v}_{i2}) + \dots + \text{var}(\tilde{v}_{iT}) \\ &\quad + 2\text{cov}(\tilde{v}_{i1}, \tilde{v}_{i2}) + \dots + 2\text{cov}(\tilde{v}_{iT-1}, \tilde{v}_{iT})].\end{aligned}\tag{10.28}$$

La fórmula de la varianza heterocedástico-robusta del Capítulo 5 obvia todas las covarianzas de la parte final de la Ecuación (10.28), por lo que si existe correlación serial, el estimador heterocedástico-robusto de la varianza es inconsistente.

La segunda razón es que si T es pequeño, la estimación de los efectos fijos introduce sesgo en el estimador heterocedástico-robusto de la varianza del Capítulo 5. Este problema no surge en la regresión de sección cruzada.

El único caso en el que pueden utilizarse los errores estándar heterocedástico-robustos habituales con datos de panel es con la regresión de efectos fijos con $T = 2$ observaciones. En este caso, la regresión de efectos fijos es equivalente a la regresión de las diferencias «antes y después» de la Sección 10.2, y los errores estándar heterocedástico-robustos y los errores estándar agrupados son equivalentes.

Para ver ejemplos empíricos que demuestran la importancia de utilizar los errores estándar agrupados en los datos económicos de panel, se puede recurrir a Bertrand, Duflo y Mullainathan (2004).

Errores estándar cuando u_{it} está correlacionada entre las distintas entidades individuales. En algunos casos, u_{it} puede estar correlacionado entre las distintas entidades individuales. Por ejemplo, en un estudio de los ingresos salariales, supongamos que el plan de muestreo selecciona familias mediante muestreo aleatorio simple, y a continuación, se realiza un seguimiento de todos los hermanos dentro de una familia. Debido a que los factores omitidos que constituyen el término de error podrían contener elementos comunes a los hermanos, no es razonable suponer que los errores son independientes para los hermanos (incluso a pesar de que sean independientes entre las familias).

En el ejemplo de los hermanos, las familias son grupos naturales, o agrupaciones (*clusters*), de observaciones, donde u_{it} está correlacionado dentro de un grupo, pero no entre los grupos o *clusters*. La deducción que lleva a la Ecuación (10.27) puede ser modificada para permitir agrupaciones entre las entidades individuales (por ejemplo, las familias) o tanto entre entidades como en el tiempo, siempre y cuando haya muchos grupos

Distribución y errores estándar cuando n es pequeño

Si n es pequeño y T es grande, sigue siendo posible utilizar los errores estándar agrupados; sin embargo, es necesario comparar los estadísticos t con los valores críticos de las tablas de la t_{n-1} , y el estadístico F para el contraste de q restricciones tiene que ser comparado con el valor crítico de la distribución $F_{q,n-q}$ multiplicado por $(n-1)/(n-q)$. Estas distribuciones son válidas bajo los supuestos del Concepto clave 10.3, además de algunos supuestos adicionales sobre la distribución conjunta de X_{it} y u_{it} en el tiempo dentro de una entidad individual. Aunque la validez de la distribución t en la regresión de sección cruzada requiere normalidad y homocedasticidad en los errores de regresión (Sección

5.6), no es necesario ninguno de estos requisitos para justificar el uso de la distribución t con errores estándar agrupados para datos de panel cuando T es grande.

Para ver por qué el estadístico t agrupado presenta una distribución t_{n-1} cuando n es pequeño y T es grande, incluso si u_{it} ni está normalmente distribuido ni es homocedástico, en primer lugar hay que darse cuenta de que si T es grande, bajo los supuestos adicionales η_i de la Ecuación (10.24) obedecerá al teorema central del límite, por lo que $\eta_i \xrightarrow{d} N(0, \sigma_\eta^2)$. (Los supuestos adicionales necesarios para este resultado son importantes y técnicos y debe aplazarse el debate hasta haber tratado los datos de series temporales en el Capítulo 14). Por lo tanto, si T es grande, entonces $\sqrt{nT}(\hat{\beta}_1 - \beta_1)$ en la Ecuación (10.24) es una media escalada de las n variables aleatorias normales η_i . Además, la fórmula agrupada s_η^2 de la Ecuación (10.27) es la fórmula habitual para la varianza muestral, y si pudiera ser calculada utilizando η_i , entonces $(n-1)s_\eta^2/\sigma_\eta^2$, tendría una distribución χ_{n-1}^2 , por lo que el estadístico t tendría una distribución t_{n-1} [véase la Sección (3.6)]. La utilización de los residuos para calcular $\hat{\eta}_i$ y s_η^2 no cambia esta conclusión. En el caso con múltiples variables explicativas, un razonamiento análogo lleva a la conclusión de que el estadístico F para contrastar q restricciones, calculado utilizando el estimador de la varianza agrupada, se distribuye como $\frac{(n-1)}{(n-q)}F_{q, n-q}$. [Por ejemplo, el valor crítico del 5 % para este estadístico F cuando $n = 10$ y $q = 4$ es $\frac{(10-1)}{(10-4)} \times 4,53 = 6,80$, donde 4,53 es el valor crítico al 5 % de la distribución $F_{4,6}$ dado en la Tabla 5B del Apéndice.] Nótese que, a medida que n aumenta, las distribuciones t_{n-1} y $\frac{(n-1)}{(n-q)}F_{q, n-q}$ se aproximan las distribuciones normal estándar y $F_{q, \infty}$ habituales⁶.

Si tanto n como T son pequeños, entonces en general $\hat{\beta}_1$ no se distribuye normalmente y los errores estándar agrupados no proporcionan inferencia fiable.

⁶ No todo el software lleva a cabo errores estándar agrupados utilizando las distribuciones t_{n-1} y $\frac{(n-1)}{(n-q)}F_{q, n-q}$ que se aplican si n es pequeño, por lo que debe comprobarse cómo el software realiza y trata los errores estándar agrupados.

Regresión con variable dependiente binaria

Do^s personas idénticas, excepto en su etnia, caminan hacia un banco y solicitan una hipoteca, un préstamo grande con el que cada uno puede comprar una casa idéntica. ¿Les trata el banco de la misma forma? ¿Tienen ambos la misma probabilidad de que su solicitud de hipoteca les sea concedida? Legalmente deben recibir un tratamiento idéntico. Pero el hecho de si en realidad lo obtienen o no es un asunto de gran preocupación para los reguladores bancarios.

Los préstamos se otorgan y se deniegan por muchas razones legítimas. Por ejemplo, si los pagos del préstamo propuestos constituyen la mayor parte o la totalidad de los ingresos mensuales del solicitante, un empleado de un banco podría denegar con razón la concesión del préstamo. Además, hasta los empleados de los bancos son humanos y pueden cometer errores sinceros, por lo que la denegación a un solicitante que pertenece a una minoría no prueba nada acerca de la discriminación. Muchos estudios sobre discriminación, por lo tanto, buscan evidencia estadística de la existencia de discriminación, es decir, pruebas que figuran en grandes conjuntos de datos que muestren que los blancos y las minorías son tratados de manera diferente.

Pero, ¿cómo se verificaría exactamente la evidencia estadística acerca de la discriminación en el mercado hipotecario? Un comienzo es comparar la proporción de solicitantes blancos y de minorías a los que se les negó un préstamo hipotecario. En los datos analizados en este capítulo, tomados de las solicitudes de hipotecas en 1990 en el área de Boston, Massachusetts, al 28 % de los solicitantes negros se les denegó el préstamo hipotecario, pero solamente se les denegó al 9 % de los solicitantes blancos. Pero esta comparación no responde realmente a la pregunta que abre este capítulo, debido a que los solicitantes negros y blancos no eran candidatos necesariamente «idénticos, excepto en su etnia». En su lugar, necesitamos un método para comparar las tasas de rechazo, *manteniendo las otras características constantes*.

Esto suena como una tarea para el análisis de regresión múltiple, y lo es, pero con una peculiaridad. La peculiaridad es que la variable dependiente —si la solicitud se deniega— es binaria. En la Parte II, se utilizan habitualmente variables binarias como regresores, lo que no causa problemas particulares. Sin embargo, cuando la variable dependiente es binaria, las cosas son más difíciles: ¿Qué significa ajustar una recta a una variable dependiente que solo puede adoptar dos valores, 0 y 1?

La respuesta a esta pregunta es la de interpretar la función de regresión como una predicción de probabilidad. Esta interpretación se discute en la Sección 11.1, y nos permite aplicar los modelos de regresión múltiple de la Parte II al caso de variables dependientes binarias. La Sección 11.1 examina este «modelo de probabilidad lineal». Pero la interpretación como predicción de probabilidad asimismo sugiere que la alternativa, los modelos de regresión no lineales pueden hacerlo mejor al modelizar estas probabilidades. Estos métodos, denominados regresiones «probit» y «logit», se examinan en la Sección 11.2. La Sección 11.3, que es opcional, analiza el método utilizado para estimar los coeficientes de las regresiones probit y logit, el método de estimación de máxima verosimilitud. En la Sección 11.4, se aplican estos métodos al conjunto de datos de solicitudes de préstamos hipotecarios de Boston para comprobar si existe evidencia de la existencia de prejuicios raciales en la concesión de préstamos hipotecarios.

La variable dependiente binaria considerada en este capítulo es un ejemplo de una variable dependiente con rango limitado; en otras palabras, se trata de una **variable dependiente limitada**. Los mo-

delos para otros tipos de variables dependientes limitadas, por ejemplo, las variables dependientes que toman varios valores discretos, se examinan en el Apéndice 11.3.

11.1 Variables dependientes binarias y modelo de probabilidad lineal

La aceptación o denegación de una solicitud de hipoteca es un ejemplo de variable binaria. Asimismo, otras muchas cuestiones importantes presentan alternativas binarias. ¿Cuál es el efecto de una subvención a la matrícula sobre la decisión de un individuo de ir a la universidad? ¿Qué determina que un adolescente empiece o no a fumar? ¿Qué determina si un país recibe o no ayuda exterior? ¿Qué determina si un solicitante de empleo tiene o no éxito? En todos estos ejemplos, el resultado relevante es binario: el estudiante irá o no irá a la universidad, el adolescente empezará o no a fumar, un país recibirá o no ayuda extranjera, el solicitante conseguirá o no un trabajo.

Esta sección trata sobre la distinción entre la regresión con una variable dependiente binaria y la regresión con una variable dependiente continua, y trata posteriormente el modelo más sencillo que se puede utilizar con variables dependientes binarias, el modelo de probabilidad lineal.

Variables dependientes binarias

El caso que se examina en este capítulo es si la pertenencia a un grupo étnico constituye un factor relevante para la denegación de una solicitud de hipoteca; la variable dependiente binaria es si se deniega o no una solicitud de hipoteca. Los datos son un subconjunto de una base de datos mayor recopilado por los investigadores del Banco de la Reserva Federal de Boston, bajo el amparo de la Ley de Divulgación de Hipotecas (HMDA) y se refieren a las solicitudes de hipotecas recogidas en el área de Boston, Massachusetts, en 1990. Los datos HMDA de Boston se describen en el Apéndice 11.1.

Las solicitudes de hipotecas son complicadas y por tanto lo es el proceso por el cual el empleado del banco toma una decisión. El empleado del banco debe predecir si el solicitante cumplirá con los pagos de su préstamo. Un aspecto importante de la información es la cuantía de los pagos requeridos por el préstamo en relación con la renta del solicitante. Como cualquier persona que haya pedido prestado dinero, sabe que resulta mucho más fácil realizar los pagos si suponen un 10 % de los ingresos ¡que si suponen un 50 %! Por lo tanto, comenzaremos por examinar la relación entre dos variables: la variable dependiente binaria *denegar*, que es igual a 1 si la solicitud de hipoteca fue denegada y es igual a 0 si fue aceptada, y la variable continua *ratio P/I*, que es el cociente entre el total de los pagos mensuales del préstamo del solicitante y sus ingresos mensuales.

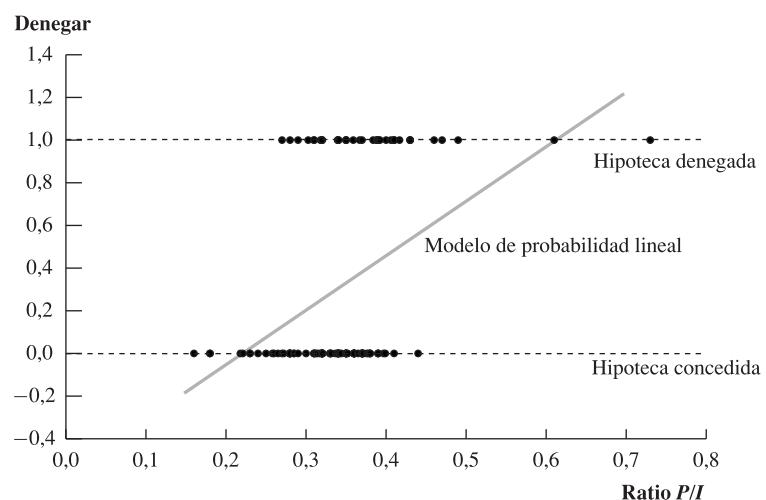
La Figura 11.1 representa un diagrama de dispersión de la variable *denegar* versus la variable *ratio P/I* para 127 de las 2.380 observaciones del conjunto de datos. (El diagrama de dispersión resulta más fácil de leer utilizando este subconjunto de datos). Este diagrama de dispersión tiene un aspecto diferente de los diagramas de dispersión de la Parte II, debido a que la variable *denegar* es binaria. Aun así, parece mostrar que existe una relación entre la variable *denegar* y la variable *ratio P/I*: unos pocos solicitantes con una ratio de pagos sobre ingresos inferior a 0,3 ven denegadas sus solicitudes, pero a la mayoría de los solicitantes con una proporción de pagos-ingresos superior a 0,4 se les deniega.

Esta relación positiva entre las variables *ratio P/I* y *denegar* (cuanto mayor es la variable *ratio P/I*, mayor es la proporción de denegaciones) se recoge en la Figura 11.1 mediante la recta de regresión estimada MCO utilizando estas 127 observaciones. Como de costumbre, esta recta representa gráficamente los valores de predicción de la variable *denegar* como función del regresor, la proporción entre los pagos y los ingresos. Por ejemplo, cuando la variable *ratio P/I* = 0,3, el valor estimado para la variable *denegar* es de 0,20. ¿Pero qué significa exactamente que el valor de predicción para la variable binaria *denegar* sea 0,20?

La clave para responder a esta pregunta —y en general para comprender la regresión con una variable dependiente binaria— consiste en interpretar la regresión como la modelización de la *probabilidad* de que la variable dependiente sea igual a 1. Por tanto, el valor de 0,20 estimado se puede interpretar en el sentido

FIGURA 11.1 Diagrama de dispersión de las denegaciones en solicitudes de hipoteca y ratio pagos-ingresos

Los solicitantes de hipotecas con una proporción de pagos de la deuda sobre ingresos (ratio P/I) elevada son más propensos a obtener una denegación de su solicitud (denegar = 1 si es denegada, denegar = 0 si es aprobada). El modelo de probabilidad lineal utiliza una línea recta para modelizar la probabilidad de denegación.



de que, cuando la variable *ratio P/I* es 0,3, se estima que la probabilidad de denegación es un 20 %. Dicho de otra manera, si existen muchas solicitudes con *ratio P/I* = 0,3, se denegarán el 20 % de ellas.

Esta interpretación se desprende de dos hechos. En primer lugar, de acuerdo con lo tratado en la Parte II, la función de regresión poblacional es el valor esperado de Y dados los regresores, $E(Y|X_1, \dots, X_k)$. En segundo lugar, de acuerdo con la Sección 2.2, si Y es una variable binaria 0 – 1, su valor esperado (o media) es la probabilidad de que $Y = 1$; es decir, $E(Y) = 0 \times \Pr(Y = 0) + 1 \times \Pr(Y = 1) = \Pr(Y = 1)$. En el contexto de regresión el valor esperado está condicionado al valor de los regresores, por lo que la probabilidad está condicionada a X . Por tanto, para una variable binaria, $E(Y|X_1, \dots, X_k) = \Pr(Y = 1|X_1, \dots, X_k)$. En resumen, para una variable dependiente binaria, el valor de predicción calculado a partir de la regresión poblacional es la probabilidad de que $Y = 1$, dado X .

El modelo lineal de regresión múltiple aplicado a una variable dependiente binaria se denomina modelo de probabilidad lineal: «lineal», debido a que es una línea recta y «modelo de probabilidad» porque modeliza la probabilidad de que la variable dependiente sea igual a 1; en nuestro ejemplo, la probabilidad de denegación del préstamo.

El modelo de probabilidad lineal

El **modelo de probabilidad lineal** es el nombre que se utiliza para el modelo de regresión múltiple de la Parte II cuando la variable dependiente es binaria en vez de continua. Debido a que la variable dependiente Y es binaria, la función de regresión poblacional corresponde a la probabilidad de que la variable dependiente sea igual a 1, dado X . El coeficiente poblacional β_1 de un regresor X es *variación en la probabilidad* de que $Y = 1$ asociada con una *variación unitaria* en X . Del mismo modo, el valor de predicción MCO, \hat{Y}_i , calculado mediante la función de regresión estimada, es la probabilidad estimada de que la variable dependiente sea igual a 1, y el estimador MCO $\hat{\beta}_1$ estima la variación de la probabilidad de que $Y = 1$ asociada a un cambio unitario en X .

Casi todas las herramientas de la Parte II son traspasables al modelo de probabilidad lineal. Los coeficientes pueden estimarse por MCO. Pueden construirse los intervalos de confianza al noventa y cinco por ciento como $\pm 1,96$ los errores estándar, las hipótesis sobre varios coeficientes pueden contrastarse utilizando el estadístico F del Capítulo 7, y las interacciones entre las variables se pueden modelizar utilizando los métodos de la Sección 8.3. Debido a que los errores del modelo de probabilidad lineal son siempre heterocedásticos (Ejercicio 11.8), es esencial que los errores estándar que se utilicen para inferencia sean los heterocedástico-robustos.

**CONCEPTO
CLAVE**
El modelo de probabilidad lineal

El modelo de probabilidad lineal es el modelo lineal de regresión múltiple,

11.1

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad (11.2)$$

aplicado a una variable dependiente binaria Y_i . Debido a que Y es binaria, $E(Y|X_1, X_2, \dots, X_k) = \Pr(Y = 1|X_1, X_2, \dots, X_k)$, por lo que el modelo de probabilidad lineal,

$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k.$$

El coeficiente de regresión β_1 es la variación de la probabilidad de que $Y = 1$ asociada con una variación unitaria en X_1 , manteniendo constantes las otras variables explicativas, y sucesivamente para β_2, \dots, β_k . Los coeficientes de la regresión se pueden estimar por MCO, y se pueden utilizar los errores MCO habituales (heterocedástico-robustos) para los intervalos de confianza y los contrastes de hipótesis.

Una herramienta que no es transferible es el R^2 . Cuando la variable dependiente es continua, es posible imaginar una situación en la que el R^2 sea igual a 1: todos los datos se encuentran exactamente en la recta de regresión. Esto es imposible cuando la variable dependiente es binaria, a menos que los regresores sean igualmente binarios. Por tanto, el R^2 no es un estadístico particularmente útil aquí. Volveremos a las medidas de ajuste en la sección siguiente. El modelo de probabilidad lineal se resume en el Concepto clave 11.1.

Aplicación a los datos HMDA de Boston. La regresión MCO de la variable dependiente binaria, *denegar*, sobre la proporción de pagos sobre ingresos, *ratio P/I*, estimada utilizando el total de las 2.380 observaciones de nuestra base de datos es

$$\widehat{\text{denegar}} = -0,080 + 0,604 \text{ ratio P/I}. \quad (11.1)$$

(0,032) (0,098)

El coeficiente estimado para la variable *ratio P/I* es positivo, y el coeficiente poblacional es estadística y significativamente diferente de cero al nivel del 1 % (el estadístico *t* es 6,13). Por tanto, los solicitantes con pagos de deuda más altos en proporción a los ingresos tienen más probabilidad de que su solicitud sea denegada. Este coeficiente puede utilizarse para calcular la variación prevista en la probabilidad de denegación, dado una variación en el regresor. Por ejemplo, de acuerdo con la Ecuación (11.1), si la variable *ratio P/I* aumenta en 0,1, la probabilidad de denegación aumenta en $0,064 \times 0,1 \cong 0,060$, es decir, 6,0 puntos porcentuales.

El modelo de probabilidad lineal estimado en la Ecuación (11.1) se puede utilizar para calcular las probabilidades de denegación estimada en función de la variable *ratio P/I*. Por ejemplo, si se prevén unos pagos de deuda de un 30 % de los ingresos de un solicitante, la variable *ratio P/I* es 0,3 y el valor teórico según la Ecuación (11.1) es $-0,080 + 0,604 \times 0,3 = 0,101$. Es decir, de acuerdo con este modelo de probabilidad lineal, un solicitante cuyos pagos de la deuda se espera que constituyan un 30 % de sus ingresos tiene una probabilidad del 10,1 % de que su solicitud sea rechazada. [La cual es diferente de la probabilidad del 20 % basada en la recta de regresión de la Figura 11.1, debido a que esta recta se calculaba utilizando únicamente 127 de las 2.380 observaciones utilizadas para estimar la Ecuación (11.1).]

¿Cuál es el efecto de la etnia sobre la probabilidad de denegación, manteniendo constante la variable *ratio P/I*? Para simplificar las cosas, nos centramos en las diferencias entre los solicitantes negros y blancos. Para estimar el efecto de la etnia manteniendo constante la variable *ratio P/I*, la Ecuación (11.1) se amplía con un regresor binario que es igual a 1 si el solicitante es negro y es igual a 0 si el solicitante es blanco. El modelo de probabilidad lineal estimado es

$$\widehat{\text{denegar}} = -0,091 + 0,559 \text{ratioP/I} + 0,177 \text{negro}. \quad (11.3)$$

(0,029) (0,089) (0,025)

El coeficiente de *negro*, 0,177, indica que un solicitante afroamericano tiene una probabilidad mayor en un 17,7 % de obtener una solicitud de hipoteca denegada que un blanco, manteniendo constante la proporción de pagos sobre ingresos. Este coeficiente es significativo al nivel del 1 % (el estadístico t es 7,11).

Interpretada de modo literal, esta estimación sugiere que podría haber un sesgo racial en las decisiones sobre concesiones de hipoteca, pero tal conclusión sería prematura. Aunque el cociente pagos-ingresos desempeña un papel en la decisión del empleado del banco, asimismo lo hacen otros muchos factores, como por ejemplo los ingresos potenciales del solicitante y el historial crediticio del individuo. Si alguna de estas variables estuviera correlacionada con los regresores *negro* o *ratio P/I*, su omisión de la Ecuación (11.3) causaría sesgo de variable omitida. Por tanto debe posponerse la extracción de cualquier conclusión acerca de la discriminación en la concesión de préstamos hipotecarios hasta que se complete un análisis en mayor profundidad en la Sección 11.3.

Deficiencias del modelo de probabilidad lineal. La linealidad que hace que el modelo de probabilidad lineal sea fácil de utilizar es del mismo modo su mayor defecto. Debido a que las probabilidades no pueden exceder el valor 1, el efecto sobre la probabilidad de que $Y = 1$ de una variación en X determinada debe ser no lineal: a pesar de que una variación en la variable *ratio P/I* desde 0,3 hasta 0,4 podría tener un gran efecto en la probabilidad de denegación, una vez que la variable *ratio P/I* sea tan grande que el préstamo es muy probable que se deniegue, el aumento en la variable *ratio P/I* tendrá en cambio un efecto escaso. Por el contrario, en el modelo de probabilidad lineal, el efecto de una variación determinada de la variable *ratio P/I* es constante, lo que conduce a predicciones sobre las probabilidades en la Figura 11.1 que son ¡inferiores a 0 para valores muy bajos de la variable *ratio P/I* y superiores a 1 y para los valores altos! Pero esto es un disparate: la probabilidad no puede ser menor que 0 o mayor que 1. Esta característica sin sentido es una consecuencia inevitable de la regresión lineal. Para hacer frente a este problema, se introducen nuevos modelos no lineales diseñados específicamente para variables dependientes binarias, los modelos de regresión probit y logit.

11.2 Regresión probit y logit

Las regresiones **probit** y **logit**¹ son modelos de regresión no lineales diseñados específicamente para variables dependientes binarias. Debido a que una regresión con una variable dependiente binaria Y modeliza la probabilidad de que $Y = 1$, tiene sentido adoptar una formulación no lineal que obligue a que los valores estimados estén entre 0 y 1. Debido a que las funciones de distribución de probabilidad acumuladas (f.d.a.) dan lugar a probabilidades entre 0 y 1 (Sección 2.1), éstas se utilizan en las regresiones logit y probit. La regresión probit utiliza la f.d.a. normal estándar. La regresión logit, asimismo denominada **regresión logística**, utiliza la f.d.a. «logística».

Regresión probit

Regresión probit con un único regresor. El modelo de regresión probit con un solo regresor X es

$$\Pr(Y = 1 | X) = \Phi(\beta_0 + \beta_1 X), \quad (11.4)$$

donde Φ es la función de distribución de probabilidad acumulada normal estándar (tabulada en la Tabla 1 del Apéndice).

Por ejemplo, supongamos que Y es la variable binaria de denegación de la solicitud de hipoteca, *denegar*, X es la proporción de pagos-ingresos (*ratio P/I*), $\beta_0 = -2$, y $\beta_1 = 3$. ¿Cuál es la probabilidad de denegación, si *ratio P/I* = 0,4? De acuerdo con la Ecuación (11.4), esta probabilidad es $\Phi(\beta_0 + \beta_1 \text{ratio P/I}) = \Phi(-2 + 3 \times 0,4) = \Phi(-0,8)$. De acuerdo con la tabla de la distribución normal acumulada (Tabla 1 del

¹ Pronunciado pro-bit y lo-yit.

Apéndice), $\Phi(-0,8) = \Pr(Z \leq -0,8) = 21,2\%$. Es decir, cuando *ratio P/I* es igual a 0,4, la probabilidad estimada de que la solicitud sea denegada es del 21,2 %, calculada utilizando el modelo probit con los coeficientes $\beta_0 = -2$, y $\beta_1 = 3$.

En el modelo probit, el término $\beta_0 + \beta_1 X$, desempeña el papel de «*z*» en la distribución normal estándar acumulada de la Tabla 1 del Apéndice. Por tanto, el cálculo del párrafo anterior puede realizarse, de forma equivalente, calculando en primer lugar el «*z*-valor», $z = \beta_0 + \beta_1 X = -2 + 3 \times 0,4 = -0,8$, y más tarde buscando la probabilidad en la cola de la distribución normal a la izquierda de $z = -0,8$, que es el 21,2 %.

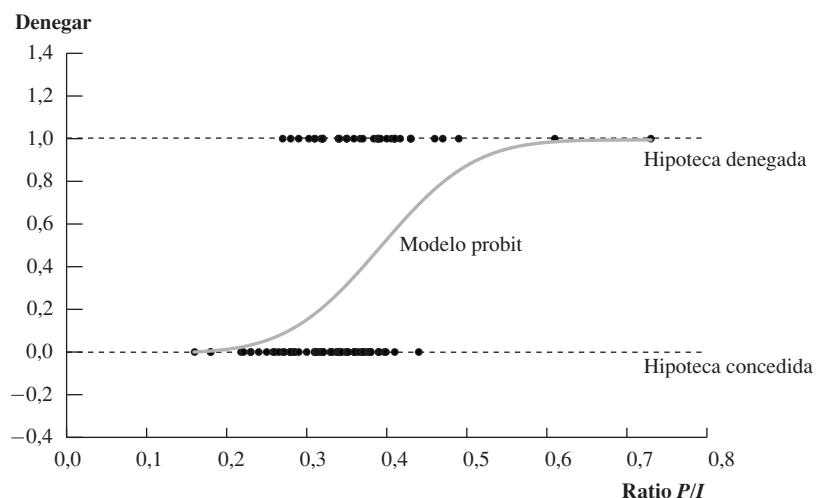
El coeficiente probit β_1 en la Ecuación (11.4) es la variación en el *z*-valor asociada con una variación unitaria en *X*. Si β_1 es positivo, un aumento en *X* aumenta el *z*-valor y por lo tanto aumenta la probabilidad de que $Y = 1$; si β_1 es negativo, un incremento de *X* disminuye la probabilidad de que $Y = 1$. Aunque el efecto de *X* sobre el *z*-valor es lineal, su efecto sobre la probabilidad es no lineal. Por tanto en la práctica la manera más fácil de interpretar los coeficientes de un modelo probit es calcular la probabilidad estimada, o la variación en la probabilidad estimada, para uno o más valores de los regresores. Cuando solo hay un regresor, la probabilidad estimada puede representarse como una función de *X*.

La Figura 11.2 representa la función de regresión estimada a la que da lugar la regresión probit de la variable *denegar* sobre la variable *ratio P/I* para las 127 observaciones del diagrama de dispersión. La función de regresión probit estimada tiene una forma de «S» alargada: es cercana a 0 y plana para valores pequeños de la variable *ratio P/I*, se inclina y aumenta para los valores intermedios, y se aplana de nuevo y se hace cercana a 1 para valores altos. Para valores pequeños de la proporción de pagos sobre ingresos, la probabilidad de denegación es pequeña. Por ejemplo, para un valor de *ratio P/I* = 0,2, la probabilidad estimada de denegación sobre la base de la función probit estimada de la Figura 11.2 es $\Pr(\text{denegar} = 1 | \text{ratio P/I} = 0,2) = 2,1\%$. Cuando la variable *ratio P/I* = 0,3, la probabilidad estimada de denegación es del 16,1 %. Cuando la variable *ratio P/I* = 0,4, la probabilidad de denegación aumenta bruscamente hasta el 51,9 %, y cuando *ratio P/I* = 0,6, la probabilidad de rechazo es del 98,3 %. De acuerdo con este modelo probit estimado, para los solicitantes con una ratio alta de pagos-ingresos, la probabilidad de rechazo es cercana a 1.

Regresión probit con regresores múltiples. En todos los problemas de regresión que hemos estudiado hasta ahora, no tener en cuenta un factor determinante de *Y* que esté correlacionado con los regresores incluidos desemboca en un sesgo de variable omitida. La regresión probit no es una excepción. En la regresión lineal, la solución consiste en incluir la variable adicional como regresor. Esta es igualmente la solución a un sesgo de variable omitida en regresión probit.

FIGURA 11.2 Modelo probit de la probabilidad de denegación. Dada la ratio *P/I*

El modelo probit utiliza la función de distribución acumulada normal para modelizar la probabilidad de denegación dada la proporción pagos-ingresos o, de forma más general, para modelizar $\Pr(Y = 1 | X)$. A diferencia del modelo de probabilidad lineal, las probabilidades condicionales del probit toman siempre valores entre 0 y 1.



El modelo probit con varios regresores amplía el modelo probit de un solo regresor mediante la adición de variables explicativas para calcular el z -valor. En consecuencia, el modelo de regresión poblacional probit con dos variables explicativas, X_1 y X_2 , es

$$\Pr(Y = 1|X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2). \quad (11.5)$$

Por ejemplo, supongamos que $\beta_0 = -1,6$, $\beta_1 = 2$, y $\beta_2 = 0,5$. Si $X_1 = 0,4$ y $X_2 = 1$, el z -valor es $z = -1,6 + 2 \times 0,4 + 0,5 \times 1 = -0,3$. Por lo tanto, la probabilidad de que $Y = 1$ dado $X_1 = 0,4$ y $X_2 = 1$ es $\Pr(Y = 1|X_1 = 0,4, X_2 = 1) = \Phi(-0,3) = 38\%$.

Efecto de una variación en X . En general, el efecto sobre Y de un cambio en X es el cambio esperado en Y que surge de un cambio en X . Cuando Y es binaria, su esperanza condicional es la probabilidad condicional de que sea igual a 1, por lo que el cambio esperado en Y que surge de un cambio en X es el cambio en la probabilidad de que $Y = 1$.

Recordemos de la Sección 8.1 que, cuando la función de regresión poblacional es una función no lineal de X , este cambio esperado se estima en tres pasos: en primer lugar, se calcula el valor esperado en el valor original de X utilizando la función de regresión estimada; más tarde, se calcula el valor esperado sobre el valor de X tras el cambio, $X + \Delta X$; por último, se calcula la diferencia entre los dos valores esperados. Este procedimiento se resume en el Concepto clave 8.1. Como se destacó en la Sección 8.1, este método *siempre* funciona para el cálculo de los efectos esperados de un cambio en X , sin importar lo complicado que sea el modelo. Cuando se aplica al modelo probit, el método del Concepto clave 8.1 da lugar al efecto estimado sobre la probabilidad de que $Y = 1$ de un cambio en X .

El modelo de regresión probit, las probabilidades esperadas, y los efectos estimados se resumen en el Concepto clave 11.2.

**CONCEPTO
CLAVE**

El modelo probit, probabilidades estimadas y efectos estimados

El modelo probit poblacional con varios regresores es

11.2

$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k), \quad (11.6)$$

donde la variable dependiente Y es binaria, Φ es la función de distribución normal estándar acumulada, y X_1 y X_2 , etc., son regresores. El modelo se interpreta mejor calculando las probabilidades esperadas y el efecto de un cambio en un regresor.

La probabilidad esperada de que $Y = 1$, dados los valores de X_1, X_2, \dots, X_k , se calcula mediante el cómputo del z -valor, $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$, y buscando luego este valor de z en la tabla de la distribución normal (Tabla 1 del Apéndice).

El coeficiente β_1 es el cambio en el z -valor derivado de un cambio unitario en X_1 , manteniendo constantes X_2, \dots, X_k .

El efecto sobre la probabilidad esperada de un cambio en un regresor se calcula (1) calculando la probabilidad esperada para el valor inicial de las variables explicativas, (2) calculando la probabilidad esperada para el nuevo o modificado valor de los regresores, y (3) tomando su diferencia.

Aplicación a los datos de las hipotecas. A modo de ejemplo, podemos ajustar un modelo probit a las 2.380 observaciones de nuestra base de datos sobre la denegación de hipotecas (*denegar*) y la ratio de pagos-ingresos (*ratio P/I*):

$$\Pr(\text{denegar} = 1 | \text{ratio P/I}) = \Phi(-2,19 + 2,97 \text{ratio P/I}). \quad (11.7)$$

(0,16)	(0,47)
--------	--------

Los coeficientes estimados de $-2,19$ y $2,97$ son difíciles de interpretar debido a que afectan a la probabilidad de denegación a través del z -valor. De hecho, lo único que puede concluirse fácilmente a partir de la regresión probit estimada en la Ecuación (11.7) es que la proporción de pagos sobre ingresos está relacionada positivamente con la probabilidad de denegación (el coeficiente de la variable *ratio P/I* es positivo) y que esta relación es estadísticamente significativa ($t = 2,97/0,47 = 6,32$).

¿Cuál es el cambio en la probabilidad estimada de que una solicitud se deniegue, cuando la proporción de pagos-ingresos aumenta de $0,3$ a $0,4$? Para responder a esta pregunta, se sigue el procedimiento del Concepto clave 8.1: se calcula la probabilidad de denegación para una *ratio P/I* = $0,3$, y a continuación para un valor de la variable *ratio P/I* = $0,4$, posteriormente se calcula la diferencia. La probabilidad de denegación cuando la variable *ratio P/I* = $0,3$ es $\Phi(-2,19 + 2,97 \times 0,3) = \Phi(-1,30) = 0,097$. La probabilidad de denegación cuando la variable *ratio P/I* = $0,4$ es $\Phi(-2,19 + 2,97 \times 0,4) = \Phi(-1,00) = 0,159$. El cambio estimado en la probabilidad de denegación es $0,159 - 0,097 = 0,062$. Es decir, un aumento en la ratio de pagos-ingresos de $0,3$ a $0,4$ está asociado a un aumento en la probabilidad de denegación de $6,2$ puntos porcentuales, del $9,7\%$ al $15,9\%$.

Debido a que la función de regresión probit es no lineal, el efecto de un cambio en X depende del valor inicial de X . Por ejemplo, si la variable *ratio P/I* = $0,5$, la probabilidad de denegación estimada en base a la Ecuación (11.7) es $\Phi(-2,19 + 2,97 \times 0,5) = \Phi(-0,71) = 0,239$. Por tanto, el cambio en la probabilidad estimada cuando la variable *ratio P/I* aumenta de $0,4$ a $0,5$ es $0,239 - 0,159$, o $8,0$ puntos porcentuales, mayor que el incremento de $6,2$ puntos porcentuales si la variable *ratio P/I* aumenta de $0,3$ a $0,4$.

¿Cuál es el efecto de la etnia en la probabilidad de denegación de hipotecas, manteniendo constante la proporción entre pagos e ingresos? Para estimar este efecto, se estima una regresión probit tanto con la variable *ratio P/I* como con la variable *negro* como regresores:

$$\overline{\Pr(\text{denegar} = 1 | \text{ratio P/I}, \text{negro})} = \Phi(-2,26 + 2,74 \text{ ratio P/I} + 0,71\text{negro}). \quad (11.8)$$

(0,16)	(0,44)	(0,083)
--------	--------	---------

Una vez más, los valores de los coeficientes son difíciles de interpretar, pero el signo y la significación estadística no lo son. El coeficiente de la variable *negro* es positivo, lo que indica que un solicitante afroamericano tiene una mayor probabilidad de rechazo que un solicitante blanco, manteniendo constante su proporción de pagos-ingresos. Este coeficiente es estadísticamente significativo al nivel del 1% (el estadístico t del coeficiente que multiplica a la variable *negro* es $8,55$). Para un solicitante blanco con un valor de la variable *ratio P/I* = $0,3$ la probabilidad de denegación prevista es del $7,5\%$, mientras que para un candidato negro con un valor de la variable *ratio P/I* = $0,3$, es de un $23,3\%$; la diferencia de las probabilidades de denegación entre estos dos candidatos hipotéticos es de $15,8$ puntos porcentuales.

Estimación de los coeficientes probit. Los coeficientes probit presentados aquí se calculan mediante el método de máxima verosimilitud, que da lugar a estimadores eficientes (con varianza mínima) en una amplia variedad de aplicaciones, incluyendo la regresión con una variable dependiente binaria. El estimador de máxima verosimilitud es consistente y se distribuye normalmente en muestras grandes, por lo que los estadísticos t y los intervalos de confianza se pueden construir de la forma habitual.

El *software* de regresión para estimar modelos probit utiliza habitualmente la estimación máximo verosímil, por lo que este es un método que resulta fácil de aplicar en la práctica. Los errores estándar que calcula dicho *software* pueden utilizarse de la misma manera que los errores estándar de los coeficientes de regresión, por ejemplo; un intervalo de confianza al 95% para el verdadero coeficiente probit se puede construir como el coeficiente estimado $\pm 1,96$ errores estándar. Del mismo modo, los estadísticos F calculados utilizando los estimadores de máxima verosimilitud se pueden utilizar para contrastar hipótesis conjuntas. La estimación de máxima verosimilitud se analiza con más profundidad en la Sección 11.3, con detalles adicionales en el Apéndice 11.2.

**CONCEPTO
CLAVE**
11.3

Regresión logit

El modelo de regresión logit poblacional de la variable dependiente binaria Y con varios regresores es

$$\begin{aligned} \Pr(Y = 1 | X_1, X_2, \dots, X_k) &= F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) \\ &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}. \end{aligned} \quad (11.9)$$

La regresión logit es similar a la regresión probit excepto porque la función de distribución acumulada es diferente.

Regresión logit

El modelo de regresión logit. El modelo de regresión logit es similar al modelo de regresión probit, excepto que la función de distribución acumulada normal estándar Φ de la Ecuación (11.6) se sustituye por la función de distribución acumulada logística estándar, que se expresa mediante F . La regresión logit se resume en el Concepto clave 11.3. La función de distribución logística acumulada presenta una forma funcional específica, que se define en términos de la función exponencial, y que se recoge en la última expresión de la Ecuación (11.9).

Al igual que en el modelo probit, los coeficientes logit se interpretan mejor mediante el cálculo de las probabilidades estimadas y las diferencias en las probabilidades estimadas.

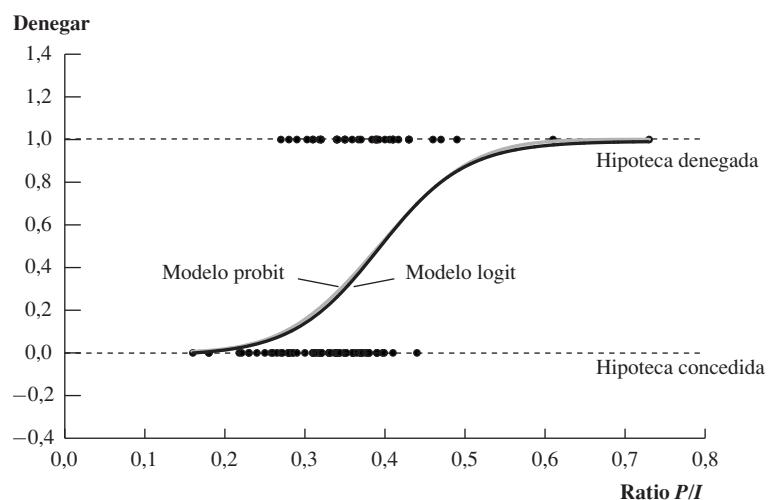
Los coeficientes del modelo logit se pueden estimar por máxima verosimilitud. El estimador de máxima verosimilitud es consistente y está distribuido normalmente en muestras grandes, por lo que los estadísticos t y los intervalos de confianza de los coeficientes pueden construirse de la forma habitual.

Las funciones de regresión logit y probit son similares. Esto se ilustra en la Figura 11.3, que representa gráficamente las funciones de regresión probit y logit para la variable dependiente *denegar* y el regresor único *ratio P/I*, estimadas por máxima verosimilitud utilizando las mismas 127 observaciones que en las Figuras 11.1 y 11.2. Las diferencias entre las dos funciones son pequeñas.

Históricamente, la principal motivación para la regresión logística era que la función de distribución logística se podía calcular más rápido que la función de distribución normal acumulada. Desde la aparición de equipos informáticos más eficientes, esta distinción ya no es importante.

FIGURA 11.3 Modelos probit y logit para la probabilidad de denegación, dada la variable *ratio P/I*

Estos modelos logit y probit dan lugar a estimaciones prácticamente idénticas de la probabilidad de que una solicitud de préstamo hipotecario sea denegada, dada la proporción pagos-ingresos.



Aplicación a los datos HMDA de Boston. Una regresión logit de la variable *denegar* frente a las variables *ratio P/I* y *negro*, utilizando las 2.380 observaciones de la base de datos, ofrece los siguientes resultados para la función de regresión estimada

$$\Pr(\text{denegar} = 1 | \text{ratio P/I}, \text{negro}) = F(-4,13 + 5,37\text{ratio P/I} + 1,27\text{negro}). \quad (11.10)$$

(0,35)	(0,96)	(0,15)
--------	--------	--------

El coeficiente de la variable *negro* es positivo y estadísticamente significativo al nivel del 1 % (el estadístico *t* es 8,47). La probabilidad estimada de denegación para un solicitante blanco con un valor de la variable *ratio P/I* = 0,3 es $1/[1 + e^{(-4,13 + 5,37 \times 0,3 + 1,27 \times 0)}] = 1/[1 + e^{2,52}] = 0,074$, o el 7,4 %. La probabilidad de denegación estimada para un solicitante afroamericano con valor de la variable *ratio P/I* = 0,3 es $[1/e^{1,25}] = 0,222$, o el 22,2 % por lo que la diferencia entre las dos probabilidades es de 14,8 puntos porcentuales.

Comparativa de los modelos de probabilidad lineal, probit y logit

Los tres modelos de probabilidad lineal, probit y logit, son solamente aproximaciones para la función de regresión poblacional desconocida $E(Y/X) = \Pr(Y = 1|X)$. El modelo de probabilidad lineal es más fácil de usar y de interpretar, pero no puede captar la naturaleza no lineal de la verdadera función de regresión poblacional. Los modelos de regresión probit y logit son no lineales en las probabilidades, pero sus coeficientes de regresión son más difíciles de interpretar. Por lo que ¿cuál debe utilizarse en la práctica?

No hay una respuesta correcta, y los distintos investigadores utilizan diferentes modelos. Las regresiones probit y logit con frecuencia producen resultados similares. Por ejemplo, de acuerdo con el modelo probit estimado en la Ecuación (11.8), la diferencia en las probabilidades de denegación entre un solicitante negro y un solicitante blanco con un valor de la variable *ratio P/I* = 0,3 se estimaba en 15,8 puntos porcentuales, mientras que la estimación logit de esta diferencia, en base a la Ecuación (11.10), era de 14,9 puntos porcentuales. A efectos prácticos, las dos estimaciones son muy similares. Una forma de elegir entre logit y probit es escoger el método que resulte más fácil de utilizar con el *software* estadístico concreto.

El modelo de probabilidad lineal proporciona la aproximación menos sensible a la función de regresión poblacional no lineal. Aun así, para algunos conjuntos de datos en los que puede haber unos pocos valores extremos de los regresores, el modelo de probabilidad lineal puede proporcionar una aproximación adecuada. En la regresión de la probabilidad de denegación de la Ecuación (11.3), la brecha estimada negro/blanco a partir del modelo de probabilidad lineal es de 17,7 puntos porcentuales, mayor que las estimaciones probit y logit, pero todavía cuantitativamente similar. No obstante, la única manera de saberlo, es estimando tanto un modelo lineal como un modelo no lineal y comparando sus probabilidades estimadas.

11.3 Estimación e inferencia en los modelos logit y probit²

Los modelos no lineales estudiados en las Secciones 8.2 y 8.3 son funciones no lineales de las variables independientes, pero son funciones lineales de los coeficientes desconocidos («parámetros»). En consecuencia, los coeficientes desconocidos de las funciones de regresión no lineales se pueden estimar por MCO. Por contra, las funciones de regresión probit y logit son funciones no lineales de los coeficientes. Es decir, los coeficientes probit $\beta_0, \beta_1, \dots, \beta_k$ de la Ecuación (11.6) aparecen *dentro* de la función de distribución acumulada normal, Φ , y los coeficientes logit de la Ecuación (11.9) aparecen dentro de la función de distribución logística estándar acumulada, F . Debido a que la función de regresión poblacional es una función no lineal de estos coeficientes $\beta_0, \beta_1, \dots, \beta_k$, los coeficientes no se pueden estimar por MCO.

En esta sección se ofrece una introducción al método estándar para la estimación de los coeficientes probit y logit, la máxima verosimilitud, los detalles matemáticos adicionales se recogen en el Apéndice 11.2. Debido a que está incluido en el *software* estadístico moderno, la estimación máximo verosímil de los

² Este apartado contiene material más avanzado que se puede omitir sin pérdida de continuidad.

coeficientes probit resulta fácil en la práctica. La teoría de la estimación de máxima verosimilitud, sin embargo, es más complicada que la teoría de mínimos cuadrados. Por lo tanto, se trata primero otro método de estimación, el de mínimos cuadrados no lineales, antes de pasar a la máxima verosimilitud.

Estimación por mínimos cuadrados no lineales

Mínimos cuadrados no lineales es un método general para estimar los parámetros desconocidos de una función de regresión en la que, como los coeficientes probit, los parámetros entran en la función de regresión poblacional de forma no lineal. El estimador de mínimos cuadrados no lineales, que se introdujo en el Apéndice 8.1, extiende el estimador MCO a las funciones de regresión que son funciones no lineales de los parámetros. Al igual que MCO, los mínimos cuadrados no lineales seleccionan los valores de los parámetros que minimizan la suma de los errores de predicción al cuadrado producidos por el modelo.

Para concretar, consideremos el estimador de mínimos cuadrados no lineales de los parámetros del modelo probit. La esperanza condicional de Y dadas las X es $E(Y|X_1, \dots, X_k) = \Pr(Y = 1|X_1, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$. La estimación por mínimos cuadrados no lineales ajusta esta función de esperanza condicional, que es una función no lineal de los parámetros, para la variable dependiente. Es decir, el estimador de mínimos cuadrados no lineales de los coeficientes probit son los valores de b_0, \dots, b_k que minimizan la suma de los errores de predicción al cuadrado:

$$\sum_{i=1}^n [Y_i - \Phi(b_0 + b_1 X_{1i} + \dots + b_k X_{ki})]^2. \quad (11.11)$$

El estimador de mínimos cuadrados no lineales comparte dos propiedades clave con el estimador MCO de la regresión lineal: es consistente (la probabilidad de que se encuentra cerca del verdadero valor se aproxima a 1 cuando el tamaño de la muestra se hace grande), y está normalmente distribuido en muestras grandes. Existen, sin embargo, estimadores que tienen una varianza menor que el estimador de mínimos cuadrados no lineales; es decir, el estimador de mínimos cuadrados no lineales es ineficiente. Por esta razón, el estimador de mínimos cuadrados no lineales de los coeficientes probit raramente se utiliza en la práctica, y en su lugar los parámetros se estiman por máxima verosimilitud.

Estimación máximo verosímil

La **función de verosimilitud** es la distribución de probabilidad conjunta de los datos, considerada como una función de los coeficientes desconocidos. El **estimador de máxima verosimilitud (EMV)** de los coeficientes desconocidos está compuesto por los valores de los coeficientes que maximizan la función de verosimilitud. Debido a que el EMV escoge los coeficientes desconocidos que maximizan la función de verosimilitud, que es a su vez la distribución de probabilidad conjunta, de hecho el EMV elige los valores de los parámetros que maximizan la probabilidad de haber obtenido los datos que realmente se observan. En este sentido, el EMV son los valores de los parámetros que «más probablemente» hayan generado los datos.

Para ilustrar la estimación de máxima verosimilitud, consideremos dos observaciones i.i.d., Y_1 e Y_2 , de una variable dependiente binaria sin regresores. Como Y es una variable aleatoria de Bernoulli, y el único parámetro desconocido a estimar es la probabilidad p de que $Y = 1$, que es asimismo la media de Y .

Para obtener el estimador de máxima verosimilitud, es necesaria una expresión para la función de verosimilitud, lo que a su vez requiere una expresión para la distribución de probabilidad conjunta de los datos. La distribución de probabilidad conjunta de las dos observaciones Y_1 e Y_2 es $\Pr(Y_1 = y_1, Y_2 = y_2)$. Debido a que Y_1 e Y_2 están distribuidas independientemente, la distribución conjunta es el producto de las distribuciones individuales [Ecuación (2.23)], por lo que $\Pr(Y_1 = y_1, Y_2 = y_2) = \Pr(Y_1 = y_1) \Pr(Y_2 = y_2)$. La distribución de Bernoulli se puede resumir mediante la fórmula $\Pr(Y = y) = p^y(1 - p)^{1-y}$: cuando $y = 1$, $\Pr(Y = 1) = p^1(1 - p)^0 = p$, y cuando $y = 0$, $\Pr(Y = 0) = p^0(1 - p)^1 = 1 - p$. Por tanto, la distribución de probabilidad conjunta de Y_1 e Y_2 es $\Pr(Y_1 = y_1, Y_2 = y_2) = [p^{y_1}(1 - p)^{1-y_1}] \times [p^{y_2}(1 - p)^{1-y_2}] = p^{(y_1+y_2)}(1 - p)^{2-(y_1+y_2)}$.

La función de verosimilitud es la distribución de probabilidad conjunta, considerada en forma de función de los coeficientes desconocidos. Para $n = 2$ observaciones i.i.d. de la variable aleatoria de Bernoulli, la función de verosimilitud es

$$f(p; Y_1, Y_2) = p^{(Y_1+Y_2)}(1-p)^{2-(Y_1+Y_2)} \quad (11.12)$$

El estimador de máxima verosimilitud de p es el valor de p que maximiza la función de verosimilitud de la Ecuación (11.12). Como en todos los problemas de maximización o minimización, esto se puede llevar a cabo mediante un procedimiento de prueba y error; es decir, se pueden probar diferentes valores de p y calcular el valor de la verosimilitud de $f(p; Y_1, Y_2)$ hasta quedar satisfecho de que esta función se maximiza. No obstante, en este ejemplo, la maximización de la función de verosimilitud mediante cálculo da lugar a una fórmula sencilla para el EMV: el EMV es $\hat{p} = \frac{1}{2}(Y_1 + Y_2)$. En otras palabras, el EMV de p es simplemente ¡la media de la muestra! De hecho, para una n general, el EMV \hat{p} de la probabilidad de Bernoulli p es la media muestral; es decir $\hat{p} = \bar{Y}$ (lo que se demuestra en el Apéndice 11.2). En este ejemplo, el EMV es el estimador de p habitual, la proporción de veces que $Y_i = 1$ en la muestra.

Este ejemplo es similar al problema de la estimación de los coeficientes desconocidos de los modelos de regresión probit y logit. En estos modelos, la probabilidad de éxito p no es constante, sino que depende de X ; es decir, es la probabilidad de éxito condicionada a X , lo cual aparece en la Ecuación (11.6) para el modelo probit y en la Ecuación (11.9) para el modelo logit. Por lo tanto, las funciones de verosimilitud probit y logit son similares a la función de verosimilitud de la Ecuación (11.12), excepto en que la probabilidad de éxito varía entre una observación y otra (debido a que dependen de X_i). Las expresiones de las funciones de verosimilitud para las funciones probit y logit se recogen en el Apéndice 11.2.

Al igual que el estimador de mínimos cuadrados no lineales, el EMV es consistente y está normalmente distribuido en muestras grandes. Debido a que habitualmente el *software* de regresión calcula el EMV de los coeficientes probit, este estimador resulta fácil de utilizar en la práctica. Todos los coeficientes probit y logit estimados presentados en este capítulo son EMV.

Inferencia estadística basada en el EMV. Debido a que el EMV se distribuye normalmente en muestras grandes, la inferencia estadística sobre los coeficientes probit y logit basada en el EMV se lleva a cabo de la misma manera que la inferencia acerca de los coeficientes de la función de regresión lineal basada en el estimador MCO. Es decir, los contrastes de hipótesis se realizan mediante el estadístico t y los intervalos de confianza al 95 % se construyen como $\pm 1,96$ errores estándar. Los contrastes de hipótesis conjuntas sobre varios coeficientes utilizan el estadístico F de forma similar a la expuesta en el Capítulo 7 para el modelo de regresión lineal. Todo esto es totalmente análogo a la inferencia estadística en el modelo de regresión lineal.

Un aspecto práctico importante es que algunos paquetes de *software* estadístico realizan los contrastes de hipótesis conjuntas con el estadístico F , mientras que otros paquetes de *software* utilizan el estadístico chi-cuadrado. El estadístico chi-cuadrado es $q \times F$, donde q es el número de restricciones que están siendo contrastadas. Debido a que el estadístico F , bajo la hipótesis nula, está distribuido como χ^2_q/q para muestras grandes, $q \times F$ se distribuye χ^2_q en muestras grandes. Debido a que los dos métodos se diferencian solamente en si dividen por q o no, dan lugar a inferencias idénticas, pero es necesario saber cuál es el método aplicado por el *software* para que sean utilizados los valores críticos correctos.

Medidas de ajuste

En la Sección 11.1, se mencionó que el R^2 es una medida de ajuste deficiente para el modelo de probabilidad lineal. Esto es válido igualmente para las regresiones probit y logit. Existen dos medidas de ajuste para los modelos con variable dependiente binaria, la «proporción correctamente estimada» y el «pseudo- R^2 ». La **proporción correctamente estimada** utiliza la regla siguiente: si $Y_i = 1$ y la probabilidad estimada supera el 50 % o si $Y_i = 0$ y la probabilidad estimada es inferior al 50 %, entonces se dice que Y_i está correctamente estimada. De lo contrario, se dice que Y_i está incorrectamente estimada. La «proporción correctamente estimada» es la proporción de las n observaciones Y_1, \dots, Y_n que está correctamente estimada.

Una ventaja de esta medida de ajuste es que resulta fácil de comprender. Una desventaja es que no refleja la calidad de la predicción: si $Y_i = 1$, la observación se considera como correctamente estimada si la probabilidad estimada es del 51 % o del 90 %.

El **pseudo- R^2** mide el ajuste del modelo mediante la función de verosimilitud. Debido a que el EMV maximiza la función de verosimilitud, la adición de otro regresor a un probit o logit aumenta el valor de la verosimilitud maximizada, al igual que la adición de un regresor necesariamente reduce la suma de los cuadrados de los residuos en la regresión lineal por MCO. Este hecho sugiere medir la calidad de ajuste de un modelo probit mediante la comparación del valor de la función de verosimilitud maximizada con todas las variables explicativas con el valor de la función de verosimilitud sin regresores. Es decir, de hecho, lo que hace el pseudo- R^2 . Se proporciona una fórmula para el pseudo- R^2 en el Apéndice 11.2.

11.4 Aplicación a los datos HMDA de Boston

Las regresiones de las dos secciones anteriores indican que las tasas de denegación fueron superiores para los solicitantes negros que para los solicitantes blancos, manteniendo constante su proporción pagos-ingresos. No obstante, los empleados de banca, al decidir sobre una solicitud de hipoteca, sopesan de forma legítima muchos factores, y, si alguno de esos otros factores difiere de forma sistemática en función de la etnia, los estimadores considerados hasta el momento presentarán sesgo de variable omitida.

En esta sección, echaremos un vistazo más cercano a la cuestión de si existe evidencia estadística de discriminación en los datos HMDA de Boston. En concreto, nuestro objetivo es estimar el efecto de la etnia en la probabilidad de denegación, manteniendo constantes aquellas características de los candidatos que un empleado de banca podría considerar tener en cuenta de forma legal al decidir acerca de una solicitud de hipoteca.

Las variables más importantes que se encuentran a disposición de los empleados de las entidades de crédito por medio de las solicitudes de hipoteca en la base de datos HMDA de Boston se muestran en la Tabla 11.1, esas son las variables en las que se centrarán los modelos empíricos sobre las decisiones de concesión de préstamo. Las dos primeras variables son medidas directas de la carga financiera que supondría el préstamo para el solicitante, medidas en términos de su renta. La primera de ellas es la *ratio P/I*; la segunda es la proporción entre los gastos relacionados con la vivienda y la renta. La siguiente variable es el tamaño del préstamo, en relación con el valor de tasación de la vivienda, si la relación préstamo-valor es cercana a 1, el banco podría tener problemas para recuperar el monto total del préstamo si el solicitante no pagara el préstamo y el banco ejecutase la hipoteca. Las tres últimas variables financieras resumen el historial crediticio del solicitante. Si un solicitante ha sido poco fiable en el pasado respecto al pago de deudas, el empleado de la entidad de crédito legítimamente podría preocuparse por la capacidad del solicitante o su deseo de cumplir con los pagos de la hipoteca en el futuro. Las tres variables miden los diferentes tipos de históricos crediticios, que el empleado de la entidad de crédito podría sopesar de forma diferente. La primera se refiere al crédito al consumo, por ejemplo la deuda de tarjetas de crédito; la segunda es el historial previo de pagos hipotecarios; y la tercera mide los problemas crediticios lo suficientemente graves como para aparecer en un registro legal público, tales como la declaración de quiebra.

La Tabla 11.1 enumera asimismo otras variables relevantes para la toma de decisión del empleado de la entidad de crédito. A veces, el solicitante debe solicitar un seguro hipotecario privado³. El empleado de la entidad de crédito sabe si esa solicitud fue denegada, y esta denegación ponderaría negativamente en la toma de decisión por parte del empleado. Las tres variables siguientes, que se refieren a la situación laboral, el estado civil y el nivel educativo del solicitante, están relacionadas con la capacidad potencial del solicitante para devolver el crédito. En el caso de una ejecución hipotecaria, las características de la propiedad resultan asimismo relevantes, y la variable siguiente indica si la propiedad es un condominio. Las dos últi-

³ El seguro hipotecario es una póliza de seguro bajo la cual la compañía de seguros efectúa el pago mensual al banco si el prestatario no cumple. Durante el periodo de este estudio, si la proporción préstamo-valor superaba el 80 %, en general, el solicitante estaba obligado a contratar un seguro hipotecario.

TABLA 11.1 Variables incluidas en los modelos de regresión de las decisiones sobre concesión de hipotecas

Variable	Definición	Media muestral
Variables financieras		
ratio <i>P/I</i>	Proporción total de pagos mensuales de la deuda sobre el total de ingresos mensuales	0,331
ratio gastos vivienda-ingresos	Proporción de los gastos domésticos mensuales sobre el total de ingresos mensuales	0,225
ratio préstamo-valor	Proporción del tamaño del préstamo sobre el valor tasado de la propiedad	0,738
calificación crédito consumo	1 si no hay pagos atrasados o morosidad 2 si uno o dos pagos atrasados o morosidad 3 si más de dos pagos atrasados 4 si no hay suficiente historial para determinar 5 si la historia de crédito en mora con los pagos atrasados 60 días 6 si la historia de crédito en mora con los pagos atrasados 90 días	2,1
calificación crédito hipotecario	1 si no pagos atrasados hipoteca 2 si no historial pagos hipoteca 3 si uno o dos pagos atrasados hipoteca 4 si más de dos pagos atrasados hipoteca	1,7
registro público insolvencia	1 si está incluido en cualquier registro público de problemas crediticios (quiebra, anulación de cargo, acciones de cobro) 0 en otro caso	0,074
Características adicionales del solicitante		
seguro hipotecario denegado	1 si el solicitante solicitó seguro hipotecario y le fue denegado 0 en otro caso	0,020
autoempleado	1 si es autoempleado, 0 en otro caso	0,116
soltero	1 si el solicitante declaró ser soltero, 0 en otro caso	3,393
titulado secundaria	1 si el solicitante posee título de graduado en una escuela secundaria, 0 en otro caso	0,984
tasa de desempleo	Tasa de desempleo en Massachusetts en 1989 en el sector del solicitante	3,8
condominio	1 si la unidad es un condominio, 0 en otro caso	0,288
negro	1 si el solicitante es negro, 0 si es blanco	0,142
denegación	1 si la solicitud de hipoteca se deniega, 0 en otro caso	0,120

mas variables de la Tabla 11.1 se refieren a si el solicitante es blanco o negro y si la solicitud fue denegada o aceptada. En estos datos, el 14,2 % de los solicitantes son negros y el 12,0 % de las solicitudes fueron denegadas.

En la Tabla 11.2 se presentan los resultados de la regresión basada en estas variables. Las especificaciones base, en las columnas (1) a (3), incluyen las variables financieras de la Tabla 11.1 además de las variables que indican si el seguro hipotecario privado fue denegado y si el solicitante trabajaba por cuenta propia. En la década de 1990 los empleados de entidades de crédito habitualmente utilizaban umbrales, o valores de corte, para la proporción préstamo-valor, por lo que la especificación base que incluye esta variable utiliza variables binarias para indicar si el valor de la relación préstamo-valor es alto ($\geq 0,95$), medio (entre 0,8 y 0,95), o bajo ($< 0,8$); este caso se omite para evitar la multicolinealidad perfecta). Los regresores de las tres primeras columnas son similares a los de la especificación base considerada por los investigadores del Ban-

TABLA 11.2 Regresión de denegaciones de hipoteca utilizando los datos HMDA de Boston

Variable dependiente: denegar = 1 si la solicitud de hipoteca es denegada = 0 si es aceptada; 2.380 observaciones.

Regresor del modelo de regresión	MPL (1)	Logit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
negro	0,084** (0,023)	0,688** (0,182)	0,389** (0,098)	0,371** (0,099)	0,363** (0,100)	0,246 (0,448)
ratio P/I	0,449** (0,114)	4,76** (1,33)	2,44** (0,61)	2,46** (0,60)	2,62** (0,61)	2,57** (0,66)
ratio gastos de la vivienda-ingresos	-0,048 (0,110)	-0,11 (1,29)	-0,18 (0,68)	-0,30 (0,68)	-0,50 (0,70)	-0,54 (0,74)
ratio préstamo-valor medio (0.80 ≤ ratio préstamo-valor ≤ 0.95)	0,031* (0,013)	0,46** (0,16)	0,21** (0,08)	0,22** (0,08)	0,22** (0,08)	0,22** (0,08)
ratio préstamo-valor alto (ratio préstamo-valor > 0.95)	0,189** (0,050)	1,49** (0,32)	0,79** (0,18)	0,79** (0,18)	0,84** (0,18)	0,79** (0,18)
calificación crédito consumo	0,031** (0,005)	0,29** (0,04)	0,15** (0,02)	0,16** (0,02)	0,34** (0,11)	0,16** (0,02)
calificación crédito hipotecario	0,021 (0,011)	0,28* (0,14)	0,15* (0,07)	0,11 (0,08)	0,16 (0,10)	0,11 (0,08)
registro público insolvencia	0,197** (0,035)	1,23** (0,20)	0,70** (0,12)	0,70** (0,12)	0,72** (0,12)	0,70** (0,12)
seguro hipotecario denegado	0,702** (0,045)	4,55** (0,57)	2,56** (0,30)	2,59** (0,29)	2,59** (0,30)	2,59** (0,29)
autoempleado	0,060** (0,021)	0,67** (0,21)	0,36** (0,11)	0,35** (0,11)	0,34** (0,11)	0,35** (0,11)
soltero				0,23** (0,08)	0,23** (0,08)	0,23** (0,08)
título secundaria				-0,61** (0,23)	-0,60* (0,24)	-0,62** (0,23)
tasa de desempleo				0,03 (0,02)	0,03 (0,02)	0,03 (0,02)
condominio					-0,05 (0,09)	
negro × ratio P/I						-0,58 (1,47)
negro × ratio gastos de la vivienda-ingresos						1,23 (1,69)
Variables indicadores adicionales de calificación crediticia	no	no	no	no	sí	no
constante	-0,183** (0,028)	-5,71** (0,48)	-3,04** (0,23)	-2,57** (0,34)	-2,90** (0,39)	-2,54** (0,35)
Solicitante soltero; título secundaria; tasa de desempleo del sector				5,85 (< 0,001)	5,22 (0,001)	5,79 (< 0,001)
Variables indicadores adicionales de calificación crediticia					1,22 (0,291)	
Interacciones raciales y negro						4,96 (0,002)

(continúa)

(Continuación Tabla 11.2)

Estadísticos *F* y *p*-valores del contraste de exclusión de grupos de variables

Regresor del modelo de regresión	MPL (1)	Logit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
Interacciones raciales y negro						4,96 (0,002)
Solamente interacciones raciales						0,27 (0,766)
Diferencia en probabilidades de denegación estimadas, blanco vs. negro (puntos porcentuales)	8,4 %	6,0 %	7,1 %	6,6 %	6,3 %	6,5 %
Estas regresiones se estimaron utilizando las $n = 2.380$ observaciones de la base de datos HMDA de Boston descrita en el Apéndice 11.1. El modelo de probabilidad lineal se estimó por MCO y las regresiones probit y logit fueron estimadas por máxima verosimilitud. Los errores estándar se ofrecen entre paréntesis bajo los coeficientes y los <i>p</i> -valores se indican entre paréntesis bajo los estadísticos <i>F</i> . La variación en la probabilidad estimada de la última fila se calculó para un solicitante hipotético cuyos valores de los regresores, distintos de la etnia, son iguales a la media muestral. Los coeficientes individuales son estadísticamente significativos al nivel del 5%* o del 1%**.						

co de la Reserva Federal de Boston en el análisis original de estos datos⁴. Las regresiones de las columnas (1) a (3) solo se diferencian en la forma de modelizar la probabilidad de denegación, utilizando un modelo de probabilidad lineal, un modelo logit, y un modelo probit, respectivamente.

Debido a que la regresión de la columna (1) es un modelo de probabilidad lineal, sus coeficientes son las variaciones en las probabilidades estimadas que aparecen debido a una variación unitaria en la variable independiente. De esta manera, un aumento en la variable *ratio P/I* de 0,1 se estima que aumente la probabilidad de denegación en 4,5 puntos porcentuales (el coeficiente de la variable *ratio P/I* en la columna (1) es 0,449, y $0,449 \times 0,1 \cong 0,045$). Del mismo modo, presentar un valor alto para la proporción entre el préstamo y el valor de la vivienda aumenta la probabilidad de denegación: una relación préstamo-valor superior al 95 % se asocia con un aumento porcentual de 18,9 puntos (el coeficiente es de 0,189) en la probabilidad de denegación, en relación al caso omitido de una proporción préstamo-valor inferior al 80 %, manteniendo constantes las otras variables de la columna (1). Los solicitantes que obtienen una mala calificación crediticia asimismo tienen más dificultades para obtener un préstamo, manteniendo todo lo demás constante, aunque curiosamente el coeficiente del crédito al consumo es estadísticamente significativo, pero no el coeficiente del crédito hipotecario. Los solicitantes que aparecen en un registro público de problemas crediticios, tales como la declaración de quiebra, tienen mucha mayor dificultad para obtener un préstamo: manteniendo todo lo demás igual, aparecer en un registro público de insolvencia se estima que aumenta la probabilidad de denegación en 0,197, o 19,7 puntos porcentuales. La denegación de seguro hipotecario privado se estima que es virtualmente decisiva: el coeficiente estimado de 0,702 significa que la denegación del seguro hipotecario incrementa la probabilidad de denegación de un préstamo hipotecario en 70,2 puntos porcentuales, manteniendo todo lo demás igual. De las nueve variables (distintas de la etnia) de la regresión, los coeficientes de todas excepto de dos son estadísticamente significativos al nivel del 5 %, lo cual es coherente con que los empleados de las instituciones de crédito tomen sus decisiones teniendo en cuenta muchos factores.

El coeficiente de la variable *negro* en la regresión (1) es 0,084, lo que indica que la diferencia en las probabilidades de denegación para los solicitantes negros y blancos es de 8,4 puntos porcentuales, manteniendo constantes las otras variables de la regresión. Esto es estadísticamente significativo al nivel de significación del 1 % ($t = 3,65$).

⁴ La diferencia entre las variables explicativas de las columnas (1) a (3) y las de Munnell *et al.* (1996), Tabla 2(1), es que Munnell *et al.* incluyen indicadores adicionales para la ubicación de la casa y la identidad de la entidad crediticia, datos que no están disponibles al público; un indicador para una vivienda multifamiliar, que aquí es irrelevante, debido a que nuestro subconjunto se centra en las viviendas unifamiliares; y la riqueza neta, que se omite porque esta variable tiene unos valores positivos y negativos muy grandes y, por lo tanto, se corre el riesgo de que los resultados sean sensibles a observaciones atípicas específicas.

De las estimaciones logit y probit presentadas en las columnas (2) y (3) se extraen conclusiones similares. En las regresiones logit y probit, ocho de los nueve coeficientes de otras variables distintas de la etnia son estadística y significativamente distintos de cero individualmente al 5 % de nivel, y el coeficiente de la variable *negro* es estadísticamente significativo al 1 %. Tal y como se reflejó en la Sección 11.2, debido a que estos modelos son no lineales, deben elegirse unos valores concretos para todas las variables explicativas a fin de calcular la diferencia en las probabilidades estimadas para los solicitantes negros y blancos. Un modo convencional de tomar esta decisión es considerar un solicitante «promedio» que presenta como valores para todas las otras variables distintas de la etnia el promedio muestral de todas ellas. La última fila de la Tabla 11.2 presenta esta diferencia estimada de las probabilidades, evaluada para este solicitante promedio. Las diferencias raciales estimadas son similares entre sí: 8,4 puntos porcentuales para el modelo de probabilidad lineal [columna (1)], 6,0 puntos porcentuales para el modelo logit [columna (2)], y 7,1 puntos porcentuales para el modelo probit [columna (3)]. Estos efectos raciales estimados y los coeficientes de la variable *negro* son menores que en las regresiones de las secciones anteriores, en las que los únicos regresores eran las variables *ratio P/I* y *negro*, lo que indica que las estimaciones anteriores presentaban sesgo de variable omitida.

Las regresiones de las columnas (4) a (6) investigan la sensibilidad de los resultados de la columna (3) a los cambios en la especificación de la regresión. La columna (4) modifica la columna (3) mediante la inclusión de características adicionales del solicitante. Estas características ayudan a predecir si el préstamo será denegado; por ejemplo, tener al menos un diploma de secundaria reduce la probabilidad de denegación (la estimación es negativa y el coeficiente es estadísticamente significativo al nivel del 1 %). Sin embargo, al tener en cuenta estas características personales no cambia el coeficiente estimado de la variable *negro* o la diferencia estimada en las probabilidades de denegación (6,6 %) de una manera importante.

La columna (5) separa las seis categorías de crédito al consumo y las cuatro categorías de crédito hipotecario para contrastar la hipótesis nula de que estas dos variables entran de forma lineal; esta regresión asimismo agrega una variable que indica si la propiedad es un condominio. La hipótesis nula de que las variables de calificación crediticia entran en la expresión para el *z*-valor linealmente no se rechaza, ni el indicador de condominio es significativo, al nivel del 5 %. Lo más importante, la diferencia racial estimada en las probabilidades de denegación (6,3 %) es esencialmente la misma que en las columnas (3) y (4).

La columna (6) examina si existen interacciones. ¿Se aplican diferentes criterios en la evaluación de las ratios pagos-ingresos y gastos de la vivienda-ingresos para los solicitantes negros frente a los solicitantes blancos? La respuesta parece ser no: los términos de interacción no son de manera conjunta estadísticamente significativos al nivel del 5 %. No obstante, la etnia sigue teniendo un efecto significativo, debido a que el indicador de la etnia y el término de interacción son conjunta y estadísticamente significativos al nivel del 1 %. Una vez más, la diferencia racial estimada en las probabilidades de denegación (6,5 %) es esencialmente la misma que en las otras regresiones probit.

En las seis especificaciones, el efecto de la etnia en la probabilidad de denegación, manteniendo constantes las otras características del solicitante, es estadísticamente significativo al nivel del 1 %. La diferencia estimada en las probabilidades de denegación entre los solicitantes negros y los solicitantes blancos oscila entre los 6,0 puntos porcentuales y los 8,4 puntos porcentuales.

Una forma de evaluar si esta diferencia es grande o pequeña es volver a una variante de la pregunta planteada al comienzo de este capítulo. Supongamos que dos personas solicitan una hipoteca, una blanca y otra negra, pero que por lo demás presentan los mismos valores para las otras variables independientes de la regresión (3); en concreto, dejando a un lado la etnia, los valores de las otras variables de la regresión (3) son los valores de las medias muestrales para la base de datos HMDA. El solicitante blanco se enfrenta a un 7,4 % de posibilidades de denegación, pero el solicitante negro se enfrenta a un 14,5 % de posibilidades de denegación. La diferencia racial estimada en las probabilidades de denegación, 7,1 puntos porcentuales, significa que el candidato negro tiene casi el doble de probabilidad de ver su solicitud denegada que el solicitante blanco.

Los resultados de la Tabla 11.2 (y los del estudio original de la Fed de Boston) proporcionan evidencia estadística acerca de la existencia de patrones raciales en la denegación de hipotecas lo que, por ley, no debería ocurrir. Esta evidencia desempeñó un papel importante como estímulo a los cambios de política por

parte de los reguladores bancarios⁵. Pero a los economistas les encanta un buen argumento, y no es sorprendente que estos resultados hayan generado un intenso debate.

Debido a que la sugerencia de que existe (o existía) discriminación racial en los préstamos, resulta acusatoria, se revisan brevemente algunos aspectos de este debate. Para ello, resulta útil adoptar el marco del Capítulo 9, es decir, es necesario tener en cuenta la validez interna y externa de los resultados de la Tabla 11.2, que son representativos de los análisis previos de los datos HMDA de Boston. Algunas de las críticas hechas al estudio original del Banco de la Reserva Federal de Boston se refieren a la validez interna: posibles errores en los datos, formas funcionales no lineales alternativas, interacciones adicionales, etc. Los datos originales fueron sometidos a una auditoría cuidadosa, se encontraron algunos errores, y los resultados presentados aquí (y en el último estudio publicado por la Fed de Boston) se basan en el conjunto de datos una vez «limpio». La estimación de otras especificaciones —diferentes formas funcionales y/o regresores adicionales— genera asimismo estimaciones de las diferencias raciales comparables a las de la Tabla 11.2. Un aspecto acerca de la validez interna que puede resultar potencialmente más difícil de tratar es si existe información no racial relevante obtenida durante las entrevistas personales para la concesión del préstamo, no registrada en la solicitud misma de préstamo, que esté correlacionada con la etnia; si es así, aún podría existir sesgo de variable omitida en las regresiones de la Tabla 11.2. Finalmente, hay quien ha cuestionado la validez externa: incluso si hubo discriminación racial en Boston en 1990, resulta erróneo hoy en día implicar en ella a los prestamistas de otros lugares. Por otra parte, la discriminación racial puede que sea menos probable al utilizar las solicitudes modernas *on line*, debido a que la hipoteca puede ser aprobada o rechazada sin una reunión cara a cara. La única forma de resolver la cuestión de la validez externa es considerar los datos de otros lugares y años⁶.

11.5 Conclusión

Cuando la variable dependiente Y es binaria, la función de regresión poblacional es la probabilidad de que $Y = 1$, condicionada a las variables explicativas. La estimación de esta función de regresión poblacional implica la búsqueda de una forma funcional que justifique su interpretación en términos de probabilidad, la estimación de los parámetros desconocidos de esa función, y la interpretación de los resultados. Los valores de predicción resultantes son probabilidades estimadas, y el efecto estimado de una variación en un regresor X es la variación estimada en la probabilidad de que $Y = 1$ debida a la variación en X .

Una forma natural de modelizar la probabilidad de que $Y = 1$, dados los regresores es mediante una función de distribución acumulada, en la que el argumento de la f.d.a. depende de los regresores. La regresión probit utiliza una f.d.a. normal como función de regresión, y la regresión logit utiliza una f.d.a. logística. Debido a que estos modelos son funciones no lineales de los parámetros desconocidos, estos parámetros son más difíciles de estimar que los coeficientes de regresión lineal. El método de estimación habitual es el de máxima verosimilitud. En la práctica, la inferencia estadística mediante las estimaciones de máxima verosimilitud se lleva a cabo de la misma forma que mediante regresión lineal múltiple; por ejemplo, los intervalos de confianza al 95 % para un coeficiente se construyen como el coeficiente estimado $\pm 1,96$ errores estándar.

A pesar de su no linealidad intrínseca, a veces la función de regresión poblacional puede aproximarse adecuadamente mediante un modelo de probabilidad lineal, es decir, mediante la línea recta obtenida por

⁵ Estos cambios de política incluyen cambios en la forma en que se realizan los exámenes sobre la equidad en los préstamos por parte de los reguladores bancarios federales, los cambios en las investigaciones realizadas por el Departamento de Justicia de EE.UU., y la mejora de los programas educativos para los bancos y otras instituciones emisoras de préstamos para la vivienda.

⁶ Si está interesado en leer más acerca de esta cuestión, un buen comienzo es el simposio sobre discriminación racial y economía en su edición de Primavera de la revista *Journal of Economic Perspectives* de 1998. El artículo para el simposio de Helen Ladd (1998) examina la evidencia y el debate sobre la discriminación racial en el préstamo hipotecario. Se presenta un tratamiento más detallado en el trabajo de Goering y Wienk (1996). El mercado hipotecario de EE.UU. ha cambiado drásticamente desde el estudio de la Fed de Boston, lo que incluye una flexibilización de las normas que rigen los préstamos, una burbuja en los precios de la vivienda, la crisis financiera de 2008-2009, y el regreso a normas más estrictas sobre la concesión de préstamos. Para una introducción a los cambios en los mercados hipotecarios, véase Green y Wachter (2007).

James Heckman y Daniel McFadden, ganadores del Premio Nobel

El Premio Nobel de Economía del año 2000 fue otorgado conjuntamente a dos económetras, James J. Heckman, de la Universidad de Chicago y Daniel L. McFadden, de la Universidad de California en Berkeley, por sus fundamentales contribuciones al análisis de los datos sobre las personas y las empresas. Gran parte de su trabajo estaba enfocado a resolver las dificultades que se presentan con las variables dependientes limitadas.

Heckman obtuvo el premio por el desarrollo de herramientas para el tratamiento de la selección muestral. Tal y como se trató en la Sección 9.2, el sesgo de selección muestral se produce cuando la disponibilidad de los datos está influida por un proceso de selección que está relacionado con el valor de la variable dependiente. Por ejemplo, supongamos que se quiere estimar la relación entre los ingresos y algún regresor, X , mediante una muestra aleatoria de la población. Si se estima la regresión con la submuestra de trabajadores empleados —es decir, aquellos que declaran ingresos positivos— la estimación MCO podría estar sujeta a un sesgo de selección. La solución de Heckman fue la de especificar una ecuación preliminar con una variable dependiente binaria que indica si el trabajador se encuentra dentro o fuera de la fuerza laboral (dentro o fuera de la submuestra) y posteriormente tratar esta ecuación y la ecuación de los ingresos como un sistema de ecuaciones simultáneas. Esta estrategia general se ha extendido a los problemas de selección que surgen en muchos campos, que van desde la economía laboral a la organización industrial y las finanzas.

McFadden fue galardonado con el premio por el desarrollo de modelos de análisis de datos de selección discreta (¿se unirá al ejército un graduado de secundaria, irá a la universidad, o conseguirá un trabajo?). Empezó por estudiar el problema de un individuo que maximiza la utilidad esperada de cada opción posible, lo cual podría depender de variables observables (como los salarios, las características del trabajo y los antecedentes familiares). A continuación, obtuvo modelos para las probabilidades de elección individual con coeficientes desconocidos, que a su vez podrían ser estimados por máxima verosimilitud. Estos modelos y sus extensiones se han demostrado ampliamente útiles en el análisis de los datos de elección discreta en muchos campos, incluyendo la economía laboral, la economía de la salud y la economía del transporte.

Para más información sobre estos y otros Premios Nobel de Economía, visite la Fundación Nobel. Página web, www.nobel.se/economics.



James J. Heckman



Daniel L. McFadden

regresión lineal múltiple. El modelo de probabilidad lineal, la regresión probit y la regresión logit proporcionan todos ellos respuestas «en definitiva» similares, cuando se aplican a los datos HMDA de Boston: los tres métodos estiman diferencias sustanciales en las tasas de denegación de préstamos hipotecarios para los solicitantes negros y para los solicitantes blancos que en lo demás son similares.

Las variables dependientes binarias son el ejemplo más común de variables dependientes limitadas, que son variables dependientes con rango limitado. El último cuarto del siglo XX fue testigo de importantes avances en los métodos económétricos para el análisis de otras variables dependientes limitadas (véase el recuadro de los premios Nobel). Se revisan algunos de estos métodos en el Apéndice 11.3.

Resumen

1. Cuando Y es una variable binaria, el modelo de regresión lineal múltiple se denomina modelo de probabilidad lineal. La recta de regresión poblacional muestra la probabilidad de que $Y = 1$, dados los valores de las variables explicativas, X_1, X_2, \dots, X_k .
2. Los modelos de regresión probit y logit son modelos de regresión no lineales que se utilizan cuando Y es una variable binaria. A diferencia del modelo de probabilidad lineal, las regresiones probit y logit aseguran que la probabilidad estimada de que $Y = 1$ se encuentre entre 0 y 1 para todos los valores de X .

3. La regresión probit utiliza la función de distribución acumulada normal estándar. La regresión logística utiliza la función de distribución acumulada logística. Los coeficientes logit y probit se estiman por máxima verosimilitud.
4. Los valores de los coeficientes de las regresiones probit y logit no resultan fáciles de interpretar. Las variaciones en la probabilidad de que $Y = 1$ asociadas a las variaciones en una o más de las X se pueden calcular utilizando el procedimiento general para los modelos no lineales descrito en el Concepto clave 8.1.
5. Los contrastes de hipótesis sobre los coeficientes de los modelos de probabilidad lineal, logit y probit se llevan a cabo mediante los estadísticos t y F habituales.

Términos clave

variable dependiente limitada (275)	función de verosimilitud (285)
modelo de probabilidad lineal (277)	estimador de máxima verosimilitud (EMV) (285)
probit (279)	proporción correctamente estimada (286)
logit (279)	pseudo- R^2 (287)
regresión logística (279)	

Revisión de conceptos

- 11.1** Supóngase que un modelo de probabilidad lineal obtiene un valor esperado de Y que es igual a 1,3. Explique por qué esto no tiene sentido.
- 11.2** En la Tabla 11.2 el coeficiente estimado de la variable *negro* es 0,084 en la columna (1), 0,688 en la columna (2), y 0,389 en la columna (3). A pesar de estas grandes diferencias, los tres modelos obtienen estimaciones similares del efecto marginal de la etnia sobre la probabilidad de denegación de la hipoteca. ¿Cómo es posible esto?
- 11.3** Una de sus amigas está utilizando datos sobre personas para estudiar los determinantes del consumo de tabaco en su universidad. Ella le pregunta si debe usar un modelo probit, logit, o un modelo de probabilidad lineal. ¿Qué le aconsejaría? ¿Por qué?
- 11.4** ¿Por qué los coeficientes de los modelos probit y logit se estiman por máxima verosimilitud en lugar de MCO?

Ejercicios

Los Ejercicios del 11.1 al 11.5 se basan en el siguiente escenario: cuatrocientos candidatos a obtener el permiso de conducir fueron seleccionados aleatoriamente y se les preguntó si aprobaron el examen de conducir ($Aprobar_i = 1$) o lo suspendieron ($Aprobar_i = 0$); asimismo se registraron los datos sobre su género ($Masculino_i = 1$ si es hombre, y $Masculino_i = 0$ si es mujer) y sus años de experiencia en conducción ($Experiencia_i$, en años). Las tablas siguientes resumen los modelos estimados.

- 11.1** Utilizando los resultados de la columna (1):
 - a) ¿La probabilidad de aprobar el examen depende de la experiencia? Explíquelo.
 - b) Matthew tiene 10 años de experiencia conduciendo. ¿Cuál es la probabilidad de que apruebe el examen?
 - c) Christopher es un conductor novel (cero años de experiencia). ¿Cuál es la probabilidad de que apruebe el examen?
 - d) La muestra incluye valores de la variable *Experiencia* entre 0 y 40 años, y solo cuatro personas de la muestra tienen más de 30 años de experiencia como conductor. Jed tiene 95 años y ha estado conduciendo desde que tenía 15 años. ¿Cuál es la predicción del modelo para la probabilidad de que Jed apruebe el examen? ¿Cree que esta predicción es fiable? ¿Por qué o por qué no?

Variable dependiente: Aprobar	Probit (1)	Logit (2)	Probabilidad lineal (3)	Probit (4)	Logit (5)	Probabilidad lineal (6)	Probit (7)
Experiencia	0,031 (0,009)	0,040 (0,016)	0,006 (0,002)				0,041 (0,156)
Masculino				-0,333 (0,161)	-0,622 (0,303)	-0,071 (0,34)	-1,174 (0,259)
Masculino × Experiencia							-0,015 (0,019)
Constante	0,712 (0,126)	1,059 (0,221)	0,774 (0,034)	1,282 (0,124)	2,197 (0,242)	0,900 (0,022)	0,806 (0,200)

- 11.2** a) Responda a los apartados (a) a (c) del Ejercicio 11.1 utilizando los resultados de la columna (2).
 b) Represente gráficamente las probabilidades estimadas por los modelos probit y logit en las columnas (1) y (2) para los valores de la variable *Experiencia* entre 0 y 60. ¿Son similares el probit y el logit?
- 11.3** a) Responda a los apartados (a) a (c) del Ejercicio 11.1 utilizando los resultados de la columna (3).
 b) Represente gráficamente las probabilidades estimadas por los modelos probit y de probabilidad lineal de las columnas (1) y (3) en función de los valores de la variable *Experiencia*, para sus valores entre 0 y 60. ¿Piensa que el modelo de probabilidad lineal resulta apropiado en este caso? ¿Por qué o por qué no?
- 11.4** Utilizando los resultados de las columnas (4) a (6):
 a) Calcule las probabilidades estimadas de aprobar el examen para hombres y mujeres.
 b) ¿Son diferentes los modelos (4) al (6)? ¿Por qué o por qué no?
- 11.5** Utilizando los resultados de la columna (7):
 a) Akira es un hombre con 10 años de experiencia conduciendo. ¿Cuál es la probabilidad de que apruebe el examen?
 b) Jane es una mujer con 2 años de experiencia como conductora. ¿Cuál es la probabilidad de que apruebe el examen?
 c) ¿Depende del género el efecto de la variable *Experiencia* sobre el rendimiento en el examen? Explíquelo.
- 11.6** Utilice el modelo probit estimado en la Ecuación (11.8) para responder a las siguientes preguntas:
 a) Un solicitante de crédito hipotecario de raza negra presenta un valor de la variable *ratio P/I* de 0,35. ¿Cuál es la probabilidad de que su solicitud sea rechazada?
 b) Supongamos que el solicitante reduce esta proporción a 0,30. ¿Qué efecto tendría en su probabilidad de serle denegado un préstamo hipotecario?
 c) Repita (a) y (b) para un solicitante blanco.
 d) ¿Depende el efecto marginal de la variable *ratio P/I* sobre la probabilidad de denegación de la hipoteca de la etnia? Explíquelo.
- 11.7** Repita el Ejercicio 11.6 utilizando el modelo logit de la Ecuación (11.10). ¿Son similares los resultados de los modelos logit y probit? Explíquelo.
- 11.8** Considérese el modelo de probabilidad lineal $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $\Pr(Y_i = 1 | X_i) = \beta_0 + \beta_1 X_i$.
 a) Demuestre que $E(u_i | X_i) = 0$.
 b) Demuestre que $\text{var}(u_i | X_i) = (\beta_0 + \beta_1 X_i)[1 - (\beta_0 + \beta_1 X_i)]$. [Pista: repasar la Ecuación (2.7).]
 c) ¿Es u_i heteroscedástico? Explíquelo.
 d) (Necesita la Sección 11.3). Obtenga la función de verosimilitud.

- 11.9** Utilícese el modelo de probabilidad lineal estimado que se muestra en la columna (1) de la Tabla 11.2 para responder a lo siguiente:
- Dos candidatos, uno blanco y otro negro, solicitan una hipoteca. Presentan los mismos valores para el resto de regresores distintos de la etnia. ¿Cuánto más probable es que al solicitante negro se le deniegue un préstamo hipotecario?
 - Construya un intervalo de confianza al 95 % para su respuesta en (a).
 - Piense en una variable omitida importante que podría sesgar la respuesta en (a). ¿Cuál es y en qué sentido podría sesgar los resultados?
- 11.10** (Necesita la Sección 11.3 y cálculo). Supóngase que una variable aleatoria Y presenta la siguiente distribución de probabilidad: $\Pr(Y = 1) = p$, $\Pr(Y = 2) = q$ y $\Pr(Y = 3) = 1 - p - q$. Se extrae una muestra aleatoria de tamaño n de esta distribución, y las variables aleatorias se expresan por Y_1, Y_2, \dots, Y_n .
- Obtenga la función de verosimilitud para los parámetros p y q .
 - Obtenga las fórmulas para el EMV de p y q .
- 11.11** (Necesita el Apéndice 11.3). ¿Qué modelo se utiliza para:
- Un estudio que explique el número de minutos que una persona pasa al mes hablando por teléfono móvil?
 - Un estudio que explique las notas (de sobrealiente a suspenso) en una clase grande de Principios de Economía?
 - Un estudio sobre la decisión de los consumidores entre Coca-Cola, Pepsi, o cola genérica?
 - Un estudio del número de teléfonos móviles que posee una familia?

Ejercicios empíricos

- E11.1** Se ha planteado que la prohibición de fumar en el lugar de trabajo puede inducir a los fumadores a dejar de fumar debido a la reducción de sus oportunidades de fumar. En este trabajo se estima el efecto de la prohibición de fumar en el trabajo sobre el consumo de tabaco a partir de los datos de una muestra de 10.000 trabajadores que trabajaban en un espacio interior en EE.UU. desde 1991 a 1993, disponible en la página web del libro de texto http://www.pearsonhighered.com/stock_watson en el archivo **Smoking**. El conjunto de datos contiene información sobre si las personas estaban o no estaban sujetas a una prohibición de fumar en el trabajo, si las personas fumaban, y otras características individuales⁷. Se ofrece una descripción detallada en el archivo **Smoking_Description**, disponible en la página web.
- Estime la probabilidad de fumar para (i) todos los trabajadores, (ii) los trabajadores afectados por la prohibición de fumar en el trabajo, y (iii) los trabajadores no afectados por la prohibición de fumar en el lugar de trabajo.
 - ¿Cuál es la diferencia en la probabilidad de fumar entre los trabajadores afectados por la prohibición de fumar en el trabajo y los trabajadores no afectados por una prohibición de fumar en el lugar de trabajo? Utilice un modelo de probabilidad lineal para determinar si esta diferencia es estadísticamente significativa.
 - Estime un modelo de probabilidad lineal con la variable *smoker* como variable dependiente y las variables explicativas siguientes: *smkban*, *female*, *age*, *age*², *hsdrop*, *hsgrad*, *colsome*, *colgrad*, *black*, y *hispanic*. Compare el efecto estimado de la prohibición de fumar mediante esta regresión con su respuesta en el apartado (b). Sugiera una razón, en base a esta regresión que explique el cambio en la estimación del efecto de la prohibición de fumar entre (b) y (c).

⁷ Estos datos fueron proporcionados por el profesor William Evans de la Universidad de Maryland y se utilizaron en su artículo con Matthew Farrelly y Edward Montgomery, «Do Workplace Smoking Bans Reduce Smoking?», *American Economic Review*, 1999, 89 (4): 728-747.

- d) Contraste la hipótesis de que el coeficiente de la variable *smkban* es igual a cero en la versión poblacional de la regresión de (c) frente a la alternativa de que es distinto de cero, al nivel de significación del 5 %.
- e) Contraste la hipótesis de que la probabilidad de fumar no depende del nivel de educación en la regresión del apartado (c). ¿La probabilidad de fumar aumenta o disminuye con el nivel de educación?
- f) En base a la regresión del apartado (c), ¿existe una relación no lineal entre la variable *age* y la probabilidad de fumar? Represente gráficamente la relación entre la probabilidad de fumar y la variable *age* para $18 \leq age \leq 65$ para un blanco, hombre, titulado universitario, no hispano y sin prohibición de fumar en el lugar de trabajo.

E11.2 Este ejercicio utiliza los mismos datos que el Ejercicio empírico 11.1.

- a) Estime un modelo probit utilizando los mismos regresores que en el Ejercicio empírico 11.1(c).
- b) Contraste la hipótesis de que el coeficiente de la variable *smkban* es igual a cero en la versión poblacional de esta regresión probit frente a la alternativa de que es distinto de cero, al nivel de significación del 5 %. Compare los estadísticos *t* y las conclusiones obtenidas con las del Ejercicio empírico 11.1(d) en base al modelo de probabilidad lineal.
- c) Contraste la hipótesis de que la probabilidad de fumar no depende del nivel de educación en este modelo probit. Compare los resultados con los del Ejercicio empírico 11.1(e) utilizando el modelo de probabilidad lineal.
- d) El señor A es blanco, no hispano, de una edad de 20 años, y abandonó la escuela secundaria. Mediante la regresión probit del apartado (a) y suponiendo que el Sr. A no está sujeto a una prohibición de fumar en el lugar de trabajo, calcule la probabilidad de que el Sr. A fume. Lleve a cabo un nuevo cálculo suponiendo que está sujeto a una prohibición de fumar. ¿Cuál es el efecto de la prohibición de fumar sobre la probabilidad de fumar?
- e) Repita el apartado (d) para la señora B, una mujer negra de 40 años, con título universitario.
- f) Repita los apartados (d) y (e) utilizando el modelo de probabilidad lineal del Ejercicio empírico 11.1(c).
- g) Sobre la base de las respuestas a los apartados (d) a (f), ¿son diferentes los resultados de los modelos probit y de probabilidad lineal? Si lo son, ¿cuál de ellos tiene más sentido? ¿Son elevados los efectos estimados en un sentido real?
- h) ¿Sigue habiendo amenazas importantes a la validez interna?

E11.3 En este ejercicio se estudia el seguro de salud, el estado de salud, y el empleo utilizando una muestra aleatoria de más de 8.000 trabajadores de Estados Unidos encuestados en 1996. Los datos están disponibles en la página web del libro http://www.pearsonhighered.com/stock_watson⁸. Se ofrece una descripción detallada de los mismos en el archivo **Insurance_Description**, asimismo disponible en la página web.

- a) ¿Es menos probable que los trabajadores por cuenta propia tengan seguro de salud en comparación con los asalariados? Si es así, ¿es elevada la diferencia en un sentido real? ¿Es la diferencia estadísticamente significativa?
- b) Los trabajadores por cuenta propia pueden ser sistemáticamente distintos a los asalariados en cuanto a su edad, educación, etc. Tras tener en cuenta estos otros factores, ¿es menos probable que los trabajadores por cuenta propia tengan seguro de salud?
- c) ¿De qué manera varía con la edad la situación en lo que respecta al seguro de salud? ¿Presentan mayor probabilidad de tener un seguro de salud los trabajadores de mayor edad? ¿Menor probabilidad?

⁸ Estos datos fueron proporcionados por el profesor Harvey Rosen de la Universidad de Princeton y se utilizaron en su artículo junto con Craig Perry, «The Self-Employed Are Less Likely Than Wage-Earners to Have Health Insurance So What?», en Douglas Holtz-Eakin y Harvey S. Rosen, eds., *Entrepreneurship and Public Policy* (Cambridge, MA: MIT Press, 2004).

- d) ¿Es distinto el efecto de ser trabajador por cuenta propia sobre la situación en cuanto al seguro para los trabajadores de mayor edad que para los trabajadores más jóvenes?
- e) Se ha argumentado que es menos probable que los trabajadores por cuenta propia estén asegurados, pero a pesar de ello, están tan sanos como los asalariados. ¿Es eso así? ¿Se sostiene el argumento para los trabajadores jóvenes? ¿Y para los trabajadores de mayor edad? ¿Existen problemas de causalidad simultánea que puedan socavar la validez interna de este tipo de análisis estadístico?

APÉNDICE

11.1 La base de datos HMDA de Boston

La base de datos HMDA de Boston fue recogida por investigadores del Banco de la Reserva Federal de Boston. La base de datos combina la información proviniente de las solicitudes de préstamos hipotecarios y de una encuesta de seguimiento de los bancos y de otras instituciones de crédito que recibieron estas solicitudes. Los datos se refieren a las solicitudes de hipotecas realizadas en 1990 en el área metropolitana de Boston. El conjunto de datos completo tiene 2.925 observaciones, en las que constan todas las solicitudes de hipotecas por parte de ciudadanos negros e hispanos además de una muestra aleatoria de las solicitudes de hipotecas realizadas por ciudadanos blancos.

Con el fin de limitar el alcance del análisis realizado en este capítulo, se utiliza solamente el subconjunto de los datos provenientes de residencias unifamiliares (lo que excluye los datos sobre viviendas multifamiliares) y solamente por parte de solicitantes negros y blancos (lo que excluye los datos sobre solicitantes pertenecientes a otros grupos minoritarios). Esto deja 2.380 observaciones. Las definiciones de las variables utilizadas en este capítulo se ofrecen en la Tabla 11.1.

Estos datos fueron proporcionados amablemente por Geoffrey Tootell del Departamento de Investigación del Banco de la Reserva Federal de Boston. Existe información adicional acerca de esta base de datos, junto con las conclusiones alcanzadas por los investigadores del Banco de la Reserva Federal de Boston, recogida en el artículo de Alicia H. Munnell, Geoffrey M. B. Tootell, Geoffrey, Lynne E. Browne y James McEneaney, «Mortgage Lending in Boston: Interpreting HMDA Data», *American Economic Review*, 1996, pp. 25-53.

APÉNDICE

11.2 Estimación máximo verosímil

En este apéndice se ofrece una breve introducción sobre la estimación de máxima verosimilitud en el contexto de los modelos de respuesta binaria tratados en este capítulo. Comenzamos por obtener la probabilidad de éxito p para n observaciones i.i.d. de una variable aleatoria de Bernoulli. Se abordan más tarde los modelos probit y logit y el análisis del pseudo R^2 . Concluimos con el estudio de los errores estándar de las probabilidades estimadas. En este apéndice se utilizan razonamientos de cálculo matemático en dos puntos.

EMV de n variables aleatorias i.i.d. de Bernouilli

El primer paso para el cálculo del EMV consiste en obtener la distribución de probabilidad conjunta. Para n observaciones i.i.d. de una variable aleatoria de Bernoulli, esta distribución de probabilidad conjunta es la extensión del caso $n = 2$ de la Sección 11.3 para un n general:

$$\begin{aligned} \Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\ = [p^{y_1}(1-p)^{(1-y_1)}] \times [p^{y_2}(1-p)^{(1-y_2)}] \times \dots \times [p^{y_n}(1-p)^{(1-y_n)}] \\ = p^{(y_1 + \dots + y_n)}(1-p)^{n - (y_1 + \dots + y_n)}. \end{aligned} \tag{11.13}$$

La función de verosimilitud es la distribución de probabilidad conjunta, considerada como una función de los coeficientes desconocidos. Sea $S = \sum_{i=1}^n Y_i$; por tanto la función de verosimilitud es

$$f_{\text{Bernoulli}}(p; Y_1, \dots, Y_n) = p^S(1-p)^{n-S}. \tag{11.14}$$

El EMV de p es el valor de p que maximiza la probabilidad o verosimilitud en la Ecuación (11.14). La función de verosimilitud se puede maximizar por medio de cálculo matemático. Resulta conveniente maximizar no la verosimilitud, sino su logaritmo (debido a que el logaritmo es una función estrictamente creciente, la maximización de la verosimilitud o de su logaritmo proporciona el mismo estimador). El logaritmo de la verosimilitud es $S \ln(p) + (n - S) \ln(1 - p)$, y la derivada del logaritmo de la verosimilitud respecto a p es

$$\frac{d}{dp} \ln [f_{\text{Bernoulli}}(p; Y_1, \dots, Y_n)] = \frac{S}{p} - \frac{n - S}{1 - p} \quad (11.15)$$

Igualando la derivada de la Ecuación (11.15) a cero y resolviendo para p se obtiene el EMV $\hat{p} = S/n = \bar{Y}$.

EMV del modelo probit

En el modelo probit, la probabilidad de que $Y_i = 1$, condicionada a X_{1i}, \dots, X_{ki} , es $p_i = \Phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$. La distribución de probabilidad condicional para la observación i -ésima es $\Pr[Y_i = y_i | X_{1i}, \dots, X_{ki}] = p_i^{y_i} (1 - p_i)^{1-y_i}$. Suponiendo que $(X_{1i}, \dots, X_{ki}, Y_i)$ son i.i.d., $i = 1, \dots, n$, la distribución de probabilidad conjunta de Y_1, \dots, Y_n , condicionada a las X , es

$$\begin{aligned} \Pr(Y_1 = y_1, \dots, Y_n = y_n | X_{1i}, \dots, X_{ki}, i = 1, \dots, n) \\ = \Pr(Y_1 = y_1 | X_{11}, \dots, X_{k1}) \times \dots \times \Pr(Y_n = y_n | X_{1n}, \dots, X_{kn}) \\ = p_1^{y_1} (1 - p_1)^{1-y_1} \times \dots \times p_n^{y_n} (1 - p_n)^{1-y_n}. \end{aligned} \quad (11.16)$$

La función de verosimilitud es la distribución de probabilidad conjunta, considerada como función de los coeficientes desconocidos. Es habitual considerar el logaritmo de la verosimilitud. De este modo, la función del logaritmo de la verosimilitud es

$$\begin{aligned} \ln [f_{\text{probit}}(\beta_0, \dots, \beta_k; Y_1, \dots, Y_n | X_{1i}, \dots, X_{ki}, i = 1, \dots, n)] \\ = \sum_{i=1}^n Y_i \ln [\Phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})] \\ + \sum_{i=1}^n (1 - Y_i) \ln [1 - \Phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})], \end{aligned} \quad (11.17)$$

donde esta expresión incorpora la fórmula probit para la probabilidad condicional, $p_i = \Phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$.

El EMV para el modelo probit maximiza la función de verosimilitud o, de manera equivalente, el logaritmo de la función de verosimilitud dado en la Ecuación (11.17). Debido a que no existe una fórmula sencilla para el EMV, la función de verosimilitud probit debe maximizarse mediante un algoritmo numérico utilizando el ordenador.

Bajo condiciones generales, los estimadores de máxima verosimilitud son consistentes y presentan una distribución muestral normal en muestras grandes.

EMV del modelo logit

La verosimilitud en el modelo logit se obtiene del mismo modo que la verosimilitud en el modelo probit. La única diferencia es que la probabilidad condicional de éxito p_i del modelo logit está dada por la Ecuación (11.9). En consecuencia, el logaritmo de la verosimilitud del modelo logit viene dado por la Ecuación (11.17), en la que se sustituye $\Phi(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$ por $[1 + e^{-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}]^{-1}$. Al igual que para el modelo probit, no existe una fórmula sencilla para el EMV de los coeficientes del logit, por lo que el logaritmo de la verosimilitud debe maximizarse por métodos numéricos.

Pseudo R^2

El pseudo R^2 compara el valor de la verosimilitud del modelo estimado con el valor de la verosimilitud cuando ninguna de las X se incluye como variable explicativa. En concreto, el pseudo R^2 para el modelo probit es

$$\text{pseudo-}R^2 = 1 - \frac{\ln(f_{\text{probit}}^{\max})}{\ln(f_{\text{Bernoulli}}^{\max})} \quad (11.18)$$

donde f_{probit}^{\max} es el valor máximo de la verosimilitud del probit (que incluye las X) y $f_{Bernoulli}^{\max}$ es el valor máximo de la verosimilitud de Bernoulli (el modelo probit que excluye todas las X).

Los errores estándar de las probabilidades estimadas

Por simplicidad, se considera el caso de un único regresor en el modelo probit. Por tanto la predicción para la probabilidad dado un valor fijo del regresor, x , es $\hat{p}(x) = \Phi(\hat{\beta}_0^{EMV} + \hat{\beta}_1^{EMV}x)$, donde $\hat{\beta}_0^{EMV}$ y $\hat{\beta}_1^{EMV}$ son los EMV de los dos coeficientes probit. Debido a que esta probabilidad estimada depende de los estimadores $\hat{\beta}_0^{EMV}$ y $\hat{\beta}_1^{EMV}$, y debido a que estos estimadores tienen una distribución muestral, la probabilidad estimada tendrá asimismo una distribución muestral.

La varianza de la distribución muestral de $\hat{p}(x)$ se calcula mediante la aproximación de la función de $\Phi(\hat{\beta}_0^{EMV} + \hat{\beta}_1^{EMV}x)$, una función no lineal de $\hat{\beta}_0^{EMV}$ y $\hat{\beta}_1^{EMV}$, mediante una función lineal de $\hat{\beta}_0^{EMV}$ y $\hat{\beta}_1^{EMV}$. En concreto, sea

$$\hat{p}(x) = \Phi(\hat{\beta}_0^{EMV} + \hat{\beta}_1^{EMV}x) \cong c + a_0(\hat{\beta}_0^{EMV} - \beta_0) + a_1(\hat{\beta}_1^{EMV} - \beta_1) \quad (11.19)$$

donde la constante c y los factores a_0 y a_1 dependen de x y se obtienen mediante cálculo. [La Ecuación (11.19) es una expansión de Taylor de primer orden, $c = \Phi(\beta_0 + \beta_1x)$; y a_0 y a_1 son las derivadas parciales $a_0 = \partial\Phi(\beta_0 + \beta_1x)/\partial\beta_0|_{\beta_0^{EMV}, \beta_1^{EMV}}$, y $a_1 = \partial\Phi(\beta_0 + \beta_1x)/\partial\beta_1|_{\beta_0^{EMV}, \beta_1^{EMV}}$. La varianza de $\hat{p}(x)$ puede calcularse ahora mediante la aproximación de la Ecuación (11.19) y la expresión para la varianza de la suma de dos variables aleatorias de la Ecuación (2.31):

$$\begin{aligned} \text{var}[\hat{p}(x)] &\cong \text{var}[c + a_0(\hat{\beta}_0^{EMV} - \beta_0) + a_1(\hat{\beta}_1^{EMV} - \beta_1)] = \\ &= a_0^2 \text{var}(\hat{\beta}_0^{EMV}) + a_1^2 \text{var}(\hat{\beta}_1^{EMV}) + 2a_0a_1 \text{cov}(\hat{\beta}_0^{EMV}, \hat{\beta}_1^{EMV}) \end{aligned} \quad (11.20)$$

Utilizando la Ecuación (11.20), el error estándar de $\hat{p}(x)$ se puede calcular mediante las estimaciones de las varianzas y covarianza de los de EMV.

APÉNDICE

11.3 Otros modelos de variable dependiente limitada

Este apéndice examina algunos de los modelos para variables dependientes limitadas, distintos de las variables binarias, que pueden encontrarse en las aplicaciones económicas. En la mayoría de los casos los estimadores MCO de los parámetros de los modelos con variable dependiente limitada son inconsistentes, y la estimación se realiza habitualmente mediante máxima verosimilitud. Existen varios ejemplos avanzados que se encuentran disponibles para el lector interesado en obtener más información, véase, por ejemplo, Ruud (2000) y Wooldridge (2002).

Modelos de regresión truncados y censurados

Supongamos que se dispone de datos de sección cruzada sobre la compra de coches por parte de individuos en un año determinado. Los compradores de coches tienen unos gastos positivos, que razonablemente pueden ser tratados como variables aleatorias, pero los no compradores tienen gastos por importe de 0 \$. Por tanto, la distribución del gasto en coche es una combinación de una distribución discreta (en cero) y una distribución continua.

El premio Nobel James Tobin desarrolló un modelo útil para una variable dependiente con una distribución en parte continua y en parte discreta (Tobin, 1958). Tobin propuso modelizar el nivel deseado de gasto por parte del individuo i -ésimo de la muestra, Y_i^* que está relacionado con las variables explicativas (por ejemplo, el tamaño de la familia), por medio de un modelo de regresión lineal. Es decir, si existe un único regresor, el nivel deseado de gasto es

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n. \quad (11.21)$$

Si Y_i^* (lo que el consumidor desea gastar) supera cierto límite, como el precio mínimo de un automóvil, el consumidor compra el coche y gasta $Y_i = Y_i^*$, lo cual es observable. Sin embargo, si Y_i^* es menor que el umbral, el gasto que se observa es 0 en lugar de Y_i^* .

Cuando se estima la Ecuación (11.21) mediante los gastos observados Y_i en lugar de Y_i^* , el estimador MCO es inconsistente. Tobin resolvió este problema obteniendo una función de verosimilitud con el supuesto adicional de que u_i presenta una distribución normal, y el EMV resultante ha sido utilizado por los especialistas en econometría aplicada para analizar muchos problemas de economía. En honor a Tobin, la Ecuación (11.21), combinada con el supuesto de normalidad de los errores, se denomina modelo de regresión tobit. El modelo tobit es un ejemplo de un *modelo de regresión censurado*, denominado así debido a que la variable dependiente está «censurada» por encima o por debajo de un límite o umbral determinado.

Modelos de selección muestral

En el modelo de regresión censurada, existen datos sobre los compradores y los que no compran, como sería el caso si los datos se obtuvieran mediante un muestreo aleatorio simple de la población adulta. Sin embargo, si se recogen los datos de los registros de los impuestos sobre las ventas, los datos solamente incluirán a los compradores: no habrá ningún dato para los que no realizan la compra. Los datos que presentan observaciones que no están disponibles por encima o por debajo de un umbral (los datos que se refieren solamente a compradores) se denominan datos truncados. El *modelo de regresión truncado* es un modelo de regresión que se aplica a los datos en los que las observaciones simplemente no están disponibles cuando la variable dependiente está por encima o por debajo de un límite determinado.

El modelo de regresión truncado es un ejemplo de un modelo de selección muestral, en el que el mecanismo de selección (un individuo se encuentra en la muestra en virtud de la compra de un coche) está relacionado con el valor de la variable dependiente (el gasto en un coche). Tal y como se trató en el recuadro de la Sección 11.4, un método de estimación de los modelos de selección muestral consiste en desarrollar dos ecuaciones, una para Y_i^* y otra para cuando Y_i^* es observable. Los parámetros del modelo se pueden estimar entonces mediante máxima verosimilitud, o mediante un procedimiento paso a paso, estimando primero la ecuación de selección y posteriormente estimando la ecuación para Y_i^* . Para un análisis adicional, consulte Ruud (2000, Capítulo 28), Greene (2000, Sección 20.4), o Wooldridge (2002, Capítulo 17).

Datos de recuento

Los *datos de recuento* surgen cuando la variable dependiente es un número que surge de un recuento, por ejemplo, el número de menús consumidos por un cliente en un restaurante en una semana. Cuando estos números son lo suficientemente grandes, la variable puede ser considerada como aproximadamente continua, pero si son pequeños, la aproximación continua resulta inadecuada. El modelo de regresión lineal que se estima por MCO, se puede utilizar para datos de recuento, incluso si el número del recuento es pequeño. Los valores de predicción de la regresión se interpretan como los valores esperados de la variable dependiente, condicionados a los regresores. Por lo tanto, si la variable dependiente es el número de menús consumidos, un valor de predicción de 1,7 significa, en media, 1,7 menús de restaurante a la semana. Sin embargo, como en el modelo de regresión binaria, MCO no aprovecha la estructura particular de los datos de recuento y puede dar lugar a predicciones sin sentido, por ejemplo, -0,2 menús de restaurante a la semana. Sin embargo, así como los modelos probit y logit eliminan las predicciones sin sentido cuando la variable dependiente es binaria, existen modelos particulares que lo hacen para el caso de datos de recuento. Los dos modelos más ampliamente utilizados son el de Poisson y los modelos de regresión binomial negativa.

Respuesta ordenada

Los *datos de respuesta ordenada* surgen cuando las categorías cualitativas mutuamente excluyentes presentan una ordenación natural, tales como la obtención de un diploma de escuela secundaria, educación universitaria (pero sin terminar), o graduarse en la universidad. Al igual que los datos de recuento, los datos de respuesta ordenada tienen un orden natural, pero a diferencia de los datos de recuento, no tienen valores numéricos naturales.

Debido a que no hay valores numéricos naturales en los datos de respuesta ordenada, MCO resulta inapropiado. En vez de eso, los datos ordenados son analizados a menudo mediante una generalización de probit denominada *modelo probit ordenado*, en el que las probabilidades de cada resultado (por ejemplo, la educación universitaria), condicionada a las variables independientes (tales como el ingreso de los padres), se modelizan utilizando la distribución normal acumulada.

Datos de elección discreta

Una variable de *elección discreta* o de *opción múltiple* puede tomar varios valores cualitativos sin orden. Un ejemplo en economía es el medio de transporte elegido por un viajero: se puede tomar el metro, el autobús, conducir un coche, o hacer el camino por los propios medios (a pie, en bicicleta). Si tuviéramos que analizar estas opciones, la variable dependiente tendría cuatro resultados posibles (metro, autobús, coche, tracción humana). Estos resultados no están ordenados de ninguna manera natural. En cambio, los resultados son una elección entre distintas alternativas cualitativas.

La tarea econométrica consiste en modelizar la probabilidad de elegir entre las distintas opciones, teniendo en cuenta diversos regresores como las características individuales (distancia de la vivienda hasta la estación del metro) y las características de cada opción (el precio del billete de metro). Como se analizó en el recuadro de la Sección 11.3, pueden desarrollarse modelos para el análisis de los datos de elección discreta a partir de los fundamentos de la maximización de utilidad. Las probabilidades de elección individual pueden expresarse en forma de probit o logit, y esos modelos se denominan modelos de regresión *probit multinomial* y *logit multinomial*.

Regresión con variables instrumentales

En el Capítulo 9 se analizaron algunos problemas, entre ellos las variables omitidas, los errores en las variables, y la causalidad simultánea, que hacen que pueda ocurrir que el término de error esté correlacionado con el regresor. El sesgo de variable omitida puede tratarse de forma directa mediante la inclusión de la variable omitida en una regresión múltiple, pero esto es factible solamente si se dispone de datos de la variable omitida. Y a veces, como cuando la causalidad va tanto de X hacia Y como de Y hacia X , es decir, cuando existe sesgo de causalidad simultánea, la regresión múltiple sencillamente no puede eliminar el sesgo. Si no es posible una solución directa para estos problemas, se necesita un nuevo procedimiento.

La **regresión con variables instrumentales (VI)** es un método general para la obtención de un estimador consistente de los coeficientes desconocidos de la función de regresión poblacional cuando la variable explicativa, X , está correlacionada con el término de error, u . Para comprender cómo funciona la regresión VI, imaginemos que la variación de X tiene dos partes: una parte que, por alguna razón, está correlacionada con u (esta es la parte que causa los problemas) y una segunda parte que está incorrelacionada con u . Si se dispone de información que permita aislar la segunda parte de X , podríamos centrarnos en esas variaciones de X que no están correlacionadas con u y desocuparnos de las variaciones de X que sesgan las estimaciones MCO.

Esto es, de hecho, lo que hace la regresión VI. La información acerca de los movimientos de X que no están correlacionados con u se obtiene a partir de una o más variables adicionales, que se denominan **variables instrumentales** o sencillamente **instrumentos**. La regresión de variables instrumentales utiliza estas variables adicionales como herramientas o «instrumentos» para aislar los movimientos de X que están incorrelacionados con u , lo que a su vez permite una estimación consistente de los coeficientes de la regresión.

En las dos primeras secciones de este capítulo se describen los mecanismos y los supuestos de la regresión VI: por qué funciona la regresión VI, qué es un instrumento válido, y cómo se lleva a cabo y se interpreta el método más común de regresión VI, los mínimos cuadrados en dos etapas. La clave de un análisis empírico exitoso mediante variables instrumentales consiste en encontrar instrumentos válidos, y la Sección 12.3 trata la cuestión de cómo establecer si un conjunto de instrumentos es válido. A modo de ilustración, la Sección 12.4 utiliza la regresión VI para estimar la elasticidad de la demanda de cigarrillos. Finalmente, la Sección 12.5 aborda la difícil cuestión de dónde buscar en primer término unos instrumentos válidos.

12.1 El estimador VI con regresor único e instrumento único

Empezamos con el caso de un único regresor, X , que puede estar correlacionado con el error de la regresión, u . Si X y u están correlacionadas, el estimador MCO es inconsistente; es decir, puede no estar cercano al verdadero valor del coeficiente de regresión, incluso cuando la muestra es muy grande [véase la Ecuación (6.1)]. Tal y como se estudió en la Sección 9.2, esta correlación entre X y u puede provenir de varias fuentes, incluyendo variables omitidas, errores en las variables (errores de medición en las variables explicativas), y causalidad simultánea (cuando la causalidad va tanto «hacia atrás», de Y hacia X , como «hacia de-

lante», de X hacia Y). Cualquiera que sea el origen de la correlación entre X y u , si existe una variable instrumental válida, Z , el efecto sobre Y de un cambio unitario en X puede estimarse utilizando el estimador de variables instrumentales.

El modelo VI y los supuestos

El modelo de regresión poblacional que relaciona la variable dependiente Y_i y la variable independiente X_i es

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n, \quad (12.1)$$

donde u_i como de costumbre, es el término de error que representa los factores omitidos que determinan Y_i . Si X_i y u_i están correlacionadas, el estimador MCO es inconsistente. La estimación con variables instrumentales utiliza una variable «instrumental» Z adicional para aislar esa parte de X que no está correlacionada con u_i .

Endogeneidad y exogeneidad. La regresión con variables instrumentales emplea terminología especializada para distinguir las variables que están correlacionadas con el término de error poblacional u de las que no lo están. Las variables correlacionadas con el término de error se denominan **variables endógenas**, mientras que las variables que no están correlacionadas con el término de error se denominan **variables exógenas**. El origen histórico de estos términos se remonta a los modelos con varias ecuaciones, en los que una variable «endógena» se determina dentro del modelo, mientras que una variable «exógena» se determina fuera del modelo. Por ejemplo, en la Sección 9.2 se consideraba la posibilidad de que si las calificaciones en los exámenes produjeran una disminución de la ratio estudiantes-maestros a causa de la intervención política y el aumento de la financiación, la causalidad funcionaría tanto desde la ratio estudiantes-maestros hacia las calificaciones como desde las calificaciones en los exámenes hacia la ratio estudiantes-maestros. Esto se representaba matemáticamente mediante un sistema de dos ecuaciones simultáneas [Ecuaciones (9.3) y (9.4)], una para cada relación de causalidad. Como se estudió en la Sección 9.2, debido a que tanto las calificaciones en los exámenes como la ratio estudiantes-maestros se determinan dentro del modelo, ambas variables están correlacionadas con el término de error poblacional u ; es decir, en este ejemplo, ambas variables son endógenas. Por el contrario, una variable exógena, que se determina fuera el modelo, no está correlacionada con u .

Las dos condiciones para un instrumento válido. Una variable instrumental válida («instrumento») debe cumplir dos condiciones, conocidas como **condición de relevancia del instrumento** y **condición de exogeneidad del instrumento**:

1. Relevancia del instrumento: $\text{corr}(Z_i, X_i) \neq 0$.
2. Exogeneidad del instrumento: $\text{corr}(Z_i, u_i) = 0$.

Si un instrumento es relevante, entonces la variación en el instrumento está relacionada con la variación en X_i . Si, además, el instrumento es exógeno, entonces esa parte de la variación de X_i captada por la variable instrumental es exógena. Por tanto, un instrumento que sea relevante y exógeno puede captar los movimientos de X_i que son exógenos. Esta variación exógena a su vez puede ser utilizada para estimar el coeficiente poblacional β_1 .

Las dos condiciones para que un instrumento sea válido son vitales para la regresión de variables instrumentales, y volveremos a ellas (y su extensión para varios regresores y varios instrumentos) en repetidas ocasiones a lo largo de este capítulo.

El estimador de mínimos cuadrados en dos etapas

Si el instrumento Z cumple los requisitos de relevancia y exogeneidad, el coeficiente β_1 puede ser estimado mediante un estimador VI denominado de **mínimos cuadrados en dos etapas** (MC2E). Como el nombre sugiere, el estimador de mínimos cuadrados en dos etapas se calcula en dos fases. La primera etapa descompone X en dos componentes: una componente problemática que puede estar correlacionada con el

error de la regresión y otra componente sin problemas que no está correlacionada con el error. La segunda etapa utiliza la componente sin problemas para estimar β_1 .

La primera etapa comienza con una regresión poblacional que liga a X con Z :

$$X_i = \pi_0 + \pi_1 Z_i + v_i, \quad (12.2)$$

donde π_0 es el término independiente o intercepto, π_1 es la pendiente, y v_i es el término de error. Esta regresión proporciona la necesaria descomposición de X_i . Una componente es $\pi_0 + \pi_1 Z_i$, la parte de X_i que puede predecirse mediante Z_i . Debido a que Z_i es exógena, esta componente de X_i está incorrelacionada con u_i , el término de error de la Ecuación (12.1). El otro componente de X_i es v_i , que es la componente problemática de X_i que está correlacionada con u_i .

La idea que está detrás de MC2E es utilizar la componente no problemática de X_i , $\pi_0 + \pi_1 Z_i$ y hacer caso omiso de v_i . La única complicación es que los valores de π_0 y π_1 son desconocidos, por lo que $\pi_0 + \pi_1 Z_i$ no puede calcularse. En consecuencia, la primera etapa de MC2E consiste en aplicar MCO a la Ecuación (12.2) y utilizar los valores de predicción de la regresión MCO, $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$, donde $\hat{\pi}_0$ y $\hat{\pi}_1$ son las estimaciones MCO.

La segunda etapa de MC2E es sencilla: la regresión de Y_i sobre \hat{X}_i por MCO. Los estimadores resultantes de la regresión de la segunda etapa son los estimadores MC2E, $\hat{\beta}_0^{MC2E}$ y $\hat{\beta}_1^{MC2E}$.

¿Por qué funciona la regresión VI?

Dos ejemplos proporcionan algo de intuición sobre por qué la regresión VI resuelve el problema de la correlación entre X_i y u_i .

Ejemplo #1: el problema de Philip Wright. El método de estimación con variables instrumentales fue publicado por primera vez en 1928 en un apéndice de un libro escrito por Philip G. Wright (Wright, 1928), aunque parece que las ideas clave de la regresión VI fueron desarrolladas en colaboración con su hijo, Sewall Wright (véase el recuadro). Philip Wright estaba preocupado por un problema económico importante de su época: cómo establecer un arancel a las importaciones (un impuesto sobre los bienes importados), sobre los aceites vegetales y las grasas animales, como la mantequilla y el aceite de soja. En la década de 1920, los aranceles a las importaciones fueron una fuente muy importante de ingresos fiscales para Estados Unidos. La clave para comprender el efecto económico de un arancel era disponer de estimaciones cuantitativas sobre las curvas de demanda y oferta de los bienes. Recordemos que la elasticidad de la oferta es la variación porcentual en la cantidad ofertada derivada de un aumento del 1 % en el precio y que la elasticidad de la demanda es la variación porcentual en la cantidad demandada que resulta de un 1 % de aumento en el precio. Philip Wright necesitaba estimaciones de las elasticidades de la oferta y la demanda.

Para concretar, consideremos el problema de la estimación de la elasticidad de la demanda de la mantequilla. Recordemos del Concepto clave 8.2 que el coeficiente de una ecuación lineal que relaciona $\ln(Y_i)$ con $\ln(X_i)$ puede interpretarse como la elasticidad de Y con respecto a X . En el problema de Wright, esto sugiere la ecuación de demanda

$$\ln(Q_i^{\text{mantequilla}}) = \beta_0 + \beta_1 \ln(P_i^{\text{mantequilla}}) + u_i \quad (12.3)$$

donde $Q_i^{\text{mantequilla}}$ es la observación i -ésima de la cantidad de mantequilla consumida, $P_i^{\text{mantequilla}}$ es su precio, y u_i representa los otros factores que afectan a la demanda, como la renta y los gustos del consumidor. En la Ecuación (12.3), un aumento del 1 % en el precio de la mantequilla supone una variación porcentual de β_1 en la demanda, por lo que β_1 es la elasticidad de la demanda.

Philip Wright tenía datos sobre el consumo total anual de la mantequilla y su precio medio anual en Estados Unidos desde 1912 hasta 1922. Hubiera sido fácil utilizar estos datos para estimar la elasticidad de la demanda aplicando MCO a la Ecuación (12.3), pero tenía una intuición clave: debido a las interacciones entre la oferta y la demanda, el regresor, $\ln(P_i^{\text{mantequilla}})$, es probable que estuviera correlacionado con el término de error.

¿Quién inventó la regresión de variables instrumentales?

La regresión con variables instrumentales fue en primer lugar propuesta como una solución al problema de causalidad simultánea en econometría en el apéndice del libro de 1928 de Philip G. Wright, *El Arancel sobre los Aceites Animales y Vegetales*¹. Si se quiere saber cómo los aceites animales y vegetales se producían, transportaban y vendían a principios del siglo xx, las primeras 285 páginas del libro son la respuesta. No obstante, los económetras están más interesados en el Apéndice B. El apéndice ofrece dos deducciones del «método para introducir factores externos» —lo que ahora se conoce como el estimador de variables instrumentales— y utiliza la regresión VI para estimar las elasticidades de la oferta y la demanda de mantequilla y de aceite de linaza. Philip hubiese sido un solitario economista con un legado intelectual escaso de no ser por este apéndice, pero su hijo Sewall se convirtió en un prominente genetista de poblaciones y estadístico. Debido a que el material matemático del apéndice es tan diferente del resto del libro, muchos económetras suponen que el hijo de Philip, Sewall Wright escribió el apéndice de forma anónima. Entonces, ¿quién escribió el Apéndice B?

De hecho, tanto el padre como el hijo podrían haber sido el autor. Philip Wright (1861-1934) obtuvo un título de máster en economía por la Universidad de Harvard en 1887, y fue profesor de matemáticas y economía (así como de literatura y educación física) en una pequeña facultad de Illinois. En una reseña del libro [Wright (1915)], utilizó un gráfico como los de las Figuras 12.1a y 12.1b para mostrar cómo una regresión de la cantidad sobre el precio no estimaría, en general, una curva de demanda, sino que estima una combinación de las curvas de oferta y demanda. A principios de la década de 1920, Sewall Wright (1889-1988) estaba investigando el análisis estadístico de ecuaciones múltiples con varias variables causales en el contexto de la genética, investigación que en parte le llevó a ostentar un puesto de catedrático en 1930 en la Universidad de Chicago.

Aunque sea demasiado tarde para preguntar a Philip o a Sewall quién escribió el Apéndice B, nunca es demasiado tarde para realizar algo de trabajo de detective estadístico. La estilometría es el subcampo de la estadística, inventado por Frederick Mosteller y David Wallace (1963), que utiliza las diferencias sutiles e inconscientes en los estilos de escritura para identificar la autoría de textos dudosos utilizando el análisis estadístico de las estructuras gramaticales y la elección de las palabras. El campo ha obtenido éxitos verificados, como el de Donald Foster (1996) descubriendo a Joseph Klein como el autor de la novela política *Primary Colors*. Cuando se compara estadísticamente el Apéndice B con los textos conocidos que escribieron de forma independiente Philip y Sewall, los resultados son claros: Philip fue el autor. ¿Quiere esto decir que Philip G. Wright inventó la regresión VI? no del todo. Recientemente, la correspondencia entre Philip y Sewall de mediados de la década de 1920 ha salido a la luz, y esta correspondencia demuestra que el desarrollo de la regresión VI fue una colaboración intelectual conjunta entre padre e hijo. Para obtener más información, véase Stock y Trebbi (2003).

¹ N. del T.: *The Tariff on Animal and Vegetable Oils* es el título original de la obra.



Philip G. Wright

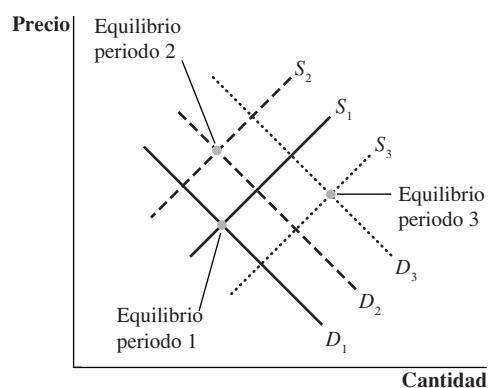


Sewall Wright

Para ver esto, observemos la Figura 12.1a, que muestra las curvas de demanda y oferta en el mercado de la mantequilla para tres años diferentes. Las curvas de oferta y de demanda del primer periodo se designan por D_1 y S_1 , y el precio y la cantidad de equilibrio se determinan mediante su intersección. En el año 2, se traslada la demanda desde D_1 hasta D_2 (por ejemplo, debido a un aumento de la renta) y se traslada la oferta de S_1 a S_2 (debido a un aumento en el coste de producción de la mantequilla); el precio de equilibrio y la cantidad se determinan por la intersección de las nuevas curvas de oferta y demanda. En el año 3, los factores que afectan a la demanda y a la oferta cambian de nuevo; la demanda se traslada de nuevo hasta D_3 , la oferta se traslada hasta S_3 , y se alcanza una nueva cantidad y precio de equilibrio. La Figura 12.1b muestra los pares de cantidad y precio de equilibrio de estos tres períodos y para los ocho años siguientes, donde en cada año las curvas de oferta y de demanda están sujetas a los desplazamientos o cambios asociados con factores distintos del precio que afectan a la oferta y a la demanda de este mercado. Este diagrama de dispersión es como el que Wright habría visto cuando representó sus datos. Tal y como el mismo razonó, el ajuste de una recta a estos puntos mediante MCO no estimaría ni una curva de demanda ni una curva de oferta, ya que los puntos han sido determinados tanto por cambios en la demanda como en la oferta.

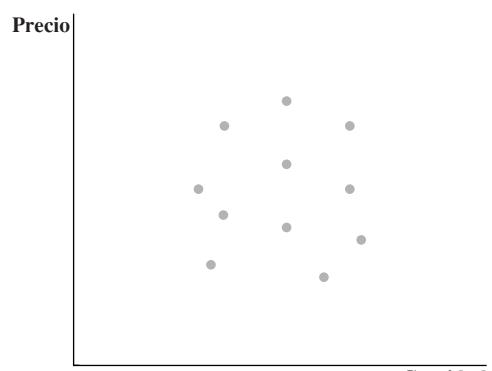
FIGURA 12.1 Datos sobre los precios y las cantidades de equilibrio

(A) El precio y la cantidad se determinan mediante la intersección de las curvas de oferta y demanda. El equilibrio en el primer periodo está determinado por la intersección de la curva de demanda D_1 y la curva de oferta S_1 . El equilibrio en el segundo periodo es la intersección de D_2 y S_2 , y el equilibrio en el tercer periodo es la intersección entre D_3 y S_3 .



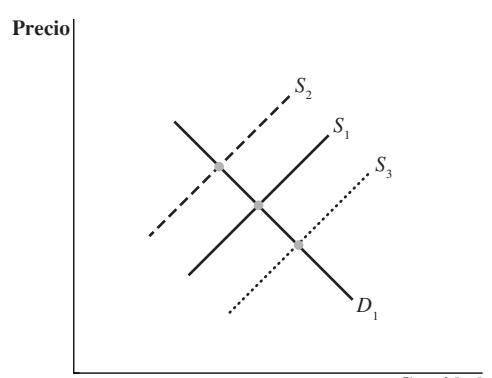
(a) Demanda y oferta en tres períodos

(B) El diagrama de dispersión muestra el precio y la cantidad de equilibrio para 11 períodos diferentes. Las curvas de oferta y demanda están ocultas. ¿Se pueden determinar las curvas de oferta y demanda a partir de los puntos del diagrama de dispersión?



(b) Precio y cantidad de equilibrio para 11 períodos

(C) Cuando la curva de oferta se desplaza de S_1 a S_2 y S_3 , pero la curva de demanda se mantiene en D_1 , los precios de equilibrio y cantidades dibujan la curva de demanda.



(c) Precio y cantidad de equilibrio cuando solo se desplaza la curva de oferta

Wright se dio cuenta de que una manera de solucionar este problema era encontrar una tercera variable que desplazara la oferta pero dejará sin desplazar la demanda. La Figura 12.1c muestra lo que sucede cuando una variable de este tipo desplaza la curva de oferta, pero la demanda permanece estable. Ahora todos los pares de precios y cantidades de equilibrio se encuentran dentro de una curva de demanda estable, y la pendiente de la curva de demanda resulta fácil de estimar. En la formulación de variable instrumental del problema de Wright, esta tercera variable, la variable instrumental, está correlacionada con el precio (desplaza la curva de oferta, lo que conlleva una variación en el precio), pero no está correlacionada con u (la curva de demanda se mantiene estable). Wright consideró distintas variables instrumentales potenciales, una

fue las condiciones meteorológicas. Por ejemplo, unas precipitaciones por debajo de la media en una región lechera podrían poner en peligro el pastoreo y por lo tanto reducir la producción de mantequilla a un precio determinado (desplazaría la curva de oferta hacia la izquierda y aumentaría el precio de equilibrio), por lo que la lluvia en una región lechera satisface la condición de relevancia del instrumento. Pero las lluvias en una región lechera no deberían tener una influencia directa sobre la demanda de mantequilla, por lo que la correlación entre las precipitaciones en una región lechera y u_i sería igual a cero; es decir, la variable de precipitaciones en una región lechera satisface la condición de exogeneidad del instrumento.

Ejemplo #2: la estimación del efecto sobre las calificaciones en las pruebas del tamaño de las clases. A pesar de tener en cuenta las características del estudiante y del distrito, las estimaciones del efecto sobre las calificaciones en las pruebas del tamaño de las clases, presentadas en la Parte II, todavía podrían presentar sesgo de variables omitidas proveniente de variables no observables tales como las oportunidades de aprendizaje fuera de la escuela o la cualificación de los maestros. Si los datos sobre estas variables no estuvieran disponibles, este sesgo de variables omitidas no podría ser abordado mediante la inclusión de las variables en las regresiones múltiples.

La regresión con variables instrumentales proporciona un enfoque alternativo de este problema. Consideremos el siguiente ejemplo hipotético: algunas escuelas de California se ven obligadas a cerrar por reparaciones a causa de un terremoto ocurrido en verano. Los distritos más cercanos al epicentro resultan los más afectados. Un distrito que tenga algunas escuelas cerradas necesita «duplicar» su número de estudiantes de manera temporal, lo que aumenta el tamaño de las clases. Esto significa que la distancia al epicentro satisface la condición de relevancia del instrumento, ya que está correlacionada con el tamaño de la clase. Pero si la distancia al epicentro no está correlacionada con cualquiera de los otros factores que afectan al rendimiento estudiantil (como el hecho de si los estudiantes están aprendiendo inglés), entonces será exógena, debido a que no está correlacionada con el término de error. Por tanto, la variable instrumental, distancia al epicentro, podría ser utilizada para eludir el sesgo de variables omitidas y para estimar el efecto del tamaño de las clases en las calificaciones obtenidas.

La distribución muestral del estimador MC2E

La distribución exacta del estimador MC2E para muestras pequeñas es complicada. No obstante, como la del estimador MCO, su distribución en muestras grandes es muy sencilla: el estimador MC2E es consistente y se distribuye normalmente.

Fórmula del estimador MC2E. A pesar de que las dos etapas de MC2E hacen que el estimador parezca complicado, cuando hay una sola X y un único instrumento Z , tal y como se supone en esta sección, existe una fórmula sencilla para el estimador MC2E. Sea s_{ZY} la covarianza muestral entre Z e Y y sea s_{ZX} la covarianza muestral entre Z y X . Como se muestra en el Apéndice 12.2, el estimador MC2E con un único instrumento es

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}} \quad (12.4)$$

Es decir, el estimador MC2E de β_1 es el cociente entre la covarianza muestral entre Z e Y y la covarianza muestral entre Z y X .

Distribución muestral de $\hat{\beta}_1^{MC2E}$ cuando el tamaño de la muestra es grande. La fórmula de la Ecuación (12.4) se puede utilizar para demostrar que $\hat{\beta}_1^{MC2E}$ es consistente y, en muestras grandes, se distribuye normalmente. El argumento se recoge aquí, y los detalles matemáticos en el Apéndice 12.3.

El argumento para demostrar que $\hat{\beta}_1^{MC2E}$ es consistente combina el supuesto de que Z_i es relevante y exógeno con la consistencia de las covarianzas muestrales respecto de las covarianzas poblacionales. Para comenzar, considérese que debido a que $Y_i = \beta_0 + \beta_1 X_i + u_i$ en la Ecuación (12.1),

$$\text{cov}(Z_i, Y_i) = \text{cov}[Z_i, (\beta_0 + \beta_1 X_i + u_i)] = \beta_1 \text{cov}(Z_i, X_i) + \text{cov}(Z_i, u_i), \quad (12.5)$$

donde la segunda igualdad se deduce a partir de las propiedades de las covarianzas [Ecuación (2.33)]. Por el supuesto de exogeneidad del instrumento, $\text{cov}(Z_i, u_i) = 0$, y por el supuesto de relevancia del instrumento, $\text{cov}(Z_i, X_i) \neq 0$. Por tanto, si el instrumento es válido, la Ecuación (12.5) implica que

$$\beta_1 = \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} \quad (12.6)$$

Es decir, el coeficiente poblacional β_1 es el cociente entre la covarianza poblacional entre Z e Y y la covarianza poblacional entre Z y X .

Como se analizó en la Sección 3.7, la covarianza muestral es un estimador consistente de la covarianza poblacional; es decir, $s_{ZY} \xrightarrow{p} \text{cov}(Z_i, Y_i)$ y $s_{ZX} \xrightarrow{p} \text{cov}(Z_i, X_i)$. Se deduce de las Ecuaciones (12.4) y (12.6) que el estimador MC2E es consistente:

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} = \beta_1 \quad (12.7)$$

La fórmula de la Ecuación (12.4) puede utilizarse asimismo para demostrar que la distribución muestral de $\hat{\beta}_1^{MC2E}$ es normal en muestras grandes. La razón es la misma que para todos los otros estimadores mínimo cuadrático que se han considerado: el estimador MC2E es una media de variables aleatorias, y cuando el tamaño de la muestra es grande, el teorema central del límite nos dice que las medias de variables aleatorias se distribuyen normalmente. En concreto, el numerador de la expresión para $\hat{\beta}_1^{MC2E}$ en la Ecuación (12.4) es $s_{ZY} = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})$, una media de $(Z_i - \bar{Z})(Y_i - \bar{Y})$. Un poco de álgebra, como la que se esboza en el Apéndice 12.3, muestra que debido a este promedio el teorema central del límite implica que, en muestras grandes, $\hat{\beta}_1^{MC2E}$ presenta una distribución muestral que se aproxima a $N(\beta_1, \sigma_{\hat{\beta}_1^{MC2E}}^2)$, donde

$$\sigma_{\hat{\beta}_1^{MC2E}}^2 = \frac{1}{n} \frac{\text{var}[(Z_i - \mu_Z)u_i]}{[\text{cov}(Z_i, X_i)]^2} \quad (12.8)$$

Inferencia estadística mediante la distribución para muestras grandes. La varianza $\sigma_{\hat{\beta}_1^{MC2E}}^2$ se puede estimar mediante la estimación de los términos de varianza y covarianzas que aparecen en la Ecuación (12.8), y la raíz cuadrada de la estimación de $\sigma_{\hat{\beta}_1^{MC2E}}^2$ es el error estándar del estimador VI. Esto se obtiene automáticamente mediante los comandos de la regresión MC2E de los paquetes de *software* econométrico. Debido a que $\hat{\beta}_1^{MC2E}$ se distribuye normalmente en muestras grandes, los contrastes de hipótesis acerca de β_1 se pueden realizar mediante el cálculo del estadístico t , y un intervalo de confianza al 95 % para muestras grandes viene dado por $\hat{\beta}_1^{MC2E} \pm 1,96ES(\hat{\beta}_1^{MC2E})$.

Aplicación a la demanda de cigarrillos

Philip Wright estaba interesado en la elasticidad de la demanda de mantequilla, pero en la actualidad otros productos básicos, tales como los cigarrillos, figuran de una forma más destacada en el centro de los debates sobre las políticas públicas. Una herramienta en la búsqueda de la reducción de las enfermedades y las muertes causadas por el tabaco —y en los costes o externalidades, en los que se incurre debido a esas enfermedades y que son soportados por el resto de la sociedad— consiste en gravar los cigarrillos tan duramente como para conseguir que el número de fumadores actuales se reduzca y se disuada a los potenciales nuevos fumadores de adquirir el hábito. Pero exactamente, ¿cuál es la cuantía de aumento de impuestos necesaria para conseguir hacer mella en el consumo de cigarrillos? Por ejemplo, ¿cuál debería ser el precio de venta después de impuestos de los cigarrillos necesario para lograr una reducción del 20 % en el consumo de cigarrillos?

La respuesta a esta pregunta depende de la elasticidad de la demanda de cigarrillos. Si la elasticidad es -1 , entonces el objetivo del 20 % en el consumo se puede lograr mediante un 20 % de aumento en el precio. Si la elasticidad es $-0,5$, entonces el precio debe subir un 40 % para lograr reducir el consumo en

un 20 %. Por descontado, la elasticidad de la demanda de cigarrillos no es conocida: debe estimarse a partir de los datos sobre los precios y las ventas. Pero, como con la mantequilla, debido a las interacciones que existen entre oferta y demanda, la elasticidad de la demanda de cigarrillos no puede estimarse consistentemente mediante una regresión MCO del logaritmo de la cantidad sobre el logaritmo del precio.

Por lo tanto, se utiliza MC2E para estimar la elasticidad de la demanda de cigarrillos utilizando los datos anuales de los 48 estados contiguos de EE.UU. desde 1985 hasta 1995 (los datos se describen en el Apéndice 12.1). Por ahora, todos los resultados son para la sección cruzada de los estados en 1995; los resultados para los datos de años anteriores (datos de panel) se presentan en la Sección 12.4.

La variable instrumental, $ImpuestoVentas_i$, es la parte del impuesto sobre los cigarrillos que proviene del impuesto general a las ventas, medido en dólares por paquete (en dólares constantes, deflactado por el Índice de Precios al Consumidor). El consumo de cigarrillos, $Q_i^{cigarrillos}$ es el número de paquetes de cigarrillos vendidos per cápita en el estado, y el precio, $P_i^{cigarrillos}$ es el precio real promedio por paquete de cigarrillos incluidos todos los impuestos.

Antes de utilizar MC2E es esencial cuestionarse si se cumplen las dos condiciones para la validez del instrumento. Volveremos a tratar este tema en detalle en la Sección 12.3, en la que se proporcionan algunas herramientas estadísticas que ayudan en esta valoración. Aún con estas herramientas estadísticas, los criterios interpretan un papel importante, por lo que resulta útil preguntarse acerca de si resulta verosímil que el impuesto sobre las ventas de cigarrillos pueda cumplir los dos requisitos.

En primer lugar se considera la relevancia del instrumento. Debido a que un impuesto sobre las ventas elevado aumenta el precio de venta después de impuestos $P_i^{cigarrillos}$, resulta verosímil pensar que el impuesto sobre las ventas por paquete satisface la condición de relevancia del instrumento.

Se considera ahora la exogeneidad. Para que el impuesto sobre las ventas sea exógeno, debe estar inco-
rrelacionado con el error en la ecuación de la demanda; es decir, el impuesto sobre ventas debe afectar a la demanda de cigarrillos solo de forma indirecta a través del precio. Esto parece verosímil: los tipos impositivos generales sobre las ventas varían de un estado a otro, pero lo hacen principalmente porque los diferentes estados eligen diferentes combinaciones de impuestos, sobre las ventas, la renta, la propiedad, y otros, para financiar los compromisos públicos. Estas decisiones acerca de las finanzas públicas se toman en función de consideraciones políticas, no por factores relacionados con la demanda de cigarrillos. Se tratará con más profundidad la credibilidad de este supuesto en la Sección 12.4, pero por ahora se mantiene como hipótesis de trabajo.

En el software estadístico moderno, la primera etapa del MC2E se calcula de forma automática, por lo que no es necesario llevar a cabo esta regresión de forma específica para calcular el estimador MC2E. Aun así, es una buena idea echar un vistazo a la regresión de la primera etapa. A partir de datos de los 48 estados en 1995, es

$$\widehat{\ln(P_i^{cigarrillos})} = 4,63 + 0,031 \text{ImpuestoVentas}_i. \quad (12.9)$$

(0,03) (0,005)

Como era de esperar, impuestos sobre las ventas elevados significan precios después de impuestos más altos. El R^2 de esta regresión es del 47 %, por lo que la variación en el impuesto sobre las ventas de cigarrillos explica el 47 % de la varianza de los precios de los cigarrillos entre los distintos estados.

En la segunda etapa de MC2E, $\ln(Q_i^{cigarrillos})$ se regresa sobre $\widehat{\ln(P_i^{cigarrillos})}$ mediante MCO. La función de regresión resultante estimada es

$$\widehat{\ln(Q_i^{cigarrillos})} = 9,72 - 1,08 \widehat{\ln(P_i^{cigarrillos})}. \quad (12.10)$$

Esta función de regresión estimada se expresa a partir del regresor de la segunda etapa, el valor estimado $\widehat{\ln(P_i^{cigarrillos})}$. Sin embargo, es habitual y menos engorroso presentar simplemente la función de regresión estimada con $\ln(P_i^{cigarrillos})$ en lugar de $\widehat{\ln(P_i^{cigarrillos})}$. Presentadas en esta notación, las estimaciones MC2E y los errores estándar heterocedástico-robustos son

$$\widehat{\ln(Q_i^{cigarrillos})} = 9,72 - 1,08 \ln(P_i^{cigarrillos}). \quad (12.11)$$

(1,53) (0,32)

La estimación MC2E sugiere que la demanda de cigarrillos es sorprendentemente elástica, en virtud de su naturaleza adictiva: un aumento en el precio de un 1 % reduce el consumo en un 1,08 %. Pero, recordando la discusión acerca de la exogeneidad de los instrumentos, tal vez esta estimación no debería todavía ser tomada demasiado en serio. Aunque la elasticidad se estimara mediante una variable instrumental, podrían aún existir variables omitidas que estuvieran correlacionadas con el impuesto sobre las ventas por paquete. Una de las principales candidatas es la variable de renta o ingresos: los estados con mayores ingresos podrían depender relativamente menos de un impuesto sobre las ventas y más de un impuesto sobre la renta a la hora de financiar al gobierno estatal. Además, la demanda de cigarrillos supuestamente depende de la renta. Por tanto nos gustaría reestimar nuestra ecuación de demanda incluyendo los ingresos como regresor adicional. No obstante, para hacerlo, es necesario en primer lugar extender el modelo de regresión VI para incluir variables explicativas adicionales.

12.2 El modelo general de regresión VI

El modelo general de regresión VI tiene cuatro tipos de variables: la variable dependiente, Y ; los regresores endógenos problemáticos, como el precio de los cigarrillos, que está correlacionado con el término de error y que etiquetaremos como X ; regresores adicionales, denominados **variables exógenas incluidas**, que etiquetaremos como W ; y variables instrumentales, Z . En general, puede haber varios regresores endógenos (X), varios regresores exógenos incluidos (W), y varias variables instrumentales (Z).

Para que la regresión VI sea posible, debe haber al menos tantas variables instrumentales (Z) como regresores (X). En la Sección 12.1, había un único regresor endógeno y un único instrumento. Tener (al menos) un instrumento para este único regresor endógeno resultaba esencial. Sin el instrumento no podría haberse calculado el estimador de variables instrumentales: no existiría la regresión de la primera etapa del MC2E.

La relación entre el número de instrumentos y el número de regresores endógenos tiene su propia terminología. Se dice que los coeficientes de regresión están **exactamente identificados** si el número de instrumentos (m) es igual al número de regresores endógenos (k); es decir, $m = k$. Los coeficientes están **sobreidentificados** si el número de instrumentos supera al número de regresores endógenos; es decir, $m > k$. Están **subidentificados** si el número de instrumentos es menor que el número de regresores endógenos; es decir, $m < k$. Los coeficientes deben estar o bien exactamente identificados o bien sobreidentificados si han de ser estimados mediante la regresión VI.

El modelo general de regresión VI y su terminología se resumen en el Concepto clave 12.1.

Variables exógenas incluidas y variables de control en la regresión VI. Las variables W de la Ecuación (12.12) pueden ser variables exógenas, en cuyo caso $E(u_i|W_i) = 0$, o pueden ser variables de control que no necesitan tener una interpretación causal, sino que se incluyen para garantizar que el instrumento no esté correlacionado con el término de error. Por ejemplo, en la Sección 12.1 se planteaba la posibilidad de que el impuesto sobre las ventas pudiera estar correlacionado con la renta, la cual, según establece la teoría económica, es un factor determinante de la demanda de cigarrillos. Si es así, el impuesto sobre las ventas estaría correlacionado con el término de error en la ecuación de demanda de cigarrillos $\ln(Q_i^{\text{cigarrillos}}) = \beta_0 + \beta_1 \ln(P_i^{\text{cigarrillos}}) + u_i$, y por lo tanto no sería un instrumento exógeno. La inclusión de la renta en la regresión, o la inclusión de las variables que tienen en cuenta la renta, eliminaría esta fuente potencial de correlación entre el instrumento y el término de error. En general, si W es una variable de control efectiva en la regresión VI, entonces la inclusión de W hace que el instrumento no esté correlacionado con u , por lo que el estimador MC2E del coeficiente de X es consistente; sin embargo, si W está correlacionada con u , el coeficiente MC2E de W está sujeto a un sesgo de variable omitida y no tiene una interpretación causal. La lógica de las variables de control en la regresión VI por tanto, es paralela a la lógica de las variables de control en MCO, considerada en la Sección 7.5.

La condición matemática para que W sea una variable de control efectiva en la regresión VI es similar a la condición sobre las variables de control en MCO analizada en la Sección 7.5. En concreto, la inclusión de W debe asegurar que la media condicional de u no dependa de Z , por lo que se cumple la independencia de

**CONCEPTO
CLAVE**
12.1
El modelo general de regresión de variables instrumentales y su terminología

El modelo general de regresión VI es

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri} + u_i, \quad (12.12)$$

$i = 1, \dots, n$, donde

- Y_i es la variable dependiente;
- $\beta_0, \beta_1, \dots, \beta_{k+r}$ son coeficientes de regresión desconocidos; y
- X_{1i}, \dots, X_{ki} son k regresores endógenos, que están potencialmente correlacionados con u_i ;
- W_{1i}, \dots, W_{ri} son r regresores exógenos incluidos, que no están correlacionados con u_i o son variables de control;
- u_i es el término de error, que representa un error de medición y/o factores omitidos;
- Z_{1i}, \dots, Z_{mi} son m variables instrumentales.

Los coeficientes están sobreidentificados si existen más instrumentos que regresores endógenos ($m > k$), están subidentificados si $m < k$, y están exactamente identificados si $m = k$. La estimación del modelo de regresión VI requiere la identificación exacta o la sobreidentificación.

la media condicional; es decir, $E(u_i|Z_i, W_i) = E(u_i|W_i)$. Para mayor claridad, en la parte principal de este capítulo nos centramos en el caso en el que las variables W son exógenas por lo que $E(u_i|W_i) = 0$. En el Apéndice 12.6 se explica cómo se pueden extender los resultados de este capítulo para el caso en el que W sea una variable de control, en cuyo caso la condición de media condicional igual a cero, $E(u_i|W_i) = 0$, se sustituye por la condición de independencia en media condicional, $E(u_i|Z_i, W_i) = E(u_i|W_i)$.

MC2E en el modelo general VI

MC2E con un único regresor endógeno. Cuando existe un único regresor endógeno X y algunas variables exógenas incluidas adicionales, la ecuación de interés es

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i, \quad (12.13)$$

donde, como antes, X_i podría estar correlacionada con el término de error, pero W_{1i}, \dots, W_{ri} no lo están.

La regresión poblacional de la primera etapa de MC2E relaciona X con las variables exógenas, es decir, las W y los instrumentos (Z):

$$X_i = \pi_0 + \pi_1 Z_{1i} + \cdots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \cdots + \pi_{m+r} W_{ri} + v_i, \quad (12.14)$$

donde $\pi_0, \pi_1, \dots, \pi_{m+r}$ son los coeficientes de regresión desconocidos y v_i es un término de error.

La Ecuación (12.14) a veces se denomina ecuación de la **forma reducida** para X . Relaciona la variable endógena X con todas las variables exógenas disponibles, tanto las que se incluyen en la regresión de interés (W) como los instrumentos (Z).

En la primera etapa de MC2E, los coeficientes desconocidos de la Ecuación (12.14) se estiman por MCO, y los valores de predicción de esta regresión son $\hat{X}_1, \dots, \hat{X}_n$.

En la segunda etapa de MC2E, la Ecuación (12.13) se estima por MCO, excepto que X_i se sustituye por su valor estimado en la primera etapa. Es decir, Y_i se regresa sobre $\hat{X}_i, W_{1i}, \dots, W_{ri}$ mediante MCO. El estimador resultante de $\beta_0, \beta_1, \dots, \beta_{1+r}$ es el estimador MC2E.

Extensión a múltiples regresores endógenos. Cuando existen varios regresores endógenos X_{1i}, \dots, X_{ki} el algoritmo MC2E es similar, excepto que cada regresor endógeno requiere su propia regresión en la

primera etapa. Cada una de estas regresiones de la primera etapa tiene la misma forma que la Ecuación (12.14); es decir, la variable dependiente es una de las X , y las variables explicativas son todos los instrumentos (Z) y todas las variables exógenas incluidas (W). En conjunto, estas regresiones de la primera etapa dan lugar a valores de predicción para cada uno de los regresores endógenos.

En la segunda etapa de MC2E, la Ecuación (12.12) se estima por MCO, excepto que los regresores endógenos (X) se sustituyen por sus valores estimados respectivos (\hat{X}). El estimador resultante de $\beta_0, \beta_1, \dots, \beta_{k+r}$ es el estimador MC2E.

En la práctica, las dos etapas de MC2E se realizan de forma automática con los comandos de estimación MC2E del *software* econométrico moderno. El estimador general MC2E se resume en el Concepto clave 12.2.

CONCEPTO CLAVE

12.2

Mínimos cuadrados en dos etapas

El estimador MC2E del modelo de regresión general VI de la Ecuación (12.12) con múltiples variables instrumentales se calcula en dos etapas:

1. **Regresión(es) de la primera etapa:** regresar X_{1i} sobre las variables instrumentales (Z_{1i}, \dots, Z_{mi}) y las variables exógenas incluidas (W_{1i}, \dots, W_{ri}) por MCO, incluyendo un término independiente o intercepto. Calcular los valores estimados mediante esta regresión, denominándolos \hat{X}_{1i} . Repetir esto para todos los regresores endógenos X_{2i}, \dots, X_{ki} , calculando así los valores estimados $\hat{X}_{1i}, \dots, \hat{X}_{ki}$.
2. **Regresión(es) de la segunda etapa:** regresar por MCO Y_i sobre los valores estimados de las variables endógenas ($\hat{X}_{1i}, \dots, \hat{X}_{ki}$) y sobre las variables exógenas incluidas (W_{1i}, \dots, W_{ri}), incluyendo un intercepto. El estimador MC2E $\hat{\beta}_0^{MC2E}, \dots, \hat{\beta}_{k+r}^{MC2E}$ son los estimadores de la regresión de la segunda etapa.

En la práctica, las dos etapas se realizan automáticamente con el comando de estimación MC2E en el *software* econométrico moderno.

Relevancia y exogeneidad de los instrumentos en el modelo general VI

Las condiciones de relevancia y exogeneidad de instrumentos necesitan ser modificadas para el modelo de regresión VI general.

Cuando existe una única variable endógena incluida pero varios instrumentos, la condición para la relevancia de los instrumentos es que al menos una Z sea útil para predecir X , dado W . Cuando existen varias variables endógenas incluidas, esta condición es más complicada porque hay que descartar multicolinealidad perfecta en la regresión poblacional de la segunda etapa. Intuitivamente, cuando existen varias variables endógenas incluidas, los instrumentos deben proporcionar suficiente información sobre los movimientos exógenos de estas variables para aislar sus efectos por separado sobre Y .

La condición general del requisito de exogeneidad del instrumento es que cada instrumento debe estar incorrelacionado con el término de error u_i . Las condiciones generales para la validez de los instrumentos están dadas en el Concepto clave 12.3.

Los supuestos de la regresión VI y la distribución muestral del estimador MC2E

Bajo los supuestos de la regresión VI, el estimador MC2E es consistente y tiene una distribución muestral que, en muestras grandes, es aproximadamente normal.

Los supuestos de la regresión VI. Los supuestos de la regresión VI son modificaciones de los supuestos de mínimos cuadrados para el modelo de regresión múltiple del Concepto clave 6.4.

El primer supuesto de la regresión VI modifica el supuesto de media condicional del Concepto clave 6.4 para aplicarlo solamente a las variables exógenas incluidas. Al igual que el segundo supuesto de mínimos cuadrados para el modelo de regresión múltiple, el segundo supuesto de la regresión VI es que las extracciones

**CONCEPTO
CLAVE**
12.3**Las dos condiciones para la validez de los instrumentos**

Un conjunto de m instrumentos Z_{1i}, \dots, Z_{mi} debe cumplir las dos condiciones siguientes para ser válido:

1. Relevancia del instrumento

- En general, sea \hat{X}_{1i}^* el valor de predicción de X_{1i} a partir de la regresión poblacional de X_{1i} sobre los instrumentos (Z) y los regresores exógenos incluidos (W), y sea «1» la expresión del regresor constante que toma el valor 1 para todas las observaciones. Entonces $(\hat{X}_{1i}^*, \dots, \hat{X}_{ki}^*, W_{1i}, W_{ri}, 1)$ no son perfectamente multicolineales.
- Si solo hay una X , entonces para que se cumpla la condición anterior, al menos una Z debe tener un coeficiente distinto de cero en la regresión poblacional de X sobre las Z y las W .

2. Exogeneidad del instrumento

Los instrumentos no están correlacionados con el término de error; es decir, $\text{corr}(Z_{1i}, u_i) = 0, \dots, \text{corr}(Z_{mi}, u_i) = 0$.

nes son i.i.d., como lo son si los datos se recogen mediante un muestreo aleatorio simple. Del mismo modo, el tercer supuesto es que los valores extremos grandes son poco probables.

El cuarto supuesto de la regresión VI es que se satisfagan las dos condiciones para la validez de los instrumentos del Concepto clave 12.3. La condición de relevancia del instrumento del Concepto clave 12.3 implica el cuarto supuesto de mínimos cuadrados del Concepto clave 4.6 (ausencia de multicolinealidad perfecta) suponiendo que las variables explicativas de la regresión de la segunda etapa no son perfectamente multicolineales. Los supuestos de la regresión VI se resumen en el Concepto clave 12.4.

**CONCEPTO
CLAVE**
12.4**Los supuestos de la regresión VI**

Las variables y los errores del modelo de regresión VI del Concepto clave 12.1 satisfacen lo siguiente:

1. $E(u_i|W_{1i}, \dots, W_{ri}) = 0;$
2. $(X_{1i}, \dots, X_{ki}, W_{1i}, \dots, W_{ri}, Z_{1i}, \dots, Z_{mi}, Y_i)$ son extracciones i.i.d. de su distribución conjunta;
3. Los valores extremos elevados son poco probables: las X , W , Z , e Y tienen momentos de cuarto orden finitos y distintos de cero; y
4. Se cumplen las dos condiciones para que un instrumento sea válido del Concepto clave 12.3.

Distribución muestral del estimador MC2E. Bajo los supuestos de la regresión VI, el estimador MC2E es consistente y tiene una distribución normal en muestras grandes. Esto se muestra en la Sección 12.1 (y en el Apéndice 12.3) para el caso particular de un único regresor endógeno, un único instrumento, y sin variables exógenas incluidas. Conceptualmente, el razonamiento expuesto en la Sección 12.1 se traslada al caso general de varios instrumentos y varias variables endógenas incluidas. Sin embargo, las expresiones para el caso general son complicadas, y se trasladan al Capítulo 18.

Inferencia mediante el estimador MC2E

Debido a que la distribución muestral del estimador MC2E es normal en muestras grandes, los procedimientos generales para la inferencia estadística (contrastos de hipótesis e intervalos de confianza) de los modelos de regresión se extienden a la regresión MC2E. Por ejemplo, los intervalos de confianza del 95 % se construyen como el estimador MC2E $\pm 1,96$ errores estándar. Del mismo modo, las hipótesis conjuntas

sobre los valores de los coeficientes poblacionales se pueden contrastar mediante el estadístico F , tal como se describió en la Sección 7.2.

Cálculo de los errores estándar MC2E. Hay dos cuestiones a tener en cuenta sobre los errores estándar MC2E. En primer lugar, los errores estándar obtenidos mediante la estimación MCO de la regresión de la segunda etapa son incorrectos porque no se tiene en cuenta que es la segunda etapa de un proceso de dos etapas. En concreto, los errores estándar MCO de la segunda etapa no sirven para el ajuste de la regresión de la segunda etapa mediante los valores estimados de las variables endógenas incluidas. Las fórmulas de los errores estándar que realizan los ajustes necesarios están incorporadas en (y se utilizan automáticamente por) los comandos de la regresión MC2E del *software* econométrico. Por lo tanto, esta cuestión no supone una preocupación en la práctica si se utiliza el comando particularizado para la regresión MC2E.

En segundo lugar, como siempre, el error u podría ser heterocedástico. Por ello es importante utilizar las versiones de los errores estándar heterocedástico-robustos exactamente por la misma razón que es importante la utilización de errores estándar heterocedástico-robustos para los estimadores MCO del modelo de regresión múltiple.

Aplicación a la demanda de cigarrillos

En la Sección 12.1, se estimó la elasticidad de la demanda de cigarrillos a partir de datos sobre el consumo anual en 48 estados de los EE.UU. en 1995 mediante MC2E con un único regresor (el logaritmo del precio real por paquete) y un único instrumento (el impuesto real sobre las ventas por paquete). No obstante, la renta afecta asimismo a la demanda, por lo que forma parte del término de error de la regresión poblacional. Tal y como se discutió en la Sección 12.1, si el impuesto estatal sobre las ventas está relacionado con la renta del estado, estará correlacionado con una variable que forma parte del término de error de la ecuación de demanda de cigarrillos, lo cual viola la condición de exogeneidad del instrumento. Si es así, el estimador de VI de la Sección 12.1 es inconsistente. Es decir, la regresión VI presenta una versión del sesgo de variable omitida. Para resolver este problema, es necesario incluir la renta en la regresión.

Por tanto, se considera una especificación alternativa en la que se incluye el logaritmo de la renta en la ecuación de demanda. En la terminología del Concepto clave 12.1, la variable dependiente Y es el logaritmo del consumo, $\ln(Q_i^{\text{cigarrillos}})$; el regresor endógeno X es el logaritmo del precio real después de impuestos, $\ln(P_i^{\text{cigarrillos}})$; la variable exógena incluida W es el logaritmo de la renta real per cápita del estado, $\ln(Renta_i)$, y el instrumento Z es el impuesto real sobre ventas por paquete, ImpuestoVentas_i . Las estimaciones MC2E y los errores estándar (heterocedástico-robustos) son

$$\widehat{\ln(Q_i^{\text{cigarrillos}})} = 9,43 - 1,14 \ln(P_i^{\text{cigarrillos}}) + 0,21 \ln(Renta_i) + u_i \quad (12.15)$$

(1,26)	(0,37)	(0,31)
--------	--------	--------

Esta regresión utiliza un único instrumento ImpuestoVentas_i , pero en realidad se dispone de otros instrumentos candidatos. Además de los impuestos generales sobre ventas, los estados cobran impuestos especiales que solamente se aplican a los cigarrillos y a otros productos del tabaco. Estos impuestos específicos sobre los cigarrillos (ImpuestoCig_i) constituyen una posible segunda variable instrumental. El impuesto específico sobre los cigarrillos aumenta el precio de los cigarrillos que paga el consumidor, por lo que podría decirse que cumple con la condición para la relevancia del instrumento. Si está incorrelacionado con el término de error de la ecuación de demanda de cigarrillos del estado, es un instrumento exógeno.

Con este instrumento adicional disponible, ahora tenemos dos variables instrumentales, el impuesto real sobre las ventas por paquete y el impuesto estatal real específico para cigarrillos por paquete. Con dos instrumentos y un único regresor endógeno, la elasticidad de la demanda está sobreidentificada; es decir, el número de instrumentos (ImpuestoVentas_i e ImpuestoCig_i , por lo que $m = 2$) supera al número de variables endógenas incluidas ($P_i^{\text{cigarrillos}}$, por lo que $k = 1$). Se puede estimar la elasticidad de la demanda mediante MC2E, donde las variables explicativas de la regresión de la primera etapa son la variable exógena incluida, $\ln(Renta_i)$, y los dos instrumentos (ImpuestoVentas_i e ImpuestoCig_i).

La estimación MC2E resultante de la función de regresión utilizando los dos instrumentos $ImpuestoVentas_i$ e $ImpuestoCig_i$ es

$$\ln(Q_i^{\text{cigarrillos}}) = 9,89 - 1,28 \ln(P_i^{\text{cigarrillos}}) + 0,28 \ln(Renta_i)$$

(0,96)	(0,25)	(0,25)
--------	--------	--------

(12.16)

Comparando las Ecuaciones (12.15) y (12.16): el error estándar de la elasticidad precio estimada es menor en un tercio en la Ecuación (12.16) [0,25 en la Ecuación (12.16) frente a 0,37 en la Ecuación (12.15)]. La razón de que el error estándar sea menor en la Ecuación (12.16) es que esta estimación utiliza más información que la Ecuación (12.15): en la Ecuación (12.15), solamente se utiliza un instrumento (el impuesto sobre las ventas), pero en la Ecuación (12.16) se utilizan dos instrumentos (el impuesto sobre las ventas y el impuesto específico sobre los cigarrillos). El uso de dos instrumentos explica un mayor porcentaje de la variación de los precios de los cigarrillos que la utilización de solamente uno de ellos, y esto se refleja en menores errores estándar de las estimaciones de la elasticidad de la demanda.

¿Son creíbles estas estimaciones? En última instancia, la credibilidad depende de si el conjunto de variables instrumentales —aquí, los dos impuestos— cumplen de forma verosímil los dos requisitos para la validez de los instrumentos. Por tanto, resulta esencial evaluar si estos instrumentos son válidos, y este es el tema que ahora se aborda.

12.3 Verificación de la validez de los instrumentos

El hecho de que la regresión de variables instrumentales resulte útil en un caso concreto depende de si los instrumentos son válidos: los instrumentos no válidos dan lugar a resultados que carecen de sentido. Por lo tanto, resulta esencial evaluar si un determinado conjunto de instrumentos es válido para una aplicación particular.

Supuesto #1: relevancia de los instrumentos

El papel de la condición de relevancia de los instrumentos en la regresión VI es sutil. Una de las formas de entender la relevancia de los instrumentos es que interpreta un papel similar al tamaño de la muestra: cuanto más relevantes sean los instrumentos —es decir, cuanta más variación de la X se explique por medio de los instrumentos— más información está disponible para su uso en la regresión VI. Un instrumento más relevante da lugar a un estimador más preciso, así como un tamaño muestral más grande da lugar a un estimador más preciso. Por otra parte, la inferencia estadística mediante MC2E se basa en que el estimador MC2E tenga una distribución muestral normal, pero de acuerdo con el teorema central del límite, la distribución normal es una buena aproximación para muestras grandes, pero no necesariamente para muestras pequeñas. Si el hecho de disponer de una mayor relevancia de los instrumentos es como disponer de un tamaño de muestra mayor, esto sugiere, correctamente, que cuanto más relevante sea el instrumento, mejor es la aproximación normal para la distribución muestral del estimador MC2E y su estadístico t .

Los instrumentos que explican una pequeña proporción de la variación de X se denominan **instrumentos débiles**. En el ejemplo de los cigarrillos, la distancia del estado a las fábricas de cigarrillos sin duda sería un instrumento débil: aunque una mayor distancia aumente los costes de envío (lo cual desplaza la curva de oferta y aumenta el precio de equilibrio), los cigarrillos son ligeros, por lo que los gastos de envío constituyen una pequeña componente del precio de los cigarrillos. Por tanto, la proporción de la variación en los precios que se explica por los costes de envío, y por tanto por la distancia a las fábricas, probablemente sea muy pequeña.

En esta sección se analiza por qué los instrumentos débiles son un problema, cómo buscar instrumentos débiles, y qué hacer si se tienen instrumentos débiles. Se supone que los instrumentos son exógenos.

Por qué los instrumentos débiles son un problema. Si los instrumentos son débiles, entonces la distribución normal proporciona una aproximación pobre para la distribución muestral del estimador

MC2E, incluso si el tamaño de la muestra es grande. Por tanto, no existe justificación teórica para los métodos habituales de llevar a cabo la inferencia estadística, incluso en muestras grandes. De hecho, si los instrumentos son débiles, el estimador MC2E puede estar erróneamente sesgado en la dirección del estimador MCO. Además, los intervalos de confianza al 95 % construidos como el estimador $MC2E \pm 1,96$ errores estándar pueden contener al verdadero valor del coeficiente muchas menos veces que el 95 %. En pocas palabras, si los instrumentos son débiles, MC2E ya no resulta fiable.

Para comprobar que existe un problema con la aproximación normal para muestras grandes de la distribución muestral del estimador MC2E, considérese el caso particular, presentado en la Sección 12.1, de una sola variable endógena incluida, un único instrumento, y sin regresores exógenos incluidos. Si el instrumento es válido, entonces $\hat{\beta}_1^{MC2E}$ es consistente ya que las covarianzas muestrales s_{ZY} y s_{ZX} son consistentes; es decir, $\hat{\beta}_1^{MC2E} = s_{ZY}/s_{ZX} \xrightarrow{P} \text{cov}(Z_i, Y_i)/\text{cov}(Z_i, X_i) = \beta_1$ [Ecuación (12.7)]. Pero supongamos ahora que el instrumento no solo es débil, sino que además es irrelevante por lo que $\text{cov}(Z_i, X_i) = 0$. Entonces, $s_{ZX} \xrightarrow{P} \text{cov}(Z_i, X_i) = 0$, por lo que, tomado literalmente, el denominador de la parte derecha del límite $\text{cov}(Z_i, Y_i)/\text{cov}(Z_i, X_i)$ ¡es cero! Evidentemente, el argumento de que $\hat{\beta}_1^{MC2E}$ es consistente se quiebra cuando la condición de relevancia del instrumento no se cumple. Como se muestra en el Apéndice 12.4, esto se traduce en que el estimador MC2E tiene una distribución muestral distinta de la normal, incluso si el tamaño de la muestra es muy grande. De hecho, cuando el instrumento es irrelevante, la distribución para muestras grandes de $\hat{\beta}_1^{MC2E}$ no es la de una variable aleatoria normal, ¡sino más bien la distribución de un cociente de dos variables aleatorias normales!

Aunque esta circunstancia de instrumentos totalmente irrelevantes no puede encontrarse en la práctica, plantea una pregunta: ¿en qué medida deben ser relevantes los instrumentos para que la distribución normal sea una buena aproximación en la práctica? La respuesta a esta pregunta en el modelo VI general es complicada. Afortunadamente, sin embargo, existe una regla práctica sencilla para la situación más común en la práctica, el caso de un único regresor endógeno.

Comprobación de la debilidad de los instrumentos cuando existe un único regresor endógeno. Una forma de comprobar los instrumentos débiles cuando existe un único regresor endógeno consiste en calcular el estadístico F para el contraste de la hipótesis de que todos los coeficientes de los instrumentos son iguales a cero en la regresión de la primera etapa de MC2E. Este **estadístico F de la primera etapa** proporciona una medida del contenido de la información incluida en los instrumentos: cuanta más información contengan, mayor es el valor esperado del estadístico F . Una regla práctica sencilla es que no es necesario preocuparse de los instrumentos débiles si el estadístico F de la primera etapa es mayor que 10. (¿Por qué 10? Véase el Apéndice 12.5). Esto se resume en el Concepto clave 12.5.

¿Qué debe hacerse si se tienen instrumentos débiles? Si se tienen muchos instrumentos, probablemente algunos de esos instrumentos sean más débiles que otros. Si se tiene un número pequeño de instrumentos fuertes y muchos débiles, será mejor descartar el más débil de los instrumentos y utilizar el subconjunto de los más relevantes para el análisis MC2E. Los errores estándar MC2E podrían aumentar cuando se quitan los instrumentos débiles, pero es necesario tener en cuenta que los errores estándar originales ¡de todos modos no eran significativos!

Sin embargo, si los coeficientes están exactamente identificados, no se pueden descartar los instrumentos débiles. Aun cuando los coeficientes estén sobreidentificados, puede que no se disponga de suficientes

CONCEPTO
CLAVE

12.5

Una regla práctica para la verificación de instrumentos débiles

El estadístico F de la primera etapa es el estadístico F para contrastar la hipótesis de que los coeficientes de los instrumentos Z_{1i}, \dots, Z_{mi} son iguales a cero en la primera etapa de los mínimos cuadrados en dos etapas. Cuando existe un único regresor endógeno, un estadístico F en la primera etapa menor que 10 indica que los instrumentos son débiles, en cuyo caso el estimador MC2E es sesgado (incluso en muestras grandes) y los estadísticos t MC2E y los intervalos de confianza son poco fiables.

instrumentos fuertes para lograr la identificación, por lo que desechar algunos instrumentos débiles no ayudará. En este caso, existen dos opciones. La primera opción es encontrar instrumentos adicionales, fuertes. Esto resulta más fácil decirlo que hacerlo: se requiere una conocimiento profundo del problema en cuestión y puede implicar el rediseño del conjunto de datos y de la naturaleza del estudio empírico. La segunda opción consiste en continuar el análisis empírico con los instrumentos débiles, pero empleando métodos distintos de MC2E. Aunque este capítulo se ha centrado en MC2E algunos otros métodos de análisis de variables instrumentales son menos sensibles a los instrumentos débiles que MC2E, y algunos de estos métodos se tratan en el Apéndice 12.5.

Supuesto #2: exogeneidad de los instrumentos

Si los instrumentos no son exógenos, entonces MC2E es inconsistente: el estimador MC2E converge en probabilidad a algo distinto del coeficiente poblacional de la regresión. Después de todo, la idea de la regresión con variables instrumentales es que el instrumento contenga información sobre la variación de X_i que no esté correlacionada con el término de error u_i . Si, de hecho, el instrumento no es exógeno, no se puede identificar esta variación exógena en X_i , y es lógico pensar que la regresión VI no proporcione un estimador consistente. Las matemáticas que respaldan este argumento están resumidas en el Apéndice 12.4.

¿Puede contrastarse estadísticamente la hipótesis de que los instrumentos son exógenos? Sí y no. Por un lado, no es posible contrastar la hipótesis de que los instrumentos son exógenos cuando los coeficientes están exactamente identificados. Por otro lado, si los coeficientes están sobreidentificados, es posible contrastar la sobreidentificación de las restricciones, es decir, contrastar la hipótesis de que los instrumentos «extras» son exógenos bajo el cumplimiento del supuesto de que existen suficientes instrumentos válidos para identificar los coeficientes de interés.

Consideremos en primer lugar el caso en que los coeficientes están exactamente identificados, por lo que se dispone de muchos instrumentos como regresores endógenos. Por tanto es imposible desarrollar un contraste estadístico para la hipótesis de que los instrumentos son en realidad exógenos. Es decir, no puede utilizarse la evidencia empírica para resolver la cuestión de si estos instrumentos satisfacen el requisito de exogeneidad. En este caso, la única forma de evaluar si los instrumentos son exógenos es recurrir a una opinión experta y al conocimiento personal de los problemas empíricos que se están analizando. Por ejemplo, el conocimiento de Philip Wright sobre la oferta y la demanda agrícolas le llevó a sugerir que las lluvias por debajo de la media posiblemente desplazarían la curva de oferta de la mantequilla, pero que no desplazarían directamente la curva de demanda.

Para evaluar si los instrumentos son exógenos se requiere necesariamente un criterio técnico basado en el conocimiento personal del caso concreto. Sin embargo, si hay más instrumentos que regresores endógenos, entonces existe una herramienta estadística que puede ser útil en este proceso: el conocido como contraste de sobreidentificación de restricciones.

El contraste de sobreidentificación de restricciones. Supongamos que se dispone de un único regresor endógeno y de dos instrumentos. Entonces se podrían calcular dos estimadores MC2E diferentes: uno que utilice el primer instrumento, y el otro que utilice el segundo. Estos dos estimadores no serán iguales debido a la variación muestral, pero si ambos instrumentos son exógenos, entonces tenderán a estar cerca el uno del otro. ¿Pero qué ocurre si estos dos instrumentos dan lugar a estimaciones muy diferentes? Que razonablemente se puede concluir que hay algo malo en uno u otro instrumento, o en los dos. Es decir, sería razonable concluir que un instrumento u otro, o ambos, no son exógenos.

El **contraste de sobreidentificación de restricciones** realiza implícitamente esta comparación. Se dice implícitamente, debido a que el contraste se lleva a cabo sin tener que calcular realmente todas las posibles estimaciones VI diferentes. Esta es la idea. La exogeneidad de los instrumentos significa que no están correlacionados con u_i . Esto sugiere que los instrumentos deberían estar aproximadamente incorrelacionados con \hat{u}_i^{MC2E} , donde $\hat{u}_i^{MC2E} = Y_i - (\hat{\beta}_0^{MC2E} + \hat{\beta}_1^{MC2E}X_{1i} + \dots + \hat{\beta}_{k+r}^{MC2E}X_{ri})$ es el residuo de la regresión estimada MC2E utilizando todos los instrumentos (aproximadamente en vez de exactamente debido a la variación muestral). (Téngase en cuenta que estos residuos se construyen utilizando las verdaderas X , en lugar de sus

Una regresión terrible

Una forma de estimar el aumento porcentual en los ingresos salariales por asistir un año más a la escuela (el «rendimiento de la educación») consiste en realizar una regresión del logaritmo de los ingresos salariales sobre los años de educación a partir de los datos de los individuos. Pero si las personas más capaces presentan tanto un mayor éxito en el mercado laboral como una asistencia a la escuela más prolongada (tal vez porque les resulte más fácil), entonces los años de educación estarán correlacionados con la variable omitida, la capacidad innata, y el estimador MCO del rendimiento de la educación será sesgado. Debido a que la capacidad innata es extremadamente difícil de medir y por lo tanto no se puede utilizar como regresor, algunos economistas laborales se han pasado a la regresión VI para estimar el rendimiento de la educación. ¿Pero qué variable está correlacionada con los años de educación, pero no con el término de error en la regresión de los ingresos salariales? Es decir, ¿qué constituye una variable instrumental válida?

El cumpleaños, sugirieron los economistas laborales Joshua Angrist y Alan Krueger. Debido a las leyes de escolaridad obligatoria, pensaron, el cumpleaños está correlacionado con los años de educación: si la ley obliga a asistir a la escuela hasta cumplir 16 años y se cumplen los 16 años en enero, es posible abandonar mientras se está cursando el décimo curso, pero si se cumple el 16 de julio ya se habrá terminado el décimo curso. Si es así, el cumpleaños satisface la condición de relevancia del instrumento. Pero haber nacido en enero o julio no debe tener ningún efecto *directo* sobre los ingresos salariales (si no es a través de los años de educación), por lo que el cumpleaños satisface la condición de exogeneidad del instrumento. Llevaron a cabo esta idea mediante la utilización del trimestre de nacimiento del individuo (periodo de tres meses) como variable instrumental. Utilizaron una muestra muy grande de los datos del Censo de los EE.UU. (¡sus regresiones tenían al menos 329.000 observaciones!), y tuvieron en cuenta otras variables como la edad del trabajador.

Pero John Bound, otro economista laboral, era escéptico. Sabía que los instrumentos débiles causan que MC2E sean poco fiables y estaba preocupado porque, a pesar del gran tam-

ño de la muestra, el trimestre de nacimiento pudiera ser un instrumento débil en algunas de sus especificaciones. Por lo que cuando Bound y Krueger se encontraron durante el almuerzo, la conversación inevitablemente se dirigió a si los instrumentos de Angrist-Krueger eran débiles. Krueger no lo creía y sugirió una forma creativa de averiguarlo: ¿por qué no volver a llevar a cabo las regresiones utilizando un instrumento verdaderamente irrelevante —reemplazando el trimestre de nacimiento verdadero de cada individuo por un trimestre de nacimiento falso, generado aleatoriamente por un ordenador— y comparando los resultados con los instrumentos reales y los instrumentos falsos? Lo que encontraron fue sorprendente: no importaba si se utilizaba el trimestre de nacimiento verdadero o el falso como instrumento, ¡MC2E proporcionaba básicamente, la misma respuesta!

Esta fue una regresión terrible para los económetras laborales. El error estándar MC2E calculado utilizando los datos verdaderos sugería que el rendimiento de la educación se estimaba de manera precisa, pero del mismo modo lo hacía el error estándar calculado utilizando los datos falsos. Por supuesto, los datos falsos *no pueden* estimar el rendimiento de la educación de forma precisa, debido a que el instrumento falso es totalmente irrelevante. La preocupación entonces, es que las estimaciones MC2E basadas en los datos reales son tan poco fiables como las basadas en los datos falsos.

El problema es que los instrumentos son de hecho muy débiles en algunas de las regresiones de Angrist y Krueger. En algunas de sus especificaciones, el estadístico *F* de la primera etapa es menor que 2, mucho menos que el límite de 10 de la regla práctica. En otras especificaciones, Angrist y Krueger obtenían un estadístico *F* de la primera etapa mayor, y en esos casos la inferencia MC2E no estaba sujeta al problema de la debilidad de los instrumentos. Por cierto, en estas especificaciones el rendimiento de la educación se estima en aproximadamente un 8 %, algo *mayor* que lo estimado por MCO¹.

¹ Las regresiones VI originales se presentaron en Angrist y Krueger (1991), y el re-análisis utilizando los instrumentos falsos se publicó en Bound, Jaeger y Baker (1995).

valores estimados en la primera etapa). En consecuencia, si los instrumentos son en realidad exógenos, entonces los coeficientes de los instrumentos en una regresión de \hat{u}_i^{MC2E} sobre los instrumentos y las variables exógenas incluidas deberían ser todos ellos iguales a cero, y esta hipótesis se puede contrastar.

Este método para calcular el contraste de sobreidentificación de restricciones se resume en el Concepto clave 12.6. Este estadístico se calcula utilizando el estadístico *F* válido con homocedasticidad. El contraste estadístico se denomina comúnmente estadístico *J*.

En muestras grandes, si los instrumentos no son débiles y los errores son homocedásticos, entonces, bajo la hipótesis nula de que los instrumentos son exógenos, el estadístico *J* presenta una distribución chi-cuadrado con $m - k$ grados de libertad (χ^2_{m-k}). Es importante recordar que aunque el número de restricciones que se contrastan sea m , los grados de libertad de la distribución asintótica del estadístico *J* son $m - k$. La razón

**CONCEPTO
CLAVE**

12.6

El contraste de sobreidentificación de restricciones (el estadístico J)

Sea \hat{u}_i^{MC2E} el residuo de la estimación MC2E de la Ecuación (12.12). Se utiliza MCO para la estimación de los coeficientes de regresión en

$$\hat{u}_i^{MC2E} = \delta_0 + \delta_1 Z_{1i} + \dots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \dots + \delta_{m+r} W_{ri} + e_i, \quad (12.17)$$

donde e_i es el término de error de la regresión. Sea F la expresión del estadístico F válido con homocedasticidad para el contraste de la hipótesis de que $\delta_1 = \dots = \delta_m = 0$. El estadístico para el contraste de sobreidentificación de restricciones es $J = mF$. Bajo la hipótesis nula de que todos los instrumentos son exógenos, si e_i es homocedástico, en muestras grandes J se distribuye χ^2_{m-k} , donde $m - k$ es el «grado de sobreidentificación», es decir, el número de instrumentos menos el número de regresores endógenos.

es que solo es posible contrastar las restricciones sobreidentificadas, de las que hay $m - K$. La modificación del estadístico J para errores heterocedásticos, se ofrece en la Sección 18.7.

La forma más fácil de comprobar que no se puede contrastar la exogeneidad de los regresores cuando los coeficientes están exactamente identificados ($m = k$) es considerar el caso de una sola variable endógena incluida ($k = 1$). Si hay dos instrumentos, entonces se pueden calcular dos estimadores MC2E, uno por cada instrumento, que se pueden comparar para comprobar si están próximos. Pero si se dispone solamente de un instrumento, entonces se puede calcular un solo estimador MC2E y no se dispone de otro con el que compararlo. De hecho, si los coeficientes están exactamente identificados, por lo que $m = k$, entonces el estadístico J de contraste de sobreidentificación es exactamente igual a cero.

12.4 Aplicación a la demanda de cigarrillos¹

Nuestro intento de estimar la elasticidad de la demanda de cigarrillos se quedó en las estimaciones MC2E recogidas por la Ecuación (12.16), en las que la renta era una variable exógena incluida y había dos instrumentos, el impuesto general sobre ventas y el impuesto específico sobre los cigarrillos. Ahora puede llevarse a cabo una evaluación más completa de estos instrumentos.

Como en la Sección 12.1, tiene sentido que los dos instrumentos sean relevantes debido a que los impuestos constituyen una parte importante de los precios después de impuestos de los cigarrillos, lo que en breve se comprobará empíricamente. En primer lugar, no obstante, nos centramos en la difícil cuestión de si es razonable que las dos variables fiscales sean exógenas.

El primer paso para establecer si un instrumento es exógeno, consiste en pensar en los argumentos acerca de por qué puede serlo o no. Esto requiere plantearse cuáles son los factores que forman parte del término de error en la ecuación de la demanda de cigarrillos y si es razonable pensar que estos factores pueden estar relacionados con los instrumentos.

¿Por qué algunos estados tienen mayor consumo de cigarrillos per cápita que otros? Una de las razones podría ser la distinta renta entre los distintos estados, pero la renta de cada estado está incluida en la Ecuación (12.16), por lo que no forma parte del término de error. Otra de las razones es que existan factores históricos que influyen en la demanda. Por ejemplo, los estados que cultivan tabaco tienen mayores índices de tabaquismo que la mayoría de los otros estados. ¿Podría este factor estar relacionado con los impuestos? Es muy posible: si el cultivo de tabaco y la producción de cigarrillos son industrias importantes en un estado, entonces estas industrias podrían ejercer su influencia para mantener bajos los impuestos específicos sobre los cigarrillos. Esto sugiere que un factor omitido de la demanda de cigarrillos —si en el estado se cultiva tabaco y se producen cigarrillos— podría estar correlacionado con los impuestos específicos sobre los cigarrillos.

¹ Este apartado supone el conocimiento de la materia de las Secciones 10.1 y 10.2 sobre datos de panel con $T=2$ períodos de tiempo.

Las externalidades del consumo de tabaco

Fumar conlleva costes que no son soportados en su totalidad por el fumador; es decir, genera externalidades. Por tanto, una justificación económica para gravar los cigarrillos es la de «internalizar» esas externalidades. En teoría, el impuesto sobre un paquete de cigarrillos debería ser igual al valor en dólares de las externalidades creadas por un paquete de tabaco. ¿Pero a cuánto ascienden exactamente las externalidades de fumar, medidas en dólares por paquete?

Algunos estudios han utilizado métodos econométricos para estimar las externalidades del consumo de tabaco. Las externalidades negativas, los costes, soportados por otros incluyen los gastos médicos pagados por el estado para atender a los fumadores enfermos, los costes sanitarios de los no fumadores asociados con el humo de segunda mano, y los incendios causados por cigarrillos.

Pero, desde un punto de vista puramente económico, el tabaquismo tiene asimismo externalidades positivas o beneficios. El beneficio económico más importante del hábito de fumar consiste en que los fumadores suelen pagar mucho más a la Seguridad Social (sistemas públicos de pensiones) en forma de impuestos de lo que alguna vez utilizarán. Existe asimismo un gran ahorro en los gastos de cuidados en el hogar para los ancianos, los fumadores tienden a no vivir tanto tiempo. Debido a que las externalidades negativas del consumo de tabaco se producen mientras el fumador está vivo, pero las positivas se acumulan tras la muerte, el valor actual neto de las externali-

lidades por paquete (el valor de los costes netos por paquete, descontados hasta la actualidad) depende de la tasa de descuento.

Los estudios no se ponen de acuerdo en un valor en dólares determinado acerca de las externalidades netas. Algunos sugieren que las externalidades netas, debidamente descontadas, son bastante pequeñas, menos que los impuestos en el presente. De hecho, las estimaciones más extremas sugieren que los factores externos netos son *positivos*, por lo que ¡fumar debería estar subvencionado! Otros estudios, que incorporan costes que probablemente son importantes pero a la vez son difíciles de cuantificar (como el cuidado de bebés que no son sanos debido a que sus madres fuman), sugieren que las externalidades podrían ser de 1 \$ por paquete, posiblemente incluso más. Pero todos los estudios coinciden en que, dado que suelen morir en la edad madura, los fumadores pagan muchos más impuestos de lo que alguna vez recuperaran durante su breve jubilación¹.

¹ Willard G. Manning *et al.* (1989) proporcionaron un cálculo inicial de las externalidades del consumo de tabaco. Un cálculo que sugería que los costes del cuidado de salud aumentarían si todo el mundo dejara de fumar se presenta en el trabajo de Barendregt *et al.* (1997). Otros estudios sobre las externalidades del consumo de tabaco fueron recopilados y revisados por Chaloupka y Warner (2000).

Una solución a esta posible correlación entre el término de error y el instrumento podría ser la de incluir información sobre el tamaño de la industria del tabaco y los cigarrillos en el estado; este es el método que se adoptó cuando se incluía la renta como regresor en la ecuación de la demanda. Pero debido a que se dispone de datos de panel sobre el consumo de cigarrillos, existe otro método que no requiere esta información. Tal y como se analizó en el Capítulo 10, los datos de panel hacen que sea posible eliminar la influencia de variables que cambian entre las distintas entidades individuales (estados), pero que no cambian a lo largo del tiempo, tales como el clima o las circunstancias históricas que conducen a una gran industria de tabaco y cigarrillos dentro de un estado. Se presentaron dos métodos para llevar esto a cabo en el Capítulo 10: construir datos sobre los *cambios* en las variables entre dos períodos de tiempo diferentes y utilizar la regresión de efectos fijos. Con el fin de mantener el análisis lo más sencillo posible en este punto, se adopta el método anterior y se llevan a cabo las regresiones del tipo descrito en la Sección 10.2, en base a los cambios en las variables entre dos años diferentes.

El lapso de tiempo entre los dos años diferentes influye sobre el modo en el que se deben interpretar las elasticidades estimadas. Debido a que los cigarrillos son adictivos, las variaciones en el precio requieren de algún tiempo para lograr modificar la conducta. En un primer momento, un aumento en el precio de los cigarrillos podría tener un efecto escaso en la demanda. Con el tiempo, sin embargo, el aumento de los precios podría contribuir al deseo de algunos de los fumadores de dejar de fumar, y, más importante, podría disuadir a los no fumadores de adquirir el hábito. Por tanto, la respuesta de la demanda al incremento de los precios podría ser pequeña en el corto plazo, pero grande en el largo plazo. Dicho de otra manera, para un producto adictivo como el tabaco, la demanda podría ser inelástica en el corto plazo, es decir, que podría tener una elasticidad de corto plazo cercana a cero, pero podría ser más elástica en el largo plazo.

En este análisis, nos centramos en la estimación de la elasticidad precio de largo plazo. Esto se lleva a cabo considerando cambios en la cantidad y en el precio que se producen a lo largo de períodos de 10 años. En concreto, en las regresiones que se consideran aquí, la variación en 10 años en el logaritmo de la cantidad, $\ln(Q_{i,1995}^{\text{cigarrillos}}) - \ln(Q_{i,1985}^{\text{cigarrillos}})$, se regresa sobre la variación en 10 años en el logaritmo del precio $\ln(P_{i,1995}^{\text{cigarrillos}}) - \ln(P_{i,1985}^{\text{cigarrillos}})$ y la variación en 10 años en el logaritmo de la renta, $\ln(Renta_{i,1995}) - \ln(Renta_{i,1985})$. Se utilizan dos instrumentos: la variación en el impuesto sobre las ventas en 10 años, $ImpuestoVentas_{i,1995} - ImpuestoVentas_{i,1985}$, y la variación en el impuesto específico sobre los cigarrillos en 10 años, $ImpuestoCig_{i,1995} - ImpuestoCig_{i,1985}$.

Los resultados se presentan en la Tabla 12.1. Como de costumbre, cada columna de la tabla presenta los resultados de una regresión diferente. Todas las regresiones tienen los mismos regresores, y todos los coeficientes se calculan utilizando MC2E; la única diferencia entre las tres regresiones es el conjunto de instrumentos utilizados. En la columna (1), el único instrumento es el impuesto sobre ventas; en la columna (2), el único instrumento son los impuestos específicos sobre los cigarrillos; y en la columna (3), ambos impuestos son utilizados como instrumentos.

En la regresión VI, la fiabilidad de los coeficientes estimados gira en torno a la validez de los instrumentos, por lo que lo primero que hay que mirar de la Tabla 12.1 son los estadísticos de diagnóstico para evaluar la validez de los instrumentos.

En primer lugar, ¿son relevantes los instrumentos? Tenemos que mirar los estadísticos F de la primera etapa. La regresión de la primera etapa en la columna (1) es

$$\begin{aligned} \widehat{\ln(P_{i,1995}^{\text{cigarrillos}}) - \ln(P_{i,1985}^{\text{cigarrillos}})} &= 0,53 - 0,22[\ln(Renta_{i,1995}) - \ln(Renta_{i,1985})] \\ &\quad (0,03) \quad (0,22) \\ &+ 0,0255(ImpuestoVentas_{i,1995} - ImpuestoVentas_{i,1985}) \\ &\quad (0,0044) \end{aligned} \quad (12.18)$$

TABLA 12.1 Estimaciones de mínimos cuadrados en dos etapas de la demanda de cigarrillos utilizando datos de panel para 48 estados de EE.UU

Regresor	(1)	(2)	(3)
$\ln(P_{i,1995}^{\text{cigarrillos}}) - \ln(P_{i,1985}^{\text{cigarrillos}})$	-0,94** (0,21)	-1,34** (0,23)	-1,20** (0,20)
$\ln(Renta_{i,1995}) - \ln(Renta_{i,1985})$	0,53 (0,34)	0,43 (0,30)	-0,05 (0,06)
Intercepto	-0,12 (0,07)	-0,02 (0,07)	-0,05 (0,06)
Variable(s) instrumental(es)	Impuesto ventas	Impuesto específico cigarrillos	Impuesto ventas e Impuesto específico cigarrillos
Estadístico F de la primera etapa	33,70	107,20	889,60
Contraste J de sobreidentificación de restricciones y p -valor	—	—	4,93 (0,026)

Estas regresiones se estimaron utilizando datos sobre 48 estados de EE.UU. (48 observaciones sobre 10 años distintos). Los datos se describen en el Apéndice 12.1. El contraste J de sobreidentificación de restricciones se describe en el Concepto clave 12.6 (su p -valor se expresa entre paréntesis), el estadístico F de la primera etapa se describe en el Concepto clave 12.5. Los coeficientes individuales son estadísticamente significativos al nivel designación del 5 % * o del 1 % **.

Debido a que solamente hay un instrumento en esta regresión, el estadístico F de la primera etapa es el cuadrado del estadístico t para el contraste de que el coeficiente de la variable instrumental, $ImpuestoVentas_{i,1995} - ImpuestoVentas_{i,1985}$, es igual a cero; es decir $F = t^2 = (0,0255/0,0044)^2 = 33,7$. Para las regresiones de las columnas (2) y (3), los estadísticos F de la primera etapa son 107,2 y 88,6, por lo que en los tres casos los estadísticos F de la primera etapa son mayores que 10. Se concluye que los instrumentos no son débiles, por lo que se puede confiar en los métodos estándar para realizar la inferencia estadística (contrastos de hipótesis, intervalos de confianza), utilizando los coeficientes y los errores estándar MC2E.

En segundo lugar, ¿son exógenos los instrumentos? Debido a que las regresiones de las columnas (1) y (2) tienen cada una un único instrumento y un único regresor endógeno incluido, los coeficientes de estas regresiones están exactamente identificados. Por lo tanto no se puede realizar el contraste J en ninguna de estas regresiones. La regresión de la columna (3), no obstante, está sobreidentificada porque tiene dos instrumentos y un solo regresor endógeno incluido, por lo que existe una ($m - k = 2 - 1 = 1$) restricción sobreidentificada. El estadístico J es 4,93: tiene una distribución χ^2_1 , por lo que el valor crítico al 5 % es 3,84 (Tabla 3 del Apéndice) y la hipótesis nula de que los dos instrumentos son exógenos se rechaza al nivel de significación del 5 % (esta deducción puede asimismo hacerse directamente mediante el p -valor de 0,026 que aparece en la tabla).

La razón por la que el estadístico J rechaza la hipótesis nula de que ambos instrumentos son exógenos es que los dos instrumentos proporcionan coeficientes estimados muy diferentes. Si el único instrumento es el impuesto sobre las ventas [columna (1)], la elasticidad precio estimada es $-0,94$, pero cuando el único instrumento es el impuesto específico sobre los cigarrillos, la elasticidad precio estimada es $-1,34$. Recorremos la idea básica del estadístico J : si ambos instrumentos son exógenos, entonces los dos estimadores MC2E que utilizan los instrumentos individuales son consistentes y se diferencian entre sí solamente debido a la variación de la muestra aleatoria. Sin embargo, si uno de los instrumentos es exógeno y el otro no lo es, entonces el estimador basado en el instrumento endógeno es inconsistente, lo cual se detecta por el estadístico J . En esta aplicación, la diferencia entre las dos elasticidades precio estimadas es lo suficientemente grande como para que sea poco probable que sea el resultado de la pura variación muestral, por lo que el estadístico J rechaza la hipótesis nula de que ambos instrumentos son exógenos.

El rechazo del estadístico J significa que la regresión de la columna (3) se basa en instrumentos no válidos (la condición de exogeneidad del instrumento no se cumple). ¿Qué implica esto sobre las estimaciones de las columnas (1) y (2)? El rechazo del estadístico J dice que al menos uno de los instrumentos es endógeno, por lo que existen tres posibilidades lógicas: el impuesto sobre ventas es exógeno, pero el impuesto específico sobre los cigarrillos no lo es, en cuyo caso la regresión de la columna (1) es fiable; el impuesto específico sobre los cigarrillos es exógeno, pero el impuesto sobre ventas no lo es, por lo que la regresión de la columna (2) es fiable; o bien ninguno de los impuestos es una variable exógena, por lo que ninguna regresión es fiable. La evidencia estadística no puede decir cuál de las posibilidades es la correcta, por lo que debe utilizarse una opinión y un juicio propios.

Creemos que el caso de la exogeneidad del impuesto general sobre ventas es más fuerte que el de la del impuesto específico sobre los cigarrillos, debido a que el proceso político puede vincular los cambios en el impuesto específico sobre los cigarrillos a los cambios en las políticas que afectan al mercado de cigarrillos y el tabaco. Por ejemplo, si el consumo de tabaco disminuye en un estado, debido a que pasa de moda, habrá menos fumadores y un grupo de presión debilitado contra el incremento de los impuestos específicos sobre los cigarrillos, lo que a su vez podría conllevar unos mayores impuestos específicos sobre los cigarrillos. Por tanto los cambios en los gustos (que forman parte de la u) podrían estar relacionados con los cambios en los impuestos específicos sobre los cigarrillos (el instrumento). Esto sugiere descartar las estimaciones VI que utilizan como instrumento el impuesto específico sobre los cigarrillos, y la adopción de la elasticidad precio estimada utilizando como instrumento el impuesto general sobre ventas, $-0,94$.

La estimación de $-0,94$ indica que el consumo de cigarrillos es bastante elástico: un aumento en el precio del 1 % conduce a una disminución en el consumo del 0,94 %. Esto puede parecer sorprendente para un producto adictivo como son los cigarrillos. Pero recuérdese que esta elasticidad se calcula utilizando las variaciones a lo largo de un periodo de diez años, por lo que se trata de una elasticidad de largo plazo. Esta

estimación sugiere que el aumento de impuestos puede hacer una mella considerable en el consumo de cigarrillos, al menos en el largo plazo.

Cuando la elasticidad se calcula utilizando las variaciones en 5 años desde 1985 hasta 1990 en vez de las variaciones a lo largo de 10 años de la Tabla 12.1, la elasticidad (calculada con el impuesto general sobre ventas como instrumento) es $-0,79$; para las variaciones desde 1990 hasta 1995, la elasticidad es de $-0,68$. Estas estimaciones sugieren que la demanda es menos elástica para horizontes de 5 años que para los de 10 años. Este hallazgo de una mayor elasticidad precio a un horizonte más largo es coherente con la gran cantidad de investigaciones realizadas sobre la demanda de cigarrillos. Las estimaciones de la elasticidad de la demanda en esta literatura suelen estar en un rango entre $-0,3$ y $-0,5$, pero se trata principalmente de elasticidades a corto plazo; algunos estudios sugieren que la elasticidad a largo plazo podría ser tal vez el doble de la elasticidad a corto plazo².

12.5 ¿De dónde provienen los instrumentos válidos?

En la práctica, el aspecto más difícil de la estimación VI es el de encontrar instrumentos que sean relevantes y exógenos. Existen dos métodos principales, que reflejan dos perspectivas diferentes sobre la modelización en Econometría y Estadística.

El primer método consiste en el uso de la teoría económica para sugerir instrumentos. Por ejemplo, el conocimiento de Philip Wright sobre la economía de los mercados agrícolas le llevó en busca de un instrumento que desplazara la curva de oferta pero no la curva de demanda; lo que a su vez le llevó a considerar las condiciones climáticas de las regiones agrícolas. Una de las áreas en la que este enfoque ha sido particularmente exitoso es el campo de la economía financiera. Algunos de los modelos económicos acerca del comportamiento de los inversores incluyen consideraciones acerca de cómo pronostican los inversores, lo que a su vez incluye conjuntos de variables que están incorrelacionadas con el término de error. Estos modelos a veces son no lineales en los datos y en los parámetros, en cuyo caso los estimadores VI estudiados en este capítulo no pueden ser utilizados. Se utiliza en su lugar una extensión de los métodos de VI a los modelos no lineales, denominada estimación del método generalizado de momentos. No obstante, las teorías económicas son abstracciones que a menudo no tienen en cuenta los matices y los detalles necesarios para el análisis de una base de datos en particular. Por lo que este método no siempre funciona.

El segundo método para construir instrumentos consiste en la búsqueda de alguna fuente exógena de variación en X que surja de lo que es, en efecto, un fenómeno aleatorio que induce cambios en el regresor endógeno. Por ejemplo, en nuestro ejemplo hipotético de la Sección 12.1, los daños de un terremoto aumentan el promedio del tamaño de las clases en algunos distritos escolares, y esta variación en el tamaño promedio de las clases no estaba correlacionada con las potenciales variables que afectan al rendimiento estudiantil. Este método habitualmente requiere el conocimiento del problema que está siendo estudiado y una cuidadosa atención a los detalles de los datos, lo cual se explica mejor a través de ejemplos.

Tres ejemplos

Pasamos ahora a presentar tres aplicaciones empíricas de la regresión VI que proporcionan ejemplos de cómo los diferentes investigadores utilizaron su conocimiento experto acerca de sus problemas empíricos para encontrar variables instrumentales.

¿Meter en la cárcel a los criminales reduce el crimen? Esta es una pregunta que solo formularía un economista. Después de todo, un criminal no puede cometer un delito fuera de la cárcel mientras se encuentra en prisión, y el hecho de que algunos delincuentes sean capturados y encarcelados sirve para

² Un aleccionador estudio económico llevado a cabo por Adda y Cornaglia (2006) sugiere que los fumadores compensan los mayores impuestos fumando con mayor intensidad, extrayendo por tanto más nicotina de cada cigarrillo. En caso de estar interesado en conocer más sobre la economía del consumo de tabaco, consulte Chaloupka y Warner (2000), Gruber (2001), y Carpenter y Cook (2008).

disuadir a otros. Pero la magnitud del efecto combinado, la variación en la tasa de criminalidad asociada con un aumento del 1 % en la población reclusa, es una cuestión empírica.

Una estrategia para la estimación de este efecto es la de realizar una regresión de las tasas de criminalidad (delitos por cada 100.000 miembros de la población general) sobre las tasas de encarcelación (los presos por cada 100.000 habitantes), utilizando datos anuales en un nivel adecuado de jurisdicción (por ejemplo, los estados de EE.UU.). Esta regresión podría incluir algunas variables de control que midan las condiciones económicas (la delincuencia aumenta cuando empeoran las condiciones económicas generales), la demografía (los jóvenes cometen más delitos que los adultos), etc. Sin embargo, existe un serio potencial de sesgo por causalidad simultánea que socava tal análisis: si la tasa de criminalidad sube y la policía hace su trabajo, habrá más presos. Por un lado, el aumento de la encarcelación reduce la tasa de criminalidad; por otro lado, una tasa de criminalidad creciente aumenta la encarcelación. Como en el ejemplo de la mantequilla de la Figura 12.1, debido a esta causalidad simultánea una regresión MCO de la tasa de criminalidad sobre la tasa de encarcelación estimará una combinación complicada de estos dos efectos. Este problema no puede resolverse mediante la búsqueda de mejores variables de control.

No obstante, este sesgo de causalidad simultánea puede ser eliminado encontrando una variable instrumental adecuada y utilizando MC2E. El instrumento debe estar correlacionado con la tasa de encarcelación (debe ser relevante), pero del mismo modo debe estar incorrelacionado con el término de error de la ecuación de interés de la tasa de criminalidad (debe ser exógeno). Es decir, debe afectar a la tasa de encarcelación, pero debe estar incorrelacionado con cualquiera de los factores no observables que determinan la tasa de criminalidad.

¿Dónde se puede encontrar algo que afecte a la encarcelación, pero no tenga efecto directo sobre la tasa de criminalidad? Un sitio es la variación exógena en la capacidad de las prisiones en funcionamiento en la actualidad. Debido a que se necesita tiempo para construir una cárcel, las restricciones de capacidad a corto plazo pueden obligar a los estados a liberar a los prisioneros antes de tiempo o de lo contrario reducir las tasas de encarcelación. Utilizando este razonamiento, Levitt (1996) sugería que el número de demandas dirigidas a reducir el hacinamiento penitenciario podría servir como una variable instrumental, y puso en práctica esta idea con datos de panel para los estados de EE.UU. desde 1972 hasta 1993.

¿Son instrumentos válidos las variables que miden los pleitos por hacinamiento? Aunque Levitt no expuso los estadísticos F de la primera etapa, los litigios por hacinamiento en las cárceles desaceleraron el aumento de los encarcelamientos de prisioneros según sus datos, lo que sugiere que este instrumento es relevante. En la medida en que los litigios por hacinamiento son inducidos por las condiciones carcelarias pero no por el tipo de delito o por sus factores determinantes, este instrumento es exógeno. Debido a que Levitt desglosa los litigios por hacinamiento en varios tipos y por lo tanto dispone de varios instrumentos, es capaz de contrastar las restricciones sobreidentificadas y no las rechaza con el estadístico J , lo que refuerza el hecho de que sus instrumentos son válidos.

Mediante estos instrumentos y MC2E, Levitt estimó que el efecto sobre la tasa de criminalidad de la encarcelación es sustancial. Este efecto estimado era tres veces mayor que el efecto estimado por MCO, lo que sugiere que MCO presentaba un gran sesgo por causalidad simultánea.

¿La reducción del tamaño de las clases incrementa las calificaciones en las pruebas? Como se vio en el análisis empírico de la Parte II, las escuelas con clases pequeñas suelen ser más ricas, y sus estudiantes tienen acceso a mejores oportunidades de aprendizaje tanto dentro como fuera del aula. En la Parte II, se utilizaba la regresión múltiple para hacer frente a la amenaza del sesgo de variable omitida teniendo en cuenta distintas medidas de las circunstancias económicas de los estudiantes, la capacidad de hablar Inglés, etc. Además, un escéptico podría preguntarse si se hizo lo suficiente: si se dejó de lado algo importante, las estimaciones del efecto de tamaño de las clases todavía estarían sesgadas.

Este sesgo potencial de variables omitidas podría resolverse mediante la inclusión de las variables de control adecuadas, pero si estos datos no estuvieran disponibles (algunos, como las oportunidades de aprendizaje fuera del aula, resultan difíciles de medir), entonces, un método alternativo consiste en utilizar la regresión de VI. Esta regresión requiere una variable instrumental que esté correlacionada con el tamaño de las clases (relevancia), pero que no esté correlacionada con los factores determinantes omitidos del desem-

peño en la prueba que forman parte del término de error, tales como el interés de los padres en el aprendizaje, las oportunidades de aprendizaje fuera del aula, la calidad de los profesores y las instalaciones escolares, etc. (exogeneidad).

¿Dónde se puede encontrar un instrumento que induzca variaciones aleatorias y exógenas en el tamaño de las clases, pero que no esté correlacionado con los otros factores determinantes del rendimiento en el examen? Hoxby (2000) sugería la biología. Debido a las fluctuaciones aleatorias en las fechas de nacimiento, el tamaño de las clases de jardín de infancia varía de un año a otro. Aunque el número real de niños que ingresan en el jardín de infancia puede ser endógeno (las noticias recientes acerca de la escuela pueden influir en si los padres envían a sus hijos a una escuela privada), sostenía que el número *potencial* de niños que ingresan en el jardín de infancia, el número de niños de 4 años de edad en el distrito, se trata principalmente de una cuestión de fluctuaciones aleatorias en las fechas de nacimiento de los niños.

¿Es la matrícula potencial un instrumento válido? El hecho de que sea exógena depende de si está correlacionada con los factores determinantes no observables del rendimiento en el examen. Sin duda, las fluctuaciones biológicas en la matrícula potencial son exógenas, pero las inscripciones potenciales podrían fluctuar asimismo debido a que los padres con niños pequeños elijan mudarse hacia un distrito escolar que está mejorando desde uno que tiene problemas. Si es así, un aumento en la matrícula potencial podría estar correlacionado con factores no observables, tales como la calidad de la gestión escolar, haciendo que este instrumento no sea válido. Hoxby abordó este problema mediante el razonamiento de que el crecimiento o la disminución en el número potencial de estudiantes que se debe a este motivo se llevaría a cabo de forma suave a lo largo de varios años, mientras que las fluctuaciones aleatorias en las fechas de nacimiento se producen a corto plazo con «picos» en la matrícula potencial. Por tanto, utilizó como instrumento no la matrícula potencial, sino la desviación de la matrícula potencial respecto de su tendencia a largo plazo. Estas desviaciones satisfacen el criterio de relevancia del instrumento (todos los estadísticos *F* de la primera etapa superan el 100). Ella hace una buena defensa de que este instrumento es exógeno, pero, como en todos los análisis VI, la credibilidad de esta suposición es, en última instancia, una cuestión de criterio.

Hoxby llevó a la práctica esta estrategia con los datos de panel detallados sobre las escuelas de primaria de Connecticut en la década de 1980 y 1990. El conjunto de datos de panel le permitió incluir los efectos fijos individuales de la escuela, lo cual, sumado a la estrategia de variables instrumentales, atacaba el problema del sesgo de variables omitidas a nivel de la escuela. Sus estimaciones MC2E sugieren que el efecto sobre las calificaciones en la prueba del tamaño de las clases es pequeño; la mayoría de sus estimaciones fueron estadísticamente no significativamente distintas de cero.

¿El tratamiento agresivo para los ataques al corazón prolonga la vida? Los tratamientos agresivos para las víctimas de ataques al corazón (técnicamente, los infartos de miocardio agudos, o IMA) tienen el potencial de salvar vidas. Antes de que un nuevo procedimiento médico, en este ejemplo, el cateterismo cardiaco³, sea aprobado para uso generalizado, pasa por ensayos clínicos, una serie de experimentos aleatorizados controlados diseñados para medir sus efectos y sus efectos secundarios. Sin embargo, unos buenos resultados en un ensayo clínico son una cosa; y el rendimiento en el mundo real es otra.

Un punto de partida natural para la estimación del efecto real de un cateterismo cardiaco consiste en comparar a los pacientes que recibieron el tratamiento con los que no lo recibieron. Esto conduce a la regresión de la duración de la supervivencia de los pacientes sobre la variable binaria del tratamiento (si el paciente recibió cateterismo cardiaco) y sobre otras variables de control que influyen en la mortalidad (edad, peso, otras medidas de las condiciones de salud, etc.) El coeficiente poblacional de la variable indicador es el incremento de la esperanza de vida del paciente proporcionada por el tratamiento. Desafortunadamente, el estimador de MCO está sujeto a un sesgo: el cateterismo cardiaco no se aplica «por casualidad» a un paciente al azar; sino que se realiza debido a que el médico y el paciente deciden que podría resultar eficaz. Si sus decisiones se basan en parte sobre factores no observables relevantes para los resultados en la salud fuera del conjunto de datos, la decisión de tratamiento estará correlacionada con el término de error de la

³ El cateterismo cardiaco es un procedimiento por el cual se inserta un catéter, o tubo, en un vaso sanguíneo y se guía a lo largo de todo el camino hasta el corazón para obtener información sobre el corazón y las arterias coronarias.

regresión. Si los pacientes más sanos son los únicos que reciben el tratamiento, el estimador MCO estará sesgado (el tratamiento está correlacionado con una variable omitida), y el tratamiento parecerá ser más efectivo de lo que realmente es.

Este sesgo potencial puede ser eliminado mediante la regresión VI con una variable instrumental válida. El instrumento debe estar correlacionado con el tratamiento (debe ser relevante) pero debe estar incorrelacionado con los factores de salud omitidos que afectan a la supervivencia (debe ser exógeno).

¿Dónde se puede encontrar algo que afecte al tratamiento, pero no al resultado de salud, que no sea a través de su efecto sobre el tratamiento? McClellan, McNeil y Newhouse (1994) sugirieron la geografía. La mayoría de los hospitales de su base de datos no estaban especializados en cirugía de cateterismo cardiaco, por lo que muchos pacientes estaban más cerca de hospitales «normales» que no ofrecían este tratamiento que de los hospitales que sí ofrecían cateterismo cardiaco. McClellan, McNeil, y Newhouse utilizaron por tanto como variable instrumental la diferencia entre la distancia desde el domicilio del paciente con IMA al hospital más cercano que disponía de técnicas de cirugía de cateterismo cardiaco y la distancia al hospital más cercano de cualquier tipo, esta distancia es cero si el hospital más cercano es un hospital con técnica de cateterismo cardiaco, y de lo contrario, es positiva. Si esta distancia relativa afecta a la probabilidad de recibir este tratamiento, entonces es relevante. Si está distribuida de forma aleatoria entre las víctimas de IMA, entonces es exógena.

¿Es la distancia relativa al hospital más cercano con técnica de cateterismo cardiaco un instrumento válido? McClellan, McNeil y Newhouse no presentan los estadísticos F de la primera etapa, pero sí proporcionan evidencia empírica de que no es débil. ¿Es esta medida de la distancia exógena? Ellos formulan dos argumentos. En primer lugar, recurren a su experiencia médica y al conocimiento del sistema de atención de salud para argumentar que la distancia a un hospital estará verosímilmente incorrelacionada con cualquiera de las variables no observables que determinan los resultados del IMA. En segundo lugar, disponen de datos sobre algunas de las variables adicionales que afectan los resultados del IMA, tales como el peso del paciente, y en su muestra la distancia no está correlacionada con estos determinantes *observables* de la supervivencia; lo que, argumentan, hace que sea más creíble del mismo modo que la distancia no esté correlacionada con los determinantes inobservables del término de error.

Utilizando 205.021 observaciones sobre estadounidenses con al menos 64 años de edad que padecieron un IMA en 1987, McClellan, McNeil y Newhouse llegaron a una conclusión sorprendente: sus estimaciones MC2E sugieren que el cateterismo cardíaco tiene un pequeño, posiblemente igual a cero, efecto sobre los resultados en la salud; es decir, el cateterismo cardiaco no prolonga sustancialmente la vida. Por el contrario, las estimaciones MCO, sugieren un efecto positivo grande. Ellos interpretan esta diferencia como una evidencia del sesgo en las estimaciones de MCO.

El método de VI de McClellan, McNeil, y Newhouse tiene una interpretación interesante. El análisis MCO utiliza como regresor el tratamiento real, pero debido a que el tratamiento real en sí mismo es el resultado de una decisión por parte del paciente y el médico, sostienen que el tratamiento en sí mismo está correlacionado con el término de error. En su lugar, MC2E utiliza la *predicción* del tratamiento, donde la variación en la predicción del tratamiento se debe a la variación en la variable instrumental: los pacientes más cercanos a un hospital que utiliza la técnica de cateterismo cardiaco tienen más probabilidades de recibir este tratamiento.

Esta interpretación tiene dos implicaciones. En primer lugar, la regresión VI estima en realidad el efecto del tratamiento no en un paciente «típico» seleccionado aleatoriamente, sino más bien en pacientes para los que la distancia es un factor importante en la decisión del tratamiento. El efecto sobre estos pacientes pueden diferir de los efectos sobre un paciente típico, lo cual proporciona una explicación acerca de la mayor eficacia estimada del tratamiento en los ensayos clínicos que la obtenida en el estudio de VI de McClellan, McNeil y Newhouse. En segundo lugar, se sugiere una estrategia general para encontrar los instrumentos en este tipo de marco: encontrar un instrumento que afecte a la probabilidad de tratamiento, pero que lo haga por razones que no estén correlacionadas con el resultado, excepto por su efecto a través de la probabilidad de tratamiento. Ambas implicaciones son de aplicación a los estudios experimentales y «cuasi experimentales», el tema del Capítulo 13.

12.6 Conclusión

Desde el humilde comienzo de la estimación acerca de cuánta menos mantequilla compraría la gente si su precio aumentase, los métodos VI han evolucionado hacia un método general para la estimación de regresiones cuando una o más variables están correlacionadas con el término de error. La regresión de variables instrumentales utiliza los instrumentos para aislar la variación en los regresores endógenos que no está correlacionada con el error en la regresión de interés; esta es la primera etapa de los mínimos cuadrados en dos etapas. Esto a su vez permite la estimación del efecto de interés en la segunda etapa de los mínimos cuadrados en dos etapas.

El éxito de la regresión VI requiere de instrumentos válidos, es decir, instrumentos que sean tanto relevantes (no débiles) como exógenos. Si los instrumentos son débiles, entonces el estimador MC2E puede ser sesgado, incluso en muestras grandes, y las inferencias estadísticas basadas en los estadísticos t y los intervalos de confianza MC2E pueden resultar engañosas. Afortunadamente, cuando existe un único regresor endógeno, es posible comprobar si los instrumentos son débiles simplemente mirando al estadístico F de la primera etapa.

Si los instrumentos no son exógenos, es decir, si uno o más instrumentos están correlacionados con el término de error, el estimador MC2E es inconsistente. Si hay más instrumentos que regresores endógenos, la exogeneidad de los instrumentos puede ser analizada mediante el estadístico J para contrastar las restricciones sobreidentificadas. Sin embargo, el supuesto central, acerca de que existen al menos tantos instrumentos exógenos como regresores endógenos, no puede ser contrastado. Por lo tanto, corresponde tanto al analista empírico como al lector crítico utilizar su conocimiento particular de la aplicación empírica para evaluar si este supuesto es razonable.

La interpretación de la regresión VI como una forma de explotar la variación exógena conocida en el regresor endógeno puede utilizarse para guiar la búsqueda de posibles variables instrumentales en una aplicación particular. Esta interpretación subyace en la mayor parte de los análisis empíricos que se encuentran bajo el título amplio de la evaluación de programas, en la que se utilizan los experimentos o quasi experimentos para estimar el efecto de los programas, las políticas, u otras intervenciones sobre alguna medida de los resultados. Varias cuestiones adicionales surgen en estas aplicaciones, por ejemplo, la interpretación de los resultados VI cuando, como en el ejemplo del cateterismo cardiaco, diferentes «pacientes» podrían tener respuestas diferentes al mismo «tratamiento». El análisis de estos y otros aspectos de la evaluación de programas empíricos se aborda en el Capítulo 13.

Resumen

1. La regresión con variables instrumentales constituye un método de estimar los coeficientes de regresión cuando una o más variables explicativas están correlacionadas con el término de error.
2. Las variables endógenas están correlacionadas con el término de error en la ecuación de interés; las variables exógenas no están correlacionadas con este término de error.
3. Para que un instrumento sea válido, debe (1) estar correlacionado con las variables endógenas incluidas y (2) ser exógeno.
4. La regresión VI requiere al menos tantos instrumentos como variables endógenas incluidas.
5. El estimador MC2E tiene dos etapas. En primer lugar, se hace una regresión de las variables endógenas incluidas sobre las variables exógenas incluidas y los instrumentos. En segundo lugar, se realiza la regresión de la variable dependiente sobre las variables exógenas incluidas y los valores estimados de las variables endógenas incluidas obtenidos mediante la(s) regresión(es) de la primera etapa.
6. Los instrumentos débiles (los instrumentos que casi están incorrelacionados con las variables endógenas incluidas) hacen que el estimador MC2E sea sesgado y que los intervalos de confianza y los contrastes de hipótesis sean poco fiables.
7. Si un instrumento no es exógeno, el estimador MC2E es inconsistente.

Términos clave

regresión con variables instrumentales (VI) (303)	sobreidentificado (311)
variables instrumentales (instrumentos) (303)	subidentificado (311)
variable endógena (304)	forma reducida (312)
variable exógena (304)	regresión de la primera etapa (313)
condición de relevancia del instrumento (304)	regresión de la segunda etapa (313)
condición de exogeneidad del instrumento (304)	instrumentos débiles (316)
mínimos cuadrados en dos etapas (304)	estadístico F de la primera etapa (317)
variables exógenas incluidas (311)	contraste de sobreidentificación
exactamente identificado (311)	de restricciones (318)

Revisión de conceptos

- 12.1** En el modelo de regresión de la curva de demanda de la Ecuación (12.3), ¿está $\ln(P_i^{\text{mantequilla}})$ correlacionado positiva o negativamente con el error, u_i ? Si se estima β_1 por MCO, se espera que el valor estimado sea mayor o menor que el verdadero valor de β_1 ? Explíquelo.
- 12.2** En el estudio de la demanda de cigarrillos de este capítulo, supóngase que se utiliza como instrumento el número de árboles por habitante en el estado. ¿Es relevante este instrumento? ¿Es exógeno? ¿Se trata de un instrumento válido?
- 12.3** En su estudio sobre los efectos de la encarcelación sobre las tasas de criminalidad, supóngase que Levitt hubiera utilizado el número de abogados por habitante como instrumento. ¿Es relevante este instrumento? ¿Es exógeno? ¿Se trata de un instrumento válido?
- 12.4** En su estudio acerca de la eficacia del cateterismo cardiaco, McClellan, McNeil, y Newhouse (1994) utilizaron como instrumento la diferencia de la distancia a un hospital con cateterismo cardiaco y a un hospital normal. ¿Cómo podría determinarse si este instrumento es relevante? ¿Cómo podría determinarse si este instrumento es exógeno?

Ejercicios

- 12.1** Esta pregunta se refiere a las regresiones de datos de panel que se resumen en la Tabla 12.1.
- Supóngase que el gobierno federal está considerando un nuevo impuesto sobre los cigarrillos que se estima que aumentará el precio de venta al por menor en 0,50 \$ por paquete. Si el precio actual por paquete es de 7,50 \$, utilice la regresión de la columna (1) para predecir la variación en la demanda. Construya un intervalo de confianza al 95 % para la variación de la demanda.
 - Supóngase que Estados Unidos entra en recesión y la renta se reduce en un 2 %. Utilice la regresión de la columna (1) para predecir la variación en la demanda.
 - Supóngase que la recesión dura menos de 1 año. ¿Cree que la regresión de la columna (1) proporcionará una respuesta fiable a la pregunta (b)? ¿Por qué o por qué no?
 - Supóngase que el estadístico F de la columna (1) fuese 3,6 en lugar de 33,6. ¿Proporcionaría la regresión una respuesta fiable a la pregunta planteada en (a)? ¿Por qué o por qué no?
- 12.2** Considérese el modelo de regresión con un único regresor: $Y_i = \beta_0 + \beta_1 X_i + u_i$. Supóngase que se cumplen los supuestos del Concepto clave 4.3.
- Demuestre que X_i es un instrumento válido. Es decir, demuestre que se cumple el Concepto clave 12.4 con $X_i = Z_i$.
 - Demuestre que los supuestos de la regresión VI del Concepto clave 12.4 se cumplen con esta elección de Z_i .
 - Demuestre que el estimador VI construido con $Z_i = X_i$ es idéntico al estimador MCO.
- 12.3** Un compañero de clase está interesado en la estimación de la varianza del término de error de la Ecuación (12.1).

- a) Supóngase que se utiliza el estimador de la regresión de la segunda etapa de MC2E $\hat{\sigma}_a^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0^{MC2E} - \hat{\beta}_1^{MC2E} \hat{X}_i)^2$: donde \hat{X}_i es valor ajustado tomado de la regresión de la primera etapa. ¿Es consistente este estimador? (Para los efectos de esta pregunta suponga que la muestra es muy grande y que los estimadores MC2E son esencialmente idénticos a β_0 y β_1).
b) ¿Es consistente $\hat{\sigma}_b^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0^{MC2E} - \hat{\beta}_1^{MC2E} \hat{X}_i)^2$?
- 12.4** Considérese la estimación MC2E con una sola variable endógena incluida y un único instrumento. Entonces, el valor de predicción de la regresión de la primera etapa es $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$. Utilice la definición de la varianza y la covarianza muestral para demostrar que $s_{\hat{X}Y} = \hat{\pi}_1 s_{ZY}$ y $s_{\hat{X}}^2 = \hat{\pi}_1^2 s_Z^2$. Utilice este resultado para completar los pasos de la deducción de la Ecuación (12.4) en el Apéndice 12.2.
- 12.5** Considérese el modelo de regresión de variables instrumentales
- $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$$
- donde X_i está correlacionada con u_i y Z_i es un instrumento. Supóngase que se cumplen los tres primeros supuestos del Concepto clave 12.4. ¿Qué supuesto VI no se cumple cuando:
- a) Z_i es independiente de (Y_i, X_i, W_i) ?
b) $Z_i = W_i$?
c) $Z_i = 1$ para todo i ?
d) $Z_i = X_i$?
- 12.6** En un modelo de regresión de variables instrumentales con un único regresor, X_i , y un único instrumento, Z_i , la regresión de X_i sobre Z_i presenta $R^2 = 0,05$ y $n = 100$. ¿Es Z_i un instrumento fuerte? [Sugerencia: véase la Ecuación (7.14)]. ¿Cambiaría su respuesta si $R^2 = 0,05$ y $n = 500$?
- 12.7** En un modelo de regresión de variables instrumentales con un único regresor, X_i , y dos instrumentos, Z_{1i} y Z_{2i} , el valor del estadístico J es 18,2
- a) Sugiere esto que $E(u_i|Z_{1i}, Z_{2i}) \neq 0$? Explíquelo.
b) Sugiere esto que $E(u_i|Z_{1i}) \neq 0$? Explíquelo.
- 12.8** Considérese el mercado de un producto con una función de oferta $Q_i^s = \beta_0 + \beta_1 P_i + u_i^s$, una función de demanda $Q_i^d = \gamma_0 + u_i^d$, y una condición de equilibrio de mercado $Q_i^s = Q_i^d$, donde u_i^s y u_i^d son variables aleatorias i.i.d. independientes entre sí, ambas con media igual a cero.
- a) Demuestre que P_i y u_i^s están correlacionados.
b) Demuestre que el estimador MCO de β_1 es inconsistente.
c) ¿Cómo estimaría β_0 , β_1 y γ_0 ?
- 12.9** Un investigador está interesado en el efecto del servicio militar sobre el capital humano. Recoge datos de una muestra aleatoria de 4.000 trabajadores de 40 años de edad y realiza una regresión MCO, $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde Y_i son los ingresos salariales anuales de los trabajadores, y X_i es una variable binaria que es igual a 1 si la persona sirvió en el ejército y es igual a 0 en caso contrario.
- a) Explique por qué las estimaciones MCO resultan probablemente poco fiables. (*Pista*: ¿qué variables se han omitido en la regresión? ¿Están correlacionadas con el servicio militar?)
b) Durante la guerra de Vietnam hubo un reclutamiento, en el cual la prioridad para ser reclutado estaba determinada por una lotería nacional. (Se seleccionaron las fechas de nacimiento y se ordenaron desde el 1 hasta el 365. Los nacidos en fechas que estaban ordenadas en primer lugar fueron reclutados antes que aquellos con fechas de nacimiento ordenadas en segundo lugar, y así sucesivamente). Explique cómo podría utilizarse la lotería como instrumento para estimar el efecto del servicio militar en los ingresos salariales. (Para más información sobre este tema, véase Joshua D. Angrist, «Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records», American Economic Review, junio 1990: 313-336).

12.10 Considérese el modelo de regresión de variables instrumentales $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$, donde Z_i es un instrumento. Supóngase que los datos sobre W_i no están disponibles y que se estima el modelo omitiendo W_i en la regresión.

- a) Supóngase que Z_i y W_i no están correlacionadas. ¿Es consistente el estimador VI?
- b) Supóngase que Z_i y W_i están correlacionadas. ¿Es consistente el estimador VI?

Ejercicios empíricos

E12.1 Durante la década de 1880, un cártel conocido como el Comité Ejecutivo Conjunto (JEC, en sus siglas originales) controlaba el transporte ferroviario de cereales desde las ciudades del medio oeste hacia las del oeste de los Estados Unidos. El cártel precedió a la Ley Antimonopolio Sherman de 1890, y operó legalmente para aumentar el precio del grano por encima de lo que habría sido el precio competitivo. De vez en cuando, los engaños de los miembros del cártel provocaban un colapso temporal en el acuerdo de fijación de precio colusivo. En este ejercicio, se utilizarán las variaciones de la oferta asociadas con los colapsos del cártel para estimar la elasticidad de la demanda de transporte de grano por ferrocarril. En la página web del libro http://www.pearsonhighered.com/stock_watson se encuentra el archivo de datos **JEC** que contiene las observaciones semanales sobre el precio del transporte en tren y de otros factores desde 1880 hasta 1886⁴. Se encuentra disponible una descripción detallada de los datos en el archivo **JEC_Description** disponible en la página web.

Supóngase que la curva de demanda de transporte ferroviario de cereales se especifica como $\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + \beta_2 Ice_i + \sum_{j=1}^{12} \beta_{2+j} Seas_{j,i} + u_i$ donde Q_i es el total de toneladas de grano enviado en la semana i , P_i es el precio de envío de una tonelada de grano por ferrocarril. La variable Ice_i es una variable binaria que es igual a 1 si los Grandes Lagos no son navegables a causa de hielo, y la variable $Seas_j$ es una variable binaria que capta la variación estacional de la demanda. La variable Ice se incluye porque el cereal podría igualmente transportarse por barco, cuando los Grandes Lagos son navegables.

- a) Estime la ecuación de la demanda por MCO. ¿Cuál es el valor estimado de la elasticidad de la demanda y su error estándar?
 - b) Explique por qué la interacción entre la oferta y la demanda podría hacer que el estimador MCO de la elasticidad fuera sesgado.
 - c) Considérese la utilización de la variable *cartel* como variable instrumental para $\ln(P)$. Utilice un razonamiento económico para analizar si resulta verosímil que la variable *cartel* satisfaga las dos condiciones para que un instrumento sea válido.
 - d) Estime la regresión de la primera etapa. ¿Es la variable *cartel* un instrumento débil?
 - e) Estime la ecuación de la demanda mediante regresión de variables instrumentales. ¿Cuál es la elasticidad de la demanda estimada y su error estándar?
 - f) ¿Sugiere la evidencia que el cártel estuviera fijando el precio que maximizaba el beneficio de monopolio? Explíquelo. (*Pista:* ¿Qué debería hacer un monopolista si la elasticidad precio fuese menor que 1?).
- E12.2** ¿Cómo afecta la fertilidad a la oferta de trabajo? Es decir, ¿cuánto disminuye la oferta de trabajo de una mujer cuando tiene un hijo más? En este ejercicio, se estimará este efecto a partir de los datos de las mujeres casadas del censo de los EE.UU. de 1980⁵. Los datos están disponibles en la página web del libro de texto http://www.pearsonhighered.com/stock_watson, en el archivo **Fertility** y se describen en el archivo **Fertility_Description**. La base de datos contiene información sobre las mujeres casadas de entre 21 y 35 años con dos o más hijos.

⁴ Estos datos fueron proporcionados por el profesor Robert Porter de la Universidad de Northwestern y fueron utilizados en su artículo «A Study of Cartel Stability: The Joint Executive Committee, 1880-1886», *The Bell Journal of Economics*, 1983, 14(2), 301-314.

⁵ Estos datos fueron proporcionados por el profesor William Evans de la Universidad de Maryland y se utilizaron en su artículo en colaboración con Joshua Angrist, «Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size», *American Economic Review*, 1998, 88(3): 450-477.

- a) Realice la regresión de la variable *weeksworked* sobre la variable indicador *morekids* mediante MCO. En promedio, ¿trabajan menos las mujeres con más de dos hijos que las mujeres con dos hijos? ¿Cuánto menos?
- b) Explique por qué la regresión MCO estimada en el apartado (a) no resulta apropiada para estimar el efecto causal de la fertilidad (*morekids*) sobre la oferta de trabajo (*weeksworked*).
- c) La base de datos contiene la variable *samesex*, que es igual a 1 si los dos primeros hijos son del mismo sexo (chico-chico o chica-chica) y es igual a 0 en caso contrario. ¿Tienen las parejas cuyos dos primeros hijos son del mismo sexo más probabilidades de tener un tercer hijo? ¿Es grande este efecto? ¿Es estadísticamente significativo?
- d) Explique por qué la variable *samesex* es un instrumento válido para la regresión de variables instrumentales de la variable *weeksworked* sobre la variable *morekids*.
- e) ¿Es la variable *samesex* un instrumento débil?
- f) Estime la regresión de la variable *weeksworked* sobre la variable *morekids* utilizando la variable *samesex* como instrumento. ¿Cuánto es el efecto de la fertilidad sobre la oferta de trabajo?
- g) ¿Cambian los resultados cuando se incluyen las variables *agem1*, *black*, *hispan*, y *othrace* en la regresión de la oferta de trabajo (considerando estas variables como variables exógenas)? Explique por qué o por qué no.
- E12.3** (Requiere el Apéndice 12.5). En la página web del libro de texto http://www.pearsonhighered.com/stock_watson se encuentra la base de datos **WeakInstrument** que contiene 200 observaciones sobre (Y_i, X_i, Z_i) para la regresión instrumental $Y_i = \beta_0 + \beta_1 X_i + u_i$.
- Obtenga $\hat{\beta}_1^{MC2E}$, su error estándar, y el intervalo de confianza habitual del 95 % para β_1 .
 - Calcule el estadístico F para la regresión de X_i sobre Z_i . ¿Existe evidencia de la presencia de un problema de «instrumento débil»?
 - Calcule un intervalo de confianza del 95 % para β_1 utilizando el procedimiento de Anderson-Rubin. (Para aplicar el procedimiento, se supone que $-5 \leq \beta_1 \leq 5$).
 - Comente las diferencias en los intervalos de confianza de los apartados (a) y (c). ¿Cuál es más fiable?

APÉNDICE

12.1 La base de datos de panel sobre consumo de cigarrillos

La base de datos consta de los datos anuales sobre los 48 estados contiguos de EE.UU. desde 1985 hasta 1995. La cantidad consumida se mide mediante las ventas anuales de cigarrillos en cajetillas per cápita por año fiscal, tal y como se desprende de los datos de la recaudación de impuestos del estado. El precio es el precio medio de venta al por menor por cajetilla en términos reales (es decir, ajustado por la inflación) durante el año fiscal, con impuestos incluidos. La renta es la renta real per cápita. El impuesto general sobre ventas es el impuesto promedio, en centavos de dólar por paquete, debido a la amplia gama de impuestos estatales que se aplican a todos los bienes de consumo. El impuesto específico sobre los cigarrillos es el impuesto que se aplica solamente a los cigarrillos. Todos los precios, rentas e impuestos utilizados en las regresiones de este capítulo están deflactados por el Índice de Precios al Consumidor y por lo tanto están expresados en dólares constantes (términos reales). Estamos muy agradecidos al profesor Jonathan Gruber del MIT por habernos proporcionado estos datos.

APÉNDICE

12.2 Obtención de la fórmula del estimador MC2 de la Ecuación (12.4)

La primera etapa de MC2E consiste en realizar una regresión de X_i sobre el instrumento Z_i por MCO y luego calcular el valor de predicción MCO \hat{X}_i , y la segunda etapa consiste en realizar la regresión de Y_i sobre \hat{X}_i por MCO. En

consecuencia, la fórmula para el estimador MC2E, expresada en términos del valor de predicción \hat{X}_i , es la fórmula para el estimador MCO del Concepto clave 4.2, con \hat{X}_i sustituyendo a X_i . Es decir, $\hat{\beta}_1^{MC2E} = s_{\hat{X}Y}/s_{\hat{X}}^2$, donde $s_{\hat{X}}^2$ es la varianza muestral de \hat{X}_i y $s_{\hat{X}Y}$ es la covarianza muestral entre Y_i y \hat{X}_i .

Debido a que \hat{X}_i es el valor de predicción de X_i procedente de la regresión de la primera etapa, $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$, las definiciones de las varianzas y covarianzas muestrales implican que $s_{\hat{X}Y} = \hat{\pi}_1 s_{ZY}$ y $s_{\hat{X}}^2 = \hat{\pi}_1^2 s_Z^2$ (Ejercicio 12.4). Por lo tanto, el estimador MC2E se puede escribir como $\hat{\beta}_1^{MC2E} = s_{\hat{X}Y}/s_{\hat{X}}^2 = s_{ZY}/(\hat{\pi}_1 s_Z^2)$. Por último, $\hat{\pi}_1$ es el coeficiente MCO de la pendiente de la primera etapa de MC2E, por lo que $\hat{\pi}_1 = s_{ZX}/s_Z^2$. Mediante la sustitución de esta fórmula para $\hat{\pi}_1$ en la fórmula $\hat{\beta}_1^{MC2E} = s_{ZY}/(\hat{\pi}_1 s_Z^2)$ se obtiene la fórmula para el estimador MC2E de la Ecuación (12.4).

APÉNDICE

12.3 Distribución del estimador MC2E para grandes muestras

En este apéndice se estudia la distribución en muestras grandes del estimador MC2E para el caso considerado en la Sección 12.1, es decir, con un único instrumento, una única variable endógena incluida, y sin variables exógenas incluidas.

Para empezar, se obtiene una fórmula para el estimador MC2E en términos de los errores que constituye la base para el análisis restante, similar a la expresión para el estimador MCO en la Ecuación (4.30) del Apéndice 4.3. De la ecuación (12.1), $Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + (u_i - \bar{u})$. En consecuencia, la covarianza muestral entre Z e Y se puede expresar como

$$\begin{aligned} s_{ZY} &= \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) \\ &= \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})] \\ &= \beta_1 s_{ZX} + \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})(u_i - \bar{u}) \\ &= \beta_1 s_{ZX} + \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})u_i \end{aligned} \tag{12.19}$$

donde $s_{ZX} = [1/(n-1)]\sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})$ y donde la última igualdad aparece debido a que $\sum_{i=1}^n (Z_i - \bar{Z}) = 0$. Sustituyendo la definición de s_{ZX} y la última expresión de la Ecuación (12.19) en la definición de $\hat{\beta}_1^{MC2E}$ y multiplicando el numerador y el denominador por $(n-1)/n$ resulta

$$\hat{\beta}_1^{MC2E} = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})u_i}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})} \tag{12.20}$$

Distribución para muestras grandes de $\hat{\beta}_1^{MC2E}$ cuando se cumplen los supuestos de la regresión VI del Concepto clave 12.4

La Ecuación (12.20) para el estimador MC2E es similar a la Ecuación (4.30) del Apéndice 4.3 para el estimador MCO, con la excepción de que aparece Z en lugar de X en el numerador y de que el denominador es la covarianza entre Z y X en lugar de la varianza de X . Debido a estas similitudes, y debido a que Z es exógeno, el argumento del Apéndice 4.3 en el sentido de que el estimador MCO se distribuye normalmente en muestras grandes se extiende al caso de $\hat{\beta}_1^{MC2E}$.

En concreto, cuando la muestra es grande $\bar{Z} \approx \mu_Z$, por lo que el numerador es aproximadamente $\bar{q} = (1/n) \sum_{i=1}^n q_i$, donde $q_i = (Z_i - \mu_Z)u_i$. Debido a que el instrumento es exógeno, $E(q_i) = 0$. Por los supuestos de la regresión VI del Concepto clave 12.4, q_i es i.i.d. con varianza $\sigma_q^2 = \text{var}[(Z_i - \mu_Z)u_i]$. De ello se desprende que $\text{var}(\bar{q}) = \sigma_q^2 = \sigma_q^2/n$, y, por el teorema central del límite, $\bar{q}/\sigma_{\bar{q}}$ está, en muestras grandes, distribuido $N(0, 1)$.

Debido a que la covarianza muestral es un estimador consistente de la covarianza muestral, $s_{ZX} \xrightarrow{P} \text{cov}(Z_i, X_i)$, la cual, debido a que el instrumento es relevante, es distinta de cero. Por tanto, por la Ecuación (12.20) $\hat{\beta}_1^{MC2E} \cong \beta_1 + \bar{q}/\text{cov}(Z_i, X_i)$, por lo que en muestras grandes $\hat{\beta}_1^{MC2E}$ se distribuye aproximadamente $N(\beta_1, \sigma_{\hat{\beta}_1^{MC2E}}^2)$, donde $\sigma_{\hat{\beta}_1^{MC2E}}^2 = \sigma_q^2 / [\text{cov}(Z_i, X_i)]^2 = (1/n) \text{var}[(Z_i - \mu_Z)u_i]/[\text{cov}(Z_i, X_i)]^2$, que es la expresión dada en la Ecuación (12.8).

APÉNDICE

12.4 La distribución del estimador MC2E para muestras grandes cuando el instrumento no es válido

En este apéndice se considera la distribución para muestras grandes del estimador MC2E en las condiciones de la Sección 12.1 (una X , una Z), cuando una u otra de las condiciones para la validez de los instrumentos no se cumple. Si la condición de relevancia del instrumento no se cumple, la distribución para grandes muestras del estimador MC2E no es normal; de hecho, su distribución es la del cociente de dos variables aleatorias normales. Si la condición de exogeneidad del instrumento no se cumple, el estimador MC2E es inconsistente.

Distribución para muestras grandes cuando el instrumento es débil

Consideremos primero el caso de que el instrumento es irrelevante por lo que $\text{cov}(Z_i, X_i) = 0$. Entonces, el argumento del Apéndice 12.3 entraña la división entre cero. Para evitar este problema, es necesario echar un vistazo más detallado al comportamiento del término del denominador de la Ecuación (12.20), cuando la covarianza poblacional es cero.

Empezamos por reescribir la Ecuación (12.20). Debido a la consistencia de la media muestral, para muestras grandes, \bar{Z} está cercana a μ_Z y \bar{X} está cercana a μ_X . Por tanto, el término en el denominador de la Ecuación (12.20) es aproximadamente igual a $(1/n)\sum_i^n(Z_i - \mu_Z)(X_i - \mu_X) = \sum_{i=1}^n r_i = \bar{r}$, donde $r_i = (Z_i - \mu_Z)(X_i - \mu_X)$. Sea $\sigma_r^2 = \text{var}(Z_i - \mu_Z)(X_i - \mu_X)$, sea $\sigma_{\bar{r}}^2 = \sigma_r^2/n$, y sean \bar{q} , $\sigma_{\bar{q}}^2$, y σ_q^2 tal y como se definen en el Apéndice 12.3. Entonces la Ecuación (12.20) implica que, para muestras grandes,

$$\hat{\beta}_1^{MC2E} \cong \beta_1 + \frac{\bar{q}}{\bar{r}} = \beta_1 + \left(\frac{\sigma_{\bar{q}}}{\sigma_{\bar{r}}} \right) \left(\frac{\bar{q}/\sigma_{\bar{q}}}{\bar{r}/\sigma_{\bar{r}}} \right) = \beta_1 + \left(\frac{\sigma_q}{\sigma_r} \right) \left(\frac{\bar{q}/\sigma_{\bar{q}}}{\bar{r}/\sigma_{\bar{r}}} \right) \quad (12.21)$$

Si el instrumento es irrelevante, entonces $E(r_i) = \text{cov}(Z_i, X_i) = 0$. Por tanto \bar{r} es la media muestral de las variables aleatorias r_i , $i = 1, \dots, n$, que son i.i.d. (por el segundo supuesto de mínimos cuadrados), tiene una varianza $\sigma_r^2 = \text{var}(Z_i - \mu_Z)(X_i - \mu_X)$ (que es finita por el tercer supuesto de la regresión VI), y tienen una media igual a cero (debido a que los instrumentos son irrelevantes). De ello se desprende que es aplicable el teorema central del límite a \bar{r} , en concreto, $\bar{r}/\sigma_{\bar{r}}$ se distribuye aproximadamente $N(0, 1)$. Por lo tanto, la última expresión de la Ecuación (12.21) implica que, en grandes muestras, la distribución de $\hat{\beta}_1^{MC2E} - \beta_1$ es la distribución de aS , donde $a = \sigma_q/\sigma_r$, y S es el cociente de dos variables aleatorias, cada una de las cuales tiene una distribución normal estándar (estas dos variables aleatorias normales estándar están correlacionadas).

En otras palabras, cuando el instrumento es irrelevante, el teorema central del límite es aplicable al denominador, así como al numerador del estimador MC2E, por lo que en muestras grandes la distribución del estimador MC2E es la distribución del cociente de dos variables aleatorias normales. Debido a que X_i y u_i están correlacionadas, estas variables aleatorias normales están correlacionadas, y la distribución para grandes muestras del estimador MC2E cuando el instrumento es irrelevante es complicada. De hecho, la distribución para muestras grandes del estimador MC2E con instrumentos irrelevantes está centrada en el límite de probabilidad del estimador MCO. Por tanto, cuando el instrumento es irrelevante, MC2E no elimina el sesgo de MCO y, además, presenta una distribución distinta de la normal, incluso en muestras grandes.

Un instrumento débil representa un caso intermedio entre un instrumento irrelevante y la distribución normal obtenida en el Apéndice 12.3. Cuando el instrumento es débil, pero no irrelevante, la distribución del estimador MC2E sigue siendo distinta de la normal, por lo que la lección general que puede extraerse de este caso extremo de un instrumento irrelevante es trasladable a los instrumentos débiles.

Distribución muestral de $\hat{\beta}_1^{\text{MC2E}}$ cuando el instrumento es endógeno

El numerador de la última expresión de la Ecuación (12.20) converge en probabilidad a $\text{cov}(Z_i, u_i)$. Si el instrumento es exógeno, este es igual a cero y el estimador MC2E es consistente (suponiendo que el instrumento no es débil). Sin embargo, si el instrumento no es exógeno, entonces, si el instrumento no es débil, $\hat{\beta}_1^{\text{MC2E}} \xrightarrow{P} \beta_1 + \text{cov}(Z_i, u_i)/\text{cov}(Z_i, X_i) \neq \beta_1$. Es decir, si el instrumento no es exógeno, el estimador MC2E es inconsistente.

APÉNDICE

12.5 Análisis de variables instrumentales con instrumentos débiles

Este apéndice presenta algunos métodos para el análisis de variables instrumentales en presencia de instrumentos potencialmente débiles. El apéndice se centra en el caso de un único regresor endógeno incluido [Ecuaciones (12.13) y (12.14)].

Contraste de instrumentos débiles

La regla práctica del Concepto clave 12.5 dice que un estadístico F de la primera etapa menor a 10 indica que los instrumentos son débiles. Una de las motivaciones de esta regla práctica surge de una expresión aproximada para el sesgo del estimador MC2E. Sea β_1^{MCO} la expresión del límite de probabilidad del estimador MCO de β_1 , y sea $\beta_1^{\text{MCO}} - \beta_1$ la expresión del sesgo asintótico del estimador MCO (si el regresor es endógeno, entonces $\hat{\beta}_1 \xrightarrow{P} \beta_1^{\text{MCO}} \neq \beta_1$). Es posible demostrar que, cuando existen muchos instrumentos, el sesgo de MC2E es aproximadamente $E(\hat{\beta}_1^{\text{MC2E}}) - \beta_1 \approx (\beta_1^{\text{MCO}} - \beta_1)/[E(F) - 1]$, donde $E(F)$ es la esperanza del estadístico F de la primera etapa. Si $E(F) = 10$, entonces el sesgo de MC2E, en relación con el sesgo de MCO, es de aproximadamente 1/9, o simplemente de más del 10 %, que es lo suficientemente pequeño como para ser aceptable en muchas aplicaciones. Sustituyendo $E(F) > 10$ por $F > 10$ se obtiene la regla práctica del Concepto clave 12.5.

El razonamiento del párrafo anterior incluía una fórmula aproximada para el sesgo del estimador MC2E cuando existen muchos instrumentos. En la mayoría de las aplicaciones, sin embargo, el número de instrumentos, m , es pequeño. Stock y Yogo (2005) proporcionan un contraste formal para los instrumentos débiles, que evita la aproximación con m grande. En el contraste de Stock-Yogo, la hipótesis nula es que los instrumentos son débiles y la hipótesis alternativa es que los instrumentos son fuertes, donde los instrumentos fuertes se definen como los instrumentos para los que el sesgo del estimador MC2E es de al menos el 10 % del sesgo del estimador MCO. El contraste implica la comparación del estadístico F de la primera etapa (por razones técnicas, la versión válida con homocedasticidad) con un valor crítico que depende del número de instrumentos. Como es el caso, para un contraste con un nivel de significación del 5 %, este rango de valores críticos está entre 9,08 y 11,52, por lo que la regla práctica de la comparación de F con 10 es una buena aproximación al contraste de Stock-Yogo.

Contraste de hipótesis y conjuntos de confianza para β

Si los instrumentos son débiles, el estimador MC2E es sesgado y tiene una distribución distinta de la normal. Por tanto, el contraste t MC2E de $\beta_1 = \beta_{1,0}$ es poco fiable, como lo es el intervalo de confianza MC2E para β_1 . Existen, no obstante, otros contrastes para $\beta_1 = \beta_{1,0}$, así como intervalos de confianza basados en estos contrastes, que resultan válidos si los instrumentos son fuertes, débiles, o incluso irrelevantes. Cuando existe un único regresor endógeno, el contraste preferido es el contraste de razón de verosimilitud condicional de Moreira (2003) (RVC). Un contraste más antiguo, que funciona para cualquier número de regresores endógenos, es el que se basa en el estadístico Anderson-Rubin (1949). Debido a que el estadístico Anderson-Rubin (1949) resulta conceptualmente menos complicado, se describe en primer lugar.

El contraste de Anderson-Rubin $\beta_1 = \beta_{1,0}$ procede en dos pasos. En el primer paso, se calcula una nueva variable $Y_i^* = Y_i - \beta_{1,0}X_i$. En el segundo paso, se realiza una regresión de Y_i^* sobre los regresores exógenos incluidos (W) y los instrumentos (Z). El estadístico de Anderson-Rubin es el estadístico F que contrasta la hipótesis de que los coeficientes de las Z son todos iguales a cero. Bajo la hipótesis nula de que $\beta_1 = \beta_{1,0}$, si los instrumentos cumplen la condición de exogeneidad (la condición 2 en el Concepto clave 12.3), estos estarán incorrelacionados con el término de error de esta regresión y la hipótesis nula será rechazada en el 5 % de todas las muestras.

Tal y como se analizó en las Secciones 3.3 y 7.4, se puede construir un conjunto de confianza como el conjunto de valores de los parámetros que no son rechazados por un contraste de hipótesis. En consecuencia, el conjunto de valores de β_1 que no son rechazados por un contraste de Anderson-Rubin al 5 % constituye un conjunto confianza del 95 % para β_1 . Cuando el estadístico F de Anderson-Rubin se calcula utilizando la fórmula válida con homocedasticidad, el conjunto de confianza de Anderson-Rubin se puede construir resolviendo una ecuación cuadrática (véase el Ejercicio empírico 12.3). La lógica que se encuentra tras el estadístico de Anderson-Rubin nunca supone la relevancia de los instrumentos, y el conjunto de la confianza de Anderson-Rubin tendrá una cobertura de probabilidad del 95 % en muestras grandes si los instrumentos son fuertes, débiles o incluso irrelevantes.

El estadístico RVC contrasta asimismo la hipótesis de que $\beta_1 = \beta_{1,0}$. Los estadísticos de razón de verosimilitud comparan el valor de la verosimilitud (véase el Apéndice 11.2) bajo la hipótesis nula con su valor bajo la alternativa y la rechazarán si la verosimilitud bajo la alternativa es lo suficientemente más grande que bajo la hipótesis nula. Los contrastes habituales en este libro, tales como el contraste F válido con homocedasticidad en regresión múltiple, se pueden obtener como contrastes de razón de verosimilitud bajo el supuesto de errores homocedásticos normalmente distribuidos. Sin embargo, a diferencia de cualquiera de los contrastes examinados en este libro, el valor crítico del contraste de RVC depende de los datos, concretamente de un estadístico que mide la fortaleza de los instrumentos. Utilizando el valor crítico correcto, el contraste RVC es válido tanto si los instrumentos son fuertes, débiles, o irrelevantes. Los intervalos de confianza RVC pueden calcularse como el conjunto de valores de β_1 que no son rechazados por el contraste RVC.

El contraste RVC es equivalente al contraste t MC2E cuando los instrumentos son fuertes y tiene potencia cuando los instrumentos son válidos. Con el *software* adecuado, el contraste RVC resulta fácil de utilizar. La desventaja del contraste RVC consiste en que no resulta fácilmente generalizable al caso de que haya más de un regresor endógeno. En ese caso, se recomienda el contraste (y el conjunto de confianza) de Anderson-Rubin; sin embargo, cuando los instrumentos son fuertes (por lo que MC2E es válido) y los coeficientes están sobreidentificados, el contraste de Anderson-Rubin es ineficiente en el sentido de que tiene menos potencia que el contraste t MC2E.

Estimación de β

Si los instrumentos son irrelevantes, no es posible obtener un estimador insesgado de β_1 , incluso en muestras grandes. Sin embargo, si los instrumentos son débiles, algunos estimadores VI tienden a estar más centrados en el verdadero valor de β_1 de lo que lo está MC2E. Un estimador de este tipo es el estimador de máxima verosimilitud con información limitada (MVIL). Tal y como su nombre indica, el estimador MVIL es el estimador de máxima verosimilitud de β_1 en el sistema formado por las Ecuaciones (12.13) y (12.14) (para un análisis de la estimación de máxima verosimilitud, véase el Apéndice 11.2). El estimador MVIL es asimismo el valor de β_1 que minimiza el estadístico de contraste de Anderson-Rubin válido con homocedasticidad. Por lo tanto, si el conjunto de confianza de Anderson-Rubin es distinto del vacío, contendrá el estimador MVIL. Además, el intervalo de confianza de RVC contiene al estimador MVIL.

Si los instrumentos son débiles, el estimador MVIL está centrado de forma más precisa en el verdadero valor de β_1 de lo que lo está MC2E. Si los instrumentos son fuertes, los estimadores MVIL y MC2E coinciden en muestras grandes. Una desventaja del estimador MVIL es que puede originar valores atípicos extremos. Los intervalos de confianza construidos en torno al estimador MVIL utilizando el error estándar MVIL son más fiables que los intervalos construidos en torno al estimador MC2E utilizando el error estándar MC2E, pero son menos fiables que los intervalos Anderson-Rubin o los intervalos RVC cuando los instrumentos son débiles.

Los problemas de estimación, contrastación, y los intervalos de confianza en la regresión VI con instrumentos débiles constituyen un área de investigación. Para una más amplia información, visite la página web de este libro.

APÉNDICE

12.6 MC2E con variables de control

En el Concepto clave 12.4, las variables W se suponían exógenas. En este apéndice se considera el caso en el que W no es exógena, sino que es una variable de control incluida para lograr que Z sea exógena. La lógica de las variables de control en MC2E es paralela a la lógica en MCO: si una variable de control efectivamente tiene en cuenta (controla por) un factor omitido, entonces el instrumento no estará correlacionado con el término de error. Debido a que la variable de control está correlacionada con el término de error, el coeficiente de una variable de control no tiene una interpretación causal. Las matemáticas de las variables de control en MC2E son análogas asimismo a las matemáticas de las variables

de control en MCO y conllevan la relajación del supuesto de que el error tiene media condicional igual a cero, dados Z y W , para que la media condicional del error no dependa de Z . Este apéndice se basa en el Apéndice 7.2 (independencia en media condicional), que debe ser revisado antes.

Consideremos el modelo de regresión VI de la Ecuación (12.12) con una única X y una única W :

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i. \quad (12.22)$$

El supuesto #1 de la regresión VI del Concepto clave 12.4 [que establece que $E(u_i|W_i) = 0$] se reemplaza por el supuesto de que, condicionado a W_i , la media de u_i no depende de Z_i :

$$E(u_i|W_i, Z_i) = E(u_i|W_i). \quad (12.23)$$

De acuerdo con el Apéndice 7.2, se supone además que $E(u_i|W_i)$ es lineal en W_i , por lo que $E(u_i|W_i) = \gamma_0 + \gamma_2 W_i$, donde γ_0 y γ_2 son coeficientes. Sea $\varepsilon_i = u_i - E(u_i|W_i, Z_i)$ y aplicando el álgebra de la Ecuación (7.25) a la Ecuación (12.22), se obtiene

$$Y_i = \delta_0 + \beta_1 X_i + \delta_2 W_i + \varepsilon_i, \quad (12.24)$$

donde $\delta_0 = \beta_0 + \gamma_0$ y $\delta_2 = \beta_2 + \gamma_2$. Ahora $E(\varepsilon_i|W_i, Z_i) = E[u_i - E(u_i|W_i, Z_i)|W_i, Z_i] = E(u_i|W_i, Z_i) - E(u_i|W_i, Z_i) = 0$, lo que a su vez implica que $\text{corr}(Z_i, \varepsilon_i) = 0$. Por lo tanto el supuesto #1 de la regresión VI así como el requisito de exogeneidad del instrumento (condición #2 del Concepto clave 12.3) se cumplen para la Ecuación (12.24) con el término de error ε_i . Por lo tanto, si el supuesto #1 de la regresión VI se sustituye por la independencia en media condicional en la Ecuación (12.23), los supuestos de la regresión VI originales del Concepto clave 12.4 son aplicables a la regresión modificada de la Ecuación (12.24).

Debido a que los supuestos de la regresión VI del Concepto clave 12.4 se cumplen para la Ecuación (12.24), todos los métodos de inferencia (tanto para los instrumentos débiles como para los fuertes) estudiados en este capítulo son aplicables a la Ecuación (12.24). En particular, si los instrumentos son fuertes, los coeficientes de la Ecuación (12.24) se estimarán consistentemente mediante MC2E y los contrastes MC2E y los intervalos de confianza será válidos.

Al igual que en MCO con variables de control, en general, el coeficiente MC2E de la variable de control W no tiene una interpretación causal. MC2E estima de forma consistente δ_2 en la Ecuación (12.24), pero δ_2 es la suma del efecto causal directo de $W(\beta_2)$ y γ_2 , lo que refleja la correlación entre W y los factores omitidos dentro de u_i que W tiene en cuenta.

En las regresiones de consumo de cigarrillos de la Tabla 12.1, resulta tentador interpretar el coeficiente de la variación a lo largo de 10 años en el logaritmo de la renta como la elasticidad renta de la demanda. Sin embargo, si el crecimiento de la renta está correlacionado con el aumento en la educación y si más educación reduce el consumo de tabaco, el crecimiento de la renta tendría su propio efecto causal (β_2 , la elasticidad-renta), además de un efecto causado por su correlación con la educación (γ_2). Si este último efecto es negativo ($\gamma_2 < 0$), los coeficientes de la renta en la Tabla 12.1 (que estima $\delta_2 = \beta_2 + \gamma_2$) subestimarían la elasticidad-renta, pero si se cumple el supuesto de independencia en media condicional de la Ecuación (12.23), el estimador MC2E de la elasticidad-precio es consistente.

Experimentos y cuasi experimentos

En muchas áreas como la psicología y la medicina, los efectos causales se estiman mediante la utilización de experimentos. Por ejemplo, antes de ver aprobada su utilización médica generalizada, un nuevo medicamento debe someterse a pruebas experimentales en las que algunos pacientes son seleccionados para recibir el medicamento de forma aleatoria mientras que a otros se les proporciona un sustituto ineficaz e inofensivo (un «placebo»); el medicamento es aprobado solamente si el experimento aleatorizado controlado proporciona evidencia estadística convincente de que ese medicamento es seguro y efectivo.

Existen tres razones para el estudio de los experimentos aleatorizados controlados en un curso de econometría. En primer lugar, un experimento aleatorizado controlado ideal proporciona un punto de referencia conceptual que permite juzgar las estimaciones de los efectos causales realizadas con datos procedentes de observaciones. En segundo lugar, los resultados de los experimentos aleatorizados controlados, cuando se llevan a cabo, pueden ser muy influyentes, por lo que es importante entender las limitaciones y amenazas a la validez de los experimentos reales así como sus puntos fuertes. En tercer lugar, las circunstancias externas a veces originan que parezca que existe aleatoriedad; es decir, debido a los sucesos externos, algunos individuos son sometidos al tratamiento «como si» fuese de forma aleatoria, de forma posiblemente condicional a algunas variables de control. Este «como si» fuese aleatorio da lugar a un «cuasi experimento» o «experimento natural», y muchos de los métodos desarrollados para analizar los experimentos aleatorizados pueden ser aplicados (con algunas modificaciones) a los cuasi experimentos.

En este capítulo se analizan los experimentos y los cuasi experimentos en economía. Las herramientas estadísticas utilizadas en este capítulo son el análisis de regresión múltiple, el análisis de datos de panel, y la regresión con variables instrumentales (VI). Lo que distingue el análisis de este capítulo no son las herramientas utilizadas, sino el tipo de datos analizados y las oportunidades o desafíos particulares que se plantean cuando se analizan los experimentos y los cuasi experimentos.

Los métodos desarrollados en este capítulo se utilizan a menudo para la evaluación de programas económicos y sociales. La **evaluación de programas** es el área de estudio que trata la estimación del efecto de un programa, decisión política, o alguna otra intervención, o «tratamiento». ¿Cuál es el efecto sobre los ingresos salariales de acudir a un programa de formación laboral? ¿Cuál es el efecto sobre el empleo de los trabajadores con baja cualificación de un aumento en el salario mínimo? ¿Cuál es el efecto sobre la asistencia a la universidad de la habilitación de préstamos en condiciones ventajosas como ayuda a los estudiantes de clases medias? En este capítulo se analiza cómo pueden evaluarse estos programas o políticas utilizando experimentos o cuasi experimentos.

Comenzamos en la Sección 13.1 explicando en detalle los argumentos de los Capítulos 1, 3 y 4 de la estimación de los efectos causales con experimentos aleatorizados controlados. En la realidad, los experimentos reales con seres humanos se enfrentan a problemas prácticos que constituyen amenazas a su validez interna y externa; estas amenazas y algunas de las herramientas econométricas que se utilizan para afrontarlas se tratan en la Sección 13.2. En la Sección 13.3 se analiza un experimento aleatorizado controlado importante en el que los estudiantes de primaria fueron asignados aleatoriamente a clases de diferentes tamaños en el estado de Tennessee a finales de la década de 1980.

La sección 13.4 se centra en la estimación de los efectos causales utilizando cuasi experimentos. Las amenazas a la validez de los cuasi experimentos se tratan en la Sección 13.5. Uno de los problemas

que surgen, tanto en los experimentos como en los cuasi experimentos, es que los efectos del tratamiento pueden diferir de un miembro de la población a otro, y la cuestión de la interpretación de las estimaciones resultantes de los efectos causales cuando la población es heterogénea; se recoge en la Sección 13.6.

13.1 Variables respuesta, efectos causales y experimentos ideales

En esta sección se explica cómo la media poblacional de los efectos causales a nivel individual puede estimarse mediante un experimento aleatorizado controlado y cómo pueden analizarse los datos de un experimento así utilizando el análisis de regresión múltiple.

Variables respuesta y efecto causal promedio

Suponga que está considerando la opción de tomar un medicamento para una patología médica, inscribirse en un programa de formación laboral, o hacer una lista optativa de problemas de econometría. Resulta razonable preguntarse, ¿cuáles son los beneficios que recibo por hacerlo (recibir el tratamiento)? Se pueden imaginar dos situaciones hipotéticas, una en la que se recibe el tratamiento y la otra en la que no se recibe. Bajo cada situación hipotética, se obtendría un resultado medible (el progreso en el estado de salud, la obtención de un trabajo, la calificación en el curso de econometría). La diferencia entre estos dos resultados posibles sería el efecto causal, para uno mismo, del tratamiento.

En términos más generales, una **variable respuesta** es el resultado que obtiene un individuo al recibir un tratamiento potencial. El efecto causal para este individuo es la diferencia entre la variable respuesta si el tratamiento se recibe y la variable respuesta si no se recibe. En general, el efecto causal puede variar de un individuo a otro. Por ejemplo, el efecto de un medicamento puede depender de la edad, de si se fuma o no, o de otras condiciones de salud. El problema es que no hay manera de medir el efecto causal para una sola persona. Debido a que el individuo o bien recibe el tratamiento o no lo recibe, se puede observar una de las variables respuesta o resultado, pero no ambas.

Aunque el efecto causal no puede ser medido para un único individuo, en muchas aplicaciones es suficiente conocer el efecto causal promedio en una población. Por ejemplo, en la evaluación de un programa de formación laboral debería considerarse la disyuntiva entre el gasto medio por alumno en prácticas frente al éxito promedio en la búsqueda de trabajo de estos alumnos. La media de los efectos causales individuales en la población estudiada se denomina **efecto causal promedio** o **efecto promedio del tratamiento**.

El efecto causal promedio para una población dada se puede estimar, al menos en teoría, mediante un experimento aleatorizado controlado ideal. Para ver cómo, en primer lugar se supone que los sujetos son seleccionados al azar a partir de la población de interés. Debido a que los sujetos son seleccionados por muestreo aleatorio simple, sus variables respuesta (resultados potenciales), y por lo tanto sus efectos causales, se han extraído de la misma distribución, por lo que el valor esperado del efecto causal muestral es el efecto causal promedio de la población. Supóngase ahora que los sujetos son asignados aleatoriamente al grupo de tratamiento o al grupo de control. Debido a que el estado de un individuo respecto al tratamiento es asignado aleatoriamente, se distribuye de forma independiente de sus posibles resultados (variables respuesta). Por tanto el valor esperado de los resultados para los pacientes tratados menos el valor esperado de los resultados para los no tratados es igual al valor esperado del efecto causal. Por lo tanto, cuando el concepto de las variables respuesta (resultados potenciales) se combina con (1) la selección aleatoria de los individuos a partir de una población y (2) la asignación aleatoria del tratamiento experimental a las personas, el valor esperado de la diferencia en los resultados entre los grupos de tratamiento y de control es el efecto causal promedio poblacional. Es decir, como se dijo en la Sección 3.5, el efecto causal sobre Y_i del tratamiento ($X_i = 1$) frente a ningún tratamiento ($X_i = 0$) es la diferencia en las esperanzas condicionales, $E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$, donde $E(Y_i|X_i = 1)$ y $E(Y_i|X_i = 0)$, son, respectivamente, los valores esperados de Y para los grupos de tratamiento y de control en un experimento aleatorizado controlado ideal. El Apéndice 13.3 proporciona un tratamiento matemático para los razonamientos anteriores.

En general, un efecto causal individual puede considerarse como dependiente tanto de las variables observables como de las variables no observables. Ya nos hemos encontrado con la idea de que un efecto causal puede depender de variables observables; por ejemplo, en el Capítulo 8 se examinaba la posibilidad de que el efecto de una reducción del tamaño de las clases podría depender de si un estudiante todavía está aprendiendo inglés. En la mayor parte de este capítulo, nos centraremos en el caso de que la variación en los efectos causales solo dependa de variables observables. La Sección 13.6 recoge la heterogeneidad no observable en los efectos causales.

Métodos econométricos para el análisis de datos experimentales

Los datos procedentes de un experimento aleatorizado controlado pueden ser analizados mediante la comparación de las diferencias en las medias o mediante una regresión que incluya la variable indicador del tratamiento y otras variables de control adicionales. Esta última especificación, el estimador de las diferencias con regresores adicionales, puede utilizarse asimismo en esquemas de aleatorización más complicados, en los que las probabilidades de aleatorización dependen de las covariables observables.

El estimador de las diferencias. El estimador de las diferencias es la diferencia en las medias muestrales entre los grupos de tratamiento y de control (Sección 3.5), que puede calcularse mediante la regresión de la variable respuesta Y sobre un indicador binario de tratamiento X :

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n. \quad (13.1)$$

Tal y como se trató en la Sección 4.4, si X se asigna al azar, entonces $E(u_i|X_i) = 0$, y el estimador MCO de β_1 en la Ecuación (13.1) es un estimador insesgado y consistente del efecto causal.

El estimador de las diferencias con variables explicativas adicionales. La eficiencia del estimador de las diferencias habitualmente puede mejorarse mediante la inclusión de algunas variables de control W en la regresión; esto conduce al **estimador de las diferencias con regresores adicionales**:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \dots + \beta_{1+r} W_{ri} + u_i, \quad i = 1, \dots, n. \quad (13.2)$$

Si W ayuda a explicar la variación en Y , entonces la inclusión de W reduce el error estándar de la regresión y, por lo general, el error estándar de $\hat{\beta}_1$. Como se analizó en la Sección 7.5 y en el Apéndice 7.2, para que el estimador $\hat{\beta}_1$ del efecto causal β_1 de la Ecuación (13.2) sea insesgado, las variables de control W deben ser tales que u_i satisfaga la condición de independencia en media condicional, es decir, $E(u_i|X_i, W_i) = E(u_i, W_i)$. Esta condición se cumple si W_i son características individuales pretratamiento, tales como el género: si W_i es una característica pretratamiento y X_i es asignado al azar, entonces X_i es independiente de u_i y W_i lo que implica que $E(u_i|X_i, W_i) = E(u_i|W_i)$. Los regresores W de la Ecuación (13.2) no deberían incluir los resultados del experimento (X_i no es asignado al azar, dado un resultado experimental). Como siempre sucede con las variables de control bajo la condición de independencia en media condicional, el coeficiente de la variable de control no tiene una interpretación causal.

Estimación de efectos causales que dependen de variables observables. Como se analizó en el Capítulo 8, la variación en los efectos causales que dependen de variables observables se puede estimar mediante la inclusión de funciones no lineales adecuadas de, o de interacciones con, X_i . Por ejemplo, si W_{1i} es un indicador binario que expresa el género, entonces los efectos causales distintos para hombres y mujeres se pueden estimar mediante la inclusión de la variable de interacción $W_{1i} \times X_i$ en la regresión de la Ecuación (13.2).

Aleatorización basada en las covariables. La aleatorización en la que la probabilidad de la asignación al grupo de tratamiento depende de una o de más variables observables W se denomina **aleatorización basada en las covariables**. Si la aleatorización se basa en covariables, entonces en general el estimador de las diferencias basado en la Ecuación (13.1) presenta sesgo de variable omitida. Por ejemplo, el Apéndice 7.2 describe un experimento hipotético para estimar el efecto causal de las tareas a realizar obligatorias frente a las optativas en un curso de econometría. En ese experimento, los estudiantes de economía ($W_i = 1$)

fueron asignados al grupo de tratamiento (tareas obligatorias, $X_i = 1$) con una probabilidad más alta que los estudiantes que provienen de otras titulaciones ($W_i = 0$). Pero si de todos modos los estudiantes de economía tienden a hacerlo mejor en la asignatura que los estudiantes de otras titulaciones, entonces existe un sesgo de variable omitida debido a que el hecho de estar en el grupo de tratamiento está correlacionado con la variable omitida, ser estudiante de economía.

Debido a que X_i fue asignado aleatoriamente dado W_i , este sesgo de variable omitida puede eliminarse utilizando el estimador de las diferencias con la variable de control adicional W_i . La asignación aleatoria de X_i dado W_i (combinada con el supuesto de función de regresión lineal) implica que, dados W_i , X_i , es independiente de u_i en la Ecuación (13.2). Esta independencia condicional a su vez implica independencia en media condicional, es decir, $E(u_i|X_i, W_i) = E(u_i|W_i)$. Por tanto, el estimador MCO $\hat{\beta}_1$ en la Ecuación (13.2) es un estimador insesgado del efecto causal cuando X_i se asigna aleatoriamente en base a W_i .

13.2 Amenazas a la validez de los experimentos

Recordemos del Concepto clave 9.1 que un estudio estadístico es *internamente válido* si las inferencias estadísticas acerca de los efectos causales son válidas para la población que está siendo estudiada; si es *externamente válido* sus inferencias y conclusiones pueden generalizarse desde la población y el escenario estudiados a otras poblaciones y escenarios. Varios de los problemas del mundo real plantean amenazas a la validez interna y externa de los análisis estadísticos de los experimentos reales con seres humanos.

Amenazas a la validez interna

Las amenazas a la validez interna de los experimentos aleatorizados controlados son la ausencia de aleatoriedad, el incumplimiento del protocolo de tratamiento, la deserción o abandono, los efectos experimentales y los tamaños muestrales pequeños.

Ausencia de aleatoriedad. Si el tratamiento no se asigna al azar, sino que se basa en parte en las características o preferencias del sujeto, entonces los resultados experimentales reflejarán tanto el efecto del tratamiento como el efecto de la asignación no aleatoria. Por ejemplo, supongamos que los participantes en un programa experimental de formación laboral son asignados al grupo de tratamiento en función de si su apellido pertenece a la primera o a la segunda mitad del alfabeto. Debido a las diferencias étnicas en los apellidos, el origen étnico puede diferir sistemáticamente entre los grupos de tratamiento y de control. En la medida en que la experiencia laboral, la educación, y otras características del mercado laboral difieran según la etnia, pueden existir diferencias sistemáticas entre los grupos de tratamiento y de control en estos factores omitidos que afectan los resultados. En general, la asignación no aleatoria puede conllevar una correlación entre X_i y u_i en las Ecuaciones (13.1) y (13.2), lo que a su vez lleva a un sesgo en el estimador del efecto del tratamiento.

Es posible contrastar la aleatoriedad. Si el tratamiento se recibe de forma aleatoria, entonces X_i no estará correlacionada con las variables observables de las características individuales pretratamiento W . Por lo tanto, un **contraste de la recepción aleatoria del tratamiento** implica contrastar la hipótesis de que los coeficientes de W_{1i}, \dots, W_{ri} son iguales a cero en una regresión de X_i sobre W_{1i}, \dots, W_{ri} . En el ejemplo del programa de formación laboral, se realiza la regresión de la variable recibir el curso de formación laboral (X_i) sobre el género, la etnia y la educación previa (W), y después se calcula el estadístico F para contrastar si los coeficientes de las W son iguales a cero, lo que constituye un contraste sobre la hipótesis nula de que el tratamiento se recibe aleatoriamente, frente a la hipótesis alternativa de que la recepción del tratamiento depende del sexo, la etnia o la educación previa. Si el diseño experimental lleva a cabo la aleatorización condicionada a las covariables, entonces esas covariables deberían incluirse en la regresión y el contraste F contrastaría los coeficientes de las W restantes¹.

¹ En este ejemplo, X_i es binaria, por lo que, como se analizó en el Capítulo 11, la regresión de X_i sobre W_{1i}, \dots, W_{ri} es un modelo de probabilidad lineal y los errores estándar heterocedástico-robustos son esenciales. Otra manera de contrastar la hipótesis de que $E(X_i|W_{1i}, \dots, W_{ri})$ no depende de W_{1i}, \dots, W_{ri} , cuando X_i es binaria consiste en utilizar un modelo probit o logit (véase la Sección 11.2).

Incumplimiento del protocolo de tratamiento. En un experimento real, las personas no siempre hacen lo que dicen. En un experimento de programa de formación laboral, por ejemplo, algunos de los sujetos asignados al grupo de tratamiento podrían no presentarse a las sesiones de formación y por lo tanto no recibir el tratamiento. Del mismo modo, los sujetos asignados al grupo de control de alguna manera podrían recibir la formación igualmente, tal vez haciendo una petición especial a un instructor o administrador.

El no cumplimiento de las personas acerca de seguir el protocolo del tratamiento aleatorio al detalle se denomina **cumplimiento parcial** del protocolo de tratamiento. En algunos casos, el experimentador sabe si el tratamiento es recibido realmente (por ejemplo, si se asistió a las clases), y el tratamiento realmente recibido se registra como X_i . Con cumplimiento parcial, existe un elemento de elección por parte del sujeto que recibe el tratamiento, por lo que X_i está correlacionada con u_i , incluso si al principio hubo asignación aleatoria. Por tanto, el incumplimiento del protocolo de tratamiento lleva a un sesgo en el estimador MCO.

Si se dispone de datos tanto del tratamiento realmente recibido (X_i) como de la asignación aleatoria inicial, entonces el efecto del tratamiento puede ser estimado mediante regresión de variables instrumentales. La **estimación por variables instrumentales del efecto del tratamiento** implica la estimación de la Ecuación (13.1) —o de la Ecuación (13.2) si hay variables de control— utilizando la asignación aleatoria inicial (Z_i) como un instrumento para el tratamiento efectivamente recibido (X_i). Hay que recordar que una variable debe cumplir las dos condiciones, relevancia del instrumento y exogeneidad del instrumento (Concepto clave 12.3), para ser una variable instrumental válida. Siempre y cuando el protocolo se siga parcialmente, entonces el nivel de tratamiento real estará parcialmente determinado por el nivel de tratamiento asignado, por lo que la variable instrumental Z_i es relevante. Si la asignación inicial es aleatoria, entonces Z_i se distribuye de forma independiente de u_i (condicionado a W_i , si la aleatoriedad es condicional a las covariadas), por lo que el instrumento es exógeno. Por tanto, en un experimento con un tratamiento asignado aleatoriamente, con cumplimiento parcial, y con datos sobre el tratamiento real, la asignación aleatoria original es una variable instrumental válida.

Esta estrategia de variables instrumentales requiere que se disponga de los datos tanto del tratamiento asignado como del tratamiento recibido. En algunos casos, podría ocurrir que los datos sobre el tratamiento efectivamente recibido no estuvieran disponibles. Por ejemplo, si a un sujeto en un experimento médico se le proporciona el medicamento pero, sin el conocimiento de los investigadores, simplemente no lo toma, entonces el tratamiento registrado («medicamento recibido») es incorrecto. La incorrecta medición del tratamiento recibido en realidad, lleva a un sesgo en el estimador de las diferencias.

Deserción o abandono. La **deserción** se refiere a los sujetos que abandonan el estudio después de haber sido aleatoriamente asignados al grupo de tratamiento o de control. A veces la deserción se produce por razones no relacionadas con el programa de tratamiento; por ejemplo, un participante en un estudio de formación laboral que tenga que salir de la ciudad para cuidar a un familiar enfermo. Pero si la razón de la deserción está relacionada con el tratamiento en sí mismo, entonces la deserción se traduce en la presencia de sesgo en el estimador MCO del efecto causal. Por ejemplo, supongamos que los alumnos más capaces abandonan el experimento del programa de formación laboral, debido a que obtienen puestos de trabajo fuera de la ciudad obtenidos debido a las habilidades de capacitación laboral adquiridas en el programa, por lo tanto al final del experimento solo los miembros menos capaces seguirán perteneciendo el grupo de tratamiento. Entonces la distribución de las características no medibles (capacidades) será diferente entre los grupos de control y tratamiento (el tratamiento permite a los participantes más capaces salir de la ciudad). En otras palabras, el tratamiento X_i estará correlacionado con u_i (que incluye las capacidades) para los que permanecen en la muestra al final del experimento y el estimador de las diferencias estará sesgado. Debido a que la deserción da lugar a una muestra seleccionada de forma no aleatoria, la deserción que está relacionada con el tratamiento lleva a un sesgo de selección (Concepto clave 9.4).

Efectos experimentales. En los experimentos cuyos sujetos son seres humanos, su comportamiento puede cambiar, simplemente debido al hecho de que los sujetos se encuentren en un experimento un fenómeno que a veces se conoce como **efecto Hawthorne** (véase el recuadro de la página siguiente).

El efecto Hawthorne

Entre los años 1920 y 1930, la Compañía General Electric llevó a cabo una serie de estudios acerca de la productividad de los trabajadores en su planta de Hawthorne. En un conjunto de experimentos, los investigadores cambiaron la potencia de las bombillas para ver cómo la iluminación afectaba a la productividad de las mujeres que se dedicaban al montaje de las componentes eléctricas. En otros experimentos se aumentaron o disminuyeron los períodos de descanso, se cambió el diseño de la sala de trabajo y se recortaron las jornadas laborales. Los primeros resultados influyentes presentados sobre estos estudios concluyeron que la productividad seguía aumentando tanto si las luces eran más tenues como si eran más intensas, tanto si las jornadas laborales eran más largas como si eran más cortas, o si las condiciones mejoraban o empeoraban. Los investigadores concluyeron que las mejoras de productividad no eran consecuencia de los cambios en el lugar

de trabajo, sino que se producían debido al papel particular desempeñado por los trabajadores en el experimento debido a que se sentían observados y valorados, por lo que trabajaban más y mejor. Con los años, la idea de que el hecho de participar en un experimento influye sobre el comportamiento ha llegado a conocerse como el efecto Hawthorne.

Pero hay un fallo en esta historia: un examen cuidadoso de los datos reales de Hawthorne revela que ¡no hay efecto Hawthorne! (Gillespie, 1991; Jones, 1992). Sin embargo, en algunos experimentos, especialmente aquellos en los que los sujetos tienen un interés en el resultado, el simple hecho de estar en un experimento podría afectar al comportamiento. El efecto Hawthorne y los efectos experimentales en general pueden representar una amenaza a la validez interna, incluso a pesar de que el efecto Hawthorne no sea evidente en los datos originales de Hawthorne.

En algunos experimentos, un protocolo «doble ciego» puede mitigar el efecto de estar en un experimento: a pesar de que tanto los sujetos como los experimentadores saben que están en un experimento, ninguno sabe si un sujeto concreto está en el grupo de tratamiento o en el grupo de control. Por ejemplo, en un experimento médico sobre un medicamento, en ocasiones puede hacerse que el medicamento y el placebo parezcan iguales para que ni el profesional médico que dispensa el medicamento ni el paciente sepan si la medicina administrada es la real o el placebo. Si el experimento es doble ciego, tanto el grupo de tratamiento como el de control deben experimentar los mismos efectos experimentales, y por tanto las diferencias en los resultados entre los dos grupos podrán atribuirse al medicamento.

Los experimentos con doble ciego son claramente inviables en economía en la realidad: tanto el sujeto experimental como el instructor saben si el sujeto está asistiendo al programa de formación laboral. En un experimento mal diseñado, este efecto experimental podría ser considerable. Por ejemplo, los profesores de un programa experimental podría tratar de hacer de forma especialmente intensa que el programa sea un éxito si piensan que sus empleos en el futuro dependerán del resultado del experimento. La decisión acerca de si los resultados experimentales son sesgados debido a la presencia de efectos experimentales requiere de valoraciones basadas en los detalles acerca de cómo se llevó a cabo el experimento.

Muestras pequeñas. Debido a que los experimentos con seres humanos pueden resultar costosos, a veces el tamaño de la muestra es pequeño. Un tamaño de muestra pequeño no sesga los estimadores del efecto causal, pero sí supone que el efecto causal se estime de forma imprecisa. Una muestra pequeña plantea asimismo amenazas a la validez de los intervalos de confianza y los contrastes de hipótesis. Debido a que la inferencia basada en los valores críticos normales y los errores estándar heterocedástico-robustos se justifica en la utilización de aproximaciones para muestras grandes, los datos experimentales con muestras pequeñas a veces son analizados bajo el supuesto de que los errores se distribuyen normalmente (Secciones 3.6 y 5.6); sin embargo, el cumplimiento del supuesto de normalidad suele ser dudoso tanto para los datos experimentales como para los datos observacionales.

Amenazas a la validez externa

Las amenazas a la validez externa comprometen la capacidad de generalizar los resultados del estudio a otras poblaciones y entornos. Existen dos amenazas de este tipo: cuando la muestra experimental no es representativa de la población de interés y cuando el tratamiento objeto del estudio no es representativo del tratamiento que se llevaría a cabo de manera más amplia.

Muestra no representativa. La población estudiada y la población de interés deben ser lo suficientemente similares como para justificar la generalización de los resultados experimentales. Si un programa de formación laboral es evaluado mediante un experimento con exreclusos, entonces podría ser posible generalizar los resultados del estudio a otros exreclusos. No obstante, debido a que los antecedentes penales pesan en las mentes de los potenciales empleadores, los resultados podrían no ser generalizables a los trabajadores que nunca han cometido un delito.

Otro ejemplo de una muestra no representativa puede surgir cuando los participantes en el grupo experimental son voluntarios. Incluso si los voluntarios son asignados aleatoriamente a los grupos de tratamiento y control, estos voluntarios podrían estar más motivados que la población en general y, para ellos, el tratamiento podría tener un efecto mayor. En términos más generales, la selección no aleatoria de la muestra a partir de la población de interés puede comprometer la capacidad de generalizar los resultados de la población estudiada (como los voluntarios) a la población de interés.

Programa o política no representativa. La política o el programa de interés debe ser del mismo modo lo suficientemente similar al programa estudiado como para permitir la generalización de los resultados. Una característica importante es que el programa a pequeña escala, con un estrecho seguimiento del experimento podría ser muy diferente del programa llevado a cabo realmente. Si el programa realmente llevado a cabo es extenso, entonces el programa a mayor escala podría no reportar el mismo control de calidad que la versión experimental o podría estar dotado de una financiación a menor escala; ambas posibilidades podrían dar lugar a un programa a escala total que sea menos eficaz que el programa experimental más pequeño. Otra diferencia entre un programa experimental y un programa real es su duración: el programa experimental solo permanece mientras dura el experimento, mientras que el programa en cuestión en la realidad puede estar disponible durante largos períodos de tiempo.

Efectos de equilibrio general. Una cuestión relacionada con la escala y la duración implica a lo que los economistas llaman los efectos de «equilibrio general». Convertir un programa experimental pequeño y temporal en un programa experimental amplio y permanente puede cambiar el entorno económico lo suficiente como para que los resultados del experimento no se puedan generalizar. Un pequeño programa experimental de formación laboral, por ejemplo, puede complementar la capacitación proporcionada por parte de los empleadores, pero si el programa se amplía mucho, podría desplazar a la formación impartida por el empleador, lo cual reduciría los beneficios netos del programa. Del mismo modo, una amplia reforma educativa, como la de ofrecer escuelas concertadas o la de una reducción sustancial de los alumnos por clase, podría aumentar la demanda de profesores y cambiar el tipo de persona a la que le gusta la enseñanza, por lo que el eventual efecto neto de la reforma amplia podría reflejar estos cambios inducidos en el personal de la escuela. Expresado en términos estadísticos, un pequeño experimento internamente válido podría medir correctamente un efecto causal, manteniendo constantes las circunstancias del mercado o el escenario de las políticas, pero los efectos de equilibrio general significan que esos otros factores, de hecho, no se mantienen constantes cuando se lleva a cabo el programa de forma generalizada.

13.3 Estimaciones experimentales del efecto de la reducción del tamaño de las clases

En este apartado volvemos a la pregunta formulada en la Parte II: ¿Cuál es el efecto sobre las calificaciones en los exámenes de la reducción del tamaño de las clases en los primeros cursos? A finales de la década de 1980, Tennessee llevó a cabo un gran y multimillonario experimento aleatorizado controlado para determinar si la reducción del tamaño de las clases era una forma eficaz de mejorar la educación primaria. Los resultados de este experimento han influido mucho en nuestra comprensión acerca del efecto de la reducción del tamaño de las clases.

Diseño experimental

El experimento de reducción del tamaño de las clases de Tennessee, conocido como Proyecto STAR (Student-Teacher Achievement Ratio, en sus siglas originales en inglés), fue un experimento de cuatro años diseñado para evaluar el efecto en el aprendizaje de las clases con pequeño tamaño. Financiado por el parlamento del estado de Tennessee, el experimento tuvo un coste aproximado de 12 millones de dólares. El estudio comparaba tres tipos diferentes de clases de jardín de infancia hasta el tercer curso: un tamaño de clase mediano, con un número de estudiantes por clase de 22 a 25, un único maestro, sin ayudantes; una clase pequeña, con un número de alumnos por clase de 13 a 17 y sin ayuda; y una clase de tamaño mediano, más un profesor ayudante.

Cada escuela participante en el experimento tenía al menos una clase de cada tipo, y los estudiantes que ingresaron en el jardín de infancia de una escuela participante fueron asignados aleatoriamente a uno de estos tres grupos al comienzo del curso académico 1985-1986. Los maestros fueron asignados del mismo modo aleatoriamente a uno de los tres tipos de clases.

De acuerdo con el protocolo experimental original, los estudiantes debían permanecer en su tipo de clase asignado inicialmente durante los 4 años del experimento (jardín de infancia hasta el tercer curso). Sin embargo, debido a las quejas de los padres, los estudiantes inicialmente asignados a una clase mediana (con o sin un ayudante) fueron reasignados aleatoriamente al comienzo del primer curso a las clases medianas con un ayudante o a clases medianas sin ayudante; los estudiantes inicialmente asignados a una clase pequeña permanecieron en una clase pequeña. Los estudiantes que ingresaron a la escuela en primer curso (el jardín de infancia era opcional), durante el segundo año del experimento, fueron asignados aleatoriamente a uno de los tres grupos. Cada año, los estudiantes del experimento fueron sometidos a pruebas estandarizadas (el examen de aprovechamiento de Stanford) de lectura y matemáticas.

El proyecto pagaba a los profesores y los ayudantes adicionales necesarios para alcanzar el objetivo del tamaño de las clases. Durante el primer año del estudio, aproximadamente 6.400 estudiantes participaron en 108 clases pequeñas, 101 clases medianas, y 99 clases medianas con ayudantes. Durante los 4 años del estudio, un total de aproximadamente 11.600 estudiantes en 80 escuelas participaron en el estudio.

Desviaciones respecto del diseño experimental. El protocolo experimental especificaba que los estudiantes no debían cambiar entre los distintos tipos de clases, a no ser mediante reasignación aleatoria al comienzo del primer curso. Sin embargo, aproximadamente el 10 % de los estudiantes cambió en los años siguientes, entre otras razones por incompatibilidades entre los niños y problemas de comportamiento. Estos cambios representan una desviación respecto del esquema de asignación al azar y, en función de la verdadera naturaleza de los cambios, tienen el potencial de introducir un sesgo en los resultados. Los cambios exclusivamente realizados para evitar los conflictos de personalidad podrían estar lo suficientemente incorrelacionados con el experimento como para no introducir un sesgo. Sin embargo, si los cambios surgieron porque los padres más preocupados por la educación de sus hijos presionaron a la escuela para que cambiara a un niño a una clase pequeña, entonces el no seguimiento del protocolo experimental podría sesgar los resultados al exagerar la eficacia de las clases más pequeñas. Otra desviación del protocolo experimental fue que el tamaño de las clases cambió con el tiempo debido a que los estudiantes cambiaban entre las clases y abandonaban el distrito escolar o ingresaban en él.

Análisis de los datos STAR

Debido a que hay dos grupos de tratamiento —clases pequeñas y clases medianas con ayudante— la versión de la regresión del estimador de las diferencias tiene que ser modificada para manejar los dos grupos de tratamiento y el grupo de control. Esta modificación se realiza mediante la introducción de dos variables binarias, una que indica si el estudiante está en una clase pequeña y otra que indica si el alumno pertenece a una clase de tamaño mediano con un ayudante, lo que lleva al modelo de regresión poblacional

$$Y_i = \beta_0 + \beta_1 ClasePequeña_i + \beta_2 MedAyuda_i + u_i \quad (13.3)$$

donde Y_i es la calificación en la prueba, $ClasePequeña_i = 1$ si el estudiante i -ésimo pertenece a una clase pequeña e igual a 0 de lo contrario, y $MedAyuda_i = 1$ si el estudiante i -ésimo está en una clase mediana con un ayudante y 0 de lo contrario. El efecto sobre la calificación en el examen de una clase pequeña, en relación con una clase mediana, es β_1 , y el efecto de una clase mediana con un ayudante, en relación con una clase mediana, es β_2 . El estimador de las diferencias en el experimento se puede calcular mediante la estimación de β_1 y β_2 en la Ecuación (13.3) por MCO.

La Tabla 13.1 presenta las estimaciones de las diferencias del efecto sobre las calificaciones de estar en una clase pequeña o en una clase de tamaño mediano con un ayudante. La variable dependiente Y_i en las regresiones de la Tabla 13.1 es la calificación total del estudiante en los apartados combinados de matemáticas y lectura del examen de Stanford². De acuerdo con las estimaciones de la Tabla 13.1, para los alumnos del jardín de infancia, el efecto de estar en una clase pequeña es el de un aumento de 13,9 puntos en el examen, en comparación con estar en una clase mediana; el efecto estimado de estar en una clase mediana con ayudante es de 0,31 puntos en el examen. Para cada curso, la hipótesis nula de que las clases pequeñas no ofrecen ninguna mejora se rechaza al nivel de significación del 1 % (bilateral). Sin embargo, no es posible rechazar la hipótesis nula de que tener a un asistente en una clase mediana no proporciona mejora, en comparación con no tener a un asistente, excepto en el primer curso. Las magnitudes estimadas de las mejoras de las clases pequeñas son muy similares en los cursos JI, 2, y 3, aunque la estimación es mayor para el primer curso.

Las estimaciones de las diferencias en la Tabla 13.1 indican que la reducción del tamaño de las clases tiene efecto sobre el rendimiento en la prueba, pero la adición de un asistente a una clase de tamaño mediano tiene un efecto mucho menor, posiblemente igual a cero. Tal y como se ha visto en la Sección 13.1, con la ampliación de las regresiones de la Tabla 13.1 mediante regresores adicionales —los regresores W de la Ecuación (13.2)— se pueden obtener estimaciones más eficientes de los efectos causales. Por otra parte, si el tratamiento recibido no es aleatorio, debido a los incumplimientos del protocolo de tratamiento, entonces las estimaciones de los efectos experimentales basadas en las regresiones con regresores adicionales podrían ser distintas a las estimaciones de las diferencias presentadas en la Tabla 13.1. Por estas dos razones, las estimaciones de los efectos experimentales en las que se han incluido variables explicativas adicionales en la Ecuación (13.3) se presentan en la Tabla 13.2 para el jardín de infancia; la primera columna de la Tabla 13.2 repite los resultados de la primera columna (para el jardín de infancia) de la Tabla 13.1, y las tres

TABLA 13.1 Proyecto STAR: estimaciones de las diferencias del efecto sobre las calificaciones en los exámenes estandarizados del tamaño de las clases en el grupo de tratamiento

Regresor	Curso			
	JI	(1)	(2)	(3)
Clase pequeña	13,90** (2,45)	29,78** (8,83)	19,39** (2,71)	15,59** (2,40)
Tamaño mediano con ayudante	0,31 (2,27)	11,96** (2,65)	3,48 (5,54)	-0,29 (2,27)
Intercepto	918,04** (1,63)	1.039,39** (1,78)	1.157,81** (1,82)	1.228,51** (1,68)
Número de observaciones	5.786	6.379	6.049	5.967

Las regresiones se estimaron utilizando la base de datos del proyecto STAR de acceso público que se describe en el Apéndice 13.1. La variable dependiente es la puntuación combinada obtenida por los estudiantes en los apartados de matemáticas y lectura del examen de Stanford. Los errores estándar aparecen entre paréntesis bajo los coeficientes. **El coeficiente individual es estadísticamente significativo al nivel de significación del 1 % mediante un contraste bilateral.

² N. Del T.: Stanford Achievement Test.

TABLA 13.2 Proyecto STAR: estimaciones de las diferencias con regresores adicionales para el jardín de infancia

Regresor	(1)	(2)	(3)	(4)
Clase pequeña	13,90** (2,45)	14,00** (2,45)	15,93** (2,24)	15,59** (2,40)
Tamaño mediano con ayudante	0,31 (2,27)	-0,60 (2,25)	1,22 (2,04)	-0,29 (2,27)
Años de experiencia del maestro		1,47** (0,17)	0,74** (0,17)	1.228,51** (1,68)
Niño				-12,09** (1,67)
Cumplir requisitos para comedor gratuito				-34,70** (1,99)
Negro				-25,43** (3,50)
Etnia distinta de blanco o negro				-8,50** (12,52)
Intercepto	918,04** (1,63)	904,72** (2,22)		
Variables indicador de la escuela?	no	no	si	si
\bar{R}^2	0,01	0,02	0,22	0,28
Número de observaciones	5.786	5.766	5.766	5.748

Las regresiones se estimaron utilizando la base de datos del proyecto STAR de acceso público que se describe en el Apéndice 13.1. La variable dependiente es la puntuación combinada obtenida por los estudiantes en los apartados de matemáticas y lectura del examen de Stanford. El número de observaciones difiere entre las distintas regresiones, debido a que faltan algunos datos. Los errores estándar aparecen entre paréntesis debajo de los coeficientes. El coeficiente individual es estadísticamente significativo al 5 %* o al 1 %** de nivel de significación en un contraste bilateral.

restantes columnas incluyen regresores adicionales que miden las características del maestro, de la escuela y del estudiante.

La principal conclusión de la Tabla 13.2 es que las estimaciones de los efectos causales mediante regresión múltiple de los dos tratamientos (clase pequeña y clase de tamaño mediano con ayudante) en las tres últimas columnas de la Tabla 13.2 son similares a las diferencias estimadas presentadas en la primera columna. El hecho de que la adición de estos regresores observables no cambie los efectos causales estimados de los diferentes tratamientos hace que sea más verosímil que la asignación aleatoria a las clases más pequeñas tampoco dependa de las variables inobservables. Como era de esperar, estos regresores adicionales incrementan el \bar{R}^2 de la regresión, y el error estándar del efecto estimado del tamaño de las clases disminuye desde 2,45 en la columna (1) hasta 2,16 en la columna (4).

Debido a que los maestros fueron asignados al azar a los distintos tipos de clases dentro de una escuela, el experimento ofrece asimismo una oportunidad de estimar el efecto sobre las calificaciones en el examen de la experiencia de los maestros. En la terminología de la Sección 13.1, la aleatorización está condicionada a las covariables W , donde W expresa un conjunto completo de variables binarias indicativas de cada escuela; es decir, W expresa un conjunto completo de efectos fijos individuales de la escuela. Por tanto, condicionado a W , los años de experiencia son asignados al azar, lo que a su vez implica que u_i en la Ecuación (13.2) satisface la independencia en media condicional, donde las variables X son los tratamientos del tamaño de las clases y de los años de experiencia del maestro y W es el conjunto completo de efectos fijos individuales de la escuela. Debido a que los maestros no fueron asignados al azar entre las escuelas, en la regresión sin efectos fijos individuales de la escuela [Tabla 13.2, columna (2)] los años de experiencia, en general, esta-

rán correlacionados con el término de error, por ejemplo, los distritos más ricos podrían tener profesores con más años de experiencia. Cuando se incluyen los efectos fijos, el coeficiente de la experiencia estimado se reduce a la mitad, desde el 1,47 en la columna (2) de la Tabla 13.2 hasta el 0,74 en la columna (3). Debido a que los maestros fueron asignados al azar dentro de la escuela, la columna (3) ofrece un estimador insesgado del efecto sobre los resultados en el examen de un año adicional de experiencia. La estimación, 0,74, es estadísticamente significativa y moderadamente grande: a diez años de experiencia les corresponde un aumento previsto en las calificaciones de 7,4 puntos.

Resulta tentador interpretar algunos de los otros coeficientes de la Tabla 13.2, pero, al igual que los coeficientes de las variables de control en general, los coeficientes no tienen una interpretación causal. Por ejemplo, los niños de preescolar obtienen peores resultados que las niñas en estas pruebas estandarizadas. Sin embargo, estas características individuales de los estudiantes *no* están asignadas aleatoriamente (¡el género del estudiante que realiza el examen *no* se asigna aleatoriamente!), por lo que estos regresores adicionales podrían estar correlacionados con las variables omitidas. Del mismo modo, si la pertenencia a una etnia o el hecho de cumplir los requisitos para recibir una ayuda de comedor gratuito están correlacionados con una reducción de las oportunidades de aprendizaje fuera de la escuela (que se omite en las regresiones de la Tabla 13.2), entonces sus coeficientes estimados podrían reflejar estas influencias omitidas.

Interpretación de los efectos estimados del tamaño de las clases. ¿Son grandes o pequeños, en un sentido práctico, los efectos estimados del tamaño de las clases presentados en las Tablas 13.1 y 13.2? Existen dos maneras de responder a esto: en primer lugar, mediante la traducción de las variaciones estimadas en las calificaciones de la prueba a unidades de desviaciones típicas de las calificaciones en la prueba, por lo que las estimaciones de la Tabla 13.1 serán comparables entre los distintos cursos; y, en segundo lugar, mediante la comparación de la estimación del efecto del tamaño de las clases con los otros coeficientes de la Tabla 13.2.

Debido a que la distribución de las calificaciones en las pruebas no es la misma para cada curso, los efectos estimados de la Tabla 13.1 no son directamente comparables entre los diferentes cursos. Abordamos este problema en la Sección 9.4, cuando quisimos comparar el efecto sobre las calificaciones en los exámenes de una reducción de la ratio estudiantes-maestros estimada utilizando los datos de California con las estimaciones basadas en los datos de Massachusetts. Debido a que las dos pruebas difieren, los coeficientes no se pueden comparar directamente. La solución en la Sección 9.4 consistía en traducir los efectos estimados a unidades de desviaciones típicas de la prueba de modo que una unidad de disminución de la ratio estudiantes-maestros se corresponde con un cambio en la cuantía de alguna proporción estimada de una desviación típica en las calificaciones. Aquí adoptamos este método por lo que los efectos estimados de la Tabla 13.1 se pueden comparar entre los distintos cursos. Por ejemplo, la desviación típica de las calificaciones en los exámenes para niños de jardín de infancia es 73,7, por lo que el efecto de estar en una clase pequeña de preescolar, en base a la estimación de la Tabla 13.1, es $13,9/73,7 = 0,19$, con un error estándar de $2,45/73,7 = 0,03$. Los efectos estimados del tamaño de las clases en la Tabla 13.1, convertidos en unidades de desviación típica de las calificaciones en el examen entre estudiantes, se cogen de la Tabla 13.3. Expresado en unidades de desviación típica, el efecto estimado de estar en una clase pequeña es similar para los cursos JI, 2, y 3, y es de aproximadamente una quinta parte de la desviación típica de las calificaciones en los exámenes. Del mismo modo, el resultado de estar en una clase de tamaño mediano con un ayudante es de aproximadamente cero para los cursos JI, 2 y 3. Los efectos estimados del tratamiento son mayores en el primer curso; sin embargo, la diferencia estimada entre las clases pequeñas y de tamaño mediano con un ayudante es de 0,20 para el primer curso, la misma que para el resto de los otros cursos. Por tanto, una interpretación de los resultados para el primer curso es que los estudiantes en el grupo de control —clase de tamaño mediano sin ayudante— pasaron a hacerlo peor en la prueba de ese año por algún motivo inusual, tal vez simplemente debido a la variación aleatoria muestral.

Otra forma de medir la magnitud del efecto estimado de estar en una clase pequeña consiste en comparar los efectos estimados del tratamiento con los otros coeficientes de la Tabla 13.2. En preescolar, el efecto estimado de estar en una clase pequeña es de 13,9 puntos en el examen (la primera fila de la Tabla 13.2). Manteniendo constantes la etnia, los años de experiencia del maestro, el hecho de cumplir los requisitos para comedor gratuito, y el grupo de tratamiento, los niños obtienen una menor puntuación en el examen

TABLA 13.3 Efectos estimados del tamaño de las clases en unidades de desviación típica de las calificaciones en las pruebas entre estudiantes

Grupo de tratamiento	Curso			
	JI	(1)	(2)	(3)
Clase pequeña	0,19** (0,03)	0,33** (0,03)	0,23** (0,03)	0,21** (0,03)
Tamaño mediano con ayudante	0,00 (0,03)	0,13** (0,03)	0,04 (0,03)	0,00 (0,03)
Desviación típica muestral de las calificaciones (s_y)	73,70	91,30	84,10	73,30

Las estimaciones y los errores estándar de las dos primeras filas son los efectos estimados de la Tabla 13.1, divididos por la desviación típica de la muestra en el examen de Stanford para ese curso (la última fila de esta tabla), calculados utilizando los datos de los estudiantes del experimento. Los errores estándar se presentan entre paréntesis debajo de los coeficientes. ** El coeficiente individual es estadísticamente significativo al nivel de significación del 1 % en un contraste bilateral.

estandarizado que las niñas, en aproximadamente 12 puntos de acuerdo con las estimaciones de la columna (4) de la Tabla 13.2. Por tanto, el efecto estimado de estar en una clase pequeña es algo mayor que la diferencia en el resultado entre niños y niñas. Otra comparación posible es la del coeficiente estimado de los años de experiencia del maestro que en la columna (4) es de 0,66, por lo que tener un maestro con 20 años de experiencia se estima que mejora el rendimiento en la prueba en 13 puntos. Por tanto, el efecto estimado de estar en una clase pequeña es aproximadamente el mismo que el efecto de tener un maestro veterano de 20 años de experiencia, respecto a tener un maestro novel. Estas comparaciones sugieren que el efecto estimado de estar en una clase pequeña es importante.

Otros resultados adicionales. Los económetras, estadísticos y especialistas en educación primaria han estudiado ampliamente este experimento, del que aquí se ofrece un breve resumen acerca de algunos de los resultados. Uno de ellos es que el efecto de estar en una clase pequeña se concentra en los cursos tempranos, como se puede ver en la Tabla 13.3; con excepción de los resultados anómalos para el primer curso, la brecha en las calificaciones en la prueba entre las clases medianas y pequeñas presentada en la Tabla 13.3 es esencialmente constante entre los diferentes cursos (0,19 unidades de desviación típica en preescolar, 0,23 en el segundo curso, y de 0,21 en el tercer curso). Debido a que los niños inicialmente asignados a una clase pequeña permanecieron en esa misma clase pequeña, permanecer en una clase pequeña no se tradujo en ganancias adicionales; por el contrario, los logros alcanzados sobre la asignación inicial se mantuvieron en los cursos superiores, pero la brecha entre los grupos de tratamiento y de control no aumentó. Otra conclusión es que, como se indica en la segunda fila de la Tabla 13.3, este experimento muestra un escaso beneficio de tener a un ayudante en un aula de tamaño mediano. Un asunto que puede afectar potencialmente a la interpretación de los resultados del experimento es el incumplimiento del protocolo de tratamiento por parte de algunos estudiantes (algunos estudiantes cambiaron desde los grupos pequeños). Si la asignación inicial en un aula de preescolar es aleatoria y no ha tenido ningún efecto directo en las calificaciones de la prueba, entonces la asignación inicial puede ser utilizada como variable instrumental que en parte, pero no totalmente, influye en la asignación definitiva. Esta estrategia fue adoptada por Krueger (1999), quien utilizó mínimos cuadrados en dos etapas (MC2E) para estimar el efecto sobre las calificaciones del tamaño de las clases utilizando las asignaciones iniciales a las aulas como variable instrumental; halló que las estimaciones MC2E y MCO eran similares, lo cual le llevó a la conclusión de que las desviaciones respecto del protocolo experimental no implicaban un sesgo importante en la estimaciones MCO³.

³ Para obtener más información sobre el proyecto STAR, véase Mosteller (1995), Mosteller, Light, y Sachs (1996), y Krueger (1999). Ehrenberg, Brewer, Gamoran, y Willms (2001a, 2001b) analizan el proyecto STAR y lo ubican en el contexto del debate político sobre el tamaño de las clases y la investigación relacionada con el asunto. Para ver algunas críticas al proyecto STAR, véase Hanushek (1999a), y para una visión crítica de la relación entre el tamaño de las clases y el rendimiento en general, véase Hanushek (1999b).

Comparación de las estimaciones observacionales y experimentales de los efectos del tamaño de las clases

La Parte II presenta las estimaciones de regresión múltiple del efecto del tamaño de las clases basadas en los datos observacionales de los distritos escolares de California y de Massachusetts. En estos datos, el tamaño de las clases *no* se asignó aleatoriamente, sino que fue determinado por funcionarios de las escuelas locales tratando de equilibrar los objetivos de la educación con las realidades presupuestarias. ¿Cómo se pueden comparar las estimaciones observacionales con las estimaciones experimentales del proyecto STAR?

Para comparar las estimaciones de California y Massachusetts con las de la Tabla 13.3, es necesario evaluar la misma reducción del tamaño de las clases y expresar la predicción del efecto en unidades de desviaciones típicas sobre las calificaciones en las pruebas. Durante los 4 años del experimento STAR, las clases pequeñas tuvieron, en promedio, aproximadamente 7,5 menos estudiantes que las clases grandes, por lo que se utilizan las estimaciones observacionales para predecir el efecto sobre las calificaciones en las pruebas de una reducción de 7,5 alumnos por clase. En base a las estimaciones MCO de las especificaciones lineales que se resumen en la primera columna de la Tabla 9.3, las estimaciones para California predicen un aumento de 5,5 puntos en el examen para una reducción en la ratio estudiantes-maestros en 7,5 estudiantes ($0,73 \times 7,5 \cong 5,5$ puntos). La desviación típica en el examen entre los estudiantes en California es de aproximadamente 38 puntos, por lo que el efecto estimado de una reducción en 7,5 alumnos, expresada en unidades de desviaciones típicas entre los estudiantes, es de $5,5/38 \cong 0,14$ desviaciones típicas⁴. El error estándar del coeficiente de la pendiente estimado para California es 0,26 (Tabla 9.3), por lo que el error estándar del efecto estimado de una reducción de 7,5 estudiantes en unidades de desviaciones típicas es de $0,26 \times 7,5/38 \cong 0,05$. Por lo tanto, en base a los datos de California, los efectos estimados de la reducción del tamaño de las clases en 7,5 estudiantes, expresada en unidades de desviación típica de las calificaciones en los exámenes entre estudiantes, es de 0,14 desviaciones típicas, con un error estándar de 0,05. Estos cálculos y otros cálculos similares para Massachusetts se resumen en la Tabla 13.4, junto con las estimaciones STAR para jardín de infancia tomados de la columna (1) de la Tabla 13.2.

Los efectos estimados para los estudios observacionales de California y Massachusetts son algo menores que las estimaciones con los datos STAR. No obstante, una de las razones de que las estimaciones de los diferentes estudios difieran, es la variabilidad del muestreo aleatorio, por lo que tiene sentido comparar los intervalos de confianza para los efectos estimados de los tres estudios. En base a los datos de STAR para jardín de infancia, el intervalo de confianza al 95 % para el efecto de estar en una clase pequeña (presentado en la última columna de la Tabla 13.4) va desde 0,13 hasta 0,25. El intervalo de confianza al 95 % comparable en base a los datos observacionales de California va desde 0,04 a 0,24; y para Massachusetts es desde 0,02 hasta 0,22. Por tanto, los intervalos de confianza al 95 % para los estudios de California y Massachusetts contienen la mayor parte del intervalo de confianza al 95 % para los datos de jardín de infancia de STAR. Visto de esta manera, los tres estudios sorprendentemente proporcionan rangos similares para las estimaciones.

Existen muchas razones por las que las estimaciones experimentales y observacionales podrían diferir. Una razón es que, tal y como se discutió en la Sección 9.4, siguen quedando amenazas a la validez interna de los estudios observacionales. Por ejemplo, debido a que los niños se trasladan entre los distintos distritos, la ratio estudiantes-maestros del distrito podría no reflejar la ratio estudiantes-maestros en realidad experimentada por los estudiantes, por lo que el coeficiente de la ratio estudiantes-maestros en los estudios de Massachusetts y California podrían estar sesgados hacia cero debido a la presencia de sesgo de errores en las variables. Otras razones comprometen la validez externa. La ratio estudiantes maestros promedio del distrito utilizada en los estudios observacionales no es la misma que el número real de niños por clase, la variable experimental STAR. El Proyecto STAR se desarrolló en un estado del sur en la década de

⁴ En la Tabla 9.3, los efectos estimados se presentan en términos de desviaciones típicas de las calificaciones en los exámenes entre distritos; en la Tabla 13.3, los efectos estimados están expresados en términos de desviaciones típicas de las calificaciones en los exámenes entre estudiantes. La desviación típica entre los estudiantes es mayor que la desviación típica entre los distritos. Para California, la desviación típica entre estudiantes es de 38, pero la desviación típica entre los distritos es de 19,1.

TABLA 13.4

Efectos estimados de la reducción de la ratio estudiantes-maestros en 7,5 en base a los datos STAR y los datos observacionales de California y Massachusetts

Estudio	β_1	Variación en la ratio estudiantes-maestros	Desviación típica de las calificaciones entre estudiantes	Efecto estimado	Intervalo de confianza al 95%
STAR (Curso JI)	-13,90** (2,45)	Clase pequeña vs. Clase mediana	73,8	-0,19** (2,45)	(0,13, 0,25)
California	-0,73** (0,26)	-7,5	38,0	-0,14** (0,05)	(0,04, 0,24)
Massachusetts	-0,64** (0,27)	-7,5	39,0	-0,12** (0,05)	(0,02, 0,22)

El coeficiente estimado $\hat{\beta}_1$ para el estudio STAR se ha tomado de la columna (1) de la Tabla 13.2. Los coeficientes estimados de los estudios de California y Massachusetts se toman de la primera columna de la Tabla 9.3. El efecto estimado es el efecto de estar en un clase pequeña frente a una clase mediana (para STAR) o el efecto de la reducción de la ratio estudiantes-maestros en 7,5 (para los estudios de California y Massachusetts). El intervalo de confianza al 95% para la reducción de la ratio estudiantes-maestros es el efecto estimado \pm 1,96 errores estándar. Los errores estándar se presentan entre paréntesis debajo de los efectos estimados. Los efectos estimados son estadística y significativamente distintos de cero *al nivel del 5 % o **al nivel de significación del 1% mediante un contraste bilateral

1980, un contexto potencialmente diferente al de California y Massachusetts en 1998, y los cursos que se comparan son diferentes (del JI hasta 3.^º en el estudio STAR, cuarto curso en Massachusetts, quinto curso en California). A la luz de todas estas razones que existen para esperar distintas estimaciones, los resultados de los tres estudios son notablemente similares. El hecho de que los estudios observacionales sean similares a las estimaciones del proyecto STAR sugiere que las restantes amenazas a la validez interna de la estimaciones observacionales son de menor importancia.

13.4 Cuasi experimentos

Las intuiciones y los métodos estadísticos de los experimentos aleatorizados controlados pueden trasladarse a marcos no experimentales. En un **cuasi experimento**, asimismo conocido como **experimento natural**, la aleatoriedad se introduce mediante las variaciones en las circunstancias individuales que hacen que parezca *como si* el tratamiento se asignara aleatoriamente. Estas variaciones en las circunstancias individuales pueden surgir a causa de los caprichos de las instituciones jurídicas, la ubicación, el calendario de aplicación de las políticas o de los programas, la aleatoriedad natural como las fechas de nacimiento, la lluvia u otros factores no relacionados con el efecto causal a estudio.

Existen dos tipos de cuasi experimentos. En el primero de ellos, el hecho de que un individuo (o en general, una entidad individual) reciba el tratamiento se interpreta como si se hubiera determinado aleatoriamente. En este caso, el efecto causal puede ser estimado por MCO utilizando el tratamiento, X_i , como regresor. En el segundo tipo de cuasi experimento, la variación «*como si fuera*» aleatoria determina solo en parte el tratamiento. En este caso, el efecto causal se estima mediante regresión de variables instrumentales, donde la fuente de variación «*como si fuera*» aleatoria proporciona la variable instrumental.

Después de ofrecer algunos ejemplos, esta sección presenta algunas extensiones de los métodos econométricos de las Secciones 13.1 y 13.2 que pueden resultar útiles para analizar los datos de los cuasi experimentos.

Ejemplos

Se ilustran los dos tipos de cuasi experimentos mediante ejemplos. El primer ejemplo es un cuasi experimento en el que el tratamiento es «*como si fuera*» aleatoriamente determinado. El segundo y tercer ejem-

plos ilustran los cuasi experimentos en los que la variación «como si fuera» aleatoria influye, pero no determina totalmente, el nivel de tratamiento.

Ejemplo #1: los efectos de la inmigración en el mercado laboral. ¿Reduce la inmigración los salarios? La teoría económica sugiere que si la oferta de trabajo aumenta debido a la afluencia de inmigrantes, el «precio» del trabajo —el salario— debería disminuir. Sin embargo, manteniendo todo lo demás constante, los inmigrantes son atraídos por las ciudades con elevada demanda laboral, por lo que el estimador MCO del efecto sobre los salarios de la inmigración estará sesgado. Un experimento aleatorizado controlado ideal para estimar el efecto sobre los salarios de la inmigración asignaría aleatoriamente un número diferente de inmigrantes (diferentes «tratamientos») a los diferentes mercados de trabajo (los «sujetos») y mediría el efecto sobre los salarios (el «resultado» o la «respuesta»). Un experimento de ese tipo, no obstante, se enfrenta a graves problemas prácticos, financieros y éticos.

De este modo, el economista del trabajo David Card (1990), utilizó un cuasi experimento en el que un gran número de inmigrantes cubanos ingresaron en el mercado laboral de Miami, Florida, en el «éxodo del Mariel», sucedido como resultado de un levantamiento temporal de las restricciones sobre la emigración desde Cuba en 1980. La mitad de los inmigrantes se establecieron en Miami, en parte porque había una gran comunidad cubana preexistente. Card estimó el efecto causal sobre los salarios de un aumento de la inmigración mediante la comparación de la variación en los salarios de los trabajadores poco cualificados en Miami con la variación en los salarios de los trabajadores similares en otras ciudades comparables de los EE.UU. durante el mismo periodo. Llegó a la conclusión de que esta afluencia de inmigrantes tuvo un efecto insignificante en los salarios de los trabajadores menos cualificados.

Ejemplo #2: los efectos del servicio militar sobre los ingresos salariales de los civiles. ¿La prestación del servicio militar mejora las perspectivas en el mercado laboral? Las fuerzas armadas proporcionan formación que los futuros empleadores podrían considerar atractiva. Sin embargo, una regresión MCO de los ingresos salariales de los civiles sobre la realización previa del servicio militar podría dar lugar a un estimador sesgado del efecto sobre los ingresos de los civiles de haber realizado el servicio militar debido a que la realización del servicio militar está determinada, al menos en parte, por las elecciones y las características individuales. Por ejemplo, los militares solo aceptan a los solicitantes que cumplan unos requisitos físicos mínimos, y la ausencia de éxito en el mercado laboral del sector privado podría hacer que una persona presentara más probabilidades de inscribirse en el ejército.

Para evitar este sesgo de selección, Joshua Angrist (1990) utilizó un diseño cuasi experimental en el que analizaba los historiales en el mercado laboral de las personas que sirvieron en el ejército de EE.UU. durante la guerra de Vietnam. A lo largo de este periodo, el hecho de que un joven fuera reclutado por el ejército estaba determinado en parte por un sistema de lotería nacional basado en las fechas de nacimiento: los hombres a los que aleatoriamente se les asignaron números bajos en la lotería eran seleccionables para ser reclutados mientras que aquellos con números elevados no lo eran. En realidad el acceso al ejército se regía por normas complicadas, tales como un examen físico y algunas excepciones, y algunos jóvenes se presentaban voluntarios al servicio, por lo que la prestación del servicio militar dependía solo en parte de si un hombre era seleccionable para el reclutamiento. Por tanto, el hecho de ser seleccionable para el reclutamiento sirve como variable instrumental que determina en parte la prestación del servicio militar, pero que se asigna aleatoriamente. En este caso, había una verdadera asignación aleatoria de ser seleccionado para el reclutamiento mediante la lotería, pero debido a que esta aleatorización no se realizaba como parte de un experimento para evaluar el efecto del servicio militar, es un cuasi experimento. Angrist concluyó que el efecto a largo plazo del servicio militar era el de reducir los salarios de los veteranos blancos, pero no el del resto.

Ejemplo #3: el efecto del cateterismo cardiaco. La Sección 12.5 describía el estudio de McClellan, McNeil, y Newhouse (1994) en el que utilizaba la distancia desde la vivienda de un paciente con un ataque al corazón a un hospital que disponía de la técnica de cateterismo cardíaco, comparada con la distancia a un hospital que carecía de servicios de hemodinámica, como una variable instrumental para el tratamiento real mediante cateterismo cardíaco. Este estudio es un cuasi experimento con una variable que determina en parte el tratamiento. El tratamiento en sí, el cateterismo cardiaco, está determinado por las

características personales de los pacientes y por la decisión del paciente y del médico; sin embargo, se ve influido asimismo por el hecho de si un hospital cercano es capaz de realizar este procedimiento. Si la ubicación del paciente es «como si fuera» asignada aleatoriamente y no tiene efecto directo sobre los resultados en salud, que no sea a través de su efecto sobre la probabilidad de un cateterismo, entonces la distancia relativa a un hospital con técnica de cateterismo es una variable instrumental válida.

Otros ejemplos. La estrategia de investigación de los cuasi experimentos ha sido aplicada asimismo en otras áreas. Garvey y Hanka (1999) utilizaron la variación en las leyes estatales de EE.UU. para examinar el efecto sobre la estructura financiera de las empresas (por ejemplo, la utilización de la deuda por las empresas) de las leyes anti-OPAs. Meyer, Viscusi y Durbin (1995) utilizaron grandes cambios discretos en la generosidad de las prestaciones del seguro de desempleo en Kentucky y Michigan, que afectaban de manera diferenciada a los trabajadores con altos pero no con bajos ingresos salariales, para estimar el efecto sobre el tiempo que transcurría sin trabajar de una variación en las prestaciones por desempleo. Las recopilaciones sobre esta cuestión de Meyer (1995), Rosenzweig y Wolpin (2000), y Angrist y Krueger (2001) proporcionan otros ejemplos de cuasi experimentos en las áreas de economía y política social.

El estimador de diferencias en diferencias

Si el tratamiento en un cuasi experimento es «como si» fuera asignado al azar, condicionado a algunas variables observadas W , entonces el efecto del tratamiento se puede estimar utilizando la regresión de las diferencias (13.2). No obstante, debido a que el investigador no tiene control sobre la aleatoriedad, puede continuar existiendo algunas diferencias entre los grupos de tratamiento y de control, incluso después de tener en cuenta W . Una forma de ajustar por estas diferencias que permanecen entre los dos grupos consiste no en comparar los resultados Y , sino la *variación* en los resultados pre y post-tratamiento, y de este modo ajustando por las diferencias en los valores pre-tratamiento de Y entre los dos grupos. Debido a que este estimador es la diferencia entre grupos en la variación, o la diferencia en el tiempo, este estimador se denomina estimador de diferencias en diferencias. Por ejemplo, en el estudio de Card (1990) del efecto de la inmigración sobre los salarios de los trabajadores con baja cualificación, se utilizaba un estimador de diferencias en diferencias para comparar la *variación* en los salarios en Miami con la *variación* en los salarios en otras ciudades de EE.UU. Otro ejemplo del uso del estimador de diferencias en diferencias es en el recuadro «¿Cuál es el efecto sobre el empleo del salario mínimo?».

El estimador de diferencias en diferencias. Sea $\bar{Y}^{\text{tratamiento}, \text{antes}}$ la media muestral de Y para aquellos en el grupo de tratamiento antes del experimento, y sea $\bar{Y}^{\text{tratamiento}, \text{después}}$ la media muestral para el grupo de tratamiento después del experimento. Sean $\bar{Y}^{\text{control}, \text{antes}}$ e $\bar{Y}^{\text{control}, \text{después}}$ las correspondientes medias muestrales para el grupo de control. La variación promedio en Y durante el curso del experimento para aquellos en el grupo de tratamiento es $\bar{Y}^{\text{tratamiento}, \text{antes}} - \bar{Y}^{\text{tratamiento}, \text{después}}$, y la variación promedio en Y durante el periodo del experimento para los del grupo de control es $\bar{Y}^{\text{control}, \text{antes}} - \bar{Y}^{\text{control}, \text{después}}$. El **estimador de diferencias en diferencias** es la variación promedio en Y para aquellos en el grupo de tratamiento, menos el cambio promedio en Y para aquellos en el grupo control:

$$\hat{\beta}_1^{\text{difsendifs}} = (\bar{Y}^{\text{tratamiento}, \text{después}} - \bar{Y}^{\text{tratamiento}, \text{antes}}) - (\bar{Y}^{\text{control}, \text{después}} - \bar{Y}^{\text{control}, \text{antes}}) \\ = \Delta\bar{Y}^{\text{tratamiento}} - \Delta\bar{Y}^{\text{control}}, \quad (13.4)$$

donde $\Delta\bar{Y}^{\text{tratamiento}}$ es la variación promedio en Y en el grupo de tratamiento y $\Delta\bar{Y}^{\text{control}}$ es la variación promedio en Y en el grupo control. Si el tratamiento es asignado aleatoriamente, entonces $\hat{\beta}_1^{\text{difsendifs}}$ es un estimador insesgado y consistente del efecto causal.

El estimador de diferencias en diferencias puede escribirse en notación de regresión. Sea ΔY_i el valor postexperimental de Y para el individuo i -ésimo, menos el valor preexperimental. El estimador de diferencias de las diferencias es el estimador MCO de la regresión,

$$\Delta Y_i = \beta_0 + \beta_1 X_i + u_i. \quad (13.5)$$

¿Cuál es el efecto sobre el empleo del salario mínimo?

¿Cuánto reduce la demanda de trabajadores poco cualificados un aumento en el salario mínimo? La teoría económica dice que la demanda cae cuando el precio aumenta, pero determinar cuánto lo hace con precisión se trata de una cuestión empírica. Debido a que los precios y las cantidades se determinan por la oferta y la demanda, el estimador MCO de una regresión del empleo sobre los salarios presenta sesgo de causalidad simultánea (Concepto clave 9.6). Hipotéticamente, un experimento aleatorizado controlado podría asignar aleatoriamente diferentes salarios mínimos a los diferentes empleadores y luego comparar los cambios en el empleo (los resultados) en los grupos de tratamiento y de control, pero ¿cómo podría llevarse a cabo en la práctica este experimento hipotético?

Los economistas laborales David Card y Alan Krueger (1994) decidieron llevar a cabo un experimento así, pero dejando que «la naturaleza» —o, más precisamente, la geografía— realizase la asignación aleatoria por ellos. En 1992, el salario mínimo en Nueva Jersey aumentó de 4,25 \$ a 5,05 \$ por hora, pero el salario mínimo en la vecina Pensylvania se mantuvo constante. En este experimento, el «tratamiento» del aumento del salario mínimo —estar domiciliado en Nueva Jersey en lugar de en Pensylvania— se interpreta «como si» se asignara aleatoriamente en el sentido de que el hecho de

estar sujeto al aumento de los salarios se supone que no está correlacionado con los otros factores determinantes de las variaciones en el empleo durante este periodo. Cardy y Krueger recogieron datos sobre el empleo en los restaurantes de comida rápida antes y después del aumento salarial en los dos estados. Cuando calcularon el estimador de diferencias en diferencias, se encontraron con un resultado sorprendente: no había evidencia de que el empleo cayera en los restaurantes de comida rápida en Nueva Jersey, en relación con los de Pensylvania. De hecho, algunas de sus estimaciones en realidad sugieren que el empleo *aumentó* en los restaurantes de Nueva Jersey después de que su salario mínimo aumentara, en relación a Pensylvania!

Estos resultados entran en conflicto con la teoría microeconómica básica y han sido bastante controvertidos. Análisis posteriores, utilizando una fuente diferente de datos de empleo, sugieren que podría haber habido una pequeña caída del empleo en Nueva Jersey después del aumento de los salarios, pero aun así la curva de demanda de trabajo estimada es muy inelástica (Neumark y Wascher, 2000). Aunque la elasticidad de los salarios exacta en este cuasi experimento es un tema de debate, el efecto sobre el empleo de un alza en el salario mínimo parece ser menor que lo que muchos economistas habían pensado previamente.

El estimador de las diferencias en diferencias se ilustra en la Figura 13.1. En esta figura, la media muestral de Y para el grupo de tratamiento es de 40 antes del experimento, mientras que la media muestral de Y antes del tratamiento para el grupo de control es de 20. En el transcurso del experimento, la media muestral de Y aumenta en el grupo de control a 30, mientras que aumenta hasta 80 para el grupo de tratamiento. Por tanto, la diferencia media de las medias muestrales después del tratamiento es $80 - 30 = 50$. Sin embargo, parte de esta diferencia se debe a que los grupos de tratamiento y de control no tienen las mismas medias antes del tratamiento: el grupo de tratamiento comenzó por delante del grupo de control. El estimador de las diferencias en diferencias mide las ganancias del grupo de tratamiento, en comparación con el grupo de control, que en este ejemplo es $(80 - 40) - (30 - 20)$. Al estar centrado en la variación de Y en el transcurso del experimento, el estimador de las diferencias en diferencias elimina la influencia de los valores iniciales de Y , que varían entre los grupos de tratamiento y de control.

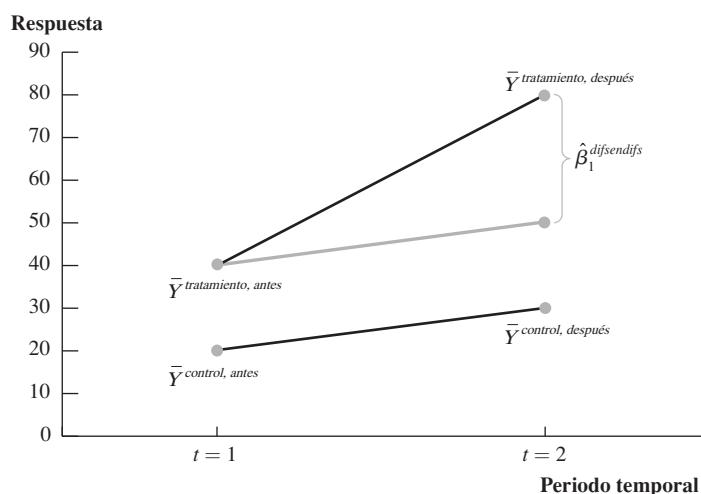
El estimador de diferencias en diferencias con regresores adicionales. El estimador de las diferencias en diferencias se puede extender para incluir regresores adicionales W_{1i}, \dots, W_{ri} , que midan las características individuales antes del experimento. Estos regresores adicionales se pueden incorporar mediante el modelo de regresión múltiple

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \dots + \beta_{1+r} W_{ri} + u_i, \quad i = 1, \dots, n. \quad (13.6)$$

El estimador MCO de β_1 en la Ecuación (13.6) es el **estimador de diferencias en diferencias con regresores adicionales**. Si X_i es «como si fuera» asignado aleatoriamente, condicionado a W_{1i}, \dots, W_{ri} , entonces u_i satisface la independencia en media condicional y el estimador MCO de $\hat{\beta}_1$ en la Ecuación (13.6) es insesgado.

FIGURA 13.1 El estimador de diferencias en diferencias

La diferencia tras el tratamiento entre los grupos de tratamiento y de control es $80 - 30 = 50$, pero esto exagera el efecto del tratamiento, debido a que antes del tratamiento \bar{Y} era mayor para el grupo de tratamiento que para el grupo de control en $40 - 20 = 20$. El estimador de las diferencias en diferencias es la diferencia entre las brechas final e inicial, por lo que $\hat{\beta}_1^{\text{difsendifs}} = (80 - 30) - (40 - 20) = 50 - 20 = 30$. De manera equivalente, el estimador de las diferencias en diferencias es la variación promedio para el grupo de tratamiento menos la variación promedio para el grupo de control, es decir, $\hat{\beta}_1^{\text{difsendifs}} = \Delta\bar{Y}^{\text{tratamiento}} - \Delta\bar{Y}^{\text{control}} = (80 - 40) - (30 - 20) = 30$.



El estimador de las diferencias en diferencias aquí descrito considera dos períodos de tiempo, antes y después del experimento. En algunos casos existen datos de panel con varios períodos de tiempo. El estimador de las diferencias en diferencias se puede extender a varios períodos de tiempo mediante los métodos de regresión de datos de panel del Capítulo 10.

Diferencias en diferencias con datos de sección cruzada repetida. Un conjunto de **datos de sección cruzada repetida** es una colección de conjuntos de datos de sección cruzada, donde cada conjunto de datos de sección cruzada corresponde a un período de tiempo diferente. Por ejemplo, el conjunto de datos podría contener observaciones sobre 400 personas en el año 2004 y sobre 500 personas diferentes en el año 2005, para un total de 900 personas diferentes. Un ejemplo de datos de sección cruzada repetida son los datos de encuestas políticas, en las que las preferencias políticas se miden por una serie de encuestas a potenciales votantes seleccionados al azar, y en los que las encuestas se realizan en fechas diferentes y cada estudio tiene diferentes encuestados.

La premisa de la utilización de datos de sección cruzada repetida es que si los individuos (o más en general, las entidades individuales) se extraen aleatoriamente a partir de la misma población, entonces los individuos de la sección cruzada más temprana se pueden utilizar como sustitutos de los individuos de los grupos de tratamiento y de control en la sección cruzada posterior.

Cuando existen dos períodos de tiempo, el modelo de regresión para datos de sección cruzada repetida es

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 G_i + \beta_3 D_t + \beta_4 W_{1it} + \dots + \beta_{3+r} W_{rit} + u_{it}, \quad (13.7)$$

donde X_{it} es el tratamiento real de la i -ésima persona (entidad individual) en la sección cruzada del período t ($t = 1, 2$), G_i es una variable binaria que indica si el individuo está en el grupo de tratamiento (o en el grupo de tratamiento sustituto, si la observación está en el período pre-tratamiento), y D_t es el indicador binario que es igual a 0 en el primer período y es igual a 1 en el segundo período. El individuo i -ésimo recibe tratamiento si él o ella está en el grupo de tratamiento en el segundo período, por lo que en la Ecuación (13.7), $X_{it} = G_i \times D_t$, es decir, X_{it} es la interacción entre G_i y D_t .

Si el cuasi experimento hace que parezca X_{it} «como si fuera» recibido aleatoriamente, condicionado a las W , entonces el efecto causal puede ser estimado mediante el estimador MCO en la Ecuación (13.7). Si existen más de dos períodos temporales, entonces la Ecuación (13.7) se modifica para que contenga $T - 1$ variables binarias que indiquen los períodos de tiempo diferentes (véase la Sección 10.4).

Estimadores de variables instrumentales

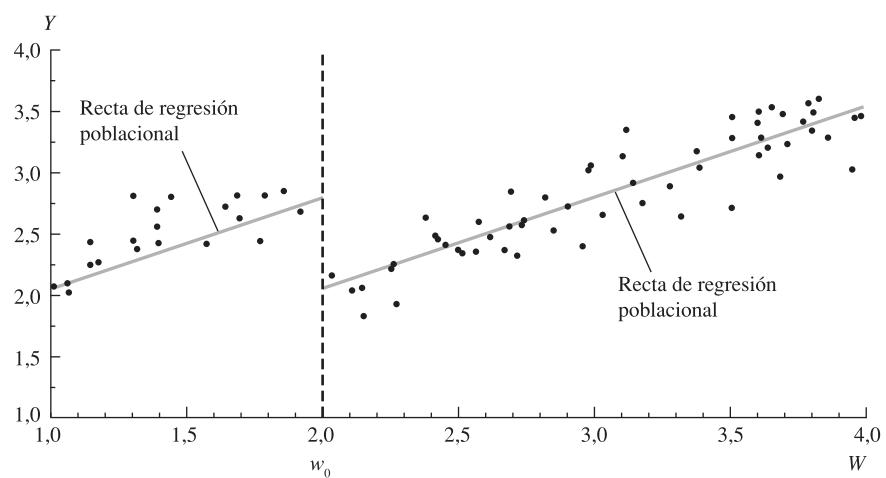
Si el cuasi experimento proporciona una variable Z_i que influye en la recepción del tratamiento, si los datos están disponibles tanto sobre Z_i como sobre el tratamiento realmente recibido (X_i), y si Z_i es como si estuviera asignado aleatoriamente (tal vez tras tener en cuenta algunas variables adicionales W_j), entonces Z_i es un instrumento válido para X_i y los coeficientes de la Ecuación (13.2) se pueden estimar mediante mínimos cuadrados en dos etapas. Cualquiera de las variables de control que aparecen en (13.2) aparecen asimismo como variables de control en la primera etapa del estimador de mínimos cuadrados en dos etapas de β_1 .

Estimadores de la regresión con discontinuidad

Una situación que surge en un cuasi experimento es cuando la recepción del tratamiento depende en su totalidad o en parte del hecho de que una variable observable W cruce un valor umbral. Por ejemplo, supongamos que los estudiantes están obligados a asistir a unos cursos de verano si su calificación media a final de curso (GPA⁵) cae por debajo de un umbral⁶. Por tanto, una manera de estimar el efecto de los cursos de verano obligatorios consiste en comparar los resultados de los estudiantes cuya GPA estaba justo por debajo del umbral (y que por tanto fueron obligados a asistir) con los resultados de los estudiantes cuyo GPA estaba justo por encima del umbral (por lo que evitaron los cursos de verano). El resultado Y podría ser la GPA del año siguiente o los ingresos salariales futuros. En tanto en cuanto no existe nada especial en el valor de umbral que no sea su utilización para la obligatoriedad de los cursos de verano, resulta razonable atribuir cualquier tipo de salto en los resultados a ese umbral para los cursos de verano. La Figura 13.2 muestra un diagrama de dispersión hipotético de un conjunto de datos en los cuales el tratamiento (los cursos de verano, X) se exigen en el caso de que GPA (W) sea menor que el valor umbral ($w_0 = 2,0$). El diagrama de dispersión muestra la GPA del año siguiente (Y) para una muestra hipotética de estudiantes como una función de la GPA de este año, junto con la función de regresión poblacional. Si la única función del umbral w_0 es la obligatoriedad de los cursos de verano, entonces el salto en la GPA del año siguiente en w_0 es una estimación del efecto de los cursos de verano en la GPA del año siguiente.

FIGURA 13.2 Un diagrama de dispersión hipotético del diseño de la discontinuidad de la regresión

Supongamos que el tratamiento binario X es obligatorio si W es menor que el valor umbral $w_0 = 2$. Siempre y cuando el único papel del umbral, w_0 sea prescribir el tratamiento, el efecto del tratamiento estará dado por la magnitud del salto o discontinuidad, en la función de regresión en $W = 2$.



⁵ N. Del T.: Siglas originales en inglés de Grade Point Average.

⁶ Este ejemplo es una versión simplificada del estudio de la discontinuidad en la regresión de los efectos de los cursos de verano para los estudiantes de primaria y secundaria de Jordan Matsudaira (2008), en el que la asistencia a los cursos de verano estaba basada en parte en los exámenes finales del curso.

Debido al salto o discontinuidad en el tratamiento en el umbral, los estudios que aprovechan una discontinuidad en la probabilidad de recibir el tratamiento en un valor umbral se denominan diseños para la **discontinuidad de la regresión**. Existen dos tipos de diseños para la discontinuidad de regresión: brusco y difuso.

Diseño brusco para la discontinuidad de la regresión. En un diseño brusco para la discontinuidad de la regresión, la recepción de tratamiento está totalmente determinada por el hecho de W exceda el umbral: todos los estudiantes con $W < w_0$ asisten a los cursos de verano, y ningún estudiante con $W \geq w_0$ asiste; es decir, $X_i = 1$ si $W < w_0$ y $X_i = 0$ si $W \geq w_0$. En este caso, el salto en Y en el umbral es igual al efecto promedio del tratamiento para la subpoblación con $W = w_0$, lo que podría ser una aproximación útil al efecto promedio del tratamiento en la población más amplia de interés. Si la función de regresión es lineal en W , con excepción de la discontinuidad inducida por el tratamiento, el efecto del tratamiento puede ser estimado por β_1 en la regresión:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i. \quad (13.8)$$

Si la función de regresión es no lineal, entonces se puede utilizar una función no lineal adecuada para W (Sección 8.2).

Diseño difuso para la discontinuidad de la regresión. En un diseño difuso para la discontinuidad de la regresión, el hecho de cruzar el umbral influye en la recepción del tratamiento, pero no es el único determinante. Por ejemplo, supongamos que algunos estudiantes cuya GPA cae por debajo del umbral están exentos de los cursos de verano, mientras que otros cuya GPA excede el umbral, no obstante asisten. Esta situación podría darse si la norma del umbral forma parte de un proceso más complicado para prescribir el tratamiento. En un diseño difuso, X_i , en general estará correlacionada con la u_i de la Ecuación (13.8). Sin embargo, si los efectos particulares del hecho de cruzar el umbral influyen únicamente aumentando la probabilidad del tratamiento, es decir, el efecto directo de cruzar el umbral está captado por el término lineal de W , entonces es posible el método de variables instrumentales. En concreto, la variable binaria Z_i que indica cruzar el umbral ($Z_i = 1$ si $W_i < w_0$ y $Z_i = 0$ si $W_i \geq w_0$) influye en el hecho de recibir el tratamiento, pero no está correlacionada con u_i , por lo que es un instrumento válido para X_i . Por tanto, en un diseño difuso para la discontinuidad de la regresión, se puede estimar β , mediante la estimación de variables instrumentales de la Ecuación (13.8), utilizando como instrumento la variable binaria que indica que $W_i < w_0$.

13.5 Problemas potenciales en cuasi experimentos

Al igual que todos los estudios empíricos, los cuasi experimentos se enfrentan a amenazas a su validez interna y externa. Una amenaza potencial muy importante a la validez interna es la cuestión de si la asignación «como si fuera» aleatoriedad, en realidad, puede ser tratada de forma fiable como verdadera aleatoriedad.

Amenazas a la validez interna

Las amenazas a la validez interna de los verdaderos experimentos aleatorizados controlados enumeradas en la Sección 13.2 son aplicables asimismo a los cuasi experimentos, pero con algunas modificaciones.

Ausencia de aleatoriedad. Los cuasi experimentos se basan en las diferencias en las circunstancias individuales, cambios legales, sucesos repentinos incorrelacionados, etc., para proporcionar la asignación «como si fuera» aleatoriedad del nivel de tratamiento. Si esta asignación «como si fuera» aleatoriedad no diera lugar a un nivel de tratamiento X (o a una variable instrumental Z) que sea aleatoria, entonces, en general, el estimador MCO es sesgado (o el estimador de variables instrumentales no es consistente).

Al igual que en un experimento real, una forma de contrastar la ausencia de aleatoriedad consiste en comprobar las diferencias sistemáticas entre los grupos de tratamiento y de control, por ejemplo mediante una regresión de X (o Z) sobre las características individuales (las W) y contrastar la hipótesis de que los coeficientes de las W son iguales a cero. Si existen diferencias que no se expliquen fácilmente por la natura-

leza del cuasi experimento, entonces existe evidencia de que el cuasi experimento no origina una verdadera aleatorización. Incluso si no existe relación entre X (o Z) y las W , seguirá existiendo la posibilidad de que X (o Z) pudiera estar relacionada con algunos de los factores no observables del término de error u . Debido a que estos factores no son observables, esta posibilidad no se puede contrastar, y la validez de la hipótesis de la asignación «como si fuera aleatoria» debe ser evaluada mediante la aplicación de un criterio experto y del conocimiento de cada caso concreto.

Incumplimiento del protocolo de tratamiento. En un experimento verdadero, el incumplimiento del protocolo de tratamiento se produce cuando los miembros del grupo de tratamiento no reciben el tratamiento, los miembros del grupo de control en realidad reciben el tratamiento, o las dos cosas a la vez; en consecuencia, el estimador MCO del efecto causal presenta sesgo de selección. El homólogo del incumplimiento del protocolo de tratamiento en un cuasi experimento es cuando la asignación «como si fuera» aleatoria influye, pero no determina, el nivel del tratamiento. En este caso, el estimador de variables instrumentales basado en la influencia cuasi experimental Z puede ser consistente a pesar de que el estimador MCO no lo sea.

Deserción o abandono. La deserción o abandono en un cuasi experimento es similar a la deserción en un verdadero experimento en el sentido de que surge debido a las decisiones o características personales, por tanto la deserción puede inducir correlación entre el nivel de tratamiento y el término de error. El resultado es el sesgo de selección muestral, por lo que el estimador MCO del efecto causal es sesgado e inconsistente.

Efectos experimentales. Una de las ventajas de los cuasi experimentos es que, debido a que no son verdaderos experimentos, habitualmente no hay ninguna razón para que las personas piensen que son sujetos experimentales. Por tanto los efectos experimentales como el efecto Hawthorne por lo general no son pertinentes en los cuasi experimentos.

Validez de los instrumentos en los cuasi experimentos. Un paso importante en la evaluación de un estudio que utiliza la regresión de variables instrumentales consiste en la reflexión cuidadosa acerca de si el instrumento es en efecto válido. Esta afirmación general, sigue siendo cierta en los estudios cuasi experimentales en los que el instrumento es «como si fuera» determinado al azar. Como se analizó en el Capítulo 12, la validez del instrumento requiere tanto la relevancia del instrumento como su exogeneidad. Debido a que la relevancia del instrumento puede ser contrastada mediante los métodos estadísticos que se resumen en el Concepto Clave 12.5, nos centraremos aquí en el segundo requisito, más sujeto a valoración de la exogeneidad del instrumento.

Aunque pudiera parecer que una variable instrumental aleatoriamente asignada es necesariamente exógena, esto no es así. Consideremos los ejemplos de la Sección 13.4. En la utilización por parte de Angrist (1990) de los números de lotería para el reclutamiento como variable instrumental en el estudio del efecto sobre los ingresos salariales de los civiles de la prestación del servicio militar, el número de la lotería era en realidad asignado aleatoriamente. Pero, como Angrist (1990) señala y analiza, si un número bajo en la lotería de reclutamiento da lugar a un comportamiento destinado a evitar el reclutamiento y esta conducta para evitarlo afecta posteriormente a los ingresos salariales civiles, entonces un número bajo en la lotería (Z_i) podría estar relacionado con factores no observables que determinen los ingresos salariales de los civiles (u_i); es decir, Z_i y u_i están correlacionados a pesar de que Z_i esté asignado aleatoriamente. Como segundo ejemplo, el estudio de McClellan, McNeil, y Newhouse (1994) acerca del efecto sobre los pacientes con un ataque al corazón de un cateterismo cardiaco consideraba la distancia relativa a un hospital con técnica de cateterismo como si estuviera asignada aleatoriamente. Pero, como los autores destacan y examinan, si los pacientes que viven cerca de un hospital con técnica de cateterismo están más sanos que los que viven lejos (tal vez debido a un mejor acceso a la atención médica en general), entonces la distancia relativa con respecto a un hospital con técnica de cateterismo estará correlacionada con las variables omitidas que están en el término de error de la ecuación de los resultados de salud. En resumen, el simple hecho de que un instrumento esté determinado aleatoriamente o sea «como si estuviera» determinado aleatoriamente, no significa necesariamente que sea exógeno en el sentido de que $\text{corr}(Z_i, u_i) = 0$. Por tanto, la exogeneidad debe ser analizada cuidadosamente, incluso si el instrumento surge de un cuasi experimento.

Amenazas a la validez externa

Los estudios cuasi experimentales se basan en datos de observaciones, y las amenazas a la validez externa de un estudio basado en un cuasi experimento son generalmente similares a las amenazas analizadas en la Sección 9.1 para los estudios de regresión convencionales basados en datos de observaciones.

Una consideración importante es que los sucesos particulares que crean la asignación «como si fuera» aleatoria en el núcleo de un estudio cuasi experimental pueden dar lugar a otras características particulares que amenacen la validez externa. Por ejemplo, el estudio de Card (1990) de los efectos en el mercado de trabajo de la inmigración discutido en la Sección 13.4 utilizaba una «como si fuera» aleatoriedad inducida por la afluencia de inmigrantes cubanos en el éxodo del Mariel. Sin embargo, existían características particulares en los inmigrantes cubanos, en Miami, y en la comunidad cubana, que podrían dificultar la generalización de estos resultados a los inmigrantes procedentes de otros países o afluentes hacia otros destinos. Del mismo modo, el estudio de Angrist (1990) sobre los efectos en el mercado de trabajo de servir en el ejército de EE.UU. durante la guerra de Vietnam, presumiblemente podrían no ser generalizables al servicio militar en tiempos de paz. Como de costumbre, el hecho de que un estudio sea generalizable a una población específica y a un escenario de interés depende de los detalles del estudio y debe evaluarse caso por caso.

13.6 Estimaciones experimentales y cuasi experimentales en poblaciones heterogéneas

Tal y como se analizó en la Sección 13.1, el efecto causal puede variar de un miembro de la población a otro. En la Sección 13.1 se analizaba la estimación de los efectos causales que variaban en función de variables observables, tales como el género. En esta sección, se consideran las consecuencias de la variación *no observable* en el efecto causal. Nos referimos a la variación no observable en el efecto causal que tiene una población heterogénea. Por motivos de simplicidad nos centramos en el papel de la heterogeneidad no observable, en esta sección se omiten las variables de control W ; las conclusiones de esta sección son trasladables a las regresiones que incluyen variables de control.

Si la población es heterogénea, entonces el individuo i -ésimo tiene ahora su propio efecto causal, β_{1i} , que (en la terminología de la Sección 13.1) es la diferencia entre los resultados potenciales para el individuo i -ésimo de si el tratamiento es recibido o no. Por ejemplo, β_{1i} podría ser cero para un programa de formación dedicado a la elaboración del currículum vítae, si el individuo i -ésimo ya sabe cómo redactar el currículum vítae. Con esta notación, la ecuación de regresión poblacional se puede escribir de la forma

$$Y_i = \beta_{0i} + \beta_{1i}X_i + u_i. \quad (13.9)$$

Debido a que β_{1i} varía en la población de un individuo a otro y los individuos se seleccionan aleatoriamente a partir de la población, β_{1i} es una variable aleatoria que, al igual que u_i , refleja la variación no observable entre los individuos (por ejemplo, la variación en los conocimientos previos de cómo redactar un currículum vítae). El efecto causal medio es el valor medio poblacional del efecto causal, $E(\beta_{1i})$; es decir, es la esperanza del efecto causal para un miembro de la población estudiada que ha sido seleccionado aleatoriamente.

¿Qué estiman los estimadores de las Secciones 13.1, 13.2, y 13.4 si existe heterogeneidad en la población del tipo de la Ecuación (13.9)? En primer lugar, se considera el estimador MCO cuando X_i es «como si fuera» determinado al azar; en este caso, el estimador MCO es un estimador consistente del efecto causal promedio. No obstante, esto no es cierto en general para el estimador VI. En cambio, si X_i está parcialmente influenciado por Z_i , entonces el estimador VI que utiliza Z como instrumento estima una media ponderada de los efectos causales, en la que reciben mayor ponderación aquellos para los cuales el instrumento es más influyente.

MCO con efectos causales heterogéneos

Si existe heterogeneidad en el efecto causal y si X_i está asignado aleatoriamente, entonces el estimador de las diferencias es un estimador consistente del efecto causal promedio. Este resultado se deduce del análisis de la Sección 13.1 y del Apéndice 13.3, que utiliza el contexto de las variables respuesta; aquí se muestra sin referencia a las variables respuesta mediante la aplicación de los conceptos de los Capítulos 3 y 4 directamente a los coeficientes aleatorios del modelo de regresión de la Ecuación (13.9).

El estimador MCO de β_1 en la Ecuación (13.1) es $\hat{\beta}_1 = s_{XY}/s_X^2$ [Ecuación (4.7)]. Si las observaciones son i.i.d., entonces la covarianza y la varianza muestrales son estimadores consistentes de la covarianza y la varianza poblacionales, por lo que $\hat{\beta}_1 \xrightarrow{P} \sigma_{XY}/\sigma_X^2$. Si X_i se asigna aleatoriamente, entonces X_i se distribuye de forma independiente de las otras características individuales, tanto observables como no observables, y en particular, se distribuye de forma independiente de β_{0i} y β_{1i} . Por lo tanto, el estimador MCO $\hat{\beta}_1$ tiene el límite

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \xrightarrow{P} \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\text{cov}(\beta_{0i} + \beta_{1i}X_i + u_i, X_i)}{\sigma_X^2} = \frac{\text{cov}(\beta_{0i} + \beta_{1i}X_i, X_i)}{\sigma_X^2} = E(\beta_{1i}), \quad (13.10)$$

donde la tercera igualdad utiliza las propiedades de las covarianzas del Concepto clave 2.3 y $\text{cov}(u_i, X_i) = 0$, lo que se deduce de que $E(u_i|X_i) = 0$ [Ecuación (2.27)], y donde la última igualdad se deriva de que β_{0i} y β_{1i} se distribuyen independientemente respecto de X_i , lo que ocurre si X_i se asigna aleatoriamente (Ejercicio 13.9). Por tanto, si X_i se asigna aleatoriamente, $\hat{\beta}_1$ es un estimador consistente del efecto causal medio $E(\beta_{1i})$.

Regresión VI con efectos causales heterogéneos

Supongamos que los efectos causales se estiman mediante una regresión de variables instrumentales de Y_i sobre X_i (tratamiento realmente recibido) con Z_i (aleatorización inicial o tratamiento «como si fuera» asignado aleatoriamente) como instrumento. Supongamos que Z_i es un instrumento válido (relevante y exógeno) y que existe heterogeneidad en el efecto sobre X_i de Z_i . En concreto, supongamos que X_i está correlacionada con Z_i mediante el modelo lineal

$$X_i = \pi_{0i} + \pi_{1i}Z_i + v_i, \quad (13.11)$$

donde los coeficientes π_{0i} y π_{1i} varían de un individuo a otro. La Ecuación (13.11) es la ecuación de la primera etapa de MC2E [Ecuación (12.2)], con la variante de que se permite que el efecto sobre X_i de un cambio en Z_i varíe de un individuo a otro.

El estimador MC2E es $\hat{\beta}_1^{MC2E} = s_{ZY}/s_{ZX}$ [Ecuación (12.4)], el cociente entre la covarianza muestral entre Z e Y y la covarianza muestral entre Z y X . Si las observaciones son i.i.d., entonces estas covarianzas muestrales son estimadores consistentes de las covarianzas poblacionales por lo que $\hat{\beta}_1^{MC2E} \xrightarrow{P} \sigma_{ZY}/\sigma_{ZX}$. Supongamos que π_{0i} , π_{1i} , β_{0i} , y β_{1i} se distribuyen de forma independiente de u_i , v_i , y Z_i ; que $E(u_i|Z_i) = E(v_i|Z_i) = 0$; y que $E(\pi_{1i}) \neq 0$ (relevancia del instrumento). Se demuestra en el Apéndice 13.2 que, bajo estos supuestos,

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{P} \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})}. \quad (13.12)$$

Es decir, el estimador MC2E converge en probabilidad al cociente entre el valor esperado del producto de β_{1i} y π_{1i} y el valor esperado de π_{1i} .

El último cociente de la Ecuación (13.12) es una media ponderada de los efectos causales individuales β_{1i} . Las ponderaciones son $\pi_{1i}/E(\pi_{1i})$ que miden la intensidad relativa con la que el instrumento influye en si el individuo i -ésimo recibe tratamiento. Por tanto el estimador MC2E es un estimador consistente de una media ponderada de los efectos causales individuales, en la que los individuos que reciben la mayor ponde-

ración son aquellos para los cuales el instrumento es más influyente. El efecto causal medio ponderado que se estima por MC2E se denomina **promedio local del efecto del tratamiento** (LATE)⁷. El término «local» hace hincapié en que es el promedio ponderado que pondera más a aquellos individuos (en general, entidades individuales), cuya probabilidad de tratamiento está más influenciada por la variable instrumental.

Existen tres casos particulares en los que el promedio local del efecto del tratamiento es igual al efecto promedio del tratamiento:

1. El efecto del tratamiento es el mismo para todos los individuos. Este caso se corresponde con $\beta_{1i} = \beta_1$ para todo i . Por tanto la última expresión de la Ecuación (13.12) se reduce a $E(\beta_{1i}\pi_{1i})/E(\pi_{1i}) = \beta_1E(\pi_{1i})/E(\pi_{1i}) = \beta_1$.
2. El instrumento afecta a cada individuo por igual. Este caso se corresponde con $\pi_{1i} = \pi_1$ para todo i . En este caso, la última expresión de la Ecuación (13.12) se reduce a $E(\beta_{1i}\pi_{1i})/E(\pi_{1i}) = E(\beta_{1i})\pi_1/\pi_1 = E(\beta_{1i})$.
3. La heterogeneidad en el efecto del tratamiento y la heterogeneidad en el efecto del instrumento están incorrelacionadas. Este caso corresponde a β_{1i} y π_{1i} aleatorios pero $\text{cov}(\beta_{1i}, \pi_{1i}) = 0$. Debido a que $E(\beta_{1i}\pi_{1i}) = \text{cov}(\beta_1, \pi_{1i}) + E(\beta_{1i})E(\pi_{1i})$ [Ecuación (2.34)], si $\text{cov}(\beta_{1i}, \pi_{1i}) = 0$ entonces $E(\beta_{1i}\pi_{1i}) = E(\beta_{1i})E(\pi_{1i})$ y la última expresión de la Ecuación (13.12) se reduce a $E(\beta_{1i}\pi_{1i})/E(\pi_{1i}) = E(\beta_{1i})E(\pi_{1i})/E(\pi_{1i}) = E(\beta_{1i})$.

En cada uno de estos tres casos, existe heterogeneidad poblacional en el efecto del instrumento, en el efecto del tratamiento, o en ambas, pero el promedio local del efecto del tratamiento es igual al efecto promedio del tratamiento. Es decir, en los tres casos, MC2E es un estimador consistente del efecto promedio del tratamiento.

A parte de estos tres casos particulares, en general, el promedio local del efecto del tratamiento es distinto del efecto promedio del tratamiento. Por ejemplo, supongamos que Z_i no tiene ninguna influencia en la decisión sobre el tratamiento para la mitad de la población (para los que $\pi_{1i} = 0$), y que Z_i tiene la misma influencia, distinta de cero sobre la decisión del tratamiento para la otra mitad (para ellos, π_{1i} es una constante distinta de cero). Entonces MC2E es un estimador consistente del promedio del efecto del tratamiento en la mitad de la población para la que el instrumento influye sobre la decisión del tratamiento. Para concretar, supongamos que los trabajadores cumplen los requisitos para recibir un programa de formación para el empleo y que se les asigna aleatoriamente un número de prioridad Z , que influye en la probabilidad de ser admitidos en el programa. La mitad de los trabajadores sabe que se beneficiarán del programa y por lo tanto pueden decidir inscribirse en el programa; para ellos, $\beta_{1i} = \beta_1^+ > 0$ y $\pi_{1i} = \pi_1^+ > 0$. La otra mitad saben que, para ellos, el programa es ineficaz, por lo que no se inscriben aunque se les admite, es decir, para ellos $\beta_{1i} = 0$ y $\pi_{1i} = 0$. El efecto promedio del tratamiento es $E(\beta_{1i}) = \frac{1}{2}(\beta_1^+ + 0) = \frac{1}{2}\beta_1^+$. El promedio local del efecto del tratamiento es $E(\beta_{1i}\pi_{1i})/E(\pi_{1i})$. Ahora $E(\pi_{1i}) = \frac{1}{2}\pi_1^+$ y $E(\beta_{1i}\pi_{1i}) = E[\beta_{1i}E(\pi_{1i}|\beta_{1i})] = \frac{1}{2}(0 + \beta_1^+\pi_1^+) = \frac{1}{2}\beta_1^+\pi_1^+$, por lo que $E(\beta_{1i}\pi_{1i})/E(\pi_{1i}) = \beta_1^+$. Por tanto, en este ejemplo, el promedio local del efecto del tratamiento es el efecto causal para aquellos trabajadores que probablemente se inscriban en el programa, y no se pondera a los que no se inscribirán bajo ninguna circunstancia. Por el contrario, el efecto promedio del tratamiento pondera del mismo modo a todos los individuos, independientemente de que se inscriban o no. Debido a que los individuos deciden inscribirse en parte en base a su conocimiento acerca de la eficacia que el programa tenga en ellos, en este ejemplo, el promedio local del efecto del tratamiento supera al efecto medio del tratamiento.

Implicaciones. Si la decisión de un individuo de recibir tratamiento depende de la efectividad del tratamiento para ese individuo, entonces el estimador MC2E no es, en general, un estimador consistente del efecto causal promedio. En cambio, MC2E estima un promedio local de los efectos del tratamiento, en el que los efectos causales de los individuos que están más influidos por el instrumento reciben las mayores ponderaciones. Esta conclusión lleva a una situación desconcertante en la que dos investigadores, armados con diferentes variables instrumentales que son válidas en el sentido de que ambas son relevantes y

⁷ Siglas de *Local Average Treatment Effect* en el idioma inglés original.

exógenas, podrían obtener estimaciones diferentes sobre «el» efecto causal, incluso en muestras grandes. La diferencia se debe a que cada investigador está estimando implícitamente una media ponderada distinta de los efectos causales individuales de la población. De hecho, un contraste J para la sobreidentificación de restricciones podría dar como resultado el rechazo si los dos instrumentos estiman diferentes promedios locales de los efectos del tratamiento, aunque ambos instrumentos sean válidos. Aunque los dos estimadores proporcionan información sobre la distribución de los efectos causales mediante sus respectivas medias ponderadas de la forma dada por la Ecuación (13.12), en general, ningún estimador es un estimador consistente del efecto causal⁸.

Ejemplo: el estudio sobre cateterismo cardiaco. Las Secciones 12.5 y 13.4 analizan el estudio de McClellan, McNeil, y Newhouse (1991) acerca del efecto sobre la mortalidad del cateterismo cardiaco en los pacientes con ataques al corazón. Los autores utilizaron la regresión de variables instrumentales, con la distancia relativa a un hospital que dispusiera de la técnica de cateterismo cardiaco, como variable instrumental. En base a sus estimaciones MC2E, hallaron que el cateterismo cardiaco tenía escaso o ningún efecto sobre los resultados en salud. Este resultado es sorprendente: los procedimientos médicos como el cateterismo cardiaco se someten a rigurosas pruebas clínicas antes de su aprobación para uso generalizado. Además, el cateterismo cardiaco permite a los cirujanos realizar intervenciones médicas, que habrían requerido una cirugía mayor en la década anterior, lo que hace que estas intervenciones sean más seguras y, presumiblemente, mejores a largo plazo para los pacientes cardíacos. ¿Cómo pudo no encontrar este estudio econométrico los efectos beneficiosos del cateterismo cardiaco?

Una posible respuesta es que existe heterogeneidad en el efecto del tratamiento del cateterismo cardiaco. Para algunos pacientes, este procedimiento resulta una intervención efectiva, pero para otros, quizás aquellos más sanos, resulta menos eficaz o, debido a los riesgos que implica cualquier cirugía, tal vez ineficaz en términos netos. Por tanto el efecto causal promedio en la población de pacientes con ataque cardíaco podría ser, y presumiblemente lo es, positivo. Sin embargo, el estimador VI, mide un efecto marginal y no un efecto medio, siendo el efecto marginal el efecto del procedimiento sobre los pacientes para los que la distancia al hospital es un factor importante en el hecho de que reciban o no tratamiento. No obstante, esos pacientes podrían ser justo los pacientes relativamente sanos para los que, en el margen, el cateterismo cardíaco resulta un procedimiento relativamente ineficaz. Si es así, el estimador MC2E de McClellan, McNeil, y Newhouse (1991) mediría el efecto del procedimiento para el paciente marginal (para el cual resulta relativamente ineficaz), no para el paciente promedio (para el cual podría resultar eficaz).

13.7 Conclusión

En el Capítulo 1, se definía el efecto causal en términos del resultado esperado de un experimento aleatorizado controlado ideal. Si está disponible un experimento aleatorizado controlado o este se puede llevar a cabo, puede proporcionar pruebas convincentes acerca del efecto causal estudiado, aunque incluso los experimentos aleatorizados controlados estén sujetos a potenciales amenazas importantes a la validez interna y externa.

A pesar de sus ventajas, en economía, los experimentos aleatorizados controlados afrontan obstáculos considerables, tales como problemas y costes éticos. Los resultados de los métodos experimentales pueden, sin embargo, aplicarse a los cuasi experimentos, en los que las circunstancias particulares hacen que parezca «como si hubiera» aleatoriedad. En los cuasi experimentos, el efecto causal puede estimarse utilizando un estimador de diferencias en diferencias, posiblemente ampliado con regresores adicionales; si la asignación

⁸ Estos son algunos estudios buenos (y avanzados) sobre el efecto de la heterogeneidad de la población sobre los estimadores en la evaluación de programas. Estos incluyen el repaso a los estudios existentes realizado por Heckman, Lalonde y Smith (1999, Sección 7) y la conferencia de James Heckman en la entrega del Premio Nobel de Economía (Heckman, 2001, Sección 7). Esta última referencia, así como el trabajo de Angrist, Graddy, e Imbens (2000), proporcionan un estudio detallado del modelo de efectos aleatorios (que considera un β_{1i} que varía entre los distintos individuos) y proporcionan versiones más generales del resultado de la Ecuación (13.12). El concepto de promedio local del efecto del tratamiento se introdujo por Angrist e Imbens (1994), que demostraron que, en general, no es igual al efecto medio del tratamiento.

«como si fuera» aleatoriedad solamente influye en parte en el tratamiento, entonces puede utilizarse en su lugar la regresión de variables instrumentales. Una ventaja importante de los cuasi experimentos es que la fuente de la «como si hubiera» aleatoriedad en los datos, habitualmente es evidente y por tanto puede evaluarse de un modo concreto. Una amenaza importante a la que se enfrentan los cuasi experimentos es que a veces lo considerado «como si fuera» aleatorio no es aleatorio en realidad, por lo que el tratamiento (o la variable instrumental) está correlacionado con variables omitidas y el estimador resultante del efecto causal está sesgado.

Los cuasi experimentos proporcionan un puente entre los conjuntos datos que proceden de observaciones y los verdaderos experimentos aleatorizados controlados. Los métodos econométricos utilizados en este capítulo para analizar los cuasi experimentos son los desarrollados, en diferentes contextos, en los capítulos anteriores: MCO, métodos de estimación con datos de panel, y regresión de variables instrumentales. Lo que diferencia a los cuasi experimentos de los casos examinados en la Parte II y en los primeros capítulos de la Parte III es la manera de interpretar las bases de datos para las que se aplican. Los cuasi experimentos proporcionan a los económetras un modo de pensar sobre cómo obtener nuevos datos, cómo pensar las variables instrumentales, y cómo evaluar la credibilidad del supuesto de exogeneidad que subyace tras la estimación MCO y de variables instrumentales⁹.

Resumen

1. El efecto causal promedio en la población estudiada es la diferencia estimada en los resultados promedio de los grupos de tratamiento y de control en un experimento aleatorizado controlado ideal. Los experimentos reales con sujetos humanos se apartan de un experimento ideal por varias razones prácticas, entre las que se encuentra el incumplimiento del protocolo experimental.
2. Si el nivel de tratamiento *real* X_i es aleatorio, entonces se puede estimar el efecto del tratamiento mediante una regresión de los resultados del tratamiento. Si el tratamiento *asignado* Z_i es aleatorio, pero el tratamiento real X_i está parcialmente determinado por una decisión individual, entonces el efecto causal puede estimarse mediante regresión de variables instrumentales utilizando Z_i como instrumento. Si el tratamiento (o el tratamiento asignado) es aleatorio condicionado a algunas variables W , esas variables de control deben incluirse en las regresiones.
3. En un cuasi experimento, las variaciones del contexto legal o en las circunstancias o los accidentes de la naturaleza son tratados «como si» indujeran la asignación aleatoria a los grupos de tratamiento y de control. Si el tratamiento real es «como si fuera» aleatorio, el efecto causal puede estimarse mediante una regresión (posiblemente con características pretratamiento adicionales como regresores); si el tratamiento asignado es «como si fuera» aleatorio, entonces, el efecto causal puede ser estimado mediante regresión de variables instrumentales.
4. Una amenaza clave a la validez interna de un estudio cuasi experimental la constituye el hecho de si la asignación «como si fuera» aleatoria en realidad da como resultado la exogeneidad. Debido a las reacciones que se producen en la conducta, el simple hecho de que el instrumento esté generado de forma «como si fuera» aleatoria no significa que sea necesariamente exógeno en el sentido requerido para que una variable instrumental sea válida.
5. Cuando el efecto del tratamiento varía de un individuo a otro, el estimador MCO es un estimador consistente del efecto causal promedio si el tratamiento real se asigna aleatoriamente o es «como si fuera» asignado aleatoriamente. Sin embargo, el estimador de variables instrumentales es una media ponderada de los efectos individuales del tratamiento, en el que los individuos para los cuales el instrumento es más influyente reciben mayor ponderación.

⁹ Shadish, Cook y Campbell (2002) proporcionan un tratamiento integral para los experimentos y cuasi-experimentos en ciencias sociales y en psicología. Una línea de investigación importante en economía del desarrollo es la que se centra en las evaluaciones experimentales de programas de salud y educación en los países en desarrollo. Para algunos ejemplos, véase Kremer, Miguel, y Thornton (2009) y la página web del Poverty Action Laboratory del MIT (<http://www.povertyactionlab.org>).

Términos clave

- evaluación de programas (339)
- variables respuesta (340)
- efecto causal promedio (340)
- efecto promedio del tratamiento (340)
- estimador de las diferencias (341)
- estimador de las diferencias con regresores adicionales (341)
- aleatorización basada en las covariables (341)
- contraste de la recepción aleatoria del tratamiento (342)
- cumplimiento parcial (343)
- estimación por variables instrumentales del efecto del tratamiento (343)
- deserción o abandono (343)
- efecto Hawthorne (343)
- cuasi experimento (352)
- experimento natural (352)
- estimador de diferencias en diferencias (354)
- estimador de diferencias en diferencias con regresores adicionales (355)
- datos de sección cruzada repetida (356)
- discontinuidad de la regresión (358)
- promedio local del efecto del tratamiento (362)

Revisión de conceptos

- 13.1** Un investigador que estudia los efectos de un nuevo fertilizante sobre el rendimiento de los cultivos planea llevar a cabo un experimento en el que se aplicarían diferentes cantidades de fertilizante a 100 parcelas diferentes de 1 acre de tierra. Habría cuatro niveles de tratamiento. El nivel uno de tratamiento es sin fertilizante, el nivel dos de tratamiento es una cantidad igual al 50 % de la cantidad recomendada por el fabricante de fertilizantes, el nivel tres de tratamiento es el 100 % de esa cantidad, y el nivel cuatro de tratamiento es el 150 %. El investigador planea aplicar el nivel uno de tratamiento a las primeras 25 parcelas de tierra, el nivel dos de tratamiento a las segundas 25 parcelas, y así sucesivamente. ¿Puede sugerir una mejor manera de asignar los niveles de tratamiento? ¿Por qué su propuesta es mejor que el método del investigador?
- 13.2** Se lleva a cabo un ensayo clínico para un nuevo medicamento reductor de colesterol. El medicamento se dispensa a 500 pacientes y un placebo a otros 500 pacientes, mediante asignación aleatoria de los pacientes a uno u otro grupo. ¿Cómo estimaría el efecto del tratamiento del medicamento? Supóngase que se dispone de datos sobre el peso, la edad y el género de cada paciente. ¿Podrían utilizarse estos datos para mejorar la estimación? Explíquelo. Supóngase que se dispone de datos sobre los niveles de colesterol de cada paciente antes de que él o ella entraran en el experimento. ¿Podrían utilizarse estos datos para mejorar la estimación? Explíquelo.
- 13.3** Los investigadores que estudian los datos STAR presentan una evidencia anecdótica de que los directores de las escuela fueron presionados por algunos padres a ubicar a sus hijos en las clases pequeñas. Supóngase que algunos directores sucumbieron a esta presión y transfirieron a algunos niños a las clases pequeñas. ¿Cómo podrían comprometer estas transferencias la validez interna del estudio? Supóngase que se dispone de datos sobre la asignación aleatoria original de cada estudiante antes de la intervención del director. ¿Cómo podría utilizarse esta información para restaurar la validez interna del estudio?
- 13.4** Explique si los efectos experimentales (como el efecto Hawthorne) podrían ser importantes en cada uno de los experimentos de las tres preguntas anteriores.
- 13.5** La Sección 12.1 proporcionaba un ejemplo hipotético en el que algunas escuelas resultaban dañadas por un terremoto. Explique por qué esto constituye un ejemplo de cuasi experimento. ¿Cómo podrían utilizarse los cambios inducidos en el tamaño de las clases para estimar el efecto del tamaño de las clases sobre las calificaciones en la prueba?

Ejercicios

- 13.1** Utilizando los resultados de la Tabla 13.1, calcule para cada curso lo siguiente: una estimación del efecto del tratamiento de clase pequeña, en relación con la clase mediana; su error estándar; así como

su intervalo de confianza al 95 %. (Para este ejercicio, hágase caso omiso a los resultados de las clases medianas con ayudante).

- 13.2** Para los siguientes cálculos, utilice los resultados de la columna (4) de la Tabla 13.2. Considérense dos aulas, A y B, con los mismos valores de los regresores de la columna (4) de la Tabla 13.2, salvo que:

- El aula A es una «clase pequeña» y el aula B es una «clase mediana». Construya un intervalo de confianza al 95 % para la diferencia esperada en las calificaciones medias en el examen.
- El aula A tiene un maestro con 5 años de experiencia y el aula B tiene un maestro con 10 años de experiencia. Construya un intervalo de confianza al 95 % para la diferencia esperada en las medias de las calificaciones.
- El aula A es una clase pequeña con un maestro con 5 años de experiencia y la clase B es una clase mediana con un maestro con 10 años de experiencia. Construya un intervalo de confianza al 95 % para la diferencia esperada en las calificaciones medias. (*Sugerencia:* en el estudio STAR, los maestros fueron asignados aleatoriamente a los diferentes tipos de aulas).
- ¿Por qué desaparece el intercepto en la columna (4)?

- 13.3** Supóngase que, en un experimento aleatorizado controlado sobre el efecto de un curso de preparación para SAT sobre las calificaciones en el SAT, se obtienen los resultados siguientes:

- Estime el efecto promedio del tratamiento en las calificaciones obtenidas.
- ¿Existe evidencia de asignación no aleatoria? Explíquelo.

	Grupo de tratamiento	Grupo de control
Calificación media SAT (\bar{X})	1.241	1.201
Desviación típica de las calificaciones SAT (s_x)	93,2	97,1
Número de hombres	55	45
Número de mujeres	45	55

- 13.4** Lea el recuadro «¿Cuál es el efecto sobre el empleo del salario mínimo?» de la Sección 13.4. Supóngase, para concretar, que Card y Krueger recopilaron sus datos en 1991 (antes del cambio en el salario mínimo de Nueva Jersey) y en 1993 (después del cambio en el salario mínimo de Nueva Jersey). Considérese la Ecuación (13.7) excluyendo los regresores W .

- ¿Cuáles son los valores de X_{it} , G_i , y D_t para:
 - Un restaurante de Nueva Jersey en 1991?
 - Un restaurante de Nueva Jersey en 1993?
 - Un restaurante de Pensylvania en 1991?
 - Un restaurante de Pensylvania en 1993?
- En términos de los coeficientes β_0 , β_1 , β_2 , y β_3 , ¿cuál es el número esperado de empleados en:
 - Un restaurante de Nueva Jersey en 1991?
 - Un restaurante de Nueva Jersey en 1993?
 - Un restaurante de Pensylvania en 1991?
 - Un restaurante de Pensylvania en 1993?
- En términos de los coeficientes β_0 , β_1 , β_2 , y β_3 , ¿cuál es el efecto causal promedio sobre el empleo del salario mínimo?

- d)** Explique por qué Card y Krueger utilizaron un estimador de diferencias en diferencias del efecto causal en lugar del estimador de las diferencias «Nueva Jersey después – Nueva Jersey antes» o el estimador de las diferencias «Nueva Jersey 1993 – Pennsylvania 1993».
- 13.5** Considérese un estudio que evalúe el efecto sobre las calificaciones de los estudiantes universitarios de las conexiones a Internet en las habitaciones. En un dormitorio grande, la mitad de las habitaciones están cableadas para conexiones a Internet de alta velocidad (grupo de tratamiento), se recopilan las notas finales de todos los residentes. De lo siguiente, ¿qué plantea una amenaza a la validez interna, y por qué?
- A mediados de año todos los atletas de sexo masculino se trasladan a una fraternidad y abandonan el estudio (no se recogen sus calificaciones finales).
 - Los estudiantes de ingeniería asignados al grupo de control montan una red de área local por lo que pueden compartir un acceso inalámbrico privado a Internet que se paga en conjunto.
 - Los estudiantes de arte del grupo de tratamiento no saben cómo acceder a sus cuentas de Internet.
 - Los estudiantes de economía del grupo de tratamiento facilitan el acceso a sus conexiones de Internet a los del grupo de control, a cambio de que paguen una tarifa.
- 13.6** Supóngase que existen datos de panel para $T = 2$ períodos de tiempo sobre un experimento aleatorizado controlado, en el que se toma la primera observación ($t = 1$) antes del experimento y la segunda observación ($t = 2$) para el periodo postratamiento. Supóngase que el tratamiento es binario; es decir, supóngase que $X_{it} = 1$ si el individuo i -ésimo está en el grupo de tratamiento y $t = 2$, y $X_{it} = 0$ en otro caso. Supóngase además que el efecto del tratamiento puede ser modelizado utilizando la especificación

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it},$$

donde α_i son los efectos individuales específicos [véase la Ecuación (13.11)], con una media igual a cero y una varianza de σ_α^2 y u_{it} es un término de error, donde u_{it} es homocedástico, $\text{cov}(u_{i1}, u_{i2}) = 0$, y $\text{cov}(u_{it}, \alpha_i) = 0$ para todo i . Sea $\hat{\beta}_1^{\text{diferencias}}$ el estimador de las diferencias, es decir, el estimador MCO de una regresión de Y_{i2} sobre X_{i2} con un intercepto, y sea $\hat{\beta}_1^{\text{difs-en-difs}}$ el estimador de diferencias en diferencias, es decir, el estimador de β_1 basado en la regresión MCO de $\Delta Y_i = Y_{i2} - Y_{i1}$ sobre $\Delta X_i = X_{i2} - X_{i1}$ y un intercepto.

- Demuestre que $n\text{var}(\hat{\beta}_1^{\text{diferencias}}) \rightarrow (\sigma_u^2 + \sigma_\alpha^2)/\text{var}(X_{i2})$ (Pista: utilice las fórmulas válidas con homocedasticidad de la varianza del estimador MCO del Apéndice 5.1).
 - Demuestre que $n\text{var}(\hat{\beta}_1^{\text{difs-en-difs}}) \rightarrow (2\sigma_u^2)/\text{var}(X_{i2})$. (Sugerencia: téngase en cuenta que $X_i = X_{i2}$; ¿por qué?)
 - En base a las respuestas de los apartados (a) y (b), ¿cuándo se prefiere el estimador de diferencias en diferencias sobre el estimador de diferencias, considerando únicamente las cuestiones de eficiencia?
- 13.7** Supóngase que se dispone de datos de panel sobre un experimento con $T = 2$ períodos (por lo que $t = 1, 2$). Considérese el modelo de regresión de datos de panel con efectos fijos temporales y con características individuales W_i que no cambian en el tiempo, como el género. Sea el tratamiento binario, por lo que $X_{it} = 1$ para $t = 2$ para los individuos en el grupo de tratamiento y sea $X_{it} = 0$ en caso contrario. Considérese el modelo de regresión poblacional

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2(D_t \times W_i) + \beta_0 D_t + v_{it},$$

donde α_i son los efectos fijos individuales, D_t es la variable binaria que es igual a 1 si $t = 2$ y es igual a 0 si $t = 1$, $D_t \times W_i$ es el producto de D_t y W_i , y las α y β son coeficientes desconocidos. Sea $\Delta Y_i = Y_{i2} - Y_{i1}$. Deduzca la Ecuación (13.6) (en el caso de un solo regresor W , por lo que $r = 1$) a partir de este modelo de regresión poblacional.

- 13.8** Supóngase que se dispone de los mismos datos que en el Ejercicio 13.7 (datos de panel con dos periodos, n observaciones), pero se ignora el regresor W . Considérese el modelo de regresión alternativo

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 G_i + \beta_3 D_t + u_{it},$$

donde $G_i = 1$ si el individuo está en el grupo de tratamiento y $G_i = 0$ si el individuo se encuentra en el grupo de control. Demuestre que el estimador MCO de β_1 es el estimador de diferencias en diferencias de la Ecuación (13.4). (*Sugerencia:* véase la Sección 8.3).

- 13.9** Deduzca la última igualdad de la Ecuación (13.10). (*Sugerencia:* utilice la definición de la covarianza y que, debido a que el tratamiento real X_i es aleatorio, β_{1i} y X_i se distribuyen de forma independiente).

- 13.10** Considérese el modelo de regresión con coeficientes de regresión heterogéneos

$$Y_i = \beta_{0i} + \beta_{1i} X_i + v_i,$$

donde $(v_i, X_i, \beta_{0i}, \beta_{1i})$ son variables aleatorias i.i.d. con $\beta_0 = E(\beta_{0i})$ y $\beta_1 = E(\beta_{1i})$.

- a) Demuestre que el modelo puede escribirse como $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i = (\beta_{0i} - \beta_0) + (\beta_{1i} - \beta_1)X_i + v_i$.
 - b) Supóngase que $E[\beta_{0i}|X_i] = \beta_0$, que $E[\beta_{1i}|X_i] = \beta_1$ y $E[v_i|X_i] = 0$. Demuestre que $E[u_i|X_i] = 0$.
 - c) Demuestre que los supuestos 1 y 2 del Concepto clave 4.3 se cumplen.
 - d) Supóngase que los valores extremos son poco frecuentes, por lo que (u_i, X_i) tienen momentos de cuarto orden finitos. ¿Es apropiado utilizar MCO y los métodos de los Capítulos 4 y 5 para estimar y llevar a cabo inferencias acerca de los valores medios de β_{0i} y β_{1i} ?
 - e) Supóngase que β_{1i} y X_i están correlacionados positivamente por lo que las observaciones de X_i con valores más altos que la media tienden a tener valores mayores a la media de β_{1i} . ¿Se cumplen los supuestos del Concepto clave 4.3? Si no es así, ¿qué supuesto(s) se viola(n)? ¿Resulta apropiado utilizar MCO y los métodos de los Capítulos 4 y 5 para estimar y llevar a cabo inferencia sobre el valor medio de β_{0i} y β_{1i} ?
- 13.11** En el Capítulo 12, se utilizaron datos de panel a nivel estatal para estimar la elasticidad-precio de la demanda de cigarrillos, utilizando el impuesto estatal sobre las ventas como variable instrumental. Considérese en particular la regresión (1) de la Tabla 12.1. En este caso, a su juicio, ¿difiere el promedio local del efecto del tratamiento del efecto promedio del tratamiento? Explíquelo.

Ejercicios empíricos

- E13.1** Un eventual empleador recibe dos currículum vítae: un currículum de un candidato blanco y un currículum similar de un candidato afroamericano. ¿Resulta más probable que el empleador llame al candidato blanco para concertar una entrevista? Marianne Bertrand y Sendhil Mullainathan llevaron a cabo un experimento aleatorizado controlado para responder a esta pregunta. Debido a que la etnia no suele incluirse habitualmente en un currículum, diferenciaron los currículos en base a nombres que «suenan de raza blanca» (tales como Emily Walsh o Gregory Baker) y nombres «que suenan de afroamericanos» (como Lakisha Washington o Jamal Jones). Se creó una gran colección de currículos ficticios, y a los que se les asignó aleatoriamente la supuesta «etnia» (basada en el «sonido» del nombre). Estos currículum fueron enviados a los posibles empleadores para comprobar cuáles de los currículos generaban una llamada de teléfono (una «devolución de la llamada») del posible empleador. Los datos del experimento y su descripción detallada se encuentran en la página web del libro de texto http://www.pearsonhighered.com/stock_watson en los archivos **Names** y **Names_Description**¹⁰.

¹⁰ Estos datos fueron proporcionados por la profesora Marianne Bertrand, de la Universidad de Chicago y fueron utilizados en su artículo conjunto con Sendhil Mullainathan, «Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination», *American Economic Review*, 2004, 94 (4): 991-1013.

- a) Defina la «tasa de devolución de llamada» como la proporción de currículos que generaron una llamada telefónica del futuro empleador. ¿Cuál fue la tasa de devolución de llamada para los blancos? ¿Y para los afroamericanos? Construya un intervalo de confianza al 95 % para la diferencia en las tasas de devolución de llamada. ¿Es estadísticamente significativa la diferencia? ¿Es elevada en el sentido del mundo real?
- b) Es diferente la tasa de devolución de llamada afroamericana/blanca para hombres y mujeres?
- c) ¿Cuál es la diferencia en las tasas de devolución de llamada en los currículos de elevada cualificación en comparación con los currículos de baja cualificación? ¿Cuál es la diferencia entre alta cualificación/baja cualificación para los solicitantes de raza blanca? ¿Y para los solicitantes afroamericanos? ¿Existe una diferencia significativa en la diferencia alta cualificación/baja cualificación para blancos y afroamericanos?
- d) Los autores del estudio afirman que la etnia fue asignada aleatoriamente a cada currículum. ¿Existe evidencia de asignación no aleatoria?
- E13.2** A un consumidor se le da la oportunidad de comprar una tarjeta de béisbol por 1 \$, pero no acepta el trato. Si al consumidor se le da ahora la tarjeta de béisbol, ¿estaría dispuesto a venderla por 1 \$? La teoría del consumidor convencional sugiere que sí, pero los economistas del comportamiento han hallado que la «propiedad» tiende a aumentar el valor de los bienes para los consumidores. Es decir, el consumidor puede esperar una cantidad de más de 1 \$ (por ejemplo, 1,20 \$) por la venta de la tarjeta, a pesar de que él estaba dispuesto a pagar solamente una cierta cantidad menor a 1 \$ (por ejemplo, 0,88 \$) cuando considerara la compra. Los economistas del comportamiento llaman a este fenómeno el «efecto dotación». John List investigó el efecto dotación en un experimento aleatorizado con comerciantes de recuerdos deportivos en una exposición de tarjetas deportivas. A los comerciantes se les asignó aleatoriamente uno de los dos artículos deportivos de colección disponibles, sean por ejemplo los bienes A o B, que tenían aproximadamente el mismo valor de mercado¹¹. Los que recibieron el bien A tenían la opción de intercambiar el bien A por el bien B con el experimentador, a los que recibieron el bien B asimismo se les dio la opción de intercambiar el bien B por el bien A con el experimentador. Los datos del experimento así como su descripción detallada se pueden encontrar en la página web del libro de texto http://www.pearsonhighered.com/stock_watson en los archivos **Sportscards** y **Sportscards_Description**¹².
- a)
- I) Supóngase que, en ausencia de cualquier efecto dotación, todos los sujetos prefieren el bien A al bien B. ¿Qué proporción de sujetos del experimento podría esperarse que intercambiaría el bien que recibieron por el otro bien? (*Sugerencia:* asignación aleatoria significa que aproximadamente el 50 % de los sujetos recibieron el bien A y el otro 50 % recibieron el bien B).
 - II) Supongamos que, en ausencia de cualquier efecto dotación, el 50 % de los sujetos prefiere el bien A al bien B, y el otro 50 % prefiere el bien B al bien A. ¿Qué proporción de sujetos podría esperarse que intercambiara el bien que recibieron por el otro bien?
 - III) Supongamos que, en ausencia de cualquier efecto dotación, el $X\%$ de los sujetos prefieren el bien A al bien B, y el otro $(100 - X)\%$ prefiere el bien B al bien A. Demuestre que podría esperarse que el 50 % de los sujetos intercambien el bien que se les dio por el otro bien.
- b) Utilizando los datos sobre tarjetas deportivas, ¿qué proporción de los sujetos intercambiaron el bien que se les dio? Es esta proporción significativamente distinta del 50 %? ¿Qué proporción de los sujetos que recibieron el bien A lo intercambiaron por el bien B? ¿Qué proporción de los sujetos que recibieron el bien B lo intercambiaron por el bien A? ¿Existe evidencia de un efecto dotación?

¹¹ El bien A era una entrada para el partido en el que Cal Ripken, Jr., estableció el récord de partidos consecutivos jugados, y el bien B era un recuerdo del partido en el que Nolan Ryan ganó su partido número 300.

¹² Estos datos fueron proporcionados por el profesor John List, de la Universidad de Chicago y fueron utilizados en su artículo «Does Market Experience Eliminate Market Anomalies», *Quarterly Journal of Economics*, 2003, 118(1): 41-71.

- c) Algunos han argumentado que el efecto dotación puede existir, pero que es probable que desaparezca a medida que los comerciantes adquieren más experiencia comercial. La mitad de los sujetos del experimento eran comerciantes y la otra mitad no eran comerciantes. Los comerciantes tienen más experiencia que el resto. Repita el apartado (b) para los comerciantes y para los no comerciantes. ¿Existe una diferencia significativa en su comportamiento? ¿Es esta evidencia consistente con la hipótesis de que el efecto dotación desaparece a medida que los operadores adquieren más experiencia?
- d) El conjunto de datos contiene dos medidas adicionales de la experiencia: el número de transacciones por mes y el número de años operando. ¿Existe evidencia de que el efecto dotación para los no comerciantes disminuye a medida que aumenta su experiencia en transacciones?

APÉNDICE

13.1 La base de datos del proyecto STAR

La base de datos de acceso público del proyecto STAR contiene datos sobre las calificaciones en las pruebas, los grupos de tratamiento, y las características de los estudiantes y profesores para los 4 años del experimento, a partir del año académico 1985-1986 hasta el año académico 1988-1989. Los datos de las calificaciones en las pruebas analizados en este capítulo son la suma de las puntuaciones en las partes de matemáticas y lectura del examen de Stanford. La variable binaria «Niño» de la Tabla 13.2 indica si el alumno es un niño ($= 1$) o una niña ($= 0$); las variables binarias «Negro» y «Etnia distinta de blanco o negro», indican la etnia del estudiante. La variable «Cumplir requisitos para comedor gratuito» indica si el estudiante reúne los requisitos para obtener una ayuda para comedor gratuito durante el curso académico. Los años de experiencia del maestro es el número total de años de experiencia del maestro que el estudiante tenía en el curso para el que se suministran los datos de la prueba. La base de datos indica asimismo a qué escuela asistió el estudiante en un año determinado, por lo que es posible construir variables binarias indicadoras específicas de la escuela.

APÉNDICE

13.2 Estimación VI con efectos causales que varían entre individuos

En este apéndice se obtiene el límite de probabilidad del estimador MC2E de la Ecuación (13.12), cuando existe heterogeneidad en la población en el efecto del tratamiento y en la influencia del instrumento sobre la aplicación del tratamiento. En concreto, se supone que se cumplen los supuestos de la regresión VI del Concepto clave 12.4, salvo que las Ecuaciones (13.9) y (13.11) se cumplen con efectos heterogéneos. Se supone además, que $\pi_{0,i}$, $\pi_{1,i}$, β_{0i} , y β_{1i} se distribuyen de forma independiente de u_i , v_i , y Z_i ; que $E(u_i|Z_i) = E(v_i|Z_i) = 0$; y que $E(\pi_{1,i}) \neq 0$.

Debido a que (X_i, Y_i, Z_i) , $i = 1, \dots, n$, son i.i.d. con momentos de cuarto orden, es aplicable la ley de los grandes números del Concepto clave 2.6 y

$$\hat{\beta}_1^{MC2E} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\sigma_{ZY}}{\sigma_{ZX}}. \quad (13.13)$$

(Véanse el Apéndice 3.3 y el Ejercicio 17.2). La tarea por tanto consiste en obtener las expresiones para σ_{ZY} y σ_{ZX} en términos de los momentos de π_{1i} y β_{1i} . Ahora $\sigma_{ZX} = E[(Z_i - \mu_Z)(X_i - \mu_X)] = E[(Z_i - \mu_Z)X_i]$. Al sustituir la Ecuación (13.11) en esta expresión de σ_{ZX} se obtiene

$$\begin{aligned} \sigma_{ZX} &= E[(Z_i - \mu_Z)(\pi_{0i} + \pi_{1i}Z_i + v_i)] \\ &= E(\pi_{0i}) \times 0 + E[\pi_{1i}Z_i(Z_i - \mu_Z)] + \text{cov}(Z_i, v_i) \\ &= \sigma_Z^2 E(\pi_{1i}), \end{aligned} \quad (13.14)$$

donde la segunda igualdad se deduce debido a que $\text{cov}(Z_i, v_i) = 0$ [que se deriva del supuesto $E(v_i|Z_i) = 0$; véase la Ecuación (2.27)], debido a que $E[(Z_i - \mu_Z)\pi_{0,i}] = E\{E[(Z_i - \mu_Z)\pi_{0,i}|Z_i]\} = E[(Z_i - \mu_Z)E(\pi_{0,i}|Z_i)] = E(Z_i - \mu_Z) \times E(\pi_{0,i})$ (que utiliza la ley de esperanzas iteradas y el supuesto de que $\pi_{0,i}$ es independiente de Z_i), y debido a que $E[\pi_{1,i}Z_i(Z_i - \mu_Z)] = E\{E[\pi_{1,i}Z_i(Z_i - \mu_Z)|Z_i]\} = E(\pi_{1,i})E[Z_i(Z_i - \mu_Z)] = \sigma_Z^2E(\pi_{1,i})$ (que utiliza la ley de esperanzas iteradas y el supuesto de que $\pi_{1,i}$ es independiente de Z_i).

Ahora consideramos σ_{ZY} . Sustituyendo la Ecuación (13.11) en la Ecuación (13.9) se obtiene $Y_i = \beta_{0i} + \beta_{1i}(\pi_{0i} + \pi_{1i}Z_i + v_i) + u_i$, por lo que

$$\begin{aligned}\sigma_{ZY} &= E[(Z_i - \mu_Z)Y_i] \\ &= E[(Z_i - \mu_Z)(\beta_{0i} + \beta_{1i}\pi_{0i} + \beta_{1i}\pi_{1i}Z_i + \beta_{1i}v_i + u_i)] \\ &= E(\beta_{0i}) \times 0 + \text{cov}(Z_i, \beta_{1i}\pi_{0i}) \\ &\quad + E[\beta_{1i}\pi_{1i}Z_i(Z_i - \mu_Z)] + E[\beta_{1i}v_i(Z_i - \mu_Z)] + \text{cov}(Z_i, u_i).\end{aligned}\quad (13.15)$$

Debido a que $(\beta_{1i}\pi_{0,i})$ y Z_i están distribuidos de forma independiente, $\text{cov}(Z_i, \beta_{1i}\pi_{0,i}) = 0$; debido a que está distribuido independientemente de v_i y Z_i y $E(v_i, Z_i) = 0$, $E[\beta_{1i}v_i(Z_i - \mu_Z)] = E(\beta_{1i})E[v_i(Z_i - \mu_Z)] = 0$; debido a que $E(u_i|Z_i) = 0$, $\text{cov}(Z_i, u_i) = 0$; y debido a que β_{1i} y $\pi_{1,i}$ están distribuidos independientemente de Z_i , $E[\beta_{1i}\pi_{1,i}Z_i(Z_i - \mu_Z)] = \sigma_Z^2E(\beta_{1i}\pi_{1,i})$. Por lo tanto la expresión final de la Ecuación (13.15) queda como

$$\sigma_{ZY} = \sigma_Z^2E(\beta_{1i}\pi_{1i}). \quad (13.16)$$

Sustituyendo las Ecuaciones (13.14) y (13.16) en la Ecuación (13.13) se obtiene $\hat{\beta}_1^{MC2E} \xrightarrow{P} \sigma_Z^2E(\beta_{1i}\pi_{1i})/\sigma_Z^2E(\pi_{1,i}) = E(\beta_{1i}\pi_{1,i})/E(\pi_{1,i})$, que es el resultado enunciado en la Ecuación (13.12).

APÉNDICE

13.3 El marco de las variables respuesta para el análisis de datos procedentes de experimentos

En este apéndice se ofrece un tratamiento matemático del marco de análisis de las variables respuesta tratado en la Sección 13.1. El marco de las variables respuesta, en combinación con un efecto del tratamiento constante, implica el modelo de regresión de la Ecuación (13.1). Si la asignación es aleatoria condicionada a las covariables, el marco de las variables respuesta lleva a la Ecuación (13.2) y a la independencia en media condicional. Consideramos un tratamiento binario con $X_i = 1$ que indica la recepción del tratamiento.

Sea $Y_i(1)$ la variable respuesta del individuo i si recibe el tratamiento y sea $Y_i(0)$ la variable respuesta si no recibe tratamiento, por lo que el efecto del tratamiento para el individuo i es $Y_i(1) - Y_i(0)$. Debido a que el individuo es tratado o no lo es, solamente es observable uno de los dos posibles resultados posibles o variables respuesta. El resultado observado, Y_i está relacionado con las variables respuesta mediante

$$Y_i = Y_i(1)X_i + Y_i(0)(1 - X_i). \quad (13.17)$$

Si algunos individuos reciben el tratamiento y otros no, la diferencia esperada en los resultados observados entre los dos grupos es $E(Y_i|X_i = 1) - E(Y_i|X_i = 0) = E[Y_i(1)|X_i = 1] - E[Y_i(0)|X_i = 0]$. Esto es cierto sin importar cómo se determina el tratamiento y dice simplemente que la diferencia esperada es la media del resultado del tratamiento para el tratado, menos la media del resultado de la ausencia de tratamiento para el no tratado. Si además los individuos son asignados aleatoriamente a los grupos de tratamiento y control, entonces X_i se distribuye independientemente de todos los atributos personales y, en particular, es independiente de $[Y_i(1), Y_i(0)]$. Con asignación aleatoria, la diferencia de medias entre los grupos de tratamiento y de control es,

$$E(Y_i|X_i = 1) - E(Y_i|X_i = 0) = E[Y_i(1)|X_i = 1] - E[Y_i(0)|X_i = 0] = E[Y_i(1) - Y_i(0)], \quad (3.18)$$

donde la segunda igualdad utiliza el hecho de que $[Y_i(1), Y_i(0)]$ son independientes de X_i por asignación aleatoria y la linealidad de las esperanzas [Ecuación (2.28)]. Por tanto, si X_i se asigna aleatoriamente, la diferencia de medias de los resultados experimentales entre los dos grupos es el efecto promedio del tratamiento en la población de la cual se extrajeron los sujetos.

El marco de análisis de las variables respuesta se puede traducir directamente a la notación de regresión utilizada a lo largo de este libro. Sea $u_i = Y_i(0) - E[Y_i(0)]$ y denominemos $E[Y_i(0)] = \beta_0$. Asimismo se denomina $Y_i(1) - Y_i(0) = \beta_{1i}$, por lo que β_{1i} es el efecto del tratamiento para el individuo i . A partir de la Ecuación (13.17), se tiene

$$\begin{aligned} Y_i &= Y_i(1)X_i + Y_i(0)(1 - X_i) \\ &= Y_i(0) + [Y_i(1) - Y_i(0)]X_i \\ &= E[Y_i(0)] + [Y_i(1) - Y_i(0)]X_i + \{Y_i(0) - E[Y_i(0)]\} \\ &= \beta_0 + \beta_{1i}X_i + u_i. \end{aligned} \quad (13.19)$$

Por lo tanto, a partir de la relación entre los resultados observados y las variables respuesta, y simplemente cambiando la notación, se obtienen los coeficientes aleatorios del modelo de regresión de la Ecuación (13.9). [La Ecuación (13.9) presenta una β_0 que varía entre los diferentes individuos, pero que es equivalente a la Ecuación (13.19) debido a que u_i varía asimismo entre individuos.] Si X_i es asignado aleatoriamente, entonces X_i es independiente de $[Y_i(1), Y_i(0)]$ y por lo tanto es independiente de β_{1i} y u_i . Si el efecto del tratamiento es constante, entonces $\beta_{1i} = \beta_1$, y la Ecuación (13.9) se convierte en la Ecuación (13.1).

Como se analizó en el Apéndice 7.2 y en las Secciones 13.1 y 13.3, en algunos casos X_i se asigna de forma aleatoria en función del valor de una tercera variable, W_i . Si W_i y la variable respuesta no son independientes, entonces, en general, la diferencia de las medias entre los grupos no es igual el efecto medio del tratamiento, es decir, la Ecuación (13.18) no se cumple. Sin embargo, la asignación aleatoria de X_i dado W_i implica que, condicionado a W_i , X_i y $[Y_i(1), Y_i(0)]$ son independientes. Esta condición de que X_i y $[Y_i(1), Y_i(0)]$ sean independientes, condicionado a W_i , se denomina a menudo *inconfundibilidad*¹³.

Si el efecto del tratamiento no varía entre los individuos y si $E[Y_i|X_i, W_i]$ es lineal, entonces la «*inconfundibilidad*» implica independencia en media condicional del error de regresión de la Ecuación (13.2). Para comprobarlo, sea $Y_i(0) = \beta_0 + \gamma W_i + u_i$, donde γ es el efecto causal (si existe) sobre $Y_i(0)$ de W_i , y sea $Y_i(1) - Y_i(0) = \beta_1$ (efecto del tratamiento constante). Entonces la lógica lleva a que la Ecuación (13.19) dé lugar a $Y_i = \beta_0 + \beta_1 X_i + \gamma W_i + u_i$, que es la Ecuación (13.2). Ahora, $E(u_i|X_i, W_i) = E[Y_i(0) - \beta_0 - \gamma W_i|X_i, W_i] = E[Y_i(0) - \beta_0 - \gamma W_i|W_i] = E[u_i|W_i]$. Donde la segunda igualdad se deduce de la inconfundibilidad (si, $[Y_i(1), Y_i(0)]$ es independiente de X_i dado W_i , entonces $E[Y_i(0)|X_i, W_i] = E[Y_i(0)|W_i]$). Por lo tanto la inconfundibilidad implica que $E(u_i|X_i, W_i) = E(u_i|W_i)$ en la Ecuación (13.2). El razonamiento del Apéndice 7.2 implica que, si $E(u_i|W_i)$ es lineal en W_i , entonces el estimador MCO de β_1 en la Ecuación (13.2) es insesgado, aunque en general el estimador MCO de γ es sesgado porque $E(u_i|W_i) \neq 0$.

¹³ N. Del T.: «Unconfoundedness» en el original de Rubin (1990).

14 Introducción a la regresión de series temporales y predicción

Los datos de series temporales —datos registrados para una única entidad individual para varios momentos del tiempo— pueden utilizarse para responder a preguntas cuantitativas para las que los datos de sección cruzada resultan inadecuados. Una de estas preguntas es: ¿cuál es el efecto causal sobre una variable de interés, Y , de una variación en el tiempo de otra variable X ? En otras palabras, ¿cuál es el efecto causal *dinámico* sobre Y de una variación en X ? Por ejemplo, ¿cuál es el efecto sobre las muertes en accidentes de tráfico de una ley que obliga a que los pasajeros utilicen cinturones de seguridad, tanto en un momento inicial como posteriormente cuando los conductores se acomoden a la ley? Otra pregunta de este tipo es: ¿cuál es la mejor predicción sobre el valor de una variable en una fecha futura? Por ejemplo, ¿cuál es la mejor predicción de la tasa de inflación, el tipo de interés o los precios de las acciones del mes próximo? Ambas preguntas —una sobre los efectos causales dinámicos, y la otra sobre predicción económica— pueden ser respondidas utilizando datos de series temporales. Pero los datos de series temporales plantean desafíos particulares, y la superación de estos desafíos requiere la utilización de algunas nuevas técnicas.

En los Capítulos 14 a 16 se introducen algunas técnicas para el análisis econométrico de los datos de series temporales y se trata la aplicación de estas técnicas a los problemas de predicción y estimación de efectos causales dinámicos. El Capítulo 14 introduce los conceptos básicos y las herramientas de regresión con datos de series temporales y su aplicación a las predicciones económicas. En el Capítulo 15, los conceptos y las herramientas desarrolladas en el Capítulo 14 se aplican al problema de la estimación de los efectos causales dinámicos utilizando datos de series temporales. El Capítulo 16 trata algunos temas más avanzados en el análisis de series temporales, incluida la predicción de varias series temporales y la modelización de los cambios en la volatilidad a lo largo del tiempo.

El problema empírico estudiado en este capítulo es la predicción de la tasa de inflación, es decir, el aumento general en términos porcentuales de los precios. Si bien en algún sentido la predicción se trata solamente de una aplicación del análisis de regresión, la predicción es algo muy diferente de la estimación de efectos causales, que ha constituido hasta ahora el centro de atención de este libro. Como se discute en la Sección 14.1, no es necesario que los modelos que son útiles para la predicción tengan una interpretación causal: si se ven peatones llevar paraguas, se puede pronosticar lluvia, a pesar de que llevar paraguas no cause la lluvia. La Sección 14.2 introduce algunos conceptos básicos del análisis de series temporales y presenta algunos ejemplos de datos de series temporales económicas. En la Sección 14.3 se presentan los modelos de regresión de series temporales en los que los regresores son los valores pasados de la variable dependiente; estos modelos «autorregresivos» utilizan la historia de la inflación para predecir su futuro. A menudo, las predicciones basadas en autorregresiones se pueden mejorar mediante la adición de otras variables predictoras y de sus valores pasados, o «retardos», como regresores, y estos modelos, habitualmente denominados modelos autorregresivos de retardos distribuidos, se presentan en la Sección 14.4. Por ejemplo, nos encontramos con que las predicciones de inflación realizadas mediante valores retardados de la tasa de desempleo, además de la inflación retardada, es decir, las predicciones basadas en una curva de Phillips empírica, mejoran las predicciones autorregresivas de la inflación. Un problema práctico consiste en decidir cuántos valores pasados se deben incluir en los modelos autorregresivos de retardos distribuidos, y en la Sección 14.5 se describen los métodos disponibles para tomar de esta decisión.

El supuesto de que el futuro será como el pasado resulta muy importante en la regresión de series temporales, lo suficiente como para que se le dote de su propio nombre, «estacionariedad». Las variables de series temporales pueden no ser estacionarias en varios sentidos, pero dos de ellos son especialmente relevantes para el análisis de regresión de series temporales con datos económicos: (1) las series pueden tener movimientos persistentes a largo plazo, es decir, la serie puede presentar tendencias; y (2) la regresión poblacional puede ser inestable en el tiempo, es decir, la regresión poblacional puede tener puntos de ruptura («breaks»). Estas desviaciones de la estacionariedad ponen en peligro las predicciones e inferencias basadas en la regresión de series temporales. Afortunadamente, existen procedimientos estadísticos para detectar las tendencias y los puntos de ruptura y, una vez detectados, para ajustar la especificación del modelo. Estos procedimientos se presentan en las Secciones 14.6 y 14.7.

14.1 Utilización de los modelos de regresión para predicción

La aplicación empírica de los Capítulos 4 a 9 estaba centrada en la estimación del efecto causal sobre las calificaciones en el examen de la ratio estudiantes-maestros. El modelo más simple de regresión del Capítulo 4 relacionaba las calificaciones en las pruebas con la ratio estudiantes-maestros (REM):

$$\widehat{\text{CalificaciónExamen}} = 989,9 - 2,28 \times REM. \quad (14.1)$$

Como se discutió en el Capítulo 6, una directora escolar, que esté contemplando la contratación de más maestros para reducir el tamaño de las clases, no consideraría que esta ecuación pudiera resultar muy útil. El coeficiente estimado de la pendiente en la Ecuación (14.1) no proporciona una estimación útil del efecto causal sobre las calificaciones obtenidas de la ratio estudiantes-maestros, debido al probable sesgo de variable omitida derivado de la omisión de las características de la escuela y de los estudiantes que son determinantes de los resultados en las pruebas y que están correlacionadas con la ratio estudiantes-maestros.

Por el contrario, tal y como se discutió en el Capítulo 9, un padre que esté considerando mudarse a un distrito escolar puede encontrar más útil la Ecuación (14.1). A pesar de que el coeficiente no tenga una interpretación causal, la regresión puede ayudar a las predicciones de los padres sobre los resultados en las pruebas de un distrito para el que no estén disponibles al público. De manera más general, un modelo de regresión puede resultar útil para la predicción, aunque ninguno de sus coeficientes tenga una interpretación causal. Desde la perspectiva de la predicción, lo que es importante es que el modelo proporcione una predicción lo más precisa posible. Aunque no existe algo así como una predicción perfecta, los modelos de regresión, no obstante, pueden proporcionar predicciones que sean precisas y fiables.

Las aplicaciones de este capítulo difieren de las predicciones del problema calificaciones/tamaño de clase, ya que este capítulo se centra en la utilización de datos de series temporales para la predicción de sucesos futuros. Por ejemplo, en realidad el padre estaría interesado en las calificaciones en el examen del año próximo, después de que su hijo o hija se haya matriculado en una escuela. Por supuesto, estos exámenes no se han llevado a cabo, por lo que el padre debe predecir las calificaciones utilizando la información disponible en la actualidad. Si están disponibles las calificaciones de la prueba para los últimos años, entonces un buen punto de partida es el de utilizar los datos sobre las calificaciones en las pruebas actuales y pasadas para predecir las calificaciones futuras. Este razonamiento nos lleva directamente a los modelos autorregresivos presentados en la Sección 14.3, en los que se utilizan los valores pasados de una variable en una regresión lineal para predecir los valores futuros de las series. El paso siguiente, que se da en la Sección 14.4, consiste en extender estos modelos para incluir variables predictoras adicionales tales como los datos sobre el tamaño de las clases. Como la Ecuación (14.1), un modelo de regresión así puede dar lugar a predicciones precisas y fiables incluso si sus coeficientes no tienen una interpretación causal. En el Capítulo 15, volveremos a problemas como el que afronta la directora de la escuela y se estudiará la estimación de los efectos causales utilizando variables de series temporales.

14.2 Introducción a los datos de series temporales y correlación serial

Esta sección introduce algunos conceptos básicos, así como la terminología que aparece en econometría de series temporales. Un buen modo de comenzar cualquier análisis de datos de series temporales consiste en representar gráficamente los datos, de modo que es donde comenzamos.

Las tasas de inflación y desempleo en Estados Unidos

En la Figura 14.1a está representada gráficamente la tasa de inflación de EE.UU., la variación porcentual anual de los precios en los Estados Unidos, medida por el Índice de Precios al Consumo (IPC), desde 1960 hasta 2004 (los datos se describen en el Apéndice 14.1). La tasa de inflación fue baja en la década de 1960, se elevó durante la década de 1970 hasta un pico después de la II Guerra Mundial del 15,5 % en el primer trimestre de 1980 (es decir, enero, febrero y marzo de 1980), y luego se redujo a menos del 3 % al final de la década de 1990. Como se puede ver en la Figura 14.1a, la tasa de inflación puede fluctuar asimismo en un punto porcentual o más de un trimestre a otro. La tasa de desempleo en los EE.UU., la proporción de la población activa sin trabajo, medida por la Encuesta de Actualizada de Población (véase el Apéndice 3.1), está representada en la Figura 14.1b. Las variaciones de la tasa de desempleo están asociadas principalmente con el ciclo económico de los Estados Unidos. Por ejemplo, la tasa de desempleo aumentó durante las recesiones de 1960-1961, 1970, 1974-1975, las recesiones gemelas de 1980 y 1981-1982, y las recesiones de 1990-1991 y 2001, episodios representados mediante sombreado en la Figura 14.1b.

Retardos, primeras diferencias, logaritmos y tasas de crecimiento

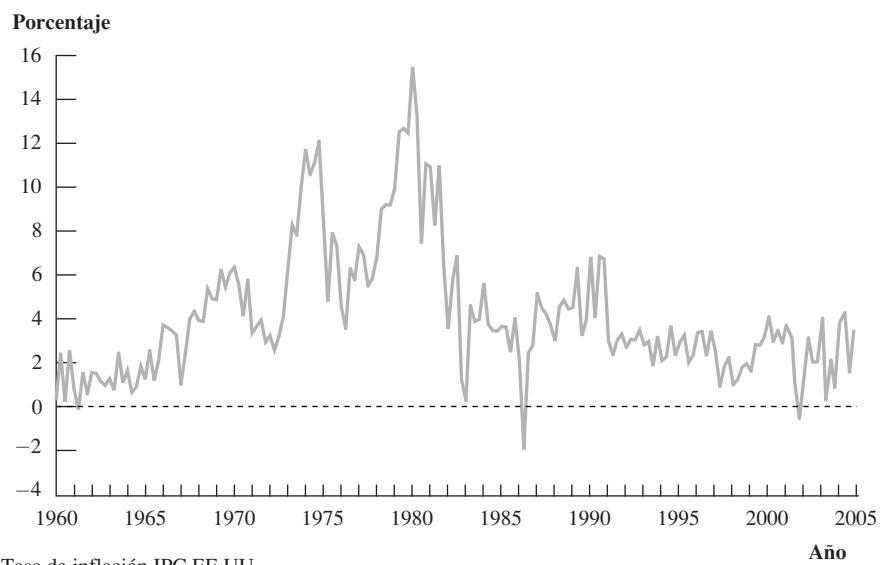
La observación de la variable de serie temporal Y realizada en el momento t se expresa como Y_t , y el número total de observaciones se expresa mediante T . El intervalo entre observaciones, es decir, el periodo de tiempo entre la observación t y la observación $t + 1$, es alguna unidad de tiempo tal como una semana, un mes, un trimestre (unidad de tres meses), o un año. Por ejemplo, los datos de inflación analizados en este capítulo son trimestrales, por lo que la unidad de tiempo (un «periodo») es un trimestre de un año.

Se utiliza una terminología y una notación particulares para indicar los valores futuros y pasados de Y . El valor de Y en el periodo anterior se denomina el *valor de su primer retardo* o, de modo más simple, su **primer retardo**, y se expresa mediante Y_{t-1} . El *valor de su j-ésimo retardo* (o simplemente su **j-ésimo retardo**) es su valor hace j periodos, que es Y_{t-j} . Del mismo modo, Y_{t+j} indica el valor de Y en el periodo siguiente.

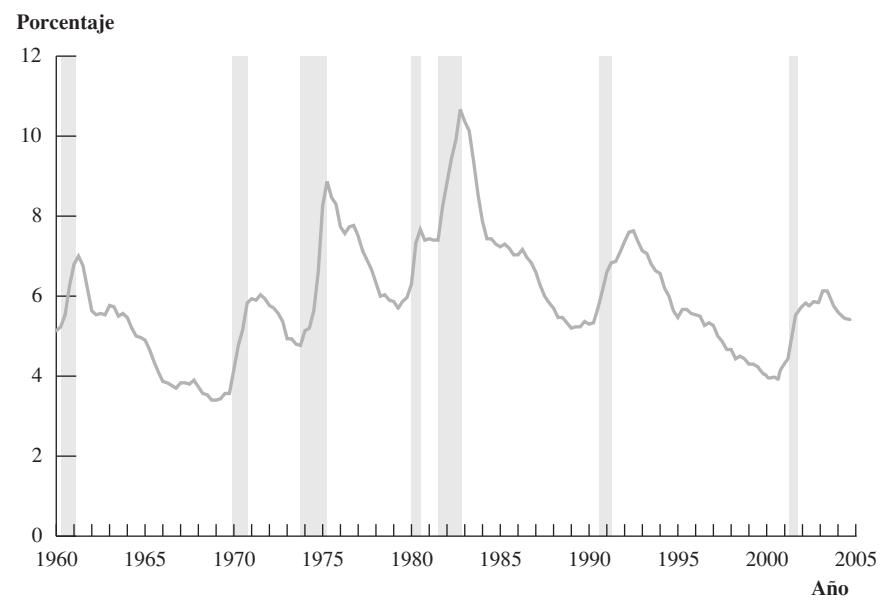
La variación en el valor de Y entre el periodo $t - 1$ y el periodo t es $Y_t - Y_{t-1}$; esta variación se denomina **primera diferencia** de la variable Y_t . En datos de series temporales se utiliza « Δ » para representar la primera diferencia, por lo que $\Delta Y_t = Y_t - Y_{t-1}$.

Las series temporales económicas a menudo son analizadas después de calcular sus logaritmos o las variaciones en sus logaritmos. Una razón para ello es que muchas series económicas, tales como el producto interior bruto (PIB), presentan un crecimiento que es aproximadamente exponencial, es decir, a largo plazo la serie tiende a crecer a un determinado porcentaje medio anual; si es así, el logaritmo de la serie crece de forma aproximadamente lineal. Otra razón es que la desviación típica de muchas series temporales económicas es aproximadamente proporcional a su nivel, es decir, la desviación típica se puede expresar correctamente un porcentaje del nivel de las series; si es así, entonces la desviación típica del logaritmo de la serie es aproximadamente constante. En cualquier caso, resulta útil transformar las series para que las variaciones en las series transformadas sean variaciones proporcionales (o porcentuales) de la serie original, y esto se logra tomando el logaritmo de las series¹.

¹ La variación en el logaritmo de una variable es aproximadamente igual a la variación proporcional de esa variable; es decir, $\ln(X + a) - \ln(X) \cong a/X$, esta aproximación funciona mejor cuando a/X es pequeño [véase la Ecuación (8.16) y el análisis que la acompaña]. Ahora, reemplazando X por Y_{t-1} , y a por ΔY_t , y teniendo en cuenta que $Y_t = Y_{t-1} + \Delta Y_t$. Esto significa que la variación pro-

FIGURA 14.1 Inflación y desempleo en Estados Unidos, 1960-2004

(a) Tasa de inflación IPC EE.UU.



(b) Tasa de desempleo EE.UU.

La inflación de los precios en Estados Unidos (Figura 14.1a) evolucionó al alza desde 1960 hasta 1980 y posteriormente disminuyó bruscamente durante los primeros años de la década de 1980. La tasa de desempleo en los Estados Unidos (Figura 14.1b) crece durante las recesiones (los períodos sombreados) y disminuye en las expansiones.

Los retardos, las primeras diferencias y las tasas de crecimiento aparecen resumidos en el Concepto clave 14.1.

Los retardos, las primeras diferencias, y las tasas de crecimiento se ilustran en la Tabla 14.1 mediante la tasa de inflación de EE.UU. La primera columna muestra el instante, o periodo, de forma que el primer

porcional en la serie Y_t entre los períodos $t-1$ y t es de aproximadamente $\ln(Y_t) - \ln(Y_{t-1}) = \ln(Y_{t-1} + \Delta Y_t) - \ln(Y_{t-1}) \cong Y_t/Y_{t-1}$. La expresión $\ln(Y_t) - \ln(Y_{t-1})$ es la primera diferencia de $\ln(Y_t)$, $\Delta \ln(Y_t)$. Por tanto, $\Delta \ln(Y_t) \cong Y_t/Y_{t-1}$. La variación porcentual es 100 veces el cambio fraccional, por lo que el porcentaje de cambio en la serie Y_t es aproximadamente $100 \Delta \ln(Y_t)$.

**CONCEPTO
CLAVE**
14.1
Retardos, primeras diferencias, logaritmos y tasas de crecimiento

- El primer retardo de una serie temporal Y_t es Y_{t-1} ; su j -ésimo retardo es Y_{t-j} .
- La primera diferencia de una serie, ΔY_t , es su variación entre los períodos $t - 1$ y t ; es decir, $\Delta Y_t = Y_t - Y_{t-1}$.
- La primera diferencia del logaritmo de Y_t es $\Delta \ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1})$.
- La variación porcentual de una serie temporal Y_t entre los períodos $t - 1$ y t es aproximadamente $100\Delta \ln(Y_t)$, siendo la aproximación más precisa cuando la variación porcentual es pequeña.

trimestre de 2004 está expresado como 2004:I, el segundo trimestre de 2004 se expresa como 2004:II, etc. La segunda columna muestra el valor del IPC en ese trimestre, y la tercera columna muestra la tasa de inflación. Por ejemplo, del primer al segundo trimestre de 2004, el índice aumentó desde 186,57 hasta 188,60, un incremento porcentual de $100 \times (188,60 - 186,57)/186,57 = 1,09\%$. Este es el aumento porcentual de un trimestre a otro. Es habitual presentar las tasas de inflación (y otras tasas de crecimiento de series temporales macroeconómicas) en una base anualizada, es decir el aumento porcentual de los precios que se produciría en un año, si la serie aumentara a la misma tasa. Debido a que en un año hay cuatro trimestres, la tasa de inflación anualizada en 2004:II es $1,09 \times 4 = 4,36$, o el 4,4% al año tras el redondeo.

Esta variación porcentual puede calcularse asimismo utilizando la aproximación de las diferencias de los logaritmos del Concepto clave 14.1. La diferencia en el logaritmo del IPC desde 2004:I hasta 2004:II es $\ln(188,60) - \ln(186,57) = 0,0108$, lo que da una diferencia porcentual trimestral aproximada de $100 \times 0,0108 = 1,08\%$. En base anualizada, es igual a $1,08 \times 4 = 4,32$, o el 4,3% después del redondeo, esencialmente la misma que la obtenida calculando directamente el porcentaje de crecimiento. Estos cálculos pueden resumirse como

$$\begin{aligned} \text{Tasa de inflación anualizada} &= \text{Inf}_t \cong 400[\ln(\text{IPC}_t) - \ln(\text{IPC}_{t-1})] \\ &= 400\Delta \ln(\text{IPC}_t), \end{aligned} \quad (14.2)$$

donde IPC_t es el valor del Índice de Precios al Consumo en el período t . El factor 400 surge de convertir la variación fraccional a términos porcentuales (multiplicando por 100) y convirtiendo la variación porcentual trimestral a una tasa anual equivalente (multiplicando por 4).

TABLA 14.1 La inflación en los Estados Unidos en el 2004 y el primer trimestre de 2005

Trimestre	IPC EE.UU.	Tasa de Inflación a una tasa anualizada (Inf_t)	Primer retardo (Inf_{t-1})	Variación de la inflación (ΔInf_t)
2004:I	186,57	3,8	0,9	2,9
2004:II	188,60	4,4	3,8	0,6
2004:III	189,37	1,6	4,4	-2,8
2004:IV	191,03	3,5	1,6	1,9
2005:I	192,17	2,4	3,5	-1,1

La tasa de inflación anualizada es la variación porcentual en el IPC en el trimestre actual respecto al trimestre anterior, multiplicada por cuatro. El primer retardo de la inflación es su valor en el trimestre anterior, y la variación de la inflación es la tasa de inflación actual menos su primer retardo. Todas las cifras se han redondeado al decimal más cercano.

Las dos últimas columnas de la Tabla 14.1 muestran los retardos y las variaciones. El primer retardo de la inflación en 2004:II es de 3,8 %, la tasa de inflación en 2004:I. La variación en la tasa de inflación desde 2004:I hasta 2004:II fue del 4,4 % – 3,8 % = 0,6 %.

Autocorrelación

En datos de series temporales, el valor de Y en un periodo por lo general está correlacionado con su valor en el periodo siguiente. La correlación de una serie con sus propios valores retardados se denomina **autocorrelación** o **correlación serial**. La primera autocorrelación (o **coeficiente de autocorrelación**) es la correlación entre Y_t e Y_{t-j} , es decir, la correlación entre los valores de Y en dos períodos adyacentes. La autocorrelación segunda es la correlación entre Y_t e Y_{t-2} , y la autocorrelación j -ésima es la correlación entre Y_t e Y_{t-j} . Del mismo modo, la **autocovarianza j -ésima** es la covarianza entre Y_t e Y_{t-j} . La autocorrelación y la autocovarianza se resumen en el Concepto clave 14.2.

Las autocovarianzas y autocorrelaciones j -ésimas poblacionales del Concepto clave 14.2 pueden ser estimadas mediante las autocovarianzas y autocorrelaciones j -ésimas muestrales, $\text{cov}(Y_t, Y_{t-j})$ y $\hat{\rho}_j$:

$$\widehat{\text{cov}}(Y_t, Y_{t-j}) = \frac{1}{T} \sum_{t=j+1}^T (Y_t - \bar{Y}_{j+1, T})(Y_{t-j} - \bar{Y}_{1, T-j}) \quad (14.5)$$

$$\hat{\rho}_j = \frac{\widehat{\text{cov}}(Y_t, Y_{t-j})}{\widehat{\text{var}}(Y_t)}, \quad (14.6)$$

donde $\bar{Y}_{j+1, T}$ expresa la media muestral de Y_t calculada para las observaciones $t = j + 1, \dots, T$ y donde $\widehat{\text{var}}(Y_t)$ es la varianza de muestral de Y .²

Las cuatro primeras autocorrelaciones muestrales de la tasa de inflación y de la variación de la tasa de inflación se presentan en la Tabla 14.2. Estas cifras muestran que la inflación está fuertemente autocorrelacionada de forma positiva: la primera autocorrelación es 0,84. La autocorrelación muestral disminuye a medida que aumenta el retardo, pero sigue siendo elevada, incluso con un retardo de cuatro trimestres. La variación de la inflación está negativamente autocorrelacionada: un aumento en la tasa de inflación en un trimestre suele estar asociado con una disminución en el trimestre siguiente.

Al principio, puede resultar contradictorio que el nivel de inflación esté fuertemente correlacionado de forma positiva, pero que su variación esté correlacionada negativamente. Sin embargo, estas dos autocorre-

**CONCEPTO
CLAVE**

14.2

Autocorrelación (correlación serial) y autocovarianza

La j -ésima autocovarianza de una serie Y_t es la covarianza entre Y_t y su j -ésimo retardo Y_{t-j} , y el coeficiente de correlación j -ésimo es la correlación entre Y_t e Y_{t-j} . Es decir,

$$j\text{-ésima autocovarianza} = \text{cov}(Y_t, Y_{t-j}) \quad (14.3)$$

$$j\text{-ésima autocorrelación} = \rho_j = \text{corr}(Y_t, Y_{t-j}) = \frac{\text{cov}(Y_t, Y_{t-j})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-j})}} \quad (14.4)$$

El coeficiente de autocorrelación j -ésimo a veces se denomina coeficiente de correlación serial j -ésimo.

² El sumatorio de la Ecuación (14.5) está dividido por T , mientras que en la fórmula habitual de la covarianza muestral [véase la Ecuación (3.24)] el sumatorio está dividido por el número de observaciones del sumatorio, menos un ajuste por los grados de libertad. La fórmula de la Ecuación (14.5) es la habitual para el cálculo de la autocovarianza. En la Ecuación (14.6) se utiliza el supuesto de que $\text{var}(Y_t)$ y $\text{var}(Y_{t-1})$ son iguales, una consecuencia del supuesto es que Y es estacionaria, lo cual se analiza en la Sección 14.4.

TABLA 14.2 Primeras cuatro autocorrelaciones muestrales de la tasa de inflación en EE.UU. y su variación, 1960:I-2004:IV

Retardo	Autocorrelación de:	
	Tasa de inflación (Inf_t)	Variación de la tasa de inflación (ΔInf_t)
1	0,84	-0,26
2	0,76	-0,25
3	0,76	-0,29
4	0,67	-0,06

laciones, miden cosas diferentes. La fuerte autocorrelación positiva de la inflación refleja la evidente tendencia de largo plazo de la inflación de la Figura 14.1: la inflación fue baja en el primer trimestre de 1965 y de nuevo en el segundo; fue elevada en el primer trimestre de 1981 y de nuevo en el segundo. Por el contrario, la autocorrelación negativa de la variación en la inflación significa que, en promedio, un aumento de la inflación en un trimestre está asociado con una disminución de la inflación en el siguiente.

Otros ejemplos de series temporales económicas

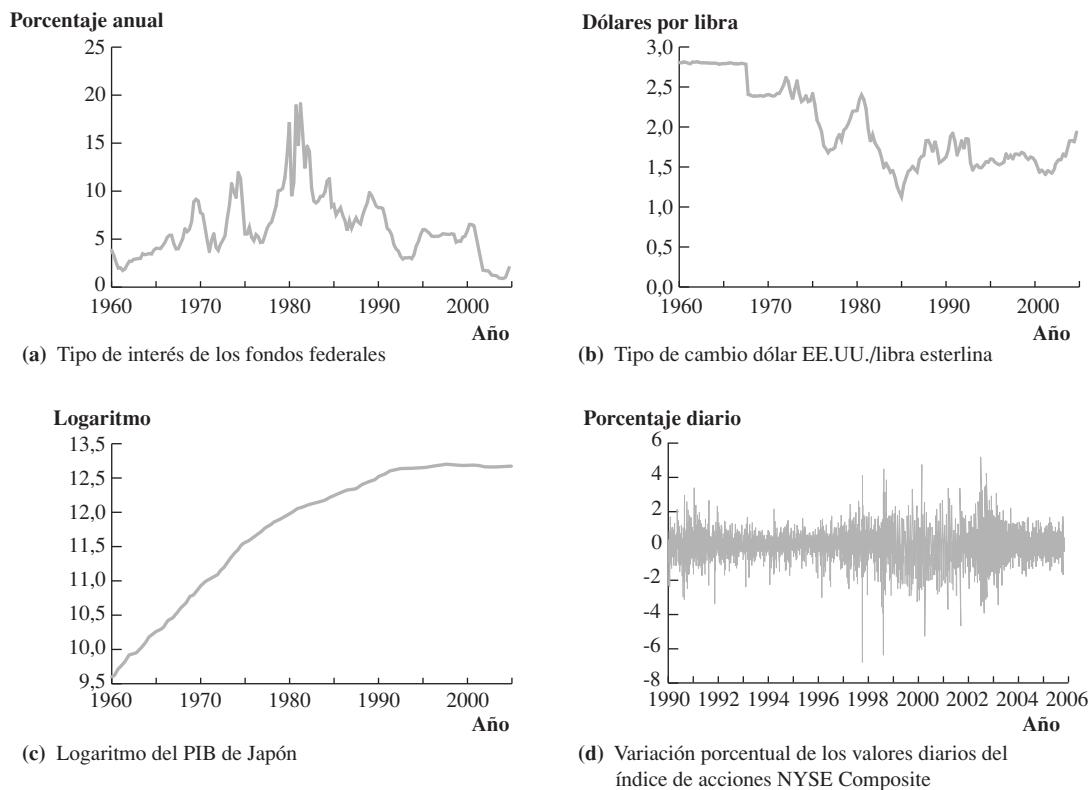
Las series temporales económicas son muy diferentes. Se representan cuatro ejemplos de series temporales económicas en la Figura 14.2: el tipo de interés de los fondos federales en EE.UU.; el tipo de cambio entre el dólar y la libra esterlina; el logaritmo del producto interior bruto de Japón; y la rentabilidad diaria en el índice del mercado de acciones Standard and Poor's 500 (S&P 500).

El tipo de interés de los fondos federales en EE.UU. (Figura 14.2a) es el tipo de interés que pagan unos bancos a otros por tomar prestado y prestarse entre sí, en financiación a un día. Este tipo de interés es importante, ya que está controlado por la Reserva Federal y es el principal instrumento de la política monetaria para la Fed. Si se comparan los gráficos del tipo de interés de los fondos federales y las tasas de desempleo e inflación de la Figura 14.1, se comprueba que un fuerte aumento en el tipo de interés de los fondos federales a menudo ha estado asociado con recesiones posteriores.

El tipo de cambio dólar/libra (Figura 14.2b) es el precio de una libra esterlina (£) en dólares de EE.UU. Antes de 1972, las economías desarrolladas seguían un sistema de tipos de cambio fijos —denominado sistema de «Bretton Woods»— bajo el cual las autoridades nacionales trabajaban para mantener los tipos de cambio evitando fluctuaciones. En 1972, las presiones inflacionistas llevaron a la ruptura de este sistema; a partir de entonces, a las principales monedas se les permitió «flotar»; es decir, sus valores estaban determinados por la oferta y la demanda de divisas en el mercado de tipos de cambio. Antes de 1972, el tipo de cambio era aproximadamente constante, con la excepción de una sola devaluación en 1968 por la que el valor oficial de la libra, con respecto al dólar, se redujo hasta 2,40 \$. Desde 1972 el tipo de cambio ha fluctuado dentro un rango muy amplio.

El PIB trimestral japonés (Figura 14.2c) es el valor total de los bienes y servicios producidos en Japón durante un trimestre. El PIB es la medida más amplia del total de la actividad económica. El logaritmo de la serie se representa en la Figura 14.2c, y las variaciones de esta serie pueden interpretarse como las tasas de crecimiento (fraccionadas). Durante la década de 1960 y los primeros años de la década de 1970, el PIB japonés creció rápidamente, pero este crecimiento se desaceleró a finales de la década de 1970 y durante la década de 1980. El crecimiento se desaceleró aún más durante la década de 1990, promediando solamente un 1,2 % anual entre 1990 y 2004.

El índice del mercado NYSE Composite es un índice general del precio de las acciones de todas las empresas que cotizan en la Bolsa de Nueva York. La Figura 14.2d representa gráficamente el porcentaje de variación diaria de este índice para los días hábiles entre el 2 de enero de 1990, y el 11 de noviembre de 2005 (un total de 4.003 observaciones). A diferencia de las otras series de la Figura 14.2, existe muy poca

FIGURA 14.2 Cuatro series temporales económicas

Las cuatro series temporales presentan patrones muy diferentes. El tipo de interés de los fondos federales (Figura 14.2a) presenta un patrón similar a la inflación de los precios. El tipo de cambio entre el dólar de EE.UU. y la libra esterlina (Figura 14.2b) muestra una variación discreta tras el colapso, en 1972, del sistema de Bretton Woods de tipos de cambio fijos. El logaritmo del PIB de Japón (Figura 14.2c) muestra un crecimiento relativamente suave, aunque la tasa de crecimiento disminuye en la década de 1970 y de nuevo en la década de 1990. La variación porcentual diaria en el índice de precios de la bolsa de valores NYSE (Figura 14.2d) es esencialmente impredecible, pero su varianza cambia: esta serie muestra «volatilidad agrupada».

correlación serial entre las variaciones porcentuales diarias: si existiera, entonces se podría predecir utilizando las variaciones diarias anteriores y se podría ganar dinero comprando cuando se espera que el mercado suba y vendiendo cuando se espera que baje. Aunque las variaciones son esencialmente impredecibles, la observación detenida de la Figura 14.2d revela patrones en su volatilidad. Por ejemplo, la desviación típica de las variaciones porcentuales diarias fue relativamente grande en 1990-1991 y 1998-2003, y relativamente pequeña en 1995 y 2005. Esta «volatilidad agrupada» se encuentra en muchas series temporales financieras, en la Sección 16.5 se tratan algunos de los modelos econométricos que existen para modelizar este tipo especial de heterocedasticidad.

14.3 Modelos autorregresivos

¿Cuál será la tasa de inflación de los precios, el aumento general de los precios en términos porcentuales, el próximo año? Los inversores de Wall Street se basan en las previsiones de inflación en el momento de decidir la cantidad a pagar por los bonos. Los economistas de los bancos centrales, como el Banco de la Reserva Federal de los EE.UU. utilizan las predicciones de inflación cuando establecen la política monetaria. Las empresas utilizan las predicciones de inflación cuando realizan la previsión de ventas de sus productos, y los gobiernos locales utilizan las predicciones de inflación cuando redactan sus presupuestos para

el año siguiente. En esta sección, se consideran las predicciones realizadas mediante **autorregresión**, un modelo de regresión que relaciona una variable de series temporales con sus valores pasados.

El modelo autorregresivo de primer orden

Si se pretende predecir el futuro de una serie temporal, un buen punto de arranque es el pasado inmediato. Por ejemplo, si se quiere pronosticar la variación de la inflación de este trimestre al siguiente, podría verse si la inflación aumentó o disminuyó el último trimestre. Un método sistemático para predecir la variación en la inflación, ΔInf_t , mediante la variación del trimestre anterior, ΔInf_{t-1} consiste en estimar una regresión MCO de ΔInf_t sobre ΔInf_{t-1} . Esta regresión estimada utilizando los datos desde 1962 hasta 2004, es

$$\widehat{\Delta Inf_t} = 0,017 - 0,238\Delta Inf_{t-1}, \quad (14.7)$$

(0,126) (0,096)

donde, como es habitual, los errores estándar están indicados entre paréntesis debajo de los coeficientes estimados, y $\widehat{\Delta Inf_t}$ es el valor estimado de ΔInf_t en base a la recta de regresión estimada. El modelo de la Ecuación (14.7) se denomina autorregresivo de primer orden: es autorregresivo porque se trata de una regresión de la serie sobre su propio retardo, ΔInf_{t-1} y es de primer orden, ya que solo se utiliza un retardo como regresor. El coeficiente de la Ecuación (14.7) es negativo, por lo que un aumento de la tasa de inflación durante en un trimestre está asociado con una disminución de la tasa de inflación en el trimestre siguiente.

Un modelo autorregresivo de primer orden se abrevia mediante AR(1), donde el «1» indica que es de primer orden. El modelo AR(1) poblacional de la serie Y_t es

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t, \quad (14.8)$$

donde u_t es un término de error.

Predicciones y errores de predicción. Supongamos que se dispone de datos históricos sobre Y y se desea predecir su valor futuro. Si Y_t sigue el modelo AR(1) de la Ecuación (14.8) y si β_0 y β_1 son conocidos, entonces la predicción de Y_{T+1} en base a Y_T es $\beta_1 + \beta_1 Y_T$.

En la práctica, β_0 y β_1 son desconocidos, por lo que las predicciones deben basarse en estimaciones de β_0 y β_1 . Se utilizan los estimadores MCO $\hat{\beta}_0$ y $\hat{\beta}_1$, que se calculan utilizando datos históricos. En general, $\hat{Y}_{T+1|T}$ indica la predicción de Y_{T+1} basada en la información hasta el periodo T utilizando un modelo estimado con datos hasta el periodo T . En consecuencia, la predicción basada en el modelo AR(1) de la Ecuación (14.8) es

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T, \quad (14.9)$$

donde $\hat{\beta}_0$ y $\hat{\beta}_1$ se calculan con los datos históricos disponibles hasta el periodo T .

El **error de predicción** es el error cometido en la predicción; es decir la diferencia entre el valor de Y_{T+1} que realmente sucede y su valor de predicción basado en Y_T :

$$\text{Error de predicción} = Y_{T+1} - \hat{Y}_{T+1|T}. \quad (14.10)$$

Predicciones versus valores de predicción. La predicción *no* es un valor de predicción MCO, y el error de predicción *no* es un residuo MCO. Los valores de predicción MCO se calculan para las observaciones de la muestra utilizada para estimar la regresión. Por el contrario, la predicción se hace para algún periodo más allá del conjunto de los datos utilizados para estimar la regresión, por lo que los datos sobre el valor real de la variable dependiente predicha no se encuentran en la muestra utilizada para estimar la regresión. Del mismo modo, el residuo MCO es la diferencia entre el valor real de Y y su valor de predicción para las observaciones de la muestra, mientras que el error de predicción es la diferencia entre el valor futuro de Y , que no está contenido en la muestra de la estimación, y la predicción de ese valor futuro. Dicho de otra manera, las predicciones y los errores de predicción se refieren a observaciones «fuera de la muestra», mientras que los valores de predicción y los residuos se refieren a observaciones «en la muestra».

Raíz del error cuadrático medio de predicción. La **raíz del error cuadrático de predicción (RECP)** es una medida de la magnitud del error de predicción, es decir, de la magnitud de un error estándar cometido con un modelo de predicción. La RECP es la raíz cuadrada de la media del error de predicción al cuadrado:

$$\text{RECP} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}. \quad (14.11)$$

La RECP tiene dos fuentes de error: el error que surge debido a que los valores futuros de u_t son desconocidos y el error cometido en la estimación de los coeficientes β_0 y β_1 . Si la primera fuente de error es mucho mayor que la segunda, tal y como puede ocurrir si el tamaño de la muestra es grande, entonces la RECP es aproximadamente $\sqrt{\text{var}(u_t)}$, la desviación típica del error u_t de la autorregresión poblacional [Ecuación (14.8)]. La desviación típica de u_t se estima a su vez mediante el error estándar de la regresión (ESR; véase la Sección 4.3). Por lo tanto, si la incertidumbre que surge de la estimación de los coeficientes de regresión es lo suficientemente pequeña como para ser ignorada, la RECP se puede estimar mediante el error estándar de la regresión. La estimación de la RECP que incluye ambas fuentes del error de predicción se recoge en la Sección 14.4.

Aplicación a la inflación. ¿Cuál es la predicción para la inflación en el primer trimestre de 2005 (2005:I) que un economista habría hecho en 2004:IV, en base a la estimación del modelo AR(1) de la Ecuación (14.7) (estimado utilizando los datos disponibles hasta 2004:IV)? A partir de la Tabla 14.1, la tasa de inflación en 2004:IV fue el 3,5 % (por lo que $\text{Inf}_{2004:\text{IV}} = 3,5$), un incremento de 1,9 puntos porcentuales a partir de 2004:III (por lo que $\Delta\text{Inf}_{2004:\text{IV}} = 1,9$). Al introducir estos valores en la Ecuación (14.7), la predicción para la variación de la inflación de 2004:IV hasta 2005:I es $\widehat{\Delta\text{Inf}}_{2005:\text{I}|2004:\text{IV}} = 0,017 - 0,238 \times \Delta\text{Inf}_{2004:\text{IV}} = 0,017 - 0,238 \times 1,9 = -0,43 \cong -0,4$ (redondeado al primer decimal). La tasa de inflación prevista es la tasa de inflación pasada más su variación prevista:

$$\widehat{\Delta\text{Inf}}_{T+1|T} = \text{Inf}_T + \widehat{\Delta\text{Inf}}_{T+1|T}. \quad (14.12)$$

Debido a que $\text{Inf}_{2004:\text{IV}} = 3,5$ % y la variación prevista en la tasa de inflación desde 2004:IV hasta 2005:I es $-0,4$, la predicción para la tasa de inflación en 2005:I es $\widehat{\text{Inf}}_{2005:\text{I}|2004:\text{IV}} = \text{Inf}_{2004:\text{IV}} + \widehat{\Delta\text{Inf}}_{2005:\text{I}|2004:\text{IV}} = 3,5\% - 0,4\% = 3,1\%$. Por lo tanto, el modelo AR(1) predice que la inflación bajará ligeramente del 3,5 % en 2004:IV al 3,1 % en 2005:I.

¿Cuál fue la precisión de la predicción AR(1)? De acuerdo con la Tabla 14.1, el valor real de la inflación en 2005:I fue de un 2,4 %, por lo que la predicción de AR(1) fue mayor en 0,7 puntos porcentuales; es decir, el error de predicción es de $-0,7$. El \bar{R}^2 del modelo AR(1) de la Ecuación (14.7) es solamente 0,05, por lo que la variación retardada de la inflación explica una proporción muy pequeña de la variación de la inflación en la muestra utilizada para el ajuste de la autorregresión. Este \bar{R}^2 tan bajo es consistente con el pobre pronóstico de la inflación en 2005:I al que da lugar la utilización de la Ecuación (14.7). De manera más general, el bajo \bar{R}^2 sugiere que este modelo AR(1) pronosticará tan solo una pequeña parte de la variabilidad de la variación de la inflación.

El error estándar de la regresión de la Ecuación (14.7) es 1,65; ignorando la incertidumbre derivada de la estimación de los coeficientes, la estimación de la RECP para las predicciones basadas en la Ecuación (14.7) por lo tanto es de 1,65 puntos porcentuales.

El modelo autorregresivo de orden p

El modelo AR(1) utiliza Y_{t-1} para predecir Y_t , pero al hacerlo ignora información potencialmente útil acerca del pasado más lejano. Una manera de incorporar esta información consiste en incluir retardos adicionales en el modelo AR(1); lo que da lugar al modelo autorregresivo de orden p , o modelo AR(p).

El **modelo autorregresivo de orden p** [el modelo **AR(p)**] representa Y_t como una función lineal de sus primeros p valores retardados, es decir, en el modelo AR(p), los regresores son $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$, más un término independiente. El número de retardos, p , incluidos en un modelo AR(p) se denomina orden, o longitud de los retardos, de la autorregresión.

**CONCEPTO
CLAVE**

14.3

Modelos autorregresivos

El modelo autorregresivo de orden p [el modelo AR(p)] representa Y_t como una función lineal de sus p primeros valores retardados:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t, \quad (14.14)$$

donde $E(u_t | Y_{t-1}, Y_{t-2}, \dots) = 0$. El número de retardos p se denomina orden, o longitud de los retardos, de la autorregresión.

Por ejemplo, un modelo AR(4) para la variación de la inflación utiliza como regresores a los cuatro primeros retardos de la variación de la inflación. Estimado por MCO para el periodo 1962-2004, el modelo AR(4) es

$$\widehat{\Delta Inf_t} = 0,02 - 0,26\Delta Inf_{t-1} - 0,32\Delta Inf_{t-2} + 0,16\Delta Inf_{t-3} - 0,03\Delta Inf_{t-4}. \quad (14.13)$$

(0,12)	(0,09)	(0,08)	(0,08)	(0,09)
--------	--------	--------	--------	--------

Los coeficientes de los tres últimos retardos adicionales de la Ecuación (14.13) son conjuntamente significativamente distintos de cero al nivel de significación del 5 %: el estadístico F es 6,91 (p -valor < 0,001). Esto se ve reflejado en una mejora en el \bar{R}^2 desde el 0,05 para el modelo AR(1) de la Ecuación (14.7) hasta el 0,18 para el modelo AR(4). Del mismo modo, el ESR del modelo AR(4) de la Ecuación (14.13) es de 1,52, lo que supone una mejora sobre el ESR del modelo AR(1), que es 1,65.

El modelo AR(p) se resume en el Concepto clave 14.3.

Propiedades de la predicción y del término de error en el modelo AR(p). El supuesto de que la esperanza condicional de u_t es igual a cero, dados los valores pasados de Y_t [es decir, $E(u_t | Y_{t-1}, Y_{t-2}, \dots) = 0$], tiene dos implicaciones importantes.

La primera implicación es que la mejor predicción para Y_{T+1} basada en su historia completa solamente depende de los p valores pasados más recientes. En concreto, sea $Y_{T+1|T} = E(Y_{T+1} | Y_T, Y_{T-1}, \dots)$ la expresión de la media condicional de Y_{T+1} dada su historia completa. Por tanto, $Y_{T+1|T}$, tiene la menor RECMPE de cualquier predicción basada en la historia de Y (Ejercicio 14.5). Si Y_t sigue un proceso AR(p), entonces la mejor predicción de Y_{T+1} basada en Y_T, Y_{T-1}, \dots es

$$Y_{T+1|T} = \beta_0 + \beta_1 Y_T + \beta_2 Y_{T-1} + \cdots + \beta_p Y_{T-p+1}, \quad (14.15)$$

lo cual se desprende del modelo AR(p) de la Ecuación (14.14) y del supuesto de que $E(u_t | Y_{t-1}, Y_{t-2}, \dots) = 0$. En la práctica, los coeficientes $\beta_0, \beta_1, \dots, \beta_p$ son desconocidos, por lo que en realidad las predicciones realizadas mediante un modelo AR(p) utilizan la Ecuación (14.15) con los coeficientes estimados.

La segunda implicación es que los errores u_t están serialmente incorrelacionados, un resultado que se deriva de la Ecuación (2.27) (Ejercicio 14.5).

Aplicación a la inflación. ¿Cuál es la predicción para la inflación en 2005:I a partir de los datos disponibles hasta 2004:IV, en base al modelo AR(4) para la inflación de la Ecuación (14.13)? Para calcular esta predicción, se sustituyen los valores de la variación en la inflación de cada uno de los cuatro trimestres de 2004 en la Ecuación (14.13): $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0,02 - 0,26\Delta Inf_{2004:IV} - 0,32\Delta Inf_{2004:III} + 0,16\Delta Inf_{2004:II} - 0,03\Delta Inf_{2004:I} = 0,02 - 0,26 \times 1,9 - 0,32 \times (-2,8) + 0,16 \times 0,6 - 0,03 \times 2,9 \cong 0,4$, donde los valores de 2004 para la variación de la inflación se han tomado de la última columna de la Tabla 14.1.

La previsión correspondiente para la inflación en 2005:I es el valor de la inflación en 2004:IV, más la variación estimada; es decir, $3,5\% + 0,4\% = 3,9\%$. El error de predicción es el valor real, 2,4 %, menos la predicción, $2,4\% - 3,9\% = 1,5$, mayor en valor absoluto que el error de predicción AR(1) de $-0,7$ puntos porcentuales.

¿Se puede batir al mercado? Parte I

¿Alguna vez ha soñado con ser rico rápidamente ganando dinero al mercado de valores? Si se cree que el mercado va a subir, deberían comprarse acciones hoy para ser vendidas más tarde, antes de que el mercado vuelva a caer. Si se es bueno al prever las variaciones de los precios de las acciones, entonces esta estrategia de mercado activa dará lugar a mejores rendimientos que una estrategia pasiva de «comprar y mantener» que consista en comprar acciones y simplemente mantenerlas. El truco, por supuesto, es disponer de una previsión fiable sobre los rendimientos futuros de las acciones.

Las predicciones basadas en los valores pasados de la rentabilidad de las acciones a veces se denominan predicciones del «momentum»: si el valor de una acción ha aumentado este mes, tal vez se haya generado un movimiento de impulso y aumente otra vez el mes siguiente. Si es así, entonces los rendimientos estarán correlacionados y el modelo autorregresivo proporcionará predicciones útiles. Se puede llevar a cabo una estrategia basada en el momentum o impulso de una acción específica o de un índice bursátil que mida el valor global del mercado.

En la Tabla 14.3 se presentan algunos modelos autorregresivos para la rentabilidad extraordinaria de un índice amplio de los precios de acciones denominado índice CRSP ponderado, utilizando datos mensuales desde 1960:1 hasta 2002:12. La rentabilidad extraordinaria mensual es la ganancia, en términos porcentuales, obtenida por la compra de una acción al final del mes anterior y su venta a finales de este mes, menos la ganancia que se hubiera obtenido de haber comprado un activo seguro (letras del tesoro norteamericano). La rentabilidad de las acciones incluye la ganancia (o pérdida) de capital debida a la variación en el precio, además de los dividendos percibidos a lo largo del mes. Los datos se describen de manera más detallada en el Apéndice 14.1.

Lamentablemente, los resultados de la Tabla 14.3 son negativos. Los coeficientes de los rendimientos retardados en el modelo AR(1) no son estadísticamente significativos, y no se puede rechazar la hipótesis nula de que los coeficientes de los rendimientos retardados son iguales a cero en los modelos AR(2) o AR(4). De hecho, el R^2 ajustado de uno de los mode-

TABLA 14.3 Modelos autorregresivos para la rentabilidad extraordinaria mensual de las acciones, 1960:1-2002:12

Variable dependiente: rentabilidad extraordinaria del índice CRSP

	(1)	(2)	(3)
Especificación	AR(1)	AR(2)	AR(4)
Variables explicativas			
Rentabilidad extraordinaria _{t-1}	0,050 (0,051)	0,053 (0,051)	0,054 (0,051)
Rentabilidad extraordinaria _{t-2}		-0,053 (0,048)	-0,054 (0,048)
Rentabilidad extraordinaria _{t-3}			0,009 (0,050)
Rentabilidad extraordinaria _{t-4}			-0,016 (0,047)
Término independiente	0,312 (0,197)	0,328 (0,199)	0,331 (0,202)
Estadístico <i>F</i> para los coeficientes de los retardos de la rentabilidad extraordinaria (<i>p</i> -valor)	0,968 (0,325)	1,342 (0,261)	0,707 (0,587)
\bar{R}^2	0,0006	0,0014	-0,0022

Notas: La rentabilidad extraordinaria se mide en términos porcentuales para un mes. Los datos se describen en el Apéndice 14.1. Todas las regresiones están estimadas para el período 1960:1-2002:12 ($T = 516$ observaciones), con valores iniciales de las variables retardadas tomados de observaciones anteriores. Las entradas en las filas de las variables explicativas son los coeficientes de esos regresores, los errores estándar aparecen entre paréntesis. Las dos últimas filas presentan los valores de los estadísticos *F* para contrastar la hipótesis de que los coeficientes de los retardos de la rentabilidad extraordinaria en la regresión, son iguales a cero, con su *p*-valor entre paréntesis, y el R^2 ajustado.

(continúa)

los es negativo y los de los otros dos son solo ligeramente positivos, lo que sugiere que ninguno de estos modelos es útil para la predicción.

Estos resultados negativos son consistentes con la teoría de que los mercados de capitales son eficientes, que sostiene que la rentabilidad extraordinaria debe ser impredecible, debido a que el precio de los valores ya refleja toda la información disponible en el mercado. El razonamiento es sencillo: si los par-

ticipantes en el mercado creen que una acción tendrá una rentabilidad extraordinaria positiva el mes que viene, entonces comprarán esa acción ahora; pero si lo hacen, subirá el precio de las acciones hasta exactamente el punto en que no exista una rentabilidad extraordinaria esperada. Como resultado, no debería ser posible pronosticar una rentabilidad extraordinaria futura mediante el uso de la información pasada públicamente disponible, al menos utilizando las regresiones de la Tabla 14.3.

14.4 Regresión de series temporales con predictores adicionales y modelo autorregresivo de retardos distribuidos

La teoría económica a menudo sugiere otras variables que podrían ayudar a predecir la variable de interés. Esas otras variables, o predictores, se pueden agregar a una autorregresión para dar lugar a un modelo de regresión de series temporales con varios predictores. Al añadir otras variables a una regresión y sus retardos, el resultado obtenido es un modelo autorregresivo de retardos distribuidos.

Predicción de la variación de la tasa de inflación mediante los valores pasados de la tasa de desempleo

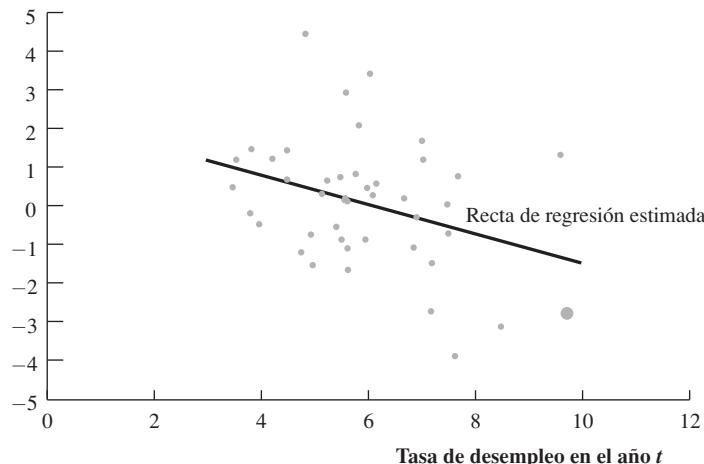
Un valor alto de la tasa de desempleo suele estar asociado con una futura disminución de la tasa de inflación. Esta relación negativa conocida como curva de Phillips de corto plazo, es evidente en el diagrama de dispersión de la Figura 14.3, en el que las variaciones año a año de la tasa de inflación de los precios están representadas junto con la tasa de desempleo del año anterior. Por ejemplo, en 1982 la tasa de desempleo promedió un 9,7 %, y durante el año siguiente la tasa de inflación cayó al 2,9 %. En total, la correlación en la Figura 14.3 es $-0,36$.

El diagrama de dispersión de la Figura 14.3 sugiere que los valores pasados de la tasa de desempleo podrían contener información acerca del curso futuro de la inflación que no se encuentra ya contenida en las

FIGURA 14.3 Diagrama de dispersión de la variación de la inflación entre el año t y el año $t+1$ versus la tasa de desempleo en el año t , 1961-2004

En 1982, la tasa de desempleo en EE.UU. fue del 9,7 % y la tasa de inflación en 1983 cayó al 2,9 % (el punto grande). En general, los valores altos de la tasa de desempleo en el año t suelen estar seguidos de disminuciones en la tasa de inflación de los precios el año siguiente, año $t + 1$, con una correlación de $-0,36$.

Variación de la inflación entre el año t y el año $t + 1$



variaciones pasadas de la inflación. Esta suposición resulta comprobable fácilmente ampliando el modelo AR(4) de la Ecuación (14.13) para incluir el primer retardo de la tasa de desempleo:

$$\begin{aligned}\widehat{\Delta Inf}_t = & 1,28 - 0,31\Delta Inf_{t-1} - 0,39\Delta Inf_{t-2} + 0,09\Delta Inf_{t-3} \\ & (0,53) \quad (0,09) \quad (0,09) \quad (0,08) \\ & - 0,08\Delta Inf_{t-4} - 0,21Desemp_{t-1} \\ & (0,09) \quad (0,09)\end{aligned}\tag{14.16}$$

El estadístico t , para la variable $Desemp_{t-1}$ es $-2,23$, por lo que este término es significativo al nivel del 5 %. El \bar{R}^2 de esta regresión es 0,21, una mejora sobre el \bar{R}^2 del modelo AR(4) de 0,18.

La predicción de la variación de la inflación en 2005:I, se obtiene mediante la sustitución de los valores de la variación de la inflación en 2004 en la Ecuación (14.16), junto con el valor de la tasa de desempleo en 2004:IV (que es del 5,4 %); la predicción resultante es $\widehat{\Delta Inf}_{2005:\text{II}|2004:\text{IV}} = 0,4$. Por lo tanto el pronóstico de la inflación en 2005:I es $3,5\% + 0,4\% = 3,9\%$, y el error de predicción es $-1,5\%$.

Si un retardo de la tasa de desempleo resulta útil para predecir la inflación, varios retardos podrían resultar aún más útiles; mediante la adición de tres retardos adicionales de la tasa de desempleo se obtiene

$$\begin{aligned}\widehat{\Delta Inf}_t = & 1,30 - 0,42\Delta Inf_{t-1} - 0,37\Delta Inf_{t-2} + 0,06\Delta Inf_{t-3} - 0,04\Delta Inf_{t-4} \\ & (0,44) \quad (0,08) \quad (0,09) \quad (0,08) \quad (0,08) \\ & - 2,64Desemp_{t-1} + 3,04Desemp_{t-2} - 0,38Desemp_{t-3} - 0,25Desemp_{t-4}. \\ & (0,46) \quad (0,86) \quad (0,89) \quad (0,45)\end{aligned}\tag{14.17}$$

El estadístico F para el contraste de significación conjunta de los retardos desde el segundo hasta el cuarto de la tasa de desempleo es igual a 10,76 (p -valor < 0,001), por lo que son conjuntamente significativos. El \bar{R}^2 de la regresión de la Ecuación (14.17) es 0,34, una sólida mejora sobre el 0,21 de la Ecuación (14.16). El estadístico F sobre todos los coeficientes de la tasa de desempleo es 8,91 (p -valor < 0,001), lo que indica que este modelo representa una mejora estadísticamente significativa con respecto al modelo AR(4) de la Sección 14.3 [Ecuación (14.13)]. El error estándar de la regresión de la Ecuación (14.17) es 1,36, una mejora sustancial sobre el ESR de 1,52 para el modelo AR(4).

La predicción de la variación de la tasa de inflación desde 2004:IV hasta 2005:I; utilizando la Ecuación (14.17) se calcula sustituyendo los valores de las variables en la ecuación. La tasa de desempleo fue del 5,7 % en 2004:I, del 5,6 % en 2004:II, y del 5,4 % en 2004:III y 2004:IV. La predicción para la variación de la inflación desde 2004:IV hasta 2005:I, en base a la Ecuación (14.17), es

$$\begin{aligned}\widehat{\Delta Inf}_{2005:\text{I}|2004:\text{IV}} = & 1,30 - 0,42 \times 1,9 - 0,37 \times (-2,8) + 0,06 \times 0,6 - 0,04 \\ & \times 2,9 - 2,66 \times 5,4 + 0,34 \times 5,4 - 0,38 \times 5,6 - 0,25 \times 5,7 = 0,1.\end{aligned}\tag{14.18}$$

Por lo tanto la predicción de la inflación para 2005:I es de $3,5\% + 0,1\% = 3,6\%$. El error de predicción es $-1,2$.

El modelo autorregresivo de retardos distribuidos. Cada uno de los modelos de las Ecuaciones (14.16) y (14.17) es un **modelo autorregresivo de retardos distribuidos (ARD)**: «autorregresivo» porque están incluidos los valores retardados de la variable dependiente como variables explicativas, como en una autorregresión, y de «retardos distribuidos» porque la regresión incluye asimismo varios retardos (una «distribución de retardos») de un predictor adicional. En general, un modelo autorregresivo de retardos distribuidos con p retardos de la variable dependiente Y_t y q retardos de un predictor adicional X , se denomina modelo **ARD(p, q)**. En esta notación, el modelo de la Ecuación (14.16) es un modelo ARD(4,1) y el modelo de la Ecuación (14.17) es un modelo ARD(4,4).

El modelo autorregresivo de retardos distribuidos se resume en el Concepto clave 14.4. Con todas estas variables explicativas, la notación de la Ecuación (14.19) resulta algo engorrosa, por lo que la notación alternativa opcional, basada en el denominado operador de retardos, se presenta en el Apéndice 14.3.

**CONCEPTO
CLAVE**

14.4

El modelo autorregresivo de retardos distribuidos

El modelo autorregresivo de retardos distribuidos con p retardos de Y_t y q retardos de X_t , denominado ARD(p, q), es

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} \\ + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \dots + \delta_q X_{t-q} + u_t, \quad (14.19)$$

donde $\beta_0, \beta_1, \dots, \beta_p, \delta_1, \dots, \delta_q$ son coeficientes desconocidos y u_t es el término de error con $E(u_t | Y_{t-1}, Y_{t-2}, \dots, X_{t-1}, X_{t-2}, \dots) = 0$.

El supuesto de que los errores en el modelo ARD tienen una media condicional igual a cero dados todos los valores pasados de Y y X , es decir, que $E(u_t | Y_{t-1}, Y_{t-2}, \dots, X_{t-1}, X_{t-2}, \dots) = 0$, implica que ningún retardo adicional ni de Y ni de X pertenece al modelo ARD. En otras palabras, las longitudes de los retardos p y q son las verdaderas longitudes de los retardos, y los coeficientes de los retardos adicionales son iguales a cero.

El modelo ARD contiene retardos de la variable dependiente (la componente autorregresiva) y un retardo distribuido de un único predictor adicional, X . Sin embargo, en general, las predicciones pueden mejorarse mediante la utilización de varios predictores. Pero antes de pasar al modelo general de series temporales con varios predictores, se introduce en primer lugar el concepto de estacionariedad, que será utilizado en este análisis.

Estacionariedad

El análisis de regresión de datos series temporales necesariamente utiliza los datos pasados para cuantificar las relaciones históricas. Si el futuro es como el pasado, entonces estas relaciones históricas pueden utilizarse para predecir el futuro. Pero si el futuro es fundamentalmente distinto del pasado, entonces esas relaciones históricas no podrán ser guías fiables para el futuro.

En el contexto de la regresión de series temporales, la idea de que las relaciones históricas pueden generalizarse para el futuro está formalizada en el concepto de **estacionariedad**. La definición precisa de la estacionariedad, que se ofrece en el Concepto clave 14.5, es que la distribución de probabilidad de la variable de series temporales no cambia en el tiempo.

**CONCEPTO
CLAVE**

14.5

Estacionariedad

Una serie temporal Y_t es *estacionaria* si su distribución de probabilidad no varía en el tiempo, es decir, si la distribución conjunta de $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ no depende de s sea cual sea el valor de T ; de lo contrario, se dice que Y_t es *no estacionaria*. Dos series temporales, X_t e Y_t , se dice que son *conjuntamente estacionarias* si la distribución conjunta de $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \dots, X_{s+T}, Y_{s+T})$ no depende de s , independientemente del valor de T . La estacionariedad requiere que el futuro sea como el pasado, al menos en un sentido probabilístico.

Regresión de series temporales con varios predictores

El modelo general de regresión de series temporales con varios predictores amplía el modelo ARD para incluir varios predictores y sus retardos. El modelo se resume en el Concepto clave 14.6. La presencia de varios predictores y de sus retardos lleva a que se utilice un doble subíndice en los coeficientes de regresión y los regresores.

**CONCEPTO
CLAVE**

14.6

Regresión de series temporales con varios predictores

El modelo general de regresión de series temporales permite k predictores adicionales, en el que se incluyen q_1 retardos del primer predictor, q_2 retardos del segundo predictor, y así sucesivamente:

$$\begin{aligned} Y_t = & \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\ & + \delta_{11} X_{1t-1} + \delta_{12} X_{1t-2} + \cdots + \delta_{1q_1} X_{1t-q_1} \\ & + \cdots + \delta_{k1} X_{kt-1} + \delta_{k2} X_{kt-2} + \cdots + \delta_{kq_k} X_{kt-q_k} + u_t, \end{aligned} \quad (14.20)$$

donde

1. $E(u_t | Y_{t-1}, Y_{t-2}, \dots, X_{1t-1}, X_{1t-2}, \dots, X_{kt-1}, X_{kt-2}, \dots) = 0$
2. (a) Las variables aleatorias $(Y_t, X_{1t}, \dots, X_{kt})$ presentan una distribución estacionaria, y
(b) $(Y_t, X_{1t}, \dots, X_{kt})$ y $(Y_{t-j}, X_{1t-j}, \dots, X_{kt-j})$ pasan a ser independientes cuando j se hace grande.
3. Los valores extremos elevados son poco probables: $X_{1t}, \dots, X_{kt}, Y_t$ presentan momentos de cuarto orden finitos y distintos de cero.
4. No existe multicolinealidad perfecta.

Los supuestos del modelo de regresión de series temporales. Los supuestos del Concepto clave 14.6 modifican los cuatro supuestos de mínimos cuadrados del modelo de regresión múltiple para datos de sección cruzada (Concepto clave 6.4) para los datos de series temporales.

El primer supuesto es que u_t tiene una media condicional igual a cero, dadas todas las variables explicativas y los retardos adicionales de las variables explicativas además de los retardos incluidos en la regresión. Este supuesto amplía el supuesto de los modelos AR y ARD e implica que la mejor predicción de Y_t utilizando todos los valores pasados de Y y de las X está dada por la regresión de la Ecuación (14.20).

El segundo supuesto de mínimos cuadrados para datos de sección cruzada (Concepto clave 6.4) es que $(X_{1i}, \dots, X_{ki}, Y_i)$, $i = 1, \dots, n$, son independientes e idénticamente distribuidas (i.i.d.). El segundo supuesto para la regresión de series temporales sustituye al supuesto de i.i.d. por otro más apropiado con dos partes. La parte (a) es que los datos se hayan obtenido a partir de una distribución estacionaria de manera que la distribución de los datos hoy es la misma que su distribución en el pasado. Este supuesto es una versión para series temporales de la parte «idénticamente distribuidas» del supuesto i.i.d.: el requisito para sección cruzada de que cada extracción sea idénticamente distribuida se sustituye por el requisito para series temporales de que la distribución conjunta de las variables, *includiendo los retardos*, no cambie en el tiempo. En la práctica, muchas series temporales económicas parecen ser no estacionarias, lo que significa que este supuesto puede no cumplirse en algunas situaciones concretas. Si las variables de series temporales son no estacionarias, entonces pueden aparecer uno o varios problemas en la regresión de series temporales: la predicción puede estar sesgada, la predicción puede ser ineficiente (pueden existir predicciones alternativas basadas en los mismos datos con una menor varianza), o la inferencia estadística basada en los MCO convencionales (por ejemplo, la realización de un contraste de hipótesis mediante la comparación del estadístico t MCO con $\pm 1,96$) puede resultar engañosa. La concreción de cuál de estos problemas aparece, así como su remedio, depende de la fuente de la ausencia de estacionariedad. En las Secciones 14.6 y 14.7, se estudian los problemas que plantean los contrastes para su detección, así como las soluciones disponibles para dos tipos importantes de no estacionariedad desde el punto de vista empírico de las series temporales económicas, las tendencias y los cambios estructurales. Sin embargo, por ahora simplemente se supone que las series son conjuntamente estacionarias y por lo tanto nos centraremos en la regresión con variables estacionarias.

La parte (b) del segundo supuesto requiere que las variables aleatorias sean independientemente distribuidas cuando son muchos los períodos de tiempo que las separan. Esto reemplaza el requisito para sección cruzada de que las variables estén independiente distribuidas entre una observación y la siguiente por el requisito para series temporales de que estén independientemente distribuidas cuando están separadas largos

periodos de tiempo. Este supuesto se denomina a veces **dependencia débil**, y asegura que en muestras grandes existe aleatoriedad suficiente en los datos como para que se cumpla la ley de los grandes números y el teorema central del límite. No se proporciona un enunciado matemático preciso de la condición de dependencia débil; en su lugar, se remite al lector a Hayashi (2000, Capítulo 2).

El tercer supuesto, que es el mismo que el tercer supuesto de mínimos cuadrados para datos de sección cruzada, es que los valores extremos elevados son poco probables, precisado en forma matemática supone que todas las variables presentan momentos de cuarto orden finitos y distintos de cero.

Por último, el cuarto supuesto, que del mismo modo es igual que para los datos de sección cruzada, es que los regresores no presenten multicolinealidad perfecta.

Inferencia estadística y contraste de causalidad de Granger. Bajo los supuestos del Concepto clave 14.6, la inferencia sobre los coeficientes de regresión utilizando MCO se realiza de la misma manera que es habitual para datos de sección cruzada.

Una aplicación útil del estadístico F para la predicción en series de temporales consiste en contrastar si los retardos de uno de los regresores incluidos tienen contenido predictivo útil, aparte de los otros regresores del modelo. La afirmación de que una variable no posee contenido predictivo corresponde a la hipótesis nula de que los coeficientes de todos los retardos de la variable son iguales a cero. El estadístico F para el contraste de esta hipótesis nula se denomina **estadístico de causalidad de Granger**, y el contraste asociado se denomina **contraste de causalidad de Granger** (Granger, 1969). Este contraste se resume en el Concepto clave 14.7.

La causalidad de Granger tiene poco que ver con la causalidad en el sentido que se utiliza en otras partes de este libro. En el Capítulo 1, se definía la causalidad en términos de un experimento aleatorizado controlado ideal, en el que los diferentes valores de X se aplicaban de forma experimental y se observaba el efecto consiguiente sobre Y . Por el contrario, la causalidad de Granger significa que si X causa a Y en el sentido de Granger, entonces X es un predictor útil de Y , dadas las otras variables de la regresión. Si bien la «predictibilidad en el sentido de Granger» es un término más preciso que la «causalidad en el sentido de Granger», esta última se ha convertido en parte de la jerga de la econometría.

Como ejemplo, puede considerarse la relación entre la variación de la tasa de inflación y sus valores pasados y los valores pasados de la tasa de desempleo. El estadístico F basado en las estimaciones MCO de la Ecuación (14.17), para el contraste de la hipótesis nula de que los coeficientes de los cuatro retardos de la tasa de desempleo son iguales a cero es 8,91 (p -valor < 0,001): en la jerga del Concepto clave 14.7, se puede concluir (al 1 % de nivel de significación) que la tasa de desempleo causa en el sentido de Granger las variaciones en la tasa de inflación. Esto *no* significa necesariamente que una variación de la tasa de desempleo causará —en el sentido del Capítulo 1— una variación posterior en la tasa de inflación. Lo que *sí* quiere decir es que los valores pasados de la tasa de desempleo parecen que contienen información que resulta útil para predecir los cambios en la tasa de inflación, más allá de la contenida en los valores pasados de la tasa de inflación.

CONCEPTO
CLAVE

14.7

Contraste de causalidad de Granger (contraste de contenido predictivo)

El estadístico para el contraste de causalidad de Granger es el estadístico F para el contraste de la hipótesis de que los coeficientes de todos los valores de una de las variables de la Ecuación (14.20) (por ejemplo, los coeficientes de X_{1t-1} , X_{1t-2} , ..., X_{1t-q_1}) son iguales a cero. Esta hipótesis nula implica que estos regresores no tienen contenido predictivo para Y_t , más allá del contenido en los otros regresores, y el contraste de esta hipótesis nula se denomina contraste de causalidad de Granger.

Incertidumbre de la predicción e intervalos de predicción

En cualquier problema de estimación, es una buena práctica proporcionar una medida de la incertidumbre de esa estimación, y la predicción no es una excepción. Una medida de la incertidumbre de una predic-

ción es la raíz del error cuadrático medio de predicción. Bajo el supuesto adicional de que los errores u_t se distribuyen normalmente, la RECM se puede utilizar para construir un intervalo de predicción, es decir, un intervalo que contenga el valor futuro de la variable con una cierta probabilidad.

Incertidumbre de la predicción. El error de predicción consta de dos componentes: la incertidumbre derivada de la estimación de los coeficientes de la regresión y la incertidumbre asociada con el valor desconocido futuro de u_t . Para las regresiones con pocos coeficientes y muchas observaciones, la incertidumbre derivada del u_t futuro puede ser mucho mayor que la incertidumbre asociada a la estimación de los parámetros. Sin embargo, en general, ambas fuentes de incertidumbre son importantes, por lo que se desarrolla ahora una expresión para la RECM que incorpora estas dos fuentes de incertidumbre.

Por simplicidad de notación, consideremos las predicciones de Y_{T+1} basadas en un modelo ARD(1,1) con un único predictor, es decir, $Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + u_t$, y supongamos que u_t es homocedástico. La predicción es $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\delta}_1 X_T$, y el error de predicción es

$$Y_{T+1} - \hat{Y}_{T+1|T} = y_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T]. \quad (14.21)$$

Debido a que u_{T+1} presenta una media condicional igual a cero y es homocedástico, u_{T+1} tiene una varianza σ_u^2 y no está correlacionado con la última expresión entre corchetes de la Ecuación (14.21). Por tanto, el error cuadrático medio de la predicción (ECMP) es

$$\begin{aligned} \text{ECMP} &= E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] \\ &= \sigma_u^2 + \text{var}[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T], \end{aligned} \quad (14.22)$$

y la RECM es la raíz cuadrada del ECMP.

La estimación del ECMP implica la estimación de las dos partes de la Ecuación (14.22). El primer término, σ_u^2 se puede estimar mediante el cuadrado del error estándar de la regresión, tal y como se estableció en la Sección 14.3. El segundo término requiere la estimación de la varianza de una media ponderada de los coeficientes de regresión, y los métodos para hacerlo se discutieron en la Sección 8.1 [véase la discusión al respecto de la Ecuación (8.7)].

Un método alternativo para la estimación del ECMP consiste en utilizar la varianza de la predicción pseudo-fuera de la muestra, procedimiento descrito en la Sección 14.7.

Intervalos de predicción. Un intervalo de predicción es como un intervalo de confianza, excepto por el hecho de que se refiere a una predicción. Es decir, un **intervalo de predicción** del 95 % es un intervalo que contiene el valor futuro de la serie en el 95 % de las realizaciones repetidas.

Una diferencia importante entre un intervalo de predicción y un intervalo de confianza es que la fórmula habitual para un intervalo de confianza al 95 % (el estimador $\pm 1,96$ errores estándar) está justificada por el teorema central del límite y por lo tanto es válida para una amplia gama de distribuciones del término de error. Por el contrario, debido a que el error de predicción de la Ecuación (14.21) incluye el valor futuro del error u_{T+1} , para calcular un intervalo de predicción se requiere o bien la estimación de la distribución del término de error, o bien la formulación de algunas hipótesis adicionales acerca de la distribución.

En la práctica, resulta conveniente suponer que u_{T+1} está normalmente distribuido. Si es así, la Ecuación (14.21) y el teorema central del límite aplicado a $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\delta}_1$ implican que el error de predicción es la suma de dos términos independientes, y normalmente distribuidos, por lo que el error de predicción se distribuye normalmente con varianza igual al ECMP. De ello se desprende que un intervalo de confianza al 95 % está dado por $\hat{Y}_{T+1|T} \pm 1,96\text{ES}(Y_{T+1} - \hat{Y}_{T+1|T})$. Donde $\text{ES}(Y_{T+1} - \hat{Y}_{T+1|T})$ es un estimador de la RECM.

Esta discusión se ha centrado en el caso en el que el término de error, u_{T+1} , es homocedástico. Si por el contrario, u_{T+1} es heterocedástico, entonces resulta necesario desarrollar un modelo para la heterocedasticidad para que el término σ_u^2 de la Ecuación (14.22) pueda estimarse, dados los valores más recientes de Y y X , y los métodos para la modelización de esta heterocedasticidad condicional se presentan en la Sección 16.5.

Debido a la incertidumbre acerca de los sucesos futuros —es decir, la incertidumbre acerca de u_{T+1} — los intervalos de predicción al 95 % pueden ser tan amplios que pueden llegar a tener una utilidad limitada en la toma de decisiones. Por eso, los expertos en previsión económica, a menudo presentan intervalos de predicción más estrechos que el 95 %, por ejemplo, un intervalo de predicción de un error estándar (que es un intervalo de predicción al 68 % si los errores se distribuyen normalmente). Por otra parte, algunos expertos en predicción económica presentan varios intervalos de predicción, tal como lo hacen los economistas del Banco de Inglaterra cuando publican sus previsiones de inflación (véase «El río de sangre» que se muestra más abajo).

El río de sangre

Como parte de los esfuerzos realizados para informar al público acerca de sus decisiones de política monetaria, el Banco de Inglaterra publica regularmente predicciones de inflación. Esas predicciones combinan los resultados obtenidos mediante los modelos estadísticos seguidos por los economistas profesionales del banco, con la opinión experta de los funcionarios veteranos del banco y el Comité de Política Monetaria. Los pronósticos se presentan en forma de conjunto de intervalos de predicción diseñados para reflejar lo que estos economistas consideran que es el rango de evolución probable que la inflación podría tomar. En su *Informe sobre inflación*, el banco imprime estos rangos en color rojo, reservando el rojo más oscuro para la banda central. Aunque el banco se refiere a esto de forma prosaica como «gráfico de abanico», la prensa ha denominado estos sombreados de tonalidades rojas como «río de sangre».

El río de sangre para noviembre de 2009 es el que se muestra en la Figura 14.4 (en esta figura la sangre es de color gris, no rojo por lo que tendrá que usarse la imaginación). Este gráfico muestra que, en noviembre de 2009, los economistas del banco esperaban que la tasa de inflación aumentase bruscamente hasta cerca del 3 % a principios de 2010, cayera aproximadamente hasta el 1 % a finales de 2010, y luego regresase a un aumento sostenido del 2 % durante 2012. No obstante, los economistas manifestaban una considerable incertidumbre acerca de la predicción. Citaban el aumento en el VAT¹ (impuesto sobre las ventas) como un factor importante del aumento de la inflación en el corto plazo y analizaban la incertidumbre asociada con la respuesta de la inflación a la ralentización de la economía y el ritmo y la fortaleza de la recuperación económica como fuentes importantes de incertidumbre inflacionaria. Como resultado, su pronóstico para un horizonte cercano estuvo muy cerca de la inflación real: la inflación en el segundo trimestre de 2010 fue del 3,5 %.

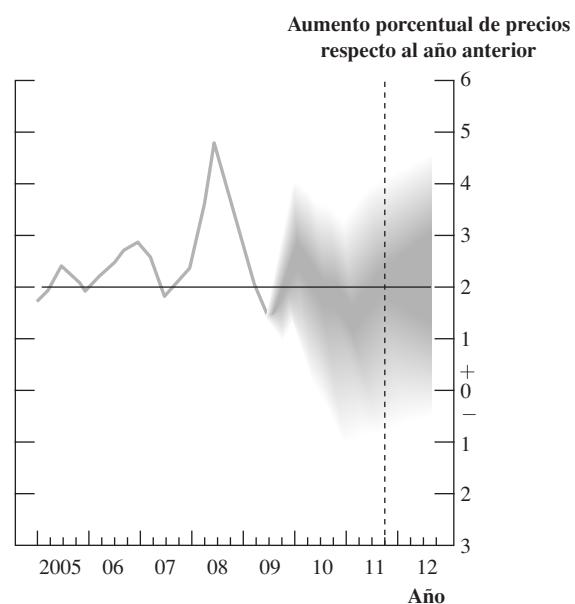
El Banco de Inglaterra ha sido pionero en el movimiento hacia una mayor apertura de los bancos centrales, y de la misma forma, otros bancos centrales publican ahora las predicciones de inflación. Las decisiones tomadas por los gestores de la política monetaria son difíciles y afectan a las vidas —y bille-

teras— de muchos de sus conciudadanos. En una democracia en la era de la información, razonan los economistas del Banco de Inglaterra, es particularmente importante que los ciudadanos entiendan las perspectivas económicas del banco y el razonamiento que subyace tras sus difíciles decisiones.

Para ver el río de sangre en su color rojo original, visite la página web del Banco de Inglaterra www.bankofengland.co.uk. Para obtener más información sobre la elaboración de predicciones de inflación por parte del Banco de Inglaterra, véase Clements (2004).

¹ N. del T.: Correspondiente al IVA.

FIGURA 14.4 El río de sangre



El gráfico de abanico del Banco de Inglaterra para noviembre 2009, muestra rangos de predicción para la inflación. La línea punteada indica el segundo trimestre de 2011, dos años después de la publicación del informe.

14.5 Selección de la longitud de los retardos mediante criterios de información

Las regresiones estimadas sobre la inflación de las Secciones 14.3 y 14.4 presentan uno o cuatro retardos de las variables predictoras. Un retardo tiene algún sentido, pero ¿por qué cuatro? De manera más general, ¿cuántos retardos se deben incluir en una regresión de series temporales? En esta sección se analizan los métodos estadísticos que se utilizan para elegir el número de retardos, en primer lugar en una autorregresión y luego, en un modelo de regresión de series temporales con varios predictores.

Determinación del orden de una autorregresión

En la práctica, en la elección del orden de una autorregresión, p , resulta necesario sopesar el beneficio marginal de incluir más retardos con el coste marginal de la incertidumbre adicional de la estimación. Por un lado, si el orden de una autorregresión estimada es demasiado bajo, se omite información potencialmente valiosa que está contenida en los valores retardados más distantes. Por otro lado, si el orden es demasiado elevado, se estiman más coeficientes que los necesarios, lo que a su vez introduce errores de estimación adicionales en las predicciones.

El método del estadístico F . Un método para la elección de p consiste en comenzar con un modelo con muchos retardos y llevar a cabo los contrastes de hipótesis sobre el último retardo. Por ejemplo, podría comenzarse por estimar un AR(6) y contrastar si el coeficiente del sexto retardo es significativo al nivel del 5 %; si no es así, se deja de lado y se estima un AR(5), contrastando el coeficiente del retardo quinto, etc. El inconveniente de este método es que va a dar lugar a un modelo demasiado grande, al menos algunas de las veces: incluso si el verdadero orden del AR es cinco, por lo que, por tanto, el coeficiente del sexto será igual a cero, un contraste al 5 % utilizando el estadístico t rechazaría incorrectamente esta hipótesis nula el 5 % de las veces en cada uno de los intentos. Por tanto, cuando el verdadero valor de p es igual a cinco, este método estimaría que p es igual a seis el 5 % de las veces.

El BIC. Una manera de sortear este problema es estimar p , mediante la minimización de un «criterio de información». Uno de estos criterios de información es el **criterio de información de Bayes (BIC)**, denominado asimismo *criterio de información Schwarz (SIC)*, que es

$$\text{BIC}(p) = \ln \left[\frac{\text{SR}(p)}{T} \right] + (p + 1) \frac{\ln(T)}{T}, \quad (14.23)$$

donde $\text{SR}(p)$ es la suma de los cuadrados de los residuos del modelo $\text{AR}(p)$ estimado. El estimador BIC de p , \hat{p} , es el valor que minimiza $\text{BIC}(p)$ entre las opciones posibles $p = 0, 1, \dots, p_{\max}$, donde p_{\max} es el mayor valor de p considerado y $p = 0$ corresponde al modelo que solamente contiene el término independiente.

La fórmula para el BIC puede parecer un poco misteriosa al principio, pero posee cierto encanto intuitivo. Consideremos el primer término de la Ecuación (14.23). Debido a que los coeficientes de regresión están estimados por MCO, la suma de los cuadrados de los residuos necesariamente disminuye (o al menos no aumenta) cuando se agrega un retardo. Por contra, el segundo término es el número de coeficientes de regresión estimados (el número de retardos, p , más uno para el término independiente) multiplicado por el factor $\ln(T)/T$. Este segundo término aumenta cuando se añade un retardo. El BIC sopesa estas dos fuerzas, por lo que el número de retardos que minimiza el BIC es un estimador consistente de la verdadera longitud de los retardos. Las matemáticas de este argumento aparecen en el Apéndice 14.5.

A modo de ejemplo, consideremos la estimación del orden del modelo AR para una autorregresión de la variación en la tasa de inflación. Las distintas etapas para el cálculo del BIC se llevan a cabo en la Tabla 14.4 para modelos autorregresivos con un orden máximo de seis ($p_{\max} = 6$). Por ejemplo, para el modelo AR(1) de la Ecuación (14.7), $SR(1)/T = 2,737$, por lo que $\ln[SR(1)/T] = 1,007$. Debido a que $T = 172$ (43 años, cuatro trimestres por año), $\ln(T)/T = 0,030$ y $(p + 1)\ln(T)/T = 2 \times 0,030 = 0,060$. Por lo tanto $BIC(1) = 1,007 + 0,060 = 1,067$.

El BIC más pequeño se alcanza cuando $p = 2$ en la Tabla 14.4. Por tanto, la estimación BIC para la longitud del retardo es 2. Como se puede observar en la Tabla 14.4, al aumentar el número de retardos aumenta el R^2 y disminuye la SR. El aumento de R^2 es grande cuando se pasa de uno a dos retardos, más pequeño de dos a tres, y bastante más pequeño de tres a cuatro. El BIC ayuda a decidir exactamente cuánto debe ser el aumento de R^2 para justificar la inclusión del retardo adicional.

El AIC. El BIC no es el único criterio de información; otro es el **criterio de información de Akaike (AIC)**:

$$AIC(p) = \ln\left[\frac{SR(p)}{T}\right] + (p + 1)\frac{2}{T} \quad (14.24)$$

La diferencia entre el criterio AIC y el BIC es que el término « $\ln(T)$ » del BIC se sustituye por «2» en el AIC, por lo que el segundo término del AIC es menor. Por ejemplo, para las 172 observaciones utilizadas para estimar los modelos autorregresivos para la inflación, $\ln(T) = \ln(172) = 5,15$, por lo que el segundo término del BIC es más del doble de grande que el término en el AIC. Por tanto, se necesita una disminución menor de la SR en el AIC para justificar la inclusión de otro retardo. Como cuestión teórica, el segundo término del AIC no es lo suficientemente grande como para asegurar que la longitud de los retardos se haya elegido correctamente, incluso en muestras grandes, por lo que el estimador AIC de p no es consistente. Tal y como se analiza en el Apéndice 14.5, en muestras grandes el criterio AIC sobreestima p con una probabilidad distinta de cero.

A pesar de este defecto teórico, el criterio AIC es utilizado habitualmente en la práctica. Si puede existir alguna preocupación acerca de que el BIC pueda dar lugar a un modelo con muy pocos retardos, el AIC ofrece una alternativa razonable.

Una nota acerca del cálculo de los criterios de información. Se evalúa mejor la bondad del ajuste a los datos de dos regresiones estimadas cuando ambas se calculan utilizando el mismo conjunto de datos. Debido a que el BIC y AIC son métodos formales para realizar esta comparación, los modelos autorregresivos en consideración deberían estimarse utilizando las mismas observaciones. Por ejemplo, en la Tabla 14.4

TABLA 14.4 El criterio de información de Bayes (BIC) y el R^2 para modelos autorregresivos de inflación en EE.UU., 1962-2004

<i>p</i>	<i>SR(p)/T</i>	<i>ln[SR(p)/T]</i>	<i>(p+1)ln(T)/T</i>	<i>BIC(p)</i>	<i>R²</i>
0	2,900	1,065	0,030	1,095	0,000
1	2,737	1,007	0,60	1,067	0,056
2	2,375	0,865	0,090	0,955	0,181
3	2,311	0,838	0,120	0,957	0,203
4	2,309	0,837	0,150	0,986	0,204
5	2,308	0,836	0,180	1,016	0,204
6	2,308	0,836	0,209	1,046	0,204

todas las regresiones se estimaron a partir de los datos desde 1962:I hasta 2004:IV, para un total de 172 observaciones. Debido a que los modelos autorregresivos incluyen retardos de la variación de la inflación, esto significa que los valores de la variación de la inflación anteriores (los valores anteriores a 1962:I) se utilizaron como variables explicativas de las observaciones preliminares. Dicho de otro modo, cada una de las regresiones examinadas en la Tabla 14.4, incluyen observaciones sobre $\Delta Inf_t, \Delta Inf_{t-2}, \dots, \Delta Inf_{t-p}$, para $t = 1962:I, \dots, 2004:IV$, lo que corresponde a 172 observaciones sobre la variable dependiente y las variables explicativas, por lo que $T = 172$ en las Ecuaciones (14.23) y (14.24).

Selección de la longitud de los retardos en una regresión de series temporales con varios predictores

El conflicto entre dos magnitudes que implica la selección de la longitud de los retardos en el modelo general de regresión de series temporales con varios predictores [Ecuación (14.20)] es similar al del modelo autorregresivo: utilizar demasiado pocos retardos puede hacer disminuir la precisión de las predicciones debido a que se pierde información valiosa, pero la adición de retardos aumenta la incertidumbre de la estimación. La elección de los retardos debe equilibrar las ventajas de utilizar información adicional con el coste de estimación de los coeficientes adicionales.

El método del estadístico F. Al igual que en el modelo autorregresivo univariante, una forma de determinar el número de retardos a incluir es mediante la utilización del estadístico F para el contraste de la hipótesis conjunta de que algún conjunto de coeficientes es igual a cero. Por ejemplo, en el análisis de la Ecuación (14.17), se contrastaba la hipótesis de que los coeficientes desde el segundo al cuarto retardo de la tasa de desempleo eran iguales a cero frente a la alternativa de que eran distintos de cero; esta hipótesis fue rechazada al nivel de significación del 1 %, lo que apoyaba la especificación con mayor longitud de retardos. Si el número de modelos que se comparan es pequeño, entonces este método del estadístico *F* resulta fácil de utilizar. Sin embargo, en general, el método del estadístico *F* puede dar lugar a modelos que sean demasiado grandes, en el sentido de que el verdadero orden de los retardos se sobreestime.

Criterios de información. Al igual que en un modelo autorregresivo, el BIC y el AIC pueden utilizarse para estimar el número de retardos y de variables del modelo de regresión de series temporales con varios predictores. Si el modelo de regresión tiene K coeficientes (incluyendo el término independiente), El BIC es

$$\text{BIC}(K) = \ln \left[\frac{SR(K)}{T} \right] + K \frac{\ln(T)}{T} \quad (14.25)$$

El AIC se define del mismo modo, pero sustituyendo el término $\ln(T)$ de la Ecuación (14.25) por el término 2. Se puede evaluar el BIC (o el AIC) para cada modelo candidato, y el modelo preferido será el del valor del BIC (o del AIC) más bajo, en base al criterio de información.

Existen dos consideraciones de orden práctico importantes a la hora de utilizar un criterio de información para calcular la longitud de los retardos. En primer lugar, como en el caso del modelo autorregresivo, deben estimarse todos los modelos candidatos a lo largo de la misma muestra; en la notación de la Ecuación (14.25), el número de observaciones utilizadas para estimar el modelo, T , debe ser el mismo para todos los modelos. En segundo lugar, cuando hay varios predictores, este método resulta exigente desde el punto de vista computacional, debido a que requiere el cálculo de muchos modelos diferentes (muchas combinaciones de los parámetros de los retardos). En la práctica, conviene utilizar un atajo que consiste en exigir que todas las variables explicativas tengan el mismo número de retardos, es decir, exigir que $p = q_1 = \dots = q_k$, por lo que solamente es necesario comparar $p_{\max} + 1$ modelos (lo que corresponde a $p = 0, 1, \dots, p_{\max}$).

14.6 Ausencia de estacionariedad I: tendencias

En el Concepto clave 14.6, se suponía que la variable dependiente y los regresores eran estacionarios. Si este no es el caso, es decir, si la variable dependiente y/o los regresores no son estacionarios, los contrastes de hipótesis, los intervalos de confianza y las predicciones habituales pueden resultar poco fiables. El problema exacto ocasionado por la ausencia de estacionariedad, y la solución a este mismo problema, depende de la naturaleza de la ausencia de estacionariedad.

En esta sección y en la siguiente, se examinan dos de los más importantes tipos de ausencia de estacionariedad en los datos de series temporales económicas: las tendencias y los cambios estructurales. En cada una de las secciones, se describe en primer lugar la naturaleza de la no estacionariedad y más tarde se discuten las consecuencias que origina sobre la regresión de series de temporales, el hecho de ignorar la presencia de este tipo de no estacionariedad. A continuación se presentan los contrastes de detección de no estacionariedad y se discuten los remedios o las soluciones a los problemas causados por ese tipo concreto de no estacionariedad. Comenzamos por analizar las tendencias

¿Qué es una tendencia?

Una **tendencia** es un movimiento persistente a largo plazo de una variable en el tiempo. Una variable de series temporales fluctúa en torno a su tendencia.

La observación de la Figura 14.1a sugiere que la tasa de inflación en los EE.UU. presenta una tendencia que consiste en una propensión general creciente hasta 1982 y una propensión descendente a partir de entonces. Las series de las Figuras 14.2a, b, c, asimismo presentan tendencias, pero sus tendencias son muy diferentes. La tendencia del tipo de interés de los fondos federales en EE.UU. es similar a la tendencia de la tasa de inflación de los EE.UU. El tipo de cambio \$/£ presenta claramente una prolongada tendencia a la baja tras el colapso del sistema de tipos de cambio fijos en 1972. El logaritmo del PIB de Japón presenta una tendencia complicada: crecimiento rápido al principio, más tarde crecimiento moderado, y finalmente crecimiento lento.

Tendencias determinísticas y estocásticas. Existe la posibilidad de observar dos tipos de tendencias en datos de series temporales: determinísticas y estocásticas. Una **tendencia determinística** es una función no aleatoria del tiempo. Por ejemplo, una tendencia determinística puede ser lineal en el tiempo; si la inflación presenta una tendencia determinística lineal, de forma que aumenta en 0.1 puntos porcentuales por trimestre, esta tendencia podría escribirse como $0.1t$, donde t se mide en períodos. Por el contrario, una **tendencia estocástica** es aleatoria y varía en el tiempo. Por ejemplo, una tendencia estocástica en la inflación podría presentar un periodo prolongado de crecimiento seguido por un periodo prolongado de decrecimiento, como el de la tendencia que presenta la inflación en la Figura 14.1.

Al igual que muchos económetras, creemos que resulta más apropiado modelizar las series temporales económicas de forma que presenten tendencias estocásticas en lugar de tendencias determinísticas. La economía resulta complicada. Resulta duro conciliar la predictibilidad que implica una tendencia determinística con las complicaciones y sorpresas que afrontan año tras año los trabajadores, las empresas y los gobiernos. Por ejemplo, aunque la inflación en EE.UU. subiera durante la década de 1970, ni estaba destinada a un aumento eterno ni estaba destinada a volver a disminuir. Por el contrario, el lento ascenso de la inflación en la actualidad se entiende que se produjo a causa de la mala suerte y los errores de política monetaria, y su control fue en gran parte una consecuencia de las difíciles decisiones tomadas por la Junta de Gobernadores de la Reserva Federal. Del mismo modo, el tipo de cambio \$/£ presentaba una tendencia bajista desde 1972 hasta 1985 y, posteriormente ascendente, pero estos movimientos fueron asimismo las consecuencias de fuerzas económicas complejas; debido a que estas fuerzas cambian de manera impredecible, se interpreta que estas tendencias poseen una componente muy imprevisible, o aleatoria.

Por estas razones, nuestro tratamiento de las tendencias en series temporales económicas se centra en las tendencias estocásticas, más que en las tendencias determinísticas, y cuando nos refiramos a las «tenden-

cias» en datos de series temporales nos referimos a las tendencias estocásticas a menos que explícitamente se diga otra cosa. Esta sección presenta el modelo más sencillo para una tendencia estocástica, el modelo de paseo aleatorio; en la Sección 16.3 se analizan otros modelos de tendencias.

El modelo de paseo aleatorio para la tendencia. El modelo más sencillo de una variable con tendencia estocástica es el paseo aleatorio. Se dice que una serie temporal Y_t sigue un **paseo aleatorio** si la variación de Y_t es i.i.d., es decir, si

$$Y_t = Y_{t-1} + u_t, \quad (14.26)$$

donde u_t es i.i.d. No obstante, utilizaremos el término *paseo aleatorio* en un sentido más general para referirnos a una serie temporal que sigue la Ecuación (14.26), donde u_t presenta una media condicional igual a cero; es decir, $E(u_t|Y_{t-1}, Y_{t-2}, \dots) = 0$.

La idea básica de un paseo aleatorio es que el valor de la serie mañana es su valor hoy, más una variación impredecible: debido a que la senda seguida por Y_t consiste en «pasos» aleatorios u_t , esa senda es un «paseo aleatorio». La media condicional de Y_t basada en los datos hasta el periodo $t-1$ es Y_{t-1} ; es decir, debido a que $E(u_t|Y_{t-1}, Y_{t-2}, \dots) = 0$, entonces $E(Y_t|Y_{t-1}, Y_{t-2}, \dots) = Y_{t-1}$. En otras palabras, si Y_{t-1} sigue un proceso de paseo aleatorio, entonces la mejor predicción del valor de mañana es su valor en la actualidad.

Algunas series, como el logaritmo del PIB japonés de la Figura 14.2c, presentan una tendencia obviamente creciente, en cuyo caso el mejor pronóstico de la serie debe incluir un ajuste por el aumento tendencial de la serie. Este ajuste lleva a una extensión del modelo de paseo aleatorio para incluir un movimiento tendencial, o «deriva», en una dirección u otra. Esta extensión se conoce como **paseo aleatorio con deriva**:

$$Y_t = \beta_0 + Y_{t-1} + u_t, \quad (14.27)$$

donde $E(u_t|Y_{t-1}, Y_{t-2}, \dots) = 0$ y β_0 es la «deriva» del paseo aleatorio. Si β_0 es positivo, entonces Y_t aumenta en promedio. En el modelo de paseo aleatorio con deriva, la mejor predicción de la serie mañana es el valor de la serie hoy, más la deriva β_0 .

El modelo de paseo aleatorio (con deriva en su caso) es sencillo y versátil, y es el principal modelo para las tendencias utilizado en este libro.

Un paseo aleatorio es no estacionario. Si Y_t sigue un paseo aleatorio, entonces no es estacionaria: la varianza de un paseo aleatorio aumenta en el tiempo, por lo que la distribución de Y_t cambia en el tiempo. Una forma de comprobarlo es reconocer que, debido a que u_t está incorrelacionado con Y_{t-1} en la Ecuación (14.26), $\text{var}(Y_t) = \text{var}(Y_{t-1}) + \text{var}(u_t)$; para que Y_t sea estacionaria $\text{var}(Y_t)$ no puede depender del tiempo, por lo que, en particular debe cumplirse $\text{var}(Y_t) = \text{var}(Y_{t-1})$, pero esto solo puede ocurrir si $\text{var}(u_t) = 0$. Otra forma de comprobarlo es imaginar que Y_t comienza en cero; es decir, $Y_0 = 0$. Por lo tanto $Y_1 = u_1$, y $Y_2 = u_1 + u_2$, y así sucesivamente de manera que $Y_t = u_1 + u_2 + \dots + u_t$. Debido a que u_t está serialmente incorrelacionado, $\text{var}(Y_t) = \text{var}(u_1 + u_2 + \dots + u_t) = t\sigma_u^2$. Por tanto, la varianza de Y_t depende de t ; de hecho, aumenta a medida que t aumenta. Debido a que la varianza de Y_t depende de t , su distribución depende de t , es decir, es no estacionaria.

Debido a que la varianza de un paseo aleatorio aumenta sin límite, sus autocorrelaciones *poblacionales* no están definidas (la primera autocovarianza y la varianza son infinitas, y el cociente entre las dos no está bien definido). Sin embargo, una característica de un paseo aleatorio es que sus autocorrelaciones *muestrales* suelen ser próximas a 1; de hecho, la j -ésima autocorrelación muestral de un paseo aleatorio converge a 1 en probabilidad.

Tendencias estocásticas, modelos autorregresivos, y raíz unitaria. El modelo de paseo aleatorio es un caso particular del modelo AR(1) [Ecuación (14.8)] en el que $\beta_1 = 1$. En otras palabras, si Y_t sigue un proceso AR(1) con $\beta_1 = 1$, entonces Y_t presenta una tendencia estocástica y es no estacionaria. Sin embargo, si $|\beta_1| < 1$, y u_t es estacionario, entonces la distribución conjunta de Y_t y sus retardos no depende de t (un resultado que se demuestra en el Apéndice 14.2), por lo que Y_t es estacionaria.

La condición análoga para que un proceso AR(p) sea estacionario es más complicada que la condición $|\beta_1| < 1$ para un AR(1). Su enunciado formal implica que las raíces del polinomio, $1 - \beta_1 z - \beta_2 z^2 -$

$-\beta_3z^3 - \dots - \beta_pz^p$. (Las raíces de este polinomio son los valores de z que satisfacen $1 - \beta_1z - \beta_2z^2 - \dots - \beta_3z^3 - \dots - \beta_pz^p = 0$). Para que un proceso AR(p) sea estacionario, todas las raíces de este polinomio deben ser mayores que 1 en valor absoluto. En el caso particular de un proceso AR(1), la raíz es el valor de z que resuelve la ecuación $1 - \beta_1z = 0$, por lo que su raíz es $z = 1/\beta_1$. Por lo que la afirmación de que la raíz debe ser mayor que 1 en valor absoluto equivale a que $\beta_1 < 1$.

Si un proceso AR(p) tiene una raíz que es igual a 1, se dice que la serie presenta una *raíz unitaria autorregresiva* o, más sencillamente, una **raíz unitaria**. Si Y_t presenta una raíz unitaria, entonces presenta una tendencia estocástica. Si Y_t es estacionaria (y por lo tanto no tiene una raíz unitaria), no presenta una tendencia estocástica. Por esta razón, utilizaremos los términos *tendencia estocástica* y *raíz unitaria* de manera intercambiable.

Problemas ocasionados por las tendencias estocásticas

Si un regresor presenta una tendencia estocástica (tiene una raíz unitaria), entonces el estimador MCO de su coeficiente y su estadístico t MCO pueden presentar una distribución no estándar (es decir, distinta de la normal), incluso en muestras grandes. Se analizan tres aspectos específicos de este problema: (1) el estimador del coeficiente autorregresivo en un modelo AR(1) está sesgado hacia 0 si su verdadero valor es 1, (2) el estadístico t de un regresor con una tendencia estocástica puede tener una distribución distinta de la normal, incluso en muestras grandes; y (3) un ejemplo extremo de los riesgos que plantean las tendencias estocásticas es que dos series que son independientes parecerá, con alta probabilidad, que están relacionadas, de forma engañosa, si las dos presentan tendencias estocásticas, una situación conocida como regresión espuria.

Problema #1: coeficientes autorregresivos que están sesgados hacia cero. Supongamos que Y_t sigue el paseo aleatorio de la Ecuación (14.26), pero que esto es desconocido por el económetra, que en su lugar estima el modelo AR(1) de la Ecuación (14.8). Debido a que Y_t es no estacionaria, los supuestos de mínimos cuadrados para la regresión de series temporales del Concepto clave 14.6 no se cumplen, por lo que, con carácter general, no puede confiarse en que los estimadores y los estadísticos de contraste presenten sus distribuciones normales habituales para grandes muestras. De hecho, en este ejemplo el estimador MCO del coeficiente autorregresivo, $\hat{\beta}_1$, es consistente, pero tiene una distribución distinta de la normal, incluso en muestras grandes: la distribución asintótica de $\hat{\beta}_1$ se desplaza hacia cero. La esperanza de $\hat{\beta}_1$ es aproximadamente $E(\hat{\beta}_1) = 1 - 5,3/T$. Esto se traduce en un gran sesgo para los tamaños muestrales que suelen encontrarse en las aplicaciones económicas. Por ejemplo, 20 años de datos trimestrales contienen 80 observaciones, en cuyo caso el valor esperado de $\hat{\beta}_1$ es $E(\hat{\beta}_1) = 1 - 5,3/80 = 0,934$. Por otra parte, esta distribución presenta un cola izquierda larga: el percentil 5 de $\hat{\beta}_1$ es aproximadamente $1 - 14,1/T$, lo que, debido a que $T = 80$, corresponde a 0,824, por lo que el 5 % de las veces $\hat{\beta}_1 < 0,824$.

Una consecuencia de este sesgo hacia cero es que si Y_t sigue un paseo aleatorio, entonces las predicciones basadas en el modelo AR(1) pueden comportarse de modo sustancialmente peor que las basadas en el modelo de paseo aleatorio, que impone el verdadero valor $\beta_1 = 1$. Esta conclusión es aplicable asimismo a modelos autorregresivos de orden superior, en los que hay ganancias en las predicciones imponiendo una raíz unitaria (es decir, estimando la autorregresión en primeras diferencias en lugar de en niveles), cuando en realidad la serie contiene una raíz unitaria.

Problema #2: distribuciones no normales del estadístico t . Si un regresor presenta una tendencia estocástica, entonces su estadístico t MCO habitual puede presentar, bajo la hipótesis nula, una distribución distinta de la normal, incluso en muestras grandes. Esta distribución no normal significa que no son válidos los intervalos de confianza habituales y los contrastes de hipótesis no pueden llevarse a cabo como de costumbre. En general, la distribución de este estadístico t no estará tabulada previamente debido a que la distribución depende de la relación entre el regresor en cuestión y el resto de los regresores. Un ejemplo importante de este problema aparece en las regresiones que tratan de predecir las rentabilidades de las acciones mediante regresores que podrían presentar tendencias estocásticas (véase el recuadro de la Sección 14.7, «¿Se puede batir al mercado? Parte II»).

Un caso importante en el que es posible tabular la distribución del estadístico t cuando el regresor tiene una tendencia estocástica es en el contexto de un modelo autorregresivo con una raíz unitaria. Volveremos a este caso particular cuando abordemos el problema de comprobar si una serie temporal contiene una tendencia estocástica.

Problema #3: regresión espuria. Las tendencias estocásticas pueden originar que dos series temporales parezcan estar relacionadas cuando en realidad no lo están, un problema que se denomina **regresión espuria**.

Por ejemplo, la inflación en EE.UU. aumentó de forma sostenida desde mediados de la década de 1960 hasta los primeros años de la década de 1980, y al mismo tiempo el PIB japonés (representado en logaritmos en la Figura 14.2c) aumentó de forma sostenida. Estas dos tendencias conspiran para dar lugar a una regresión que parece ser «significativa» utilizando las medidas convencionales. Estimada por MCO utilizando los datos desde 1965 a 1981, esta regresión es

$$\widehat{\text{InflaciónEE.UU.}_t} = -37,78 + 3,83 \times \ln(\text{PIBJaponés}_t), \bar{R}^2 = 0,56. \quad (14.28)$$

(3,99) (0,36)

El estadístico t del coeficiente de la pendiente es mayor que 10, lo que para los estándares habituales indica una fuerte relación positiva entre las dos series, y el \bar{R}^2 es elevado. Sin embargo, la realización de esta regresión con los datos desde 1982 a 2004 proporciona los siguientes resultados

$$\widehat{\text{InflaciónEE.UU.}_t} = 31,20 + 2,17 \times \ln(\text{PIBJaponés}_t), \bar{R}^2 = 0,08. \quad (14.29)$$

(10,41) (0,80)

Las regresiones de las Ecuaciones (14.28) y (14.29) no podrían ser más diferentes. Interpretadas en términos literales, la Ecuación (14.28) indica una fuerte relación positiva, mientras que la Ecuación (14.29) indica una débil relación negativa, pero en apariencia estadísticamente significativa.

La razón de estos resultados contradictorios es que ambas series presentan tendencias estocásticas. Estas tendencias se alinearon desde 1965 hasta 1981, pero no hubo alineamiento en el periodo que va desde 1982 hasta 2004. No existe, de hecho, ninguna razón de peso, ya sea desde el punto de vista económico o político, para creer que las tendencias de estas dos series estén relacionadas. En resumen, estas regresiones son espurias.

Las regresiones de las Ecuaciones (14.28) y (14.29) ilustran de forma empírica la cuestión teórica de que MCO puede resultar engañoso cuando las series contienen tendencias estocásticas (véase el Ejercicio 14.6 para una simulación por ordenador que demuestra este resultado). Un caso particular en el que algunos métodos basados en la regresión *resultan* fiables es cuando la componente de tendencia de las dos series es la misma, es decir, cuando las series contienen una tendencia estocástica *común*, si es así, se dice que las series están cointegradas. Los métodos econométricos para la detección y el análisis de series temporales económicas cointegradas se discuten en la Sección 16.4.

Detección de tendencias estocásticas: contraste de raíz unitaria AR

La tendencias en los datos de series temporales pueden detectarse mediante métodos formales e informales. Los métodos informales incluyen la inspección de los gráficos de los datos de series temporales y el cálculo de los coeficientes de autocorrelación, tal y como se llevó a cabo en la Sección 14.2. Debido a que el primer coeficiente de autocorrelación estará cerca de 1 si la serie tiene una tendencia estocástica, al menos en muestras grandes, un primer coeficiente de autocorrelación pequeño combinado con un gráfico de la serie temporal que aparentemente no tenga tendencia sugiere que la serie no tiene tendencia. Sin embargo, si queda alguna duda, existen procesos estadísticos formales que pueden utilizarse para contrastar la hipótesis de que existe una tendencia estocástica en la serie frente a la alternativa de que no existe ninguna tendencia.

En esta sección, se utiliza el contraste de Dickey-Fuller (denominado así por sus inventores David Dickey y Wayne Fuller, 1979) para contrastar la presencia de una tendencia estocástica. A pesar de que el

contraste de Dickey-Fuller no es el único contraste de presencia de una tendencia estocástica (se discute otro contraste en la Sección 16.3), es el contraste más utilizado en la práctica y es uno de los más fiables.

El contraste de Dickey-Fuller en el modelo AR(1). El punto de partida del **contraste de Dickey-Fuller** es el modelo autorregresivo. Como se señaló anteriormente, el paseo aleatorio de la Ecuación (14.27) es un caso particular del modelo AR(1) con $\beta_1 = 1$. Si $\beta_1 = 1$, Y_t es no estacionaria y contiene una tendencia (estocástica). Por tanto, en el modelo AR(1), la hipótesis de que Y_t tiene una tendencia puede ser contrastada mediante el contraste

$$H_0: \beta_1 = 1 \text{ vs. } H_1: \beta_1 < 1 \text{ en } Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t. \quad (14.30)$$

Si $\beta_1 = 1$, el AR(1) tiene una raíz autorregresiva igual a 1, por lo que la hipótesis nula en la Ecuación (14.30) es que el AR(1) tiene una raíz unitaria, y la alternativa es que es estacionario.

Este contraste resulta más fácil de llevar a la práctica mediante la estimación de una versión modificada de la Ecuación (14.30) obtenida restando Y_{t-1} de ambos lados. Sea $\delta = \beta_1 - 1$; por tanto la Ecuación (14.30) se convierte en

$$H_0: \delta = 0 \text{ vs. } H_1: \delta < 0 \text{ en } \Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t. \quad (14.31)$$

El estadístico t MCO para contrastar $\delta = 0$ en la Ecuación (14.31) se denomina **estadístico de Dickey-Fuller**. La formulación de la Ecuación (14.31) resulta conveniente debido a que el *software* de regresión reporta automáticamente el estadístico t para contrastar $\delta = 0$. Téngase en cuenta que el contraste de Dickey-Fuller es unilateral, debido a que la alternativa relevante es que Y_t sea estacionaria, por lo que $\beta_1 < 1$ o, equivalentemente, $\delta < 0$. El estadístico de Dickey-Fuller se calcula utilizando errores estándar «no robustos», es decir, los errores estándar «válidos con homocedasticidad» presentados en el Apéndice 5.1 [Ecuación (5.29) para el caso de un único regresor y en la Sección 18.4 para el modelo de regresión múltiple]³.

El contraste de Dickey-Fuller en el modelo AR(p). El estadístico de Dickey-Fuller presentado en el contexto de la Ecuación (14.31) solamente es aplicable a un AR(1). Tal y como se analizó en la Sección 14.3, en algunas series el modelo AR(1) no capta toda la correlación serial de Y_t , en cuyo caso, resulta más apropiado un modelo autorregresivo de orden superior.

La extensión del contraste de Dickey-Fuller al modelo AR(p) se resume en el Concepto clave 14.8. Bajo la hipótesis nula, $\delta = 0$ y ΔY_t es un proceso AR(p) estacionario. Bajo la hipótesis alternativa, $\delta < 0$, por lo que Y_t es estacionario. Debido a que el modelo de regresión utilizado para calcular esta versión del estadístico de Dickey-Fuller se ve aumentado por los retardos de ΔY_t , el estadístico t resultante se conoce como **estadístico de Dickey-Fuller aumentado (ADF)**.

En general, la longitud del retardo p es desconocida, pero puede estimarse mediante un criterio de información aplicado a las regresiones del tipo de la Ecuación (14.32) para distintos valores de p . Los estudios del estadístico ADF sugieren que es mejor tener demasiados retardos que demasiado pocos, por lo que se recomienda la utilización del criterio de información AIC en lugar del criterio de información BIC para estimar p para el estadístico ADF⁴.

Contraste frente a la alternativa de estacionariedad en torno a una tendencia temporal determinística. Hasta el momento, el análisis ha considerado la hipótesis nula de que la serie tiene una raíz unitaria, y la hipótesis alternativa de que es estacionaria. Esta hipótesis alternativa de estacionariedad es adecuada para las series, como la tasa de inflación, que no presentan crecimiento a largo plazo. Sin embargo, otras series temporales económicas, como el PIB japonés (Figura 14.2c), muestran crecimiento a largo plazo, y para tales series la alternativa de estacionariedad sin tendencia no es apropiada. En cambio, una alternativa que se utiliza habitualmente es que las series sean estacionarias alrededor de una tendencia temporal determinística, es decir, una tendencia que sea una función determinística del tiempo.

³ Bajo la hipótesis nula de presencia de una raíz unitaria, los errores estándar «no robustos» habituales dan lugar a un estadístico t que es en realidad robusto respecto de heterocedasticidad, un resultado sorprendente y particular.

⁴ Véase Stock (1994) y Haldrup y Jansson (2006) para una revisión de los estudios de simulación de las propiedades en muestras finitas del estadístico de Dickey-Fuller y otros estadísticos de contraste de raíz unitaria.

**CONCEPTO
CLAVE**
14.8
El contraste de Dickey-Fuller aumentado para raíz unitaria autorregresiva

El contraste de Dickey-Fuller aumentado (ADF) para una raíz unitaria autorregresiva contrasta la hipótesis nula $H_0: \delta = 0$ frente a la hipótesis alternativa unilateral $H_1: \delta < 0$ en la regresión

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t. \quad (14.32)$$

Bajo la hipótesis nula, Y_t tiene una tendencia estocástica; bajo la hipótesis alternativa, Y_t es estacionaria. El estadístico ADF es el estadístico t MCO para contrastar $\delta = 0$ en la Ecuación (14.32).

Si en su lugar, la hipótesis alternativa es que Y_t es estacionaria en torno a una tendencia temporal lineal determinística, entonces debe añadirse esta tendencia, « t » (el número de observación), como regresor adicional, en cuyo caso la regresión de Dickey-Fuller se convierte en

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t, \quad (14.33)$$

donde α es un coeficiente desconocido y el estadístico ADF es el estadístico t MCO para contrastar $\delta = 0$ en la Ecuación (14.33).

La longitud del retardo, p , se puede estimar utilizando el criterio BIC o AIC. Cuando $p = 0$, no se incluyen retardos de ΔY_t como regresores en las Ecuaciones (14.32) y (14.33), y el contraste ADF se simplifica al contraste de Dickey-Fuller en el modelo AR(1). El estadístico ADF *no* sigue una distribución normal, incluso en muestras grandes. Los valores críticos para el contraste ADF unilateral dependen de si el contraste está basado en la Ecuación (14.32) o en la (14.33) y se presentan en la Tabla 14.5.

Una formulación específica de esta hipótesis alternativa es que la tendencia temporal es lineal, es decir, la tendencia es una función lineal de t ; por lo tanto, la hipótesis nula es que la serie tiene una raíz unitaria, y la alternativa es que no tiene una raíz unitaria, pero tiene una tendencia temporal determinística. La regresión de Dickey-Fuller debe modificarse para contrastar la hipótesis nula de una raíz unitaria frente a la alternativa de que es estacionaria alrededor de una tendencia temporal lineal. Como se resume en la Ecuación (14.33) del Concepto clave 14.8, esto se logra añadiendo una tendencia temporal (el regresor $X_t = t$) a la regresión.

Una tendencia temporal lineal no es el único modo de especificar una tendencia temporal determinística; por ejemplo, la tendencia temporal determinística podría ser cuadrática, o podría ser lineal pero presentar cambios estructurales (es decir, ser lineal con pendientes que son distintas en dos partes de la muestra). La utilización de alternativas como estas, con tendencias determinísticas no lineales, debería estar motivada por la teoría económica. Para un estudio detallado de los contrastes de raíz unitaria frente a estacionariedad en torno a tendencias determinísticas no lineales, véase Maddala y Kim (1998, Capítulo 13).

Valores críticos para el estadístico ADF. Bajo la hipótesis nula de existencia de una raíz unitaria, el estadístico ADF *no* tiene una distribución normal, incluso en muestras grandes. Debido a que su distribución no es estándar, no pueden ser utilizados los valores críticos habituales de la distribución normal cuando se utiliza el estadístico ADF para el contraste de raíz unitaria; debe utilizarse en su lugar un conjunto particular de valores críticos, basados en la distribución del estadístico ADF bajo la hipótesis nula.

Los valores críticos para el contraste ADF se ofrecen en la Tabla 14.5. Debido a que la hipótesis alternativa de estacionariedad implica que en las Ecuaciones (14.32) y (14.33), $\delta < 0$, el contraste ADF es unilateral. Por ejemplo, si la regresión no incluye una tendencia temporal, entonces se rechaza la hipótesis de una raíz unitaria al nivel de significación del 5 % si el estadístico ADF es menor que $-2,86$. Si se incluye una tendencia temporal en la regresión, el valor crítico en su lugar es $-3,41$.

TABLA 14.5 Valores críticos para muestras grandes del estadístico Dickey-Fuller aumentado

Regresores determinísticos	10 %	5 %	1 %
Solamente término independiente	-2,57	-2,86	-3,43
Término independiente y tendencia temporal	-3,12	-3,41	-3,96

Los valores críticos de la Tabla 14.5 son considerablemente menores (más negativos) que los valores críticos unilaterales de $-1,28$ (al nivel del 10 %) y $-1,645$ (al 5 % de nivel) de la distribución normal estándar. La distribución no estándar del estadístico ADF es un ejemplo de cómo los estadísticos t MCO para los regresores con tendencias estocásticas pueden tener distribuciones no normales. La razón por la cual la distribución del estadístico ADF no es estándar se trata en profundidad en la Sección 16.3.

¿Presenta la inflación en EE.UU. una tendencia estocástica? Se puede contrastar la hipótesis nula de que la inflación tiene una tendencia estocástica frente a la alternativa de que es estacionaria mediante un contraste ADF para una raíz unitaria autorregresiva. La regresión ADF con cuatro retardos de ΔInf_t es

$$\widehat{\Delta Inf_t} = 0,51 - 0,11\Delta Inf_{t-1} - 0,19\Delta Inf_{t-2} - 0,26\Delta Inf_{t-3} + 0,20\Delta Inf_{t-4} + 0,01\Delta Inf_{t-5}. \quad (14.34)$$

(0,21)	(0,04)	(0,08)	(0,08)	(0,08)	(0,08)
--------	--------	--------	--------	--------	--------

El estadístico t ADF es el estadístico t para el contraste de la hipótesis de que el coeficiente de la variable ΔInf_{t-1} es igual a cero; es decir, $t = -2,69$. A partir de la Tabla 14.5, el valor crítico al 5 % es $-2,86$. Debido a que el estadístico ADF de $-2,69$ es menos negativo que $-2,86$, el contraste no rechaza la hipótesis nula al nivel de significación del 5 %. En base a la regresión de la Ecuación (14.34), por lo tanto, no puede rechazarse (al nivel de significación del 5 %) la hipótesis nula de que la inflación tiene una raíz unitaria autorregresiva, es decir, que la inflación contiene una tendencia estocástica, frente a la alternativa de que es estacionaria.

La regresión ADF de la Ecuación (14.34) incluye cuatro retardos de la variable ΔInf_t para calcular el estadístico ADF. Cuando el número de retardos se calcula utilizando el criterio AIC, con $0 \leq p \leq 5$, sin embargo, el estimador AIC de la longitud de los retardos, es tres. Cuando se utilizan tres retardos (es decir, cuando ΔInf_{t-1} , ΔInf_{t-2} , y ΔInf_{t-3} se incluyen como regresores), el estadístico ADF es $-2,72$, que es menos negativo que $-2,86$. Por lo tanto, cuando el número de retardos en la regresión ADF se elige mediante el criterio AIC, no se rechaza la hipótesis de que la inflación contiene una tendencia estocástica al nivel de significación del 5 %.

Estos contrastes se realizaron al nivel de significación del 5 %. Sin embargo, al 10 % de nivel de significación, los contrastes rechazan la hipótesis nula de una raíz unitaria: los estadísticos ADF de $-2,69$ (cuatro retardos) y $-2,72$ (tres retardos) son más negativos que el valor crítico de $-2,57$ al 10 %. Por lo tanto los estadísticos ADF pintan un cuadro más bien ambiguo, y quien lleve a cabo la predicción deberá hacer una valoración en base a la información disponible sobre la cuestión de modelizar la inflación con una tendencia estocástica. Claramente, en la Figura 14.1a la inflación muestra cambios en el largo plazo, en consonancia con el modelo de tendencia estocástica. En la práctica, muchos analistas consideran la inflación en EE.UU. como una variable que presenta una tendencia estocástica, y aquí se sigue esta estrategia.

Resolución de los problemas originados por tendencias estocásticas

La manera más fiable de tratar la presencia de una tendencia en una serie consiste en transformar la serie de forma que desaparezca la tendencia. Si la serie tiene una tendencia estocástica, es decir, si la serie tiene una raíz unitaria, entonces la primera diferencia de la serie no tendrá tendencia. Por ejemplo, si Y_t sigue un paseo aleatorio, por lo que $Y_t = \beta_0 + Y_{t-1} + u_t$, entonces, $\Delta Y_t = \beta_0 + u_t$ es estacionaria. Por tanto, mediante la utilización de las primeras diferencias se elimina la tendencia de paseo aleatorio en una serie.

En la práctica, rara vez se puede asegurar que una serie tiene una tendencia estocástica. Recordemos que, como norma general, el hecho de no rechazar la hipótesis nula no significa necesariamente que la hipótesis nula sea verdadera; sino que simplemente significa que no se dispone de suficientes pruebas que permitan concluir que es falsa. Por tanto, cuando no se rechaza la hipótesis nula de una raíz unitaria mediante el contraste ADF, no significa que la serie en realidad *tenga* una raíz unitaria. Por ejemplo, en un modelo AR(1) el verdadero coeficiente β_1 podría estar muy cercano a uno, por ejemplo 0,98, en cuyo caso el contraste ADF tiene poca potencia, es decir, presenta una probabilidad baja de rechazar de forma correcta la hipótesis nula en las muestras con un tamaño muestral como el de nuestra serie de inflación. Incluso aunque no se rechace la hipótesis nula de presencia de raíz unitaria, no significa que la serie tenga una raíz unitaria, y aun podría ser razonable aproximar la verdadera raíz autorregresiva como si fuera igual a 1 y por lo tanto utilizar las diferencias de la serie en lugar de sus niveles⁵.

14.7 Ausencia de estacionariedad II: cambios estructurales

Un segundo tipo de no estacionariedad se presenta cuando la función de regresión poblacional cambia durante el transcurso de la muestra. En economía, esto puede ocurrir por varias razones, tales como la presencia de cambios en la política económica, cambios en la estructura de la economía, o un invento que cambie una industria en concreto. Si ocurren tales «cambios estructurales», o «rupturas», entonces un modelo de regresión que no tenga en cuenta esos cambios puede proporcionar una base engañosa para la inferencia y la predicción.

En esta sección se presentan dos estrategias para verificar la presencia de cambios estructurales en el tiempo de una función de regresión con series temporales. La primera estrategia busca los potenciales cambios estructurales desde el punto de vista de los contrastes de hipótesis y consiste en la contrastación de la existencia de cambios en los coeficientes de regresión mediante estadísticos F . La segunda estrategia busca los potenciales cambios estructurales desde el punto de vista de la predicción: se simula que la muestra termina antes de lo que realmente lo hace y se evalúan las predicciones que hubiesen sido realizadas de esta forma. Los cambios estructurales se detectan cuando la capacidad de predicción es sustancialmente peor de lo esperado.

¿Qué es un cambio estructural?

Los cambios estructurales pueden surgir debido tanto a un cambio discreto en los coeficientes de regresión poblacionales en un momento evidente, como a una evolución gradual de los coeficientes a lo largo de un horizonte temporal más prolongado.

Una fuente de cambios estructurales discretos en los datos macroeconómicos puede provenir de un cambio importante en la política macroeconómica. Por ejemplo, el colapso del sistema de tipos de cambio fijos de Bretton Woods en 1972 produjo un cambio estructural en el comportamiento de las series temporales del tipo de cambio \$/£ que se hace evidente en la Figura 14.2b. Antes de 1972, el tipo de cambio era prácticamente constante, con la excepción de una única devaluación en 1968 por la que el valor oficial de la libra, con respecto al dólar, se redujo. Por el contrario, desde 1972 el tipo de cambio ha fluctuado dentro de un rango muy amplio.

Los cambios estructurales pueden ocurrir asimismo más lentamente a medida que la regresión poblacional evoluciona en el tiempo. Por ejemplo, estos cambios pueden surgir debido a la lenta evolución de la política económica y los cambios consiguientes que provocan en la estructura de la economía. Los métodos para la detección de cambios estructurales descritos en esta sección pueden detectar ambos tipos de cambios estructurales, cambios evidentes y de evolución lenta.

Problemas ocasionados por los cambios estructurales. Si se produce un cambio estructural en la función de regresión poblacional a lo largo de la muestra, entonces la regresión estimada por MCO para la muestra completa estimará una relación que se cumplirá «en promedio», en el sentido de que la estimación

⁵ Para un análisis más amplio de las tendencias estocásticas en variables de series temporales económicas y de los problemas que suponen para el análisis de regresión, véase Stock y Watson (1988).

combinará ambos períodos diferentes. Dependiendo de la ubicación y del tamaño del cambio estructural, la función de regresión «promedio» podría ser muy diferente de la verdadera función de regresión al final de la muestra, y esto llevaría a malas predicciones.

Contrastes de cambio estructural

Una forma de detectar los cambios estructurales es la de contrastar la presencia de variaciones discretas, cambios estructurales, o puntos de ruptura, en los coeficientes de regresión. La forma de hacerlo dependerá de si el momento en el que se sospecha que se produce la ruptura o el cambio estructural (**punto de ruptura**) es conocido.

Contraste de cambio estructural con punto de ruptura conocido. En algunos casos concretos, se puede sospechar que existe un punto de ruptura en un momento conocido. Por ejemplo, si se están estudiando las relaciones comerciales internacionales con datos de la década de 1970, es posible plantear la hipótesis de que existe un cambio estructural en la función de regresión poblacional de interés en 1972, cuando se abandonó el sistema de tipos de cambio fijos de Bretton Woods en favor de la libre flotación de los tipos de cambio.

Si la fecha del hipotético cambio estructural en los coeficientes es conocida, entonces la hipótesis nula de ausencia de cambio estructural se puede contrastar mediante una regresión con variable binaria de interacción del tipo descrito en el Capítulo 8 (Concepto clave 8.4). Por simplicidad, se considera un modelo ARD(1,1), por lo que existe un término independiente, un único retardo de Y_t , y un único retardo de X_t . Sea τ la expresión del periodo en el que se produce el cambio estructural hipotético y sea $D_t(\tau)$ una variable binaria que es igual a 0 antes del periodo del cambio estructural y 1 tras él, por lo que $D_t(\tau) = 0$ si $t \leq \tau$ y $D_t(\tau) = 1$ si $t > \tau$. Por tanto, la regresión que incluye el indicador binario del cambio estructural y todos los términos de interacción es

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t. \quad (14.35)$$

Si no existe cambio estructural, entonces la función de regresión poblacional es la misma en ambas partes de la muestra, por lo que los términos que incluyen la variable binaria de cambio estructural $D_t(\tau)$ no entran en la Ecuación (14.35). Es decir, bajo la hipótesis nula de ausencia de cambio estructural, $\gamma_0 = \gamma_1 = \gamma_2 = 0$. Bajo la hipótesis alternativa de que existe cambio estructural, entonces la función de regresión poblacional es diferente antes y después del punto de ruptura τ , en cuyo caso, al menos una de las γ es distinta de cero. Por tanto, la hipótesis de presencia de cambio estructural puede contrastarse mediante el estadístico F que contrasta la hipótesis de que $\gamma_0 = \gamma_1 = \gamma_2 = 0$ frente a la hipótesis de que al menos uno de los coeficientes γ es distinto de cero. Esto se suele denominar contraste de cambio estructural de Chow con punto de ruptura conocido, así denominado por su inventor, Gregory Chow (1960).

Si existen varios predictores o más retardos, este contraste puede ser ampliado mediante la adición de variables binarias de interacción para todos los regresores y contrastando la hipótesis de que todos los coeficientes de los términos que incluyen $D_t(\tau)$ son iguales a cero.

Este método puede modificarse para comprobar si existe cambio estructural en un subconjunto de coeficientes mediante la inclusión solamente de las variables binarias con interacción para el subconjunto de regresores que interese.

Contraste de cambio estructural con punto de ruptura desconocido. A menudo, el momento en el que puede producirse el posible cambio estructural es desconocido o conocido solamente dentro de un rango. Supóngase, por ejemplo, que se sospecha de que se produjo un cambio estructural entre dos períodos, τ_0 y τ_1 . El contraste de Chow puede ser modificado para abordar esta circunstancia mediante la contrastación de la existencia de cambio estructural en todos los posibles períodos τ que estén entre τ_0 y τ_1 , y a continuación, utilizando el mayor de los estadísticos F resultantes para contrastar un cambio estructural en un momento desconocido. Este contraste de Chow modificado se denomina indistintamente **estadístico de razón de verosimilitud de Quandt (QLR)** (Quandt, 1960) (la denominación que se utiliza en este libro), o *estadístico de supWald*.

Debido a que el estadístico QLR es el mayor de varios estadísticos F , su distribución no es la misma que la de un estadístico F individual. En su lugar, los valores críticos para el estadístico QLR deben ser obtenidos a partir de una distribución particular. Al igual que para el estadístico F , esta distribución depende de la cantidad de restricciones que se contrastan, q , es decir, el número de coeficientes (incluyendo el término independiente) que se permite que cambien, bajo la hipótesis alternativa. La distribución del estadístico QLR depende asimismo de τ_0/T y τ_1/T , es decir, de los últimos períodos, τ_0 y τ_1 , de las submuestras sobre las cuales se calculan los estadísticos F , expresados como una proporción del tamaño total de la muestra.

Para que la aproximación para grandes muestras de la distribución del estadístico QLR sea correcta, los últimos períodos de las submuestras, τ_0 y τ_1 , no pueden estar demasiado cerca del principio o del final de la muestra. Por esta razón, en la práctica, el estadístico QLR se calcula sobre un rango, o subconjunto, «reducido» de la muestra. Es habitual elegir un 15 % de «reducción», es decir, establecer $\tau_0 = 0,15T$ y $\tau_1 = 0,85T$ (redondeando al entero más cercano). Con el 15 % de reducción, se calcula el estadístico F para los puntos de ruptura que se encuentran en el 70 % central de la muestra.

Los valores críticos para el estadístico QLR, calculado con el 15 % de reducción, se ofrecen en la Tabla 14.6. Comparando estos valores críticos con los de la distribución $F_{q,\infty}$ (Tabla 4 del Apéndice) se hace evidente que los valores críticos para los estadísticos QLR son mayores. Esto refleja el hecho de que los

TABLA 14.6 Valores críticos del estadístico QLR con un 15 % de reducción

Número de restricciones (q)	10 %	5 %	1 %
1	7,12	8,68	12,16
2	5,00	5,86	7,78
3	4,09	4,71	6,02
4	3,59	4,09	5,12
5	3,26	3,66	4,53
6	3,02	3,37	4,12
7	2,84	3,15	3,82
8	2,69	2,98	3,57
9	2,58	2,84	3,38
10	2,48	2,71	3,23
11	2,40	2,62	3,09
12	2,33	2,54	2,97
13	2,27	2,46	2,87
14	2,21	2,40	2,78
15	2,16	2,34	2,71
16	2,12	2,29	2,64
17	2,08	2,25	2,58
18	2,05	2,20	2,53
19	2,01	2,17	2,48
20	1,99	2,13	2,43

Estos valores críticos son aplicables cuando $\tau_0 = 0,15T$ y $\tau_1 = 0,85T$ (redondeando al entero más cercano), por lo que el estadístico F se calcula para todos los puntos de ruptura potenciales del 70 % central de la muestra. El número de restricciones q es el número de restricciones contrastadas por cada estadístico F individual. Los valores críticos para otros porcentajes de reducción pueden encontrarse en Andrews (2003).

estadísticos QLR solo se fijan en el mayor de los estadísticos F . Examinando los estadísticos F para todos los posibles puntos de ruptura contemplados, el estadístico QLR tiene muchas oportunidades de rechazar la hipótesis nula, lo que conlleva que los valores críticos del QLR sean mayores que los del estadístico F individual.

Al igual que el contraste de Chow, el contraste QLR puede utilizarse enfocado a la posibilidad de que existan cambios estructurales solamente en algunos de los coeficientes de regresión. Esto se realiza calculando en primer lugar el contraste de Chow para distintos puntos de ruptura utilizando variables binarias con interacciones solamente para las variables de los coeficientes bajo sospecha, calculando a continuación el contraste de Chow máximo en el rango $\tau_0 \leq \tau \leq \tau_1$. Los valores críticos para esta versión del contraste QLR se toman asimismo de la Tabla 14.6, en la que el número de restricciones (q) es el número de restricciones contrastadas por el estadístico F correspondiente.

Si existe un cambio estructural discreto en un instante dentro del rango contrastado, entonces el estadístico QLR rechazará con alta probabilidad en muestras grandes. Por otra parte, el momento en que el estadístico F correspondiente es máximo, $\hat{\tau}$, es una estimación del punto de ruptura τ . Esta estimación es una buena estimación en el sentido de que, bajo ciertas condiciones técnicas, $\hat{\tau}/T \xrightarrow{P} \tau/T$; es decir, la proporción del tramo de la muestra en que se produce el cambio estructural se estima de manera consistente.

El estadístico QLR rechaza asimismo la hipótesis nula con una probabilidad alta en muestras grandes cuando existen varios cambios estructurales discretos o cuando el cambio estructural se presenta en forma de lenta evolución de la función de regresión. Esto significa que el estadístico QLR detecta formas de inestabilidad distintas a un único cambio estructural discreto. Como resultado, si el estadístico QLR rechaza la hipótesis nula, puede significar que existe un cambio estructural único y discreto, que existen varios cambios estructurales discretos, o que hay una evolución lenta de la función de regresión.

El estadístico QLR se recoge en el Concepto clave 14.9.

Advertencia: es probable que no se conozca el punto de ruptura incluso aunque se crea que sí. A veces un experto puede pensar que él, o ella, conoce el momento en el que se produce un posible cambio estructural por lo que puede utilizarse el contraste de Chow en lugar del contraste QLR. Pero si esta

CONCEPTO
CLAVE

14.9

El contraste QLR para la estabilidad de los coeficientes

Sea $F(\tau)$ la expresión del estadístico F para el contraste de la hipótesis de cambio estructural en los coeficientes de regresión en el momento τ ; por ejemplo, en la regresión de la Ecuación (14.35), este es el estadístico F para contrastar la hipótesis nula de que $\gamma_0 = \gamma_1 = \gamma_2 = 0$. El estadístico de contraste QLR (o de sup Wald) es el mayor de los estadísticos dentro del rango $\tau_0 \leq \tau \leq \tau_1$:

$$\text{QLR} = \max [F(\tau_0), F(\tau_0 + 1), \dots, F(\tau_1)]. \quad (14.36)$$

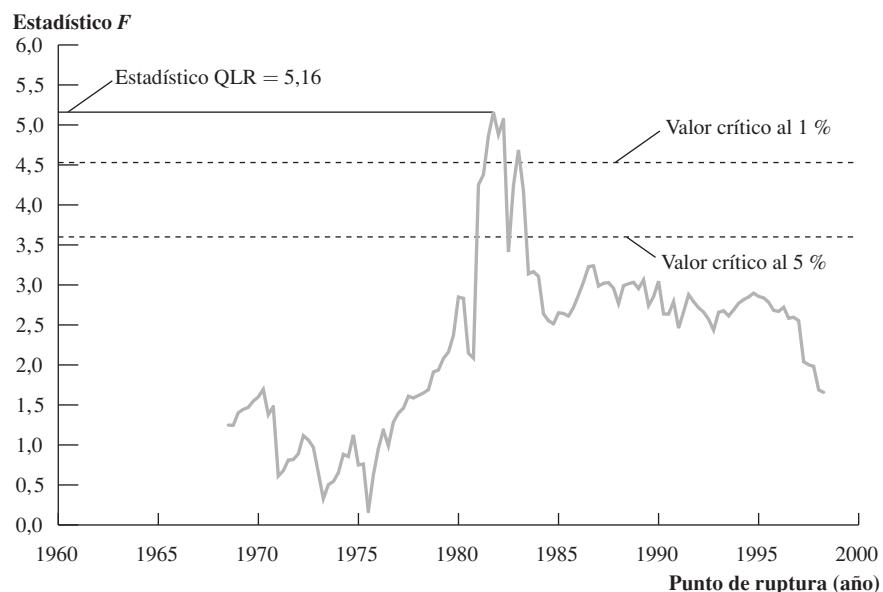
1. Al igual que el estadístico F , el estadístico QLR puede ser utilizado para contrastar la existencia de un cambio estructural en todos o solo en algunos de los coeficientes de regresión.
2. En muestras grandes, la distribución del estadístico QLR bajo la hipótesis nula depende del número de restricciones que se contrasten, q , y de los extremos τ_0 y τ_1 como proporción de T . Los valores críticos están recogidos en la Tabla 14.6 para un 15 % de reducción ($\tau_0 = 0,15T$ y $\tau_1 = 0,85T$, redondeando al entero más cercano).
3. El contraste QLR puede detectar la existencia de un único cambio estructural discreto, varios cambios estructurales discretos, y/o la evolución lenta de la función de regresión.
4. Si existe un cambio estructural evidente en la función de regresión, el periodo en el que se registra el estadístico de Chow mayor es un estimador del punto de ruptura.

información se basa en el conocimiento por parte del experto acerca de la serie que se analiza, entonces, de hecho, este momento fue estimado mediante los datos, aunque de manera informal. La estimación preliminar del punto de ruptura significa que los valores críticos habituales de F no pueden utilizarse para el contraste de Chow para un cambio estructural en ese momento. Por tanto, en estas circunstancias, sigue siendo apropiado utilizar el estadístico QLR.

Aplicación: ¿se ha mantenido estable la curva de Phillips? El contraste QLR proporciona un método para comprobar si la curva de Phillips se mantuvo estable desde 1962 hasta 2004. En concreto, nos centraremos en si hubo cambios en los coeficientes de los valores retardados de la tasa de desempleo y en el término independiente en la especificación ARD (4.4) de la Ecuación (14.17) que contiene cuatro retardos tanto de ΔInf_t como de Desemp_t .

Los estadísticos F de Chow para contrastar la hipótesis de que el término independiente y los coeficientes de Desemp_{t-1} , Desemp_{t-2} , ..., Desemp_{t-4} en la Ecuación (14.17) son constantes frente a la alternativa de que cambian en un momento determinado están representados gráficamente en la Figura 14.5 para cambios estructurales en el 70 % central de la muestra. Por ejemplo, el estadístico F para contrastar un cambio estructural en 1980:I es de 2,85, el valor representado en ese periodo en la figura. Cada estadístico F contrasta cinco restricciones (no existe cambio estructural en el término independiente ni en los cuatro coeficientes de los retardos de la tasa de desempleo), por lo que $q = 5$. El mayor de estos estadísticos F es 5,16, que se produce en 1981:IV; este es el estadístico QLR. La comparación del valor 5,16 con el valor crítico para $q = 5$ de la Tabla 14.6 indica que la hipótesis de que estos coeficientes son estables se rechaza al nivel de significación del 1 % (el valor crítico es 4,53). Por tanto, existe evidencia de que al menos uno de estos cinco coeficientes cambió a lo largo del periodo muestral.

FIGURA 14.5 Estadísticos F para el contraste de cambio estructural en la ecuación (14.17) en diferentes períodos



En un punto de ruptura dado, el estadístico F representado aquí contrasta la hipótesis nula de presencia de cambio estructural en al menos uno de los coeficientes de, Desemp_{t-1} , Desemp_{t-2} , Desemp_{t-3} , Desemp_{t-4} o el término independiente de la Ecuación (14.17). Por ejemplo, el estadístico F que contrasta la presencia de cambio estructural en 1980:I es 2,85. El estadístico QLR es el mayor de estos estadísticos F, que es 5,16. Esto supera el valor crítico al 1 % que es igual a 4,53.

Predicción pseudo fuera de la muestra

La prueba definitiva para un modelo de predicción es su rendimiento fuera la muestra, es decir, su capacidad de predecir en «tiempo real», después de que el modelo haya sido estimado. La **predicción pseudo fuera de la muestra** es un método para simular el rendimiento en tiempo real de un modelo de predicción. La idea de la predicción pseudo fuera de la muestra es sencilla: elegir un periodo cercano al final de la muestra, estimar el modelo de predicción a partir de los datos hasta ese periodo, y más tarde utilizar ese modelo estimado para realizar una predicción. La realización de este ejercicio para varios períodos, cercanos al final de la muestra, da lugar a una serie de pseudo predicciones y de pseudo errores de predicción. Los pseudo errores de predicción pueden ser examinados para comprobar si son representativos de lo que cabría esperar si la relación de predicción fuera estacionaria.

La razón por la que se denomina predicción «pseudo» fuera de la muestra es que no se trata de una verdadera predicción fuera de la muestra. La predicción fuera de la muestra se realiza en tiempo real; es decir, el pronóstico se realiza sin el beneficio de conocer los valores futuros de la serie. En la predicción pseudo fuera de muestra, se simula la predicción en tiempo real mediante el modelo, pero se dispone de los datos «futuros» frente a los que se evalúan los simulados, o pseudo predicciones. La predicción fuera de la muestra imita el proceso de predicción que se produciría en tiempo real, pero sin tener que esperar a que lleguen nuevos datos.

La predicción pseudo fuera de la muestra proporciona a quien pronostica una señal acerca de lo bien que el modelo de predicción ha realizado las predicciones al final de la muestra. Esto puede proporcionar información valiosa, ya sea reforzando la confianza porque el modelo ya se ha comportado bien al pronosticar, o bien sugiriendo que el modelo ya se ha salido de la senda correcta en el pasado reciente. La metodología de la predicción pseudo fuera de la muestra se resume en el Concepto clave 14.10.

Otras utilidades de las predicciones pseudo fuera de la muestra. Una segunda utilidad de la predicción pseudo fuera de la muestra consiste en estimar la RECM. Debido a que la predicción pseudo fuera de la muestra se calcula utilizando solamente los datos previos al momento de la predicción, los errores de predicción pseudo fuera de la muestra reflejan tanto la incertidumbre asociada a los valores futuros del término de error como la incertidumbre que aparece debido a que los coeficientes de regresión son estimados; es decir, los errores de predicción pseudo fuera de la muestra incluyen las dos fuentes de error de la Ecuación (14.21). Por tanto, la desviación típica muestral de los errores de predicción pseudo fuera de la muestra es un estimador de la RECM. Tal y como ya se ha discutido en la Sección 14.4, este estimador de la RECM se puede utilizar para cuantificar la incertidumbre de la predicción y para construir intervalos de predicción.

Una tercera utilidad de la predicción pseudo fuera de la muestra es la de comparar dos o más modelos candidatos a ser utilizados para predicción. Dos modelos que parezcan ajustarse bien a los datos pueden comportarse, sin embargo, de manera muy diferente en un ejercicio de predicción pseudo fuera de la mues-

CONCEPTO
CLAVE

14.10

Predicciones pseudo fuera de la muestra

Las predicciones pseudo fuera de la muestra se calculan siguiendo los siguientes pasos:

1. Elegir un número de observaciones, P , para las que se van a generar las predicciones pseudo fuera de la muestra; por ejemplo, P podría ser el 10 % o el 15 % del tamaño de la muestra. Sea $s = T - P$.
2. Estimar la regresión de predicción con el conjunto de datos reducido para $t = 1, \dots, s$.
3. Calcular la predicción para el primer periodo más allá de esta muestra reducida, $s + 1$; y denominarlo $\tilde{Y}_{s+1|s}$.
4. Calcular el error de predicción, $\tilde{u}_{s+1} = Y_{s+1} - \tilde{Y}_{s+1|s}$.
5. Repetir los pasos 2 a 4 para los períodos restantes, $s = T - P + 1$ hasta $T - 1$ (reestimando la regresión para cada periodo). Las predicciones pseudo fuera de la muestra son $\{\tilde{Y}_{s+1|s}, s = T - P, \dots, T - 1\}$, y los errores de predicción pseudo fuera de la muestra son $\{\tilde{u}_{s+1}, s = T - P, \dots, T - 1\}$.

¿Se puede batir al mercado? Parte II

Quizás haya escuchado alguna vez el consejo de que debe comprarse una acción cuando los dividendos que genera son elevados en relación a su precio. Comprar una acción supone, en efecto, la compra del flujo de los dividendos futuros pagados por esa compañía a cuenta de sus beneficios. Si el flujo de dividendos es inusualmente grande en relación con el precio de las acciones de la compañía, entonces podría considerarse que la empresa está infravalorada. Si ocurre que los dividendos actuales constituyen un indicador de los futuros dividendos, entonces la rentabilidad por dividendo —el cociente entre los dividendos actuales y el precio de la acción— podría predecir una rentabilidad extraordinaria futura de las acciones. Si la rentabilidad por dividendo es elevada, la acción está infravalorada y se podría prever que la rentabilidad suba.

Este razonamiento sugiere un examen mediante modelos autorregresivos de retardos distribuidos de la rentabilidad extraordinaria, en los que la variable predictora sea la rentabilidad por dividendo. Sin embargo, con este método surge una dificultad: la rentabilidad por dividendo es muy persistente e incluso podría ocurrir que contuviese una tendencia estocástica. Utilizando los datos mensuales desde 1960:1 hasta 2002:12 del logaritmo del cociente dividendos-precio para el índice CRSP (los datos están descritos en el Apéndice 14.1), un contraste de raíz unitaria de Dickey-Fuller que incluye un término independiente no rechaza la hipótesis nula de existencia de una raíz unitaria al nivel de significación del 10 %. Como siempre, el hecho de que no se rechace la hipótesis nula no significa que la hipótesis nula sea cierta, pero subraya el hecho de que la rentabilidad por dividendo es un regresor muy persistente. Siguiendo la lógica de la Sección 14.6, este resultado sugiere que debería utilizarse la primera diferencia del logaritmo de la rentabilidad por dividendo como regresor, en lugar del nivel del logaritmo de la rentabilidad por dividendo.

En la Tabla 14.7 se presentan los modelos ARD para la rentabilidad extraordinaria sobre el índice CRSP. En las columnas (1) y (2), la rentabilidad por dividendo aparece en primeras diferencias, y los estadísticos t individuales y los estadísticos F conjuntos no rechazan la hipótesis nula de impredecibilidad. Sin embargo, aunque estas especificaciones son acordes con las recomendaciones para la modelización de la Sección 14.6, no se corresponden con el razonamiento económico del párrafo introductorio, que relaciona la rentabilidad con el *nivel* de la rentabilidad por dividendo. La columna (3) de la Tabla 14.7 por tanto, presenta un modelo ARD(1,1) para la rentabilidad extraordinaria utilizando el logaritmo de la rentabilidad por dividendo, que se estima hasta 1992:12. El estadístico t es igual a 2,25, lo que supera el valor crítico habitual del 5 % que es igual a 1,96. Sin embargo, debido a que el regresor es muy persistente, la distribución de este estadístico t resulta sospechosa y el valor crítico 1,96 puede ser

inadecuado. (el estadístico F de esta regresión no se presenta debido a que no necesariamente tiene una distribución chi-cuadrado, incluso en muestras grandes, debido a la persistencia del regresor).

Una forma de evaluar la aparente predictibilidad hallada en la columna (3) de la Tabla 14.7 consiste en llevar a cabo un análisis de predicción pseudo fuera de la muestra. Haciéndolo para el periodo fuera de la muestra 1993:1-2002:12 se obtiene una raíz del error cuadrático medio de predicción muestral del 4,08 %. Por el contrario, la RECM muestral de que todas las rentabilidades extraordinarias predichas son iguales a cero es del 4,00 %, y la RECM muestral de una «predicción constante» (en la que el modelo de predicción estimado recursivamente incluye solamente un término independiente) es de 3,98 %. La predicción pseudo fuera de la muestra basada en el modelo ARD(1,1) con el logaritmo de la rentabilidad por dividendo es peor que las predicciones en las que no hay predictores!

Esta ausencia de predictibilidad es consistente con la versión fuerte de la hipótesis de los mercados eficientes, que sostiene que toda la información disponible al público está incorporada en los precios de las acciones por lo que las rentabilidades no deberían ser predecibles mediante la información disponible al público (la versión débil se refiere solamente a las previsiones basadas en las rentabilidades pasadas). El mensaje central de que las rentabilidades extraordinarias no son fácilmente predecibles tiene sentido: si lo fueran, los precios de las acciones subirían hasta el punto en que no existiera rentabilidad extraordinaria. La interpretación de resultados como los de la Tabla 14.7 es objeto de un debate acalorado entre los economistas financieros.

Algunos consideran que la ausencia de predictibilidad en las regresiones predictivas constituye una reivindicación de la hipótesis de los mercados eficientes (véase, por ejemplo, Goyal y Welch, 2003). Otros dicen que cuando se analizan las regresiones durante períodos de tiempo más largos y a más largo plazo, y se utilizan herramientas que están específicamente diseñadas para tratar regresores persistentes, aparece evidencia de la existencia de predictibilidad (véase Campbell y Yogo, 2006). Esta predictibilidad pudiera derivarse de un comportamiento económico racional, bajo el cual las actitudes de los inversores hacia el riesgo varían a lo largo del ciclo económico (Campbell, 2003), o pudiera reflejar la «exuberancia irracional» (Shiller, 2005).

Los resultados de la Tabla 14.7 se refieren a rentabilidades mensuales, pero algunos económetras financieros se han centrado en horizontes incluso más cortos. La teoría de la «microestructura del mercado» —los movimientos de las acciones en el mercado minuto a minuto— sugiere que podrían existir períodos fugaces de predictibilidad y que puede ganarse dinero si se es inteligente y ágil. Pero para ello son necesarios unos nervios templados, además de un montón de potencia computacional, así como un equipo de económetras con talento.

(continúa)

TABLA 14.7 Modelos autorregresivos de retardos distribuidos para la rentabilidad mensual extraordinaria de las acciones

Variable dependiente: rentabilidad extraordinaria en el índice CPRS por capitalización ponderado			
	(1)	(2)	(3)
Especificación	ADL (1,1)	ADL (2,2)	ADL (1,1)
Periodo de estimación	1960:1– 2002:12	1960:1– 2002:12	1960:1– 1992:12
Variables explicativas			
Rentabilidad extraordinaria _{t-1}	0,059 (0,158)	0,042 (0,162)	0,078 (0,057)
Rentabilidad extraordinaria _{t-2}		-0,213 (0,193)	
Δln (rentabilidad por dividendo) _t	0,009 (0,157)	-0,012 (0,163)	
Δln (rentabilidad por dividendo) _{t-1}		-0,161 (0,185)	
ln (rentabilidad por dividendo) _{t-2}			0,026 ^a (0,012)
Término independiente	0,0031 (0,0020)	0,0037 (0,0021)	0,090 ^a (0,039)
Estadístico <i>F</i> sobre todos los coeficientes (<i>p</i> -valor)	0,501 (0,0606)	0,843 (0,497)	
<i>R</i> ²	-0,0014	-0,0008	0,0134

Notas: Los datos se describen en el Apéndice 14.1. En las casillas de las filas de las variables explicativas se presentan los coeficientes de los regresores, con los errores estándar entre paréntesis. Las dos últimas filas presentan el estadístico *F* para el contraste de la hipótesis de que todos los coeficientes de la regresión son iguales a cero, con su *p*-valor entre paréntesis, y el *R*² ajustado.

^a|t| > 1,96.

tra. Cuando los modelos son diferentes, por ejemplo, cuando incluyen diferentes predictores, la predicción pseudo fuera de la muestra proporciona un modo apropiado de comparar los dos modelos que se centra en su potencial para proporcionar pronósticos fiables.

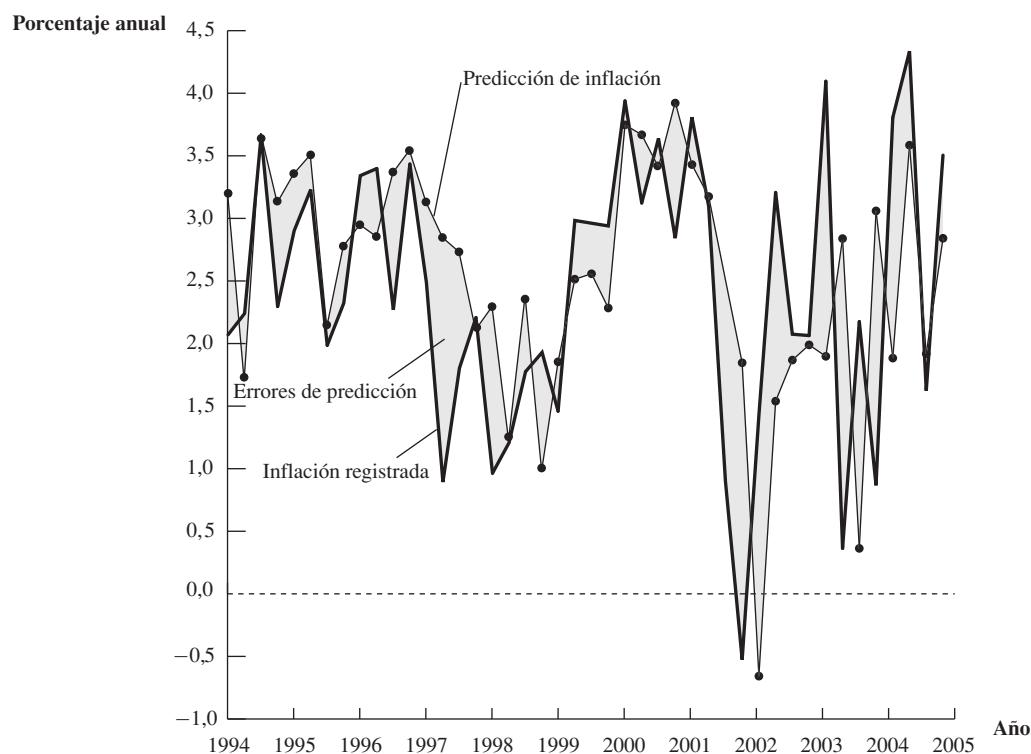
Aplicación: ¿cambió la curva de Phillips durante la década de 1990? Utilizando el estadístico QLR, se rechaza la hipótesis nula de que la curva de Phillips se mantuvo estable frente a la alternativa de presencia de cambio estructural al nivel de significación del 1 % (véase la Figura 14.5). El estadístico *F* máximo se produce en 1981:IV, lo que indica que el cambio estructural se produjo al principio de la década de 1980. Esto indica que alguien que quisiera predecir utilizando la variable de desempleo retardada debería utilizar una muestra para la estimación que comience tras el cambio estructural de 1981:IV. Aun así, quedaría un pregunta por resolver: ¿constituye la curva de Phillips un modelo de predicción estable tras el cambio estructural de 1981:IV?

Si los coeficientes de la curva de Phillips cambiaron en algún momento a lo largo del periodo 1982:I–2004:I, entonces las predicciones pseudo fuera de la muestra calculadas con los datos a partir de 1982:I no serían buenas. Las predicciones de inflación pseudo fuera de la muestra para el periodo 1999:I–2004:IV, calculadas a partir de la curva de Phillips con cuatro retardos y estimada con los datos a partir de 1982:I, están representadas en la Figura 14.6, junto con los verdaderos valores de la inflación. Por ejemplo, la previsión de inflación para 1999:I fue calculada mediante la regresión de la variable ΔInf_t sobre ΔInf_{t-1} , ..., ΔInf_{t-4} , $Desemp_{t-1}$, ..., $Desemp_{t-4}$, con un término independiente utilizando los datos hasta

1998:IV, y a continuación, calculando la predicción $\widehat{\Delta Inf}_{1999:I|1998:IV}$ utilizando estos coeficientes estimados y los datos hasta 1998:IV. La predicción de inflación para 1999:I es por lo tanto $\widehat{Inf}_{1999:I|1998:IV} = Inf_{1998:IV} + \widehat{\Delta Inf}_{1999:I|1998:IV}$. Este mismo procedimiento se repitió en su totalidad utilizando los datos hasta 1999:I para calcular la predicción $\widehat{\Delta Inf}_{1999:II|1999:I}$. Haciendo lo mismo para los 24 trimestres desde 1999:I hasta 2004:IV se generan 24 predicciones pseudo fuera de la muestra, que están representadas en la Figura 14.6. Los errores de predicción pseudo fuera de la muestra son las diferencias entre la verdadera inflación y su predicción pseudo fuera de la muestra, es decir, las diferencias entre las dos líneas de la Figura 14.6. Por ejemplo, en el periodo 2000:IV, la tasa de inflación se redujo en 0,8 puntos porcentuales, pero la predicción pseudo fuera de la muestra para $\Delta Inf_{2000:IV}$ fue de 0,3 puntos porcentuales, por lo que el error de predicción pseudo fuera de la muestra fue $\Delta Inf_{2000:IV} - \widehat{\Delta Inf}_{2000:IV|2000:III} = -0,8 - 0,3 = -1,1$ puntos porcentuales. En otras palabras, alguien que quisiera predecir mediante el modelo ARD(4,4) para la curva de Phillips, estimado hasta el periodo 2000:III, habría pronosticado que la inflación se incrementaría en 0,3 puntos porcentuales en el periodo 2000:IV, cuando en realidad se redujo en 0,8 puntos porcentuales.

¿Cómo se comparan la media y la desviación típica de los errores de predicción pseudo fuera de la muestra con el ajuste del modelo dentro de la muestra? El error estándar de la regresión de la curva de Phillips con cuatro retardos, ajustada con los datos que van desde el periodo 1982:I hasta el periodo 1998:IV, es de 1,30, por lo que en base al ajuste dentro de la muestra podría esperarse que el error de predicción fuera de la muestra tuviera una media igual a cero y una raíz cuadrada del error cuadrático medio de predicción D_t , 1,30. De hecho, a lo largo del periodo de predicción pseudo fuera de la muestra 1999:I-2004:IV, el promedio del error de predicción es de 0,11 y el estadístico t para contrastar la hipótesis de que la media del error de predicción es igual a cero es 0,41; por lo que no se rechaza la hipótesis de que las predicciones tienen una

FIGURA 14.6 Inflación en EE.UU. y predicciones pseudo fuera de muestra



Las predicciones pseudo fuera de muestra, realizadas mediante una curva de Phillips con cuatro retardos de la forma descrita por la Ecuación (14.17), en general siguen la senda de la inflación realmente registrada y son consistentes con un modelo de predicción de la curva de Phillips estable post-1982.

media igual a cero. Además, la RECM a lo largo del periodo de predicción pseudo fuera de la muestra es de 1,32, muy cerca del valor de 1,30 para el error estándar de la regresión a lo largo del periodo 1982:I-1998:IV. Por otra parte, el gráfico de las predicciones y de los errores de predicción de la Figura 14.6 no muestra valores atípicos importantes o discrepancias inusuales.

De acuerdo con el ejercicio de predicción pseudo fuera de la muestra, el comportamiento del modelo de predicción de la curva de Phillips a lo largo del periodo pseudo fuera de la muestra 1999:I-2004:IV fue comparable a su comportamiento a lo largo del periodo 1982:I-1998:IV dentro de la muestra. Aunque el contraste QLR apunta a la inestabilidad de la curva de Phillips en la primera parte de la década de 1980, este análisis pseudo fuera de la muestra sugiere que, tras el cambio estructural en los primeros años de la década de 1980, el modelo de predicción de la curva de Phillips se ha mantenido estable.

Resolución de los problemas originados por cambios estructurales

La mejor manera de adaptarse a un cambio estructural en la función de regresión poblacional depende del origen del cambio estructural. Si el cambio estructural se produce claramente en una fecha específica, este cambio estructural se podrá detectar con alta probabilidad mediante el estadístico QLR, y se podrá estimar el punto de ruptura. Por tanto, la función de regresión podrá estimarse utilizando una variable binaria que indique las dos submuestras asociadas a ese cambio estructural, con interacciones con el resto de los regresores si resulta necesario. Si se produce un cambio estructural en todos los coeficientes, entonces esta regresión tomará la forma de la Ecuación (14.35), donde τ se reemplaza por el punto de ruptura estimado, $\hat{\tau}$, mientras que si el cambio estructural se produce solamente en algunos de los coeficientes, solo aparecerán en la regresión los términos de interacción relevantes. Si existe de hecho un cambio estructural evidente, entonces la inferencia sobre los coeficientes de regresión puede realizarse como de costumbre, por ejemplo, utilizando los valores críticos de la normal para los contrastes de hipótesis basados en los estadísticos t . Además, pueden realizarse las predicciones mediante la función de regresión estimada que es aplicable al final de la muestra.

Si el cambio estructural no es evidente, sino que más bien surge debido a un cambio lento y constante en los parámetros, el remedio resulta más difícil y queda fuera del alcance de este libro⁶.

14.8 Conclusión

En datos de series temporales, en general, una variable está correlacionada entre una observación, o periodo temporal, y la siguiente. Una consecuencia de esta correlación es que la regresión lineal se puede utilizar para predecir los valores futuros de una serie temporal en base a sus valores actuales y pasados. El punto de partida para la regresión de series temporales es una autorregresión, en la que los regresores son los valores retardados de la variable dependiente. Si se dispone de predictores adicionales, entonces se pueden añadir sus retardos a la regresión.

En este capítulo se han analizado algunas cuestiones técnicas que surgen al estimar y utilizar regresiones con datos de series temporales. Una de ellas es la determinación del número de retardos a incluir en las regresiones. Tal y como se trató en la Sección 14.5, si el número de retardos se elige de forma que se minimice el criterio BIC, la longitud estimada de los retardos es consistente con la verdadera longitud de los retardos.

Otra de estas cuestiones se refiere al hecho de si las series analizadas son estacionarias. Si las series son estacionarias, entonces pueden utilizarse los métodos habituales para realizar la inferencia estadística (como la comparación de los estadísticos t con los valores críticos normales), y debido a que la función de regresión poblacional es estable en el tiempo, las regresiones estimadas utilizando datos históricos pueden ser utilizadas de forma fiable para la predicción. Sin embargo, si las series son no estacionarias, entonces las cosas se vuelven más complicadas, y la complicación específica dependerá de la naturaleza de la no estacio-

⁶ Para un estudio más avanzado de la estimación y la contrastación en presencia de cambios estructurales discretos, véase Hansen (2001). Para un análisis más avanzado sobre la estimación y la predicción cuando existen coeficientes que evolucionan lentamente, véase Hamilton (1994, Capítulo 13).

nariedad. Por ejemplo, si las series son no estacionarias debido a que tienen una tendencia estocástica, entonces el estimador MCO y el estadístico t pueden presentar unas distribuciones no estándar (no normales), incluso en muestras grandes, y el comportamiento de las predicciones puede mejorarse mediante la especificación de la regresión en primeras diferencias. En la Sección 14.6 se introdujo un contraste para la detección de este tipo de no estacionariedad, el contraste de raíz unitaria de Dickey-Fuller. Por otra parte, si la función de regresión poblacional presenta un cambio estructural, de ignorar este cambio estructural se obtiene como resultado una versión promedio a partir de la estimación de la función de regresión poblacional, lo que a su vez puede conducir a predicciones sesgadas y/o imprecisas. Los procedimientos para la detección de un cambio estructural en la función de regresión poblacional se introdujeron en la Sección 14.7.

En este capítulo, se han aplicado los métodos de regresión de series temporales a la predicción económica, y no se ha dado una interpretación causal a los coeficientes de estos modelos. No es necesaria una relación causal para la predicción, y al ignorar las interpretaciones causales se permite mayor libertad para la búsqueda de buenos pronósticos. Sin embargo, en algunas aplicaciones, el objetivo no es el de desarrollar un modelo de predicción, sino más bien el de estimar las relaciones causales entre las variables de series temporales, es decir, el de estimar el efecto causal *dinámico* sobre Y a lo largo del tiempo de un cambio en X . Bajo las condiciones adecuadas, los métodos presentados en este capítulo, u otros métodos estrechamente relacionados, se pueden utilizar para estimar los efectos causales dinámicos, y este es el tema del siguiente capítulo.

Resumen

1. No es necesario que los modelos de regresión utilizados para la predicción tengan una interpretación causal.
2. Una variable de serie de temporal generalmente está correlacionada con uno o más de sus valores retardados, es decir, está serialmente correlacionada.
3. Un modelo autorregresivo de orden p es un modelo de regresión lineal múltiple en el que los regresores son los p primeros retardos de la variable dependiente. Los coeficientes de un modelo AR(p) pueden estimarse mediante MCO, y la función de regresión estimada puede utilizarse para predicción. El orden del retardo p se puede estimar utilizando un criterio de información como el BIC.
4. La adición de otras variables y de sus retardos a una autorregresión puede mejorar el resultado de las predicciones. Bajo los supuestos de mínimos cuadrados para la regresión de series temporales (Concepto clave 14.6), los estimadores MCO tienen distribuciones normales en muestras grandes y la inferencia estadística se lleva a cabo del mismo modo que para datos de sección cruzada.
5. Los intervalos de predicción constituyen una forma de cuantificar la incertidumbre de la predicción. Si los errores se distribuyen normalmente, se puede construir un intervalo de predicción aproximadamente del 68 % como la predicción, más o menos una estimación de la raíz del error cuadrático medio de predicción.
6. Una serie que contiene una tendencia estocástica es no estacionaria, lo que viola el segundo supuesto de mínimos cuadrados del Concepto clave 14.6. El estimador MCO y el estadístico t para el coeficiente de un regresor con una tendencia estocástica pueden tener una distribución no estándar, lo que potencialmente puede conducir a estimadores sesgados, predicciones inefficientes e inferencias erróneas. El estadístico ADF se puede utilizar para contrastar la presencia de una tendencia estocástica. Una tendencia estocástica de paseo aleatorio puede eliminarse por medio de las primeras diferencias de la serie.
7. Si la función de regresión poblacional cambia a lo largo del tiempo, las estimaciones MCO que ignoran esta inestabilidad no resultan fiables para la inferencia estadística o la predicción. El estadístico QLR puede utilizarse para contrastar la presencia de un cambio estructural, y, si se halla un cambio estructural discreto, la función de regresión puede reestimarse de forma que incorpore el cambio estructural.
8. Las predicciones pseudo fuera de la muestra pueden ser utilizadas para evaluar la estabilidad del modelo al final de la muestra, para estimar la raíz del error cuadrático medio de predicción, y para comparar diferentes modelos de predicción.

Términos clave

- primer retardo (375)
- j -ésimo retardo (375)
- primera diferencia (375)
- autocorrelación (378)
- correlación serial (378)
- coeficiente de autocorrelación (378)
- autocovarianza j -ésima (378)
- autorregresión (381)
- error de predicción (381)
- raíz del error cuadrático medio de predicción (RECM) (382)
- modelo autorregresivo de orden p [AR(p)] (382)
- modelo autorregresivo de retardos distribuidos (ARD) (386)
- ARD(p, q) (386)
- estacionariedad (387)
- dependencia débil (389)
- estadístico de causalidad de Granger (389)
- contraste de causalidad de Granger (389)
- intervalo de predicción (390)
- criterio de información de Bayes (BIC) (392)
- criterio de información de Akaike (AIC) (393)
- tendencia (395)
- tendencia determinística (395)
- tendencia estocástica (395)
- paseo aleatorio (396)
- paseo aleatorio con deriva (396)
- raíz unitaria (397)
- regresión espuria (398)
- contraste de Dickey-Fuller (399)
- estadístico de Dickey-Fuller (399)
- estadístico de Dickey-Fuller aumentado (ADF) (399)
- punto de ruptura (403)
- estadístico de la razón de verosimilitud de Quandt (QLR) (403)
- predicción pseudo fuera de la muestra (407)

Revisión de conceptos

- 14.1 Observando el gráfico del logaritmo del PIB de Japón en la Figura 14.2c, ¿parece esta serie temporal estacionaria? Explíquelo. Supóngase que se calcula la primera diferencia de esta serie. ¿Podría parecer estacionaria? Explíquelo.
- 14.2 Muchos economistas financieros creen que el modelo de paseo aleatorio resulta una buena descripción del logaritmo del precio de las acciones. Esto implica que las variaciones porcentuales de los precios de las acciones son impredecibles. Un analista financiero afirma que dispone de un nuevo modelo que realiza mejores predicciones que el modelo de paseo aleatorio. Explique cómo se examinaría la afirmación del analista acerca de que su modelo resulta superior.
- 14.3 Un investigador estima un AR(1) con término independiente y halla que la estimación MCO de β_1 es 0,95, con un error estándar de 0,02. ¿Un intervalo de confianza al 95 % incluye $\beta_1 = 1$? Explíquelo.
- 14.4 Supóngase que se sospecha que el término independiente de la Ecuación (14.17) cambió en 1992:I. ¿Cómo se modificaría la ecuación para incorporar este cambio? ¿Cómo se contrastaría un cambio en el término independiente si no se conociera el momento del cambio?

Ejercicios

- 14.1 Considérese el modelo AR(1) $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$. Supóngase que el proceso es estacionario.
 - a) Demuestre que $E(Y_t) = E(Y_{t-1})$. (*Sugerencia:* leer el Concepto clave 14.5).
 - b) Demuestre que $E(Y_t) = \beta_0/(1 - \beta_1)$.
- 14.2 El índice de producción industrial (IPI_t) es una serie temporal mensual que mide la cantidad de productos industriales producidos en un mes determinado. En este problema se utilizan los datos de este índice para los Estados Unidos. Todas las regresiones se estiman para el periodo muestral 1960:1 a 2000:12 (es decir, desde enero de 1960 hasta diciembre de 2000). Sea $Y_t = 1.200 \times \ln(IPI_t/IPI_{t-1})$.
 - a) La predicción de una experta establece que Y_t muestra la variación porcentual mensual en IPI_t , medida en puntos porcentuales anuales. ¿Es esto correcto? ¿Por qué?

- b) Supóngase que la experta estima que el siguiente modelo AR(4) para Y_t :

$$\hat{Y}_t = 1,377 + 0,318Y_{t-1} + 0,123Y_{t-2} + 0,068Y_{t-3} + 0,001Y_{t-4}. \\ (0,062) \quad (0,078) \quad (0,055) \quad (0,068) \quad (0,056)$$

Utilice este AR(4) para predecir el valor de Y_t en enero de 2001 utilizando los siguientes valores de IPI para agosto de 2000 hasta diciembre de 2000:

Periodo	2000:7	2000:8	2000:9	2000:10	2000:11	2000:12
IPI	147,595	148,650	148,973	148,660	148,206	147,300

- c) Preocupada por las posibles fluctuaciones estacionales de la producción, la experta añade Y_{t-12} al modelo autorregresivo. El coeficiente estimado para Y_{t-12} es de $-0,054$ con un error estándar de $0,053$. ¿Es este coeficiente estadísticamente significativo?
- d) Preocupada por un potencial cambio estructural, lleva a cabo un contraste QLR (con un 15 % de reducción) sobre el término constante y los coeficientes AR del modelo AR(4). El estadístico QLR resultante fue de 3,45. ¿Existe evidencia de la existencia de cambio estructural? Explíquelo.
- e) Preocupada por el hecho de que podría haber incluido muy pocos o demasiados retardos en el modelo, la experta estima modelos AR(p) con $p = 1, \dots, 6$ sobre el mismo periodo muestral. La suma de los cuadrados de los residuos de cada uno de los modelos estimados se muestra en la tabla. Utilice el criterio BIC para estimar el número de retardos que deberían incluirse en la auto-regresión. ¿Cambiarían los resultados si se utilizara el criterio AIC?

Orden AR	1	2	3	4	5	6
SR	29,175	28,538	28,393	28,391	28,378	28,317

- 14.3** Utilizando los mismos datos que en el Ejercicio 14.2, un investigador contrasta la presencia de una tendencia estocástica en $\ln(IPI_t)$, utilizando la siguiente regresión:

$$\widehat{\Delta \ln(IPI_t)} = 0,061 + 0,00004t - 0,018 \ln(IPI_{t-1}) + 0,333\Delta \ln(IPI_{t-1}) + 0,162\Delta \ln(IPI_{t-2}) \\ (0,024) \quad (0,00001) \quad (0,007) \quad (0,075) \quad (0,055)$$

donde los errores estándar que aparecen entre paréntesis se calculan utilizando la fórmula válida con homocedasticidad y el regresor « t » es una tendencia temporal lineal.

- a) Utilice el estadístico ADF para contrastar la presencia de una tendencia estocástica (raíz unitaria) en la variable $\ln(IPI)$.
- b) ¿Apoyan estos resultados la especificación utilizada en el Ejercicio 14.2? Explíquelo.
- 14.4** La experta del Ejercicio 14.2 aumenta su modelo AR(4) para el crecimiento de IPI incluyendo cuatro valores retardados de ΔR_t , donde R_t es el tipo de interés de las letras del Tesoro de EE.UU. a tres meses (medido en puntos porcentuales de tasa anual).
- a) El estadístico F de causalidad de Granger para los cuatro retardos de ΔR_t es 2,35. ¿Los tipos de interés ayudan a predecir el crecimiento de la variable IPI ? Explíquelo.
- b) La investigadora realiza asimismo una regresión de ΔR_t sobre un término constante, cuatro retardos de ΔR_t y cuatro retardos del crecimiento de IPI . El estadístico F de causalidad de Granger resultante sobre los cuatro retardos del crecimiento de la variable IPI es 2,87. ¿Ayuda el crecimiento de IPI a predecir los tipos de interés? Explíquelo.
- 14.5** Demuestre los siguientes resultados sobre las medias condicionales, las predicciones y los errores de predicción:

- a)** Sea W una variable aleatoria con media μ_W y varianza σ_W^2 y sea c una constante. Demuestre que $E[(W - c)^2] = \sigma_W^2 + (\mu_W - c)^2$.
- b)** Considérese el problema de la predicción de Y_t a partir de los datos sobre Y_{t-1}, Y_{t-2}, \dots . Sea f_{t-1} la expresión de alguna predicción de Y_t en la que el subíndice $t - 1$ de f_{t-1} indica que la predicción es una función de los datos hasta el periodo $t - 1$. Sea $E[(Y_t - f_{t-1})^2 | Y_{t-1}, Y_{t-2}, \dots]$ el error cuadrático medio condicional de la predicción f_{t-1} condicionado a la observación de Y hasta el periodo $t - 1$. Demuestre que el error cuadrático medio de predicción condicional se minimiza cuando $f_{t-1} = Y_{t|t-1}$, donde $Y_{t|t-1} = E(Y_t | Y_{t-1}, Y_{t-2}, \dots)$. (*Sugerencia:* repase el Ejercicio 2.27).
- c)** Sea u_t el error de la Ecuación (14.14). Demuestre que $\text{cov}(u_t, u_{t-j}) = 0$ para $j \neq 0$. [*Sugerencia:* utilizar la Ecuación (2.27)].
- 14.6** En este ejercicio se lleva a cabo un experimento de Monte Carlo que estudia el fenómeno de la regresión espuria tratado en la Sección 14.6. En un estudio de Monte Carlo, se generan datos artificiales utilizando un ordenador, y luego se utilizan estos datos artificiales para calcular los estadísticos que están siendo analizados. Esto permite calcular la distribución de los estadísticos para los modelos conocidos cuando las expresiones matemáticas de las distribuciones son complicadas (como en este caso) o incluso desconocidas. En este ejercicio, se van a generar datos para dos series, Y_t y X_t , paseos aleatorios independientemente distribuidos. Los pasos concretos son los siguientes:
- Utilice el ordenador para generar una secuencia de $T = 100$ variables aleatorias normales estándar i.i.d. Denomine esas variables como e_1, e_2, \dots, e_{100} . Igualle $Y_1 = e_1$ e $Y_t = Y_{t-1} + e_t$ para $t = 2, 3, \dots, 100$.
 - Utilice el ordenador para generar una nueva secuencia, a_1, a_2, \dots, a_{100} , de $T = 100$ variables aleatorias normales estándar i.i.d. Haciendo $X_1 = a_1$ y $X_t = X_{t-1} + a_t$ para $t = 2, 3, \dots, 100$.
 - Realice una regresión de Y_t sobre una constante y X_t . Calcule el estimador MCO, el R^2 de la regresión, y el estadístico t (válido con homocedasticidad) para el contraste de la hipótesis nula de que β_1 (el coeficiente de X_t) es igual a cero.
- Utilice este algoritmo para contestar a las siguientes preguntas:
- a)** Ejecute el algoritmo de los apartados (I) hasta (III) una vez. Utilice el estadístico t del apartado (III) para contrastar la hipótesis nula de que $\beta_1 = 0$ utilizando los valores críticos habituales al 5 % de 1,96. ¿Cuál es el R^2 de la regresión?
- b)** Repita (a) 1.000 veces, grabando cada valor de R^2 y el estadístico t . Construya un histograma de los R^2 y los estadísticos t . ¿Cuáles son los percentiles 5, 50, y 95 de las distribuciones de los R^2 y los estadísticos t ? ¿En qué proporción de los 1.000 conjuntos de datos simulados el estadístico t supera en valor absoluto el valor 1,96?
- c)** Repita (b) para diferentes números de observaciones, por ejemplo, $T = 50$ y $T = 200$. ¿Se aproxima al 5 % la proporción de veces que se rechaza la hipótesis nula a medida que aumenta el tamaño de la muestra, tal y como debería ocurrir debido a que se han generado X e Y de forma que estén independientemente distribuidas? ¿Parece que esta proporción se aproxime a algún otro límite al aumentar T ? ¿Cuál ese límite?
- 14.7** Supóngase que Y_t sigue el modelo AR(1) estacionario $Y_t = 2,5 + 0,7Y_{t-1} + u_t$, donde u_t es i.i.d. con $E(u_t) = 0$ y $\text{var}(u_t) = 9$.
- a)** Calcule la media y la varianza de Y_t . (*Sugerencia:* véase el Ejercicio 14.1).
- b)** Calcule las dos primeras autocovarianzas de Y_t . (*Pista:* léase el Apéndice 14.2).
- c)** Calcule las dos primeras autocorrelaciones de Y_t .
- d)** Supóngase que $Y_T = 102,3$. Calcule $Y_{T+1|T} = E(Y_{T+1} | Y_T, Y_{T-1})$.
- 14.8** Supóngase que Y_t es el valor mensual de la cantidad de nuevos proyectos de construcción de viviendas iniciados en Estados Unidos. Debido al clima, Y_t presenta un patrón estacional pronunciado; por ejemplo, el número de viviendas iniciadas es bajo en enero y alto en junio. Sea μ_{En} el valor medio del número de viviendas iniciadas en enero y, μ_{Feb} , μ_{Mar} , ..., μ_{Dic} los valores medios del número de viviendas iniciadas en los otros meses. Demuestre que los valores de $\mu_{En}, \mu_{Feb}, \mu_{Mar}, \dots, \mu_{Dic}$, se pueden estimar a partir de la regresión MCO $Y_t = \beta_0 + \beta_1 Feb_t + \beta_2 Mar_t + \dots + \beta_{11} Dic_t + u_t$, don-

de Feb_t es una variable binaria igual a 1 si t es febrero, Mar_t es una variable binaria igual a 1 si t es marzo, y así sucesivamente. Demuestre que, $\beta_0 + \beta_2 = \mu_{Mar}$, etc.

- 14.9** El modelo de media móvil de orden q tiene la forma

$$Y_t = \beta_0 + e_1 + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q},$$

donde e_t es una variable aleatoria incorrelacionada serialmente con media 0 y varianza σ_e^2 .

- a) Demuestre que $E(Y_t) = \beta_0$.
 - b) Demuestre que la varianza de Y_t es $\text{var}(Y_t) = \sigma_e^2(1 + b_1^2 + \cdots + b_q^2)$.
 - c) Demuestre que $p_j = 0$ para $j > q$.
 - d) Supóngase que $q = 1$. Obtenga las autocovarianzas de Y .
- 14.10** Un investigador lleva a cabo un contraste QLR con un 25 % de reducción, y con $q = 5$ restricciones. Conteste a las siguientes preguntas utilizando los valores de la Tabla 14.6 («Valores críticos del estadístico QLR con un 15 % de reducción») y de la Tabla 4 del apéndice («Valores críticos de la distribución $F_{m, \infty}$ »).
- a) El estadístico F QLR es igual a 4,2. ¿Rechazaría el investigador la hipótesis nula al nivel del 5 %?
 - b) El estadístico F QLR es igual a 2,1. ¿Rechazaría el investigador la hipótesis nula al nivel del 5 %?
 - c) El estadístico F QLR es igual a 3,5. ¿Rechazaría el investigador la hipótesis nula al nivel del 5 %?
- 14.11** Supóngase que ΔY_t sigue un modelo AR(1) $\Delta Y_t = \beta_0 + \beta_1 \Delta Y_{t-1} + u_t$.
- a) Demuestre que Y_t sigue un modelo AR(2).
 - b) Obtenga los coeficientes del modelo AR(2) para Y_t como función de β_0 y β_1 .

Ejercicios empíricos

En la página web del libro de texto http://www.pearsonhighered.com/stock_watson, se encuentra el archivo de datos **USMacro_Quarterly** que contiene los datos trimestrales acerca de varias series macroeconómicas de los Estados Unidos; los datos se describen en el archivo **USMacro_Description**. Calcule $Y_t = \ln(GDP_t)^7$, el logaritmo del PIB real, y ΔY_t , la tasa de crecimiento trimestral del PIB. En los Ejercicios empíricos desde el 14.1 hasta el 14.6, utilice el periodo muestral 1955:1-2009:4 (pueden utilizarse los datos anteriores a 1955, si es necesario, como valores iniciales de los retardos en las regresiones).

- E14.1**
- a) Estime la media de ΔY_t .
 - b) Exprese la tasa media de crecimiento en puntos porcentuales de variación anual. [Sugerencia: multiplicar la media muestral de (a) por 400].
 - c) Estime la desviación típica de ΔY_t . Exprese la respuesta en puntos porcentuales de variación anual.
 - d) Estime las cuatro primeras autocorrelaciones de ΔY_t . ¿Cuáles son las unidades de medida de las autocorrelaciones (tasas trimestrales de crecimiento, puntos porcentuales en tasa anual, o no tiene unidades)?
- E14.2**
- a) Estime un modelo AR(1) para ΔY_t . ¿Cuál es el coeficiente estimado del AR(1)? ¿Es el coeficiente estadística y significativamente distinto de cero? Construya un intervalo de confianza al 95 % para el coeficiente poblacional del AR(1).
 - b) Estime un modelo AR(2) para ΔY_t . Es el coeficiente del AR(2) estadística y significativamente distinto de cero? ¿Es preferible este modelo al modelo AR(1)?

⁷ N. del T.: GDP (Gross Domestic Product) es el nombre original de la serie que se corresponde con el PIB (Producto Interior Bruto).

- c) Estime los modelos AR(3) y AR(4). (I) Utilizando la estimación de los modelos AR(1) hasta AR(4), utilice el criterio BIC para elegir el número de retardos del modelo AR. (II) ¿Cuántos retardos se eligen mediante el criterio AIC?
- E14.3** Utilice un estadístico de Dickey-Fuller aumentado para contrastar la presencia de una raíz unitaria en el modelo AR para Y_t . Como alternativa, suponga que Y_t es estacionaria alrededor de una tendencia determinística.
- E14.4** Contraste la existencia de un cambio estructural en el modelo AR(1) utilizando un contraste QLR.
- E14.5**
- a) Sea R_t el tipo de interés de las letras del tesoro a tres meses. Estime un modelo ARD(1,4) para ΔY_t utilizando retardos de ΔR_t como predictores adicionales. Comparando el modelo ARD(1,4) con el modelo AR(1), ¿cuánto cambia el \bar{R}^2 ?
 - b) ¿Es significativo el estadístico F de causalidad de Granger?
 - c) Contraste la presencia de un cambio estructural en el término independiente y en los coeficientes de los valores retardados de ΔR utilizando un contraste QLR. ¿Existe evidencia de cambio estructural?
- E14.6**
- a) Realice predicciones pseudo fuera de la muestra utilizando el modelo AR(1) comenzando en 1989:4 y yendo hasta el final de la muestra. (Es decir, calculando $\widehat{\Delta Y}_{1990:1|1989:4}$, $\widehat{\Delta Y}_{1990:2|1990:1}$, etc.).
 - b) Realice predicciones pseudo fuera de la muestra con el modelo ARD(1,4).
 - c) Realice predicciones pseudo fuera de la muestra utilizando el siguiente modelo «simplista»:
- $$\Delta Y_{t+1/t} = (\Delta Y_t + \Delta Y_{t-1} + \Delta Y_{t-2} + \Delta Y_{t-3})/4.$$
- d) Calcule los errores de predicción pseudo fuera de la muestra para cada uno de los modelos. ¿Son sesgadas algunas de estas predicciones? ¿Qué modelo tiene la menor raíz del error cuadrático medio de predicción (RECM)? ¿Cuál es la cuantía de la RECM (expresada en puntos porcentuales a una tasa anual) del mejor modelo?
- E14.7** Lea los recuadros «¿Se puede batir al mercado? Parte I» y «¿Se puede batir al mercado? Parte II» de este capítulo. A continuación, consulte la página web del libro, en la que se encuentra una versión extendida del conjunto de datos descrito en los recuadros; los datos están en el archivo **Stock_Returns_1931_2002** y se describen en el archivo **Stock_Returns_1931_2002_Description**.
- a) Repita los cálculos presentados en la Tabla 14.3 utilizando las regresiones estimadas para el periodo muestral 1932:1-2002:12.
 - b) Repita los cálculos presentados en la Tabla 14.7 utilizando las regresiones estimadas para el periodo muestral 1932:1-2002:12.
 - c) ¿Es muy persistente la variable $\ln(\text{dividend yield}^8)$? Explíquelo.
 - d) Realice predicciones pseudo fuera de la muestra para la rentabilidad extraordinaria a lo largo del periodo 1983:1-2002:12 mediante regresiones que comiencen en 1932:1.
 - e) ¿Los resultados obtenidos en los apartados (a) hasta (d) sugieren cambios importantes sobre las conclusiones alcanzadas en los recuadros? Explíquelo.

⁸ N. del T.: Rentabilidad por dividendo.

APÉNDICE

14.1 Los datos de series temporales utilizados en el Capítulo 14

Los datos de series temporales macroeconómicas de los Estados Unidos son recogidos y publicados por varias agencias gubernamentales. El Índice de Precios al Consumo de EE.UU. se mide mediante encuestas mensuales y es elaborado por el Bureau of Labor Statistics (BLS). La tasa de desempleo se calcula a partir de la Encuesta Actualizada de Población del BLS (véase el Apéndice 3.1). Los datos trimestrales utilizados aquí fueron calculados promediando los valores mensuales. El tipo de interés de los fondos federales son el promedio mensual de los tipos diarios de acuerdo con lo publicado por la Reserva Federal, y los tipos de cambio dólar/libra son el promedio mensual de los tipos de cambio diarios; ambos referidos al último mes del trimestre. Los datos del PIB japonés se obtuvieron de la OCDE. La variación porcentual diaria en el índice NYSE Composite se calculan como $100\Delta \ln(NYSE_t)$, donde $NYSE_t$ es el valor del índice al cierre diario de la Bolsa de Nueva York; debido a que la bolsa de valores no está abierta los fines de semana y los días festivos, el periodo de análisis es el de un día laborable. Estas y otros miles de series temporales económicas están disponibles gratuitamente en las páginas web mantenidas por diversas agencias de recogida de datos.

Las regresiones de las Tablas 14.3 y 14.7 utilizan datos financieros mensuales de Estados Unidos. Los precios de las acciones (P_t) están medidos para una base amplia (NYSE y AMEX) de índices de precios acciones ponderados por capitalización construidos por el Centre for Research in Security Prices (CRSP). La rentabilidad extraordinaria mensual en términos porcentuales, es $100 \times \{\ln[(P_t + Div_t)/P_{t-1}] - \ln(LT_t)\}$, donde Div_t son los dividendos pagados por las acciones del índice CRSP y LT_t es la rentabilidad bruta (1 más el tipo de interés) de una letra del tesoro a 30 días a lo largo del mes t . El cociente dividendos-precio se construye como los dividendos repartidos a lo largo de los últimos 12 meses, divididos por el precio en el mes actual.

Agradecemos a Motohiro Yogo su ayuda y el hecho de haber proporcionado estos datos.

APÉNDICE

14.2 Estacionariedad en el modelo AR(1)

En este apéndice se demuestra que si $|\beta_1| < 1$ y u_t es estacionario, entonces Y_t es estacionaria. Como se recordará del Concepto clave 14.5 la variable de serie temporal Y_t es estacionaria si la distribución conjunta de $(Y_{s+1}, \dots, Y_{s+T})$ no depende de s , independientemente del valor de T . Para perfilar el argumento, esto se demuestra a continuación de manera formal para $T = 2$ bajo los supuestos simplificadores de que $\beta_0 = 0$ y $\{u_t\}$ son $N(0, \sigma_u^2)$ i.i.d.

El primer paso consiste en obtener una expresión para Y_t en términos de las u_t . Debido a que $\beta_0 = 0$, la Ecuación (14.8) implica que $Y_t = \beta_1 Y_{t-1} + u_t$. Sustituyendo la expresión $Y_{t-1} = \beta_1 Y_{t-2} + u_{t-1}$ dentro de la expresión anterior se obtiene $Y_t = \beta_1(\beta_1 Y_{t-2} + u_{t-1}) + u_t = \beta_1^2 Y_{t-2} + \beta_1 u_{t-1} + u_t$. Continuando con las sustituciones un paso más, se obtiene $Y_t = \beta_1^3 Y_{t-3} + \beta_1^2 u_{t-2} + \beta_1 u_{t-1} + u_t$, y continuando indefinidamente se obtiene:

$$Y_t = u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \beta_1^3 u_{t-3} + \dots + \sum_{i=0}^{\infty} \beta_1^i u_{t-i}. \quad (14.37)$$

Por lo tanto Y_t es una media ponderada de las u_t actuales y pasadas. Debido a que las u_t están distribuidas normalmente y debido a que la media ponderada de variables aleatorias normales es normal (Sección 2.4), Y_{s+1} e Y_{s+2} tienen una distribución normal bivariante. Recordemos de la Sección 2.4, que la distribución normal bivariante está completamente determinada por las medias de las dos variables, sus varianzas, y su covarianza. Por tanto, para demostrar que Y_t es estacionaria, es necesario demostrar que las medias, varianzas y covarianza de (Y_{s+1}, Y_{s+2}) no dependen de s . Se puede utilizar una extensión del argumento que se presenta a continuación para demostrar que la distribución de $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ no depende de s .

Las medias y las varianzas de Y_{s+1} e Y_{s+2} pueden calcularse mediante la Ecuación (14.37), con el subíndice $s+1$ o $s+2$ sustituyendo a t . En primer lugar, debido a que $E(u_t) = 0$ para todo t , $E(Y_t) = E(\sum_{i=0}^{\infty} \beta_1^i u_{t-i}) = \sum_{i=0}^{\infty} \beta_1^i E(u_{t-i}) = 0$, por lo que la media de Y_{s+1} e Y_{s+2} son ambas iguales a cero y en particular no dependen de s . En segundo lugar, $\text{var}(Y_t) = \text{var}(\sum_{i=0}^{\infty} \beta_1^i u_{t-i}) = \sum_{i=0}^{\infty} (\beta_1^i)^2 \text{var}(u_{t-i}) = \sigma_u^2 \sum_{i=0}^{\infty} (\beta_1^i)^2 = \sigma_u^2 / (1 - \beta_1^2)$, donde la última igualdad se deriva del hecho de que si $|a| < 1$, $\sum_{i=0}^{\infty} a^i = 1/(1 - a)$; por lo que $\text{var}(Y_{s+1}) = \text{var}(Y_{s+2}) = \sigma_u^2 / (1 - \beta_1^2)$, que no depende de s , siempre y cuando $|\beta_1| < 1$. Finalmente, debido a que $Y_{s+2} = \beta_1 Y_{s+1} + u_{s+2}$, $\text{cov}(Y_{s+1}, Y_{s+2}) = E(Y_{s+1} Y_{s+2}) = E[Y_{s+1}(\beta_1 Y_{s+1} + u_{s+2})] = \beta_1 \text{var}(Y_{s+1}) + \text{cov}(Y_{s+1}, u_{s+2}) = \beta_1 \text{var}(Y_{s+1}) = \beta_1 \sigma_u^2 / (1 - \beta_1^2)$. La covarianza no depende de s , por lo que Y_{s+1} y Y_{s+2} tienen una distribución de probabilidad conjunta que no depende de s ; es decir, su

distribución conjunta es estacionaria. Si $|\beta_1| \geq 1$, este cálculo se quiebra debido a que la suma infinita de la Ecuación (14.37) no converge y la varianza de Y_t es infinita. Por tanto, Y_t es estacionaria si $|\beta_1| < 1$, pero no si $|\beta_1| \geq 1$.

El argumento anterior se hizo bajo el supuesto de que $\beta_0 = 0$ y u_t está normalmente distribuido. Si $\beta_0 \neq 0$, el argumento es similar, excepto en que las medias de Y_{s+1} e Y_{s+2} son $\beta_0/(1 - \beta_1)$ y la Ecuación (14.37) debe ser modificada para esta media distinta de cero. El supuesto de que u_t es normal i.i.d. puede ser reemplazado por el supuesto de que u_t es estacionaria con una varianza finita, debido a que, por la Ecuación (14.37), Y_t puede expresarse todavía como una función de las u_t actual y pasadas, por lo que la distribución de Y_t es estacionaria, siempre y cuando la distribución de u_t sea estacionaria y la expresión de la suma infinita de la Ecuación (14.37) tenga sentido en el sentido de que converja, lo que requiere que $|\beta_1| < 1$.

APÉNDICE

14.3 Notación del operador de retardos

La notación en este capítulo y los dos siguientes se simplifica considerablemente mediante la adopción de lo que se conoce como notación del operador de retardos. Sea L la expresión para el **operador de retardos**, que tiene la propiedad de transformar una variable en su retardo. Es decir, el operador de retardos L presenta la propiedad de que $LY_t = Y_{t-1}$. Al aplicar el operador de retardos dos veces, se obtiene el segundo retardo: $L^2Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}$. De manera más general, aplicando el operador de retardos j veces, se obtiene el retardo j . En resumen, el operador de retardos tiene la propiedad de que

$$LY_t = Y_{t-1}, L^2Y_t = Y_{t-2}, \text{ y } L^jY_t = Y_{t-j}. \quad (14.38)$$

La notación del operador de retardos permite definir el **polinomio de retardos**, que es un polinomio en el operador de retardos:

$$a(L) = a_0 + a_1L + a_2L^2 + \dots + a_pL^p = \sum_{j=0}^p a_jL^j, \quad (14.39)$$

donde a_0, \dots, a_p son los coeficientes del polinomio de retardos y $L^0 = 1$. El grado del polinomio de retardos $a(L)$ en la Ecuación (14.39) es p . Multiplicando Y_t por $a(L)$ se obtiene la siguiente expresión

$$a(L)Y_t = \left(\sum_{j=0}^p a_jL^j \right) Y_t = \sum_{j=0}^p a_j(L^jY_t) = \sum_{j=0}^p a_jY_{t-j} = a_0Y_t + a_1Y_{t-1} + \dots + a_pY_{t-p}. \quad (14.40)$$

La expresión de la Ecuación (14.40) implica que el modelo AR(p) de la Ecuación (14.14) se puede escribir de forma compacta como

$$a(L)Y_t = \beta_0 + u_t, \quad (14.41)$$

donde $a_0 = 1$ y $a_j = -\beta_j$, para $j = 1, \dots, p$. Del mismo modo, un modelo ARD(p, q) se puede escribir como

$$a(L)Y_t = \beta_0 + c(L)X_{t-1} + u_t, \quad (14.42)$$

donde $a(L)$ es un polinomio de retardos de grado p (con $a_0 = 1$) y $c(L)$ es un polinomio de retardos de grado $q - 1$.

APÉNDICE

14.4 Modelos ARMA

El **modelo autorregresivo de media móvil (ARMA)** extiende el modelo autorregresivo mediante la modelización de u_t como una variable serialmente correlacionada, en concreto como un proceso de retardos distribuidos (o «media móvil») de otro término de error no observable. En la notación del operador de retardos del Apéndice 14.3, sea $u_t = b(L)e_t$, donde $b(L)$ es un polinomio de retardos de grado q con $b_0 = 1$ y e_t es una variable aleatoria no observable serialmente incorrelacionada. Entonces el modelo ARMA(p, q) es

$$a(L)Y_t = \beta_0 + b(L)e_t, \quad (14.43)$$

donde $a(L)$ es un polinomio de retardos de grado p con $a_0 = 1$.

Ambos modelos AR y ARMA se pueden considerar como métodos de aproximar las autocovarianzas de Y_t . La razón de esto es que cualquier serie temporal Y_t estacionaria con varianza finita puede escribirse tanto como un modelo AR como un modelo MA con un término de error serialmente incorrelacionado, aunque puede ocurrir que sea necesario que los modelos AR o MA tengan un orden infinito. El segundo de estos resultados, es decir, que un proceso estacionario se pueda expresar en forma de media móvil, se conoce como teorema de descomposición de Wold y es uno de los resultados fundamentales que sustentan la teoría del análisis de series temporales estacionarias.

A nivel teórico, las familias de modelos AR, MA, y ARMA son igualmente ricas, siempre y cuando los polinomios de retardos tengan un grado suficientemente alto. Sin embargo, en algunos casos las autocovarianzas se puede aproximar de mejor forma mediante un modelo ARMA(p, q) siendo p y q pequeños que mediante un modelo AR puro con solo unos pocos retardos. Sin embargo, a nivel práctico, la estimación de los modelos ARMA resulta más difícil que la estimación de los modelos AR, y los modelos ARMA resultan más difíciles de extender a regresores adicionales de lo que lo son los modelos AR.

APÉNDICE

14.5 Consistencia del estimador de la longitud de los retardos BIC

En este apéndice se recoge el argumento de que el estimador BIC de la longitud de los retardos, \hat{p} , en un modelo autorregresivo es correcto en muestras grandes; es decir, $\Pr(\hat{p} = p) \rightarrow 1$. Esto no es cierto para el estimador AIC, que puede sobreestimar p incluso en muestras grandes.

BIC

Se considera en primer lugar el caso particular en el que el criterio BIC se utiliza para elegir entre distintos modelos autorregresivos con cero, uno, o dos retardos, cuando la verdadera longitud del retardo es uno. Se demuestra a continuación que (i) $\Pr(\hat{p} = 0) \rightarrow 0$ y (ii) $\Pr(\hat{p} = 2) \rightarrow 0$, de lo que se deduce que $\Pr(\hat{p} = 1) \rightarrow 1$. La extensión de este argumento para el caso general de buscar entre $0 \leq p \leq p_{\max}$ implica demostrar que $\Pr(\hat{p} < p) \rightarrow 0$ y que $\Pr(\hat{p} > p) \rightarrow 0$; la estrategia para demostrarlo es la misma que la utilizada en (i) y (ii) a continuación.

Prueba de (i) y (ii)

Prueba de (i). Para elegir $\hat{p} = 0$ debe ocurrir que $\text{BIC}(0) < \text{BIC}(1)$; es decir, $\text{BIC}(0) - \text{BIC}(1) < 0$. Ahora $\text{BIC}(0) - \text{BIC}(1) = [\ln(SR(0)/T) + (\ln T)/T] - [\ln(SR(1)/T) + 2(\ln T)/T] = \ln(SR(0)/T) - \ln(SR(1)/T) - (\ln T)/T$. Ahora $SR(0)/T = [(T-1)/T]s_Y^2 \xrightarrow{p} \sigma_Y^2$, $SR(1)/T \xrightarrow{p} \sigma_u^2$, y $(\ln T)/T \rightarrow 0$; colocando las piezas juntas, $\text{BIC}(0) - \text{BIC}(1) \xrightarrow{p} \ln \sigma_Y^2 - \ln \sigma_u^2 > 0$ porque $\sigma_Y^2 > \sigma_u^2$. Se deduce que $\Pr[\text{BIC}(0) < \text{BIC}(1)] \rightarrow 0$, por lo que $\Pr(\hat{p} = 0) \rightarrow 0$.

Prueba de (ii). Para elegir $\hat{p} = 2$ debe ocurrir que $\text{BIC}(2) < \text{BIC}(1)$ o bien $\text{BIC}(2) - \text{BIC}(1) < 0$. Ahora $T[\text{BIC}(2) - \text{BIC}(1)] = T\{\ln(SR(2)/T) + 3(\ln T)/T] - [\ln(SR(1)/T) + 2(\ln T)/T]\} = T\ln[SR(2)/SR(1)] + \ln T = -T\ln[1 + F/(T-2)] + \ln T$, donde $F = [SR(1) - SR(2)]/[SR(2)(T-2)]$ es el estadístico F válido con homocedasticidad (Ecuación 7.13) para el contraste de la hipótesis nula de que $\beta_2 = 0$ en el modelo AR(2). Si u_t es homocedástico, entonces F tiene una distribución asintótica χ_1^2 ; si no, seguirá alguna otra distribución asintótica. Por tanto, $\Pr[\text{BIC}(2) - \text{BIC}(1) < 0] = \Pr[T[\text{BIC}(2) - \text{BIC}(1)] < 0] = \Pr[-T\ln[1 + F/(T-2)] + (\ln T) < 0] = \Pr[T\ln[1 + F/(T-2)] > \ln T]$. A medida que aumenta T , $T\ln[1 + F/(T-2)] - F \xrightarrow{p} 0$ [una consecuencia de la aproximación logarítmica $\ln(1+a) \cong a$, que se convierte en exacta a medida que $a \rightarrow 0$]. Por lo tanto $\Pr[\text{BIC}(2) - \text{BIC}(1) < 0] \rightarrow \Pr(F > \ln T) \rightarrow 0$, por lo que $\Pr(\hat{p} = 2) \rightarrow 0$.

AIC

En el caso particular de un modelo AR(1) cuando se consideran cero, uno, o dos retardos, (i) es aplicable al criterio AIC si se sustituye el término $\ln T$ por 2, por lo que $\Pr(\hat{p} = 0) \rightarrow 0$. Todos los pasos en la prueba de (ii) para el criterio BIC son asimismo aplicables al criterio AIC, con la modificación de que $\ln T$ es sustituido por 2; por lo que $\Pr[\text{AIC}(2) - \text{AIC}(1) < 0] \rightarrow \Pr(F > 2) \rightarrow 0$. Si u_t es homocedástico, entonces $\Pr(F > 2) \rightarrow \Pr(\chi_1^2 > 2) = 0,16$, por lo que $\Pr(\hat{p} = 2) \rightarrow 0,16$. En general, cuando \hat{p} se elige mediante el criterio AIC, $\Pr(\hat{p} < p) \rightarrow 0$ pero $\Pr(\hat{p} > p)$ tiende a un número positivo, por lo que $\Pr(\hat{p} = p)$ no tiende a 1.

Estimación de efectos causales dinámicos

En la película de 1983 *Entre Pillos Anda el Juego*¹, los personajes de Dan Aykroyd y Eddie Murphy utilizaban información privilegiada acerca de la evolución de la cosecha de naranjas de Florida durante el invierno para hacer fortuna en el mercado de futuros de zumo de naranja concentrado, un mercado para los contratos de compra o venta de grandes cantidades de zumo de naranja concentrado a un precio determinado en una fecha futura. En la vida real, los operadores de futuros de zumo de naranja, de hecho, prestan una gran atención a las condiciones meteorológicas en Florida: el hielo en Florida destruye las naranjas de Florida, la fuente de casi todo el zumo de naranja concentrado congelado producido en los Estados Unidos, por lo que su oferta cae y aumenta el precio. ¿Pero cuánto aumenta exactamente el precio cuando empeora el clima en Florida? ¿La subida de precio se produce en su totalidad de una sola vez, o existen retardos; y si es así, durante cuánto tiempo? Estas son preguntas que en la vida real, los operadores de futuros sobre el zumo de naranja necesitan responder si quieren triunfar.

En este capítulo se aborda el problema de estimar el efecto sobre Y ahora y en el futuro de una variación en X , es decir, el **efecto causal dinámico** sobre Y de una variación en X . ¿Cuál es, por ejemplo, el efecto sobre la senda de evolución temporal de los precios del zumo de naranja de una larga temporada de heladas en Florida? El punto de partida para la modelización y la estimación de los efectos causales dinámicos es el denominado modelo de regresión de retardos distribuidos, en el que se expresa Y_t como una función de los valores actuales y pasados de X_t . La Sección 15.1 presenta el modelo de retardos distribuidos en el contexto de la estimación del efecto a lo largo del tiempo del clima frío en Florida sobre el precio del zumo de naranja concentrado. En la Sección 15.2 se echa un vistazo exhaustivo a lo que, exactamente, significa un efecto causal dinámico.

Una forma de estimar los efectos causales dinámicos consiste en estimar los coeficientes del modelo de retardos distribuidos mediante MCO. Tal y como se analizó en la Sección 15.3, este estimador es consistente si el error de la regresión tiene una media condicional igual a cero dados los valores actuales y pasados de X , una condición que (como en el Capítulo 12) se conoce como exogeneidad. Debido a que los factores determinantes de Y_t omitidos están correlacionados en el tiempo, es decir, debido a que están serialmente correlacionados, el término de error en el modelo de retardos distribuidos pueden presentar correlación serial. Esta posibilidad, a su vez requiere errores estándar «consistentes a heterocedasticidad y autocorrelación» (HAC), el tema de la Sección 15.4.

Una segunda forma de estimar los efectos causales dinámicos, que se trata en la Sección 15.5, consiste en modelizar la correlación serial en el término de error como un modelo autorregresivo y más tarde utilizar este modelo autorregresivo para obtener un modelo autorregresivo de retardos distribuidos (ARD). De forma alternativa, los coeficientes del modelo original de retardos distribuidos pueden estimarse por mínimos cuadrados generalizados (MCG). Sin embargo, tanto los métodos ARD como MCG requieren una versión más fuerte de la exogeneidad que la que se ha utilizado hasta ahora: la exogeneidad estricta, bajo la cual los errores de regresión tienen una media condicional igual a cero, dados los valores pasados, presentes, y futuros de X .

¹ N. del T.: *Trading Places* es el título original con el que se estrenó la película en EE.UU.

La Sección 15.6 proporciona un análisis más completo acerca de la relación entre los precios del zumo de naranja y el clima. En esta aplicación, el clima queda más allá del control humano y por lo tanto es exógeno (aunque, como se analiza en la Sección 15.6, la teoría económica sugiere que no es estrictamente exógeno necesariamente). Debido a que la exogeneidad es necesaria para estimar los efectos causales dinámicos, la Sección 15.7 examina este supuesto en algunos casos particulares tomados de la macroeconomía y las finanzas.

Este capítulo se basa en el contenido de las Secciones 14.1 a 14.4, no obstante con excepción de un subapartado (que puede ser omitido) del análisis empírico en la Sección 15.6, no resulta necesario el material de las Secciones 14.5 a 14.7.

15.1 Un «primer gusto en boca» de los datos del zumo de naranja

Orlando, el centro histórico de la región de cultivo de naranja en Florida, habitualmente es soleado y cálido. No obstante, de vez en cuando hay una ola de frío, y si las temperaturas bajan por debajo de cero durante mucho tiempo, los árboles pierden muchas de sus naranjas. Si la ola de frío es severa, los árboles se congelan. A consecuencia de la helada, la oferta de zumo de naranja concentrado cae y su precio aumenta. Sin embargo, el calendario del aumento de los precios resulta bastante complicado. El concentrado de zumo de naranja es una mercancía de consumo «duradera» o almacenable; es decir, se puede almacenar en estado de congelación, aunque con algunos costes (asociados a llevar a cabo el tratamiento de frío). Por tanto, el precio del zumo de naranja concentrado no solo depende de la oferta actual, sino que asimismo depende de las expectativas acerca de la oferta futura. Una helada en la actualidad significa que la oferta futura de zumo de naranja será baja, pero debido a que el zumo concentrado que actualmente se encuentra almacenado puede ser utilizado con el fin de satisfacer la demanda actual o futura, el precio del zumo concentrado existente aumenta en la actualidad. ¿Pero cuánto aumenta exactamente el precio del zumo concentrado cuando se produce una helada? La respuesta a esta pregunta no solo interesa a los operadores del mercado de zumo de naranja, sino en general, a los economistas interesados en el estudio de las operaciones en los mercados de materias primas modernos. Para saber cuánto varía el precio del zumo de naranja en respuesta a las condiciones meteorológicas, es necesario analizar los datos sobre los precios del zumo de naranja y del clima.

Los datos mensuales sobre el precio del zumo concentrado de naranja, su variación mensual en términos porcentuales, así como las temperaturas registradas en la región de cultivo de naranjas en Florida desde enero de 1950 hasta diciembre de 2000 se representan en la Figura 15.1. El precio, representado en la Figura 15.1a, es una medida del precio real medio de concentrado de zumo de naranja congelado pagado por los mayoristas. Este precio ha sido deflactado mediante el índice general de precios al productor de bienes finales con el fin de eliminar los efectos de la inflación general en los precios. La variación porcentual de los precios representada en la Figura 15.1b es la variación de los precios a lo largo del mes en términos porcentuales. Los datos de las temperaturas registradas representados en la Figura 15.1c es el «índice de heladas»² en el aeropuerto de Orlando, Florida, calculado como la suma de grados Fahrenheit que la temperatura mínima cae por debajo del nivel de congelación del agua³ en un día determinado a lo largo de todos los días del mes; por ejemplo, en noviembre de 1950 la temperatura en el aeropuerto cayó por debajo del nivel de congelación dos veces, el día 25 (31 °F) y el día 29 (29 °F), para un valor total de 4 en el índice de heladas $[(32 - 31) + (32 - 29) = 4]$. (Los datos se describen en más detalle en el Apéndice 15.1). Como puede verse mediante la comparación entre los gráficos de la Figura 15.1, el precio del zumo de naranja concentrado presenta grandes oscilaciones, algunas de las cuales parecen estar asociadas con el clima frío en Florida.

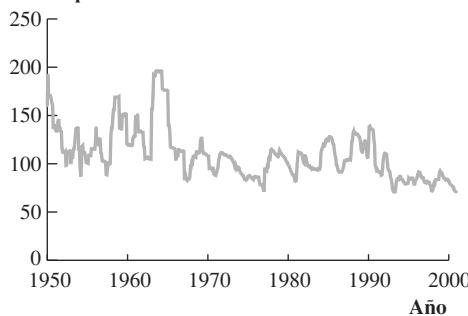
Comenzamos nuestro análisis cuantitativo de la relación entre el precio del zumo de naranja y las condiciones meteorológicas utilizando una regresión que permite estimar la cuantía en la que suben los precios del zumo de naranja cuando las condiciones meteorológicas se vuelven frías. La variable dependiente es la variación porcentual en el precio a lo largo de ese mes $[\%VP_t]$, donde $\%VP_t = 100 \times \Delta \ln(P_t^{ZN})$ y P_t^{ZN} es el precio del zumo de naranja en términos reales]. El regresor es el número de índice de heladas a lo largo de

² N. del T.: Medición combinada de la duración y magnitud de las temperaturas inferiores a 0 °C durante una determinada estación de heladas. Sus unidades son los grados-día.

³ N. del T.: 32 °F.

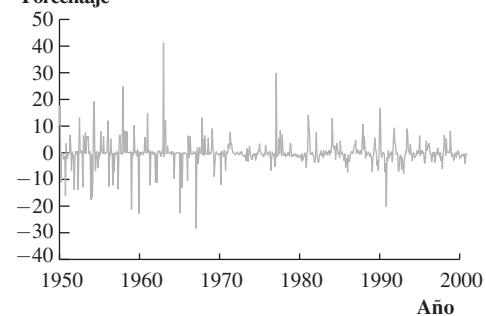
FIGURA 15.1 Precios del zumo de naranja y clima de Florida, 1950-2000

Índice de precios



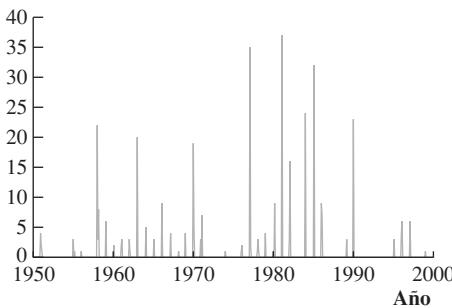
(a) Índice de precios del zumo de naranja concentrado congelado

Porcentaje



(b) Variación porcentual del precio del zumo de naranja concentrado congelado

Índice de heladas



(c) Índice de heladas mensual en Orlando, Florida

Existen grandes variaciones en el precio del concentrado de zumo de naranja. La mayor parte de las grandes variaciones coinciden con heladas en Orlando, lugar en el que se encuentran la mayor parte de los naranjos.

ese mes (IH_t). Esta regresión se estima con los datos mensuales disponibles desde enero del año 1950 hasta diciembre de 2000 (como todas las regresiones de este capítulo), para un total de $T = 612$ observaciones:

$$\widehat{\% VP_t} = -0,40 + 0,47IH_t. \quad (15.1)$$

(0,22) (0,13)

Los errores estándar presentados en esta sección no son los errores estándar MCO habituales, sino los errores estándar que son consistentes a heterocedasticidad y autocorrelación, (HAC) que son los que resultan apropiados cuando el término de error y los regresores están autocorrelacionados. Los errores estándar HAC se analizan en la Sección 15.4, y por ahora se utilizan sin una explicación más detallada.

De acuerdo con esta regresión, un aumento unitario en el índice de heladas a lo largo de un mes, aumenta el precio del zumo de naranja concentrado a lo largo de ese mes en un 0,47 %. En un mes con el índice de heladas igual a 4, como en el caso de noviembre de 1950, se estima que el precio del zumo de naranja concentrado aumentó en un 1,88 % ($4 \times 0,47\% = 1,88\%$), en relación a un mes sin heladas.

Debido a que la regresión de la Ecuación (15.1) incluye solamente una medida contemporánea de las condiciones meteorológicas, no capta los efectos persistentes de la ola de frío sobre el precio del zumo de naranja a lo largo de los meses siguientes. Para captarlos resulta necesario considerar el efecto sobre los precios de los valores de la variable IH , tanto contemporáneos como retardados, lo que a su vez puede hacerse ampliando la regresión de la Ecuación (15.1) con, por ejemplo, los valores retardados de la variable IH de los 6 últimos meses:

$$\begin{aligned} \widehat{\% VP_t} = & -0,65 + 0,47IH_t + 0,14IH_{t-1} + 0,06IH_{t-2} \\ & (0,23) (0,14) (0,08) (0,06) \\ & + 0,07IH_{t-3} + 0,03IH_{t-4} + 0,05IH_{t-5} + 0,05IH_{t-6}. \end{aligned} \quad (15.2)$$

(0,05) (0,03) (0,03) (0,04)

La Ecuación (15.2) es una regresión de retardos distribuidos. El coeficiente de la variable IH_t en la Ecuación (15.2) estima el aumento porcentual de los precios a lo largo del mes en el que se produce la helada; se estima que un aumento unitario del índice de heladas aumenta los precios durante ese mes en un 0,47 %. El coeficiente de primer retardo de la variable IH_t , IH_{t-1} , estima el aumento porcentual de los precios que resulta de un aumento unitario del índice de heladas en el mes anterior, el coeficiente del segundo retardo estima el efecto de un aumento unitario del índice de heladas hace 2 meses, etc. De forma equivalente, el coeficiente del primer retardo de la variable IH estima el efecto de un incremento unitario en la variable IH 1 mes después de que ocurra la helada. Por tanto, los coeficientes estimados de la Ecuación (15.2) son estimaciones del efecto de un aumento unitario de la variable IH_t sobre todos los valores actuales y futuros de la variable $\%VP_t$; es decir, son estimaciones del efecto dinámico de la variable IH_t sobre la variable $\%VP_t$. Por ejemplo, se estima que el índice de heladas igual a 4 en noviembre de 1950 aumentó los precios del zumo de naranja en un 1,88 % a lo largo de noviembre de 1950, en un 0,56 % ($= 4 \times 0,14$) adicional en diciembre de 1950, en un 0,24 % ($= 4 \times 0,06$) adicional en enero de 1951, y así sucesivamente.

15.2 Efectos causales dinámicos

Antes de conocer más acerca de las herramientas disponibles para la estimación de los efectos causales dinámicos, deberíamos parar un momento a pensar acerca de lo que, exactamente, se entiende por un efecto causal dinámico. Tener una idea clara acerca de lo que es un efecto causal dinámico conduce a una comprensión más clara de las condiciones bajo las cuales es posible estimarlo.

Efectos causales y datos de series temporales

En la Sección 1.2 fue definido efecto causal como aquel resultado de un experimento aleatorizado controlado ideal: cuando un horticultor aplica aleatoriamente fertilizante a algunos parcelas de tomate, pero no a otras y mide más tarde el rendimiento, la diferencia esperada en el rendimiento entre las parcelas fertilizadas y sin fertilizar es el efecto causal sobre el rendimiento del tomate de la aplicación de fertilizante. Sin embargo, este concepto de experimento es uno en el que existen varios sujetos individuales (varias parcelas de tomate o varias personas), por lo que los datos son o bien datos de sección cruzada (el rendimiento del tomate al final de la cosecha) o bien datos de panel (los ingresos salariales individuales antes y después de un programa experimental de formación profesional). Al disponer de varios sujetos, es posible tener tanto de grupo de tratamiento como de grupo de control y por lo tanto es posible estimar el efecto causal del tratamiento.

En las aplicaciones con series temporales, esta definición de los efectos causales en términos de un experimento aleatorizado controlado ideal necesita ser modificada. Para concretar, consideremos un problema importante de la macroeconomía: la estimación del efecto de una variación no anticipada en el tipo de interés a corto plazo sobre la actividad económica actual y futura en un país determinado, medida por el PIB. En términos literales, el experimento aleatorizado controlado de la Sección 1.2 implicaría la asignación aleatoria de las diferentes economías a los grupos de tratamiento y de control. Los bancos centrales del grupo de tratamiento aplicarían el tratamiento de una variación aleatoria en los tipos de interés, mientras que en el grupo de control no se aplicarían tales variaciones aleatorias; y debería medirse la actividad económica en ambos grupos (por ejemplo, en términos del PIB) a lo largo de varios años siguientes. ¿Pero y si estamos interesados en la estimación de este efecto para un país específico, por ejemplo, Estados Unidos? Entonces este experimento implicaría la necesidad de disponer de diferentes «clones» de los Estados Unidos como sujetos y asignar a algunas de estas economías clonadas al grupo de tratamiento y a otras al grupo de control. Obviamente, este experimento en un «universo paralelo» es inviable.

En cambio, con datos de series temporales, resulta útil imaginar un experimento aleatorizado controlado que consista en someter al mismo sujeto (por ejemplo, la economía de EE.UU.) a distintos tratamientos (variaciones de los tipos de interés elegidas al azar) para diferentes momentos del tiempo (la década de 1970, de 1980, etc.). En este marco, el mismo sujeto en diferentes momentos desempeña el papel tanto del grupo de tratamiento como del grupo de control: a veces la Fed varía el tipo de interés, mientras que otras

veces no lo hace. Debido a que los datos se recogen a lo largo del tiempo, es posible estimar el efecto causal dinámico, es decir, la senda temporal de los efectos del tratamiento sobre los resultados de interés. Por ejemplo, un aumento no anticipado de los tipos de interés a corto plazo en dos puntos porcentuales, mantenido durante un trimestre, podría en principio tener un efecto insignificante sobre la producción; tras dos trimestres, el crecimiento del PIB podría ralentizarse, alcanzando la mayor desaceleración después de año y medio; y más tarde en los próximos dos años, el crecimiento del PIB podría volver a ser el habitual. Esta trayectoria temporal de los efectos causales es el efecto causal dinámico sobre el crecimiento del PIB de un cambio no anticipado de los tipos de interés.

Como segundo ejemplo, consideremos el efecto causal sobre las variaciones en el precio del zumo de naranja del índice de heladas. Es posible imaginar una serie de experimentos hipotéticos, en los que cada uno de ellos diera lugar a un efecto causal diferente. Un experimento podría consistir en cambiar las condiciones meteorológicas de las plantaciones de naranjos en Florida, manteniendo constantes las condiciones meteorológicas en otros lugares —por ejemplo, manteniendo constantes las condiciones meteorológicas en las plantaciones de pomelos en Texas y en otras regiones con plantaciones de cítricos. Este experimento mediría un efecto parcial, manteniendo constantes el resto de condiciones meteorológicas. Un segundo experimento podría consistir en variar las condiciones meteorológicas en todas las regiones, el «tratamiento» consistiría en la aplicación de las condiciones meteorológicas generales. Si las condiciones meteorológicas estuvieran correlacionadas entre las distintas regiones de cultivos competidores, entonces estos dos efectos causales dinámicos serían diferentes. En este capítulo, se considera el efecto causal del segundo experimento, es decir, el efecto causal de la aplicación de las condiciones meteorológicas generales. Esto equivale a medir el efecto dinámico sobre los precios de una variación en las condiciones meteorológicas de Florida, *sin* mantener constantes las condiciones meteorológicas en otras regiones agrícolas.

Efectos dinámicos y modelo de retardos distribuidos. Debido a que los efectos dinámicos necesariamente ocurren en el tiempo, resulta necesario que el modelo estadístico utilizado para estimar los efectos causales dinámicos incorpore retardos. Para hacerlo, se puede expresar Y_t como un modelo de retardos distribuidos del valor actual de X_t y de sus r valores pasados:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (15.3)$$

donde u_t es un término de error que incluye el error de medida de Y_t y el efecto de los determinantes de Y_t omitidos. El modelo de la Ecuación (15.3) se denomina **modelo de retardos distribuidos** que relaciona X_t , y sus r retardos, con Y_t .

A modo de ilustración de la Ecuación (15.3), consideremos una versión modificada del experimento tomates/fertilizante: debido a que los fertilizantes aplicados hoy pueden permanecer en el terreno durante los próximos años, la horticultora desea determinar el efecto sobre el rendimiento del tomate *a lo largo del tiempo* de la aplicación de fertilizante. En consecuencia, diseña un experimento que dura 3 años y divide sus parcelas aleatoriamente en cuatro grupos: el primer grupo se fertiliza solamente el primer año; el segundo se fertiliza solamente el segundo año; el tercero es fertilizado solamente el tercer año; y el cuarto, el grupo de control, no se fertiliza nunca. Se cultivan anualmente tomates en cada una de las parcelas, y se pesa la cosecha del tercer año. Los tres grupos de tratamiento se expresan mediante las variables binarias, X_{t-2} , X_{t-1} , y X_t , donde t representa el tercer año (el año en que se pesa la cosecha), $X_{t-2} = 1$ si la parcela se encuentra en el primer grupo (fertilizada dos años antes), $X_{t-1} = 1$ si la parcela se fertilizó un año antes, y X_t si la parcela fue fertilizada a lo largo del último año. En el contexto de la Ecuación (15.3) (que es aplicable a una sola parcela), el efecto de ser fertilizada en el último año es β_1 , el efecto de ser fertilizada un año antes es β_2 , y el efecto de ser fertilizada hace 2 años es β_3 . Si el efecto de los fertilizantes es mayor el año en el que se aplica, entonces β_1 sería mayor que β_2 y β_3 .

De manera más general, el coeficiente del valor contemporáneo de X_t , β_1 , es el efecto contemporáneo o inmediato de una variación unitaria en X_t , sobre Y_t . El coeficiente de X_{t-1} , β_2 , es el efecto sobre Y_t de una variación unitaria en X_{t-1} o, equivalentemente, el efecto sobre Y_{t+1} de una variación unitaria en X_t ; es decir, β_2 es el efecto de una variación unitaria en X sobre Y un periodo más tarde. En general, el coeficiente de X_{t-h} es el efecto de una variación unitaria en X sobre Y después de h periodos. El efecto causal dinámico es el efecto de una variación en X_t sobre Y_t , Y_{t+1} , Y_{t+2} , etcétera; es decir, es la secuencia de efectos causales

sobre los valores actuales y los valores futuros de Y . Por lo tanto, en el contexto del modelo de retardos distribuidos de la Ecuación (15.3), el efecto causal dinámico es la secuencia de coeficientes, $\beta_1, \beta_2, \dots, \beta_{r+1}$.

Implicaciones para el análisis empírico de series temporales. Esta formulación de los efectos causales dinámicos en datos de series temporales como el resultado esperado de un experimento en el cual se aplican repetidamente diferentes niveles de tratamiento al mismo sujeto, tiene dos implicaciones para los intentos empíricos de medir el efecto causal dinámico con datos de series de temporales que provienen de observaciones. La primera implicación es que el efecto causal dinámico no debería cambiar a lo largo de la muestra sobre la que se dispone de datos. A su vez, esto está implícito en los datos que son conjuntamente estacionarios (Concepto clave 14.5). Tal y como se trató en la Sección 14.7, la hipótesis de que una función de regresión poblacional es estable en el tiempo puede contrastarse mediante el contraste QLR para un cambio estructural, y es posible estimar el efecto causal dinámico en diferentes submuestras. La segunda implicación es que X no debe estar correlacionada con el término de error, y es en esta implicación en la que ahora se centra el análisis.

Dos tipos de exogeneidad

En la Sección 12.1 se definía como variable «exógena» a una variable que no estaba correlacionada con el término de error de la regresión y como variable «endógena» a una variable que estaba correlacionada con el término de error. Esta terminología sigue la senda de los modelos de varias ecuaciones, en los que una variable «endógena» se determina dentro del modelo mientras que una variable «exógena» se determina fuera del modelo. En términos generales, si han de estimarse los efectos causales dinámicos mediante el modelo de retardos distribuidos de la Ecuación (15.3), las variables explicativas (las X) deben estar incorrelacionadas con el término de error. Por lo tanto X debe ser exógena. Sin embargo, debido a que se trabaja con datos de series de temporales, resulta necesario afinar las definiciones de exogeneidad. De hecho, existen dos conceptos diferentes de exogeneidad que aquí se utilizan.

El primer concepto de exogeneidad es que el término de error tiene una media condicional igual a cero, dados los valores actuales, y todos los anteriores de X_t , es decir, que $E(u_t|X_t, X_{t-1}, X_{t-2}, \dots) = 0$. Esto modifica el supuesto habitual de media condicional para regresión múltiple con datos de sección cruzada (Supuesto #1 del Concepto clave 6.4), que solo requiere que u_t tenga una media condicional igual a cero, dados los regresores incluidos, es decir, que $E(u_t|X_t, X_{t-1}, \dots, X_{t-r}) = 0$. La inclusión de todos los valores retardados de X_t , en la esperanza condicional implica que todos los efectos causales más distantes —todos aquellos efectos causales más allá del retardo r — son iguales a cero. Por tanto, bajo este supuesto, los coeficientes de los r retardos distribuidos de la Ecuación (15.3) constituyen todos los efectos causales dinámicos distintos de cero. Se puede denominar a este supuesto —que $E(u_t|X_t, X_{t-1}, \dots) = 0$ — como *exogeneidad pasada y presente*, pero debido a la similitud de esta definición con la definición de exogeneidad del Capítulo 12, solamente se utiliza el término **exogeneidad**.

El segundo concepto de exogeneidad es que el término de error tiene media igual a cero, dados todos los valores pasados, presentes y futuros de X_t , es decir, que $E(u_t|..., X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \dots) = 0$. Esto se denomina **exogeneidad estricta**; para mayor claridad, puede denominarse asimismo *exogeneidad pasada, presente, y futura*. La razón de introducir el concepto de exogeneidad estricta es que, cuando X es estrictamente exógena, existen estimadores más eficientes de los efectos causales dinámicos que los estimadores MCO de los coeficientes de la regresión de retardos distribuidos de la Ecuación (15.3).

La diferencia entre la exogeneidad (pasada y presente) y la exogeneidad estricta (pasada, presente y futura) es que la exogeneidad estricta incluye los valores futuros de X en la esperanza condicional. Por tanto, la exogeneidad estricta implica exogeneidad, pero no a la inversa. Una manera de entender la diferencia entre ambos conceptos es considerar las implicaciones de estas definiciones para las correlaciones entre X y u . Si X es exógena (pasada y presente), entonces u_t no está correlacionado con los valores actuales y pasados de X_t . Si X es estrictamente exógena, además u_t no está correlacionado con los valores *futuros* de X_t . Por ejemplo, si una variación en Y_t provoca variaciones en los valores *futuros* de X_t , entonces X_t no es estrictamente exógena a pesar de que podría ser exógena (pasada y presente).

A modo de ejemplo, consideremos el experimento hipotético multianual del tomate/fertilizantes, descrito de acuerdo con la Ecuación (15.3). Debido a que el fertilizante se aplica en el experimento hipotético de modo aleatorio, es exógeno. Debido a que el rendimiento del tomate en la actualidad no depende de la cantidad de fertilizante aplicado en el futuro, la serie temporal del fertilizante es asimismo estrictamente exógena.

Como segundo ejemplo, consideremos el ejemplo del precio del zumo de naranja, en el que Y_t es la variación mensual en términos porcentuales de los precios del zumo de naranja y X_t es el índice de heladas en ese mes. Desde la perspectiva de los mercados de zumo de naranja, se puede interpretar que las condiciones meteorológicas —el índice de heladas— es como si estuvieran asignadas aleatoriamente, en el sentido de que las condiciones meteorológicas están fuera del control humano. Si el efecto de la variable IH es lineal y si no tiene ningún efecto sobre los precios tras los primeros r meses, entonces se deduce que las condiciones meteorológicas son exógenas. ¿Pero son las condiciones meteorológicas estrictamente exógenas? Si la media condicional de u_t dado que la variable IH en el futuro es distinta de cero, entonces la variable IH no es estrictamente exógena. La respuesta a esta pregunta requiere reflexionar cuidadosamente acerca de lo que, exactamente, está contenido en u_t . En particular, si los participantes en el mercado de zumo de naranja utilizan las predicciones de la variable IH cuando deciden cuánto van a comprar o vender a un precio dado, entonces los precios del zumo de naranja, y por lo tanto el término de error u_t , podría incorporar información acerca de la variable IH en el futuro que podría ser un predictor útil de la variable IH . Esto significa que u_t estará correlacionado con los valores futuros de la variable IH_t . De acuerdo con esta lógica, debido a que u_t incluye las predicciones sobre las condiciones meteorológicas futuras en Florida, la variable IH sería exógena (pasada y presente) pero no *estrictamente* exógena. La diferencia entre esto y el ejemplo del tomate/fertilizantes es que, mientras que las parcelas de tomate no se ven afectadas por la fertilización futura, los participantes en el mercado de zumo de naranja se ven influidos por las predicciones acerca de las condiciones meteorológicas futuras en Florida. Volveremos a la cuestión de si la variable IH es estrictamente exógena cuando se analicen los datos de los precios del zumo de naranja con más detalle en la Sección 15.6.

Las dos definiciones de exogeneidad se recogen en el Concepto clave 15.1.

**CONCEPTO
CLAVE**

El modelo de retardos distribuidos y la exogeneidad

En el modelo de retardos distribuidos

15.1

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (15.4)$$

existen dos tipos diferentes de exogeneidad, es decir, dos condiciones diferentes de exogeneidad:

Exogeneidad pasada y presente (exogeneidad):

$$E(u_t | X_t, X_{t-1}, X_{t-2}, \dots) = 0; \quad (15.5)$$

Exogeneidad pasada, presente y futura (exogeneidad estricta):

$$E(u_t | \dots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \dots) = 0 \quad (15.6)$$

Si X es estrictamente exógena, es exógena, pero la exogeneidad no implica la exogeneidad estricta.

15.3 Estimación de efectos causales dinámicos con regresores exógenos

Si X es exógena, entonces su efecto causal dinámico sobre Y se puede estimar mediante la estimación MCO de la regresión de retardos distribuidos de la Ecuación (15.4). En esta sección se recogen las condicio-

nes bajo las cuales estos estimadores MCO dan lugar a inferencias estadísticas válidas y se introducen los conceptos de multiplicadores dinámicos y multiplicadores dinámicos acumulativos

Los supuestos del modelo de retardos distribuidos

Los cuatro supuestos del modelo de regresión de retardos distribuidos son similares a los cuatro supuestos del modelo de regresión múltiple para datos de sección cruzada (Concepto clave 6.4), modificados para los datos de series temporales.

El primer supuesto es que X es exógena, lo cual amplia el supuesto de media condicional igual a cero para los datos de sección cruzada a fin de incluir todos los valores retardados de X . Como ya se trató en la Sección 15.2, este supuesto implica que los coeficientes de los r retardos distribuidos de la Ecuación (15.3) constituyen todos los efectos causales dinámicos distintos de cero. En este sentido, la función de regresión poblacional recoge todo el efecto dinámico sobre Y de una variación en X .

El segundo supuesto tiene dos partes: la parte (a) requiere que las variables tengan una distribución estacionaria, y la parte (b) requiere que pasen a ser independientemente distribuidas a medida que el espacio temporal que las separa aumente en gran medida. Este supuesto es el mismo que el supuesto correspondiente para el modelo ARD (el segundo supuesto del Concepto clave 14.6), y el análisis de este supuesto realizado en la Sección 14.4 es del mismo modo aplicable aquí.

El tercer supuesto es que los valores extremos muy grandes son poco probables, esto se concreta de forma matemática mediante el supuesto de que las variables tienen más de ocho momentos finitos y distintos de cero. Esto es más fuerte que el supuesto de momentos de cuarto orden finitos que se utiliza en otros lugares de este libro. Tal y como se discutió en la Sección 15.4, este supuesto más fuerte es el que se utiliza en las matemáticas que se hallan tras el estimador de la varianza HAC.

El cuarto supuesto, que es el mismo que en el modelo de regresión múltiple con datos de sección cruzada, es que no existe multicolinealidad perfecta.

El modelo de regresión de retardos distribuidos y sus supuestos se recogen en el Concepto clave 15.2.

Extensión a X adicionales. El modelo de retardos distribuidos es directamente extensible al caso de varias X : las X adicionales y sus retardos simplemente se incluyen como regresores de la regresión de retardos distribuidos, y los supuestos del Concepto clave 15.2 se modifican de modo que se incluyan estos regresores adicionales. A pesar de que la extensión a varias X es conceptualmente sencilla, la notación se complica, oscureciendo las ideas principales de la estimación y la inferencia en el modelo de retardos distribuidos. Por esta razón, no se trata el caso de varias X de forma explícita en este capítulo, pero se deja como una extensión sencilla del modelo de retardos distribuidos con una única X .

CONCEPTO
CLAVE

15.2

Los supuestos del modelo de retardos distribuidos

El modelo de retardos distribuidos está recogido en el Concepto clave 15.1 [Ecuación (15.4)], donde

1. X es exógena, es decir, $E(u_t | X_t, X_{t-1}, X_{t-2}, \dots) = 0$.
2. (a) Las variables aleatorias Y_t y X_t tienen una distribución estacionaria, y
(b) (Y_t, X_t) y (Y_{t-j}, X_{t-j}) se hacen independientes a medida que j se hace grande.
3. Los valores extremos elevados son poco probables; Y_t y X_t tienen más de ocho momentos finitos distintos de cero.
4. No existe multicolinealidad perfecta.

u_t autocorrelacionados, errores estándar e inferencia

En el modelo de regresión de retardos distribuidos, el término de error u_t puede estar autocorrelacionado; es decir, u_t puede estar correlacionado con sus valores retardados. Esta autocorrelación aparece debido a que, en datos de series temporales, los factores omitidos incluidos en u_t pueden estar por sí solos serial-

mente correlacionados. Por ejemplo, supongamos que la demanda de zumo de naranja depende asimismo de la renta, con lo que un factor que influye en el precio del zumo de naranja es la renta, en concreto, la renta agregada de los consumidores potenciales de zumo de naranja. Por tanto, la renta agregada es una variable omitida de la regresión de retardos distribuidos de las variaciones del precio de zumo de naranja sobre el índice de heladas. Sin embargo, la renta agregada está serialmente correlacionada: la renta disminuye en las recesiones y aumenta en las expansiones. Por tanto, si la renta está correlacionada serialmente, y, debido a que forma parte del término de error, u_t , estará correlacionado serialmente. Este ejemplo es típico: debido a que los determinantes de Y están serialmente correlacionados por sí solos, en general en el modelo de retardos distribuidos u_t , estará correlacionado.

La presencia de autocorrelación en u_t ni afecta a la consistencia de MCO, ni introduce sesgo. Sin embargo, si los errores están autocorrelacionados, en general los errores estándar MCO habituales serán inconsistentes y debe utilizarse una fórmula diferente. Por tanto, la presencia de correlación en los errores es análoga a la presencia de heterocedasticidad: los errores estándar válidos con homocedasticidad son «erróneos» cuando los errores son en realidad heterocedásticos, en el sentido de que la utilización de los errores estándar válidos con homocedasticidad dan como resultado inferencias estadísticas engañosas, cuando los errores son heterocedásticos. Del mismo modo, cuando los errores están serialmente correlacionados, los errores estándar basados en errores i.i.d. son «erróneos», en el sentido de que dan lugar a inferencias estadísticas engañosas. La solución a este problema pasa por utilizar los errores estándar consistentes a heterocedasticidad y autocorrelación (HAC), el tema que se trata en la Sección 15.4.

Multiplicadores dinámicos y multiplicadores dinámicos acumulativos

Otra denominación del efecto causal dinámico es multiplicador dinámico. Los multiplicadores dinámicos acumulativos son los efectos causales acumulados, hasta un retardo dado; por lo que los multiplicadores dinámicos acumulativos miden el efecto acumulado sobre Y de una variación en X .

Multiplicadores dinámicos. El efecto de una variación unitaria en X sobre Y tras h periodos, que es β_{h+1} en la Ecuación (15.4), se denomina **multiplicador dinámico** del periodo h . Por tanto, los multiplicadores dinámicos que vinculan a X con Y son los coeficientes de X_t , y sus retardos en la Ecuación (15.4). Por ejemplo, β_2 es el multiplicador dinámico de un periodo, β_3 es el multiplicador dinámico de dos periodos, y así sucesivamente. En esta terminología, el multiplicador dinámico del periodo cero (o contemporáneo), o **efecto impacto**, es β_1 , el efecto sobre Y de una variación en X en el mismo periodo.

Debido a que los multiplicadores dinámicos se estiman mediante los coeficientes de regresión MCO, sus errores estándar son los errores estándar HAC de los coeficientes de la regresión MCO.

Multiplicadores dinámicos acumulativos. El **multiplicador dinámico acumulativo** del periodo h es el efecto acumulado de una variación unitaria en X sobre Y en los siguientes h periodos. Por tanto, los multiplicadores dinámicos acumulativos son la suma acumulada de los multiplicadores dinámicos. En términos de los coeficientes de la regresión de retardos distribuidos de la Ecuación (15.4), el multiplicador acumulativo del periodo cero es β_1 , el multiplicador acumulativo de un periodo es $\beta_1 + \beta_2$, y el multiplicador dinámico acumulativo de h periodos es $\beta_1, \beta_2, \dots, \beta_{h+1}$. La suma de todos los multiplicadores dinámicos individuales, $\beta_1, \beta_2, \dots, \beta_{r+1}$, es el efecto acumulado de largo plazo sobre Y de una variación en X y se denomina **multiplicador dinámico acumulativo de largo plazo**.

Por ejemplo, consideremos la regresión de la Ecuación (15.2). El efecto inmediato de una variación unitaria del índice de heladas es que el precio del zumo de naranja concentrado aumente en un 0,47 %. El efecto acumulado de una variación del precio para el mes siguiente es la suma del efecto impacto y del efecto dinámico a un mes vista; por lo que el efecto acumulado sobre los precios es el aumento inicial del 0,47 %, más el posterior aumento más pequeño del 0,14 % para un total del 0,61 %. Del mismo modo, el multiplicador dinámico acumulativo de 2 meses es $0,47\% + 0,14\% + 0,06\% = 0,67\%$.

Los multiplicadores dinámicos acumulativos pueden estimarse directamente mediante una modificación de la regresión de retardos distribuidos de la Ecuación (15.4). Esta regresión modificada es

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \delta_3 \Delta X_{t-2} + \dots + \delta_r \Delta X_{t-r+1} + \delta_{r+1} X_{t-r} + u_t. \quad (15.7)$$

Los coeficientes de la ecuación (15.7), $\delta_1, \delta_2, \dots, \delta_{r+1}$ son en realidad los multiplicadores dinámicos acumulativos. Esto puede demostrarse con un poco de álgebra (Ejercicio 15.5), que demuestra que las regresiones poblacionales de las Ecuaciones (15.7) y (15.4) son equivalentes, donde $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$, etc. El coeficiente de X_{t-r} , δ_{r+1} es multiplicador dinámico acumulativo de largo plazo; es decir, $\delta_{r+1} = \beta_1 + \beta_2 + \beta_3 + \dots + \beta_{r+1}$. Por otra parte, los estimadores MCO de los coeficientes de la Ecuación (15.7) son iguales a la suma acumulada correspondiente de los estimadores MCO de la Ecuación (15.4). Por ejemplo, $\hat{\delta}_2 = \hat{\beta}_1 + \hat{\beta}_2$. La principal ventaja de la estimación de los multiplicadores dinámicos acumulativos mediante la especificación de la Ecuación (15.7) es que, debido a que los estimadores MCO de los coeficientes de regresión son estimadores de los multiplicadores dinámicos acumulativos, los errores estándar HAC de los coeficientes de la Ecuación (15.7) son los errores estándar HAC de los multiplicadores dinámicos acumulativos.

15.4 Errores estándar consistentes en presencia de heterocedasticidad y autocorrelación

Si el término de error u_t está autocorrelacionado, entonces los estimadores MCO de los coeficientes son consistentes, pero en general los errores estándar MCO habituales para datos de sección cruzada no lo son. Esto significa que las inferencias estadísticas convencionales —los contrastes de hipótesis y los intervalos de confianza— basadas en los errores estándar MCO habituales, en general, inducen a error. Por ejemplo, los intervalos de confianza construidos como el estimador MCO $\pm 1,96$ errores estándar convencionales no contienen necesariamente el verdadero valor en el 95 % de las muestras repetidas, incluso si el tamaño de la muestra es grande. Esta sección comienza con la obtención de la fórmula correcta para la varianza del estimador MCO con errores autocorrelacionados, más tarde volveremos a los errores estándar consistentes en presencia de heterocedasticidad y autocorrelación (HAC).

Esta sección analiza los errores estándar HAC de la regresión con datos de series temporales. El Capítulo 10 introdujo un tipo de errores estándar HAC, los errores estándar agrupados, que son los apropiados para los datos de panel. Aunque los errores estándar agrupados para datos de panel y los errores estándar HAC para datos de series temporales tienen el mismo objetivo, las diferentes estructuras de datos conducen a fórmulas diferentes. Esta sección es autocontenido, por lo que la lectura del Capítulo 10 no constituye un requisito previo.

Distribución del estimador MCO con errores autocorrelacionados

Por simplicidad, se considera el estimador MCO $\hat{\beta}_1$ del modelo de regresión de retardos distribuidos sin retardos, es decir, el modelo de regresión lineal con un único regresor X_t :

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \quad (15.8)$$

en el que se cumplen los supuestos del Concepto clave 15.2. Esta sección muestra que la varianza de $\hat{\beta}_1$ se puede escribir como el producto de dos términos: la expresión de $\text{var}(\hat{\beta}_1)$, que es aplicable si u_t no está correlacionado serialmente, multiplicada por un factor de corrección que surge de la autocorrelación de u_t , o, de modo más preciso, de la autocorrelación en $(X_t - \bar{X})u_t$.

Como se muestra en el Apéndice 4.3, la fórmula para el estimador MCO del Concepto clave 4.2 se puede reescribir como

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})u_t}{\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2} \quad (15.9)$$

donde la Ecuación (15.9) es la Ecuación (4.30) con un cambio de notación por lo que i y n se sustituyen respectivamente por t y T . Debido a que $\bar{X} \xrightarrow{p} u_X$ y $\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 \xrightarrow{p} \sigma_X^2$, en muestras grandes $\hat{\beta}_1 - \beta_1$ viene dado aproximadamente por

$$\hat{\beta}_1 - \beta_1 \cong \frac{\frac{1}{T} \sum_{t=1}^T (X_t - \mu_X) u_t}{\sigma_X^2} = \frac{\frac{1}{T} \sum_{t=1}^T v_t}{\sigma_X^2} = \frac{\bar{v}}{\sigma_X^2} \quad (15.10)$$

donde $v_t = (X_t - \mu_X) u_t$ y $\bar{v} = \frac{1}{T} \sum_{t=1}^T v_t$. Por tanto

$$\text{var}(\hat{\beta}_1) = \text{var}\left(\frac{\bar{v}}{\sigma_X^2}\right) = \frac{\text{var}(\bar{v})}{(\sigma_X^2)^2} \quad (15.11)$$

Si v_t es i.i.d. —como se supone en el caso de datos de sección cruzada en el Concepto clave 4.3— entonces $\text{var}(\bar{v}) = \text{var}(v_t)/T$ y la fórmula de la varianza de $\hat{\beta}_1$ del Concepto clave 4.4 es aplicable. Sin embargo, si u_t y X_t no están distribuidas de forma independiente a lo largo del tiempo, en general v_t estará serialmente correlacionada, por lo que $\text{var}(\bar{v}) \neq \text{var}(v_t)/T$ y no es aplicable el Concepto clave 4.4. En cambio, si v_t está serialmente correlacionada, la varianza de \bar{v} está dada por

$$\begin{aligned} \text{var}(\bar{v}) &= \text{var}[(v_1 + v_2 + \dots + v_T)/T] \\ &= [\text{var}(v_1) + \text{cov}(v_1, v_2) + \dots + \text{cov}(v_1, v_T) \\ &\quad + \text{cov}(v_2, v_1) + \text{var}(v_2) + \dots + \text{var}(v_T)]/T^2 \\ &= [T \text{var}(v_t) + 2(T-1) \text{cov}(v_t, v_{t-1}) \\ &\quad + 2(T-2) \text{cov}(v_t, v_{t-2}) + \dots + 2\text{cov}(v_t, v_{t-T+1})]/T^2 \\ &= \frac{\sigma_v^2}{T} f_T \end{aligned} \quad (15.12)$$

donde

$$f_T = 1 + 2 \sum_{j=1}^{T-1} \left(\frac{T-j}{T} \right) \rho_j \quad (15.13)$$

donde $\rho_j = \text{corr}(v_t, v_{t-j})$. En muestras grandes, f_T tiende al límite, $f_T \rightarrow f_\infty = 1 + 2 \sum_{j=1}^\infty \rho_j$.

Combinando las expresiones de la Ecuación (15.10) para $\hat{\beta}_1$ y la Ecuación (15.12) para $\text{var}(\bar{v})$ se obtiene la fórmula para la varianza de $\hat{\beta}_1$ cuando v_t está autocorrelacionada:

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{T} \frac{\sigma_v^2}{(\sigma_X^2)^2} \right] f_T \quad (15.14)$$

donde f_T está dada por la Ecuación (15.13).

La Ecuación (15.14) expresa la varianza de $\hat{\beta}_1$ como el producto de dos términos. El primero, entre corchetes, es la fórmula para la varianza de $\hat{\beta}_1$ dada en el Concepto clave 4.4, que es aplicable en ausencia de correlación serial. El segundo es el factor f_T , que ajusta esta fórmula para el caso de presencia de correlación serial. Debido a este factor adicional f_T en la Ecuación (15.14), el error estándar de MCO habitual calculado utilizando la Ecuación (5.4) no es correcto si los errores están serialmente correlacionados: si $v_t = (X_t - \mu_X) u_t$ está serialmente correlacionada, el estimador de la varianza se compensa con el factor f_T .

Errores estándar HAC

Si el factor f_T , que se define en la Ecuación (15.13), fuera conocido, entonces la varianza de $\hat{\beta}_1$ podría estimarse multiplicando el estimador habitual de la varianza para sección cruzada por f_T . Sin embargo, este

factor depende de las autocorrelaciones de v_t que son desconocidas, por lo que debe ser estimado. El estimador de la varianza de $\hat{\beta}_1$ que incorpora este ajuste es consistente tanto si existe heterocedasticidad como si no y tanto si v_t está autocorrelacionada como si no. Por consiguiente, este estimador se denomina el estimador de la varianza de $\hat{\beta}_1$ **consistente a heterocedasticidad y autocorrelación** (HAC), y la raíz cuadrada del estimador HAC de la varianza es el **error estándar HAC** de $\hat{\beta}_1$.

La fórmula HAC de la varianza. El estimador de la varianza de $\hat{\beta}_1$ consistente a heterocedasticidad y autocorrelación es

$$\tilde{\sigma}_{\hat{\beta}_1}^2 = \hat{\sigma}_{\hat{\beta}_1}^2 \hat{f}_T \quad (15.15)$$

donde $\hat{\sigma}_{\hat{\beta}_1}^2$ es el estimador de la varianza de $\hat{\beta}_1$ en ausencia de correlación serial, dado por la Ecuación (5.4), y donde \hat{f}_T es un estimador del factor f_T de la Ecuación (15.13).

La tarea de construir un estimador consistente \hat{f}_T constituye un desafío. Para comprobar por qué, se consideran dos extremos. En un extremo, teniendo en cuenta la fórmula de la Ecuación (15.13), podría parecer natural reemplazar las autocorrelaciones poblacionales ρ_j por las autocorrelaciones muestrales $\tilde{\rho}_j$ [definidas en la Ecuación (14.6)], lo que daría lugar al estimador $1 + 2 \sum_{j=1}^{T-1} \left(\frac{T-j}{T}\right) \tilde{\rho}_j$. Sin embargo, este estimador contiene tantas autocorrelaciones estimadas que es inconsistente. Intuitivamente, debido a que cada una de las autocorrelaciones estimadas contiene un error de estimación, la estimación de tantas autocorrelaciones da lugar a un error de estimación para este estimador de f_T que sigue siendo grande incluso en muestras grandes. En el otro extremo, podría imaginarse la utilización de solamente unas pocas autocorrelaciones muestrales, por ejemplo, solamente la primera autocorrelación muestral, haciendo caso omiso de todas las autocorrelaciones superiores. A pesar de que este estimador elimina el problema de la estimación de demasiadas autocorrelaciones, presenta un problema diferente: es inconsistente debido a que no tiene en cuenta las autocorrelaciones adicionales que aparecen en la Ecuación (15.13). En resumen, la utilización de demasiadas autocorrelaciones muestrales provoca que el estimador tenga una varianza grande, pero si se utilizan demasiado pocas autocorrelaciones se ignoran las autocorrelaciones presentes entre los retardos más elevados, por lo que en ambos casos extremos, el estimador es inconsistente.

Los estimadores de f_T utilizados en la práctica buscan el equilibrio entre estos dos casos extremos escogiendo el número de autocorrelaciones que se incluyen de manera que dependa del tamaño de la muestra T . Si el tamaño de la muestra es pequeño, solamente se utilizan unas pocas autocorrelaciones, pero si el tamaño de la muestra es grande, se incluyen más autocorrelaciones (pero aun así muchas menos que T). En concreto, sea \hat{f}_T dada por

$$\hat{f}_T = 1 + 2 \sum_{j=1}^{m-1} \left(\frac{m-j}{m} \right) \tilde{\rho}_j, \quad (15.16)$$

donde $\tilde{\rho}_j = \sum_{t=j+1}^T \hat{v}_t \hat{v}_{t-j} / \sum_{t=1}^T \hat{v}_t^2$, donde $\hat{v}_t = (X_t - \bar{X}) \hat{u}_t$ (como en la definición de $\hat{\sigma}_{\hat{\beta}_1}^2$). El parámetro m de la Ecuación (15.16) se denomina **parámetro de truncamiento** del estimador HAC debido a que la suma de las autocorrelaciones se acorta, o se trunca, para incluir solamente $m - 1$ autocorrelaciones en lugar de las $T - 1$ autocorrelaciones que aparecen en la fórmula poblacional de la Ecuación (15.13).

Para que \hat{f}_T sea consistente, m debe ser elegido de modo que sea grande en muestras grandes, aunque sigue siendo mucho menor que T . Una pauta que puede servir para escoger m en la práctica consiste en utilizar la fórmula

$$m = 0,75T^{1/3} \quad (15.17)$$

redondeada a un número entero. Esta fórmula, que se basa en el supuesto de que existe es una cantidad moderada de autocorrelación en v_t , proporciona una regla de referencia para determinar m en función del número de observaciones de la regresión⁴.

⁴ La Ecuación (15.17) proporciona la «mejor» elección de m si u_t y X_t son procesos autorregresivos de primer orden con los coeficientes de la primera autocorrelación iguales a 0,5, donde «mejor» significa que es el estimador que minimiza $E(\hat{\sigma}_{\hat{\beta}_1}^2 - \sigma_{\hat{\beta}_1}^2)^2$. La ecuación (15.17) se basa en una fórmula más general obtenida por Andrews [1991, Ecuación (5.3)].

El valor del parámetro de truncamiento m que resulta de la Ecuación (15.17) se puede modificar utilizando el conocimiento que se tenga de la serie que se maneja. Por un lado, si existe una gran correlación serial en v_t , entonces se puede aumentar m más allá del valor que se obtiene a partir de la Ecuación (15.17). Por otro lado, si v_t presenta poca correlación serial, podría disminuirse m . Debido a la ambigüedad asociada a la elección de m , resulta aconsejable probar uno o dos valores alternativos de m para al menos una especificación con el fin de asegurarse de que los resultados no son sensibles a m .

El estimador HAC de la Ecuación (15.15), con el \hat{f}_T dado por la Ecuación (15.16), se denomina **estimador de la varianza de Newey-West**, después de que los económetras Whitney Newey y Kenneth West, fueran los que lo propusieron. Ellos demostraron que, cuando se utiliza de acuerdo a una regla como la de la Ecuación (15.17), bajo supuestos generales, este estimador es un estimador consistente de la varianza de $\hat{\beta}_1$ (Newey y West, 1987). Sus demostraciones (y las de Andrews, 1991) suponen que v_t tiene más de cuatro momentos, lo que a su vez está implícito en que X_t y u_t tengan más de ocho momentos, y esta la razón de que el tercer supuesto del Concepto clave 15.2 sea que X_t , y u_t tengan más de ocho momentos.

Otros Estimadores HAC. El estimador de la varianza de Newey-West no es el único estimador HAC. Por ejemplo, las ponderaciones $(m - j)/m$ en la Ecuación (15.16) se pueden reemplazar por otras ponderaciones diferentes. Si se utilizan diferentes ponderaciones, la regla para la elección del parámetro de truncamiento de la Ecuación (15.17) ya no es aplicable y debe utilizarse una regla diferente en su lugar, desarrollada para esas ponderaciones. El análisis de los estimadores HAC con distintas ponderaciones queda fuera del alcance de este libro. Para obtener más información sobre este tema, véase Hayashi (2000, Sección 6.6).

Extensión para la regresión múltiple. Todas las cuestiones abordadas en esta sección se pueden generalizar al modelo de regresión de retardos distribuidos del Concepto clave 15.1 con varios retardos y, en general, al modelo de regresión múltiple con errores serialmente correlacionados. En particular, si el término de error está serialmente correlacionado, los errores estándar MCO habituales no constituyen una base fiable para la inferencia y deben utilizarse en su lugar los errores estándar HAC. Si el estimador de la varianza HAC utilizado es el estimador de Newey-West [el estimador de la varianza HAC basado en las ponderaciones $(m - j)/m$], entonces el parámetro de truncamiento m puede ser elegido de acuerdo con la regla de la Ecuación (15.17) tanto si existe un único regresor como si existen varias variables explicativas. La fórmula para los errores estándar HAC en regresión múltiple está incorporada en el *software* moderno de regresión diseñado para ser utilizado con datos de series temporales. Debido a que esta fórmula incluye álgebra matricial, se omite y en su lugar se remite al lector a Hayashi (2000, Sección 6.6) para los detalles matemáticos.

Los errores estándar HAC se recogen en el Concepto clave 15.3.

CONCEPTO
CLAVE

15.3

Errores estándar HAC

El problema: El término de error u_t en el modelo de regresión de retardos distribuidos del Concepto clave 15.1 puede estar serialmente correlacionado. Si es así, los estimadores MCO de los coeficientes son consistentes, pero en general los errores estándar MCO habituales no lo son, dando lugar a contrastes de hipótesis e intervalos de confianza erróneos.

La solución: Los errores estándar deberían calcularse a partir del estimador de la varianza consistente a heterocedasticidad y autocorrelación (HAC). El estimador HAC implica la estimación de $m - 1$ autocovarianzas, así como de la varianza; en el caso de un único regresor, las fórmulas relevantes están recogidas por las Ecuaciones (15.15) y (15.16).

En la práctica, la utilización de los errores estándar HAC implica la elección del parámetro de truncamiento m . Para ello, se utiliza la fórmula de la Ecuación (15.17) como punto de referencia, y a continuación, se aumenta o se disminuye m , dependiendo de si los regresores y los errores presentan una correlación serial elevada o baja.

15.5 Estimación de efectos causales dinámicos con regresores estrictamente exógenos

Cuando X_t es estrictamente exógena, se dispone de dos estimadores alternativos para los efectos causales dinámicos. El primero de estos estimadores incluye la estimación de un modelo autorregresivo de retardos distribuidos (ARD) en lugar de un modelo de retardos distribuidos y el cálculo de los multiplicadores dinámicos a partir de los coeficientes ARD estimados. Este método puede implicar la estimación de menos coeficientes que la estimación MCO del modelo de retardos distribuidos, disminuyendo potencialmente con ello el error de estimación. El segundo método consiste en estimar los coeficientes del modelo de retardos distribuidos, utilizando **mínimos cuadrados generalizados (MCG)** en lugar de MCO. Aunque se estima el mismo número de coeficientes en el modelo de retardos distribuidos estimando tanto por MCG como por MCO, el estimador MCG tiene una varianza menor. Por simplicidad de la exposición, se presentan y analizan inicialmente estos dos métodos de estimación en el contexto de un modelo de retardos distribuidos con un único retardo y errores AR(1). Sin embargo, las ventajas potenciales de estos dos estimadores son mayores, cuando aparecen muchos retardos en el modelo de retardos distribuidos, por lo que estos estimadores son extensibles al modelo general de retardos distribuidos con errores autorregresivos de orden superior.

El modelo de retardos distribuidos con errores AR(1)

Supongamos que el efecto causal sobre Y de una variación en X tiene solamente una duración de dos períodos; es decir, tiene un efecto impacto inicial β_1 y un efecto en el periodo siguiente de β_2 , pero ningún efecto posterior. Por tanto el modelo de regresión de retardos distribuidos adecuado es el modelo de retardos distribuidos únicamente con los valores actual y pasado de X_{t-1} :

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t. \quad (15.18)$$

Tal y como se estudió en la Sección 15.2, en general, el término de error u_t de la Ecuación (15.18) está serialmente correlacionado. Una consecuencia de esta correlación serial es que, si los coeficientes de los retardos distribuidos se estiman por MCO, la inferencia basada en los errores estándar MCO habituales puede ser engañosa. Por esta razón, las Secciones 15.3 y 15.4 hicieron hincapié en la utilización de los errores estándar HAC cuando los parámetros β_1 y β_2 de la Ecuación (15.18) se estiman por MCO.

En esta sección, se adopta un método diferente acerca de la correlación serial en u_t . Este método, que es posible si X_t es estrictamente exógena, implica la adopción de un modelo autorregresivo para la correlación serial en u_t , y a continuación la utilización de este modelo AR para la obtención de algunos estimadores que pueden ser más eficientes que el estimador MCO en el modelo de retardos distribuidos.

En concreto, supongamos que u_t sigue el modelo AR(1)

$$u_t = \phi_1 u_{t-1} + \tilde{u}_t, \quad (15.19)$$

donde ϕ_1 es el parámetro autorregresivo, \tilde{u}_t no está serialmente correlacionado, y no es necesario el término independiente porque $E(u_t) = 0$. Las Ecuaciones (15.18) y (15.19) implican que el modelo de retardos distribuidos con un error serialmente correlacionado puede reescribirse como un modelo autorregresivo de retardos distribuidos con un error serialmente incorrelacionado. Para ello, se retarda cada lado de la Ecuación (15.18) y se resta ϕ_1 multiplicado por este retardo a cada uno de los lados:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} &= (\beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t) - \phi_1(\beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_{t-1}) \\ &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} - \phi_1 \beta_0 - \phi_1 \beta_1 X_{t-1} - \phi_1 \beta_2 X_{t-2} + \tilde{u}_t, \end{aligned} \quad (15.20)$$

donde la segunda igualdad utiliza el hecho de que $\tilde{u}_t = u_t - \phi_1 u_{t-1}$. Reagrupando los términos de la Ecuación (15.20), se obtiene que

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \tilde{u}_t, \quad (15.21)$$

donde

$$\alpha_0 = \beta_0(1 - \phi_1), \delta_0 = \beta_1, \delta_1 = \beta_2 - \phi_1\beta_1, \text{ y } \delta_2 = -\phi_1\beta_2, \quad (15.22)$$

donde β_0 , β_1 , y β_2 y son los coeficientes de la Ecuación (15.18) y ϕ_1 es el coeficiente de autocorrelación de la Ecuación (15.19).

La Ecuación (15.21) es un modelo ARD que incluye un valor contemporáneo de X y dos de sus retardos. Nos referiremos a la Ecuación (15.21) como la representación ARD del modelo de retardos distribuidos con errores autorregresivos dado por las Ecuaciones (15.18) y (15.19).

Los términos de la Ecuación (15.20) se pueden reorganizar de manera diferente para obtener una expresión que resulta equivalente a las Ecuaciones (15.21) y (15.22). Sea $\tilde{Y}_t = Y_t - \phi_1 Y_{t-1}$ la **cuasi diferencia** de Y_t («cuasi» porque no es la primera diferencia, la diferencia entre Y_t e Y_{t-1} ; sino que en vez de eso, es la diferencia entre Y_t y $\phi_1 Y_{t-1}$). Del mismo modo, sea $\tilde{X}_t = X_t - \phi_1 X_{t-1}$ la cuasi diferencia de X_t . Por tanto la Ecuación (15.20) se puede escribir como

$$\tilde{Y}_t = \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \tilde{u}_t. \quad (15.23)$$

Nos referiremos a la Ecuación (15.23) como la representación en cuasi diferencias del modelo de retardos distribuidos con errores autorregresivos dado en las Ecuaciones (15.18) y (15.19).

El modelo ARD de la Ecuación (15.21) [con las restricciones en los parámetros de la Ecuación (15.22)] y el modelo de las cuasi diferencias de la Ecuación (15.23) son equivalentes. En ambos modelos, el término de error, \tilde{u}_t , está serialmente incorrelacionado. Sin embargo, las dos representaciones sugieren diferentes estrategias para la estimación. Pero antes de discutir esas estrategias, prestaremos atención a los supuestos bajo los cuales dan lugar a estimadores consistentes de los multiplicadores dinámicos, β_1 y β_2 .

El supuesto de media condicional igual a cero en los modelos ARD(1,2) y de las cuasi diferencias. Debido a que las Ecuaciones (15.21) [con las restricciones de la Ecuación (15.22)] y (15.23) son equivalentes, las condiciones para su estimación son las mismas, por lo que por conveniencia se considera la Ecuación (15.23).

El modelo de las cuasi diferencias de la Ecuación (15.23) es un modelo de retardos distribuidos que incluye las variables cuasi diferenciadas con un error que no presenta correlación serial. En consecuencia, las condiciones para la estimación MCO de los coeficientes de la Ecuación (15.23) son los supuestos de mínimos cuadrados para el modelo de retardos distribuidos del Concepto clave 15.2, expresados en términos de \tilde{u}_t y \tilde{X}_t . El supuesto fundamental aquí es el primero, el cual, aplicado a la Ecuación (15.23), consiste en que \tilde{X}_t sea exógena; es decir,

$$E(\tilde{u}_t | \tilde{X}_t, \tilde{X}_{t-1}, \dots) = 0, \quad (15.24)$$

en la que permitiendo que la esperanza condicional dependa de retardos distantes de \tilde{X}_t se asegura que ningún retardo adicional de \tilde{X}_t , aparte de los que aparecen en la Ecuación (15.23), entra en la función de regresión poblacional.

Debido a que $\tilde{X}_t = X_t - \phi_1 X_{t-1}$, por lo que $X_t = \tilde{X}_t + \phi_1 X_{t-1}$, condicionar a \tilde{X}_t y a todos sus retardos es equivalente a condicionar a X_t , y a todos sus retardos. Por tanto, la condición de la esperanza condicionada de la Ecuación (15.24) es equivalente a la condición de que $E(\tilde{u}_t | X_t, X_{t-1}, \dots) = 0$. Además, debido a que $\tilde{u}_t = u_t - \phi_1 u_{t-1}$, esta condición, a su vez implica que

$$\begin{aligned} 0 &= E(\tilde{u}_t | X_t, X_{t-1}, \dots) \\ &= E(u_t - \phi_1 u_{t-1} | X_t, X_{t-1}, \dots) \\ &= E(u_t | X_t, X_{t-1}, \dots) - \phi_1 E(u_{t-1} | X_t, X_{t-1}, \dots). \end{aligned} \quad (15.25)$$

Para que la igualdad de la Ecuación (15.25) se cumpla en general para todos los valores de ϕ_1 , debe ocurrir que $E(u_t | X_t, X_{t-1}, \dots) = 0$ y que $E(u_{t-1} | X_t, X_{t-1}, \dots) = 0$. Cambiando los subíndices de tiempo, la condición de que $E(u_{t-1} | X_t, X_{t-1}, \dots) = 0$ puede reescribirse como

$$E(u_t | X_{t+1}, X_t, X_{t-1}, \dots) = 0, \quad (15.26)$$

lo cual (por la ley de esperanzas iteradas) implica que $E(u_t|X_t, X_{t-1}, \dots) = 0$. En resumen, que se cumpla el supuesto de media condicional igual a cero en la Ecuación (15.24) en general para todos los valores de ϕ_1 es equivalente a que se cumpla la condición de la Ecuación (15.26).

La condición de la Ecuación (15.26) se encuentra implícita en el hecho de que X_t es estrictamente exógena, pero no está implícita en la condición de que X_t sea exógena (pasada y presente). Por lo tanto, los supuestos de mínimos cuadrados para la estimación del modelo de retardos distribuidos de la Ecuación (15.23) se cumplen si X_t es estrictamente exógena, pero no es suficiente con que X_t sea exógena (pasada y presente).

Debido a que la representación ARD [Ecuaciones (15.21) y (15.22)] es equivalente a la representación en cuasi diferencias [Ecuación (15.23)], el supuesto de media condicional necesario para estimar los coeficientes de la representación en cuasi diferencias [que $E(u_t|X_{t+1}, X_t, X_{t-1}, \dots) = 0$] es asimismo el supuesto de media condicional para la estimación consistente de los coeficientes de la representación ARD.

Pasamos ahora a las dos estrategias de cálculo sugeridas por estas dos representaciones: la estimación de los coeficientes de ARD y la estimación de los coeficientes del modelo en cuasi diferencias.

Estimación MCO del modelo ARD

La primera estrategia consiste en utilizar MCO para estimar los coeficientes del modelo ARD de la Ecuación (15.21). Tal y como demuestra la deducción que conduce a la Ecuación (15.21), la inclusión del retardo de Y y de un retardo adicional de X como regresores hace que el término de error esté serialmente incorrelacionado (bajo el supuesto de que el error sigue un proceso autorregresivo de primer orden). Por tanto, se pueden utilizar los errores estándar MCO habituales; es decir, los errores estándar HAC no son necesarios cuando los coeficientes del modelo ARD de la Ecuación (15.21) se estiman por MCO.

Los coeficientes ARD estimados no son en sí mismos estimaciones de los multiplicadores dinámicos, pero se pueden calcular los multiplicadores dinámicos a partir de los coeficientes ARD. Un método general para calcular los multiplicadores dinámicos consiste en expresar la función de regresión estimada como función de los valores actuales y pasados de X_t , es decir, eliminando Y_t de la función de regresión estimada. Para ello, se sustituyen de forma repetida las expresiones de los valores retardados de Y_t en la función de regresión estimada. En concreto, se considera la función de regresión estimada

$$\hat{Y}_t = \hat{\phi}_1 Y_{t-1} + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2}, \quad (15.27)$$

donde el término independiente estimado ha sido omitido debido a que no forma parte de ninguna expresión de los multiplicadores dinámicos. Retardando ambos lados de la expresión de la Ecuación (15.27) se obtiene $\hat{Y}_{t-1} = \hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}$, por lo que sustituyendo \hat{Y}_{t-1} en la Ecuación (15.27) por esta expresión de \hat{Y}_{t-1} y agrupando términos se obtiene

$$\begin{aligned} \hat{Y}_t &= \hat{\phi}_1(\hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}) + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2} \\ &= \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1 \hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1) X_{t-2} + \hat{\phi}_1 \hat{\delta}_2 X_{t-3} + \hat{\phi}_1^2 Y_{t-2}. \end{aligned} \quad (15.28)$$

Repetiendo este proceso varias veces mediante la sustitución sucesiva de las expresiones de Y_{t-2} , Y_{t-3} , etc. se obtiene

$$\begin{aligned} \hat{Y}_t &= \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1 \hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0) X_{t-2} \\ &\quad + \hat{\phi}_1(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0) X_{t-3} + \hat{\phi}_1^2(\hat{\delta}_2 + \hat{\phi}_1 \hat{\delta}_1 + \hat{\phi}_1^2 \hat{\delta}_0) X_{t-4} + \dots \end{aligned} \quad (15.29)$$

Los coeficientes de la Ecuación (15.29) son los estimadores de los multiplicadores dinámicos, calculados a partir de los estimadores MCO de los coeficientes del modelo ARD de la Ecuación (15.21). Si las restricciones sobre los coeficientes de la Ecuación (15.22) se cumplieran de forma exacta para los coeficientes *estimados*, entonces todos los multiplicadores dinámicos a partir del segundo (es decir, los coeficientes

de X_{t-2} , X_{t-3} , etc.) serían iguales a cero⁵. Sin embargo, si se adopta esta estrategia de estimación, tales restricciones no se cumplirán con exactitud, por lo que los multiplicadores estimados a partir del segundo en la Ecuación (15.29) serán en general distintos de cero.

Estimación MCG

La segunda estrategia para estimar los multiplicadores dinámicos cuando X_t es estrictamente exógena consiste en utilizar mínimos cuadrados generalizados (MCG), lo que implica la estimación de la Ecuación (15.23). Para describir el estimador MCG, inicialmente se supone que ϕ_1 es conocido. Debido a que en la práctica es desconocido, este estimador no resulta factible, por lo que se denomina estimador MCG infactible. Sin embargo, se puede modificar el estimador MCG infactible, utilizando un estimador de ϕ_1 , lo que da lugar a una versión factible del estimador MCG.

MCG infactible. Supongamos que ϕ_1 es conocido; entonces las variables quasi diferenciadas \tilde{X}_t e \tilde{Y}_t pueden calcularse directamente. Tal y como se analizó en el contexto de las Ecuaciones (15.24) y (15.26), si X_t es estrictamente exógena, entonces $E(\tilde{u}_t | \tilde{X}_t, \tilde{X}_{t-1}, \dots) = 0$. Por lo tanto, si X_t es estrictamente exógena y si ϕ_1 es conocido, los coeficientes α_0 , β_1 , y β_2 de la Ecuación (15.23) se pueden estimar mediante la regresión MCO de la variable \tilde{Y}_t sobre \tilde{X}_t y \tilde{X}_{t-1} (incluyendo un término independiente). Los estimadores resultantes de β_1 y β_2 —es decir, los estimadores MCO de los coeficientes de las pendientes en la Ecuación (15.23) cuando ϕ_1 es conocido— forman parte del **estimador MCG infactible**. Estos estimadores no son factibles debido a que ϕ_1 es desconocido, por lo que \tilde{X}_t e \tilde{Y}_t no se pueden calcular y por lo tanto estos estimadores MCO en realidad no pueden calcularse.

MCG factible. El **estimador MCG factible** modifica el estimador MCG infactible utilizando un estimador preliminar de ϕ_1 , $\hat{\phi}_1$, para calcular la estimación de las quasi diferencias. En concreto, los estimadores MCG factibles de β_1 y β_2 son los estimadores MCO de β_1 y β_2 en la Ecuación (15.23), calculados mediante la regresión de \hat{Y}_t sobre \hat{X}_t y \hat{X}_{t-1} (con un término independiente), donde $\hat{X}_t = X_t - \hat{\phi}_1 X_{t-1}$ e $\hat{Y}_t = Y_t - \hat{\phi}_1 Y_{t-1}$.

El estimador preliminar, $\hat{\phi}_1$, se puede calcular estimando, en primer lugar, la regresión de retardos distribuidos de la Ecuación (15.18) por MCO, y utilizando, más tarde, MCO para estimar ϕ_1 en la Ecuación (15.19) con los residuos MCO \hat{u}_t en lugar de los errores de la regresión no observables u_t . Esta versión del estimador MCG se denomina estimador de Cochrane-Orcutt (1949).

Una posible extensión del método de Cochrane-Orcutt consiste en continuar con este proceso de forma iterativa: utilizar el estimador MCG de β_1 y β_2 para calcular los estimadores revisados de u_t ; utilizar estos nuevos residuos para re-estimar ϕ_1 ; utilizar esta estimación revisada de ϕ_1 para calcular una estimación revisada de las quasi diferencias; utilizar esta estimación revisada de las quasi diferencias para re-estimar β_1 y β_2 ; y continuar con este proceso hasta que los estimadores de β_1 y β_2 converjan. Esto se conoce como estimador iterado de Cochrane-Orcutt.

Interpretación de mínimos cuadrados no lineales del estimador MCG. Una interpretación equivalente del estimador MCG es que estima el modelo ARD de la Ecuación (15.21), imponiendo las restricciones sobre los parámetros de la Ecuación (15.22). Estas restricciones son funciones no lineales de los parámetros originales β_0 , β_1 , β_2 , y ϕ_1 , por lo que esta estimación no se puede llevar a cabo mediante MCO. En cambio, los parámetros pueden ser estimados por mínimos cuadrados no lineales (MCNL). Tal y como se explica en el Apéndice 8.1, MCNL minimiza la suma de los errores al cuadrado cometidos por la función de regresión estimada, teniendo en cuenta que la función de regresión es una función no lineal de los parámetros que se estiman. En general, la estimación de MCNL puede necesitar algoritmos sofisticados para lograr minimizar funciones no lineales de los parámetros desconocidos. Sin embargo, en el caso particular que nos ocupa, estos sofisticados algoritmos no son necesarios; sino que el estimador MCNL se puede calcular mediante el algoritmo descrito anteriormente para el estimador de Cochrane-Orcutt iterado. Por lo que

⁵ Sustituyendo las igualdades que aparecen en la Ecuación (15.22) se demuestra que, si esas igualdades se cumplen, entonces $\delta_2 + \phi_1 \delta_1 + \phi_1^2 \delta_0 = 0$.

el estimador MCG de Cochrane-Orcutt iterado es, en realidad el estimador de MCNL de los coeficientes del modelo ARD, sujeto a las restricciones no lineales de la Ecuación (15.22).

Eficiencia de MCG. La virtud del estimador MCG es que cuando X es estrictamente exógena y los errores transformados \tilde{u}_t son homocedásticos, es eficiente entre los estimadores lineales, al menos en muestras grandes. Para comprobarlo, consideremos en primer lugar el estimador MCG infactible. Si \tilde{u}_t es homocedástico, si ϕ_1 es conocido (por lo que \tilde{X}_t e \tilde{Y}_t se pueden considerar como si fueran observables), y si X_t es estrictamente exógena, entonces el teorema de Gauss Markov conlleva que el estimador MCO de α_0 , β_1 y β_2 de la Ecuación (15.23) es eficiente entre todos los estimadores lineales condicionalmente insesgados; es decir, el estimador MCO de los coeficientes de la Ecuación (15.23) es el estimador lineal insesgado óptimo, o ELIO (Sección 5.5). Debido a que el estimador MCO de la Ecuación (15.23) es el estimador MCG infactible, esto significa que el estimador MCG infactible es ELIO. El estimador MCG factible es similar al estimador MCG infactible, salvo por que ϕ_1 es estimado. Debido a que el estimador de ϕ_1 es consistente y a que su varianza es inversamente proporcional a T , los estimadores MCG factibles e infactibles tienen la misma varianza en muestras grandes. En este sentido, si X es estrictamente exógena, entonces el estimador MCG factible es ELIO en muestras grandes. En particular, si X es estrictamente exógena, entonces el estimador MCG es más eficiente que el estimador MCO para los coeficientes de retardos distribuidos analizados en la Sección 15.3.

Los estimadores de Cochrane-Orcutt y de Cochrane-Orcutt iterados presentados aquí son casos particulares de la estimación MCG. En general, la estimación MCG implica la transformación del modelo de regresión para que los errores sean homocedásticos y serialmente incorrelacionados, y posteriormente la estimación por MCO de los coeficientes del modelo de regresión transformado. En general, el estimador MCG es consistente y ELIO en muestras grandes si X es estrictamente exógena, pero no es consistente si X solamente es exógena (pasada y presente). Las matemáticas que subyacen tras la estimación MCG incluyen álgebra de matrices, por lo que se posponen hasta la Sección 18.6.

El modelo de retardos distribuidos con retardos adicionales y errores AR(p)

El análisis anterior del modelo de retardos distribuidos de las Ecuaciones (15.18) y (15.19), que presenta un único retardo de X_t y un término de error AR(1), se puede trasladar al modelo general de retardos distribuidos con varios retardos y con término de error AR(p).

El modelo general de retardos distribuidos con errores autorregresivos. El modelo general de retardos distribuidos con r retardos y un término de error AR(p) es

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (15.30)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \tilde{u}_t, \quad (15.31)$$

donde, $\beta_1, \dots, \beta_{r+1}$ son los multiplicadores dinámicos y ϕ_1, \dots, ϕ_p , son los coeficientes autorregresivos del término de error. Si el modelo para los errores es AR(p), entonces \tilde{u}_t estará serialmente incorrelacionado.

Un razonamiento algebraico similar al que condujo al modelo de ARD de la Ecuación (15.21) demuestra que las Ecuaciones (15.30) y (15.31) implican que Y_t se puede escribir en forma ARD como:

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \tilde{u}_t, \quad (15.32)$$

donde $q = r + p$ y $\delta_0, \dots, \delta_q$ son funciones de las β y de las ϕ de las Ecuaciones (15.30) y (15.31). De manera equivalente, el modelo de las Ecuaciones (15.30) y (15.31) se puede escribir en forma de cuasi diferencias como

$$\tilde{Y}_t = \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \cdots + \beta_{r+1} \tilde{X}_{t-r} + \tilde{u}_t, \quad (15.33)$$

donde $\tilde{Y}_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$ y $\tilde{X}_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$.

Condiciones para la estimación de los coeficientes ARD. El análisis anterior acerca de las condiciones para la estimación consistente de los coeficientes ARD en el caso AR(1) se puede extender al modelo general con errores AR(p). El supuesto de media condicional igual a cero para la Ecuación (15.33) consiste en que

$$E(\tilde{u}_t | \tilde{X}_t, \tilde{X}_{t-1}, \dots) = 0. \quad (15.34)$$

Debido a que $\tilde{u}_t = u_t - \phi_1 u_{t-1} - \dots - \phi_p u_{t-p}$ y a que $\tilde{X}_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}$, esta condición es equivalente a

$$\begin{aligned} E(u_t | X_t, X_{t-1}, \dots) - \phi_1 E(u_{t-1} | X_t, X_{t-1}, \dots) \\ - \dots - \phi_p E(u_{t-p} | X_t, X_{t-1}, \dots) = 0. \end{aligned} \quad (15.35)$$

Para que la Ecuación (15.35) se cumpla para cualesquiera valores de ϕ_1, \dots, ϕ_p , debe ocurrir que cada una de las esperanzas condicionales de la Ecuación (15.35) sea igual a cero; de forma equivalente, debe ocurrir que

$$E(u_t | X_{t+p}, X_{t+p-1}, X_{t+p-2}, \dots) = 0. \quad (15.36)$$

Esta condición no está implícita en el hecho de que X_t sea exógena (pasada y presente), pero sí está implícita en que X_t sea estrictamente exógena. De hecho, en el límite, cuando p es infinito (por lo que el término de error en el modelo de retardos distribuidos sigue una regresión de orden infinito), la condición de la Ecuación (15.36) se convierte en la condición del Concepto clave 15.1 para exogeneidad estricta.

Estimación del modelo ARD mediante MCO. Como en el caso del modelo de retardos distribuidos con un único retardo y un término de error AR(1), los multiplicadores dinámicos se pueden estimar a partir los estimadores MCO de los coeficientes ARD de la Ecuación (15.32). Las fórmulas generales son similares, pero más complicadas, que las de la Ecuación (15.29) y se expresan de una manera más adecuada mediante la notación del multiplicador de retardos; estas fórmulas están recogidas en el Apéndice 15.2. En la práctica, el *software* de regresión moderno está diseñado para el análisis de regresión de series temporales y es el que realiza estos cálculos.

Estimación mediante MCG. Por otra parte, los multiplicadores dinámicos se pueden estimar por MCG (factibles). Esto implica la estimación MCO de los coeficientes de la especificación en cuasi diferencias de la Ecuación (15.33), utilizando estimaciones de las cuasi diferencias. Las estimaciones de las cuasi diferencias se pueden calcular utilizando estimadores preliminares de los coeficientes autorregresivos ϕ_1, \dots, ϕ_p , como en el caso del modelo AR(1). El estimador MCG es asintóticamente ELIO, en el sentido estudiado anteriormente para el caso AR(1).

La estimación de los multiplicadores dinámicos en condiciones de exogeneidad estricta se encuentra recogida en el Concepto clave 15.4.

¿Qué utilizar: MCO o MCG? Las dos opciones de estimación, la estimación MCO de los coeficientes ARD y la estimación MCG de los coeficientes de retardos distribuidos, presentan ventajas e inconvenientes.

La ventaja del método ARD es que puede reducir el número de parámetros necesarios para calcular los multiplicadores dinámicos, en comparación con la estimación MCO del modelo de retardos distribuidos. Por ejemplo, el modelo ARD estimado de la Ecuación (15.27) daba lugar a la representación de retardos distribuidos infinitamente grande estimada en la Ecuación (15.29). En la medida en que un modelo de retardos distribuidos con solamente r retardos constituya verdaderamente una aproximación de un modelo de retardos distribuidos con mayores retardos, el modelo ARD puede proporcionar un método sencillo para estimar aquellos retardos más lejanos utilizando solamente unos pocos parámetros desconocidos. Por tanto, en la práctica podría ser posible estimar el modelo ARD de la Ecuación (15.39) con valores de p y q mucho menores que los valores de r necesarios para la estimación MCO de los coeficientes de retardos distribuidos de la Ecuación (15.37). En otras palabras, la especificación ARD puede proporcionar un resumen compacto, o parsimonioso, de una distribución de retardos larga y compleja (véase el Apéndice 15.2 para un análisis más avanzado).

**CONCEPTO
CLAVE**
Estimación de multiplicadores dinámicos con exogeneidad estricta

El modelo general de retardos distribuidos con r retardos y término de error AR(p) es

15.4

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t \quad (15.37)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \tilde{u}_t. \quad (15.38)$$

Si X_t es estrictamente exógena, entonces los multiplicadores dinámicos $\beta_1, \dots, \beta_{r+1}$ se pueden estimar utilizando en primer lugar MCO para estimar los coeficientes del modelo ARD

$$\begin{aligned} Y_t &= \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} \\ &\quad + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \tilde{u}_t, \end{aligned} \quad (15.39)$$

donde $q = r + p$ y posteriormente calculando los multiplicadores dinámicos utilizando el *software* de regresión. Por otra parte, los multiplicadores dinámicos se pueden estimar mediante la estimación de los coeficientes de los retardos distribuidos de la Ecuación (15.37) por MCG.

La ventaja del estimador MCG consiste en que, para una longitud de retardos del modelo de retardos distribuidos dada, r , el estimador MCG de los coeficientes de los retardos distribuidos es más eficiente que el estimador MCO, al menos en muestras grandes. En la práctica, por tanto, la ventaja de utilizar el método ARD se debe a que la especificación ARD puede permitir estimar menos parámetros que los que se estiman mediante el método MCG.

15.6 Los precios del zumo de naranja y el frío

Esta sección utiliza las herramientas de la regresión de series temporales para obtener otras informaciones a partir de los datos sobre las temperaturas de Florida y los precios del zumo de naranja. En primer lugar, ¿cuánto tiempo perdura el efecto de una helada sobre los precios? En segundo lugar, ¿ha sido estable este efecto dinámico o, por el contrario, ha cambiado a lo largo de los 51 años que abarcan los datos disponibles? Y, si es así, ¿cuánto?

Comenzamos este análisis con la estimación de los efectos causales dinámicos mediante el método de la Sección 15.3, es decir, mediante la estimación MCO de los coeficientes de una regresión de retardos distribuidos de la variación porcentual en los precios (%VP_t) sobre el índice de heladas en ese mes (IH_t) y sobre sus valores retardados. Para que el estimador de retardos distribuidos sea consistente, la variable IH debe ser exógena (pasada y presente). Tal y como se analizó en la Sección 15.2, este supuesto resulta razonable en este caso. Las personas no pueden influir en las condiciones meteorológicas, por lo que resulta adecuado considerar el tiempo como si fuera asignado al azar en un experimento. Debido a que la variable IH es exógena, los efectos causales dinámicos se pueden estimar mediante la estimación MCO de los coeficientes del modelo de retardos distribuidos de la Ecuación (15.4) en el Concepto clave 15.1.

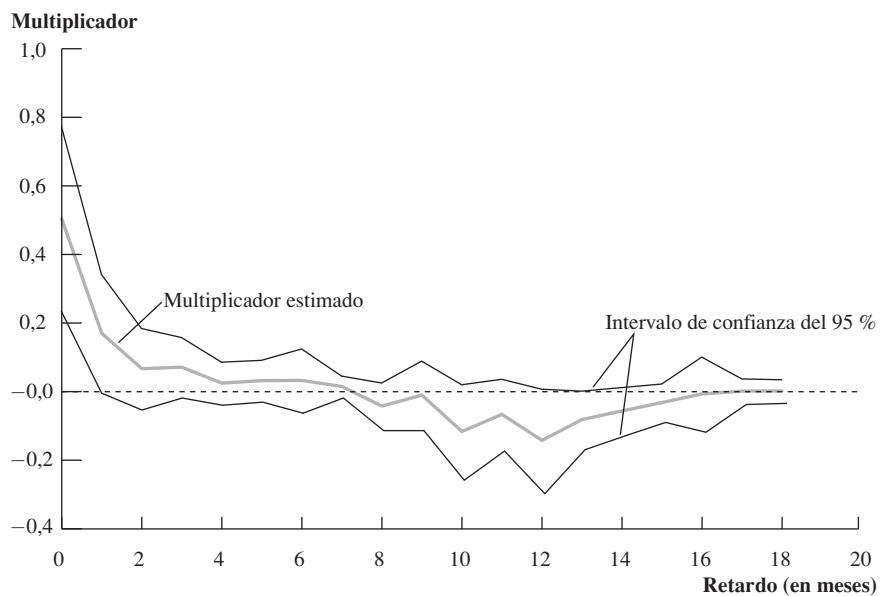
Tal y como se analizó en las Secciones 15.3 y 15.4, el término de error puede presentar correlación serial en las regresiones de retardos distribuidos, por lo que es importante utilizar los errores estándar HAC, que se ajustan a esta correlación serial. Para los resultados iniciales, el parámetro de truncamiento de los errores estándar de Newey-West (m en la notación de la Sección 15.4) fue elegido mediante la regla de la Ecuación (15.17): debido a que existen 612 observaciones mensuales, de acuerdo con esa regla $m = 0,75T^{1/3} = 0,75 \times 612^{1/3} = 6,37$, pero como m debe ser un número natural, esto se redondea a $m = 7$, la sensibilidad de los errores estándar a la elección de este parámetro de truncamiento se investiga a continuación.

Los resultados de la estimación MCO para la regresión de retardos distribuidos de la variable $\%VP_t$ sobre las variables $IH_t, IH_{t-1}, \dots, IH_{t-18}$ se recogen en la columna (1) de la Tabla 15.1. Los coeficientes de esta regresión (solo algunos de los cuales se presentan en la tabla) son estimaciones del efecto causal dinámico sobre las variaciones en los precios del zumo de naranja (en términos porcentuales) durante los primeros 18 meses siguientes al aumento unitario en el índice de heladas en un mes. Por ejemplo, se estima que una variación unitaria del índice de heladas aumenta el precio en un 0,50 % a lo largo del mes en que sucede ese índice de heladas. El efecto sobre los precios de una variación unitaria del índice de heladas en los meses siguientes es menor: después de un mes, el efecto estimado es de un aumento en el precio en un 0,17 % adicional; y después de 2 meses, el efecto estimado es el de un aumento en el precio de un 0,07 %

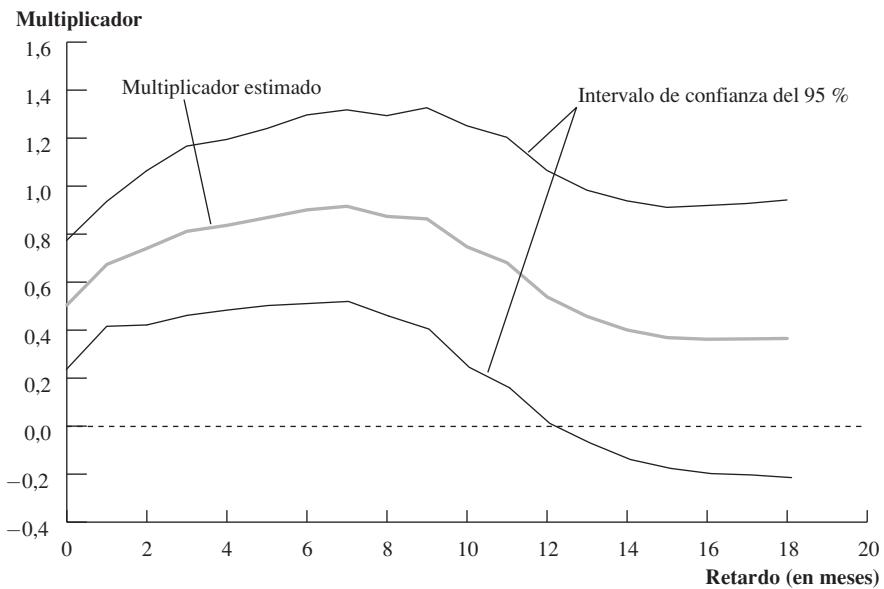
TABLA 15.1 El efecto dinámico de un grado-día en el índice de heladas (IH) sobre el precio del zumo de naranja: selección de multiplicadores dinámicos y multiplicadores dinámicos acumulativos

N.º del retardo	(1) Multiplicadores dinámicos	(2) Multilicadores acumulativos	(3) Multiplicadores acumulativos	(4) Multiplicadores acumulativos
0	0,50 (0,14)	0,50 (0,14)	0,50 (0,14)	0,51 (0,15)
1	0,17 (0,09)	0,67 (0,14)	0,67 (0,13)	0,70 (0,15)
2	0,07 (0,06)	0,74 (0,17)	0,74 (0,16)	0,76 (0,18)
3	0,07 (0,04)	0,81 (0,18)	0,81 (0,18)	0,84 (0,19)
4	0,02 (0,03)	0,84 (0,19)	0,84 (0,19)	0,87 (0,20)
5	0,03 (0,03)	0,87 (0,19)	0,87 (0,19)	0,89 (0,20)
6	0,03 (0,05)	0,90 (0,20)	0,90 (0,21)	0,91 (0,21)
.				
.				
.				
12	0,14 (0,08)	0,54 (0,27)	0,54 (0,28)	0,54 (0,28)
.				
.				
18	0,00 (0,02)	0,37 (0,30)	0,37 (0,31)	0,37 (0,30)
¿Indicadores mensuales?	No	No	No	Si $F = 1,01$ ($p = 0,43$)
Parámetro de truncamiento del error estándar HAC	7	7	14	7
Todas las regresiones fueron estimadas por MCO utilizando datos mensuales (que se describen en el Apéndice 15.1) desde enero de 1950 hasta diciembre de 2000, para un total de $T = 612$ observaciones mensuales. La variable dependiente es la variación mensual en términos porcentuales en el precio del zumo de naranja ($\%VP_t$). La regresión (1) es la regresión de retardos distribuidos con el índice de heladas y 18 de sus valores retardados, es decir, $IH_t, IH_{t-1}, \dots, IH_{t-18}$, y los coeficientes que se presentan son las estimaciones MCO de los multiplicadores dinámicos. Los multiplicadores acumulativos son la suma acumulada de los multiplicadores dinámicos estimados. Todas las regresiones incluyen un término independiente que no se presenta en la tabla. Entre paréntesis se presentan los errores estándar HAC de Newey-West, calculados utilizando el valor del parámetro de truncamiento que se presenta en la última fila.				

FIGURA 15.2 El efecto dinámico de un grado-día en el índice de heladas (IH) sobre el precio del zumo de naranja



(a) Multiplicadores dinámicos estimados e intervalos de confianza del 95 %



(b) Multiplicadores dinámicos acumulados estimados e intervalos de confianza del 95 %

Los multiplicadores dinámicos estimados muestran que una helada provoca un aumento inmediato en los precios. Los futuros aumentos en los precios son mucho menores que el impacto inicial. El multiplicador acumulativo muestra que las heladas tienen un efecto persistente en el nivel de los precios del zumo de naranja, los precios alcanzan un pico a los siete meses después de la helada.

adicional. El R^2 de esta regresión es de 0,12, lo que indica que la mayor parte de la variación mensual en los precios del zumo de naranja no se explica por los valores actuales y pasados de la variable IH .

La representación gráfica de los multiplicadores dinámicos puede transmitir información de forma más eficaz que las tablas como la Tabla 15.1. Los multiplicadores dinámicos de la columna (1) de la Tabla 15.1 se representan en la Figura 15.2a junto con sus intervalos de confianza del 95 %, calculados como el coefi-

ciente estimado $\pm 1,96$ errores estándar HAC. Tras el repentino aumento inicial de los precios, los posteriores aumentos en los precios son menores, aunque se estima que los precios aumentan suavemente en cada uno de los primeros 6 meses después de la helada. Tal y como puede apreciarse en la Figura 15.2a, exceptuando el primer mes, para los meses restantes los multiplicadores dinámicos no son estadística y significativamente distintos de cero al nivel de significación del 5 %, aunque se estima que son positivos hasta el séptimo mes.

La Columna (2) de la Tabla 15.1 contiene los multiplicadores dinámicos acumulativos para esta especificación, es decir, la suma acumulada de los multiplicadores dinámicos presentados en la columna (1). Estos multiplicadores dinámicos están representados gráficamente en la Figura 15.2b, junto con sus intervalos de confianza al 95 %. Después del primer mes, el efecto acumulado del grado-día en el índice de heladas es el de un aumento en los precios en un 0,67 %, después de 2 meses se estima que el precio aumenta en un 0,74 %, y tras 6 meses se estima que el precio aumenta en un 0,90 %. Como se puede observar en la Figura 15.2b, estos multiplicadores acumulativos aumentan hasta el séptimo mes, debido a que los multiplicadores dinámicos individuales son positivos para los primeros 7 meses. Para el octavo mes, el multiplicador dinámico es negativo, por lo que el precio del zumo de naranja comienza a caer lentamente desde su máximo. Tras 18 meses, el incremento acumulado de los precios es de solamente un 0,37 %; es decir, el multiplicador dinámico acumulativo de largo plazo es solamente 0,37 %. Este multiplicador dinámico acumulativo de largo plazo no es estadística y significativamente distinto de cero al nivel de significación del 10 % ($t = 0,37/0,30 = 1,23$).

Análisis de sensibilidad. Al igual que en cualquier análisis empírico, es importante comprobar si estos resultados son sensibles a los cambios en los detalles del análisis empírico. Por lo tanto, se examinan tres aspectos de este análisis: la sensibilidad respecto al cálculo de los errores estándar HAC; una especificación alternativa que investiga la presencia de un posible sesgo de variable omitida; y un análisis de la estabilidad a lo largo del tiempo de los multiplicadores estimados.

En primer lugar, se investiga si los errores estándar presentados en la segunda columna de la Tabla 15.1 son sensibles a las distintas opciones del parámetro de truncamiento HAC, m . En la columna (3), se presentan los resultados para $m = 14$, el doble del valor que se utiliza en columna (2). La especificación de la regresión es la misma que en la columna (2), por lo que los coeficientes estimados y los multiplicadores dinámicos son idénticos; solamente los errores estándar son diferentes pero, en este caso, no demasiado. La conclusión es que los resultados no son sensibles a los cambios en el parámetro de truncamiento HAC.

En segundo lugar, se investiga una posible fuente de sesgo de variable omitida. Las heladas en Florida no suceden aleatoriamente (se asignan aleatoriamente) a lo largo del año, sino que más bien se producen en el invierno (por supuesto). Si la demanda de zumo de naranja es estacional (¿se demanda más zumo de naranja en invierno que en verano?), entonces el patrón estacional de la demanda de zumo de naranja podría estar correlacionado con la variable IH , lo que daría lugar a un sesgo de variable omitida. La cantidad de naranjas vendidas para zumo es endógena: los precios y las cantidades se determinan al mismo tiempo por las fuerzas de oferta y demanda. Por tanto, tal y como se estudió en la Sección 9.2, la inclusión de la cantidad daría lugar a un sesgo por simultaneidad. Sin embargo, la componente estacional de la demanda se puede captar mediante la inclusión como regresores de variables estacionales. La especificación de la columna (4) de la Tabla 15.1 incluye por tanto 11 variables binarias mensuales, una que indica si el mes es enero, una que indica si es febrero, y así sucesivamente (como de costumbre, debe omitirse una variable binaria para evitar la multicolinealidad perfecta con el término independiente). Estas variables indicadoras de los meses no son estadísticamente significativas de forma conjunta al nivel del 10 % ($p = 0,43$), y los multiplicadores dinámicos acumulativos estimados son esencialmente los mismos que para las especificaciones que excluyen los indicadores mensuales. En resumen, las fluctuaciones estacionales de la demanda no son una fuente importante de sesgo de variable omitida.

¿Se han mantenido estables en el tiempo los multiplicadores dinámicos?⁶ A fin de evaluar la estabilidad de los multiplicadores dinámicos, resulta necesario comprobar si los coeficientes de la regre-

⁶ El análisis de estabilidad en este subapartado se basa en el material de la Sección 14.7 y puede pasarse por alto si ese material no ha sido estudiado.

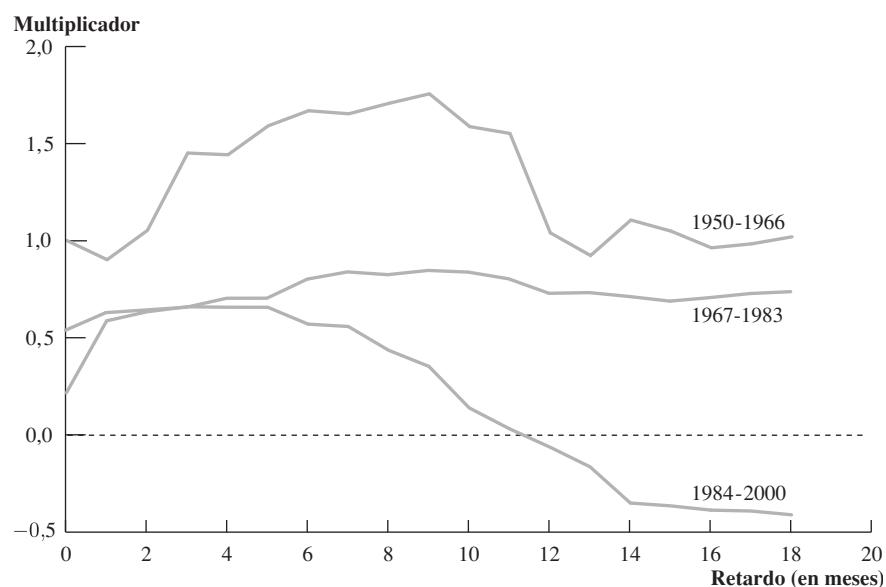
sión de retardos distribuidos se han mantenido estables en el tiempo. Debido a que no tenemos un punto de ruptura específico en mente, se contrasta la inestabilidad en los coeficientes de la regresión utilizando el estadístico de la razón de verosimilitud de Quandt (QLR) (Concepto clave 14.9). El estadístico QLR (con un 15 % de reducción y un estimador de la varianza HAC), calculado para la regresión de la columna (1) con todos los coeficientes de interacción, tiene un valor de 21,19, con $q = 20$ grados de libertad (los coeficientes de la variable IH_i , sus 18 primeros retardos, y el término independiente). El valor crítico al 1 % en la Tabla 14.6 es de 2,43, por lo que el estadístico QLR rechaza al nivel de significación del 1 %. Esas regresiones QLR tienen 40 regresores, un número grande; al recalcularlas para solamente seis retardos (de modo que haya 16 regresores y $q = 8$), asimismo da como resultado el rechazo al 1 % de nivel. Por tanto se rechaza la hipótesis de que los multiplicadores dinámicos son estables al nivel de significación del 1 %.

Una forma de comprobar cómo los multiplicadores dinámicos han cambiado a lo largo del tiempo consiste en calcularlos para diferentes partes de la muestra. La Figura 15.3 representa gráficamente los multiplicadores dinámicos acumulativos estimados para el primer tercio (1950-1966), el tercio medio (1967-1983), y el último tercio (1984-2000) de la muestra, calculados mediante la realización por separado de las regresiones para cada una de las submuestras. Estas estimaciones muestran un patrón interesante y destacable. En la década de 1950 y comienzos de la década de 1960, un grado-día del índice de heladas tenía un efecto grande y persistente sobre el precio. La magnitud del efecto sobre el precio de un grado-día del índice de heladas disminuyó en la década de 1970, aunque siguió siendo muy persistente. A finales de la década de 1980 y durante la década de 1990, el efecto a corto plazo de un grado-día del índice de heladas era el mismo que en la década de 1970, pero se hizo mucho menos persistente y prácticamente se eliminaba después de un año. Estas estimaciones sugieren que el efecto causal dinámico sobre los precios del zumo de naranja de una helada en Florida se ha hecho menor y menos persistente en la segunda mitad del siglo XX. El recuadro «Naranjos en Movimiento», trata sobre una posible explicación de la inestabilidad de los efectos causales dinámicos.

Estimaciones ARD y MCG. Tal y como se analizó en la Sección 15.5, si el término de error de la regresión de retardos distribuidos está correlacionado serialmente y la variable IH es estrictamente exógena, es posible estimar los multiplicadores dinámicos de manera más eficiente que mediante la estimación MCO de los coeficientes de retardos distribuidos. No obstante, antes de utilizar ya sea el estimador MCG o el estimador basado en el modelo ARD, debe tenerse en cuenta la cuestión de si la variable IH es, en realidad, estrictamente exógena. Ciertamente, los seres humanos no pueden afectar a las condiciones meteorológicas diarias,

FIGURA 15.3 Multiplicadores dinámicos acumulativos estimados en diferentes períodos muestrales

El efecto dinámico de las heladas sobre los precios del zumo de naranja cambió significativamente a lo largo de la segunda mitad del siglo XX. Una helada tenía un impacto mayor sobre los precios durante 1950-1966 que posteriormente, y el efecto de una helada era menos persistente durante el período 1984-2000 que con anterioridad.



Naranjos en movimiento

¿Por qué varían a lo largo del tiempo los multiplicadores dinámicos de la Figura 15.3? Una posible explicación es que son las variaciones en los mercados, pero otra posible explicación sería la traslación de los árboles hacia el sur.

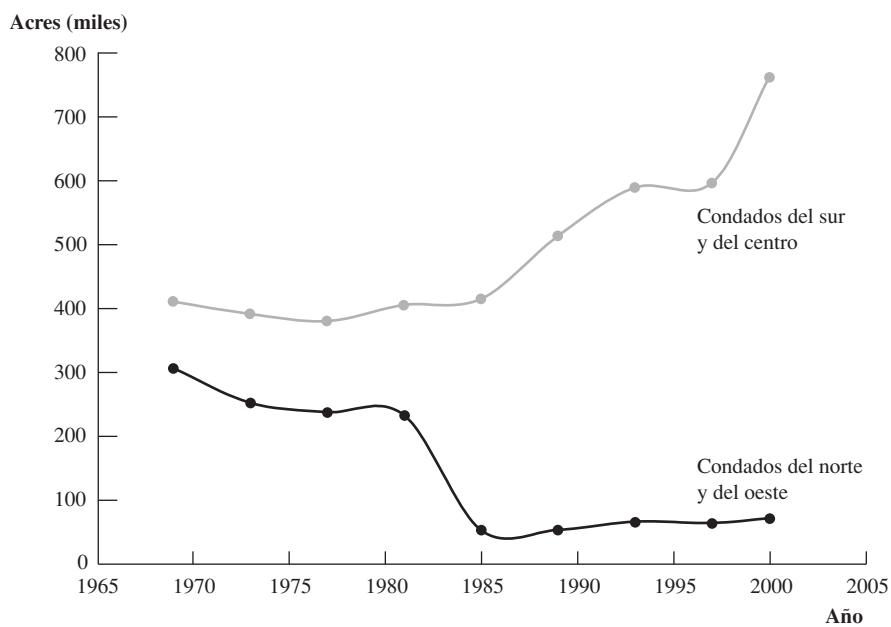
De acuerdo con el Departamento de Cítricos de Florida, las fuertes heladas registradas en la década de 1980, que pueden apreciarse en la Figura 15.1(c), estimularon a los productores de cítricos a buscar un clima más cálido. Tal y como se muestra en la Figura 15.4, el número de acres de naranjos en los condados del norte y del oeste, más propensos a las heladas, se redujo desde los 232.000 acres en 1981 hasta los 53.000 acres en 1985, y como consecuencia la superficie de naranjos en los condados del sur y del centro aumentó desde los 413.000 en 1985 hasta los 588.000 en 1993. Con los culti-

vos más al sur, las heladas del norte dañan una proporción más pequeña de la cosecha y —como indican los multiplicadores dinámicos de la Figura 15.3— el precio se vuelve menos sensible a las temperaturas de la ciudad de Orlando situada más al norte.

Está bien, los naranjos podrían no haberse movido —diría MacBeth— pero la migración hacia el sur de los naranjos proporciona un nuevo significado al término «no estacionariedad»¹.

¹ Agradecemos al profesor James Cobbe de la Universidad de Florida State su información acerca del traslado al sur de las plantaciones de naranjos.

FIGURA 15.4 Incremento de plantaciones de naranjos en Florida



pero, ¿significa eso que las condiciones meteorológicas son *estrictamente exógenas*? ¿Presenta el término de error u_t en la regresión de retardos distribuidos una media condicional igual a cero, dados los valores pasados, presentes y *futuros* de la variable IH_t ?

El término de error en la homóloga poblacional de la regresión de retardos distribuidos de la columna (1) de la Tabla 15.1 es la discrepancia entre el precio y su predicción poblacional basada en las condiciones meteorológicas de los últimos 18 meses. Esta discrepancia podría surgir por varios motivos, uno de los cuales es que los operadores utilizan las previsiones meteorológicas para Orlando. Por ejemplo, si se prevé un invierno especialmente frío, entonces los operadores lo incorporarían en el precio, por lo que el precio estaría por encima de su valor de predicción basado en la regresión poblacional; es decir, el término de error sería positivo. Si esta predicción es exacta, entonces las condiciones meteorológicas futuras podrían convertirse en unas condiciones frías en la realidad. Por lo que los grados-día del índice de heladas futuros serían

positivos ($X_{t+1} > 0$) cuando el precio actual es inusualmente alto ($u_t > 0$), por lo que $\text{corr}(X_{t+1}, u_t)$ es positiva. Expresado de una forma más sencilla, aunque los operadores del mercado de zumo de naranja no pueden influir en las condiciones meteorológicas, pueden —y lo hacen— predecirlas (véase el recuadro). En consecuencia, el término de error de la regresión precio/clima está correlacionado con las condiciones meteorológicas futuras. En otras palabras, la variable *IH* es exógena, pero si este razonamiento es cierto, no es estrictamente exógena, y los estimadores MCG y ARD no serán estimadores consistentes de los multiplicadores dinámicos. Por lo tanto, estos estimadores no se utilizan en esta aplicación.

15.7 ¿Es creíble la exogeneidad?: algunos ejemplos

Al igual que en la regresión con datos de sección cruzada, la interpretación de los coeficientes de una regresión de retardos distribuidos como efectos causales dinámicos depende del supuesto de que X sea exógena. Si X_t , o sus valores retardados están correlacionados con u_t , entonces la media condicional de u_t , dependerá de X_t o de sus retardos, en cuyo caso X no es exógena (pasada y presente). Los regresores pueden estar correlacionados con el término de error por varias razones, pero con datos económicos de series temporales, una cuestión particularmente importante es que podría ser por la existencia de causalidad simultánea, lo cual (como se explicó en las Secciones 9.2 y 12.1) da lugar a regresores endógenos. En la Sección 15.6, se trató en detalle la hipótesis de exogeneidad y exogeneidad estricta del índice de heladas. En esta sección, se examina el supuesto de exogeneidad en otras cuatro aplicaciones económicas.

La renta de EE.UU. y las exportaciones australianas

Los Estados Unidos constituyen una fuente importante de demanda para las exportaciones australianas. Podría investigarse cuál es la sensibilidad exacta de las exportaciones australianas a las fluctuaciones en la renta agregada de EE.UU. mediante la regresión de las exportaciones de Australia hacia los Estados Unidos sobre una medida de la renta de los EE.UU. En sentido estricto, debido a que la economía mundial se encuentra integrada, existe causalidad simultánea en esta relación: una disminución de las exportaciones de Australia reduce la renta de Australia, lo que reduce la demanda de las importaciones procedentes de los Estados Unidos, lo que a su vez reduce la renta de EE.UU. No obstante, en la práctica, este efecto es muy pequeño debido a que la economía australiana es mucho menor que la economía de EE.UU. Por tanto, la renta de EE.UU. puede ser considerada de forma razonable como una variable exógena de esta regresión.

Por el contrario, en una regresión de las exportaciones de la Unión Europea hacia los Estados Unidos sobre la renta de los EE.UU., el argumento para considerar la renta de los EE.UU. como una variable exógena resulta menos convincente debido a que la demanda de los residentes de la Unión Europea de las exportaciones procedentes de EE.UU. constituye una parte sustancial de la demanda total de exportaciones procedentes de EE.UU. Por lo tanto, un descenso de la demanda de EE.UU. sobre las exportaciones procedentes de la UE reduciría la renta de la UE, lo que a su vez disminuiría la demanda de las exportaciones procedentes de EE.UU. y por lo tanto una disminución de la renta de EE.UU. Debido a la existencia de estos vínculos en el comercio internacional, las exportaciones de la UE hacia Estados Unidos y la renta de los EE.UU. se determinan simultáneamente, por lo que en esta regresión la renta de EE.UU. podría decirse que no es exógena. Este ejemplo ilustra una cuestión más general, como es el hecho de que una variable es exógena dependiendo del contexto: la renta de EE.UU. es verosímilmente exógena en una regresión que explique las exportaciones australianas, pero no en una regresión que explique las exportaciones de la UE.

Los precios del petróleo y la inflación

Desde que el precio del petróleo aumentara en la década de 1970, los macroeconomistas han estado interesados en la estimación del efecto dinámico de un aumento en el precio internacional del crudo sobre la tasa de inflación de EE.UU. Debido a que los precios del petróleo se establecen en los mercados mundiales en gran parte por parte de países extranjeros que son productores de petróleo, en un principio se podría pensar que los precios del petróleo son exógenos. Sin embargo, los precios del petróleo no son como las

ÚLTIMA HORA: los operadores de materias primas hacen tiritar Disney World

Aunque el tiempo en Disney World en Orlando, Florida, suele ser agradable, de vez en cuando puede aparecer una ola de frío. Si se visita Disney World en una tarde de invierno, ¿debería llevarse una prenda de abrigo? Algunas personas pueden ver el pronóstico del tiempo en televisión, pero los que saben pueden hacer algo mejor: ¡pueden comprobar el precio de cierre de ese día del mercado de Nueva York de futuros sobre el precio del zumo de naranja!

El economista financiero Richard Roll llevó a cabo un estudio detallado sobre la relación que podría existir entre los precios del zumo de naranja y las condiciones meteorológicas. Roll (1984) examinó el efecto sobre los precios de unas condiciones meteorológicas de frío en Orlando, pero estudió asimismo el «efecto» de las variaciones en el precio de los contratos de futuros del zumo de naranja (un contrato de compra de zumo de naranja concentrado en una determinada fecha futura) sobre las condiciones meteorológicas. Roll utilizó datos diarios desde 1975 hasta 1981 sobre los precios de los contratos de futuros de zumo de naranja que cotizan en la Bolsa del Algodón de Nueva York y sobre las temperaturas diarias y nocturnas en Orlando. Halló que un aumento en el precio del contrato de futuros a lo largo de un día de negociación en Nueva York predecía un tiempo frío, en particular una ola de frío, en Orlando la noche siguiente. De hecho, el mercado era tan eficaz en la predicción de condiciones meteorológicas frías en Florida que un aumento de los precios a lo largo de un día de

negociación predecía realmente errores de predicción en las previsiones meteorológicas oficiales del gobierno de EE.UU. para esa noche.

El estudio de Roll es interesante asimismo por lo que no encontró: a pesar de que sus detallados datos meteorológicos explicaban algunas de las variaciones diarias en los precios de los futuros sobre el zumo de naranja, la mayoría de las variaciones diarias en los precios del zumo de naranja se mantuvo sin explicación. Por lo tanto, sugería que el mercado de futuros sobre el zumo de naranja mostraba «un exceso de volatilidad», es decir, una volatilidad mayor que la que se podía atribuir a las oscilaciones en los fundamentales. Entender por qué (y si) existe un exceso de volatilidad en los mercados financieros constituye en la actualidad una importante área de investigación en economía financiera.

Los resultados de Roll ilustran asimismo la diferencia entre los efectos causales dinámicos previstos y estimados. Las variaciones en los precios en el mercado de futuros del zumo de naranja es un predictor útil de un tiempo frío, pero eso no quiere decir que los operadores de materias primas sean tan poderosos que pueden *causar* una disminución de las temperaturas. Los visitantes de Disney World puede que tiriten tras un aumento en los precios de los contratos de futuros de zumo de naranja, pero no tiembran *a causa* de la subida de los precios, a menos que, por supuesto, estuvieran posicionados cortos en el mercado de futuros del zumo de naranja.

condiciones meteorológicas: los miembros de la OPEP establecen los niveles de producción de petróleo de forma estratégica, teniendo en cuenta muchos factores, que incluyen el estado de la economía mundial. En la medida en que los precios del petróleo (o las cantidades) se establezcan en base a una evaluación de las condiciones de la economía mundial actuales y futuras, que incluyan la inflación en los Estados Unidos, los precios del petróleo serán endógenos.

La política monetaria y la inflación

Los miembros de los bancos centrales que se encuentran a cargo de la política monetaria necesitan conocer el efecto sobre la inflación de la política monetaria. Debido a que la principal herramienta de política monetaria es el tipo de interés de corto plazo (el «tipo a corto»), necesitan conocer el efecto causal dinámico sobre la inflación de una variación en el tipo de interés de corto plazo. Aunque el tipo a corto es uno de los objetivos declarados por el banco central, no se establece por parte de los bancos centrales de forma aleatoria (como lo sería en un experimento aleatorizado ideal), sino que más bien se establece de forma endógena: el banco central determina el tipo a corto en base a una evaluación acerca del estado actual y futuro de la economía, que incluye de forma particular la tasa de inflación actual y futura. La tasa de inflación depende a su vez del tipo de interés (tipos de interés más elevados reducen la demanda agregada), pero el tipo de interés depende de la tasa de inflación, sus valores pasados, y su valor futuro (esperado). Por tanto el tipo a corto es endógeno, y el efecto causal dinámico de una variación en el tipo a corto sobre la inflación futura no puede ser estimado de forma consistente mediante una regresión MCO de la tasa de inflación sobre los valores actuales y pasados de los tipos de interés.

La curva de Phillips

La curva de Phillips investigada en el Capítulo 14 es una regresión de la variación de la tasa de inflación sobre las variaciones retardadas de la tasa de inflación y los retardos de la tasa de desempleo. Debido a que los retardos de la tasa de desempleo ocurrieron en el pasado, se podría pensar que no puede existir interacción mutua entre las tasas de inflación actuales y los valores pasados de la tasa de desempleo, por lo que los valores pasados de la tasa de desempleo pueden ser considerados como exógenos. Sin embargo, los valores pasados de la tasa de desempleo no han sido asignados aleatoriamente en un experimento; sino que la tasa de desempleo pasada fue determinada simultáneamente con los valores pasados de la inflación. Debido a que la tasa inflación y la tasa de desempleo se determinan simultáneamente, el resto de los otros factores que determinan la inflación que están incluidos en u_t , están correlacionados con los valores pasados de la tasa de desempleo; es decir, la tasa de desempleo no es exógena. De ello se desprende que la tasa de desempleo no es estrictamente exógena, por lo que los multiplicadores dinámicos calculados mediante una curva de Phillips empírica [por ejemplo, el modelo ARD de la Ecuación (14.17)] no son estimaciones consistentes del efecto causal dinámico sobre la inflación de una variación de la tasa de desempleo.

15.8 Conclusión

Los datos de series temporales proporcionan la oportunidad de estimar la trayectoria temporal de los efectos sobre Y de una variación en X , es decir, el efecto causal dinámico sobre Y de un cambio en X . No obstante, para estimar los efectos causales dinámicos mediante una regresión de retardos distribuidos, X debe ser exógena, como lo sería si estuviera asignada aleatoriamente en un experimento aleatorizado ideal. Si X no solamente es exógena, sino que es además *estrictamente exógena*, entonces los efectos causales dinámicos pueden estimarse utilizando un modelo autorregresivo de retardos distribuidos o bien mediante MCG.

En algunas aplicaciones, tales como la estimación del efecto causal dinámico sobre el precio del zumo de naranja de una ola de frío en Florida, puede formularse un argumento convincente en favor de que el regresor (el índices de heladas) sea exógeno; por lo que el efecto causal dinámico puede ser estimado mediante la estimación MCO de los coeficientes de los retardos distribuidos. Sin embargo, incluso en esta aplicación, la teoría económica sugiere que el clima no es una variable estrictamente exógena, por lo que los métodos ARD y MCG no resultan adecuados. Por otra parte, en muchas relaciones que los económetras consideran interesantes, existe causalidad simultánea, por lo que el regresor en estas especificaciones no es exógeno, sea en sentido estricto o no. Determinar si el regresor es exógeno (o exógeno en sentido estricto) en última instancia, requiere la combinación de la teoría económica, el conocimiento institucional y una valoración cuidadosa.

Resumen

1. Los efectos causales dinámicos en series temporales se definen en el contexto de un experimento aleatorizado, en el que un mismo sujeto (entidad individual) recibe diferentes tratamientos asignados aleatoriamente durante diferentes períodos. Los coeficientes de una regresión de retardos distribuidos de Y sobre X y sus retardos se pueden interpretar en términos de efectos causales dinámicos cuando la trayectoria temporal de X se determina aleatoriamente y de forma independiente de otros factores que influyen en Y .
2. La variable X es exógena (pasada y presente), si la media condicional del término de error u_t de la regresión de retardos distribuidos de Y sobre los valores actuales y pasados de X no depende de los valores actuales y pasados de X . Si además la media condicional de u_t no depende de los valores futuros de X , entonces X es estrictamente exógena.
3. Si X es exógena, entonces los estimadores MCO de los coeficientes de una regresión de retardos distribuidos de Y sobre los valores actuales y pasados de X son estimadores consistentes de los efectos causales dinámicos. En general, el término de error u_t de esta regresión está correlacionado serialmente, de

modo que los errores estándar convencionales resultan engañosos, por lo que deben utilizarse en su lugar los errores estándar HAC.

4. Si X es estrictamente exógena, entonces los multiplicadores dinámicos pueden ser estimados mediante la estimación MCO de un modelo ARD o bien mediante MCG.
5. La exogeneidad es un supuesto fuerte que en muchas ocasiones no se cumple para datos económicos de series temporales debido a la existencia de causalidad simultánea, además, el supuesto de exogeneidad estricta es aún más fuerte.

Términos clave

efecto causal dinámico (421)	error estándar consistente a heterocedasticidad y autocorrelación (HAC) (432)
modelo de retardos distribuidos (425)	parámetro de truncamiento (432)
exogeneidad (426)	estimador de varianza de Newey-West (433)
exogeneidad estricta (426)	mínimos cuadrados generalizados (MCG) (434)
multiplicador dinámico (429)	cuasi diferencia (435)
efecto impacto (429)	estimador MCG infactible (437)
multiplicador dinámico acumulativo (429)	estimador MCG factible (437)
multiplicador dinámico acumulativo de largo plazo (429)	

Revisión de conceptos

- 15.1 En la década de 1970 era una práctica común estimar un modelo de retardos distribuidos que relacionara las variaciones en el producto interior bruto nominal (Y) sobre las variaciones actuales y pasadas de la oferta de dinero (X). ¿Bajo qué supuestos estimaría esta regresión los efectos causales del dinero sobre el PIB nominal? ¿Se puede pensar que existe una probabilidad elevada de que se satisfagan estos supuestos en una economía moderna como es la de los Estados Unidos?
- 15.2 Supóngase que X es estrictamente exógena. Un investigador estima un modelo ARD(1,1), calcula los residuos de la regresión, y halla que los residuos presentan una correlación serial elevada. ¿Debería estimar el investigador un nuevo modelo ARD con retardos adicionales o simplemente utilizar errores estándar HAC para los coeficientes estimados del modelo ARD(1,1)?
- 15.3 Supóngase que se estima una regresión de retardos distribuidos, en la que la variable dependiente es ΔY_t en lugar de Y_t . Explique cómo se calcularían los multiplicadores dinámicos de X_t sobre Y_t .
- 15.4 Supóngase que se añade la variable IH_{t+1} como regresor adicional en la Ecuación (15.2). Si la variable IH es estrictamente exógena, ¿podría esperarse que el coeficiente de la variable IH_{t+1} fuera igual a cero o distinto de cero? ¿Cambiaría su respuesta si la variable IH fuera exógena, pero no estrictamente exógena?

Ejercicios

- 15.1 Los aumentos en los precios del petróleo han sido señalados como los culpables de que hayan ocurrido algunas recesiones económicas en los países desarrollados. Para cuantificar el efecto de los precios del petróleo sobre la actividad económica real, los investigadores han realizado regresiones como las descritas en este capítulo. Sea PIB_t la expresión del valor del producto interior bruto trimestral de Estados Unidos y sea $Y_t = 100 \ln(PIB_t/PIB_{t-1})$ la variación porcentual del PIB trimestral. James Hamilton, un experto en econometría y macroeconomía, sugirió que los precios del petróleo afectan negativamente a la economía solamente cuando ascienden por encima de sus valores pasados recientes. En concreto, sea O_t igual a la diferencia positiva expresada en puntos porcentuales entre los precios del petróleo en el periodo t y su valor máximo a lo largo del año anterior. Una regresión de retardos distribuidos que relaciona Y_t con O_t , estimada para el periodo 1955:I-2000:IV, proporciona los siguientes resultados

$$\begin{aligned}\hat{Y}_t = & 1,0 - 0,055O_t - 0,026O_{t-1} - 0,031O_{t-2} - 0,109O_{t-3} - 0,128O_{t-4} \\ & (0,1) \quad (0,054) \quad (0,057) \quad (0,048) \quad (0,042) \quad (0,053) \\ & + 0,008O_{t-5} + 0,025O_{t-6} - 0,019O_{t-7} + 0,067O_{t-8}. \\ & (0,025) \quad (0,048) \quad (0,039) \quad (0,042)\end{aligned}$$

- a) Supóngase que los precios del petróleo se sitúan un 25 % por encima de su valor máximo anterior y permanecen en este nuevo nivel más elevado (de modo que $O_t = 25$ y $O_{t+1} = O_{t+2} = \dots = 0$). ¿Cuál es el efecto previsto sobre el crecimiento de la producción para cada trimestre de los 2 años siguientes?
- b) Construya un intervalo de confianza del 95 % para las respuestas del apartado (a).
- c) ¿Cuál es la variación acumulada prevista en el crecimiento del PIB para ocho trimestres?
- d) El estadístico F HAC para el contraste de que los coeficientes de O_t y sus retardos son iguales a cero es 3,49. ¿Son los coeficientes significativamente distintos de cero?
- 15.2** Los macroeconomistas se han dado cuenta asimismo de que tras una subida de los precios del petróleo ocurren variaciones en los tipos de interés. Sea R_t el tipo de interés de las Letras del Tesoro a tres meses (en puntos porcentuales de tasa anual). La regresión de retardos distribuidos que relaciona la variación en R_t (ΔR_t) con O_t estimada para el periodo 1955:I-2000:IV es
- $$\begin{aligned}\widehat{\Delta R}_t = & 0,07 + 0,062O_t + 0,048O_{t-1} - 0,014O_{t-2} - 0,086O_{t-3} - 0,000O_{t-4} \\ & (0,06) \quad (0,045) \quad (0,034) \quad (0,028) \quad (0,169) \quad (0,058) \\ & + 0,023O_{t-5} - 0,010O_{t-6} - 0,100O_{t-7} - 0,014O_{t-8}. \\ & (0,065) \quad (0,047) \quad (0,038) \quad (0,025)\end{aligned}$$
- a) Supóngase que los precios del petróleo se sitúan en un nivel del 25 % por encima de su valor máximo anterior y permanecen en este nuevo nivel más elevado (de modo que $O_t = 25$ y $O_{t+1} = O_{t+2} = \dots = 0$). ¿Cuál es la variación prevista en los tipos de interés para cada trimestre de los 2 años siguientes?
- b) Construya un intervalo de confianza al 95 % para las respuestas del apartado (a).
- c) ¿Cuál es el efecto de esta variación de los precios del petróleo sobre el nivel de los tipos de interés en el periodo $t + 8$? ¿Cuál será la respuesta en relación a los multiplicadores acumulativos?
- d) El estadístico F HAC para contrastar si los coeficientes de O_t y de sus retardos son iguales a cero es de 4,25. ¿Son los coeficientes significativamente distintos de cero?
- 15.3** Considérense dos experimentos aleatorizados distintos. En el experimento A, los precios del petróleo se establecen de manera aleatoria y el banco central reacciona de acuerdo con sus habituales reglas de política en respuesta a las condiciones económicas, que incluyen las variaciones en el precio del petróleo. En el experimento B, los precios del petróleo se fijan de manera aleatoria y el banco central mantiene los tipos de interés constantes y en concreto no responde a las variaciones en el precio del petróleo. Se observa en ambos experimentos un crecimiento del PIB. Supóngase ahora que los precios del petróleo son exógenos en la regresión del Ejercicio 15.1. ¿A qué experimento, A o B, corresponde el efecto causal dinámico estimado en el Ejercicio 15.1?
- 15.4** Supóngase que los precios del petróleo son estrictamente exógenos. Analice cómo se pueden mejorar las estimaciones de los multiplicadores dinámicos del Ejercicio 15.1.
- 15.5** Deduzca la Ecuación (15.7) a partir de la Ecuación (15.4) y demuestre que, $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$ (etc.). (Sugerencia: téngase en cuenta que $X_t = \Delta X_t + \Delta X_{t-1} + \dots + \Delta X_{t-p+1} + X_{t-p}$).
- 15.6** Considérese el modelo de regresión $Y_t = \beta_0 + \beta_1 X_t + u_t$, donde u_t sigue un proceso estacionario de acuerdo con el modelo AR(1) $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ con \tilde{u}_t i.i.d. con media igual a 0, varianza igual a σ_u^2 , y $|\phi_1| < 1$, el regresor X_t , sigue un proceso AR(1) estacionario $X_t = \gamma_1 X_{t-1} + e_t$ con e_t i.i.d. con media igual a 0, varianza igual a σ_e^2 , y $|\gamma_1| < 1$, y e_t es independiente de \tilde{u}_t para todo $t \in \mathbb{Z}$.

- a) Demuestre que $\text{var}(u_t) = \frac{\sigma_u^2}{1 - \phi_1^2}$ y que $\text{var}(X_t) = \frac{\sigma_e^2}{1 - \gamma_1^2}$.
- b) Demuestre que $\text{cov}(u_t, u_{t-j}) = \phi_1^j \text{ var}(u_t)$ y que $\text{cov}(X_t, X_{t-j}) = \gamma_1^j \text{ var}(X_t)$.
- c) Demuestre que $\text{corr}(u_t, u_{t-j}) = \phi_1^j$ y que $\text{corr}(X_t, X_{t-j}) = \gamma_1^j$.
- d) Considérense los términos σ_v^2 y f_T de la Ecuación (15.14).
- Demuestre que $\sigma_v^2 = \sigma_X^2 \sigma_u^2$, donde σ_X^2 es la varianza de X , y σ_u^2 es la varianza de u .
 - Obtenga una expresión para f_∞ .
- 15.7** Considérese el modelo de regresión $Y_t = \beta_0 + \beta_1 X_t + u_t$, donde u_t sigue un proceso estacionario de acuerdo con el modelo AR(1) $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ con \tilde{u}_t i.i.d. con media igual a 0, varianza igual a $\sigma_{\tilde{u}}^2$ y, $|\phi_1| < 1$.
- a) Supóngase que X_t es independiente de \tilde{u}_t para todo t y j . ¿Es X_t exógena (pasada y presente)?
¿Es X_t estrictamente exógena (pasada, presente y futura)?
- b) Supóngase que $X_t = \tilde{u}_{t+1}$. ¿Es X_t estrictamente exógena?
- 15.8** Considérese el modelo del Ejercicio 15.7 con $X_t = \tilde{u}_{t+1}$.
- a) ¿Es consistente el estimador MCO de β_1 ? Explique la respuesta.
- b) Explique por qué el estimador MCG de β_1 no es consistente.
- c) Demuestre que el estimador MCG infactible $\hat{\beta}_1^{MCG} \xrightarrow{p} \beta_1 - \frac{\phi_1}{1 + \phi_1^2}$. [Sugerencia: utilice la fórmula (6.1) de variable omitida aplicada a la regresión de cuasi diferencias de la Ecuación (15.23)].
- 15.9** Considérese el modelo de regresión «solo con término constante» $Y_t = \beta_0 + u_t$, donde u_t sigue un proceso estacionario de acuerdo con el modelo AR(1) $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ con \tilde{u}_t i.i.d. con media igual a 0, varianza igual a $\sigma_{\tilde{u}}^2$ y $|\phi_1| < 1$.
- a) Demuestre que el estimador MCO es $\hat{\beta}_0 = T^{-1} \sum_{t=1}^T Y_t$.
- b) Demuestre que el estimador MCG (infactible) es $\hat{\beta}_0^{MCG} = (1 - \phi_1)^{-1} (T - 1)^{-1} \sum_{t=2}^{T-1} (Y_t - \phi_1 Y_{t-1})$. [Sugerencia: el estimador MCG de β_0 es $(1 - \phi_1)^{-1}$ multiplicado por el estimador MCO de α_0 de la Ecuación (15.23). ¿Por qué?]
- c) Demuestre que $\hat{\beta}_0^{MCG}$ se puede escribir como $\hat{\beta}_0^{MCG} = (T - 1)^{-1} \sum_{t=2}^{T-1} Y_t + (1 - \phi_1)^{-1} (T - 1)^{-1} (Y_T - \phi_1 Y_1)$. [Sugerencia: reordenar la fórmula del apartado (b)].
- d) Deduzca la diferencia $\hat{\beta}_0 - \hat{\beta}_0^{MCG}$ y analice por qué resulta probable que sea pequeña, cuando T es grande.
- 15.10** Considérese el modelo ARD $Y_t = 3,1 + 0,4 Y_{t-1} + 2,0 X_t - 0,8 X_{t-1} + \tilde{u}_t$, con X_t estrictamente exógena.
- a) Obtenga el efecto impacto de X sobre Y .
- b) Obtenga los cinco primeros multiplicadores dinámicos.
- c) Obtenga los cinco primeros multiplicadores acumulativos.
- d) Obtenga el multiplicador dinámico acumulativo de largo plazo.

Ejercicios empíricos

- E15.1** En este ejercicio se realizará una estimación del efecto de los precios del petróleo sobre la actividad macroeconómica a partir de datos mensuales sobre el Índice de Producción Industrial (IPI) (variable IP en la base de datos) y una medida mensual de la variable O_t descrita en el Ejercicio 15.1. Los datos pueden encontrarse en la página web del libro de texto http://www.pearsonhighered.com/stock_watson en el archivo **USMacro_Monthly**.
- a) Calcule la tasa de crecimiento mensual de la variable IPI, expresada en términos porcentuales, $\text{crec_ipi}_t = 100 \ln(IPI_t/IPI_{t-1})$. ¿Cuáles son la media y la desviación típica de la variable crec_ipi durante el periodo muestral 1952:1-2009:12?

- b) Represente gráficamente el valor de la variable O_t . ¿Por qué existen tantos valores de O_t iguales a cero? ¿Por qué no existen valores de O_t que sean negativos?
- c) Estime un modelo de retardos distribuidos para la variable *crec_ipi* sobre el valor actual y los 18 valores retardados de la variable O_t . ¿Qué valor del parámetro de truncamiento HAC estándar m elegiría? ¿Por qué?
- d) Considerados de forma conjunta, ¿son los coeficientes de O_t estadística y significativamente distintos de cero?
- e) Realice unos gráficos como los de la Figura 15.2 que muestren los multiplicadores dinámicos estimados, los multiplicadores acumulativos, y los intervalos de confianza al 95 %. Comente su opinión acerca del tamaño de los multiplicadores en la realidad.
- f) Supóngase que la elevada demanda de Estados Unidos (evidenciada por los elevados valores de la variable *crec_ipi*) conduce a un aumento en los precios del petróleo. ¿Es exógena la variable O_t ? ¿Resultan fiables los multiplicadores estimados que se muestran en los gráficos del apartado (e)? Explíquelo.

E15.2 En el archivo de datos **USMacro_Monthly**, se encuentran los datos de dos series agregadas de precios para Estados Unidos: el Índice de Precios al Consumidor (IPC, ICP en la base de datos) y el Deflactor del Gasto de Consumo Personal, variable DGCP (PCED en la base de datos). Estas series son medidas alternativas de los precios al consumo en los Estados Unidos. Los precios del IPC son los precios de una cesta de bienes cuya composición se actualiza cada 5-10 años. El DGCP utiliza una ponderación encadenada de los precios de una cesta de bienes cuya composición cambia mes a mes. Los economistas argumentan que el IPC exagera la inflación, ya que no tiene en cuenta la sustitución que se produce cuando los precios relativos cambian. Si este sesgo de sustitución es importante, la inflación media del IPC debe ser sistemáticamente más elevada que la inflación DGCP. Sea $\pi_t^{IPC} = 1.200 \times \ln [IPC(t)/IPC(t-1)]$, $\pi_t^{DGCP} = 1.200 \times \ln [DGCP(t)/DGCP(t-1)]$, e $Y_t = \pi_t^{IPC} - \pi_t^{DGCP}$, por lo que π_t^{IPC} es la tasa de inflación mensual de los precios (medida en términos porcentuales de tasa anual) basada en el IPC, π_t^{DGCP} es la tasa mensual de inflación de los precios a partir de la DGCP, e Y_t es la diferencia. Utilizando los datos desde 1970:1 hasta 2009:12, realice los siguientes ejercicios.

- a) Calcule las medias muestrales de las variables π_t^{IPC} y π_t^{DGCP} . ¿Son coherentes estas estimaciones puntuales con la presencia de un sesgo de sustitución en el IPC económicamente significativo?
- b) Calcule la media muestral de la variable Y_t . Explique por qué es numéricamente igual a la diferencia de las medias calculada en el apartado (a).
- c) Demuestre que la media poblacional de la variable Y es igual a la diferencia de las medias poblacionales entre las dos tasas de inflación.
- d) Considérese la regresión «solo con término constante»: $Y_t = \beta_0 + u_t$. Demuestre que $\beta_0 = E(Y)$. ¿Cree que u_t está serialmente correlacionada? Explíquelo.
- e) Construya un intervalo de confianza al 95 % para β_0 . ¿Qué valor del parámetro de truncamiento HAC estándar, m , elegiría? ¿Por qué?
- f) ¿Existe evidencia estadísticamente significativa de que la tasa media de inflación para el IPC sea mayor que la tasa de inflación para el DGCP?
- g) ¿Existe evidencia de inestabilidad en β_0 ? Lleve a cabo un contraste QLR.

APÉNDICE

15.1 La base de datos del zumo de naranja

Los datos sobre los precios del zumo de naranja son la componente del zumo de naranja congelado del grupo de alimentación y alimentos procesados del Índice de Precios al Productor (IPP), recogidos por la Oficina de Estadísticas Laborales de EE.UU. (BLS series wpu02420301). La serie de precios del zumo de naranja fue dividida por el IPP general de productos terminados para ajustar por la inflación del nivel general de precios. La serie del índice de heladas fue construida a partir de las temperaturas mínimas diarias registradas en los aeropuertos del área de Orlando, obtenidas de la Administración Nacional Oceánica y Atmosférica (NOAA) del Departamento de Comercio de los EE.UU. La serie IH fue construida de manera que su calendario y el calendario de los datos de precios del zumo de naranja estuvieran aproximadamente sincronizados. En concreto, los datos sobre el precio del zumo de naranja congelado son recogidos por una encuesta realizada sobre una muestra de productores a mediados de cada mes, aunque la fecha exacta varía de mes en mes. En consecuencia, la serie de la variable IH fue construida para que fuera el índice de heladas desde el día 11 de un mes hasta el día 10 del mes siguiente, es decir, IH es el máximo entre cero y 32, menos la temperatura mínima diaria, sumada para todos los días entre el 11 y el 10. Por tanto, la variable $\%VP_t$ de febrero es la variación porcentual en los precios reales del zumo de naranja desde mediados de enero hasta mediados de febrero, y la variable IH_t en febrero es el índice de heladas desde el 11 de enero hasta el 10 de febrero.

APÉNDICE

15.2 Modelo ARD y mínimos cuadrados generalizados en notación del operador de retardos

En este apéndice se presenta el modelo de retardos distribuidos en notación del operador de retardos, se obtienen las representaciones ARD y en cuasi diferencias del modelo de retardos distribuidos, y se analizan las condiciones bajo las cuales el modelo ARD puede tener un menor número de parámetros que el modelo de retardos distribuidos original.

Los modelos de retardos distribuidos, ARD, y en cuasi diferencias, en notación del operador de retardos

Tal como se define en el Apéndice 14.3, el operador de retardo, L , tiene la propiedad de que $L^j X_t = X_{t-j}$, y los retardos distribuidos $\beta_1 X_t + \beta_2 X_{t-1} + \dots + \beta_r X_{t-r}$ se puede expresar como $\beta(L)X_t$, donde $\beta(L) = \sum_{j=0}^r \beta_j L^j$, con $L^0 = 1$. Por tanto, el modelo de retardos distribuidos del Concepto clave 15.1 [Ecuación (15.4)] se puede escribir en notación del operador de retardos como

$$Y_t = \beta_0 + \beta(L)X_t + u_t. \quad (15.40)$$

Además, si el término de error u_t sigue un proceso $AR(p)$, entonces se puede escribir como

$$\phi(L)u_t = \tilde{u}_t, \quad (15.41)$$

donde $\phi(L) = \sum_{j=0}^p \phi_j L^j$, donde $\phi_0 = 1$ y \tilde{u}_t está serialmente incorrelacionado [téngase en cuenta que ϕ_1, \dots, ϕ_p tal y como se definen aquí son los opuestos (cambiados de signo) de ϕ_1, \dots, ϕ_p en la notación de la Ecuación (15.31)].

Para obtener el modelo ARD, se premultiplica cada lado de la Ecuación (15.40) por por $\phi(L)$ lo que

$$\phi(L)Y_t = \phi(L)[\beta_0 + \beta(L)X_t + u_t] = \alpha_0 + \delta(L)X_t + \tilde{u}_t, \quad (15.42)$$

donde

$$\alpha_0 = \phi(1)\beta_0 \text{ y } \delta(L) = \phi(L)\beta(L), \text{ donde } \phi(1) = \sum_{j=0}^p \phi_j. \quad (15.43)$$

Para obtener el modelo en cuasi diferencias, debe tenerse en cuenta que $\phi(L)\beta(L)X_t = \beta(L)\phi(L)X_t = \beta(L)\tilde{X}_t$, donde $\tilde{X}_t = \phi(L)X_t$. Por tanto reordenando la Ecuación (15.42) se obtiene

$$\tilde{Y}_t = \alpha_0 + \beta(L)\tilde{X}_t + \tilde{u}_t, \quad (15.44)$$

donde \tilde{Y}_t es la cuasi diferencia de Y_t ; es decir, $\tilde{Y}_t = \phi(L)Y_t$.

Los estimadores ARD y MCG

El estimador MCO de los coeficientes ARD se obtiene mediante la estimación MCO de la Ecuación (15.42). Los coeficientes originales de retardos distribuidos son $\beta(L)$, los cuales, en términos de coeficientes estimados, son $\hat{\beta}(L) = \delta(L)/\phi(L)$; es decir, los coeficientes de $\beta(L)$ satisfacen las restricciones implícitas en $\phi(L)\beta(L) = \delta(L)$. Por tanto, el estimador de los multiplicadores dinámicos basado en los estimadores MCO de los coeficientes del modelo ARD, $\hat{\delta}(L)$ y $\hat{\phi}(L)$, es

$$\hat{\beta}^{ARD}(L) = \frac{\hat{\delta}(L)}{\hat{\phi}(L)} \quad (15.45)$$

Las expresiones de los coeficientes de la Ecuación (15.29) del texto se han obtenido como un caso particular de la Ecuación (15.45) cuando $r = 1$ y $p = 1$.

El estimador MCG factible se calcula mediante la obtención de un estimador preliminar de $\phi(L)$, calculando las cuasi diferencias estimadas, estimando $\beta(L)$ en la Ecuación (15.44) mediante estas cuasi diferencias estimadas, y (si se desea) iterando hasta lograr la convergencia. El estimador MCG iterado es el estimador MCNL calculado mediante la estimación MCNL del modelo ARD de la Ecuación (15.42), sujeto a las restricciones no lineales sobre los parámetros contenidas en la Ecuación (15.43).

Tal y como se subrayó en el debate en torno a la Ecuación (15.36) de este capítulo, no es suficiente con que X_t sea exógena (pasada y presente) para poder utilizar cualquiera de estos métodos de estimación, la exogeneidad por sí sola no asegura que la Ecuación (15.36) se cumpla. Sin embargo, si X es estrictamente exógena, entonces la Ecuación (15.36) se cumple, y, suponiendo que se cumplen los supuestos 2 a 4 del Concepto clave 14.6, estos estimadores son consistentes y asintóticamente normales. Por otra parte, los errores estándar MCO habituales (heterocedástico-robustos para sección cruzada) proporcionan una base válida para la inferencia estadística.

Reducción del número de parámetros mediante el modelo ARD. Supongamos que el polinomio de retardos distribuidos $\beta(L)$ se puede expresar como un cociente de polinomios de retardos, $\theta_1(L)/\theta_2(L)$, donde tanto $\theta_1(L)$ como $\theta_2(L)$ son polinomios de retardos de grado bajo. Por tanto, $\phi(L)\beta(L)$ en la Ecuación (15.43) es $\phi(L)\beta(L) = \phi(L)\theta_1(L)/\theta_2(L) = [\phi(L)/\theta_2(L)]\theta_1(L)$. Si se da la circunstancia de que $\phi(L) = \theta_2(L)$, entonces $\delta(L) = \phi(L)\beta(L) = \theta_1(L)$. Si el grado de $\theta_1(L)$ es bajo, entonces q , el número de retardos de X_t en el modelo ARD, puede ser mucho menor que r . Por tanto, bajo estos supuestos, la estimación del modelo ARD implica estimar potencialmente menos parámetros que el modelo de retardos distribuidos original. En este sentido el modelo ARD puede lograr parametrizaciones más parsimoniosas (es decir, utiliza menos parámetros desconocidos) que el modelo de retardos distribuidos.

Tal y como se ha desarrollado aquí, el supuesto de que ocurre que $\phi(L)$ y $\theta_2(L)$ son iguales parece una coincidencia que no ocurriría en una aplicación. Sin embargo, el modelo ARD es capaz de captar un gran número de formas de los multiplicadores dinámicos con tan solo unos pocos coeficientes.

ARD o MCG: sesgo versus varianza. Una buena manera de pensar acerca de si estimar multiplicadores dinámicos estimando en primer lugar un modelo ARD y posteriormente calculando los multiplicadores dinámicos a partir de los coeficientes ARD o, alternativamente, mediante la estimación directa del modelo de retardos distribuidos mediante MCG consiste en interpretar la decisión en términos de un equilibrio entre sesgo y varianza. La estimación de los multiplicadores dinámicos mediante un modelo ARD aproximado introduce un sesgo; sin embargo, debido a que existen pocos coeficientes, la varianza del estimador de los multiplicadores dinámicos puede ser pequeña. Por contra, la estimación de un modelo largo de retardos distribuidos mediante MCG da lugar a un menor sesgo en los multiplicadores; sin embargo, debido a que existen tantos coeficientes, su varianza puede ser grande. Si la aproximación ARD para los multiplicadores dinámicos es buena, entonces el sesgo de los multiplicadores dinámicos implicados será pequeño, por lo que el método ARD tendrá una menor varianza que el método MCG con solamente un pequeño aumento en el sesgo. Por esta razón, la estimación sin restricciones de un modelo ARD con un pequeño número de retardos de Y y X es una forma atractiva de aproximar una distribución de retardos larga cuando X es estrictamente exógena.

Otros temas relacionados con la regresión en series temporales

Este capítulo recoge algunos temas adicionales relacionados con la regresión de series temporales, comenzando con la predicción. El Capítulo 14 consideraba la predicción de una sola variable. Sin embargo, en la práctica podría ocurrir que se deseé predecir dos o más variables, como la tasa de inflación y la tasa de crecimiento del PIB. La Sección 16.1 introduce un modelo para predecir varias variables, los vectores autorregresivos (VAR's), en los que se utilizan los valores retardados de dos o más variables para predecir los valores futuros de esas variables. El Capítulo 14 se centraba asimismo en la formulación de predicciones sobre un periodo futuro (por ej. sobre un trimestre) pero la formulación de predicciones con un horizonte de dos, tres, o más periodos futuros es igualmente importante. Los métodos para la realización de predicciones multiperiodo se tratan en la Sección 16.2.

Las Secciones 16.3 y 16.4 regresan a la cuestión de la Sección 14.6, las tendencias estocásticas. La Sección 16.3 introduce modelos adicionales de tendencias estocásticas y un contraste de raíces unitarias alternativo. En la Sección 16.4 se introduce el concepto de cointegración, que aparece cuando dos variables comparten una tendencia estocástica común, es decir, cuando cada una de las variables contiene una tendencia estocástica, pero existe alguna diferencia ponderada de las dos variables que no la contiene.

En algunos datos de series temporales, especialmente en el caso de datos financieros, la varianza cambia a lo largo del tiempo: algunas veces las series presentan una elevada volatilidad, mientras que otras veces la volatilidad es baja, por lo que los datos presentan agrupaciones de volatilidad. La Sección 16.5 analiza la volatilidad agrupada y en ella se introducen modelos en los que la varianza del error de predicción varía a lo largo del tiempo, es decir, modelos en los que el error de predicción es condicionalmente heterocedástico. Los modelos de heterocedasticidad condicional tienen varias aplicaciones. Una de las aplicaciones es el cálculo de intervalos de predicción, en los que la anchura del intervalo varía a lo largo del tiempo de forma que refleja los períodos con alta o baja incertidumbre. Otra aplicación es la predicción de la incertidumbre acerca de la rentabilidad de un activo, como puede ser una acción, lo que a su vez puede ser útil para evaluar el riesgo de tenencia de un activo.

16.1 Vectores autorregresivos

El Capítulo 14 se centraba en la predicción de la tasa de inflación, pero en la realidad los expertos en predicción asimismo se dedican a predecir otras variables macroeconómicas clave, como son la tasa de desempleo, la tasa de crecimiento del PIB, y los tipos de interés. Una manera de hacerlo consiste en desarrollar un modelo de predicción para cada una de las variables por separado mediante los procedimientos de la Sección 14.4. Otra manera consiste en desarrollar un único modelo que pueda predecir todas las variables, lo que puede ayudar a que las predicciones sean mutuamente consistentes. Un método para predecir variables con un único modelo consiste en utilizar vectores autorregresivos (VAR). Un VAR amplía un modelo autorregresivo univariante al caso de varias variables de series temporales, es decir, extiende el modelo autorregresivo univariante a un «vector» de variables de series temporales.

El modelo VAR

Un **vector autorregresivo** (VAR) con dos variables de series temporales, Y_t y X_t , consta de dos ecuaciones: en una de ellas, la variable dependiente es Y_t ; en la otra, la variable dependiente es X_t . Los regresores de ambas ecuaciones son los valores retardados de ambas variables. En términos más generales, un VAR con k variables de series temporales consta de k ecuaciones, una para cada una de las variables; donde las variables explicativas de todas las ecuaciones son los valores retardados de todas las variables. Los coeficientes del VAR se estiman mediante la estimación de cada una de las ecuaciones por MCO.

Los VAR se recogen en el Concepto clave 16.1.

Inferencia en VAR. Bajo los supuestos de VAR, los estimadores MCO son consistentes y tienen una distribución conjunta normal en muestras grandes. En consecuencia, la inferencia estadística se lleva a cabo de la forma habitual; por ejemplo, los intervalos de confianza al 95 % sobre los coeficientes pueden construirse como el coeficiente estimado $\pm 1,96$ errores estándar.

En los VAR aparece un nuevo aspecto de los contrastes de hipótesis debido a que un VAR con k variables es una colección, o sistema, de k ecuaciones. Por tanto es posible contrastar hipótesis conjuntas que incluyen restricciones sobre varias ecuaciones.

Por ejemplo, en el VAR(p) de dos variables de las Ecuaciones (16.1) y (16.2), cabe preguntarse si la longitud correcta de los retardos es p , o $p - 1$; es decir, cabría preguntarse si los coeficientes de Y_{t-p} y X_{t-p} son iguales a cero en esas dos ecuaciones. La hipótesis nula de que estos coeficientes son iguales a cero es

$$H_0: \beta_{1p} = 0, \beta_{2p} = 0, \gamma_{1p} = 0, \text{ y } \gamma_{2p} = 0. \quad (16.3)$$

La hipótesis alternativa es que al menos uno de estos cuatro coeficientes es distinto de cero. Por tanto, la hipótesis nula incluye coeficientes de *ambas* ecuaciones, dos de cada una de las ecuaciones.

Debido a que los coeficientes estimados tienen una distribución conjunta normal en muestras grandes, es posible contrastar las restricciones sobre estos coeficientes mediante el cálculo de un estadístico F . La fórmula exacta de este estadístico es complicada debido a que la notación debe manejar varias ecuaciones, por lo que la omitimos. En la práctica, los paquetes de *software* más modernos cuentan con procedimientos automatizados para la contrastación de hipótesis sobre los coeficientes de los sistemas de varias ecuaciones.

¿Cuántas variables deben incluirse en un VAR? El número de coeficientes en cada una de las ecuaciones de un VAR es proporcional al número de variables en el VAR. Por ejemplo, un VAR con cinco variables y cuatro retardos tendrá 21 coeficientes (cuatro retardos de cada una de las cinco variables, ade-

**CONCEPTO
CLAVE**

16.1

Vectores autorregresivos

Un vector autorregresivo (VAR) es un conjunto de k regresiones de series temporales, en las que los regresores son los valores retardados de las k series. Un VAR extiende el modelo autorregresivo univariante a una lista, o «vector», de variables de series temporales. Cuando el número de retardos en cada una de las ecuaciones es el mismo y es igual a p , el sistema de ecuaciones se denomina VAR(p).

En el caso de dos variables de series temporales, Y_t y X_t , el VAR(p) consta de dos ecuaciones

$$Y_t = \beta_{10} + \beta_{11}Y_{t-1} + \cdots + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1} + \cdots + \gamma_{1p}X_{t-p} + u_{1t} \quad (16.1)$$

$$X_t = \beta_{20} + \beta_{21}Y_{t-1} + \cdots + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1} + \cdots + \gamma_{2p}X_{t-p} + u_{2t}, \quad (16.2)$$

donde las β y las γ son coeficientes desconocidos y u_{1t} y u_{2t} son los términos de error.

Los supuestos VAR son los supuestos de regresión de series temporales del Concepto clave 14.6, aplicados a cada una de las ecuaciones. Los coeficientes de un VAR se calculan mediante la estimación de cada una de las ecuaciones por MCO.

más del término independiente) en cada una de las cinco ecuaciones, para un total de ¡105 coeficientes! La estimación de todos estos coeficientes aumenta la cuantía del error de estimación de una predicción, lo cual puede dar lugar a un deterioro de la precisión de la predicción.

La implicación práctica es que resulta necesario mantener controlado el número de variables de un VAR y, sobre todo, asegurarse de que las variables están relacionadas unas con otras de forma verosímil para que puedan resultar útiles para el pronóstico del resto. Por ejemplo, sabemos a partir de una combinación de evidencia empírica (como la discutida en el Capítulo 14) y teoría económica, que la tasa de inflación, la tasa de desempleo, y el tipo de interés de corto plazo están relacionados entre sí, lo que sugiere que estas variables podrían ayudar a pronosticarse unas a otras en un VAR. La inclusión de una variable que no esté relacionada con el resto en un VAR, no obstante, introduce un error de estimación, sin que añada un contenido predictivo, lo que reduce la precisión de la predicción.

Determinación de la longitud de los retardos en los VAR¹. La longitud de los retardos en un VAR puede determinarse utilizando tanto los contrastes F como los criterios de información.

El criterio de información para un sistema de ecuaciones amplía el criterio de información para una única ecuación de la Sección 14.5. Para definir este criterio de información es necesario adoptar la notación matricial. Sea Σ_u la matriz de covarianzas $k \times k$ de los errores VAR y sea $\hat{\Sigma}_u$ la matriz de covarianzas estimada, en la que el elemento i, j de $\hat{\Sigma}_u$ es $\frac{1}{T} \sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}$, donde \hat{u}_{it} es el residuo MCO de la ecuación i -ésima y \hat{u}_{jt} es el residuo MCO de la ecuación j -ésima. El criterio BIC para el VAR es

$$\text{BIC}(p) = \ln [\det(\hat{\Sigma}_u)] + k(kp + 1) \frac{\ln(T)}{T} \quad (16.4)$$

donde $\det(\hat{\Sigma}_u)$ es el determinante de la matriz $\hat{\Sigma}_u$. El criterio AIC se calcula utilizando la Ecuación (16.4), modificada mediante la sustitución del término « $\ln(T)$ » por « 2 ».

La expresión del BIC para las k ecuaciones del VAR de la Ecuación (16.4) amplía la expresión para una única ecuación de la Sección 14.5. Si existe una única ecuación, el primer término se simplifica a $\ln [SR(p)/T]$. El segundo término de la Ecuación (16.4) es la penalización por la adición de variables explicativas adicionales; $k(kp + 1)$ es el número total de coeficientes de regresión del VAR (hay k ecuaciones, cada una de las cuales tiene un término independiente y p retardos de cada una de las k variables de series de temporales).

La estimación de la longitud de los retardos en un VAR mediante el criterio BIC se lleva a cabo de forma análoga al caso de una sola ecuación: de entre un conjunto de valores candidatos de p , la longitud estimada de los retardos es el valor de p que minimiza $\text{BIC}(p)$.

Utilización de VAR para el análisis causal. Hasta ahora el análisis se ha centrado en el uso de los VAR para la predicción. Otra utilidad de los modelos VAR consiste en analizar las relaciones causales que existen entre las variables económicas de series temporales; de hecho, fue para este propósito para lo que los VAR se introdujeron por primera vez en economía por el económetra y macroeconomista Christopher Sims (1980). La utilización de los VAR para la inferencia causal se conoce como modelización VAR estructural; estructural, debido a que en esta aplicación los VAR se utilizan para modelizar la estructura subyacente de la economía. El análisis VAR estructural utiliza las técnicas descritas en esta sección en el contexto de la predicción, además de algunas herramientas adicionales. No obstante, la principal diferencia conceptual entre el uso de los VAR para predicción y su uso para la modelización estructural es que la modelización estructural requiere de supuestos muy específicos, obtenidos a partir de la teoría económica y el conocimiento institucional, acerca de lo que es exógeno y de lo que no lo es. El análisis de los VAR estructurales se lleva a cabo de mejor modo en el contexto de la estimación de sistemas de ecuaciones simultáneas, que queda más allá del alcance de este libro. Para una introducción a la utilización de los VAR para la predicción y el análisis de políticas, véase Stock y Watson (2001). Para los detalles matemáticos adicionales sobre la modelización VAR estructural, véase Hamilton (1994) o Watson (1994).

¹ En esta sección se utilizan matrices y puede ser omitida en el caso de enfoques menos matemáticos.

Un modelo VAR para las tasas de inflación y desempleo

A modo de ejemplo, consideremos un VAR de dos variables para la tasa de inflación, Inf_t , y la tasa de desempleo, $Desemp_t$. Al igual que en el Capítulo 14, se considera que la tasa de inflación presenta una tendencia estocástica, por lo que resulta apropiado transformarla mediante el cálculo de su primera diferencia, ΔInf_t .

El VAR para las variables ΔInf_t y $Desemp_t$ consta de dos ecuaciones: una en la que la variable ΔInf_t es la variable dependiente, y otra en la que la variable $Desemp_t$ es la variable dependiente. En ambas ecuaciones las variables explicativas son los valores retardados de ΔInf_t y $Desemp_t$. A causa de la aparente existencia de un cambio estructural en la curva de Phillips en los primeros años de la década de 1980 hallado en la Sección 14.7 mediante el contraste QLR, el VAR se estima utilizando los datos desde 1982:I hasta 2004:IV.

La primera ecuación del VAR es la ecuación de la inflación:

$$\widehat{\Delta Inf_t} = 1,47 - 0,64\Delta Inf_{t-1} - 0,64\Delta Inf_{t-2} - 0,13\Delta Inf_{t-3} - 0,13\Delta Inf_{t-4} \\ (0,55) \quad (0,12) \quad (0,10) \quad (0,11) \quad (0,09) \\ - 3,49Desemp_{t-1} + 2,80Desemp_{t-2} + 2,44Desemp_{t-3} - 2,03Desemp_{t-4}. \quad (16.5) \\ (0,58) \quad (0,94) \quad (1,07) \quad (0,55)$$

El R^2 ajustado es $\bar{R}^2 = 0,44$.

La segunda ecuación del VAR es la ecuación del desempleo, en la que las variables explicativas son las mismas que en la ecuación de la inflación, pero la variable dependiente es la tasa de desempleo:

$$\widehat{Desemp_t} = 0,22 + 0,005\Delta Inf_{t-1} + 0,004\Delta Inf_{t-2} - 0,007\Delta Inf_{t-3} - 0,003\Delta Inf_{t-4} \\ (0,12) \quad (0,017) \quad (0,018) \quad (0,018) \quad (0,014) \\ + 1,52Desemp_{t-1} - 0,29Desemp_{t-2} - 0,43Desemp_{t-3} + 0,16Desemp_{t-4}. \quad (16.6) \\ (0,11) \quad (0,18) \quad (0,21) \quad (0,11)$$

El R^2 ajustado es $\bar{R}^2 = 0,982$.

Las Ecuaciones (16.5) y (16.6), consideradas de forma conjunta, son un modelo VAR(4) de la variación de la tasa de inflación, ΔInf_t , y la tasa de desempleo, $Desemp_t$.

Estas ecuaciones VAR pueden utilizarse para llevar a cabo un contraste de causalidad de Granger. El estadístico F para el contraste de la hipótesis nula de que los coeficientes de $Desemp_{t-1}$, $Desemp_{t-2}$, $Desemp_{t-3}$, y $Desemp_{t-4}$ son iguales a cero en la ecuación de la inflación [Ecuación (16.5)] toma el valor 11,04, cuyo p -valor es menor que 0,001. Por lo tanto se rechaza la hipótesis nula, por lo que se puede concluir que la tasa de desempleo es un predictor útil de las variaciones de la inflación, dados los retardos de la inflación (es decir, la tasa de desempleo causa en el sentido de Granger a las variaciones de la inflación). El estadístico F para el contraste de la hipótesis de que los coeficientes de los cuatro retardos de la variable ΔInf_t son iguales a cero en la ecuación del desempleo [Ecuación (16.6)] es igual a 0,16, con un p -valor de 0,96. Por lo que la variación de la tasa de inflación no causa en el sentido de Granger la tasa de desempleo al nivel de significación del 10 %.

Las predicciones para las tasas de inflación y desempleo para un horizonte temporal de un periodo se obtienen exactamente tal y como se indicaba en la Sección 14.4. La predicción de la variación de la inflación desde 2004:IV hasta 2005:I, en base a la Ecuación (16.5), es $\widehat{\Delta Inf}_{2005:1|2004:IV} = -0,1$ puntos porcentuales. Un cálculo similar utilizando la Ecuación (16.6) proporciona una predicción de la tasa de desempleo para 2005:I en base a los datos hasta 2004:IV de $\widehat{Desemp}_{2005:1|2004:1} = 5,4\%$, muy cercana a su valor real, $Desemp_{2005:1} = 5,3\%$.

16.2 Predicciones multiperíodo

El análisis de predicción se ha centrado hasta el momento en la formulación de predicciones a un periodo visto. No obstante, a menudo los profesionales que se dedican a realizar predicciones están llamados a

realizar pronósticos con un horizonte que va más allá de un solo periodo futuro. En esta sección se describen dos métodos para la realización de predicciones para un horizonte de varios periodos. El método habitual consiste en la realización de predicciones «iteradas», en el que un modelo de predicción a un periodo vista se repite cada vez que se adelanta la predicción un periodo más hacia adelante, de la forma que se especifica más adelante en esta sección. El segundo método consiste en hacer predicciones «directas», mediante una regresión en la que la variable dependiente es la variable multiperiodo para la que se quiere realizar la predicción. Por las razones que se discuten al final de esta sección, en la mayoría de las aplicaciones el método iterativo es más recomendable que el método directo.

Predicciones multiperiodo iteradas

La idea esencial de una predicción iterada es la de utilizar un modelo de predicción para realizar una predicción con un horizonte temporal de un periodo, para el periodo $T + 1$ se utilizan los datos hasta el periodo T . A continuación se utiliza el modelo para realizar un pronóstico para el periodo $T + 2$ dados los datos disponibles hasta T , donde el valor previsto para $T + 1$ se considera como si fuera un dato más a los efectos de realizar el pronóstico para el periodo $T + 2$. Por tanto la predicción a un periodo vista (que se conoce asimismo como predicción un paso más allá) se utiliza como paso intermedio para realizar a su vez la predicción a un horizonte temporal de dos periodos. Este proceso se repite, o itera, hasta que se realiza la predicción para el horizonte de pronóstico deseado h .

El método de predicción iterativo AR: AR(1). Una predicción iterativa AR(1) utiliza un patrón AR(1) para el modelo de un horizonte temporal de un periodo vista. Por ejemplo, consideremos el proceso autorregresivo de primer orden para la variable ΔInf_t [Ecuación (14.7)]:

$$\widehat{\Delta Inf}_t = 0,02 - 0,24\Delta Inf_{t-1}. \quad (16.7)$$

(0,13) (0,10)

El primer paso para calcular la predicción de la variable $\Delta Inf_{2005:II}$ con un horizonte de dos trimestres vista en base a la Ecuación (16.7), con datos hasta 2004:IV consiste en calcular la predicción con un horizonte temporal de un periodo vista de $\Delta Inf_{2005:I}$ en base a los hasta 2004:IV: $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0,02 - 0,24\Delta Inf_{2004:IV} = 0,02 - 0,24 \times 1,9 = -0,4$. El segundo paso consiste en sustituir esta predicción en la Ecuación (16.7) de modo que $\widehat{\Delta Inf}_{2005:II|2004:IV} = 0,02 - 0,24\widehat{\Delta Inf}_{2005:I|2004:IV} = 0,02 - 0,24 \times (-0,4) = 0,1$. Por lo tanto, en base a la información disponible hasta el cuarto trimestre de 2004, esta predicción establece que la tasa de inflación se incrementará en 0,1 puntos porcentuales entre el primer y el segundo trimestre de 2005.

El método de predicción iterativo AR: AR(p). La estrategia del AR(1) iterado es extensible a un AR(p) mediante la sustitución de Y_{T+1} por su pronóstico $\widehat{Y}_{T+1|T}$, y considerando posteriormente la predicción como un dato a efectos de la predicción AR(p) de Y_{T+2} . Por ejemplo, consideremos la predicción iterada con un horizonte temporal de dos periodos vista de la inflación basada en el modelo AR(4) de la Sección 14.3 [Ecuación (14.13)]:

$$\widehat{\Delta Inf}_t = 0,02 - 0,26\Delta Inf_{t-1} - 0,32\Delta Inf_{t-2} + 0,16\Delta Inf_{t-3} - 0,03\Delta Inf_{t-4}. \quad (16.8)$$

(0,12) (0,09) (0,08) (0,08) (0,09)

La predicción de la variable $\Delta Inf_{2005:I}$ en base a los datos disponibles hasta 2004:IV mediante este modelo AR(4), calculado en la Sección 14.3, es $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0,4$. Por tanto la predicción para un horizonte temporal de dos trimestres vista basada en el AR(4) es $\widehat{\Delta Inf}_{2005:II|2004:IV} = 0,02 - 0,26\widehat{\Delta Inf}_{2005:I|2004:IV} - 0,32\Delta Inf_{2004:IV} + 0,16\Delta Inf_{2004:III} - 0,03\Delta Inf_{2004:II} = 0,02 - 0,26 \times 0,4 - 0,32 \times 1,9 + 0,16 \times (-2,8) - 0,08 \times 0,6 = -1,1$. De acuerdo con esta predicción AR(4) iterada, basada en los datos disponibles hasta el cuarto trimestre de 2004, se prevé que la tasa de la inflación disminuya en 1,1 puntos porcentuales entre el primer y el segundo trimestre de 2005.

Predicciones iteradas multivariantes mediante un VAR iterado. Las predicciones iteradas multivariantes se pueden calcular utilizando un VAR del mismo modo que se calculan las predicciones iteradas univariantes por medio de un modelo autorregresivo. La principal característica novedosa de una predicción iterada multivariante es que la predicción con un horizonte temporal de dos periodos vista (periodo $T + 2$) de una variable depende de las predicciones de todas las variables del VAR para el periodo $T + 1$. Por ejemplo, para calcular la predicción acerca de la variación desde el periodo $T + 1$ al periodo $T + 2$ mediante un VAR con las variables ΔInf_t y $Desemp_t$, se podría predecir tanto ΔInf_{T+1} como $Desemp_{T+1}$ utilizando los datos disponibles hasta el periodo T como un paso intermedio en el pronóstico de ΔInf_{T+2} . De manera más general, para calcular las predicciones VAR multiperiodo iteradas con un horizonte temporal de h periodos vista, es necesario calcular las predicciones de todas las variables para todos los periodos intermedios entre T y $T + h$.

A modo de ejemplo, vamos a calcular la predicción VAR iterada de la variable $\Delta Inf_{2005:II}$ en base a los datos disponibles hasta 2004:IV mediante el VAR(4) para las variables ΔInf_t y $Desemp_t$ de la Sección 16.1 [Ecuaciones (16.5) y (16.6)]. El primer paso consiste en calcular la predicción con un horizonte temporal de un trimestre vista de $\widehat{\Delta Inf}_{2005:I|2004:IV}$ y $\widehat{Desemp}_{2005:I|2004:IV}$ a partir del VAR. La predicción $\widehat{\Delta Inf}_{2005:I|2004:IV}$ basada en la Ecuación (16.5) se calculó en la Sección 14.3 y es $-0,1$ puntos porcentuales [Ecuación (14.18)]. Un cálculo similar utilizando la Ecuación (16.6) muestra que $\widehat{Desemp}_{2005:I|2004:IV} = 5,4\%$. En el segundo paso, se sustituyen estas predicciones en las Ecuaciones (16.5) y (16.6) para dar lugar a la predicción a dos trimestres vista, $\widehat{\Delta Inf}_{2005:II|2004:IV}$:

$$\begin{aligned}\widehat{\Delta Inf}_{2005:II|2004:IV} &= 1,47 - 0,64\widehat{\Delta Inf}_{2005:I|2004:IV} - 0,64\Delta Inf_{2004:IV} - 0,13\Delta Inf_{2004:III} \\ &\quad - 0,13\Delta Inf_{2004:II} - 3,49\widehat{Desemp}_{2005:I|2004:IV} + 2,80Desemp_{2004:IV} \\ &\quad + 2,44Desemp_{2004:III} - 2,03Desemp_{2004:II} \\ &= 1,47 - 0,64 \times (-0,1) - 0,64 \times 1,9 - 0,13 \times (-2,8) - 0,13 \times 0,6 \\ &\quad - 3,49 \times 5,4 + 2,80 \times 5,4 + 2,44 \times 5,4 - 2,03 \times 5,6 = -1,1.\end{aligned}\tag{16.9}$$

Por tanto, la predicción VAR(4) iterada, basada en los datos disponibles hasta el cuarto trimestre de 2004, es que la inflación se reducirá en 1,1 puntos porcentuales entre el primer y el segundo trimestre de 2005.

Las predicciones multiperiodo iteradas se recogen en el Concepto clave 16.2.

Predicciones multiperiodo directas

Las predicciones multiperiodo directas se calculan sin iterar utilizando una única regresión en la que la variable dependiente es la variable en un horizonte temporal de varios periodos vista que se pretende predecir y los regresores son las variables predictoras. Las predicciones calculadas de esta forma se denominan predicciones directas debido a que se pueden utilizar directamente los coeficientes de regresión para realizar la predicción multiperiodo.

El método de predicción multiperiodo directo. Supongamos que se desea realizar una predicción de Y_{T+2} mediante de la utilización de los datos disponibles hasta el periodo temporal T . El método directo multivariante toma como punto de partida el modelo ARD, pero retarda las variables predictoras en un periodo adicional. Por ejemplo, si se utilizan dos retardos de los predictores, entonces la variable dependiente es Y_t y los regresores son, Y_{t-2} , Y_{t-3} , X_{t-2} , y X_{t-3} . Los coeficientes de esta regresión se pueden utilizar directamente para calcular la predicción de Y_{T+2} utilizando los datos sobre Y_T , Y_{T-1} , X_T , y X_{T-1} , sin necesitar ninguna iteración. De manera más general, en una regresión de predicción directa con un horizonte temporal de h -periodos vista, todos los predictores se retardan h periodos para generar la predicción a h periodos vista.

**CONCEPTO
CLAVE**
16.2
Predicciones multiperiodo iteradas

La **predicción multiperiodo iterada AR** se calcula por pasos: en primer lugar se calcula la predicción con un horizonte temporal de un periodo visto, y más tarde esta se utiliza para calcular la predicción a dos periodos vista, etc. Las predicciones iteradas con un horizonte de dos y tres períodos vista basadas en un AR(p) son

$$\hat{Y}_{T+2|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T+1|T} + \hat{\beta}_2 Y_T + \hat{\beta}_3 Y_{T-1} + \dots + \hat{\beta}_p Y_{T-p+2} \quad (16.10)$$

$$\hat{Y}_{T+3|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T+2|T} + \hat{\beta}_2 Y_{T+1|T} + \hat{\beta}_3 Y_T + \dots + \hat{\beta}_p Y_{T-p+3} \quad (16.11)$$

donde las β son las estimaciones MCO de los coeficientes AR(p). Continuando con este proceso («iterando») se generan predicciones acerca del valor de las variables en un horizonte temporal futuro de períodos más alejados.

La **predicción multiperiodo iterada VAR** se calcula asimismo por pasos: en primer lugar se calcula la predicción para un horizonte temporal de un periodo visto para todas las variables del VAR, y a continuación se utilizan estas predicciones para calcular las predicciones para un horizonte temporal de dos períodos hacia adelante, y se continúa con este proceso de forma iterativa hasta llegar al horizonte de predicción deseado h . La predicción iterada a dos períodos vista de Y_{T+2} basada en un VAR(p) de dos variables del Concepto clave 16.1 es

$$\begin{aligned} \hat{Y}_{T+2|T} = & \hat{\beta}_{10} + \hat{\beta}_{11} \hat{Y}_{T+1|T} + \hat{\beta}_{12} Y_T + \hat{\beta}_{13} Y_{T-1} + \dots + \hat{\beta}_{1p} Y_{T-p+2} \\ & + \hat{\gamma}_{11} \hat{X}_{T+1|T} + \hat{\gamma}_{12} X_T + \hat{\gamma}_{13} X_{T-1} + \dots + \hat{\gamma}_{1p} X_{T-p+2}, \end{aligned} \quad (16.12)$$

donde los coeficientes de la Ecuación (16.12) son las estimaciones MCO de los coeficientes del VAR. La iteración genera predicciones para un horizonte temporal futuro más alejado.

Por ejemplo, la predicción de la variable ΔInf_t para un horizonte temporal de dos trimestres vista, utilizando cuatro retardos tanto de la variable ΔInf_{t-2} como de la variable $Desemp_{t-2}$, se calcula estimando en primer lugar la regresión:

$$\begin{aligned} \widehat{\Delta Inf}_{t|t-2} = & -0,15 - 0,25\Delta Inf_{t-2} + 0,16\Delta Inf_{t-3} - 0,15\Delta Inf_{t-4} - 0,10\Delta Inf_{t-5} \\ (0,53) \quad (0,13) \quad (0,13) \quad (0,14) \quad (0,07) \\ & - 0,17Desemp_{t-2} + 1,82Desemp_{t-3} - 3,53Desemp_{t-4} + 1,89Desemp_{t-5}. \end{aligned} \quad (16.13)$$

$$(0,70) \quad (1,63) \quad (2,00) \quad (0,91)$$

La predicción para un horizonte temporal de dos trimestres vista de la variación de la inflación desde 2005:I hasta 2005:II se calcula mediante la sustitución de los valores de $\Delta Inf_{2004:IV}$, ..., $\Delta Inf_{2004:I}$..., $Desemp_{2004:IV}$, ..., $Desemp_{2004:I}$ en la Ecuación (16.13); lo que da lugar a

$$\begin{aligned} \widehat{\Delta Inf}_{2005:II|2004:IV} = & -0,15 - 0,25\Delta Inf_{2004:IV} + 0,16\Delta Inf_{2004:III} - 0,15\Delta Inf_{2004:II} \\ & - 0,10\Delta Inf_{2004:I} - 0,17Desemp_{2004:IV} + 1,82Desemp_{2004:III} \\ & - 3,53Desemp_{2004:III} + 1,89Desemp_{2004:I} = -1,38. \end{aligned} \quad (16.14)$$

La predicción directa con un horizonte temporal de tres períodos hacia delante de ΔInf_{T+3} se calcula retardando todas las variables explicativas de la Ecuación (16.13) en un trimestre adicional, estimando esta regresión, y más tarde calculando la predicción. La predicción directa para un horizonte temporal de h trimestres por delante de ΔInf_{T+h} se calcula utilizando ΔInf_t como variable dependiente y los regresores ΔInf_{t-h} y $Desemp_{t-h}$, además de los retardos adicionales de ΔInf_{t-h} y $Desemp_{t-h}$ que se desee.

Errores estándar en regresiones multiperiodo directas. Debido a que la variable dependiente en una regresión multiperiodo tiene lugar en dos o más períodos en el futuro, el término de error en una regresión multiperiodo está seriamente correlacionado. Para comprobarlo, consideremos la predicción de la inflación en un horizonte temporal de dos períodos vista y supongamos que se produce un salto inesperado de los precios del petróleo en el trimestre siguiente. La predicción de la inflación con un horizonte temporal de dos períodos vista realizada hoy será demasiado baja, ya que no incorpora este suceso inesperado. Debido a que el aumento en el precio del petróleo era desconocido asimismo en el trimestre anterior, la predicción con un horizonte temporal de dos períodos vista, realizada el trimestre pasado, será del mismo modo demasiado baja. Por tanto el aumento inesperado del precio del petróleo a lo largo del trimestre siguiente significa que las predicciones con un horizonte temporal de dos períodos vista, realizadas tanto en el trimestre pasado como en el trimestre actual, son demasiado bajas. Debido a tales circunstancias sobrevenidas, el término de error en una regresión multiperiodo presenta correlación serial.

Como se analizó en la Sección 15.4, si el término de error está correlacionado seriamente, los errores estándar MCO habituales son incorrectos o, dicho de forma más precisa, no constituyen una base fiable para la inferencia. Por tanto, con regresiones multiperiodo directas deben utilizarse los errores estándar consistentes a heterocedasticidad y autocorrelación (HAC). Por lo tanto, los errores estándar presentados en la Ecuación (16.13) para regresiones multiperiodo directas son los errores estándar HAC de Newey-West, donde se establece el parámetro de truncamiento m de acuerdo con la Ecuación (15.17); para estos datos (para los cuales $T = 92$), la Ecuación (15.17) obtiene $m = 3$. Para horizontes de predicción más largos, el grado de superposición —y por lo tanto el grado de correlación serial en el error— aumenta: en general, los primeros $h - 1$ coeficientes de autocorrelación de los errores en una regresión con un horizonte temporal de h períodos vista son distintos de cero. Por tanto los valores de m mayores a los indicados por la Ecuación (15.17) resultan apropiados para regresiones multiperiodo con horizontes de predicción largos.

Las predicciones multiperiodo directas se resumen en el Concepto clave 16.3.

CONCEPTO CLAVE

16.3

Predicciones multiperiodo directas

La **predicción multiperiodo directa** con un horizonte temporal de h períodos vista futuros basada en los p retardos tanto de Y_t como de un predictor adicional X_t , se calcula estimando en primer lugar la regresión,

$$Y_t = \delta_0 + \delta_1 Y_{t-h} + \cdots + \delta_p Y_{t-p-h+1} + \delta_{p+1} X_{t-h} + \cdots + \delta_{2p} X_{t-p-h+1} + u_t, \quad (16.15)$$

y utilizando más tarde los coeficientes estimados directamente para realizar la predicción de Y_{T+h} utilizando los datos disponibles hasta el periodo T .

¿Qué método debe utilizarse?

En la mayoría de las aplicaciones, el procedimiento recomendado es el método iterativo para la predicción multiperiodo, por dos razones. En primer lugar, desde una perspectiva teórica, si el modelo con un horizonte temporal de un periodo vista subyacente (el AR o VAR que se utiliza para calcular la predicción iterada) se especifica correctamente, entonces los coeficientes se estiman de manera más eficiente si están estimados mediante una regresión con un horizonte temporal de un periodo vista (iterando a continuación) que si se estiman mediante una regresión con un horizonte temporal de varios períodos vista. En segundo lugar, desde una perspectiva práctica, los profesionales que se dedican a las predicciones suelen estar interesados en predicciones efectuadas no con un único horizonte temporal, sino con varios horizontes vista. Debido a que se han generado utilizando el mismo modelo, las predicciones iteradas suelen presentar trayectorias temporales menos erráticas entre los distintos horizontes que las predicciones directas. Debido que se utiliza un modelo diferente para cada uno de los horizontes en las predicciones directas, el error muestral de los coeficientes estimados puede añadir fluctuaciones aleatorias a la trayectoria temporal de una secuencia de predicciones multiperiodo directas.

No obstante, bajo algunas circunstancias, las predicciones directas son preferibles a las predicciones iteradas. Una circunstancia de este tipo es cuando se tienen razones para creer que el modelo con un horizonte temporal de un periodo vista (el AR o el VAR) no se ha especificado correctamente. Por ejemplo, se podría pensar que la ecuación para la variable que se está tratando de prever en un VAR está especificada correctamente, pero que una o más de las otras ecuaciones del VAR están especificadas de forma incorrecta, quizás debido a la presencia de términos no lineales no considerados. Si el modelo con un horizonte temporal de un periodo vista se especifica de forma incorrecta, en general la predicción multiperiodo iterada será sesgada y el ECMP de la predicción iterada puede ser mayor que el ECMP de la predicción directa, a pesar de que la predicción directa tenga una varianza mayor. Una segunda circunstancia en la que podría resultar deseable una predicción directa aparece en los modelos de predicción multivariante con muchos predictores, en cuyo caso un VAR especificado en términos de todas las variables podría resultar poco fiable debido a que habría muchísimos coeficientes estimados.

16.3 Órdenes de integración y contraste DF-MCG de raíces unitarias

En esta sección se extiende el tratamiento de las tendencias estocásticas presentado en la Sección 14.6 abordando dos cuestiones adicionales. En primer lugar, las tendencias de algunas series temporales no se describen correctamente mediante el modelo de paseo aleatorio, por lo que se introduce una extensión de este modelo y se analizan sus implicaciones para la modelización de una regresión con una serie de ese tipo. En segundo lugar, se continúa con el análisis de contrastación de presencia de raíz unitaria en datos de series temporales y, entre otras cosas, se introduce un segundo contraste de raíz unitaria, el contraste DF-MCG.

Otros modelos de tendencias y órdenes de integración

Recordemos que el modelo de paseo aleatorio para la tendencia, introducido en la Sección 14.6, establece que la tendencia en el periodo t es igual a la tendencia en el periodo $t - 1$, más un término de error aleatorio. Si Y_t sigue un paseo aleatorio con deriva β_0 , entonces

$$Y_t = \beta_0 + Y_{t-1} + u_t, \quad (16.16)$$

donde u_t está seriamente correlacionado. Recordemos asimismo de la Sección 14.6 que, si una serie tiene una tendencia de paseo aleatorio, entonces tiene una raíz autorregresiva que es igual a 1.

Aunque el modelo de paseo aleatorio para una tendencia describe los movimientos a largo plazo de muchas series temporales económicas, algunas series temporales económicas presentan tendencias que son más suaves —es decir, que varían menos de un periodo a otro— que lo que implica la Ecuación (16.16). Es necesario un modelo diferente para describir las tendencias de estas series.

Un modelo de tendencia suave hace que la primera diferencia de la tendencia siga un paseo aleatorio; es decir,

$$\Delta Y_t = \beta_0 + \Delta Y_{t-1} + u_t, \quad (16.17)$$

donde u_t está seriamente incorrelacionada. Por tanto, si Y_t sigue la Ecuación (16.17), ΔY_t sigue un paseo aleatorio, por lo que $\Delta Y_t - \Delta Y_{t-1}$ es estacionaria. La diferencia de las primeras diferencias, $\Delta Y_t - \Delta Y_{t-1}$ se denomina la **segunda diferencia** de Y_t y se expresa por medio de $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$. En esta terminología, si Y_t sigue la Ecuación (16.17), entonces su segunda diferencia es estacionaria. Si una serie tiene una tendencia con la forma de la Ecuación (16.17), entonces la primera diferencia de la serie tiene una raíz autorregresiva que es igual a 1.

Terminología de los «órdenes de integración». Resulta útil emplear algunos términos adicionales para distinguir entre estos dos modelos de tendencias. Una serie que tiene una tendencia de paseo aleatorio se dice que es **integrada de orden uno**, o **I(1)**. Una serie que tiene una tendencia de la forma de la Ecuación

**CONCEPTO
CLAVE**
16.4
Órdenes de integración, diferenciación y estacionariedad

- Si Y_t es integrada de orden uno, es decir, si Y_t es $I(1)$, entonces Y_t tiene una raíz unitaria autorregresiva y su primera diferencia, ΔY_t , es estacionaria.
- Si Y_t es integrada de orden dos, es decir, si Y_t es $I(2)$, entonces ΔY_t tiene una raíz unitaria autorregresiva y su segunda diferencia, $\Delta^2 Y_t$, es estacionaria.
- Si Y_t es **integrada de orden d** , es decir, si Y_t es $I(d)$, entonces Y_t debe ser diferenciada d veces para eliminar su tendencia estocástica, es decir, $\Delta^d Y_t$ es estacionaria.

(16.17) se dice que es **integrada de orden dos** o $I(2)$. Una serie que no tiene una tendencia estocástica y es estacionaria se dice que es **integrada de orden cero**, o $I(0)$.

El **orden de integración** en la terminología de $I(1)$ y $I(2)$ es el número de veces que una serie debe ser diferenciada para que sea estacionaria: si Y_t es $I(1)$, entonces la primera diferencia de Y_t , ΔY_t , es estacionaria, y si Y_t es $I(2)$, entonces la segunda diferencia de Y_t , $\Delta^2 Y_t$, es estacionaria. Si Y_t es $I(0)$, entonces Y_t es estacionaria.

Los órdenes de integración se resumen en el Concepto clave 16.4.

Cómo contrastar si una serie es $I(2)$ o $I(1)$. Si Y_t es $I(2)$, entonces ΔY_t es $I(1)$, por lo que ΔY_t tiene una raíz autorregresiva que es igual a 1. Sin embargo, si Y_t es $I(1)$, entonces ΔY_t es estacionaria. Por tanto, la hipótesis nula de que Y_t es $I(2)$ puede contrastarse frente a la hipótesis alternativa de que Y_t es $I(1)$ mediante el contraste de que ΔY_t tiene una raíz unitaria autorregresiva. Si la hipótesis de que ΔY_t tiene una raíz unitaria autorregresiva es rechazada, entonces la hipótesis de que Y_t es $I(2)$ se rechaza en favor de la alternativa de que Y_t es $I(1)$.

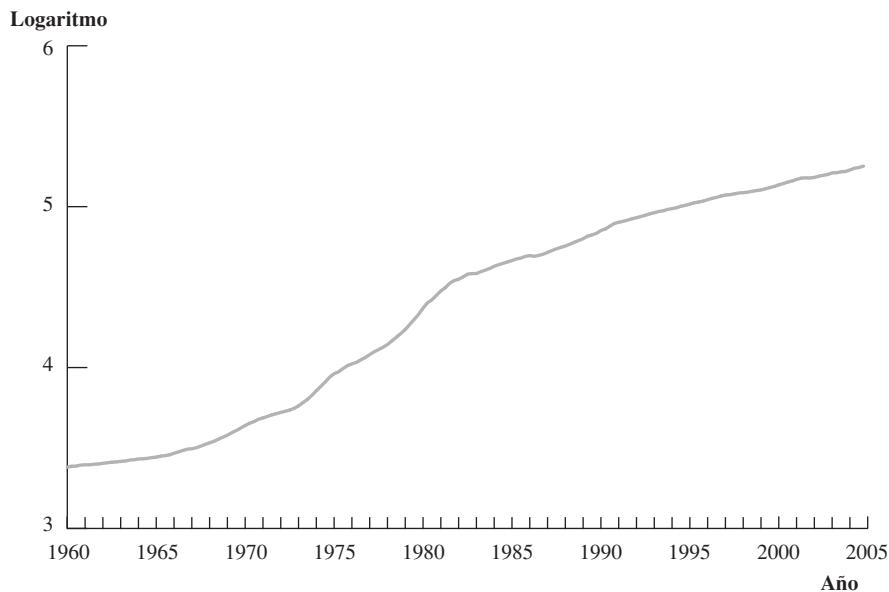
Ejemplos de series $I(2)$ y $I(1)$. El nivel de precios y la tasa de inflación. En el Capítulo 14, llegamos a la conclusión de que resultaba verosímil que la tasa de inflación en los Estados Unidos tuviera una tendencia estocástica de paseo aleatorio, es decir, que la tasa de inflación fuera $I(1)$. Si la inflación es $I(1)$, entonces su tendencia estocástica se elimina mediante la realización de su primera diferencia, por lo que la variable $\Delta \ln(p_t)$ es estacionaria. Recordemos de la Sección 14.2 [Ecuación (14.2)] que la inflación trimestral en tasa anual es la primera diferencia del logaritmo del nivel de precios, multiplicada por 400; es decir, $\ln(p_t) = 400\Delta p_t$, donde $p_t = \ln(IPC_t)$ e IPC_t indica el valor del Índice de Precios al Consumo en el trimestre t . Por lo tanto considerar la tasa de inflación como $I(1)$ es equivalente a considerar la variable Δp_t como $I(1)$, pero esto a su vez es equivalente a considerar p_t como $I(2)$. Por tanto, hemos estado considerando todo el tiempo hasta ahora el logaritmo del nivel de precios como una variable $I(2)$, incluso aunque no hayamos utilizado esa terminología.

El logaritmo del nivel de precios, p_t , y la tasa de inflación se representan gráficamente en la Figura 16.1. La tendencia a largo plazo del logaritmo del nivel de precios (Figura 16.1a) varía de forma más suave que la tendencia a largo plazo de la tasa de inflación (Figura 16.1b). La variación suave de la tendencia del logaritmo del nivel de precios es típica de las series $I(2)$.

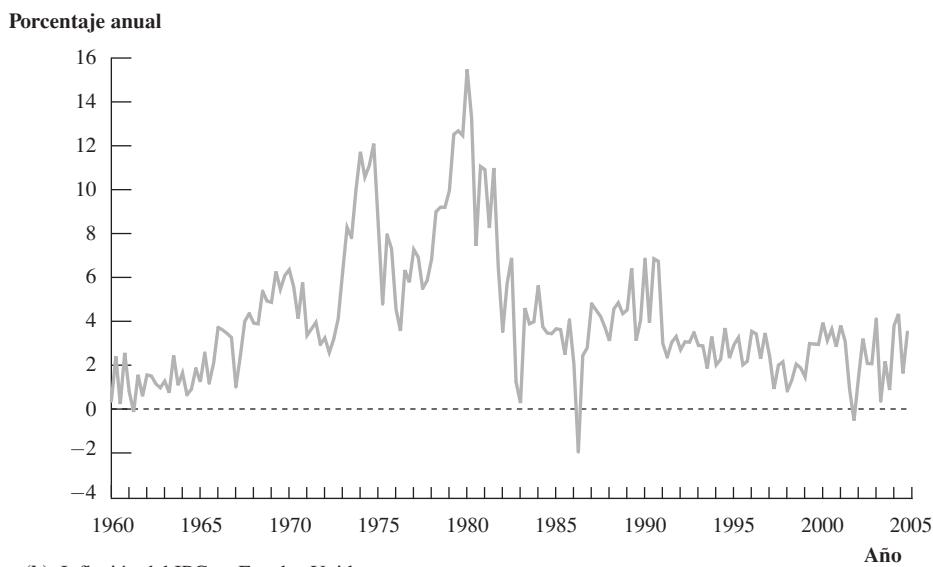
El contraste DF-MCG de raíces unitarias

Esta sección continúa con el análisis de la Sección 14.6 al respecto de la contrastación de la presencia de raíz unitaria autorregresiva. En primer lugar se describe otro contraste de raíz unitaria autorregresiva, el denominado contraste DF-MCG. A continuación, en una sección matemática opcional, se analiza por qué los estadísticos de contraste de raíz unitaria no tienen distribuciones normales, incluso en muestras grandes.

El contraste DF-MCG. El contraste ADF fue el primer contraste desarrollado para contrastar la hipótesis nula de una raíz unitaria y es el contraste más comúnmente utilizado en la práctica. No obstante, posteriormente se han propuesto otros contrastes, muchos de los cuales presentan una potencia mayor (Concepto clave 3.5) que el contraste ADF. Un contraste con mayor potencia que el contraste ADF es más probable que rechace la hipótesis nula de una raíz unitaria frente a la alternativa estacionaria cuando la alternativa es

FIGURA 16.1 El logaritmo del nivel de precios y la tasa de inflación en los Estados Unidos, 1960-2004

(a) Logaritmo del IPC de Estados Unidos



(b) Inflación del IPC en Estados Unidos

La tendencia del logaritmo de los precios (Figura 16.1a) es mucho más suave que la tendencia de la inflación (Figura 16.1b).

verdadera; por lo tanto, un contraste más potente es más capaz de distinguir entre una raíz unitaria AR y una raíz que es elevada, pero menor que 1.

Esta sección trata sobre uno de esos contrastes, el **contraste DF-MCG** desarrollado por Elliott, Rothenberg, y Stock (1996). El contraste se presenta para el caso en el que, bajo la hipótesis nula, Y_t tiene una tendencia de paseo aleatorio, posiblemente con deriva, y en el que bajo la alternativa Y_t es estacionaria en torno a una tendencia temporal lineal.

El contraste DF-MDG se calcula en dos pasos. En el primer paso, el término independiente y la tendencia se estiman por mínimos cuadrados generalizados (MCG; véase la Sección 15.5). La estimación MCG se lleva a cabo mediante el cálculo de tres nuevas variables, V_t , X_{1t} , y X_{2t} , donde $V_t = Y_t$ y $V_t = Y_t - \alpha^* Y_{t-1}$,

con $t = 2, \dots, T$, $X_{1t} = 1$ y $X_{2t} = 1 - \alpha^*$, con $t = 2, \dots, T$, y $X_{21} = 1$ y $X_{2t} = t - \alpha^*(t - 1)$, donde α^* se calcula mediante la fórmula $\alpha^* = 1 - 13,5/T$. Posteriormente V_t se regresa sobre X_{1t} y X_{2t} ; es decir, se utiliza MCO para estimar los coeficientes de la ecuación de regresión poblacional

$$V_t = \delta_0 X_{1t} + \delta_1 X_{2t} + e_t, \quad (16.18)$$

utilizando las observaciones $t = 1, \dots, T$, donde e_t es el término de error. Debe tenerse en cuenta que no existe término independiente en la regresión de la Ecuación (16.18). Los estimadores MCO $\hat{\delta}_0$ y $\hat{\delta}_1$ se utilizan posteriormente para calcular una versión «destendenciada» de Y_t , $Y_t^d = Y_t - (\hat{\delta}_0 + \hat{\delta}_1 t)$.

En el segundo paso, se utiliza el contraste de Dickey-Fuller para contrastar la existencia de una raíz unitaria autorregresiva en Y_t^d , donde la regresión de Dickey-Fuller no incluye un término independiente o una tendencia temporal. Es decir, se regresa la variable ΔY_t^d sobre las variables Y_{t-1}^d y $\Delta Y_{t-1}^d, \dots, \Delta Y_{t-p}^d$, donde el número de retardos p se determina, como es habitual, ya sea por el conocimiento experto, o bien mediante la utilización de un método basado en los datos como un criterio de información AIC o BIC tal y como se analizó en la Sección 14.5.

Si la hipótesis alternativa es que Y_t es estacionaria, con una media que podría ser distinta de cero, pero sin tendencia temporal, entonces se modifican los pasos anteriores. En concreto, α^* se calcula a partir de la fórmula $\alpha^* = 1 - 7/T$, X_{2t} se omite de la regresión de la Ecuación (16.18), y la serie Y_t^d se calcula como $Y_t^d = Y_t - \hat{\delta}_0$.

La regresión MCG del primer paso del contraste DF-MCG hace que este contraste resulte más complicado que el contraste ADF convencional, pero asimismo es esto lo que mejora su capacidad para discriminar entre la hipótesis nula de presencia de raíz unitaria autorregresiva y la alternativa de que Y_t es estacionaria. Esta mejora puede ser importante. Por ejemplo, supongamos que Y_t es en realidad un proceso AR(1) estacionario con un coeficiente autorregresivo $\beta_1 = 0,95$, que existen $T = 200$ observaciones, y que los contrastes de raíz unitaria se calculan sin tendencia temporal [es decir, se excluye t de la regresión de Dickey-Fuller, y se omite la variable X_{2t} de la Ecuación (16.18)]. Por tanto, la probabilidad de que el contraste ADF rechace correctamente la hipótesis nula al nivel de significación del 5 % es de aproximadamente el 31 % frente al 75 % del contraste DF-MCG.

Valores críticos para el contraste DF-MCG. Debido a que los coeficientes de los términos determinísticos se estiman de manera diferente en los contrastes ADF y DF-MCG, los contrastes tienen valores críticos distintos. Los valores críticos para el contraste DF-MCG están recogidos en la Tabla 16.1. Si el estadístico de contraste DF-MCG (el estadístico t de Y_{t-1}^d en la regresión del segundo paso) es menor que el valor crítico (es decir, es más negativo que el valor crítico), entonces la hipótesis nula de que Y_t tiene una raíz unitaria se rechaza. Al igual que los valores críticos para el contraste de Dickey-Fuller, el valor crítico apropiado depende de la versión del contraste utilizada, es decir, de si se incluye o no una tendencia temporal [de si se incluye o no X_{2t} en la Ecuación (16.18)].

Aplicación a la inflación. El estadístico DF-MCG, calculado para la tasa de inflación medida por el IPC, Inf_t a lo largo del periodo que va desde 1962:I hasta 2004:IV con un intercepto pero sin tendencia temporal, es $-2,06$, cuando se incluyen en la regresión de Dickey-Fuller del segundo paso, tres retardos de la variable ΔY_t^d . Este valor es menor que el valor crítico del 5 % en la Tabla 16.1, $-1,95$, por lo que la

TABLA 16.1 Valores críticos para el contraste DF-MCG

Regresores determinísticos [Regresores de la Ecuación (16.18)]	10 %	5 %	1 %
Solamente término independiente (solamente X_{1t})	-1,62	-1,95	-2,58
Término independiente y tendencia temporal (X_{1t} y X_{2t})	-2,57	-2,89	-3,48

Fuente: Fuller (1976) y Elliott, Rothenberg, y Stock (1996, Tabla 1)

utilización del contraste DF-MCG con tres retardos lleva a rechazar la hipótesis nula de una raíz unitaria al 5 % de nivel de significación. La elección de los tres retardos está basada en el criterio AIC (de un máximo de seis retardos).

Debido a que el contraste DF-MCG es capaz de discriminar mejor entre la hipótesis nula de presencia de raíz unitaria y la alternativa estacionaria, una interpretación de este hallazgo es que la inflación es en realidad estacionaria, y que el contraste de Dickey-Fuller realizado en la Sección 14.6 no pudo detectarlo (al nivel del 5 %). Esta conclusión, sin embargo, debería atemperarse señalando que si bien el contraste DF-MCG rechaza la hipótesis nula esto es, en esta aplicación, sensible a la elección de la longitud de los retardos. Si el contraste se basa en dos retardos, que es el número de retardos seleccionados mediante el criterio BIC, se rechaza la hipótesis nula al nivel del 10 % pero no al 5 %. El resultado es sensible asimismo a la elección del periodo muestral; si el estadístico se calcula en cambio para el periodo 1963:I hasta 2004:IV (es decir, eliminando exactamente el primer año), el contraste rechaza la hipótesis nula al nivel del 10 %, pero no al 5 % utilizando la longitud de los retardos resultante del criterio AIC. El cuadro general por lo tanto, es bastante ambiguo [como el que se basaba en el contraste ADF, que se analizaba de acuerdo con la Ecuación (14.34)] y requiere de una decisión informada por parte de quien realice el pronóstico acerca de si resulta mejor modelizar la inflación como $I(1)$ o como estacionaria.

¿Por qué los contrastes de raíz unitaria tienen distribuciones no normales?

En la Sección 14.6, se hizo hincapié en que la distribución normal en muestras grandes sobre la que está basado gran parte del análisis de regresión no es aplicable si los regresores son no estacionarios. Bajo la hipótesis nula de que la regresión contiene una raíz unitaria, el regresor Y_{t-1} en la regresión de Dickey-Fuller (y el regresor Y_{t-1}^d en la regresión modificada de Dickey-Fuller de la segunda etapa del contraste DF-MCG) es no estacionario. La distribución no normal del estadístico de contraste de raíz unitaria es una consecuencia de esta no estacionariedad.

Con el fin de proporcionar alguna idea intuitiva y matemática acerca de esta no normalidad, consideremos la regresión de Dickey-Fuller más sencilla posible, en la que la variable ΔY_t se regresa sobre un único regresor Y_{t-1} y se excluye el término independiente. En la notación del Concepto clave 14.8, el estimador MCO de esta regresión es $\hat{\delta} = \sum_{t=1}^T Y_{t-1} \Delta Y_t / \sum_{t=1}^T Y_{t-1}^2$, por lo que

$$T\hat{\delta} = \frac{\frac{1}{T} \sum_{t=1}^T Y_{t-1} \Delta Y_t}{\frac{1}{T^2} \sum_{t=1}^T Y_{t-1}^2} \quad (16.19)$$

Consideremos el numerador de la Ecuación (16.19). Bajo el supuesto adicional de que $Y_0 = 0$, un poco de álgebra (Ejercicio 16.5) demuestra que

$$\frac{1}{T} \sum_{t=1}^T Y_{t-1} \Delta Y_t = \frac{1}{2} \left[\left(\frac{Y_T}{\sqrt{T}} \right)^2 - \frac{1}{T} \sum_{t=1}^T (\Delta Y_t)^2 \right] \quad (16.20)$$

Bajo la hipótesis nula, $\Delta Y_t = u_t$, el cual no presenta correlación serial y tiene una varianza finita, por lo que el segundo término de la Ecuación (16.20) tiene el límite de probabilidad $\frac{1}{T} \sum_{t=1}^T (\Delta Y_t)^2 \xrightarrow{P} \sigma_u^2$. Bajo el supuesto de que $Y_0 = 0$, el primer término de la Ecuación (16.20) se puede escribir como $Y_T / \sqrt{T} = \sqrt{\frac{1}{T}} \sum_{t=1}^T \Delta Y_t = \sqrt{\frac{1}{T}} \sum_{t=1}^T u_t$, que a su vez cumple el teorema central del límite; es decir, $Y_T / \sqrt{T} \xrightarrow{d} N(0, \sigma_u^2)$. Por tanto $(Y_T / \sqrt{T})^2 - \frac{1}{T} \sum_{t=1}^T (\Delta Y_t)^2 \xrightarrow{d} \sigma_u^2(Z^2 - 1)$, donde Z es una variable aleatoria normal estándar. Recordemos, sin embargo, que el cuadrado de una distribución normal estándar tiene una distribución chi cuadrado con 1 grado de libertad. Por lo tanto de la Ecuación (16.20) se deduce que, bajo la hipótesis nula, el numerador de la Ecuación (16.19) tiene, en el límite, la distribución

$$\frac{1}{T} \sum_{t=1}^T Y_{t-1} \Delta Y_t \xrightarrow{d} \frac{\sigma_u^2}{2} (\chi_1^2 - 1) \quad (16.21)$$

La distribución para muestras grandes de la Ecuación (16.21) es diferente a la distribución normal para muestras grandes habitual cuando el regresor es estacionario. En cambio, el numerador del estimador MCO del coeficiente de Y_t en esta regresión de Dickey-Fuller tiene una distribución que es proporcional a la distribución chi-cuadrado con un grado de libertad, menos 1.

Este análisis ha tenido en cuenta solamente el numerador de $T\hat{\delta}$. Asimismo, el denominador se comporta, bajo la hipótesis nula, de manera inusual: debido a que Y_t sigue un paseo aleatorio bajo la hipótesis nula, $\frac{1}{T} \sum_{t=1}^T Y_{t-1}^2$ no converge en probabilidad a una constante. En cambio, el denominador de la Ecuación (16.19) es una variable aleatoria, incluso en muestras grandes: bajo la hipótesis nula, $\frac{1}{T^2} \sum_{t=1}^T Y_{t-1}^2$ converge conjuntamente en distribución con el numerador. Las distribuciones inusuales del numerador y el denominador de la Ecuación (16.19) son la fuente de la distribución no estándar del estadístico de contraste de Dickey-Fuller y la razón de que el estadístico ADF tenga su propia tabla particular de valores críticos.

16.4 Cointegración

A veces dos o más series tienen en común la misma tendencia estocástica. En este caso particular, conocido como cointegración, el análisis de regresión puede revelar la existencia de relaciones a largo plazo entre las variables de series temporales, pero se necesitan algunos métodos nuevos.

Cointegración y corrección de error

Dos o más series de tiempo con tendencias estocásticas pueden evolucionar juntas de forma tan estrecha a largo plazo que puede parecer que tienen la misma componente tendencial; es decir, que parece que tengan una **tendencia común**. Por ejemplo, se representan gráficamente dos variables de tipos de interés de la deuda pública de EE.UU. en la Figura 16.2. Uno de estos tipos de interés es el tipo de interés de las Letras del Tesoro a 90 días, en tasa anual ($R90_t$); la otra es el tipo de interés de los Bonos del Tesoro de EE.UU. a un 1 año ($R1A_t$); estos tipos de interés se analizan en el Apéndice 16.1. Los tipos de interés exhiben las mismas tendencias a largo plazo: ambos fueron bajos en la década de 1960, ambos subieron en la década de 1970 hasta los máximos de los primeros años de la década de 1980, y más tarde ambos cayeron a lo largo de la década de 1990. Por otra parte, la diferencia entre las dos series, $R1A_t - R90_t$, que se denomina el «spread»² entre los dos tipos de interés y se representa asimismo en la Figura 16.2, no parece tener una tendencia. Es decir, restando el tipo de interés de 90 días al tipo de interés de 1 año parece que se eliminan las tendencias en ambos tipos de interés individuales. Dicho de otra manera, aunque ambos tipos de interés difieren, parecen compartir una tendencia estocástica común: debido a que la tendencia de cada serie individual se elimina restando una serie de la otra, las dos series deben tener la misma tendencia; es decir, deben tener una tendencia estocástica común.

Dos o más series que tienen una tendencia estocástica común se dice que están cointegradas. La definición formal de **cointegración** (debida al economista Clive Granger, 1983; véase el recuadro sobre Clive Granger y Robert Engle) se ofrece en el Concepto clave 16.5. En esta sección, presentamos un contraste para la detección de la cointegración, analizamos la estimación de los coeficientes de las regresiones que relacionan variables cointegradas, e ilustramos la utilización de la relación de cointegración para predicción.

**CONCEPTO
CLAVE**

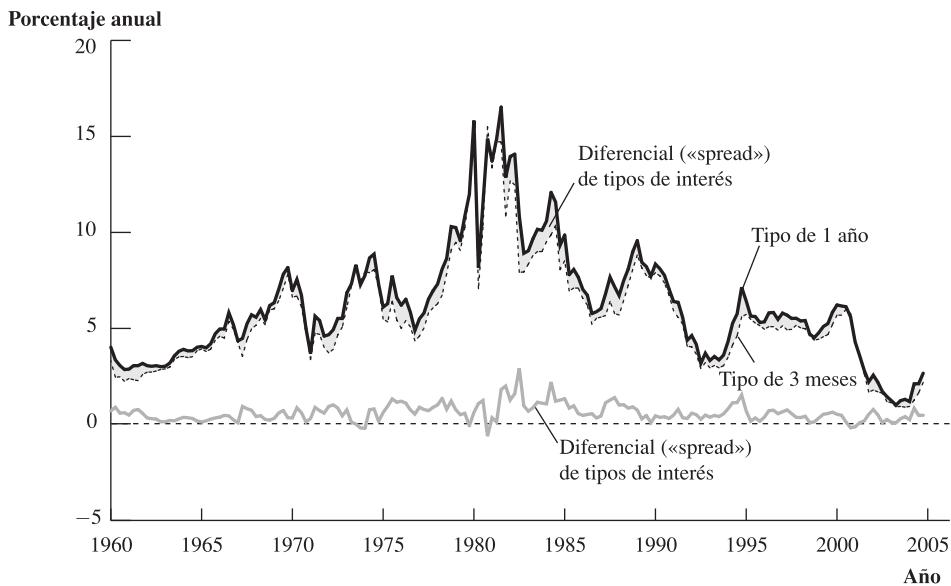
16.5

Cointegración

Supongamos que X_t e Y_t son integradas de orden uno. Si, para algún coeficiente θ , $Y_t - \theta X_t$ es integrada de orden cero, entonces X_t e Y_t se dice que están *cointegradas*. El coeficiente θ se denomina **coeficiente de cointegración**.

Si X_t e Y_t están cointegradas, entonces tienen la misma, o común, tendencia estocástica. Calculando la diferencia $Y_t - \theta X_t$ se elimina esta tendencia estocástica común.

² *N. del T.:* O bien, «diferencial» de tipos de interés.

FIGURA 16.2 Tipo de interés de un año, tipo de interés de tres meses, y diferencial de tipos de interés

Los tipos de interés de un año y de tres meses comparten una tendencia estocástica común. El diferencial o «spread», entre los dos tipos no muestra una tendencia. Estos dos tipos de interés parecen estar cointegrados.

El análisis se centra inicialmente en el caso de que solamente existen dos variables, X_t e Y_t .

Modelo del vector de corrección de error. Hasta ahora, hemos eliminado la tendencia estocástica en una variable $I(1)$, Y_t , mediante el cálculo de su primera diferencia, ΔY_t ; los problemas creados por las tendencias estocásticas se solucionaron utilizando ΔY_t en lugar de Y_t en las regresiones de series temporales. Sin embargo, si X_t e Y_t están cointegradas, otra forma de eliminar la tendencia consiste en calcular, $Y_t - \theta X_t$, donde θ se elige de manera que elimine la tendencia común de la diferencia. Debido a que el término $Y_t - \theta X_t$ es estacionario, puede utilizarse asimismo en el análisis de regresión.

De hecho, si X_t e Y_t están cointegradas, las primeras diferencias de X_t e Y_t pueden modelizarse utilizando un VAR, aumentando con la inclusión de $Y_{t-1} - \theta X_{t-1}$, como variable explicativa adicional:

$$\begin{aligned}\Delta Y_t = & \beta_{10} + \beta_{11}\Delta Y_{t-1} + \dots + \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \dots \\ & + \gamma_{1p}\Delta X_{t-p} + \alpha_1(Y_{t-1} - \theta X_{t-1}) + u_{1t}\end{aligned}\quad (16.22)$$

$$\begin{aligned}\Delta X_t = & \beta_{20} + \beta_{21}\Delta Y_{t-1} + \dots + \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \dots \\ & + \gamma_{2p}\Delta X_{t-p} + \alpha_2(Y_{t-1} - \theta X_{t-1}) + u_{2t}\end{aligned}\quad (16.23)$$

El término $Y_t - \theta X_t$ se denomina **término de corrección de error**. El modelo combinado de las Ecuaciones (16.22) y (16.23) se denomina **modelo del vector de corrección de error** (MVCE). En un MVCE, los valores pasados de $Y_t - \theta X_t$ ayudan a predecir los valores futuros de ΔY_t y/o los de ΔX_t .

¿Cómo se puede saber si dos variables aleatorias están cointegradas?

Existen tres maneras de determinar si dos variables pueden modelizarse de forma verosímil como cointegradas: utilizar el conocimiento experto y la teoría económica, representar gráficamente las series y comprobar si parecen tener una tendencia estocástica común, y llevar a cabo contrastes estadísticos de cointegración. Los tres métodos deberían utilizarse en la práctica.

En primer lugar, debe utilizarse el conocimiento experto de estas variables para decidir si la cointegración es un hecho creíble. Por ejemplo, los dos tipos de interés de la Figura 16.2 están vinculados entre sí por la denominada teoría de las expectativas de la estructura temporal de los tipos de interés. Según esta teoría, el tipo de interés el 1 de enero del Bono del Tesoro a 1 año es el promedio de los tipos de interés de una Letra del Tesoro a 90 días para el primer trimestre del año y del tipo de interés esperado en el futuro de las Letras del Tesoro a 90 días emitidas en el segundo, tercer, y cuarto trimestres del año; si no, los inversores podrían esperar obtener beneficio, bien manteniendo el Bono del Tesoro durante 1 año o, por el contrario, manteniendo una secuencia de cuatro Letras del Tesoro a 90 días, y ofertarían precios hasta que la rentabilidad esperada se igualara. Si el tipo de interés de 90 días tiene una tendencia estocástica de paseo aleatorio, esta teoría implica que esta tendencia estocástica se hereda por el tipo de interés de 1 año y que la diferencia entre los dos tipos, es decir el *spread*, es estacionaria. Por tanto, la teoría de las expectativas de la estructura temporal implica que si los tipos de interés son $I(1)$, entonces estarán cointegrados con un coeficiente de cointegración de $\theta = 1$ (Ejercicio 16.2).

En segundo lugar, una inspección visual de la serie ayuda a identificar los casos en los que la cointegración es verosímil. Por ejemplo, la gráfica de los dos tipos de interés de la Figura 16.2 muestra que cada una de las series parece ser $I(1)$, pero que el diferencial parece ser $I(0)$, por lo que parece que las dos series pueden estar cointegradas.

En tercer lugar, los procedimientos de contraste de raíz unitaria introducidos hasta el momento se pueden extender a los contrastes de cointegración. La idea en la que se basan estos contrastes es que si Y_t y X_t están cointegradas con coeficiente de cointegración θ , entonces $Y_t - \theta X_t$ es estacionaria; de lo contrario, $Y_t - \theta X_t$ es no estacionaria [es $I(1)$]. La hipótesis de que Y_t y X_t no están cointegradas [es decir, que $Y_t - \theta X_t$ es $I(1)$], puede contrastarse por tanto mediante la contrastación de la hipótesis nula de que $Y_t - \theta X_t$ tiene una raíz unitaria; si esta hipótesis se rechaza, entonces Y_t y X_t se pueden modelizar como cointegradas. Los detalles de este contraste dependen de si el coeficiente de cointegración θ es conocido.

Contraste de cointegración cuando θ es conocido. En algunos casos el conocimiento experto o la teoría económica sugieren los valores de θ . Si θ es conocido, se pueden utilizar los contrastes de raíz unitaria de Dickey-Fuller y DF-MCG para contrastar la existencia de cointegración, en primer lugar construyendo la serie $z_t = Y_t - \theta X_t$ y contrastando a continuación la hipótesis nula de que z tiene una raíz unitaria autorregresiva.

Contraste de cointegración cuando θ es desconocido. Si el coeficiente de cointegración θ es desconocido, entonces debe ser estimado antes de contrastar la presencia de raíz unitaria en el término de corrección de error. Este paso preliminar hace necesaria la utilización de valores críticos diferentes en el contraste de raíz unitaria posterior.

En concreto, en el primer paso el coeficiente de cointegración se estima mediante la estimación MCO de la regresión

$$Y_t = \alpha + \theta X_t + z_t. \quad (16.24)$$

En el segundo paso, se utiliza un contraste t de Dickey-Fuller (con término independiente, pero sin tendencia temporal) para contrastar la presencia de raíz unitaria en el residuo de esta regresión, \hat{z}_t . Este procedimiento en dos etapas se denomina contraste de cointegración de Dickey-Fuller Aumentado de Engle-Granger, o **contraste EG-ADF** (Engle y Granger, 1987).

Los valores críticos del estadístico EG-ADF figuran en la Tabla 16.2³. Los valores críticos de la primera fila son aplicables cuando hay un solo regresor en la Ecuación (16.26), por lo que existen dos variables cointegradas (X_t e Y_t). Las siguientes filas son aplicables al caso de varias variables cointegradas, que se analiza al final de esta sección.

³ Los valores críticos de la Tabla 16.2 se han tomado de Fuller (1976) y Phillips y Ouliaris (1990). Siguiendo una sugerencia de Hansen (1992), los valores críticos de la Tabla 16.2 se han elegido de manera que sean aplicables tanto si X_t e Y_t tienen componentes de derivada (término independiente) como si no.

Robert Engle y Clive Granger, ganadores del Premio Nobel

En 2003, dos económetras, Robert F. Engle y Clive W. J. Granger, ganaron el Premio Nobel de Economía por su investigación sobre los fundamentos teóricos de la econometría de series temporales que llevaron a cabo a finales de la década de 1970 y principios de la década de 1980.



Clive W. J. Granger

El trabajo de Granger se centró sobre cómo manejar tendencias estocásticas en datos de series temporales. A partir de un trabajo anterior realizado por él mismo y por otros autores, sabía que dos series no correlacionadas con tendencias estocásticas podrían, de acuerdo con las medidas estadísticas habituales de los estadísticos t y los R^2 de regresión, aparentar estar significativamente correlacionadas sin estarlo realmente; este es el problema de «regresión espuria». En la década de 1970, la práctica habitual consistía en utilizar las diferencias de los datos de series temporales para evitar el riesgo de una regresión espuria. Por esta razón, Granger se mostraba escéptico ante algunos trabajos recientes realizados por algunos económetras británicos (Davidson, Hendry, Srba, y Yeo, 1978), quienes afirmaban que la diferencia retardada entre el logaritmo del consumo y el logaritmo de la renta ($\ln C_{t-1} - \ln Y_{t-1}$) era un predictor valioso de la tasa de crecimiento del consumo ($\Delta \ln C_t$). Debido a que $\ln C_t$ y $\ln Y_t$ presentan por separado una raíz unitaria, el sentido común dictaba que deberían ser incluidos en primeras diferencias, debido a que el hecho de incluirlos en niveles podría dar lugar a una versión de una regresión espuria. Granger se propuso demostrar matemáticamente que el equipo británico había cometido un error, pero en su lugar demostró que su especificación era correcta: existe una representación matemática bien definida —el modelo del vector de corrección de error— para las series temporales que individualmente son $I(1)$ pero cuya combinación lineal es $I(0)$. Él

llamó a esta situación «cointegración». En los siguientes trabajos con su colega de la Universidad de California en San Diego, Robert Engle, Granger propuso varios contrastes de cointegración, sobre todo el contraste ADF de Engle-Granger descrito en la página anterior. Los métodos del análisis de cointegración son ahora un elemento básico en la macroeconometría moderna.

Por la misma época, Robert Engle estaba reflexionando acerca del notable incremento de la volatilidad en la inflación de los EE.UU. durante los últimos años de la década de 1970 (véase la Figura 16.1b). Si la volatilidad de la inflación había crecido, razonó, entonces los intervalos de predicción para las predicciones de inflación debían ser más amplios que lo que indicaban los modelos de la época, ya que los modelos mantenían la varianza de la inflación constante. Pero, ¿cómo se puede predecir exactamente una varianza que cambia a lo largo del tiempo (que no es observable) de un término de error (que tampoco es observable)?



Robert F. Engle

La respuesta de Engle fue desarrollar el modelo de heterocedasticidad condicional autorregresiva (ARCH), que se describe en la Sección 16.5. El modelo ARCH y sus extensiones, desarrollado principalmente por Engle y sus estudiantes, ha demostrado ser especialmente útil para modelizar la volatilidad de los rendimientos de los activos, y la volatilidad de las predicciones resultante puede utilizarse para valorar derivados financieros y para evaluar los cambios a lo largo del tiempo del riesgo de mantener activos financieros. En la actualidad, las medidas y las predicciones de la volatilidad constituyen una componente central de la econometría financiera, y el modelo ARCH y sus descendientes son las herramientas de trabajo básicas para la modelización de la volatilidad.

TABLA 16.2 Valores críticos del estadístico ADF de Engle-Granger

Número de X en la Ecuación (16.24)	10 %	5 %	1 %
1	-3,12	-3,41	-3,96
2	-3,52	-3,80	-4,36
3	-3,84	-4,16	-4,73
4	-4,20	-4,49	-5,07

Estimación de los coeficientes de cointegración

Si X_t e Y_t están cointegradas, entonces el estimador MCO de los coeficientes de la regresión de cointegración de la Ecuación (16.24) es consistente. Sin embargo, en general, el estimador MCO tiene una distri-

bución no normal, y las inferencias basadas en sus estadísticos t pueden resultar engañosas tanto si los estadísticos t se calculan utilizando errores estándar HAC como si no. Debido a estos inconvenientes del estimador MCO de θ , los económetras han desarrollado una serie de estimadores distintos para el coeficiente de cointegración.

Un estimador de θ de este tipo que resulta fácil de utilizar en la práctica es el **estimador MCO dinámico (MCOD)** (Stock y Watson, 1993). El estimador MCOD está basado en una versión modificada de la Ecuación (16.24), que incluye los valores pasados, presente y futuros de la variación de X_t :

$$Y_t = \beta_0 + \theta X_t + \sum_{j=-p}^p \delta_j \Delta X_{t-j} + u_t \quad (16.25)$$

Por tanto, en la Ecuación (16.25), las variables explicativas son $X_t, \Delta X_{t+p}, \dots, \Delta X_{t-p}$. El estimador MCOD de θ es el estimador MCO de θ en la regresión de la Ecuación (16.25).

Si X_t e Y_t están cointegradas, entonces el estimador MCOD es eficiente en muestras grandes. Por otra parte, las inferencias estadísticas acerca de θ y de las δ de la Ecuación (16.25) basadas en los errores estándar HAC son válidas. Por ejemplo, el estadístico t construido utilizando el estimador MCOD con errores estándar HAC tiene una distribución normal estándar en muestras grandes.

Una manera de interpretar la Ecuación (16.25) consiste en recordar de la Sección 15.3 que los multiplicadores dinámicos acumulativos se pueden calcular mediante la modificación de la regresión de retardos distribuidos de Y_t sobre X_t y sus retardos. En concreto, en la Ecuación (15.7), los multiplicadores dinámicos acumulativos fueron calculados mediante una regresión de la variable Y_t sobre la variable ΔX_t , los retardos de ΔX_t , y X_{t-r} ; el coeficiente de X_{t-r} en esta especificación es el multiplicador dinámico acumulativo de largo plazo. Del mismo modo, si X_t fuera estrictamente exógena, entonces en la Ecuación (16.25) el coeficiente de X_t , θ sería el multiplicador acumulativo de largo plazo, es decir, el efecto a largo plazo sobre Y de una variación en X . Si X_t no es estrictamente exógena, entonces los coeficientes no tienen esta interpretación. Sin embargo, debido a que X_t e Y_t tienen una tendencia estocástica común si están cointegradas, el estimador MCOD es consistente, incluso si X_t es endógena.

El estimador MCOD no es el único estimador eficiente del coeficiente de cointegración. El primero de estos estimadores fue desarrollado por Søren Johansen (Johansen, 1988). Para un análisis sobre el método de Johansen y de otras maneras de estimar el coeficiente de cointegración, véase Hamilton (1994, Capítulo 20).

Incluso si la teoría económica no sugiere un valor específico para el coeficiente de cointegración, es importante comprobar si la relación de cointegración estimada tiene sentido en la práctica. Debido a que los contrastes de cointegración pueden resultar engañosos (pueden rechazar erróneamente la hipótesis nula de no cointegración con más frecuencia de lo que deberían, y con frecuencia no rechazan la hipótesis nula de forma errónea), resulta particularmente importante basarse en la teoría económica, el conocimiento institucional, y el sentido común en la estimación y la utilización de las relaciones de cointegración.

Extensión a varias variables cointegradas

Los conceptos, los contrastes y los estimadores analizados aquí son extensibles al caso de más de dos variables. Por ejemplo, si hay tres variables, Y_t , X_{1t} y X_{2t} , cada una de las cuales es $I(1)$, entonces están cointegradas con coeficientes de cointegración θ_1 y θ_2 si $Y_t - \theta_1 X_{1t} - \theta_2 X_{2t}$ es estacionaria. Cuando existen tres o más variables, pueden existir varias relaciones de cointegración. Por ejemplo, consideremos la modelización de la relación entre tres variables de tipos de interés: el tipo de 3 meses, el tipo de 1 año, y el tipo de 5 años ($R5A$). Si son $I(1)$, entonces la teoría de las expectativas de la estructura temporal de los tipos de interés sugiere que todas ellas estarán cointegradas. Una relación de cointegración que sugiere la teoría es $R1A_t - R90_t$, y una segunda relación es $R5A_t - R90_t$. (La relación $R5A_t - R1A_t$ es asimismo una relación de cointegración, pero no contiene información adicional más allá de la contenida en las otras dos relaciones, ya que es perfectamente multicolineal con las otras dos relaciones de cointegración).

El procedimiento EG-ADF para el contraste de una única relación de cointegración entre varias variables es el mismo que para el caso de dos variables, salvo que la regresión de la Ecuación (16.24) se modifica para que tanto X_1 , como X_2 , sean regresores; los valores críticos para el contraste EG-ADF se muestran en la Tabla 16.2, en la que la fila adecuada depende del número de regresores en la primera etapa de la regresión de cointegración MCO. El estimador MCOD de una única relación de cointegración entre varias X implica la inclusión del nivel de cada una de las X , junto con los valores adelantados y retardados de la primera diferencia de cada una de las X . Los contrastes para varias relaciones de cointegración se pueden llevar a cabo mediante métodos sistemáticos, tales como el método de Johansen (1988), y el estimador MCOD se puede extender a varias relaciones de cointegración mediante la estimación de varias ecuaciones, una para cada relación de cointegración. Para un análisis adicional de los métodos de cointegración para varias variables, véase Hamilton (1994).

Una nota de advertencia. Si dos o más variables están cointegradas, entonces el término de corrección de error puede ayudar a predecir estas variables y, posiblemente, otras variables relacionadas. Sin embargo, la cointegración requiere que las variables tengan la misma tendencia estocástica. Las tendencias de las variables económicas, por lo general, surgen debido a la existencia de interacciones complejas entre fuerzas dispares, y las series que están estrechamente relacionadas pueden tener tendencias diferentes, por razones sutiles. Si las variables que no están cointegradas se modelizan de forma incorrecta mediante un MVCE, entonces el término de corrección de error será $I(1)$; lo cual introduce una tendencia en la predicción que puede dar lugar a un rendimiento pobre de la predicción fuera de muestra. Por tanto, la predicción mediante MVCE debe estar basada en una combinación de argumentos teóricos convincentes a favor de la cointegración y en el análisis empírico cuidadoso.

Aplicación a los tipos de interés

Tal y como se señaló anteriormente, la teoría de las expectativas de la estructura temporal de los tipos de interés implica que si dos tipos de interés para diferentes vencimientos son $I(1)$, entonces estarán cointegrados con un coeficiente de cointegración de $\theta = 1$; es decir, el diferencial entre los dos tipos será estacionario. La observación de la Figura 16.2 proporciona apoyo cualitativo a la hipótesis de que los tipos de interés de 1 año y 3 meses están cointegrados. En primer lugar, se utilizan estadísticos de contraste de raíces unitarias con el fin de aportar evidencia formal adicional acerca de esta hipótesis, y a continuación, se estima un modelo vectorial de corrección de error para estos tipos de interés.

Contrastes de raíces unitarias y cointegración. En la Tabla 16.3 se presentan varios estadísticos de contraste de raíces unitarias y cointegración para estas dos series. Los estadísticos de contraste de raíces unitarias de las dos primeras filas examinan la hipótesis de que los dos tipos de interés, el tipo de interés de 3 meses ($R90$) y el tipo de un año ($R1A$), tienen por separado una raíz unitaria. Dos de los cuatro estadísticos de las dos primeras filas no rechazan esta hipótesis al nivel del 10 %, y tres de los cuatro no la rechazan al 5 % de nivel. La excepción es el estadístico ADF evaluado para el tipo de las Letras del Tesoro a 90 días ($-2,96$), que rechaza la hipótesis de raíz unitaria al 5 % de nivel. Los estadísticos ADF y DF-MCG dan lugar a diferentes conclusiones acerca de esta variable (el contraste ADF rechaza la hipótesis de raíz unitaria al nivel del 5 %, mientras que el contraste DF-MCG no la rechaza), lo que significa que es necesario hacer un juicio de valor para decidir si estas variables se pueden modelizar de forma verosímil como $I(1)$. En conjunto, estos resultados sugieren que los tipos de interés son verosímilmente modelizables como $I(1)$.

Los estadísticos de raíz unitaria para el diferencial, $R1A_t - R90_t$, contrastan otra hipótesis adicional, el hecho de que estas variables no están cointegradas frente a la alternativa de que sí lo están. La hipótesis nula de que el diferencial presenta una raíz unitaria se rechaza al nivel del 1 % utilizando ambos contrastes de raíz unitaria. Por tanto, se rechaza la hipótesis de que las series no están cointegradas frente a la alternativa de que sí lo están, con un coeficiente de cointegración $\theta = 1$. En conjunto, la evidencia a partir de las tres primeras filas de la Tabla 16.3 indica que estas variables pueden ser modelizadas de forma verosímil como cointegradas con $\theta = 1$.

TABLA 16.3 Estadísticos de contraste de raíces unitarias y cointegración sobre dos tipos de interés

Series	Estadístico ADF	Estadístico DF-MCE
$R90$	-2,96*	-1,88
$R1A$	-2,22	-1,37
$R1A - R90$	-6,31**	-5,59**
$R1A - 1,046 R90$	-6,97**	-

$R90$ es el tipo de interés de las Letras Tesoro a 90 días de Estados Unidos, en tasa anual, y $R1A$ es el tipo de interés de los Bonos del Tesoro de Estados Unidos a 1 año. Las regresiones se estimaron a partir de datos trimestrales a lo largo del período 1962:I-1999:IV. El número de retardos en las regresiones del estadístico de contraste de raíz unitaria fue elegido mediante el criterio AIC (máximo seis retardos). Los estadísticos de contraste de raíz unitaria son significativos al 5 % * o al 1 % ** de nivel de significación.

Debido a que en esta aplicación la teoría económica sugiere un valor para θ (la teoría de las expectativas de la estructura temporal sugiere que $\theta = 1$) y debido a que el término de corrección de error es $I(0)$ cuando se impone este valor (el diferencial es estacionario), en principio no es necesaria la utilización del contraste EG-ADF, en el que se estima θ . No obstante, el contraste se calcula a modo de ilustración. La primera etapa del contraste EG-ADF consiste en estimar θ mediante la regresión MCO de una variable sobre la otra; el resultado es

$$\widehat{R1A}_t = 0,361 + 1,046R90_t, \quad \bar{R}^2 = 0,973. \quad (16.26)$$

El segundo paso consiste en calcular el estadístico ADF para los residuos de esta regresión, \hat{z}_t . El resultado, presentado en la última fila de la Tabla 16.3, es inferior al valor crítico del 1 % que es igual a -3,96 en la Tabla 16.2, por lo que la hipótesis nula de que \hat{z}_t tiene una raíz unitaria autorregresiva es rechazada. Este estadístico apunta asimismo hacia el tratamiento de los dos tipos de interés como cointegrados. Nótese que no se presentan los errores estándar en la Ecuación (16.26) debido a que, como se mencionó anteriormente, el estimador MCO del coeficiente de cointegración tiene una distribución no normal y su estadístico t no está normalmente distribuido, por lo que la presentación de los errores estándar (HAC o no) resultaría confusa.

Un modelo de vector de corrección de error para los dos tipos de interés. Si Y_t y X_t están cointegradas, entonces las predicciones sobre las variables ΔY_t y ΔX_t se pueden mejorar mediante la ampliación de un VAR de las variables ΔY_t y ΔX_t , con los valores retardados del término de corrección de error, es decir, mediante el cálculo de predicciones mediante el MVCE de las Ecuaciones (16.22) y (16.23). Si θ es conocido, entonces los coeficientes desconocidos del MVCE pueden estimarse por MCO, incluyendo $z_{t-1} = Y_{t-1} - \theta X_{t-1}$ como regresor adicional. Si θ es desconocido, entonces el MVCE puede calcularse utilizando \hat{z}_{t-1} como regresor, donde $\hat{z}_t = Y_t - \hat{\theta}X_t$, donde $\hat{\theta}$ es un estimador de θ .

En el caso de los dos tipos de interés, la teoría sugiere que $\theta = 1$, y los contrastes de raíz unitaria apoyan la modelización de los dos tipos de interés como cointegrados con un coeficiente de cointegración igual a 1. Por lo tanto, se especifica el MVCE con el valor sugerido por la teoría de $\theta = 1$, es decir, añadiendo el retardo del diferencial, $R1A_{t-1} - R90_{t-1}$, a un VAR de $\Delta R1A_t$ y $\Delta R90_t$. Especificado con dos retardos de las primeras diferencias, el MVCE resultante es

$$\begin{aligned} \widehat{\Delta R90}_t &= 0,14 - 0,24\Delta R90_{t-1} - 0,44\Delta R90_{t-2} - 0,01\Delta R1A_{t-1} \\ &\quad (0,17) \quad (0,32) \quad (0,34) \quad (0,39) \\ &\quad + 0,15\Delta R1A_{t-2} - 0,18(R1A_{t-1} - R90_{t-1}) \\ &\quad (0,27) \quad (0,27) \end{aligned} \quad (16.27)$$

$$\begin{aligned} \widehat{\Delta R1A}_t &= 0,36 - 0,14\Delta R90_{t-1} - 0,33\Delta R90_{t-2} - 0,11\Delta R1A_{t-1} \\ &\quad (0,16) \quad (0,30) \quad (0,29) \quad (0,35) \\ &\quad + 0,10\Delta R1A_{t-2} - 0,52(R1A_{t-1} - R90_{t-1}) \\ &\quad (0,25) \quad (0,24) \end{aligned} \quad (16.28)$$

En la primera ecuación, ninguno de los coeficientes es individualmente significativo al nivel del 5 % y los coeficientes de las primeras diferencias retardadas de los tipos de interés no son conjuntamente significativos al nivel del 5 %. En la segunda ecuación, los coeficientes de los retardos de las primeras diferencias no son conjuntamente significativos, pero el coeficiente del retardo del diferencial (el término de corrección de error), que se estima que es igual a $-0,52$, tiene un estadístico t de $-2,17$, por lo que es estadísticamente significativo al nivel del 5 %. Aunque los valores retardados de la primera diferencia de los tipos de interés no son útiles para predecir los tipos de interés futuros, el retardo del diferencial ayuda a predecir la variación en el tipo de los bonos del Tesoro a 1 año. Cuando el tipo a 1 año supera el tipo a 90 días, se prevé que el tipo a 1 año disminuya en el futuro.

16.5 Volatilidad agrupada y heterocedasticidad condicional autorregresiva

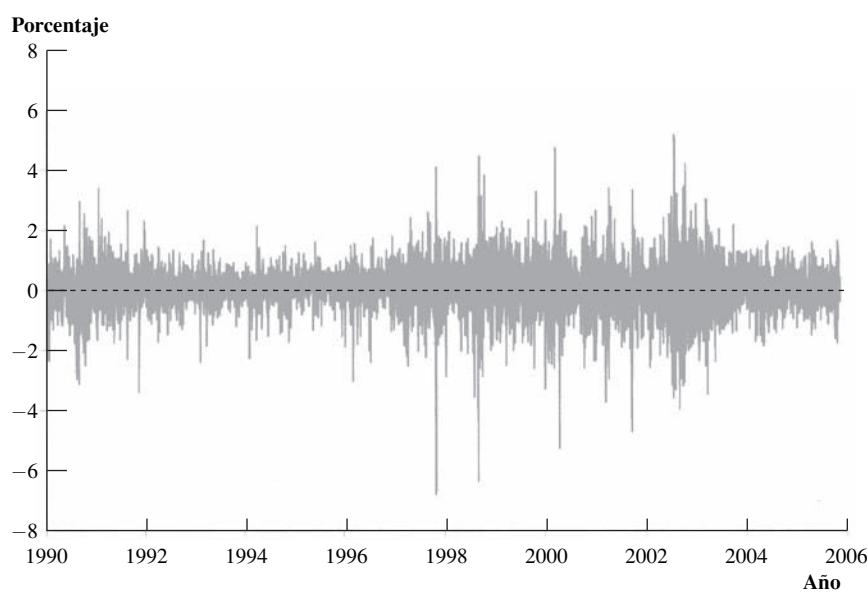
El fenómeno que a veces es tranquilo, y a veces no —es decir, que la volatilidad aparece agrupada—, se manifiesta en muchas series temporales económicas. En esta sección se presentan dos modelos para la cuantificación de la volatilidad agrupada o, como del mismo modo se conoce, la heterocedasticidad condicional.

Volatilidad agrupada

La volatilidad de muchas variables financieras y macroeconómicas varía a lo largo del tiempo. Por ejemplo en las variaciones porcentuales diarias en el índice de precios de la bolsa de Nueva York (NYSE), que se muestra en la Figura 16.3, aparecen períodos de alta volatilidad, como los de 1990 y 2003, junto con otros períodos de baja volatilidad, como en 1993. Una serie con algunos períodos de baja volatilidad y algunos otros períodos de alta volatilidad se dice que presenta **volatilidad agrupada**. Debido a que la volatilidad aparece concentrada en algunos períodos, la varianza de la variación porcentual diaria en el índice de precios NYSE de las acciones puede predecirse, aunque la variación diaria del precio en sí misma sea muy difícil de pronosticar.

FIGURA 16.3 Variación porcentual diaria del índice NYSE, 1990-2005

La variación porcentual diaria en los precios del índice NYSE, 1990-2005 presenta volatilidad agrupada, de tal forma que existen algunos períodos de alta volatilidad, como en los últimos años de la década de 1990, y otros períodos de relativa tranquilidad, como a mediados de la década de 1990.



La predicción de la varianza de una serie resulta interesante por varias razones. En primer lugar, la varianza del precio de un activo constituye una medida del riesgo de poseer ese activo: cuanto mayor sea la varianza de las variaciones diarias de los precios de las acciones, más puede ganar —o perder— un participante en el mercado de valores en un día cualquiera. Un inversor al que le preocupa el riesgo estaría menos dispuesto a acudir al mercado de valores durante un periodo de alta volatilidad que durante los periodos de baja volatilidad.

En segundo lugar, el valor de algunos derivados financieros, como son las opciones, depende de la varianza del valor del activo subyacente. Un operador en el mercado de opciones desea las mejores predicciones disponibles acerca de la volatilidad futura que le puedan ayudar a conocer el precio al que compra o vende opciones.

En tercer lugar, la predicción de las varianzas hace posible disponer de intervalos de confianza precisos. Supongamos que se está realizando una predicción sobre la tasa de inflación. Si la varianza del error de predicción es constante, puede construirse intervalo de confianza de la predicción aproximado de acuerdo con las directrices estudiadas en la Sección 14.4, es decir, como la predicción más o menos un múltiplo del ESR. Sin embargo, si la varianza del error de predicción cambia a lo largo del tiempo, la amplitud del intervalo de predicción debería cambiar a lo largo del tiempo: en los periodos en los que la inflación está sujeta a perturbaciones o *shocks* particularmente grandes, el intervalo debería ser amplio; durante los periodos de relativa tranquilidad, el intervalo debería ser más estrecho.

La volatilidad agrupada puede interpretarse como un agrupamiento de la varianza del término de error a lo largo del tiempo: si el error de regresión presenta una varianza pequeña en un periodo, su varianza tiende asimismo a ser pequeña en el periodo siguiente. En otras palabras, la volatilidad agrupada implica que el error muestra una heterocedasticidad cambiante a lo largo del tiempo.

Heterocedasticidad condicional autorregresiva

Dos de los modelos planteados para la volatilidad agrupada son el **modelo de heterocedasticidad condicional autorregresiva (ARCH)** y su extensión, el **modelo ARCH generalizado (GARCH)**.

ARCH. Considérese la regresión ARD(1,1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t. \quad (16.29)$$

En el modelo ARCH, que fue desarrollado por el económetra Robert Engle (Engle, 1982; véase el recuadro sobre Clive Granger y Robert Engle), el error u_t se modeliza de forma que esté normalmente distribuido con media igual a cero y varianza igual a σ_t^2 , donde σ_t^2 depende de los valores pasados de u_t elevados al cuadrado. En concreto, el modelo ARCH de orden p , que se expresa como ARCH(p), es

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_p u_{t-p}^2, \quad (16.30)$$

donde $\alpha_0, \alpha_1, \dots, \alpha_p$ son coeficientes desconocidos. Si estos coeficientes son positivos, y si los últimos errores al cuadrado son elevados el modelo ARCH predice que el actual error al cuadrado será de cuantía elevada en el sentido de que su varianza, σ_t^2 , será grande.

A pesar de que aquí está descrito para el modelo ARD(1,1) de la Ecuación (16.29), el modelo ARCH se puede aplicar a la varianza del error de cualquier modelo de regresión de series temporales con un error que tenga una media condicional igual a cero, incluyendo los modelos ARD de orden superior, los modelos autorregresivos, y las regresiones de series de temporales con varios predictores.

GARCH. El modelo ARCH generalizado (GARCH), desarrollado por el económetra Tim Bollerslev (1986), es una extensión del modelo ARCH que permite que σ_t^2 dependa de sus propios retardos, así como de los retardos del cuadrado del error. El modelo GARCH (p, q) es

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2 + \phi_1 \sigma_{t-1}^2 + \dots + \phi_q \sigma_{t-q}^2, \quad (16.31)$$

donde $\alpha_0, \alpha_1, \dots, \alpha_p, \phi_1, \dots, \phi_q$ son coeficientes desconocidos.

El modelo ARCH es análogo a un modelo de retardos distribuidos, y el modelo GARCH es análogo a un modelo ARD. Tal y como se analizó en el Apéndice 15.2, el modelo ARD (cuando corresponda) puede proporcionar un modelo más parsimonioso de multiplicadores dinámicos que el modelo de retardos distribuidos. Del mismo modo, mediante la incorporación de los retardos de σ_t^2 , el modelo GARCH puede captar las varianzas que cambian lentamente con un menor número de parámetros que el modelo ARCH.

Una aplicación importante de los modelos ARCH y GARCH es la de medir y predecir la volatilidad que cambia en el tiempo de la rentabilidad de los activos financieros, en particular de los activos que se observan con altas frecuencias muestrales como es la rentabilidad diaria de las acciones de la Figura 16.3. En esos casos, a menudo se modeliza la rentabilidad en sí misma como impredecible, por lo que la regresión de la Ecuación (16.29) solo incluye el término independiente.

Estimación e inferencia. Los modelos ARCH y GARCH se estiman mediante el método de máxima verosimilitud (Apéndice 11.2). Los estimadores de los coeficientes ARCH y GARCH se distribuyen normalmente en muestras grandes, por lo que en muestras grandes los estadísticos t presentan distribuciones normales estándar, y se pueden construir los intervalos de confianza como la estimación de máxima verosimilitud $\pm 1,96$ errores estándar.

Aplicación a la volatilidad de las cotizaciones de valores

Un modelo GARCH (1,1) para las variaciones diarias en términos porcentuales del índice NYSE de los precios de las acciones, R_t estimado utilizando datos de todos los días hábiles desde el 2 de enero de 1990 hasta el 11 de noviembre de 2005, es

$$\hat{R}_t = 0,049 \quad (16.32)$$

$$(0,012)$$

$$\hat{\sigma}_t^2 = 0,0079 + 0,072u_{t-1}^2 + 0,919\sigma_{t-1}^2 \quad (16.33)$$

$$(0,0014) \quad (0,005) \quad (0,006)$$

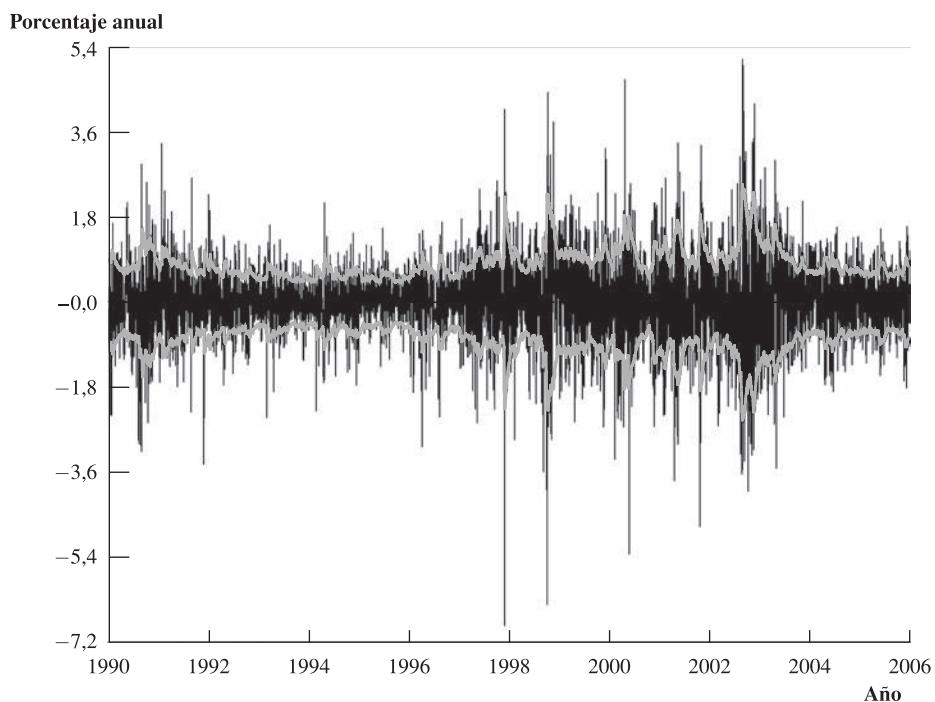
No aparecen predictores retardados en la Ecuación (16.32), debido a que las variaciones diarias de los precios del NYSE son en esencia impredecibles.

Los dos coeficientes del modelo GARCH (los coeficientes de u_{t-1}^2 y σ_{t-1}^2) son ambos estadísticamente significativos considerados de forma individual al nivel de significación del 5 %. Una medida de la persistencia de los movimientos de la varianza es la suma de los coeficientes de u_{t-1}^2 y σ_{t-1}^2 del modelo GARCH (Ejercicio 16.9). Esta suma (0,991) es grande, lo que indica que las variaciones de la varianza condicional son persistentes. Dicho de otro modo, el modelo GARCH estimado implica que los períodos de alta volatilidad en los precios de las acciones de la Bolsa de Nueva York serán duraderos. Esta conclusión es coherente con los largos períodos de volatilidad agrupada que se observan en la Figura 16.3.

La varianza condicional estimada para el periodo t , $\hat{\sigma}_t^2$, se puede calcular utilizando los residuos de la Ecuación (16.32) y los coeficientes de la Ecuación (16.33). En la Figura 16.4 se representan gráficamente las bandas que indican más y menos una desviación típica condicional (es decir, $\pm \hat{\sigma}_t^2$), en base al modelo GARCH (1,1), junto con las desviaciones de la serie de variaciones en los precios en términos porcentuales respecto de su media. Las bandas de desviación típica condicional cuantifican la volatilidad que varía a lo largo del tiempo de las variaciones diarias en los precios. A mediados de la década de 1990, las bandas de desviación típica condicional son estrechas, lo que indica menores niveles de riesgo para los inversores en el índice NYSE. Por contra, alrededor del cambio de siglo, estas bandas de desviación típica condicional son amplias, lo que indica un periodo de mayor volatilidad diaria en los precios de las acciones.

16.6 Conclusión

Esta parte del libro ha cubierto el análisis de algunas de las herramientas más utilizadas, así como de algunos conceptos de regresión de series temporales. Muchas otras herramientas para el análisis de series

FIGURA 16.4 Variación porcentual diaria del índice NYSE y bandas GARCH (1,1)

Las bandas GARCH (1,1), que son $\pm \hat{\sigma}_t$, donde $\hat{\sigma}_t$ se calcula utilizando la Ecuación (16.33), son estrechas cuando la varianza condicional es pequeña y anchas cuando es grande. La volatilidad condicional de las variaciones en el precio de las acciones varía considerablemente a lo largo del periodo 1990-2005.

temporales han sido desarrolladas para casos y aplicaciones específicas. En caso de estar interesado en aprender más sobre predicción económica, véanse los libros de texto de nivel introductorio de Enders (1995) y Diebold (2007). Para un tratamiento avanzado de la econometría de datos de series temporales, véase Hamilton (1994).

Resumen

1. El modelo de vectores autorregresivos es un «vector» de k variables de series temporales, cada una de las cuales depende de sus propios retardos y de los retardos de las $k - 1$ series restantes. Las predicciones de cada una de las series temporales resultantes de un VAR son mutuamente consistentes, en el sentido de que están basadas en la misma información.
2. Las predicciones con un horizonte temporal de dos o más periodos vista se puede calcular ya sea mediante la iteración hacia delante de un modelo con un horizonte temporal de un periodo vista (un AR o un VAR), o mediante la estimación de una regresión con un horizonte temporal de varios periodos vista (multiperíodo).
3. Dos series que comparten una tendencia estocástica común están cointegradas; es decir, Y_t y X_t están cointegradas si Y_t y X_t son $I(1)$, pero $Y_t - \theta X_t$ es $I(0)$. Si Y_t y X_t están cointegradas, el término de corrección de error $Y_t - \theta X_t$ puede ayudar a predecir la variable Y_t y/o la variable X_t . Un modelo de vector de corrección de error es un modelo VAR de Y_t y X_t , ampliado para incluir el término de corrección de error retardado.
4. La volatilidad agrupada —cuando la varianza de una serie es elevada en algunos periodos y baja en otros— es habitual en las series temporales económicas, sobre todo en series temporales financieras.

5. El modelo ARCH de volatilidad agrupada expresa la varianza condicional del error de regresión como función de los errores de regresión recientes al cuadrado. El modelo GARCH amplía el modelo ARCH para incluir asimismo los retardos de la varianza condicional. Los modelos ARCH y GARCH estimados dan lugar a intervalos de predicción con amplitudes que dependen de la volatilidad de los errores de regresión más recientes.

Términos clave

vector autorregresivo (VAR) (456)	cointegración (468)
predicción multiperodo iterada AR(461)	coeficiente de cointegración (468)
predicción multiperodo iterada VAR (461)	término de corrección de error (469)
predicción multiperiodo directa (462)	modelo del vector de corrección de error (469)
segunda diferencia (463)	contraste EG-ADF (470)
integrada de orden d , $I(d)$ (464)	estimador MCO dinámico (MCOD) (472)
integrada de orden cero [$I(0)$], uno [$I(1)$], o dos [$I(2)$] (464)	volatilidad agrupada (475)
orden de integración (464)	heterocedasticidad condicional autorregresiva (ARCH) (476)
contraste DF-MCG (465)	ARCH generalizado (GARCH) (476)
tendencia común (468)	

Revisión de conceptos

- 16.1** Un macroeconomista quiere realizar previsiones macroeconómicas para las siguientes variables: PIB, consumo, inversión, consumo público, exportaciones, importaciones, tipos de interés a corto plazo, tipos de interés de largo plazo y tasa de inflación de los precios. Dispone de series temporales trimestrales para cada una de estas variables desde 1970 hasta 2010. ¿Debería estimar un VAR para estas variables y utilizarlo para la predicción? ¿Por qué o por qué no? ¿Puede sugerir un método alternativo?
- 16.2** Supóngase que Y_t sigue un modelo AR(1) estacionario con $\beta_0 = 0$ y $\beta_1 = 0,7$. Si $Y_t = 5$, ¿cuál es la predicción para Y_{t+2} (es decir, cómo es $Y_{t+2|t}$)? ¿Cómo es $Y_{t+h|t}$ para $h = 30$? ¿Le parece razonable la predicción para $h = 30$?
- 16.3** Una versión de la teoría del consumo de la renta permanente implica que el logaritmo del PIB real (Y) y el logaritmo del consumo real (C) están cointegrados con un coeficiente de cointegración igual a 1. Explique cómo se podría investigar esta implicación mediante (a) la representación gráfica de los datos y (b) la utilización de un contraste estadístico.
- 16.4** Considérese el modelo ARCH, $\sigma_t^2 = 1,0 + 0,8u_{t-1}^2$. Explique por qué esto lleva a la volatilidad agrupada. (Sugerencia: ¿Qué ocurre cuando u_{t-1}^2 es inusualmente grande?).
- 16.5** El contraste de raíz unitaria DF-MCG tiene una potencia mayor que la del contraste de Dickey-Fuller. ¿Por qué debe utilizarse un contraste más potente?

Ejercicios

- 16.1** Supóngase que Y_t sigue un proceso AR(1) estacionario $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$.
- Demuestre que la predicción a h periodos vista de Y_t está dada por $Y_{t+h|t} = \mu_Y + \beta_1^h(Y_t - \mu_Y)$, donde $\mu_Y = \beta_0/(1 - \beta_1)$.
 - Supóngase que X_t está relacionada con Y_t mediante $X_t = \sum_{i=0}^{\infty} \delta^i Y_{t+i|t}$, donde $|\delta| < 1$. Demuestre que $X_t = \mu_Y/(1 - \delta) + (Y_t - \mu_Y)/(1 - \beta_1 \delta)$.
- 16.2** Una versión de la teoría de las expectativas de la estructura temporal de tipos de interés sostiene que el tipo de largo plazo es igual al promedio de los valores esperados de los tipos de interés de corto plazo en el futuro, más una prima que es $I(0)$. En concreto, sea Rk_t la expresión del tipo de interés del

periodo k , sea $R1_t$ la expresión del tipo de interés de un periodo, y sea e_t la expresión de la prima $I(0)$. Entonces $Rk_t = \frac{1}{k} \sum_{i=0}^{k-1} R1_{t+1|t} + e_t$, donde $R1_{t+1|t}$ es la predicción realizada en el periodo t acerca del valor de $R1$ en el periodo $t+1$. Supóngase que $R1_t$ sigue un paseo aleatorio, por lo que $R1_t = R1_{t-1} + u_t$.

- a) Demuestre que $Rk_t = R1_t + e_t$.
 - b) Demuestre que Rk_t y $R1_t$ están cointegrados. ¿Cuál es el coeficiente de cointegración?
 - c) Supóngase ahora que $\Delta R1_t = 0,5\Delta R1_{t-1} + u_t$. ¿De qué manera cambia su respuesta en el apartado (b)?
 - d) Supóngase ahora que $R1_t = 0,5R1_{t-1} + u_t$. ¿De qué manera cambia su respuesta en el apartado (b)?
- 16.3** Supóngase que u_t sigue el proceso ARCH, $\sigma_t^2 = 1,0 + 0,5u_{t-1}^2$.
- a) Sea $E(u_t^2) = \text{var}(u_t)$ la varianza no condicionada de u_t . Demuestre que $\text{var}(u_t) = 2$. (*Sugerencia:* utilice la ley de las esperanzas iteradas $E(u_t^2) = E[E(u_t^2|u_{t-1})]$).
 - b) Supóngase que la distribución de u_t condicionada a los valores retardados de u_t es $N(0, \sigma_t^2)$. Si $u_{t-1} = 0,2$, ¿cuál es la $\Pr(-3 \leq u_t \leq 3)$? Si $u_{t-1} = 2,0$, ¿cuál es la $\Pr(-3 \leq u_t \leq 3)$?
- 16.4** Supóngase que Y_t sigue el modelo AR(p) $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + u_t$, donde $E(u_t|Y_{t-1}, Y_{t-2}, \dots) = 0$. Sea $Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, \dots)$. Demuestre que $Y_{t+h|t} = \beta_0 + \beta_1 Y_{t-1+h|t} + \dots + \beta_p Y_{t-p+h|t}$ para $h > p$.
- 16.5** Verifique la Ecuación (16.20). [*Sugerencia:* utilice $\sum_{i=1}^T Y_t^2 = \sum_{t=1}^T (Y_{t-1} + \Delta Y_t)^2$ para demostrar que $\sum_{t=1}^T Y_t^2 = \sum_{t=1}^T Y_{t-1}^2 + 2 \sum_{i=1}^T Y_{t-1} \Delta Y_t + \sum_{t=1}^T \Delta Y_t^2$ y resuelva para $\sum_{t=1}^T Y_{t-1} \Delta Y_t$].
- 16.6** Una regresión de Y_t sobre los valores actuales, pasados y futuros de X_t da lugar a

$$Y_t = 3,0 + 1,7X_{t+1} + 0,8X_t - 0,2X_{t-1} + u_t.$$

- a) Reordene la regresión de modo que adquiera la forma mostrada en la Ecuación (16.25). ¿Cuáles son los valores de, $\theta, \delta_{-1}, \delta_0$, y δ_1 ?
 - b) I) Supóngase que X_t es $I(1)$ y u_t es $I(1)$. ¿Están X e Y cointegradas?
II) Supóngase que X_t es $I(0)$ y u_t es $I(1)$. ¿Están X e Y cointegradas?
III) Supóngase que X_t es $I(1)$ y u_t es $I(0)$. ¿Están X e Y cointegradas?
- 16.7** Supóngase que $\Delta Y_t = u_t$, donde u_t es $N(0, 1)$ e i.i.d., y considérese la regresión $Y_t = \beta X_t + \text{error}$, donde $X_t = \Delta Y_{t+1}$ y error es el error de la regresión. Demuestre que $\hat{\beta} \xrightarrow{d} \frac{1}{2}(\chi_1^2 - 1)$. [*Sugerencia:* analice el numerador de $\hat{\beta}$ mediante un análisis como el de la Ecuación (16.21). Analice el denominador utilizando la ley de los grandes números].
- 16.8** Considérese el siguiente modelo VAR de dos variables con un retardo y sin término independiente:

$$Y_t = \beta_{11} Y_{t-1} + \gamma_{11} X_{t-1} + u_{1t}$$

$$X_t = \beta_{21} Y_{t-1} + \gamma_{21} X_{t-1} + u_{2t}$$

- a) Demuestre que la predicción iterada a dos periodos vista para Y se puede escribir como $Y_{t/2} = \delta_1 Y_{t-2} + \delta_2 X_{t-2}$ y obtenga valores para δ_1 y δ_2 en términos de los coeficientes del VAR.
 - b) A la luz de la respuesta en el apartado (a) ¿serán distintas las predicciones multiperiodo iteradas de las predicciones multiperiodo directas? Explíquelo.
- 16.9**
- a) Supóngase que $E(u_t|u_{t-1}, u_{t-2}, \dots) = 0$, que $\text{var}(u_t|u_{t-1}, u_{t-2}, \dots)$ sigue el modelo ARCH(1) $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$, y que el proceso de u_t es estacionario. Demuestre que $\text{var}(u_t) = \alpha_0/(1 - \alpha_1)$. (*Sugerencia:* utilice la ley de esperanzas iteradas $E(u_t^2) = E[E(u_t^2|u_{t-1})]$).
 - b) Extienda el resultado obtenido en el apartado (a) al modelo ARCH(p).
 - c) Demuestre que $\sum_{i=1}^p \alpha_i < 1$ para un modelo ARCH(p) estacionario.
 - d) Amplíe el resultado obtenido en el apartado (a) al modelo GARCH (1,1).
 - e) Demuestre que $\alpha_1 + \phi_1 < 1$ para un modelo GARCH (1,1) estacionario.

- 16.10** Considérese el modelo cointegrado $Y_t = \theta X_t + v_{1t}$ y $X_t = X_{t-1} + v_{2t}$, con v_{1t} y v_{2t} variables aleatorias serialmente incorrelacionadas con media igual a cero y con $E(v_{1t}v_{2j}) = 0$ para todo t y j . Deduzca el modelo del vector de corrección de error [Ecuaciones (16.22) y (16.23)] para X e Y .

Ejercicios empíricos

Estos ejercicios se basan en las series de datos de los archivos de datos **USMacro_Quarterly** y **USMacro_Monthly** descritas en los Ejercicios empíricos de los Capítulos 14 y 15. Sea $Y_t = \ln(PIB_t)$ (GDP_t en la serie original de la base de datos), R_t es el tipo de interés de las letras del Tesoro a 3 meses, y π_t^{IPC} y π_t^{DGCP} son las tasas de inflación del IPC y del Deflactor del Gasto de Consumo Personal (DGCP), respectivamente.

- E16.1** Utilizando los datos trimestrales disponibles desde 1955:1 hasta 2009:4, estime un VAR(4) (un VAR con cuatro retardos) para las variables ΔY_t y ΔR_t .
- ¿La variable ΔR causa en el sentido de Granger a la variable ΔY ? ¿La variable ΔY causa en el sentido de Granger a la variable ΔR ?
 - ¿Debería incluir el VAR más de cuatro retardos?
- E16.2** En este ejercicio se calculan las predicciones pseudo fuera de la muestra con un horizonte de dos trimestres vista para la variable ΔY comenzando en 1989:4 hasta el final de la muestra. (Es decir, se calcula $\Delta Y_{1990:2|1989:4}$, $\Delta Y_{1990:3|1990:1}$, etc.)
- Formule predicciones iteradas pseudo fuera de la muestra para un horizonte de dos trimestres vista utilizando un modelo AR(1).
 - Formule predicciones iteradas pseudo fuera de la muestras para un horizonte de dos trimestres vista utilizando un modelo VAR(4) para ΔY y ΔR .
 - Formule predicciones iteradas pseudo fuera de la muestras para un horizonte de dos trimestres vista utilizando la predicción simplista $\Delta Y_{t+2|t} = (\Delta Y_t + \Delta Y_{t-1} + \Delta Y_{t-2} + \Delta Y_{t-3})/4$.
 - ¿Qué modelo tiene la menor raíz del error cuadrático medio de predicción?
- E16.3** Utilice el contraste DF-MCG para contrastar la presencia de raíz unitaria autorregresiva en la variable Y_t . Como alternativa, supongamos que Y_t es estacionaria alrededor de una tendencia determinística. Compare los resultados con los resultados obtenidos en el Ejercicio empírico 14.3.
- E16.4** En el Ejercicio empírico 15.2, se estudiaba el comportamiento de $\pi_t^{IPC} - \pi_t^{DGCP}$ a lo largo del periodo muestral 1970:1 hasta 2009:12. Ese análisis se basaba en la suposición de que $\pi_t^{IPC} - \pi_t^{DGCP}$ es $I(0)$.
- Contraste la presencia de raíz unitaria en la autorregresión para $\pi_t^{IPC} - \pi_t^{DGCP}$. Realice el contraste utilizando el contraste ADF que incluye una constante y 12 retardos de la primera diferencia de $\pi_t^{IPC} - \pi_t^{DGCP}$. Realice asimismo el contraste utilizando el procedimiento DF-MCG.
 - Contraste la presencia de raíz unitaria en la autorregresión de π_t^{IPC} y en la autorregresión de π_t^{DGCP} . Al igual que en el apartado (a), utilice tanto el contraste ADF como el DF-MCG incluyendo una constante y los 12 retardos de las primeras diferencias.
 - ¿Qué dicen los resultados de los apartados (a) y (b) acerca de la cointegración entre estas dos tasas de inflación? ¿Cuál es el valor del coeficiente de cointegración (θ) implícito en las respuestas de los apartados (a) y (b)?
 - Supongamos que se desconoce que el coeficiente de cointegración es $\theta = 1$. ¿Cómo contrastaría la presencia de cointegración? Realice el contraste. ¿Cómo estimaría θ ? Estime el valor de θ mediante la regresión MCOD de π_t^{IPC} sobre π_t^{DGCP} y sobre seis adelantos y retardos de $\Delta \pi_t^{DGCP}$. ¿Está el valor estimado de θ cercano a 1?
- E16.5**
- Utilizando los datos de la variable ΔY (la tasa de crecimiento del PIB) desde 1955:1 hasta 2009:4, estime un modelo AR(1) con errores GARCH(1,1).
 - Represente gráficamente los residuos del modelo AR(1), junto con las bandas $\pm \hat{\sigma}_t$ como en la Figura 16.4.

- c) Algunos macroeconomistas han afirmado que hubo una fuerte caída en la variabilidad de la variable ΔY alrededor de 1983, que denominan la «Gran Moderación». ¿Es evidente esta Gran Moderación en el gráfico realizado en el apartado (b)?

APÉNDICE

16.1 Datos financieros de EE.UU. utilizados en el Capítulo 16

Los tipos de interés de las Letras del Tesoro de EE.UU. a 3 meses y de los Bonos del Tesoro de Estados Unidos a 1 año son los promedios mensuales de los tipos diarios, convertidos a una base anual, de acuerdo con lo expresado por la Junta de Gobernadores de la Reserva Federal de EE.UU. Los datos trimestrales utilizados en este capítulo son la media mensual de los tipos de interés para el último mes del trimestre.

¿Por qué debería un económetra aplicado interesarse en aprender algo de teoría econométrica? Existen varias razones. Aprender teoría econométrica convierte la «caja negra» que es el software estadístico en un conjunto de herramientas flexibles del que se puede extraer la herramienta más adecuada para realizar el trabajo que nos ocupe. La comprensión de la teoría econométrica ayuda a apreciar por qué estas herramientas funcionan y permite comprender cuáles son los supuestos necesarios para que cada una de las herramientas funcione correctamente. Quizás lo más importante sea que el conocimiento de la teoría econométrica ayuda a reconocer cuándo una herramienta no va a funcionar correctamente en un caso concreto y cuándo debe acudirse a un método econométrico diferente.

Este capítulo proporciona una introducción a la teoría econométrica de regresión lineal con un único regresor. Esta introducción tiene por objeto complementar, no sustituir, el material de los Capítulos 4 y 5, que por tanto deben haber sido leídos con anterioridad.

En este capítulo se amplían los contenidos de los Capítulos 4 y 5 en dos sentidos.

En primer lugar, proporcionando un tratamiento matemático a la distribución muestral del estimador MCO y del estadístico t , tanto en muestras grandes bajo los tres supuestos de mínimos cuadrados del Concepto clave 4.3 como en muestras finitas bajo los dos supuestos adicionales de homocedasticidad y normalidad en los errores. Estos cinco supuestos ampliados de mínimos cuadrados se encuentran en la Sección 17.1. En las Secciones 17.2 y 17.3, además de en el Apéndice 17.2, se desarrollan matemáticamente las distribuciones normales para muestras grandes del estimador MCO y del estadístico t bajo los tres primeros supuestos (los supuestos de mínimos cuadrados del Concepto clave 4.3). En la Sección 17.4 se deducen las distribuciones exactas del estimador MCO y del estadístico t bajo los dos supuestos adicionales de homocedasticidad y errores normalmente distribuidos.

En segundo lugar, este capítulo amplía los Capítulos 4 y 5 al ofrecer un método alternativo para el tratamiento de la heterocedasticidad. El método de los Capítulos 4 y 5 consistía en utilizar los errores estándar heterocedástico-robustos para asegurar que la inferencia estadística es válida incluso si los errores son heterocedásticos. Sin embargo, este método tiene un coste: si los errores son heterocedásticos, entonces en teoría existe un estimador más eficiente que MCO. Este estimador, denominado de mínimos cuadrados ponderados, se presenta en la Sección 17.5. Los mínimos cuadrados ponderados requieren una gran cantidad de información previa acerca de la naturaleza exacta de la heterocedasticidad, es decir, sobre la varianza condicional de u dado X . Cuando esta información se encuentra disponible, los mínimos cuadrados ponderados mejoran el método MCO. Sin embargo, en la mayor parte de las ocasiones, esta información no se encuentra disponible; en esos casos, resulta preferible utilizar el método MCO con errores estándar heterocedástico-robustos.

17.1 Los supuestos ampliados de mínimos cuadrados y el estimador MCO

En esta sección se presenta un conjunto de supuestos que amplían y refuerzan los tres supuestos de mínimos cuadrados del Capítulo 4. Estos supuestos más fuertes se utilizan en las secciones siguientes para obtener mejores resultados teóricos acerca del estimador MCO de lo que resulta posible en el marco de los supuestos más débiles (pero más realistas) del Capítulo 4.

Los supuestos ampliados de mínimos cuadrados

Supuestos ampliados de mínimos cuadrados #1, #2 y #3. Los tres primeros supuestos ampliados de mínimos cuadrados son los tres supuestos que figuran en el Concepto clave 4.3: que la media condicional de u_i , dado X_i , es igual a cero; que (X_i, Y_i) , donde $i = 1, \dots, n$ son extracciones i.i.d. de su distribución conjunta; y que X_i y u_i tienen momentos de cuarto orden (finitos).

Bajo estos tres supuestos, el estimador MCO es insesgado, es consistente, y tiene una distribución muestral asintóticamente normal. Si estos tres supuestos se cumplen, entonces los métodos de inferencia introducidos en el Capítulo 4 —la contrastación de hipótesis mediante el estadístico t y la construcción de intervalos de confianza del 95 % como $\pm 1,96$ errores estándar— están justificados si el tamaño de la muestra es grande. No obstante, para desarrollar una teoría de estimación eficiente mediante MCO o para caracterizar la distribución muestral exacta del estimador MCO, son necesarios unos supuestos más fuertes.

Supuesto ampliado de mínimos cuadrados #4. El cuarto supuesto ampliado de mínimos cuadrados es que u_i es homocedástico; es decir, $\text{var}(u_i|X_i) = \sigma_u^2$, donde σ_u^2 es una constante. Como se analizó en la Sección 5.5, si este supuesto adicional se cumple, el estimador MCO es eficiente entre todos los estimadores lineales e insesgados, condicionado a X_1, \dots, X_n .

Supuesto ampliado de mínimos cuadrados #5. El quinto supuesto ampliado de mínimos cuadrados es que la distribución condicional de u_i , dado X_i , es normal.

Bajo los supuestos de mínimos cuadrados #1 y #2 y los supuestos ampliados de mínimos cuadrados #4 y #5, u_i es $N(0, \sigma_u^2)$ i.i.d., y u_i y X_i están independientemente distribuidas. Para comprobarlo, debe tenerse en cuenta que el quinto supuesto ampliado de mínimos cuadrados establece que la distribución condicional de $u_i|X_i$ es $N(0, \text{var}(u_i|X_i))$, donde la distribución tiene media igual a cero por el primer supuesto ampliado de mínimos cuadrados. No obstante, por el cuarto supuesto de mínimos cuadrados, $\text{var}(u_i|X_i) = \sigma_u^2$, por lo que la distribución condicional de $u_i|X_i$ es $N(0, \sigma_u^2)$. Debido a que esta distribución condicional no depende de X_i , u_i y X_i están distribuidas independientemente. Por el segundo supuesto de mínimos cuadrados, u_i está independientemente distribuido de u_j para todo $j \neq i$. Se deduce que, bajo los supuestos ampliados de mínimos cuadrados #1, #2, #4, y #5, u_i y X_i son variables que están independientemente distribuidas y u_i es $N(0, \sigma_u^2)$ i.i.d.

En la Sección 17.4 se demuestra que, si se cumplen los cinco supuestos ampliados de mínimos cuadrados, el estimador MCO tiene una distribución muestral exacta normal y el estadístico t válido con homocedasticidad tiene un distribución exacta t de Student.

Los supuestos ampliados de mínimos cuadrados cuarto y quinto son mucho más restrictivos que los tres primeros. Aunque podría resultar razonable suponer que se cumplen los tres primeros supuestos en un caso concreto, los dos últimos supuestos son menos realistas. Incluso aunque en la práctica estos dos últimos supuestos no se cumplan, tendrían un interés teórico debido a que si uno de los dos o los dos se cumplen, el estimador MCO presenta propiedades adicionales además de las estudiadas en los Capítulos 4 y 5. Por tanto se puede enriquecer la comprensión del estimador MCO, y en general de la teoría de la estimación en el modelo de regresión lineal, analizando la estimación bajo estos supuestos más fuertes.

Los cinco supuestos ampliados de mínimos cuadrados para el modelo con una única variable explicativa se recogen en el Concepto clave 17.1.

El estimador MCO

Para facilitar la consulta, repetimos los estimadores MCO de β_0 y β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (17.2)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (17.3)$$

Las Ecuaciones (17.2) y (17.3) se deducen en el Apéndice 4.2.

**CONCEPTO
CLAVE**
17.1
Los supuestos ampliados de mínimos cuadrados para el modelo de regresión con regresor único

El modelo de regresión con una única variable explicativa es

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n. \quad (17.1)$$

Los supuestos ampliados de mínimos cuadrados son

1. $E(u_i|X_i) = 0$ (media condicional igual a cero).
2. $(X_i, Y_i), i = 1, \dots, n$, son extracciones independientes e idénticamente distribuidas (i.i.d.) obtenidas a partir de su distribución conjunta.
3. (X_i, u_i) tienen momentos de cuarto orden finitos y distintos de cero.
4. $\text{var}(u_i|X_i) = \sigma_u^2$ (homocedasticidad); y
5. La distribución condicional de u_i dado X_i es normal (errores normales).

17.2 Fundamentos de teoría de distribución asintótica

La teoría de distribución asintótica es la teoría de la distribución de estadísticos —estimadores, estadísticos de contraste, e intervalos de confianza— cuando el tamaño muestral es grande. De modo formal, esta teoría supone la caracterización del comportamiento de la distribución muestral de un estadístico de acuerdo con una secuencia de muestras aun más grandes. La teoría es asintótica en el sentido de que caracteriza el comportamiento del estadístico en el límite, a medida que $n \rightarrow \infty$.

Aunque las muestras grandes no son nunca, por supuesto, infinitas, la teoría de distribución asintótica interpreta un papel central en econometría y en estadística por dos razones. En primer lugar, si el número de observaciones utilizadas en una aplicación empírica es grande, el límite asintótico puede proporcionar una aproximación de alta calidad a la distribución en muestras finitas. En segundo lugar, las distribuciones muestrales asintóticas habitualmente son más sencillas, y por tanto más fáciles de utilizar en la práctica, que las distribuciones exactas en muestras finitas. En conjunto, estas dos razones suponen que los métodos fiables y directos para la realizar inferencia estadística —contrastación mediante estadísticos t e intervalos de confianza al 95 % calculados como $\pm 1,96$ errores estándar— pueden estar basados en las distribuciones muestrales aproximadas obtenidas mediante la teoría asintótica.

Las dos piedras angulares de la teoría de distribución asintótica son la ley de los grandes números y el teorema central del límite, que se presentaron en la Sección 2.6. Esta sección comienza con la continuación del análisis de la ley de los grandes números y con el teorema central del límite, incluyendo una demostración de la ley de los grandes números. Posteriormente se introducen dos herramientas adicionales, el teorema de Slutsky y el teorema de la función continua, que amplían la utilidad de la ley de los grandes números y del teorema central del límite. A modo de ilustración, se utilizan estas herramientas para demostrar que la distribución del estadístico t basado en \bar{Y} para el contraste de la hipótesis de que $E(Y) = \mu_0$ presenta una distribución normal estándar bajo la hipótesis nula.

La convergencia en probabilidad y la ley de los grandes números

Los conceptos de convergencia en probabilidad y la ley de los grandes números se introdujeron en la Sección 2.6. Ahora se proporciona una definición matemática exacta de la convergencia en probabilidad, seguida del enunciado y la demostración de la ley de los grandes números.

Consistencia y convergencia en probabilidad. Sea $S_1, S_2, \dots, S_n, \dots$ una secuencia de variables aleatorias. Por ejemplo, S_n podría ser la media muestral \bar{Y} de una muestra de n observaciones de la variable aleatoria Y . La secuencia de variables aleatorias $\{S_n\}$ se dice que **converge en probabilidad** a un límite, μ

(es decir, $S_n \xrightarrow{P} \mu$), si la probabilidad de que S_n se encuentre a una distancia menor o igual a $\pm\delta$ de μ tiende a 1 a medida que $n \rightarrow \infty$, siendo δ positivo. Es decir,

$$S_n \xrightarrow{P} \mu \quad \text{si y solo si} \quad \Pr(|S_n - \mu| \geq \delta) \rightarrow 0 \quad (17.4)$$

Si $n \rightarrow \infty$ para todo $\delta > 0$. Si $S_n \xrightarrow{P} \mu$, entonces S_n se dice que es un **estimador consistente** de μ .

La ley de los grandes números. La ley de los grandes números establece que, bajo ciertas condiciones sobre Y_1, \dots, Y_n , la media muestral \bar{Y} converge en probabilidad a la media poblacional. Dentro de la teoría de probabilidad se han desarrollado distintas versiones de la ley de los grandes números, que se corresponden con las distintas condiciones sobre Y_1, \dots, Y_n . La versión de la ley de los grandes números que se utiliza en este libro es que Y_1, \dots, Y_n son extracciones i.i.d. de una distribución con una varianza finita. La ley de los grandes números (enunciada asimismo en el Concepto clave 2.6) es

$$\text{si } Y_1, \dots, Y_n \text{ son i.i.d., } E(Y_i) = \mu_Y, \text{ y } \text{var}(Y_i) < \infty, \text{ entonces } \bar{Y} \xrightarrow{P} \mu_Y \quad (17.5)$$

La idea de la ley de los grandes números puede verse en la Figura 2.8: a medida que aumenta el tamaño muestral, la distribución muestral de \bar{Y} se concentra en torno a la media poblacional, μ . Una característica de la distribución muestral es que la varianza de \bar{Y} disminuye al aumentar el tamaño muestral; otra característica es que la probabilidad de que \bar{Y} esté más allá de una distancia $\pm\delta$ de μ , se desvanece a medida que n aumenta. Estas dos características de la distribución muestral están en realidad vinculadas, y la demostración de la ley de los grandes números aprovecha este vínculo.

Demostración de la ley de los grandes números. El vínculo entre la varianza de \bar{Y} y la probabilidad de que \bar{Y} se encuentre a una distancia de μ_Y menor de $\pm\delta$ lo proporciona la desigualdad de Chebychev, que se enuncia y se demuestra en el Apéndice 17.2 [véase la Ecuación (17.42)]. Expresada en términos de \bar{Y} , la desigualdad de Chebychev es

$$\Pr(|\bar{Y} - \mu_Y| \geq \delta) \leq \frac{\text{var}(\bar{Y})}{\delta^2} \quad (17.6)$$

para cualquier constante positiva δ . Debido a que Y_1, \dots, Y_n son i.i.d. con una varianza σ_Y^2 , $\text{var}(\bar{Y}) = \sigma_Y^2/n$; por tanto, para cualquier $\delta > 0$, $\text{var}(\bar{Y})/\delta^2 = \sigma_Y^2/(\delta^2 n) \rightarrow 0$. Se desprende de la Ecuación (17.6) que $\Pr(|\bar{Y} - \mu_Y| \geq \delta) \rightarrow 0$ para todo $\delta > 0$, lo que demuestra la ley de los grandes números.

Algunos ejemplos. La consistencia es un concepto fundamental en teoría de distribución asintótica, por lo que a continuación se presentan algunos ejemplos de estimadores consistentes e inconsistentes de la media poblacional, μ_Y . Supongamos que $Y_i, i = 1, \dots, n$ son i.i.d. con una varianza σ_Y^2 que es positiva y finita. Consideremos los tres estimadores siguientes de μ_Y : (1) $m_a = Y_1$; (2) $m_b = \left(\frac{1-a^n}{1-a}\right) \sum_{i=1}^n a^{i-1} Y_i$, donde $0 < a < 1$; y (3) $m_c = \bar{Y} + 1/n$. Son consistentes estos estimadores?

El primer estimador m_a es exactamente la primera observación, por lo que $E(m_a) = E(Y_1) = \mu_Y$ y m_a es insesgado. No obstante, m_a no es consistente: $\Pr(|m_a - \mu_Y| \geq \delta) = \Pr(|Y_1 - \mu_Y| \geq \delta)$, que debe ser positiva para un δ suficientemente grande (debido a que $\sigma_Y^2 > 0$), por lo que $\Pr(|m_a - \mu_Y| \geq \delta)$ no tiende a cero a medida que $n \rightarrow \infty$, por lo que m_a no es consistente. Esta inconsistencia no debería resultar sorprendente: debido a que m_a utiliza la información contenida en una única observación, su distribución no puede concentrarse en torno a μ_Y a medida que aumenta el tamaño muestral.

El segundo estimador, m_b , es insesgado pero no es consistente. Es insesgado debido a que

$$\begin{aligned} E(m_b) &= E\left[\left(\frac{1-a^n}{1-a}\right)^{-1} \sum_{i=1}^n a^{i-1} Y_i\right] = \left(\frac{1-a^n}{1-a}\right)^{-1} \sum_{i=1}^n a^{i-1} \mu_Y = \mu_Y \\ \text{dado que } \sum_{i=1}^n a^{i-1} &= 1 - a^n \sum_{i=0}^{\infty} a^i = \frac{1-a^n}{1-a} \end{aligned}$$

La varianza de m_b es

$$\text{var}(m_b) = \left(\frac{1-a^n}{1-a}\right)^{-2} \sum_{i=1}^n a^{2(i-1)} \sigma_Y^2 = \sigma_Y^2 \frac{(1-a^{2n})(1-a)^2}{(1-a^2)(1-a^n)^2} = \sigma_Y^2 \frac{(1+a^n)(1-a)}{(1-a^n)(1+a)}$$

cuyo límite, $\text{var}(m_b) \xrightarrow{p} \sigma_Y^2(1 - a)/(1 + a)$, a medida que $n \rightarrow \infty$. Por tanto la varianza de este estimador no tiende a cero, la distribución no se concentra en torno a μ_Y , y el estimador, aunque es insesgado, no es consistente. Esto quizás resulta sorprendente, debido a que en este estimador están presentes todas las observaciones. Pero la mayoría de las observaciones reciben una ponderación muy pequeña (la ponderación de la i -ésima observación es proporcional a a^{i-1} , un número muy pequeño cuando i es grande), y por esta razón existe una cantidad insuficiente de cancelación de errores muestrales para que el estimador sea consistente.

El tercer estimador, m_c , es sesgado pero consistente. Su sesgo es $1/n: E(m_c) = E(\bar{Y} + 1/n) = \mu_Y + 1/n$. Sin embargo, el sesgo tiende a cero a medida que aumenta el tamaño de la muestra y m_c es consistente: $\Pr(|m_c - \mu_Y| \geq \delta) = \Pr(|\bar{Y} + 1/n - \mu_Y| \geq \delta)$. A continuación, a partir de la Ecuación (17.43) del Apéndice 17.2, una generalización de la desigualdad de Chebychev implica que para cualquier variable aleatoria W , $\Pr(|W| \geq \delta) \leq E(W^2)/\delta^2$ para cualquier constante positiva δ . Por tanto $\Pr(|\bar{Y} + 1/n - \mu_Y| \geq \delta) \leq E[(\bar{Y} + 1/n - \mu_Y)^2]/\delta^2$. Pero $E[(\bar{Y} + 1/n - \mu_Y)^2] = \text{var}(\bar{Y}) + 1/n^2 = \sigma_Y^2/n + 1/n^2 \rightarrow 0$ a medida que n aumenta de tamaño. Se deduce que $\Pr(|\bar{Y} + 1/n - \mu_Y| \geq \delta) \rightarrow 0$, y m_c es consistente. Este ejemplo ilustra el hecho general de que un estimador puede ser sesgado en muestras finitas pero, si su sesgo desaparece a medida que el tamaño de la muestra se hace grande, el estimador todavía puede ser consistente (Ejercicio 17.10).

El teorema central del límite y la convergencia en distribución

Si la distribución de una sucesión de variables aleatorias converge a un límite cuando $n \rightarrow \infty$, entonces se dice que la sucesión de variables aleatorias converge en distribución. El teorema central del límite dice que, bajo ciertas condiciones generales, la media muestral estandarizada converge en distribución a una variable aleatoria normal.

Convergencia en distribución. Sea $F_1, F_2, \dots, F_n \dots$ una sucesión de funciones de distribución acumuladas que corresponde a una sucesión de variables aleatorias, $S_1, S_2, \dots, S_n, \dots$. Por ejemplo, S_n puede ser la media muestral estandarizada, $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$. Entonces se dice que la sucesión de variables aleatorias S_n **converge en distribución** a S (y se expresa en forma de $S_n \xrightarrow{d} S$) si las funciones de distribución $\{F_n\}$ convergen a F , la distribución de S . Es decir,

$$S_n \xrightarrow{d} S \quad \text{si y solo si} \quad \lim_{n \rightarrow \infty} F_n(t) = F(t) \quad (17.7)$$

donde el límite se cumple en todos los instantes t en los cuales la distribución límite F es continua. La distribución F se denomina **distribución asintótica** de S_n .

Resulta útil comparar los conceptos de convergencia en probabilidad (\xrightarrow{p}) y convergencia en distribución (\xrightarrow{d}). Si $S_n \xrightarrow{p} \mu$, entonces S_n se acerca a μ con alta probabilidad a medida que n aumenta. En cambio, si $S_n \xrightarrow{d} S$, entonces la *distribución* de S_n se acerca a la *distribución* de S a medida que aumenta n .

El teorema central del límite. Reformulamos ahora el teorema central del límite utilizando el concepto de convergencia en distribución. El teorema central del límite del Concepto clave 2.7 establece que si Y_1, \dots, Y_n son i.i.d. y $0 < \sigma_Y^2 < \infty$, entonces la distribución asintótica de $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ es $N(0, 1)$. Debido a que $\sigma_{\bar{Y}} = \sigma_Y/\sqrt{n}$, $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}} = \sqrt{n}(\bar{Y} - \mu_Y)/\sigma_Y$. Por tanto, el teorema central del límite puede ser reformulado como $\sqrt{n}(\bar{Y} - \mu_Y) \xrightarrow{d} \sigma_Y Z$, donde Z es una variable aleatoria normal estándar. Esto significa que la distribución de $\sqrt{n}(\bar{Y} - \mu_Y)$ converge a $N(0, \sigma_Y^2)$ cuando $n \rightarrow \infty$. La expresión compacta habitual de este límite es

$$\sqrt{n}(\bar{Y} - \mu_Y) \xrightarrow{d} N(0, \sigma_Y^2) \quad (17.8)$$

Es decir, si Y_1, \dots, Y_n son i.i.d. y $0 < \sigma_Y^2 < \infty$, entonces la distribución de $\sqrt{n}(\bar{Y} - \mu_Y)$ converge a una distribución normal con media igual a cero y varianza igual a σ_Y^2 .

Extensiones para datos de series temporales. La ley de los grandes números y el teorema central del límite formulados en la Sección 2.6 son aplicables a observaciones i.i.d. Tal y como se analizó en el

Capítulo 14, el supuesto i.i.d. no resulta adecuado para datos de series temporales, y resulta necesario ampliar estos teoremas antes de que puedan aplicarse a observaciones de series temporales. Estas ampliaciones son de carácter técnico, en el sentido de que la conclusión es la misma —las versiones de la ley de los grandes números y el teorema central del límite son aplicables a los datos de series temporales—, pero las condiciones bajo las cuales son aplicables son distintas. Esto se analiza brevemente en la Sección 16.4, pero el tratamiento matemático de la teoría de distribución asintótica para variables de series temporales queda fuera del alcance de este libro y se recomienda a los lectores interesados que acudan a Hayashi (2000, Capítulo 2).

El teorema de Slutsky y el teorema de la función continua

El **teorema de Slutsky** combina la consistencia y la convergencia en distribución. Supongamos que $a_n \xrightarrow{p} a$, donde a es una constante, y $S_n \xrightarrow{d} S$. Entonces

$$a_n + S_n \xrightarrow{d} a + S, \quad a_n S_n \xrightarrow{d} aS \quad \text{y, si } a \neq 0, \quad S_n/a_n \xrightarrow{d} S/a \quad (17.9)$$

Estos tres resultados se denominan de forma conjunta teorema de Slutsky.

El **teorema de la función continua** se refiere a las propiedades asintóticas de una función continua, g , de una sucesión de variables aleatorias, S_n . El teorema tiene dos partes. La primera es que si S_n converge en probabilidad a una constante a , entonces $g(S_n)$ converge en probabilidad a $g(a)$; la segunda es que si S_n converge en distribución a S , entonces $g(S_n)$ converge en distribución a $g(S)$. Es decir, si g es una función continua, entonces

- (i) si $S_n \xrightarrow{p} a$, entonces $g(S_n) \xrightarrow{p} g(a)$, y
 - (ii) si $S_n \xrightarrow{d} S$, entonces $g(S_n) \xrightarrow{d} g(S)$
- (17.10)

Como ejemplo de (i), si $s_Y^2 \xrightarrow{p} \sigma_Y^2$, entonces $\sqrt{s_Y^2} = s_Y \xrightarrow{p} \sigma_Y$. Como ejemplo de (ii), supongamos que $S_n \xrightarrow{d} Z$, donde Z es una variable aleatoria normal estándar, y sea $g(S_n) = S_n^2$. Debido a que g es continua, el teorema de la función continua es aplicable y $g(S_n) \xrightarrow{d} g(Z)$; es decir, $S_n^2 \xrightarrow{d} Z^2$. En otras palabras, la distribución de S_n^2 converge a la distribución del cuadrado de una variable aleatoria normal estándar, que a su vez tiene una distribución χ_1^2 ; es decir, $S_n^2 \xrightarrow{d} \chi_1^2$.

Aplicación al estadístico t basado en la media muestral

Se utilizan ahora el teorema central del límite, la ley de los grandes números y el teorema de Slutsky para demostrar que, bajo la hipótesis nula, el estadístico t basado en \bar{Y} presenta una distribución normal estándar cuando Y_1, \dots, Y_n son i.i.d. y $0 < E(Y_i^4) < \infty$.

El estadístico t para contrastar la hipótesis nula de que $E(Y_i) = \mu_0$ basado en la media muestral \bar{Y} está dado por las Ecuaciones (3.8) y (3.11), y se puede escribir de la forma

$$t = \frac{\bar{Y} - \mu_0}{s_Y / \sqrt{n}} = \frac{\sqrt{n}(\bar{Y} - \mu_0)}{\sigma_Y} \div \frac{s_Y}{\sigma_Y}, \quad (17.11)$$

donde la segunda igualdad utiliza el truco de dividir tanto el numerador como el denominador por σ_Y .

Debido a que Y_1, \dots, Y_n tienen momentos de segundo orden (lo que está implícito en el hecho de que tengan momentos de cuarto orden; véase Ejercicio 17.5), y debido a que Y_1, \dots, Y_n son i.i.d., el primer término tras la última igualdad de la Ecuación (17.11) cumple el teorema central del límite: bajo la hipótesis nula, $\sqrt{n}(\bar{Y} - \mu_0)/\sigma_Y \xrightarrow{d} N(0, 1)$. Además, $s_Y^2 \xrightarrow{p} \sigma_Y^2$ (tal y como se demuestra en el Apéndice 3.3), por lo que $s_Y^2/\sigma_Y^2 \xrightarrow{p} 1$ y el cociente del segundo término de la Ecuación (17.11) tiende a 1 (Ejercicio 17.4). Por tanto la expresión tras la última igualdad de la Ecuación (17.11) tiene la forma de la última expresión de la Ecuación (17.9), donde [en notación de la Ecuación (17.9)] $S_n = \sqrt{n}(\bar{Y} - \mu_0)/\sigma_Y \xrightarrow{d} N(0, 1)$ y $a_n = s_Y/\sigma_Y \xrightarrow{p} 1$. Se deduce mediante la aplicación del teorema de Slutsky que $t \xrightarrow{d} N(0, 1)$.

17.3 Distribución asintótica del estimador MCO y del estadístico t

Recordemos del Capítulo 4, que, bajo los supuestos del Concepto clave 4.3 (los tres primeros supuestos del Concepto clave 17.1), el estimador MCO $\hat{\beta}_1$ es consistente y $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ tiene una distribución asintótica normal. Por otra parte, el estadístico t para contrastar la hipótesis nula $\beta_1 = \beta_{1,0}$ tiene una distribución asintótica normal estándar bajo la hipótesis nula. En esta sección se resumen estos resultados y se proporcionan detalles adicionales acerca de sus demostraciones.

Consistencia y normalidad asintótica de los estimadores MCO

La distribución para muestras grandes de $\hat{\beta}_1$, inicialmente formulada en el Concepto clave 4.4, es

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0, \frac{\text{var}(v_i)}{[\text{var}(X_i)]^2}\right) \quad (17.12)$$

donde $v_i = (X_i - \mu_X)u_i$. La prueba de este resultado fue esbozada en el Apéndice 4.3, pero en esa prueba se omitieron algunos detalles y se incluía una aproximación que no se demostró de manera formal. Los pasos que faltan en esta demostración se dejan como Ejercicio 17.3.

Una de las implicaciones de la Ecuación (17.12) es que $\hat{\beta}_1$ es consistente (Ejercicio 17.4).

Consistencia de los errores estándar heterocedástico-robustos

Bajo los tres primeros supuestos de mínimos cuadrados, los errores estándar heterocedástico-robustos para $\hat{\beta}_1$ constituyen la base para realizar inferencias estadísticas válidas. En concreto,

$$\frac{\hat{\sigma}_{\hat{\beta}_1}^2}{\sigma_{\hat{\beta}_1}^2} \xrightarrow{p} 1 \quad (17.13)$$

donde $\hat{\sigma}_{\hat{\beta}_1}^2 = \text{var}(v_i)/\{n[\text{var}(X_i)]^2\}$ y $\hat{\sigma}_{\hat{\beta}_1}^2$ es el cuadrado del error estándar heterocedástico-robusto definido en la Ecuación (5.4); es decir,

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} \quad (17.14)$$

Para demostrar el resultado de la Ecuación (17.13), se utilizan en primer lugar las definiciones de $\hat{\sigma}_{\hat{\beta}_1}^2$ y $\sigma_{\hat{\beta}_1}^2$ para reescribir el cociente de la Ecuación (17.13) como

$$\frac{\hat{\sigma}_{\hat{\beta}_1}^2}{\sigma_{\hat{\beta}_1}^2} = \left[\frac{n}{n-2} \right] \left[\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\text{var}(v_i)} \right] \div \left[\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}{\text{var}(X_i)} \right]^2 \quad (17.15)$$

Es necesario demostrar que cada uno de los tres términos entre corchetes que se encuentran en el lado derecho de la Ecuación (17.15) converge en probabilidad a 1. Es evidente que el primer término converge a 1, y por la propiedad de consistencia de la varianza muestral (Apéndice 3.3) el último término converge en probabilidad a 1. Por lo tanto, lo único que queda es demostrar que el segundo término converge en probabilidad a 1, es decir, que $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2 \xrightarrow{p} \text{var}(v_i)$.

La demostración de que $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2 \xrightarrow{p} \text{var}(v_i)$ se desarrolla en dos pasos. El primero muestra que $\frac{1}{n} \sum_{i=1}^n v_i^2 \xrightarrow{p} \text{var}(v_i)$; el segundo muestra que $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2 - \frac{1}{n} \sum_{i=1}^n v_i^2 \xrightarrow{p} 0$.

Por el momento, supongamos que X_i y u_i tienen momentos de orden ocho [es decir $E(X_i^8) < \infty$ y $E(u_i^8) < \infty$], lo cual es un supuesto más fuerte que el de los momentos de cuarto orden requeridos por el tercer supuesto de mínimos cuadrados. Para demostrar el primer paso, es necesario demostrar que $\frac{1}{n} \sum_{i=1}^n v_i^2$ cumple la ley de los grandes números de la Ecuación (17.5). Para hacerlo, v_i^2 deben ser i.i.d. (lo son por el segundo supuesto de mínimos cuadrados) y $\text{var}(v_i^2)$ debe ser finita. Para demostrar que $\text{var}(v_i^2) < \infty$, se aplica la desigualdad de Cauchy-Schwarz (Apéndice 17.2): $\text{var}(v_i^2) \leq E(v_i^4) = E[(X_i - \mu_X)^4 u_i^4] \leq \{E[(X_i - \mu_X)^8]E(u_i^8)\}^{1/2}$. Por lo tanto, si X_i y u_i tienen momentos de orden ocho, entonces v_i^2 tiene una varianza finita y por lo tanto satisface la ley de los grandes números de la Ecuación (17.5).

El segundo paso consiste en demostrar que $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2 - \frac{1}{n} \sum_{i=1}^n v_i^2 \xrightarrow{P} 0$. Debido a que $v_i = (X_i - \mu_X)u_i$, este segundo paso es lo mismo que demostrar que

$$\frac{1}{n} \sum_{i=1}^n [(X_i - \bar{X})^2 \hat{u}_i^2 - (X_i - \mu_X)^2 u_i^2] \xrightarrow{P} 0 \quad (17.16)$$

La demostración de este resultado implica establecer $\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)X_i$, expandiendo el término de la Ecuación (17.16) entre corchetes, aplicando repetidamente la desigualdad de Cauchy-Schwarz, y utilizando la consistencia de $\hat{\beta}_0$ y $\hat{\beta}_1$. Los detalles de álgebra se dejan como Ejercicio 17.9.

El argumento anterior supone que X_i y u_i tienen momentos de orden ocho. No obstante, esto no es necesario, y el resultado $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2 \xrightarrow{P} \text{var}(v_i)$ se puede demostrar bajo el supuesto más débil de que X_i y u_i tienen momentos de cuarto orden, tal y como se estableció en el tercer supuesto de mínimos cuadrados. Sin embargo, esta demostración queda fuera del alcance de este libro de texto; véase Hayashi (2000, Sección 2.5) para más detalles.

Normalidad asintótica del estadístico t heterocedástico-robusto

Ahora vamos a demostrar que, bajo la hipótesis nula, el estadístico t MCO heterocedástico-robusto para contrastar la hipótesis de que $\beta_1 = \beta_{1,0}$ tiene una distribución asintótica normal estándar si se cumplen los supuestos de mínimos cuadrados #1, #2 y #3.

El estadístico t construido utilizando el error estándar heterocedástico-robusto $ES(\hat{\beta}_1) = \hat{\sigma}_{\hat{\beta}_1}$ [definido en la Ecuación (17.14)] es

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\sqrt{n}(\hat{\beta}_1 - \beta_{1,0})}{\sqrt{n\hat{\sigma}_{\hat{\beta}_1}^2}} \div \sqrt{\frac{\hat{\sigma}_{\hat{\beta}_1}^2}{\sigma_{\hat{\beta}_1}^2}} \quad (17.17)$$

Se deduce de la Ecuación (17.12) que el primer término tras la segunda igualdad de la Ecuación (17.17) converge en distribución a una variable aleatoria normal estándar. Además, debido a que el error estándar heterocedástico-robusto es consistente [Ecuación (17.13)], $\sqrt{\hat{\sigma}_{\hat{\beta}_1}^2/\sigma_{\hat{\beta}_1}^2} \xrightarrow{P} 1$ (Ejercicio 17.4). Se deduce del teorema de Slutsky que $t \xrightarrow{d} N(0, 1)$.

17.4 Distribuciones muestrales exactas con errores normalmente distribuidos

En muestras pequeñas, la distribución del estimador MCO y del estadístico t depende de la distribución del error de regresión y por lo general es complicada. Sin embargo, tal y como se trató en la Sección 5.6, si los errores de regresión son homocedásticos y normalmente distribuidos, estas distribuciones son sencillas. En concreto, si los cinco supuestos ampliados de mínimos cuadrados del Concepto clave 17.1 se cumplen, entonces el estimador MCO tiene una distribución muestral normal, condicionada a X_1, \dots, X_n . Además, el estadístico t tiene una distribución t de Student. Estos resultados se presentan aquí para $\hat{\beta}_1$.

Distribución de $\hat{\beta}_1$ con errores normales

Si los errores son i.i.d. con distribución normal e independientes de los regresores, entonces la distribución de $\hat{\beta}_1$ condicionada a X_1, \dots, X_n , es $N(\beta_1, \sigma_{\hat{\beta}_1|X}^2)$, donde

$$\sigma_{\hat{\beta}_1|X}^2 = \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (17.18)$$

La deducción de la distribución normal $N(\beta_1, \sigma_{\hat{\beta}_1|X}^2)$, condicionada a X_1, \dots, X_n , implica (i) establecer que la distribución es normal; (ii) demostrar que $E(\hat{\beta}_1|X_1, \dots, X_n) = \beta_1$; y (iii) verificar la Ecuación (17.18).

Para demostrar (i), debe tenerse en cuenta que condicionado a X_1, \dots, X_n , $\hat{\beta}_1 - \beta_1$ es una media ponderada de u_1, \dots, u_n :

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (17.19)$$

Esta ecuación se dedujo en el Apéndice 4.3 [Ecuación (4.30) y se vuelve a mostrar aquí para mayor comodidad]. Por los supuestos ampliados de mínimos cuadrados #1, #2, #4, y #5, u_i es $N(0, \sigma_u^2)$ i.i.d., y u_i y X_i están independientemente distribuidas. Debido a que las medias ponderadas de las variables normalmente distribuidas se distribuyen asimismo normalmente, se deduce que $\hat{\beta}_1$ se distribuye normalmente, condicionado a X_1, \dots, X_n .

Para demostrar (ii), se toman las esperanzas condicionadas de ambos lados de la Ecuación (17.19): $E[(\hat{\beta}_1 - \beta_1)|X_1, \dots, X_n] = E[\sum_{i=1}^n (X_i - \bar{X})u_i / \sum_{i=1}^n (X_i - \bar{X})^2 | X_1, \dots, X_n] = \sum_{i=1}^n (X_i - \bar{X})E(u_i|X_1, \dots, X_n) / \sum_{i=1}^n (X_i - \bar{X})^2 = 0$, donde la última igualdad se obtiene debido a que $E(u_i|X_1, \dots, X_n) = E(u_i|X_i) = 0$. Por tanto $\hat{\beta}_1$ es condicionalmente insesgado; es decir,

$$E(\hat{\beta}_1|X_1, \dots, X_n) = \beta_1 \quad (17.20)$$

Para demostrar (iii), se utiliza el hecho de que los errores se distribuyen de forma independiente, condicionado a X_1, \dots, X_n para calcular la varianza condicional de $\hat{\beta}_1$ utilizando la Ecuación (17.19):

$$\begin{aligned} \text{var}(\hat{\beta}_1|X_1, \dots, X_n) &= \text{var}\left[\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} | X_1, \dots, X_n\right] \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \text{var}(u_i|X_1, \dots, X_n)}{\left[\sum_{i=1}^n (X_i - \bar{X})^2\right]^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_u^2}{\left[\sum_{i=1}^n (X_i - \bar{X})^2\right]^2} \end{aligned} \quad (17.21)$$

Cancelando el término del numerador en la última expresión de la Ecuación (17.21) se obtiene la fórmula de la varianza condicional de la Ecuación (17.18).

Distribución del estadístico t válido con homocedasticidad

El estadístico t válido con homocedasticidad para contrastar la hipótesis nula de que $\beta_1 = \beta_{1,0}$ es

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{ES(\hat{\beta}_1)} \quad (17.22)$$

donde $ES(\hat{\beta}_1)$ se calcula utilizando el error estándar válido con homocedasticidad de $\hat{\beta}_1$. Sustituyendo la fórmula de $ES(\hat{\beta}_1)$ [Ecuación (5.29) del Apéndice 5.1] en la Ecuación (17.22) y reordenando se obtiene

$$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{s_u^2 / \sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\sigma_u^2 / \sum_{i=1}^n (X_i - \bar{X})^2}} \div \sqrt{\frac{s_u^2}{\sigma_u^2}} \\ &= \frac{(\hat{\beta}_1 - \beta_{1,0})/\sigma_{\hat{\beta}_{1|X}}}{\sqrt{W/(n-2)}} \end{aligned} \quad (17.23)$$

donde $s_u^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2$ y $W = \sum_{i=1}^n \hat{u}_i^2 / \sigma_u^2$. Bajo la hipótesis nula, $\hat{\beta}_1$ tiene una distribución $N(\beta_{1,0}, \sigma_{\hat{\beta}_{1|X}}^2)$ condicionada a X_1, \dots, X_n por lo que la distribución del numerador de la última expresión de la Ecuación (17.23) es $N(0,1)$. Se demuestra en la Sección 18.4 que W tiene una distribución chi-cuadrado con $n-2$ grados de libertad y además que W se distribuye de forma independiente del estimador MCO estandarizado del numerador de la Ecuación (17.23). Se desprende de la definición de la distribución t de Student (Apéndice 17.1) que, bajo los cinco supuestos ampliados de mínimos cuadrados, el estadístico t válido con homocedasticidad tiene una distribución t de Student con $n-2$ grados de libertad.

¿Dónde encaja el ajuste por los grados de libertad? El ajuste por los grados de libertad en s_u^2 asegura que s_u^2 es un estimador insesgado de σ_u^2 y que el estadístico t tiene una distribución t de Student cuando los errores se distribuyen normalmente.

Debido a que $W = \sum_{i=1}^n \hat{u}_i^2 / \sigma_u^2$ es una variable aleatoria chi-cuadrado con $n-2$ grados de libertad, su media es $E(W) = n-2$. Por tanto $E[W/(n-2)] = (n-2)/(n-2) = 1$. Reorganizando la definición de W , se obtiene que $E\left(\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2\right) = \sigma_u^2$. Por tanto, la corrección por los grados de libertad hace que s_u^2 sea un estimador insesgado de σ_u^2 . Además, al dividir por $n-2$ en lugar de por n , el término del denominador de la última expresión de la Ecuación (17.23) coincide con la definición de una variable aleatoria con distribución t de Student que se ofrece en el Apéndice 17.1. Es decir, utilizando el ajuste por los grados de libertad para calcular el error estándar, el estadístico t tiene la distribución t de Student cuando los errores se distribuyen normalmente.

17.5 Mínimos cuadrados ponderados

Bajo los cuatro primeros supuestos ampliados de mínimos cuadrados, el estimador MCO es eficiente entre la clase de los estimadores lineales (en Y_1, \dots, Y_n), y condicionalmente (a X_1, \dots, X_n) insesgados; es decir, el estimador MCO es ELIO. Este resultado es el teorema de Gauss-Markov, que fue estudiado en la Sección 5.5 y demostrado en el Apéndice 5.2. El teorema de Gauss-Markov proporciona una justificación teórica a la utilización del estimador MCO. Una limitación importante del teorema de Gauss-Markov es que necesita errores homocedásticos. Si, como a menudo sucede en la práctica, los errores son heterocedásticos, el teorema de Gauss-Markov no se cumple y el estimador MCO no es ELIO.

En esta sección se presenta una modificación del estimador MCO, denominada **mínimos cuadrados ponderados (MCP)**, que es más eficiente que el MCO cuando los errores son heterocedásticos.

MCP requiere conocer un poco más acerca de la función de la varianza condicional, $\text{var}(u_i|X_i)$. Se consideran dos casos. En el primer caso, $\text{var}(u_i|X_i)$ es conocida con un factor de proporcionalidad, y MCP es ELIO. En el segundo caso, la forma funcional de $\text{var}(u_i|X_i)$ es conocida, pero esta forma funcional contiene

algunos parámetros desconocidos que deben ser estimados. Bajo ciertas condiciones adicionales, la distribución asintótica de MCP en el segundo caso es la misma que si los parámetros de la función de la varianza condicional fueran en realidad conocidos, y en este sentido el estimador MCP es asintóticamente ELIO. La Sección concluye con un análisis de las ventajas e inconvenientes prácticos del tratamiento de la heterocedasticidad mediante MCP o, alternativamente, utilizando los errores estándar heterocedástico-robustos.

MCP con heterocedasticidad conocida

Supongamos que la varianza condicional $\text{var}(u_i|X_i)$ es conocida con un factor de proporcionalidad; es decir,

$$\text{var}(u_i|X_i) = \lambda h(X_i) \quad (17.24)$$

donde λ es una constante y h es una función conocida. En este caso, el estimador MCP es el estimador obtenido dividiendo en primer lugar la variable dependiente y el regresor por la raíz cuadrada de h y posteriormente realizando una regresión MCO de esta variable dependiente transformada sobre el regresor transformado. En concreto, se dividen ambos lados del modelo de una sola variable independiente por $\sqrt{h(X_i)}$ para obtener

$$\tilde{Y}_i = \beta_0 \tilde{X}_{0i} + \beta_1 \tilde{X}_{1i} + \tilde{u}_i, \quad (17.25)$$

donde $\tilde{Y}_i = Y_i/\sqrt{h(X_i)}$, $\tilde{X}_{0i} = 1/\sqrt{h(X_i)}$, $\tilde{X}_{1i} = X_i/\sqrt{h(X_i)}$, y $\tilde{u}_i = u_i/\sqrt{h(X_i)}$.

El **estimador MCP** es el estimador MCO de β_1 en la Ecuación (17.25); es decir, es el estimador obtenido mediante la regresión MCO de \tilde{Y}_i sobre \tilde{X}_{0i} y \tilde{X}_{1i} , donde el coeficiente de \tilde{X}_{0i} toma el lugar del término independiente o intercepto de la regresión sin ponderar.

Bajo los tres primeros supuestos de mínimos cuadrados del Concepto clave 17.1 más el supuesto de heterocedasticidad conocida de la Ecuación (17.24), MCP es ELIO. La razón de que el estimador MCP sea ELIO es que la ponderación de las variables hace que el término de error \tilde{u}_i de la regresión ponderada sea homocedástico. Es decir,

$$\text{var}(\tilde{u}_i|X_i) = \text{var}\left[\frac{u_i}{\sqrt{h(X_i)}}|X_i\right] = \frac{\text{var}(u_i|X_i)}{h(X_i)} = \frac{\lambda h(X_i)}{h(X_i)} = \lambda \quad (17.26)$$

por lo que la varianza condicional de \tilde{u}_i , $\text{var}(\tilde{u}_i|X_i)$, es constante. Por lo tanto los cuatro primeros supuestos de mínimos cuadrados son aplicables a la Ecuación (17.25). En rigor, el teorema de Gauss-Markov fue probado en el Apéndice 5.2 para la Ecuación (17.1), que incluye el término independiente β_0 , por lo que no es aplicable a la Ecuación (17.25), en la que se sustituye el término independiente por $\beta_0 \tilde{X}_{0i}$. Sin embargo, la generalización del teorema de Gauss-Markov para regresión múltiple (Sección 18.5) es aplicable a la estimación de β_1 en la regresión poblacional ponderada, Ecuación (17.25). En consecuencia, el estimador MCO de β_1 en la Ecuación (17.25)—es decir, el estimador MCP de β_1 —es ELIO.

En la práctica, la función h por lo general es desconocida, por lo que ni las variables ponderadas de la Ecuación (17.25), ni el estimador MCP se pueden calcular. Por esta razón, el estimador MCP que se describe aquí a veces se denomina estimador **MCP infactible**. Para llevar a cabo MCP en la práctica, la función h debe ser estimada, cuestión a la que nos referiremos ahora.

MCP con heterocedasticidad de forma funcional conocida

Si la heterocedasticidad tiene una forma funcional conocida, entonces la función de heterocedasticidad h puede ser estimada y el estimador MCP puede calcularse utilizando la función estimada.

Ejemplo #1: la varianza de u es cuadrática en X . Supongamos que la varianza condicional es conocida y que toma la forma de la función cuadrática

$$\text{var}(u_i|X_i) = \theta_0 + \theta_1 X_i^2 \quad (17.27)$$

donde θ_0 y θ_1 son parámetros desconocidos, $\theta_0 > 0$, y $\theta_1 \geq 0$.

Debido a que θ_0 y θ_1 son desconocidos, no es posible construir las variables ponderadas \tilde{Y}_i , \tilde{X}_{0i} , y \tilde{X}_{1i} . Sin embargo, sí es posible estimar θ_0 y θ_1 , y utilizar entonces estas estimaciones para calcular la estimación de $\text{var}(u_i|X_i)$. Sean $\hat{\theta}_0$ y $\hat{\theta}_1$ los estimadores de θ_0 y θ_1 , y sea $\widehat{\text{var}}(u_i|X_i) = \hat{\theta}_0 + \hat{\theta}_1 X_i^2$. Se definen las variables explicativas ponderadas $\tilde{Y}_i = Y_i / \sqrt{\widehat{\text{var}}(u_i|X_i)}$, $\tilde{X}_{0i} = 1 / \sqrt{\widehat{\text{var}}(u_i|X_i)}$, y $\tilde{X}_{1i} = X_{1i} / \sqrt{\widehat{\text{var}}(u_i|X_i)}$. El estimador MCP es el estimador MCO de los coeficientes de la regresión de \tilde{Y}_i sobre \tilde{X}_{0i} y \tilde{X}_{1i} (donde $\beta_0 \tilde{X}_{0i}$ ocupa el lugar del término independiente β_0).

La aplicación de este estimador requiere la estimación de la función de la varianza condicional, es decir, la estimación de θ_0 y θ_1 en la Ecuación (17.27). Una forma de estimar θ_0 y θ_1 de modo consistente es regresar \hat{u}_i^2 sobre X_i^2 mediante MCO, donde \hat{u}_i^2 es el cuadrado del i -ésimo residuo MCO.

Supongamos que la varianza condicional tiene la forma de la Ecuación (17.27) y que $\hat{\theta}_0$ y $\hat{\theta}_1$ son estimadores consistentes de θ_0 y θ_1 . Bajo los supuestos del #1 al #3 del Concepto clave 17.1, además de algunas condiciones adicionales sobre los momentos que aparecen debido a que θ_0 y θ_1 son estimados, la distribución asintótica del estimador MCP es la misma que en el caso de que θ_0 y θ_1 fueran conocidos. Por tanto el estimador MCP con θ_0 y θ_1 estimados tiene la misma distribución asintótica que el estimador MCP infactible y es en este sentido asintóticamente ELIO.

Debido a que este método de MCP se puede llevar a cabo mediante la estimación de los parámetros desconocidos de la función de la varianza condicional, este método a veces se denomina **MCP factibles** o **MCP estimados**.

Ejemplo #2: la varianza depende de una tercera variable. Asimismo MCP puede utilizarse cuando la varianza condicional depende de una tercera variable, W_i , que no aparece en la función de regresión. En concreto, supongamos que se recogen datos acerca de tres variables, Y_i , X_i , y W_i , $i = 1, \dots, n$; la función de regresión poblacional depende de X_i , pero no de W_i ; y la varianza condicional depende de W_i pero no de X_i . Es decir, la función de regresión poblacional es $E(Y_i|X_i, W_i) = \beta_0 + \beta_1 X_i$ y la varianza condicional es $\text{var}(u_i|X_i, W_i) = \lambda h(W_i)$, donde λ es una constante y h es una función que debe estimarse.

Por ejemplo, supongamos que un investigador está interesado en modelizar la relación entre la tasa de desempleo en un estado y una variable de política económica estatal (X_i). No obstante, la tasa de desempleo medida (Y_i) es una estimación de la tasa de desempleo real (Y_i^*) basada en una encuesta. Por tanto, Y_i mide Y_i^* con un error, de tal forma que el origen del error es el error de aleatoriedad de la encuesta, por lo que $Y_i = Y_i^* + v_i$, donde v_i es el error de medición que surge de la encuesta. En este ejemplo, es factible que el tamaño muestral de la encuesta, W_i , no sea en sí mismo un factor determinante de la verdadera tasa de desempleo estatal. Por tanto la función de regresión poblacional no depende de W_i ; es decir, $E(Y_i^*|X_i, W_i) = \beta_0 + \beta_1 X_i$. Por consiguiente, tenemos las dos ecuaciones

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i^* \quad (17.28)$$

$$Y_i = Y_i^* + v_i \quad (17.29)$$

donde la Ecuación (17.28) modeliza la relación entre la variable de política económica del estado y la verdadera tasa de desempleo y la Ecuación (17.29) representa la relación entre la tasa de desempleo medida Y_i y la verdadera tasa de desempleo Y_i^* .

El modelo de las Ecuaciones (17.28) y (17.29) puede conducir a una regresión poblacional en la que la varianza condicional del error dependa de W_i , pero no de X_i . El término de error u_i^* de la Ecuación (17.28) representa otros factores omitidos de esta regresión, mientras que el término de error v_i de la Ecuación (17.29) representa el error de medición que procede de la encuesta de la tasa de desempleo. Si u_i^* es homocedástico, entonces $\text{var}(u_i^*|X_i, W_i) = \sigma_{u^*}^2$ es constante. La varianza del error de la encuesta, no obstante, depende inversamente del tamaño muestral de la encuesta W_i ; es decir, $\text{var}(v_i|X_i, W_i) = a/W_i$, donde a es una constante. Debido a que v_i es el error aleatorio de la encuesta, se supone con seguridad que no está correlacionado con u_i^* , por lo que $\text{var}(u_i^* + v_i|X_i, W_i) = \sigma_{u^*}^2 + a/W_i$. Por tanto, sustituyendo la Ecuación (17.28) en la Ecuación (17.29) se obtiene un modelo de regresión con heterocedasticidad

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (17.30)$$

$$\text{var}(u_i|X_i, W_i) = \theta_0 + \theta_1 \left(\frac{1}{W_i} \right) \quad (17.31)$$

donde $u_i = u_i^* + v_i$, $\theta_0 = \sigma_{u^*}^2$, $\theta_1 = a$, y $E(u_i|X_i, W_i) = 0$.

Si ocurriese que θ_0 y θ_1 fueran conocidos, la función de la varianza condicional de la Ecuación (17.31) podría utilizarse para estimar β_0 y β_1 mediante MCP. En este ejemplo, θ_0 y θ_1 son desconocidos, pero pueden ser estimados mediante la regresión de los residuos MCO al cuadrado [procedentes de la estimación MCO de la Ecuación (17.30)] sobre $1/W_i$. A continuación, la función de varianza condicional estimada puede utilizarse para construir las ponderaciones en MCP factibles.

Cabe destacar que es muy importante que $E(u_i|X_i, W_i) = 0$; si no, los errores ponderados no tendrían una media condicional igual a cero y MCP será inconsistente. Dicho de otra manera, si W_i es en realidad un factor determinante de Y_i , entonces la Ecuación (17.30) debería ser una ecuación de regresión múltiple que incluyera tanto a X_i como a W_i .

Método general de MCP factibles. En general, el método de los MCP factibles consta de cinco pasos:

1. Regresión de Y_i sobre X_i mediante MCO y obtención de los residuos MCO, \hat{u}_i , $i = 1, \dots, n$.
2. Estimación de un modelo para la función de la varianza condicional $\text{var}(u_i|X_i)$. Por ejemplo, si la función de la varianza condicional tiene la forma de la Ecuación (17.27), esto implica la regresión de \hat{u}_i^2 sobre X_i^2 . En general, este paso implica la estimación de una función para la varianza condicional $\text{var}(u_i|X_i)$.
3. Utilización de la función estimada para calcular los valores esperados de la función de la varianza condicional, $\text{var}(u_i|X_i)$.
4. Ponderación de la variable dependiente y el regresor (incluido el término independiente) por la inversa de la raíz cuadrada de la función de la varianza condicional estimada.
5. Estimación de los coeficientes de la regresión ponderada mediante MCO; los estimadores resultantes son los estimadores MCP.

Los paquetes de *software* de regresión suelen incluir comandos opcionales para la estimación de mínimos cuadrados ponderados que automatizan los pasos cuarto y quinto.

¿Errores estándar heterocedástico-robustos o MCP?

Existen dos maneras de actuar en presencia de heterocedasticidad: estimar β_0 y β_1 mediante MCP o estimar β_0 y β_1 mediante MCO y utilizar los errores estándar heterocedástico-robustos. La decisión acerca de qué método utilizar en la práctica requiere sopesar las ventajas y desventajas de cada uno de ellos.

La ventaja de MCP consiste en que es más eficiente que el estimador MCO de los coeficientes de los regresores originales, al menos asintóticamente. La desventaja de MCP consiste en que es necesario conocer la función de la varianza condicional y estimar sus parámetros. Si la función de la varianza condicional tiene la forma cuadrática de la Ecuación (17.27), esto se hace fácilmente. En la práctica, no obstante, la forma funcional de la función de la varianza condicional rara vez es conocida. Por otra parte, si la forma funcional es incorrecta, entonces los errores estándar calculados mediante rutinas de regresión MCP no son válidos en el sentido de que conducen a inferencias estadísticas incorrectas (los contrastes presentan el tamaño incorrecto).

La ventaja de utilizar errores estándar heterocedástico-robustos es que dan lugar a inferencias asintóticamente válidas incluso si no se conoce la forma de la función de la varianza condicional. Una ventaja adicional es que los errores estándar heterocedástico-robustos se calculan fácilmente como una opción dentro de los paquetes informáticos modernos de regresión, por lo que no es necesario ningún esfuerzo adicional para protegerse frente a esa amenaza. La desventaja de los errores estándar heterocedástico-robustos consiste en que el estimador MCO tendrá un mayor varianza que el estimador MCP (basado en la verdadera función de la varianza condicional), al menos asintóticamente.

En la práctica, la forma funcional de $\text{var}(u_i|X_i)$ casi nunca es conocida, lo cual plantea un problema para el uso de MCP en las aplicaciones del mundo real. Este problema resulta bastante complicado con un único regresor, pero en los casos con regresores múltiples resulta aún más difícil conocer la forma funcional de la varianza condicional. Por esta razón, la utilización de MCP en la práctica se enfrenta a retos importantes. Por el contrario, en los paquetes estadísticos modernos resulta sencillo utilizar errores estándar heterocedástico-robustos, y las inferencias resultantes son fiables bajo condiciones muy generales; en particular, los errores estándar heterocedástico-robustos se pueden utilizar sin necesidad de especificar una forma funcional para la varianza condicional. Por estas razones, nuestra opinión es que, a pesar del atractivo teórico de los MCP, los errores estándar heterocedástico-robustos proporcionan una mejor manera de manejar la heterocedasticidad potencial en la mayoría de las aplicaciones.

Resumen

1. La normalidad asintótica del estimador MCO, junto con la consistencia de los errores estándar heterocedástico-robustos, implica que, si los tres primeros supuestos de mínimos cuadrados del Concepto clave 17.1 se cumplen, entonces el estadístico t heterocedástico-robusto tiene una distribución asintótica normal estándar bajo la hipótesis nula.
2. Si los errores de la regresión son i.i.d. con distribución normal, condicionada a los regresores, entonces $\hat{\beta}_1$ tiene una distribución muestral exacta normal, condicionada a los regresores. Además el estadístico t válido con homocedasticidad tiene una distribución muestral exacta t_{n-2} de Student bajo la hipótesis nula.
3. El estimador de mínimos cuadrados ponderados (MCP) es el estimador MCO aplicado a una regresión ponderada, en la que todas las variables están ponderadas por la raíz cuadrada de la inversa de la varianza condicional, $\text{var}(u_i|X_i)$, o de su estimación. Aunque el estimador MCP es asintóticamente más eficiente que el MCO, para poder ser llevado a cabo MCP en la práctica, la forma funcional de la función de la varianza condicional debe ser conocida, lo cual por lo general resulta una tarea difícil.

Términos clave

convergencia en probabilidad (485)	mínimos cuadrados ponderados (MCP) (492)
estimador consistente (486)	estimador MCP (493)
convergencia en distribución (487)	MCP infactible (493)
distribución asintótica (487)	MCP factible (494)
teorema de Slutsky (488)	f.d.p. normal (499)
teorema de la función continua (488)	f.d.p. normal bivariante (500)

Revisión de conceptos

- 17.1** Supóngase que el Supuesto #4 del Concepto clave 17.1 es cierto, pero se construye un intervalo de confianza al 95 % para β_1 utilizando el error estándar heterocedástico-robusto en una muestra grande. ¿Podría este intervalo de confianza ser asintóticamente válido en el sentido de que contenga el verdadero valor de β_1 en el 95 % de todas las muestras repetidas para n grande? Supóngase por el contrario que el Supuesto #4 del Concepto clave 17.1 no se cumple, pero que se construye un intervalo de confianza al 95 % para β_1 mediante la fórmula del error estándar válido con homocedasticidad para una muestra grande. ¿Sería este intervalo de confianza asintóticamente válido?
- 17.2** Supóngase que A_n es una variable aleatoria que converge en probabilidad a 3. Supóngase que B_n es una variable aleatoria que converge en distribución a una normal estándar. ¿Cuál es la distribución asintótica de $A_n B_n$? Utilice esta distribución asintótica para calcular un valor aproximado de $\Pr(A_n B_n < 2)$.
- 17.3** Supóngase que Y y X están relacionadas por la regresión $Y = 1,0 + 2,0X + u$. Un investigador dispone de observaciones sobre X e Y , donde $0 \leq X \leq 20$, donde la varianza condicional es $\text{var}(u_i|X_i = x) = 1$ para $0 \leq x \leq 10$ y $\text{var}(u_i|X_i = x) = 16$ para $10 \leq x \leq 20$. Represente gráfica-

mente un diagrama de dispersión hipotético de las observaciones (X_i, Y_i) , $i = 1, \dots, n$. ¿Pondera MCP más las observaciones con $x \leq 10$ o aquellas con $x > 10$? ¿Por qué?

- 17.4** En lugar de utilizar MCP, el investigador del problema anterior decide calcular el estimador MCO utilizando solamente las observaciones para las que $x \leq 10$, a continuación, utilizando solamente las observaciones para las cuales $x > 10$, y más tarde obtiene la media de los dos estimadores MCO. ¿Es este estimador más eficiente que MCP?

Ejercicios

- 17.1** Considérese el modelo de regresión sin término independiente, $Y_i = \beta_1 X_i + u_i$ (por lo que el verdadero valor del intercepto β_0 , es cero).
- Obtenga el estimador de mínimos cuadrados de β_1 para el modelo de regresión restringida $Y_i = \beta_1 X_i + u_i$. Esto se conoce como estimador de mínimos cuadrados restringidos ($\hat{\beta}_1^{MCR}$) de β_1 , ya que se estima bajo una restricción, que en este caso es $\beta_0 = 0$.
 - Obtenga la distribución asintótica de $\hat{\beta}_1^{MCR}$ bajo los Supuestos #1 a #3 del Concepto clave 17.1.
 - Demuestre que $\hat{\beta}_1^{MCR}$ es lineal [Ecuación (5.24)] y, bajo los Supuestos #1 y #2 del Concepto clave 17.1, condicionalmente insesgado [Ecuación (5.25)].
 - Obtenga la varianza condicional de $\hat{\beta}_1^{MCR}$ bajo las condiciones de Gauss-Markov (Supuestos #1 a #4 del Concepto clave 17.1).
 - Compare la varianza condicional de $\hat{\beta}_1^{MCR}$ obtenida en el apartado (d) con la varianza condicional del estimador MCO $\hat{\beta}_1$ (de la regresión que incluye un término independiente) bajo las condiciones de Gauss-Markov. ¿Qué estimador es más eficiente? Utilice las fórmulas de las varianzas para explicar por qué.
 - Obtenga la distribución muestral exacta de $\hat{\beta}_1^{MCR}$ bajo los Supuestos #1 a #5 del Concepto clave 17.1.
 - Considérese ahora el estimador $\tilde{\beta}_1 = \sum_{i=1}^n Y_i / \sum_{i=1}^n X_i$. Obtenga una expresión para $\text{var}(\tilde{\beta}_1 | X_1, \dots, X_n) - \text{var}(\hat{\beta}_1^{MCR} | X_1, \dots, X_n)$ bajo las condiciones de Gauss-Markov y utilice esta expresión para demostrar que $\text{var}(\tilde{\beta}_1 | X_1, \dots, X_n) \geq \text{var}(\hat{\beta}_1^{MCR} | X_1, \dots, X_n)$.
- 17.2** Supóngase que (X_i, Y_i) son i.i.d. con momentos de cuarto orden finitos. Demuestre que la covarianza muestral es un estimador consistente de la covarianza poblacional; es decir $s_{XY} \xrightarrow{P} \sigma_{XY}$, donde s_{XY} está definida en la Ecuación (3.24). (Sugerencia: utilice la estrategia del Apéndice 3.3 y la desigualdad de Cauchy-Schwarz).
- 17.3** Este ejercicio completa los detalles de la obtención de la distribución asintótica de $\hat{\beta}_1$ dada en el Apéndice 4.3.
- Utilice la Ecuación (17.19) para obtener la expresión
- $$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n v_i}}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} - \frac{(\bar{X} - \mu_X) \sqrt{\frac{1}{n} \sum_{i=1}^n u_i}}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$
- donde $v_i = (X_i - \mu_X)u_i$.
- Utilice el teorema central del límite, la ley de grandes números, y el teorema de Slutsky para demostrar que el último término de la ecuación converge en probabilidad a cero.
 - Utilice la desigualdad de Cauchy-Schwarz y el tercer supuesto de mínimos cuadrados del Concepto clave 17.1 para demostrar que $\text{var}(v_i) < \infty$. ¿Satisface el término $\sqrt{\frac{1}{n} \sum_{i=1}^n v_i} / \sigma_v$ el teorema central del límite?
 - Aplique el teorema central del límite y el teorema de Slutsky para obtener la resultado de la Ecuación (17.12).

17.4 Demuestre los siguientes resultados:

- a) Demuestre que $\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, a^2)$, donde a^2 es una constante, implica $\hat{\beta}_1$ que es consistente. [Sugerencia: utilice el teorema de Slutsky].
- b) Demuestre que $s_u^2/\sigma_u^2 \xrightarrow{P} 1$ implica que $s_u/\sigma_u \xrightarrow{P} 1$.

17.5 Supóngase que W es una variable aleatoria con $E(W^4) < \infty$. Demuestre que $E(W^2) < \infty$.

17.6 Demuestre que si $\hat{\beta}_1$ es condicionalmente insesgado, entonces es insesgado; es decir, demuestre que si $E(\hat{\beta}_1|X_1, \dots, X_n) = \beta_1$, entonces $E(\hat{\beta}_1) = \beta_1$.

17.7 Supóngase que X y u son variables aleatorias continuas y (X_i, u_i) , $i = 1, \dots, n$, son i.i.d.

- a) Demuestre que la función de densidad de probabilidad conjunta (f.d.p.) de (u_i, u_j, X_i, X_j) se puede escribir como $f(u_i, X_i)f(u_j, X_j)$ para $i \neq j$, donde $f(u_i, X_i)$ es la f.d.p. conjunta de u_i y X_i .
- b) Demuestre que $E(u_i u_j | X_i X_j) = E(u_i | X_i)E(u_j | X_j)$ para $i \neq j$.
- c) Demuestre que $E(u_i | X_1, \dots, X_n) = E(u_i | X_i)$.
- d) Demuestre que $E(u_i u_j | X_1, X_2, \dots, X_n) = E(u_i | X_i)E(u_j | X_j)$ para $i \neq j$.

17.8 Considérese el modelo de regresión del Concepto clave 17.1 y supóngase que los Supuestos #1, #2, #3, y #5 se cumplen. Supóngase que el Supuesto #4 se reemplaza por el supuesto de que $\text{var}(u_i | X_i) = \theta_0 + \theta_1 |X_i|$, donde $|X_i|$ es el valor absoluto de X_i , y $\theta_0 > 0$ y $\theta_1 \geq 0$.

- a) ¿Es ELIO el estimador MCO de β_1 ?
- b) Supóngase que θ_0 y θ_1 son conocidos. ¿Cuál es el estimador ELIO de β_1 ?
- c) Obtenga la distribución muestral exacta del estimador MCO, $\hat{\beta}_1$, condicionada a X_1, \dots, X_n .
- d) Obtenga la distribución muestral exacta del estimador MCP (considerando conocidos θ_0 y θ_1) de β_1 , condicionada a X_1, \dots, X_n .

17.9 Demuestre la Ecuación (17.16), bajo los Supuestos #1 y #2 del Concepto clave 17.1 además del supuesto de que X_i y u_i tienen momentos de orden ocho.

17.10 Sea $\hat{\theta}$ un estimador del parámetro θ , donde $\hat{\theta}$ podría ser sesgado. Demuestre que si $E[(\hat{\theta} - \theta)^2] \rightarrow 0$ cuando $n \rightarrow \infty$ (es decir, el error cuadrático medio de $\hat{\theta}$ tiende a cero), entonces $\hat{\theta} \xrightarrow{P} \theta$. [Sugerencia: utilice la Ecuación (17.43) con $W = \hat{\theta} - \theta$].

17.11 Supóngase que X e Y tienen una distribución normal bivariante cuya densidad está dada en la Ecuación (17.38).

- a) Demuestre que la densidad de Y dado $X = x$ se puede escribir como

$$f_{Y|X=x}(y) = \frac{1}{\sigma_{Y|X}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y - \mu_{Y|X}}{\sigma_{Y|X}}\right)^2\right]$$

donde $\sigma_{Y|X} = \sqrt{\sigma_Y^2(1 - \rho_{XY}^2)}$ y $\mu_{Y|X} = \mu_Y - (\sigma_{XY}/\sigma_X)(x - \mu_X)$. [Sugerencia: utilice la definición de la densidad de probabilidad condicional $f_{Y|X=x}(y) = [g_{X,Y}(x, y)]/[f_X(x)]$, donde $g_{X,Y}$ es la densidad conjunta de X e Y , y f_X es la densidad marginal de X].

- b) Utilice el resultado del apartado (a) para demostrar que $Y|X = x \sim N(\mu_{Y|X}, \sigma_{Y|X}^2)$.
- c) Utilice el resultado del apartado (b) para demostrar que $E(Y|X = x) = a + bx$ para las constantes a y b escogidas de la forma adecuada.

17.12 a) Supóngase que $u \sim N(0, \sigma_u^2)$. Demuestre que $E(e^u) = e^{\frac{1}{2}\sigma_u^2}$.

- b) Supóngase que la distribución condicional de u dado $X = x$ es $N(0, a + bx^2)$, donde a y b son constantes positivas. Demuestre que $E(e^u|X = x) = e^{\frac{1}{2}(a + bx^2)}$.

17.13 Considérese el modelo de regresión heterogéneo $Y_i = \beta_{0i} + \beta_{1i}X_i + u_i$ donde β_{0i} y β_{1i} son variables aleatorias que varían de una observación a otra. Supóngase que $E(u_i | X_i) = 0$ y que (β_{0i}, β_{1i}) se distribuyen independientemente de X_i .

- a) Sea $\hat{\beta}_1^{MCO}$ el estimador MCO de β_1 dado en la Ecuación (17.2). Demuestre que $\hat{\beta}_1^{MCO} \xrightarrow{P} E(\beta_1)$, donde $E(\beta_1)$ es el valor medio de β_{1i} en la población. [Sugerencia: véase la Ecuación (13.10)].

- b) Supóngase que $\text{var}(u_i|X_i) = \theta_0 + \theta_1 X_i^2$, donde θ_0 y θ_1 son constantes positivas conocidas. Sea $\hat{\beta}_1^{MCP}$ el estimador de mínimos cuadrados ponderados. ¿Ocurre que $\hat{\beta}_1^{MCP} \xrightarrow{P} E(\beta_1)$? Explíquelo.

APÉNDICE

17.1 La distribución normal y sus afines y los momentos de las variables aleatorias continuas

En este apéndice se definen y analizan la distribución normal y sus afines. Las definiciones de las distribuciones chi-cuadrado, F , y t de Student, dadas en la Sección 2.4, se redefinen aquí como referencia. Comenzaremos por presentar las definiciones de las probabilidades y de los momentos que se refieren a variables aleatorias continuas.

Probabilidades y momentos de variables aleatorias continuas

Tal y como se analizó en la Sección 2.1, si Y es una variable aleatoria continua, su probabilidad queda caracterizada por su función de densidad de probabilidad (f.d.p.). La probabilidad de que Y se encuentre entre dos valores es el área bajo su f.d.p. entre estos dos valores. No obstante, debido a que Y es continua, las expresiones matemáticas de las probabilidades incluyen integrales en lugar de los sumatorios que resultan apropiados para variables aleatorias discretas.

Sea f_Y la función de densidad de probabilidad de Y . Debido a que las probabilidades no pueden ser negativas, $f_Y(y) \geq 0$, para todo y . La probabilidad de que Y se encuentre entre a y b (donde $a < b$) es

$$\Pr(a \leq Y \leq b) = \int_a^b f_Y(y) dy \quad (17.32)$$

Debido a que Y debe tomar valores en la recta real, $\Pr(-\infty \leq Y \leq \infty) = 1$ lo que implica que $\int_{-\infty}^{\infty} f_Y(y) dy = 1$.

Las esperanzas y los momentos de las variables aleatorias continuas, como los de las variables aleatorias discretas, son las medias de sus valores ponderados por probabilidad, excepto que los sumatorios [por ejemplo, el sumatorio de la Ecuación (2.3)] se sustituyen por integrales. Por lo tanto, el valor esperado de Y es

$$E(Y) = \mu_Y = \int y f_Y(y) dy \quad (17.33)$$

donde el rango de integración es el conjunto de valores para los cuales f_Y es distinta de cero. La varianza es la esperanza de $(Y - \mu_Y)^2$, y el momento r -ésimo de una variable aleatoria es la esperanza de Y^r . Por tanto

$$\text{var}(Y) = E(Y - \mu_Y)^2 = \int (y - \mu_Y)^2 f_Y(y) dy, \quad (17.34)$$

$$E(Y^r) = \int y^r f_Y(y) dy \quad (17.35)$$

La distribución normal

La distribución normal de una sola variable. La función densidad de probabilidad de una variable aleatoria normalmente distribuida (la **f.d.p. normal**) es

$$f_Y(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{y - \mu}{\sigma} \right)^2 \right] \quad (17.36)$$

donde $\exp(x)$ es la función exponencial de x . El factor $1/\sigma \sqrt{2\pi}$ de la Ecuación (17.36) asegura que $\Pr(-\infty \leq Y \leq \infty) = \int_{-\infty}^{\infty} f_Y(y) dy = 1$.

La media de la distribución normal es μ y su varianza es σ^2 . La distribución normal es simétrica, por lo que todos los momentos centrales impares de orden tres y superior son iguales a cero. El momento central cuarto es $3\sigma^4$. En general, si Y se distribuye $N(\mu, \sigma^2)$, sus momentos centrales pares están dados por

$$E[(Y - \mu)^k] = \frac{k!}{2^{k/2}(k/2)!} \sigma^k \quad (k \text{ par}) \quad (17.37)$$

Si $\mu = 0$ y $\sigma^2 = 1$, la distribución normal se denomina distribución normal estándar. La f.d.p. normal estándar se expresa mediante ϕ y la f.d.a. se expresa por Φ . Por tanto la densidad normal estándar es $\phi(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})$ y $\Phi(y) = \int_{-\infty}^y \phi(s) ds$.

La distribución normal bivariante. La **f.d.p. normal bivariante** para las dos variables aleatorias X e Y es

$$g_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho_{XY}^2}} \times \exp\left\{\frac{1}{-2(1 - \rho_{XY}^2)} \left[\left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2 \right]\right\} \quad (17.38)$$

donde ρ_{XY} es la correlación entre X e Y .

Cuando X e Y están incorrelacionadas ($\rho_{XY} = 0$), $g_{XY}(x, y) = f_X(x)f_Y(y)$, donde f es la densidad normal dada por la Ecuación (17.36). Esto demuestra que si X e Y están normalmente distribuidas de forma conjunta y están incorrelacionadas, entonces están independientemente distribuidas. Esta es una característica particular de la distribución normal que no suele ser cierta para otras distribuciones.

La distribución normal multivariante amplía la distribución normal bivariante para más de dos variables aleatorias. Esta distribución se formula de forma más adecuada utilizando matrices y se presenta en el Apéndice 18.1.

La distribución normal condicional. Supongamos que X e Y están normalmente distribuidas de forma conjunta. Entonces la distribución condicional de Y dado X es $N(\mu_{Y|X}, \sigma_{Y|X}^2)$, con una media $\mu_{Y|X} = \mu_Y + (\sigma_{XY}/\sigma_X^2)(X - \mu_X)$ y una varianza $\sigma_{Y|X}^2 = (1 - \rho_{XY}^2)\sigma_Y^2$. La media de esta distribución condicional, condicionada a $X = x$, es una función lineal de x , y la varianza no depende de x (Ejercicio 17.11).

Distribuciones afines

La distribución chi-cuadrado. Sean Z_1, Z_2, \dots, Z_n n variables aleatorias normales estándar i.i.d. La variable aleatoria

$$W = \sum_{i=1}^n Z_i^2 \quad (17.39)$$

tiene una distribución chi-cuadrado con n grados de libertad. Esta distribución se expresa mediante χ_n^2 . Debido a que $E(Z_i^2) = 1$, y $E(Z_i^4) = 3$, $E(W) = n$ y $\text{var}(W) = 2n$.

La distribución t de Student. Sea Z con distribución normal estándar, sea W con distribución χ_m^2 , y sean Z y W variables distribuidas de forma independiente. Entonces la variable aleatoria

$$t = \frac{Z}{\sqrt{W/m}} \quad (17.40)$$

tiene una distribución t de Student con m grados de libertad, y se expresa como t_m . La distribución t_∞ es la distribución normal estándar.

La distribución F . Sean W_1 y W_2 variables aleatorias independientes con distribuciones chi-cuadrado con grados de libertad respectivos n_1 y n_2 . Entonces la variable aleatoria

$$F = \frac{W_1/n_1}{W_2/n_2} \quad (17.41)$$

tiene una distribución F con (n_1, n_2) grados de libertad. La distribución se expresa mediante F_{n_1, n_2} .

La distribución F depende de los grados de libertad del numerador n_1 y de los grados de libertad del denominador n_2 . A medida que el número de grados de libertad del denominador aumenta, la distribución F_{n_1, n_2} se aproxima correctamente mediante una distribución $\chi^2_{n_1}$, dividida por n_1 . En el límite, la distribución $F_{n_1, \infty}$ es la misma que la distribución $\chi^2_{n_1}$, dividida por n_1 ; es decir, es la misma que la distribución $\chi^2_{n_1}/n_1$.

APÉNDICE

17.2 Dos desigualdades

En este anexo se enuncian y demuestran la desigualdad de Chebychev y la desigualdad de Cauchy-Schwarz.

La desigualdad de Chebychev

La desigualdad de Chebychev utiliza la varianza de la variable aleatoria V con el fin de acotar la probabilidad de que V se encuentre a más distancia que $\pm \delta$ respecto de su media, donde δ es una constante positiva:

$$\Pr(|V - \mu_V| \geq \delta) \leq \frac{\text{var}(V)}{\delta^2} \quad (\text{desigualdad de Chebychev}) \quad (17.42)$$

Para demostrar la Ecuación (17.42), sea $W = V - \mu_V$, sea f la f.d.p. de W , y sea δ cualquier número positivo. Ahora

$$\begin{aligned} E(W^2) &= \int_{-\infty}^{\infty} w^2 f(w) dw \\ &= \int_{-\infty}^{-\delta} w^2 f(w) dw + \int_{-\delta}^{\delta} w^2 f(w) dw + \int_{\delta}^{\infty} w^2 f(w) dw \\ &\geq \int_{-\infty}^{-\delta} w^2 f(w) dw + \int_{\delta}^{\infty} w^2 f(w) dw \\ &\geq \delta^2 \left[\int_{-\infty}^{-\delta} f(w) dw + \int_{\delta}^{\infty} f(w) dw \right] \\ &= \delta^2 \Pr(|W| \geq \delta), \end{aligned} \quad (17.43)$$

donde la primera igualdad es la definición de $E(W^2)$, la segunda igualdad se cumple debido a que los rangos de integración fraccionan la recta real, la primera desigualdad se cumple debido a que el término que no es tenido en cuenta es no negativo, la segunda desigualdad se cumple debido a que $w^2 \geq \delta^2$ a lo largo de todo el rango de integración y la última igualdad se cumple por la definición de $\Pr(|W| \geq \delta)$. Sustituyendo $W = V - \mu_V$ en la última expresión, teniendo en cuenta que $E(W^2) = E[(V - \mu_V)^2] = \text{var}(V)$, y reordenando se obtiene la desigualdad que aparece en la Ecuación (17.42). Si V es discreta, esta prueba es aplicable con sumatorios en sustitución de las integrales.

La desigualdad de Cauchy-Schwarz

La desigualdad de Cauchy-Schwarz es una generalización de la desigualdad de la correlación $|\rho_{XY}| \leq 1$, para incorporar medias distintas de cero. La desigualdad de Cauchy-Schwarz es

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)} \quad (\text{desigualdad de Cauchy-Schwarz}) \quad (17.44)$$

La demostración de la Ecuación (17.44) es similar a la prueba de la desigualdad de la correlación del Apéndice 2.1. Sea $W = Y + bX$, donde b es una constante. Entonces $E(W^2) = E(Y^2) + 2bE(XY) + b^2E(X^2)$. Ahora sea $b = -E(XY)/E(X^2)$ por lo que (después de la simplificación) la expresión se convierte en $E(W^2) = E(Y^2) - [E(XY)]^2/E(X^2)$. Debido a que $E(W^2) = 0$ (debido a que $W^2 \geq 0$), debe ocurrir que $[E(XY)]^2 = E(X^2)E(Y^2)$, y la desigualdad de Cauchy-Schwarz se deduce tomando la raíz cuadrada.

Teoría de regresión múltiple

Este capítulo proporciona una introducción a la teoría del análisis de regresión múltiple. El capítulo persigue cuatro objetivos. El primero consiste en presentar el modelo de regresión múltiple en forma matricial, lo cual conlleva la obtención de fórmulas compactas para el estimador MCO y los estadísticos de contraste. El segundo objetivo es el de caracterizar la distribución muestral del estimador MCO, tanto en muestras grandes (mediante teoría asintótica) como en muestras pequeñas (si los errores son homocedásticos y están normalmente distribuidos). El tercer objetivo es el de estudiar la teoría de estimación eficiente de los coeficientes en regresión múltiple y describir los mínimos cuadrados generalizados (MCG), un método para estimar los coeficientes de regresión de forma eficiente cuando los errores son heterocedásticos y/o están correlacionados entre distintas observaciones. El cuarto objetivo es el de llevar a cabo un tratamiento breve de la teoría de distribución asintótica de la regresión de variables instrumentales (VI) en el modelo lineal, incluyendo una introducción a la estimación por el método generalizado de momentos (MGM) del modelo de regresión lineal VI con errores heterocedásticos.

El capítulo comienza definiendo el modelo de regresión múltiple y el estimador MCO en forma matricial en la Sección 18.1. Esta Sección presenta asimismo los supuestos ampliados de mínimos cuadrados para el modelo de regresión múltiple. Los cuatro primeros supuestos son los mismos que los supuestos de mínimos cuadrados del Concepto clave 6.4 y son los que subyacen tras las distribuciones asintóticas utilizadas para justificar los procedimientos utilizados en los Capítulos 6 y 7. Los dos supuestos ampliados restantes de mínimos cuadrados son más fuertes y nos permitirán explorar en más detalle las propiedades teóricas del estimador MCO en el modelo de regresión múltiple.

Las tres secciones siguientes examinan la distribución muestral del estimador MCO y los estadísticos de contraste. La Sección 18.2 presenta las distribuciones asintóticas del estimador MCO y el estadístico t bajo los supuestos de mínimos cuadrados del Concepto clave 6.4. La Sección 18.3 unifica y generaliza los contrastes de hipótesis sobre varios coeficientes presentados en las Secciones 7.2 y 7.3, y proporciona la distribución asintótica de los estadísticos F resultantes. En la Sección 18.4, se examinan las distribuciones muestrales exactas del estimador MCO y de los estadísticos de contraste en el caso particular en el que los errores son homocedásticos y están distribuidos normalmente. Aunque el supuesto de homocedasticidad y de errores normales resulta poco verosímil en la mayoría de las aplicaciones econométricas, las distribuciones muestrales exactas tienen un interés teórico, y los p -valores calculados mediante estas distribuciones aparecen frecuentemente en las salidas proporcionadas por el software de regresión.

Las dos siguientes secciones se centran en la teoría de la estimación eficiente de los coeficientes del modelo de regresión múltiple. La Sección 18.5 generaliza el teorema de Gauss-Markov para regresión múltiple. La Sección 18.6 desarrolla el método de mínimos cuadrados generalizados (MCG).

La última sección aborda la estimación VI en el modelo de regresión VI general cuando los instrumentos son válidos y son fuertes. En esta sección se deduce la distribución asintótica del estimador MC2E cuando los errores son heterocedásticos y se proporcionan las expresiones para el error estándar del estimador MC2E. El estimador MC2E es uno de los muchos estimadores MGM posibles, y en esta sección se proporciona una introducción a la estimación MGM en el modelo de regresión lineal VI. Se demuestra que el estimador MC2E es el estimador MGM eficiente si los errores son homocedásticos.

Requisitos matemáticos previos. Para el tratamiento del modelo lineal en este capítulo se utiliza la notación matricial y las herramientas de álgebra lineal básica, se supone además que el lector ha cursado una asignatura previa de introducción al álgebra lineal. En el Apéndice 18.1 se repasan los vectores, las matrices, y las operaciones matriciales que se utilizan en este capítulo. Además, se utiliza el cálculo multivariante en la Sección 18.1 para obtener el estimador MCO.

18.1 El modelo lineal de regresión múltiple y el estimador MCO en forma matricial

Tanto el modelo lineal de regresión múltiple como el estimador MCO se pueden representar de manera compacta mediante notación matricial.

El modelo de regresión múltiple en forma matricial

El modelo de regresión múltiple poblacional (Concepto clave 6.2) es

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n \quad (18.1)$$

Con el fin de expresar el modelo de regresión múltiple en forma matricial, se definen los vectores y matrices siguientes:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \\ \vdots \\ \mathbf{X}'_n \end{pmatrix}, \quad \text{y } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad (18.2)$$

por lo tanto \mathbf{Y} es $n \times 1$, \mathbf{X} es $n \times (k+1)$, \mathbf{U} es $n \times 1$, y $\boldsymbol{\beta}$ es $(k+1) \times 1$. A lo largo del capítulo se expresan las matrices y vectores en negrita. En esta notación,

- \mathbf{Y} es el vector de dimensión $n \times 1$ de las n observaciones de la variable dependiente.
- \mathbf{X} es la matriz de dimensión $n \times (k+1)$ de las n observaciones sobre los $k+1$ regresores (incluyendo el regresor «constante» para el intercepto).
- El vector columna X_i de dimensión $(k+1) \times 1$ es la observación i -ésima de los $k+1$ regresores; es decir, $X'_i = (1, X_{1i}, \dots, X_{ki})$, donde X'_i es la traspuesta de X_i .
- \mathbf{U} es el vector de dimensión $n \times 1$ de los n términos de error.
- $\boldsymbol{\beta}$ es el vector de dimensión $(k+1) \times 1$ de los $k+1$ coeficientes de regresión desconocidos.

El modelo de regresión múltiple de la Ecuación (18.1) para la observación i -ésima, expresado mediante los vectores $\boldsymbol{\beta}$ y \mathbf{X}_i , es

$$Y_i = \mathbf{X}'_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n \quad (18.3)$$

En la Ecuación (18.3), el primer regresor es el regresor «constante» que siempre es igual a 1, y su coeficiente es el intercepto. Por tanto el intercepto no aparece de forma separada en la Ecuación (18.3); sino que es el primer elemento del vector de coeficientes $\boldsymbol{\beta}$.

Recopilando el total de las n observaciones de la Ecuación (18.3) se obtiene el modelo de regresión múltiple en forma matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \quad (18.5)$$

Los supuestos ampliados de mínimos cuadrados

Los supuestos ampliados de mínimos cuadrados para el modelo con varios regresores son los cuatro supuestos del modelo de regresión múltiple del Concepto clave 6.4, además de los dos supuestos adiciona-

**CONCEPTO
CLAVE**
18.1
Los supuestos ampliados de mínimos cuadrados para el modelo de regresión múltiple

El modelo de regresión lineal con varias variables explicativas es

$$Y_i = X'_i \beta + u_i, \quad i = 1, \dots, n. \quad (18.4)$$

Los supuestos ampliados de mínimos cuadrados son

1. $E(u_i|X_i) = 0$ (u_i tiene media condicional igual a cero).
2. $(X_i, Y_i), i = 1, \dots, n$ son extracciones independientes e idénticamente distribuidas (i.i.d.) a partir de su distribución conjunta.
3. X_i , y u_i tienen momentos de cuarto orden finitos y distintos de cero.
4. X tiene rango completo de columnas (no existe multicolinealidad perfecta).
5. $\text{var}(u_i|X_i) = \sigma_u^2$ (homocedasticidad).
6. La distribución condicional de u_i dado X_i es normal (errores normales).

les de errores homocedásticos y normalmente distribuidos. El supuesto de homocedasticidad se utiliza cuando se estudia la eficiencia del estimador MCO, y el supuesto de normalidad se utiliza cuando se estudia la distribución muestral exacta del estimador MCO y de los estadísticos de contraste.

Los supuestos ampliados de mínimos cuadrados se recogen en el Concepto clave 18.1.

Salvo por las diferencias en la notación, los tres primeros supuestos del Concepto clave 18.1 son idénticos a los tres primeros supuestos del Concepto clave 6.4.

El cuarto supuesto de los Conceptos clave 6.4 y 18.1 puede parecer diferente, pero en realidad es el mismo: simplemente son diferentes formas de expresar que no puede existir multicolinealidad perfecta. Recordemos que la multicolinealidad perfecta surge cuando un regresor se puede escribir como combinación lineal perfecta del resto de los regresores. En la notación matricial de la Ecuación (18.2), la multicolinealidad perfecta significa que una columna de X es una combinación lineal perfecta del resto de las otras columnas de X , pero si esto es cierto, entonces X no tiene rango de columnas completo. Por tanto decir que X tiene rango $k + 1$, es decir, tiene un rango igual al número de columnas de X , es tan solo otra manera de decir que los regresores no son perfectamente multicolineales.

El quinto supuesto de mínimos cuadrados del Concepto clave 18.1 es que el término de error es condicionalmente homocedástico, y el sexto supuesto es que la distribución condicional de u_i , dado X_i , es normal. Estos dos supuestos son iguales a los dos últimos supuestos del Concepto clave 17.1, excepto que ahora se enuncian para regresores múltiples.

Implicaciones para el vector de medias y la matriz de covarianzas de U . Los supuestos de mínimos cuadrados del Concepto clave 18.1 incluyen expresiones sencillas para el vector de medias y la matriz de covarianzas de la distribución condicional de U dada la matriz de regresores X . (El vector de medias y la matriz de covarianzas de un vector de variables aleatorias se definen en el Apéndice 18.2). En concreto, el primer y segundo supuestos del Concepto clave 18.1 implican que $E(u_i|X) = E(u_i|X_i) = 0$ y que $\text{cov}(u_i, u_j|X) = E(u_i u_j|X_i X_j) = E(u_i|X_i)E(u_j|X_j) = 0$ para $i \neq j$ (Ejercicio 17.7). Los supuestos primero, segundo, y quinto implican que $E(u_i^2|X) = E(u_i^2|X_i) = \sigma_u^2$. Combinando estos resultados, se obtiene que

$$\text{bajo los supuestos } \#1 \text{ y } \#2, E(U|X) = \mathbf{0}_n, \text{ y} \quad (18.6)$$

$$\text{bajo los supuestos } \#1, \#2, \text{ y } \#5, E(UU'|X) = \sigma_u^2 \mathbf{I}_n \quad (18.7)$$

donde $\mathbf{0}_n$ es el vector n -dimensional de ceros e \mathbf{I}_n es la matriz identidad $n \times n$.

Del mismo modo, los supuestos primero, segundo, quinto, y sexto del Concepto clave 18.1 implican que la distribución condicional del vector aleatorio n -dimensional, \mathbf{U} condicionada a \mathbf{X} , es la distribución normal multivariante (que se define en el Apéndice 18.2). Es decir,

$$\begin{aligned} &\text{bajo los Supuestos \#1, \#2, \#5, y \#6, la} \\ &\text{distribución condicional de } \mathbf{U} \text{ dada } \mathbf{X} \text{ es } N(\mathbf{0}_n, \sigma_u^2 \mathbf{I}_n) \end{aligned} \quad (18.8)$$

El estimador MCO

El estimador MCO minimiza la suma de los errores de predicción al cuadrado, $\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki})^2$ [Ecuación (6.8)]. La fórmula del estimador MCO se obtiene tomando la derivada de la suma de los errores de predicción al cuadrado respecto de cada elemento del vector de coeficientes, igualando estas derivadas a cero y resolviendo para el estimador $\hat{\beta}$.

La derivada de la suma de los errores de predicción al cuadrado con respecto al j -ésimo coeficiente de regresión b_j , es

$$\begin{aligned} &\frac{\partial}{\partial b_j} \sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki})^2 = \\ &-2 \sum_{i=1}^n X_{ji} (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki}) \end{aligned} \quad (18.9)$$

para $j = 0, \dots, k$, donde, para $j = 0$, $X_{0i} = 1$ para todo i . La derivada de la parte derecha de la Ecuación (18.9) es el elemento j -ésimo del vector $k + 1$ dimensional, $-2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\mathbf{b})$, donde \mathbf{b} es el vector $k + 1$ dimensional que consta de b_0, \dots, b_k . Existen $k + 1$ derivadas de este tipo, cada una de ellas correspondiente a un elemento de \mathbf{b} . Combinadas, dan lugar al sistema de $k + 1$ ecuaciones que, cuando se igualan a cero, constituyen las condiciones de primer orden del estimador MCO $\hat{\beta}$. Es decir, $\hat{\beta}$ resuelve el sistema de $k + 1$ ecuaciones

$$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \mathbf{0}_{k+1} \quad (18.10)$$

o, equivalentemente, $\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\hat{\beta}$.

Resolviendo el sistema de ecuaciones (18.10) se obtiene el estimador MCO $\hat{\beta}$ en forma matricial:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (18.11)$$

donde $(\mathbf{X}'\mathbf{X})^{-1}$ es la inversa de la matriz $\mathbf{X}'\mathbf{X}$.

El papel de la «multicolinealidad no perfecta». El cuarto supuesto de mínimos cuadrados del Concepto clave 18.1 establece que \mathbf{X} tiene rango completo de columnas. A su vez, esto implica que la matriz $\mathbf{X}'\mathbf{X}$ tiene rango completo, es decir, que $\mathbf{X}'\mathbf{X}$ es no singular. Debido a que $\mathbf{X}'\mathbf{X}$ es no singular, es invertible. Por tanto, el supuesto de que no existe multicolinealidad perfecta asegura que $(\mathbf{X}'\mathbf{X})^{-1}$ existe, por lo que la Ecuación (18.10) tiene una solución única y la fórmula de la Ecuación (18.11) para el estimador MCO puede, en realidad, calcularse. Dicho de otro modo, si \mathbf{X} no tiene rango completo de columnas, no existe una única solución para la Ecuación (18.10) y $\mathbf{X}'\mathbf{X}$ es singular. Por lo tanto, $(\mathbf{X}'\mathbf{X})^{-1}$ no puede calcularse y por lo tanto, no se puede calcular el $\hat{\beta}$ de la Ecuación (18.11).

18.2 Distribución asintótica del estimador MCO y del estadístico t

Si el tamaño de la muestra es grande y se cumplen los cuatro primeros supuestos del Concepto clave 18.1, entonces el estimador MCO tiene una distribución conjunta asintóticamente normal, el estimador heterocedástico-robusto de la matriz de covarianzas es consistente, y el estadístico t MCO heterocedástico-robusto presenta una distribución asintótica normal estándar. Estos resultados utilizan la distribución normal multivariante (Apéndice 18.2) y una extensión multivariante del teorema central del límite.

El teorema central del límite multivariante

El teorema central del límite del Concepto clave 2.7 es aplicable a una variable aleatoria unidimensional. Para obtener la distribución asintótica conjunta de los elementos de $\hat{\beta}$, es necesario un teorema central del límite multivariante que sea aplicable a vectores de variables aleatorias.

El teorema central del límite multivariante extiende el teorema central del límite univariante a las medias de las observaciones de un vector de variables aleatorias, \mathbf{W} , donde \mathbf{W} es m -dimensional. La diferencia entre el teorema central del límite para un escalar frente al de una variable aleatoria vectorial son las condiciones sobre las varianzas. Para el caso escalar del Concepto clave 2.7, el requisito es que la varianza sea a la vez finita y distinta de cero. En el caso vectorial, el requisito es que la matriz de covarianzas sea a la vez finita y definida positiva. Si la variable aleatoria vectorial \mathbf{W} tiene una matriz de covarianzas finita y definida positiva, entonces $0 < \text{var}(\mathbf{c}'\mathbf{W}) < \infty$ para todos los vectores \mathbf{c} m -dimensionales distintos de cero (Ejercicio 18.3).

El teorema central del límite multivariante que vamos a utilizar está recogido en el Concepto clave 18.2.

**CONCEPTO
CLAVE**

18.2

El teorema central del límite multivariante

Supongamos que $\mathbf{W}_1, \dots, \mathbf{W}_n$ son variables aleatorias i.i.d. m -dimensionales con vector de medias $E(\mathbf{W}_i) = \boldsymbol{\mu}_{\mathbf{W}}$ y matriz de covarianzas $E[(\mathbf{W}_i - \boldsymbol{\mu}_{\mathbf{W}})(\mathbf{W}_i - \boldsymbol{\mu}_{\mathbf{W}})'] = \Sigma_{\mathbf{W}}$, donde $\Sigma_{\mathbf{W}}$ es finita y definida positiva. Sea $\bar{\mathbf{W}} = \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i$. Entonces $\sqrt{n}(\bar{\mathbf{W}} - \boldsymbol{\mu}_{\mathbf{W}}) \xrightarrow{d} N(\mathbf{0}_m, \Sigma_{\mathbf{W}})$.

Normalidad asintótica de $\hat{\beta}$

En muestras grandes, el estimador MCO presenta la distribución asintótica normal multivariante

$$\sqrt{n}(\hat{\beta} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}_{k+1}, \Sigma_{\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})}) \text{ donde } \Sigma_{\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})} = \mathbf{Q}_X^{-1} \Sigma_V \mathbf{Q}_X^{-1} \quad (18.12)$$

donde \mathbf{Q}_X es la matriz de dimensión $(k+1) \times (k+1)$ de los momentos de segundo orden de los regresores, es decir, $\mathbf{Q}_X = E(\mathbf{X}_i \mathbf{X}_i')$, y Σ_V es la matriz de covarianzas de dimensión $(k+1) \times (k+1)$ de $\mathbf{V}_i = \mathbf{X}_i u_i$, es decir, $\Sigma_V = E(\mathbf{V}_i \mathbf{V}_i')$. Téngase en cuenta que el segundo supuesto de mínimos cuadrados del Concepto clave 18.1 implica que las \mathbf{V}_i , $i = 1, \dots, n$, son i.i.d.

Expresado en términos de $\hat{\beta}$ en lugar de $\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})$, la aproximación normal de la Ecuación (18.12) es

$\hat{\beta}$ está distribuido, en muestras grandes, $N(\boldsymbol{\beta}, \Sigma_{\hat{\beta}})$

$$\text{donde } \Sigma_{\hat{\beta}} = \Sigma_{\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})}/n = \mathbf{Q}_X^{-1} \Sigma_V \mathbf{Q}_X^{-1}/n \quad (18.13)$$

La matriz de covarianzas $\Sigma_{\hat{\beta}}$ en la Ecuación (18.13) es la matriz de covarianzas de la distribución normal aproximada de $\hat{\beta}$, mientras que $\Sigma_{\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})}/n$ en la Ecuación (18.12) es la matriz de covarianzas de la distribución asintótica normal de $\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})$. Estas dos matrices de covarianzas se diferencian por un factor n , dependiendo de si el estimador MCO está escalado por \sqrt{n} .

Obtención de la Ecuación (18.12). Para deducir la Ecuación (18.12), se utilizan en primer lugar las Ecuaciones (18.4) y (18.11) para escribir $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{U})$, de manera que

$$\hat{\beta} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} \quad (18.14)$$

Por lo tanto $\hat{\beta} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}$, por lo que

$$\sqrt{n}(\hat{\beta} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \left(\frac{\mathbf{X}'\mathbf{U}}{\sqrt{n}}\right) \quad (18.15)$$

La obtención de la Ecuación (18.12) implica, en primer lugar, argumentar que la matriz «denominador» de la Ecuación (18.15), $\mathbf{X}'\mathbf{X}/n$ es consistente y, en segundo lugar, que la matriz «numerador», cumple el teorema central del límite multivariante del Concepto clave 18.2. Los detalles figuran en el Apéndice 18.3.

Errores estándar heterocedástico-robustos

El estimador heterocedástico-robusto de $\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$ se obtiene mediante la sustitución, en su definición, [Ecuación (18.12)] de los momentos poblacionales por los momentos muestrales. En consecuencia, el estimador heterocedástico-robusto de la matriz de covarianzas de $\sqrt{n}(\hat{\beta}-\beta)$ es

$$\Sigma_{\sqrt{n}(\hat{\beta}-\beta)} = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \hat{\Sigma}_{\hat{V}} \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}, \text{ donde } \hat{\Sigma}_{\hat{V}} = \frac{1}{n-k-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \hat{u}_i^2 \quad (18.16)$$

El estimador $\hat{\Sigma}_{\hat{V}}$ incorpora el mismo ajuste por los grados de libertad que aparece en el ESR para el modelo de regresión múltiple (Sección 6.4) para corregir el riesgo potencial de sesgo a la baja debido a la estimación de $k+1$ coeficientes de regresión.

La demostración de que $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)} \xrightarrow{P} \Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$ es conceptualmente similar a la demostración, presentada en la Sección 17.3, de la consistencia de los errores estándar heterocedástico-robustos en el modelo de un único regresor.

Errores estándar heterocedástico-robustos. El estimador heterocedástico-robusto de la matriz de covarianzas de $\hat{\beta}$, $\hat{\Sigma}_{\hat{\beta}}$ es

$$\hat{\Sigma}_{\hat{\beta}} = n^{-1} \hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)} \quad (18.17)$$

El error estándar heterocedástico-robusto para el coeficiente de regresión j -ésimo es la raíz cuadrada del elemento j -ésimo de la diagonal de $\hat{\Sigma}_{\hat{\beta}}$. Es decir, el error estándar heterocedástico-robusto del coeficiente j -ésimo es

$$ES(\hat{\beta}_j) = \sqrt{(\hat{\Sigma}_{\hat{\beta}})_{jj}} \quad (18.18)$$

donde $(\hat{\Sigma}_{\hat{\beta}})_{jj}$ es el elemento (j,j) de $\hat{\Sigma}_{\hat{\beta}}$.

Intervalos de confianza para los efectos previstos

En la Sección 8.1 se describían dos métodos para calcular el error estándar de los efectos previstos que provocados por las variaciones en dos o más regresores. Existen expresiones compactas en forma matricial para estos errores estándar y, por lo tanto, para los intervalos de confianza de los efectos previstos.

Consideremos una variación en el valor de las variables explicativas para la observación i -ésima desde un valor inicial, por ejemplo $\mathbf{X}_{i,0}$, hasta un nuevo valor, $\mathbf{X}_{i,0} + \mathbf{d}$, por lo que la variación en \mathbf{X}_i es $\Delta\mathbf{X}_i = \mathbf{d}$, donde \mathbf{d} es un vector con dimensión $k+1$. Esta variación en \mathbf{X} puede afectar a varios regresores (es decir, a varios elementos de \mathbf{X}_i). Por ejemplo, si dos de los regresores son los valores de una variable independiente y su cuadrado, entonces \mathbf{d} es la diferencia entre los valores posteriores e iniciales de estas dos variables.

El efecto esperado de esta variación en \mathbf{X}_i es $\mathbf{d}'\beta$, y el estimador de este efecto es $\mathbf{d}'\hat{\beta}$. Debido a que las combinaciones lineales de variables aleatorias normalmente distribuidas están a su vez normalmente distribuidas, $\sqrt{n}(\mathbf{d}'\hat{\beta} - \mathbf{d}'\beta) = \mathbf{d}'\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbf{d}'\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}\mathbf{d})$. Por tanto, el error estándar de este efecto previsto es $(\mathbf{d}'\hat{\Sigma}_{\hat{\beta}}\mathbf{d})^{1/2}$. Un intervalo de confianza al 95 % para este efecto previsto es

$$\mathbf{d}'\hat{\beta} \pm 1,96 \sqrt{\mathbf{d}'\hat{\Sigma}_{\hat{\beta}}\mathbf{d}} \quad (18.19)$$

Distribución asintótica del estadístico t

El estadístico t para el contraste de la hipótesis de que $\beta_j = \beta_{j,0}$, construido mediante el error estándar heterocedástico-robusto de la Ecuación (18.18), está dado en el Concepto clave 7.1. El argumento de que este estadístico t tiene una distribución asintótica normal estándar es similar al argumento dado en la Sección 17.3 para el modelo con un único regresor.

18.3 Contrastes de hipótesis conjuntas

En la Sección 7.2 se consideraban los contrastes de hipótesis conjuntas que incluían varias restricciones, en las que cada restricción incluía un único coeficiente, y en la Sección 7.3 se consideraban los contrastes de una sola restricción que incluían dos o más coeficientes. La matriz enunciada en la Sección 18.1 permite una representación unificada de estos dos tipos de hipótesis de restricciones lineales sobre el vector de coeficientes, en la que cada restricción puede incluir varios coeficientes. Bajo los cuatro primeros supuestos de mínimos cuadrados del Concepto clave 18.1, el estadístico F MCO heterocedástico-robusto para contrastar estas hipótesis tiene una distribución asintótica $F_{q,\infty}$ bajo la hipótesis nula.

Hipótesis conjuntas en notación matricial

Consideremos una hipótesis conjunta que es lineal en los coeficientes y que impone q restricciones, donde $q \leq k + 1$. Cada una de estas q restricciones puede incluir uno o más coeficientes de regresión. Esta hipótesis nula conjunta se puede expresar en notación matricial como

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad (18.20)$$

donde \mathbf{R} es una matriz no aleatoria $q \times (k + 1)$ con rango de filas completo y \mathbf{r} es un vector no aleatorio $q \times 1$. El número de filas de \mathbf{R} es q , que es el número de restricciones que se imponen bajo la hipótesis nula.

La hipótesis nula de la Ecuación (18.20) recoge todas las hipótesis nulas consideradas en las Secciones 7.2 y 7.3. Por ejemplo, una hipótesis conjunta del tipo considerado en la Sección 7.2 es que $\beta_0 = 0, \beta_1 = 0, \dots, \beta_{q-1} = 0$. Para escribir este conjunto de hipótesis en la forma de la Ecuación (18.20), definimos $\mathbf{R} = [\mathbf{I}_q \mathbf{0}_{q \times (k+1-q)}]$ y $\mathbf{r} = \mathbf{0}_q$.

La formulación de la Ecuación (18.20) refleja asimismo las restricciones de la Sección 7.3 que incluyen varios coeficientes de regresión múltiple. Por ejemplo, si $k = 2$, entonces la hipótesis de que $\beta_1 + \beta_2 = 1$ se puede escribir en la forma de la Ecuación (18.20) haciendo $\mathbf{R} = [0 \ 1 \ 1]$, $\mathbf{r} = 1$ y $q = 1$.

Distribución asintótica del estadístico F

El estadístico F heterocedástico-robusto para el contraste de la hipótesis conjunta de la Ecuación (18.20) es

$$F = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}\hat{\Sigma}_{\hat{\boldsymbol{\beta}}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q \quad (18.21)$$

Si los primeros cuatro supuestos del Concepto clave 18.1 se cumplen, entonces bajo la hipótesis nula

$$F \xrightarrow{d} F_{q,\infty} \quad (18.22)$$

Este resultado se deduce mediante la combinación de la normalidad asintótica de $\hat{\boldsymbol{\beta}}$ con la consistencia del estimador heterocedástico-robusto $\hat{\Sigma}_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}$ de la matriz de covarianzas. En concreto, en primer lugar debe tenerse en cuenta que la Ecuación (18.12) y la Ecuación (18.74) del Apéndice 18.2 implican que, bajo la hipótesis nula, $\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{R}\Sigma_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}\mathbf{R}')$. Se deduce de la Ecuación (18.77) que, bajo la hipótesis nula, $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}\Sigma_{\hat{\boldsymbol{\beta}}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = [\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[\mathbf{R}\Sigma_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}\mathbf{R}']^{-1}[\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \xrightarrow{d} \chi_q^2$. No obstante, debido a que $\hat{\Sigma}_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})} \xrightarrow{P} \Sigma_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}$, se desprende del teorema de Slutsky que $[\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[\mathbf{R}\hat{\Sigma}_{\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}\mathbf{R}']^{-1}[\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \xrightarrow{d} \chi_q^2$ o, equivalentemente (debido a que $\hat{\Sigma}_{\boldsymbol{\beta}} = \hat{\Sigma}_{\boldsymbol{\beta}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}/n$), que $F \xrightarrow{d} \chi_q^2/q$, que a su vez se distribuye como una $F_{q,\infty}$.

Conjuntos de confianza para varios coeficientes

Tal y como se analizó en la Sección 7.4, se puede construir un conjunto de confianza asintóticamente válido para dos o más elementos de β como un conjunto de valores que, cuando se toman como hipótesis nula, no son rechazados por el estadístico F . En principio, este conjunto podría calcularse evaluando repetidamente el estadístico F para muchos valores de β , pero, tal y como ocurre en el caso de un intervalo de confianza para un único coeficiente, resulta más sencillo manipular la fórmula del estadístico de contraste para obtener una fórmula explícita para el conjunto de confianza.

He aquí el procedimiento para construir un conjunto de confianza para dos o más elementos de β . Sea δ el vector q -dimensional formado por los coeficientes para los cuales se desea construir un conjunto de confianza. Por ejemplo, si se está construyendo un conjunto de confianza para los coeficientes de regresión β_1 y β_2 , entonces $q = 2$ y $\delta = (\beta_1 \ \beta_2)'$. En general, se puede escribir $\delta = R\beta$, donde la matriz R se compone de ceros y unos [tal y como se trató de acuerdo con la Ecuación (18.20)]. El estadístico F para el contraste de la hipótesis de que $\delta = \delta_0$ es $F = (\hat{\delta} - \delta_0)'[R\hat{\Sigma}_\beta R']^{-1}(\hat{\delta} - \delta_0)/q$, donde $\hat{\delta} = R\hat{\beta}$. Un conjunto de confianza del 95 % para δ es el conjunto de valores de δ_0 que no son rechazados por el estadístico F . Es decir, cuando $\delta = R\beta$, un conjunto de confianza al 95 % para δ es

$$\{\delta : (\hat{\delta} - \delta)'[R\hat{\Sigma}_\beta R']^{-1}(\hat{\delta} - \delta)/q \leq c\} \quad (18.23)$$

donde c es el percentil 95 (el valor crítico al 5 %) de la distribución $F_{q,\infty}$.

El conjunto de la Ecuación (18.23) se compone de todos los puntos contenidos en el interior de la elipse definida cuando la desigualdad de la Ecuación (18.23) es una igualdad (esto es un elipsoide cuando $q > 2$). Por tanto el conjunto de confianza para δ se puede calcular mediante la resolución de la Ecuación (18.23) para los límites de la elipse.

18.4 Distribución de los estadísticos de regresión con errores normales

Las distribuciones presentadas en las Secciones 18.2 y 18.3, que están justificadas por la aplicación de la ley de los grandes números y del teorema central del límite, son aplicables cuando el tamaño de la muestra es grande. Sin embargo, si los errores son homocedásticos y están distribuidos normalmente, condicionados a X , entonces el estimador MCO tiene una distribución normal multivariante en muestras finitas, condicionada a X . Además, la distribución en muestras finitas del cuadrado del error estándar de la regresión es proporcional a la distribución chi-cuadrado con $n - k - 1$ grados de libertad, el estadístico t MCO válido con homocedasticidad tiene una distribución t de Student con $n - k - 1$ grados de libertad, y el estadístico F válido con homocedasticidad tiene una distribución $F_{q,n-k-1}$. Los argumentos de esta sección emplean algunas fórmulas matriciales especializadas para los estadísticos de regresión MCO, que se presentan por primera vez.

Representación matricial de los estadísticos de regresión MCO

Los valores de predicción MCO, los residuos, y la suma de los cuadrados de los residuos tienen representaciones matriciales compactas. Estas representaciones utilizan dos matrices P_X y M_X .

Las matrices P_X y M_X . El álgebra de MCO en el modelo multivariante se basa en dos matrices $n \times n$ simétricas, P_X y M_X :

$$P_X = X(X'X)^{-1}X' \text{ y} \quad (18.24)$$

$$M_X = I_n - P_X \quad (18.25)$$

Una matriz \mathbf{C} es idempotente si \mathbf{C} es cuadrada y $\mathbf{CC} = \mathbf{C}$ (véase el Apéndice 18.1). Debido a que $\mathbf{P}_X = \mathbf{P}_X\mathbf{P}_X$ y $\mathbf{M}_X = \mathbf{M}_X\mathbf{M}_X$ (Ejercicio 18.5), y debido a que \mathbf{P}_X y \mathbf{M}_X son simétricas, \mathbf{P}_X y \mathbf{M}_X son matrices simétricas e idempotentes.

Las matrices \mathbf{P}_X y \mathbf{M}_X tienen algunas propiedades adicionales útiles, que se derivan directamente de las definiciones de las Ecuaciones (18.24) y (18.25):

$$\begin{aligned}\mathbf{P}_X\mathbf{X} &= \mathbf{X} \text{ y } \mathbf{M}_X\mathbf{X} = \mathbf{0}_{n \times (k+1)} \\ \text{rango}(\mathbf{P}_X) &= k+1 \text{ y } \text{rango}(\mathbf{M}_X) = n-k-1\end{aligned}\quad (18.26)$$

donde $\text{rango}(\mathbf{P}_X)$ es el rango de la matriz \mathbf{P}_X .

Las matrices \mathbf{P}_X y \mathbf{M}_X puede utilizarse para descomponer un vector n -dimensional \mathbf{Z} en dos partes: una parte generada por las columnas de \mathbf{X} y una parte ortogonal a las columnas de \mathbf{X} . En otras palabras, $\mathbf{P}_X\mathbf{Z}$ es la proyección de \mathbf{Z} en el espacio generado por las columnas de \mathbf{X} , $\mathbf{M}_X\mathbf{Z}$ es la parte de \mathbf{Z} ortogonal a las columnas de \mathbf{X} , y $\mathbf{Z} = \mathbf{P}_X\mathbf{Z} + \mathbf{M}_X\mathbf{Z}$.

Valores de predicción y residuos MCO. Las matrices \mathbf{P}_X y \mathbf{M}_X proporcionan algunas expresiones sencillas para los valores de predicción y los residuos MCO. Los valores de predicción MCO, $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, y los residuos MCO, $\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}}$ se pueden expresar de la siguiente manera (Ejercicio 18.5):

$$\hat{\mathbf{Y}} = \mathbf{P}_X\mathbf{Y} \text{ y} \quad (18.27)$$

$$\hat{\mathbf{U}} = \mathbf{M}_X\mathbf{Y} = \mathbf{M}_X\mathbf{U} \quad (18.28)$$

Las expresiones de las Ecuaciones (18.27) y (18.28) proporcionan una prueba sencilla de que los residuos y los valores de predicción MCO son ortogonales, es decir, que la Ecuación (4.37) se cumple: $\hat{\mathbf{Y}}'\hat{\mathbf{U}} = \mathbf{Y}'\mathbf{P}_X'\mathbf{M}_X\mathbf{Y} = 0$, donde la segunda igualdad se deriva de $\mathbf{P}_X'\mathbf{M}_X = \mathbf{0}_{n \times m}$, lo que a su vez se deriva de que $\mathbf{M}_X\mathbf{X} = \mathbf{0}_{n \times (k+1)}$ en la Ecuación (18.26).

El error estándar de la regresión. El ESR, que se define en la Sección 4.3, es $s_{\hat{u}}$, donde

$$s_{\hat{u}}^2 = \frac{1}{n-k-1} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n-k-1} \hat{\mathbf{U}}'\hat{\mathbf{U}} = \frac{1}{n-k-1} \mathbf{U}'\mathbf{M}_X\mathbf{U} \quad (18.29)$$

donde la última igualdad se deriva de que $\hat{\mathbf{U}}'\hat{\mathbf{U}} = (\mathbf{M}_X\mathbf{U})'(\mathbf{M}_X\mathbf{U}) = \mathbf{U}'\mathbf{M}_X\mathbf{M}_X\mathbf{U} = \mathbf{U}'\mathbf{M}_X\mathbf{U}$ (por ser \mathbf{M}_X simétrica e idempotente).

Distribución de $\hat{\boldsymbol{\beta}}$ con errores normales

Debido a que $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}$ [Ecuación (18.14)] y debido a que la distribución de \mathbf{U} condicionada a \mathbf{X} es, por hipótesis, $N(\mathbf{0}_n, \sigma_u^2\mathbf{I}_n)$ [Ecuación (18.8)], la distribución condicional de $\hat{\boldsymbol{\beta}}$ dado \mathbf{X} es normal multivariante con media $\boldsymbol{\beta}$. La matriz de covarianzas de $\hat{\boldsymbol{\beta}}$, condicionada a \mathbf{X} , es $\Sigma_{\hat{\boldsymbol{\beta}}|\mathbf{X}} = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'|X] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}\mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|X] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma_u^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$. Por tanto, bajo los seis supuestos del Concepto clave 18.1, la distribución condicional para muestras finitas de $\hat{\boldsymbol{\beta}}$ dado \mathbf{X} es

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \Sigma_{\hat{\boldsymbol{\beta}}|\mathbf{X}}) \text{ donde } \Sigma_{\hat{\boldsymbol{\beta}}|\mathbf{X}} = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} \quad (18.30)$$

Distribución de $s_{\hat{u}}^2$

Si se cumplen los seis supuestos del Concepto clave 18.1, entonces $s_{\hat{u}}^2$ tiene una distribución muestral exacta que es proporcional a una distribución chi-cuadrado con $n-k-1$ grados de libertad:

$$s_{\hat{u}}^2 \sim \frac{\sigma_u^2}{n-k-1} \times \chi_{n-k-1}^2 \quad (18.31)$$

La prueba de la Ecuación (18.31) comienza con la Ecuación (18.29). Debido a que \mathbf{U} está distribuido normalmente condicionado a \mathbf{X} y debido a que \mathbf{M}_X es una matriz simétrica e idempotente, la forma cuadrática $\mathbf{U}'\mathbf{M}_X\mathbf{U}/\sigma_u^2$ tiene una distribución exacta chi-cuadrado con grados de libertad igual al rango de \mathbf{M}_X [Ecuación (18.78) del Apéndice 18.2]. A partir de la Ecuación (18.26), el rango de \mathbf{M}_X es $n - k - 1$. Por tanto $\mathbf{U}'\mathbf{M}_X\mathbf{U}/\sigma_u^2$ tiene una distribución exacta χ_{n-k-1}^2 , de la cual se deduce la Ecuación (18.31).

El ajuste por los grados de libertad asegura que $s_{\hat{u}}^2$ es insesgado. La esperanza de una variable aleatoria con una distribución χ_{n-k-1}^2 es $n - k - 1$; por lo que $E(\mathbf{U}'\mathbf{M}_X\mathbf{U}) = (n - k - 1)\sigma_u^2$, por lo que $E(s_{\hat{u}}^2) = \sigma_u^2$.

Errores estándar válidos con homocedasticidad

El estimador de $\tilde{\Sigma}_{\hat{\beta}}$ válido con homocedasticidad de la matriz de covarianzas de $\hat{\beta}$, condicionada a \mathbf{X} , se obtiene sustituyendo la varianza muestral $s_{\hat{u}}^2$ por la varianza poblacional σ_u^2 en la expresión para $\Sigma_{\hat{\beta}|X}$ de la Ecuación (18.30). En consecuencia,

$$\tilde{\Sigma}_{\hat{\beta}} = s_{\hat{u}}^2(\mathbf{X}'\mathbf{X})^{-1} \quad (\text{válido con homocedasticidad}) \quad (18.32)$$

El estimador de la varianza de la distribución normal condicional de $\hat{\beta}_j$, dado \mathbf{X} , es el elemento (j, j) de $\tilde{\Sigma}_{\hat{\beta}}$. Por tanto el error estándar válido con homocedasticidad de $\hat{\beta}_j$ es la raíz cuadrada del j -ésimo elemento de la diagonal principal de $\tilde{\Sigma}_{\hat{\beta}}$. Es decir, el error estándar válido con homocedasticidad de $\hat{\beta}_j$ es

$$\tilde{ES}(\hat{\beta}_j) = \sqrt{(\tilde{\Sigma}_{\hat{\beta}})_{jj}} \quad (\text{válido con homocedasticidad}) \quad (18.33)$$

Distribución del Estadístico t

Sea \tilde{t} el estadístico t para contrastar la hipótesis $\beta_j = \beta_{j,0}$, construido mediante el error estándar válido con homocedasticidad; es decir, sea

$$\tilde{t} = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{(\tilde{\Sigma}_{\hat{\beta}})_{jj}}} \quad (18.34)$$

Bajo los seis supuestos ampliados de mínimos cuadrados del Concepto clave 18.1, la distribución muestral exacta de \tilde{t} es la distribución t de Student con $n - k - 1$ grados de libertad; es decir,

$$\tilde{t} \sim t_{n-k-1} \quad (18.35)$$

La demostración de la Ecuación (18.35) se ofrece en el Apéndice 18.4.

Distribución del estadístico F

Si se cumplen los seis supuestos de mínimos cuadrados del Concepto clave 18.1, el estadístico F para contrastar la hipótesis de la Ecuación (18.20), construido con el estimador válido con homocedasticidad de la matriz de covarianzas, tiene una distribución exacta $F_{q, n-k-1}$ bajo la hipótesis nula.

El estadístico F válido con homocedasticidad. El estadístico F válido con homocedasticidad es similar al estadístico F heterocedástico-robusto de la Ecuación (18.21), salvo que se utiliza el estimador válido con homocedasticidad $\tilde{\Sigma}_{\hat{\beta}}$ en lugar del estimador heterocedástico-robusto $\hat{\Sigma}_{\hat{\beta}}$. Sustituyendo la expresión $\tilde{\Sigma}_{\hat{\beta}} = s_{\hat{u}}^2(\mathbf{X}'\mathbf{X})^{-1}$ en la expresión para el estadístico F de la Ecuación (18.21) se obtiene el estadístico F válido con homocedasticidad para el contraste de la hipótesis nula de la Ecuación (18.20):

$$\tilde{F} = \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})/q}{s_{\hat{u}}^2} \quad (18.36)$$

Si se cumplen los seis supuestos del Concepto clave 18.1, entonces bajo la hipótesis nula

$$\tilde{F} \sim F_{q, n-k-1} \quad (18.37)$$

La prueba de la Ecuación (18.37) se ofrece en el Apéndice 18.4.

El estadístico F de la Ecuación (18.36) se denomina versión de Wald del estadístico F (debe su nombre al del estadístico Abraham Wald). Aunque la fórmula para el estadístico F válido con homocedasticidad dada en la Ecuación (7.13) parece bastante distinta de la fórmula para el estadístico de Wald de la Ecuación (18.36), el estadístico F válido con homocedasticidad y el estadístico F de Wald son dos versiones de un mismo estadístico. Es decir, las dos expresiones son equivalentes, un resultado que se demuestra en el Ejercicio 18.13.

18.5 Eficiencia del estimador MCO con errores homocedásticos

Bajo las condiciones de Gauss-Markov para regresión múltiple, el estimador MCO de β es eficiente entre todos los estimadores lineales condicionalmente insesgados; es decir, el estimador MCO es ELIO.

Las condiciones de Gauss-Markov para regresión múltiple

Las **condiciones de Gauss-Markov para regresión múltiple** son

- (i) $E(\mathbf{U}|\mathbf{X}) = \mathbf{0}_n$,
- (ii) $E(\mathbf{U}\mathbf{U}'|\mathbf{X}) = \sigma_u^2 \mathbf{I}_n$, y
- (iii) X tiene rango completo de columnas. (18.38)

Las condiciones de Gauss-Markov para regresión múltiple a su vez están implícitas en los primeros cinco supuestos del Concepto clave 18.1 [véanse las Ecuaciones (18.6) y (18.7)]. Las condiciones de la Ecuación (18.38) generalizan las condiciones de Gauss-Markov para el modelo de un único regresor al modelo de regresión múltiple. [Utilizando la notación matricial, las condiciones segunda y tercera de Gauss-Markov de la Ecuación (5.31) están recogidas mediante una única condición (ii) en la Ecuación (18.38)].

Estimadores lineales condicionalmente insesgados

Comenzamos con la descripción del tipo de estimadores lineales e insesgados, y demostramos que los estimadores MCO pertenecen a la clase de este tipo de estimadores.

El tipo de estimadores lineales condicionalmente insesgados. Un estimador $\tilde{\beta}$ se dice que es lineal si es una función lineal de Y_1, \dots, Y_n . Por lo tanto, el estimador $\tilde{\beta}$ es lineal en Y si se puede escribir de la forma

$$\tilde{\beta} = A'Y \quad (18.39)$$

donde A es una matriz de dimensión $n \times (k + 1)$ de ponderaciones que pueden depender de X y de constantes no aleatorias, pero no de Y .

Un estimador es condicionalmente insesgado si la media de su distribución muestral condicional, dado X , es β . Es decir, $\tilde{\beta}$ es condicionalmente insesgado si $E(\tilde{\beta}|X) = \beta$.

El estimador MCO es lineal y condicionalmente insesgado. La comparación de las Ecuaciones (18.11) y (18.39) muestra que el estimador MCO es lineal en Y ; en concreto, $\hat{\beta} = \hat{A}'Y$, donde $\hat{A} = X(X'X)^{-1}$. Para demostrar que $\hat{\beta}$ es condicionalmente insesgado, recordemos de la Ecuación (18.14) que $\hat{\beta} = \beta + (X'X)^{-1}X'U$. Tomando la esperanza condicional a ambos lados de esta expresión se obtiene $E(\hat{\beta}|X) = \beta + E[(X'X)^{-1}X'U|X] = \beta + (X'X)^{-1}X'E(U|X) = \beta$. Donde la última igualdad se deriva de que $E(U|X) = 0$ por la primera condición de Gauss-Markov.

El teorema de Gauss-Markov para regresión múltiple

El **teorema de Gauss-Markov para regresión múltiple** proporciona las condiciones bajo las cuales el estimador MCO es eficiente entre la clase de estimadores lineales condicionalmente insesgados. No obstante

**CONCEPTO
CLAVE**

18.3

El Teorema de Gauss-Markov para regresión múltiple

Supongamos que se cumplen las condiciones de Gauss-Markov para regresión múltiple de la Ecuación (18.38). Entonces el estimador $\hat{\beta}$ MCO es ELIO. Es decir, sea un estimador lineal condicionalmente insesgado de β y sea c un vector no aleatorio de dimensión $k + 1$. Entonces, $\text{var}(c'\hat{\beta}|X) \leq \text{var}(c'\tilde{\beta}|X)$ para todo vector c distinto de cero, y donde la desigualdad se cumple con igualdad para todo c solamente si $\tilde{\beta} = \hat{\beta}$.

surge una cuestión sutil debido a que $\hat{\beta}$ es un vector y su «varianza» es una matriz de covarianzas. Si la «varianza» de un estimador es una matriz, ¿qué significa exactamente decir que un estimador tiene una varianza menor que otro?

El teorema de Gauss-Markov aborda este problema comparando la varianza de un estimador candidato de una *combinación lineal* de los elementos de β con la varianza de la combinación lineal correspondiente de $\hat{\beta}$. En concreto, sea c un vector con dimensión $k + 1$ y consideremos el problema de estimar la combinación lineal $c'\beta$ mediante el estimador candidato $c'\tilde{\beta}$ (donde $\tilde{\beta}$ es un estimador lineal condicionalmente insesgado), por un lado y $c'\hat{\beta}$ por el otro lado. Debido a que $c'\tilde{\beta}$ y $c'\hat{\beta}$ son ambos escalares y ambos son estimadores lineales condicionalmente insesgados de $c'\beta$, ahora tiene sentido comparar sus varianzas.

El teorema de Gauss-Markov para regresión múltiple, dice que el estimador MCO de $c'\beta$ es eficiente; es decir, el estimador MCO $c'\hat{\beta}$ tiene la menor varianza condicional de todos los estimadores lineales condicionalmente insesgados $c'\tilde{\beta}$. Sorprendentemente, esto es cierto, sin que importe cuál sea la combinación lineal. Es en este sentido en el que el estimador MCO es ELIO en regresión múltiple.

El teorema de Gauss-Markov se enuncia en el Concepto clave 18.3 y se demuestra en el Apéndice 18.5.

18.6 Mínimos cuadrados generalizados¹

El supuesto de muestreo i.i.d. se ajusta a muchas aplicaciones. Por ejemplo, supongamos que Y_i y X_i corresponden a información acerca de individuos, tal como sus ingresos, educación, y características personales, y donde los individuos han sido seleccionados a partir de una población mediante muestreo aleatorio simple. En este caso, debido al esquema de muestreo aleatorio simple (X_i, Y_i), son necesariamente i.i.d. Debido a que (X_i, Y_i) y (X_j, Y_j) están independientemente distribuidas para $i \neq j$, u_i y u_j están independientemente distribuidas para $i \neq j$. Esto a su vez implica que u_i y u_j están incorrelacionados para $i \neq j$. En el contexto de los supuestos de Gauss-Markov, el supuesto de que $E(UU'|X)$ es diagonal es por tanto apropiado si los datos se recogen de manera que las observaciones estén distribuidas de forma independiente.

Sin embargo, algunos métodos de muestreo que aparecen en econometría no dan lugar a observaciones independientes y pueden conducir en su caso a términos de error u_i que están correlacionados entre una observación y otra. El ejemplo más claro es cuando los datos se recogen a lo largo del tiempo sobre la misma entidad individual, es decir, cuando los datos son datos de series temporales. Tal y como se trató en la Sección 15.3, en las regresiones que incluyen datos de series temporales, muchos de los factores omitidos están correlacionados entre un periodo a otro, y esto puede dar lugar a términos de error de la regresión (que representan los factores omitidos) que estén correlacionados de un periodo de observación al siguiente. En otras palabras, el término de error de un periodo no estará, en general, distribuido de forma independiente del término de error del periodo siguiente. En cambio, el término de error de un periodo podría estar correlacionado con el término de error del periodo siguiente.

La presencia de términos de error correlacionados crea dos problemas para la inferencia basada en MCO. En primer lugar, *ni* los errores estándar heterocedástico-robustos *ni* los errores estándar válidos con homocedasticidad provinientes de MCO proporcionan una base válida para la inferencia. La solución a este

¹ El estimador MCG se introdujo en la Sección 15.5 en el contexto de la regresión de series temporales con retardos distribuidos. Su presencia aquí se debe a un tratamiento matemático autocontenido de MCG que permite su lectura independiente de la Sección 15.5, no obstante la lectura previa de esta sección ayuda a concretar esas ideas.

problema consiste en utilizar errores estándar que sean robustos, tanto en presencia de heterocedasticidad como de correlación entre los términos de error entre observaciones. Este tema —la estimación de la matriz de covarianzas consistente a heterocedasticidad y autocorrelación (HAC)— es el tema de la Sección 15.4 y no continuaremos aquí con él.

En segundo lugar, si el término de error está correlacionado entre observaciones, entonces $E(\mathbf{U}\mathbf{U}'|\mathbf{X})$ es no diagonal, no se cumple la segunda condición de Gauss-Markov de la Ecuación (18.38), y MCO no es ELIO. En esta sección se estudia un estimador, el de **mínimos cuadrados generalizados** (MCG), que es ELIO (al menos asintóticamente) cuando la matriz de covarianzas condicional de los errores ya no es proporcional a la matriz identidad. Un caso particular de MCG son los mínimos cuadrados ponderados, analizados en la Sección 17.5, en los que la matriz de covarianzas condicional es diagonal y el i -ésimo elemento de la diagonal principal es una función de X_i . Al igual que MCP, MCG transforma el modelo de regresión para que los errores del modelo transformado satisfagan las condiciones de Gauss-Markov. El estimador MCG es el estimador MCO de los coeficientes del modelo transformado.

Los supuestos de MCG

Existen cuatro supuestos bajo los cuales MCG resulta válido. El primer supuesto MCG es que u_i tiene una media igual a cero, condicionada a $\mathbf{X}_1, \dots, \mathbf{X}_n$; es decir,

$$E(\mathbf{U}|\mathbf{X}) = \mathbf{0}_n \quad (18.40)$$

Este supuesto está implícito en los dos primeros supuestos de mínimos cuadrados del Concepto clave 18.1; es decir, si $E(u_i|\mathbf{X}_i) = 0$ y (\mathbf{X}_i, Y_i) , $i = 1, \dots, n$, son i.i.d., entonces $E(\mathbf{U}|\mathbf{X}) = \mathbf{0}_n$. Sin embargo, en MCG no pretendemos que se cumpla el supuesto de i.i.d.; después de todo, un objetivo de MCG es el de tratar los errores que están correlacionados entre observaciones. Se analizará la importancia del supuesto de la Ecuación (18.40) después de introducir el estimador MCG.

El segundo supuesto MCG es que la matriz de covarianzas condicional de \mathbf{U} dado \mathbf{X} es una función de \mathbf{X} :

$$E(\mathbf{U}\mathbf{U}'|\mathbf{X}) = \Omega(\mathbf{X}) \quad (18.41)$$

donde $\Omega(\mathbf{X})$ es una matriz $n \times n$ definida positiva de valores que son función de \mathbf{X} .

Existen dos aplicaciones principales de MCG que están cubiertas por este supuesto. La primera es el muestreo independiente con errores heterocedásticos, en cuyo caso $\Omega(\mathbf{X})$ es una matriz diagonal cuyos elementos de la diagonal principal son $\lambda h(\mathbf{X}_i)$, donde λ es una constante y h es una función. En este caso, estudiado en la Sección 17.5, MCG es MCP.

La segunda aplicación es para errores homocedásticos que están serialmente correlacionados. En la práctica, en este caso se desarrolla un modelo para la correlación serial. Por ejemplo, un posible modelo es que el término de error esté correlacionado solamente con su vecino inmediato, por lo que $\text{corr}(u_i, u_{i-1}) = \rho \neq 0$, pero $\text{corr}(u_i, u_j) = 0$ si $|i - j| \geq 2$. En este caso, $\Omega(\mathbf{X})$ tiene σ_u^2 como elemento de la diagonal principal, $\rho\sigma_u^2$ en la primera diagonal contigua a la diagonal principal y ceros en el resto. Por lo que $\Omega(\mathbf{X})$ no depende de \mathbf{X} , $\Omega_{ii} = \sigma_u^2$, $\Omega_{ij} = \rho\sigma_u^2$ para $|i - j| = 1$ y $\Omega_{ij} = 0$ para $|i - j| > 1$. Se analizarán más adelante otros modelos para la correlación serial, incluyendo el modelo autorregresivo de primer orden, en el contexto de MCG, en la Sección 15.5 (véase asimismo el Ejercicio 18.8).

Un supuesto que ha aparecido en todas las listas anteriores de supuestos de mínimos cuadrados para datos de sección cruzada es que \mathbf{X}_i y u_i tengan momentos de cuarto orden finitos y distintos de cero. En el caso de MCG, los supuestos necesarios específicos sobre los momentos para demostrar los resultados asintóticos dependen de la naturaleza de la función $\Omega(\mathbf{X})$, de si $\Omega(\mathbf{X})$ es conocido o estimado, así como del estadístico en cuestión (el estimador MCG, el estadístico t , etc.). Debido a que los supuestos son específicos para cada caso y modelo, no se presentan aquí supuestos específicos sobre los momentos y el análisis de las propiedades en muestras grandes de MCG supone que estas condiciones específicas sobre los momentos se cumplen para el caso pertinente que nos ocupe. Para completar, como tercer supuesto MCG, se supone simplemente que \mathbf{X}_i y u_i satisfacen las condiciones adecuadas sobre los momentos.

**CONCEPTO
CLAVE**

18.4

Los supuestos MCG

En el modelo de regresión lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$, los supuestos MCG son

1. $E(\mathbf{U}|\mathbf{X}) = \mathbf{0}_n$.
2. $E(\mathbf{U}\mathbf{U}'|\mathbf{X}) = \boldsymbol{\Omega}(\mathbf{X})$, donde $\boldsymbol{\Omega}(\mathbf{X})$ es una matriz $n \times n$ definida positiva que puede depender de \mathbf{X} .
3. X_i y u_i satisfacen las condiciones sobre los momentos adecuadas.
4. \mathbf{X} tiene rango completo por columnas (no existe multicolinealidad perfecta).

El cuarto supuesto MCG es que \mathbf{X} tiene rango completo de columnas; es decir, las variables explicativas no son perfectamente multicolineales.

Los supuestos MCG se recogen en el Concepto clave 18.4.

Consideramos la estimación MCG en dos casos. En el primer caso, $\boldsymbol{\Omega}(\mathbf{X})$ es conocido. En el segundo caso, la forma funcional de $\boldsymbol{\Omega}(\mathbf{X})$ es conocida con algunos parámetros que pueden estimarse. Para simplificar la notación nos referiremos a la función $\boldsymbol{\Omega}(\mathbf{X})$ como matriz $\boldsymbol{\Omega}$. Por lo que la dependencia de $\boldsymbol{\Omega}$ sobre \mathbf{X} se encuentra implícita.

MCG con $\boldsymbol{\Omega}$ conocida

Cuando $\boldsymbol{\Omega}$ es conocida, el estimador MCG utiliza $\boldsymbol{\Omega}$ para transformar el modelo de regresión original en otro con errores que cumplan las condiciones de Gauss-Markov. En concreto, sea \mathbf{F} una matriz raíz cuadrada de $\boldsymbol{\Omega}^{-1}$; es decir, sea \mathbf{F} una matriz que satisface $\mathbf{F}'\mathbf{F} = \boldsymbol{\Omega}^{-1}$ (véase el Apéndice 18.1). Una propiedad de \mathbf{F} es que $\mathbf{F}\boldsymbol{\Omega}\mathbf{F}' = \mathbf{I}_n$. Premultiplicando ahora ambos lados de la Ecuación (18.4) por \mathbf{F} se obtiene

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{U}} \quad (18.42)$$

donde $\tilde{\mathbf{Y}} = \mathbf{F}\mathbf{Y}$, $\tilde{\mathbf{X}} = \mathbf{F}\mathbf{X}$, y $\tilde{\mathbf{U}} = \mathbf{F}\mathbf{U}$.

La idea clave de MCG es que, bajos los cuatro supuestos MCG, se cumplen los supuestos de Gauss-Markov para la regresión transformada de la Ecuación (18.42). Es decir, mediante la transformación de todas las variables por la inversa de la matriz raíz cuadrada de $\boldsymbol{\Omega}$, los errores de regresión en la regresión transformada tienen una media condicional igual a cero y una matriz de covarianzas que es igual a la matriz identidad. Para demostrar esto matemáticamente, en primer lugar hay que tener en cuenta que $E(\tilde{\mathbf{U}}|\tilde{\mathbf{X}}) = E(\mathbf{F}\mathbf{U}|\mathbf{F}\mathbf{X}) = \mathbf{F}E(\mathbf{U}|\mathbf{F}\mathbf{X} = \mathbf{0}_n)$ por el primer supuesto MCG [Ecuación (18.40)]. Además, $E(\tilde{\mathbf{U}}\tilde{\mathbf{U}}'|\tilde{\mathbf{X}}) = E[(\mathbf{F}\mathbf{U})(\mathbf{F}\mathbf{U})'|\mathbf{F}\mathbf{X}] = \mathbf{F}E(\mathbf{U}\mathbf{U}'|\mathbf{F}\mathbf{X})\mathbf{F}' = \mathbf{F}\boldsymbol{\Omega}\mathbf{F}' = \mathbf{I}_n$, donde la segunda igualdad se deriva del hecho de que $(\mathbf{F}\mathbf{U})' = \mathbf{U}'\mathbf{F}'$ y la última igualdad se obtiene a partir de la definición de \mathbf{F} . De ello se desprende que el modelo de regresión transformado de la Ecuación (18.42) satisface las condiciones de Gauss-Markov del Concepto clave 18.3.

El estimador MCG, $\tilde{\boldsymbol{\beta}}^{MCG}$ es el estimador MCO de $\boldsymbol{\beta}$ en la Ecuación (18.42); es decir, $\tilde{\boldsymbol{\beta}}^{MCG} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\tilde{\mathbf{Y}})$. Debido a que el modelo de regresión transformado satisface las condiciones de Gauss-Markov, el estimador MCG es el estimador condicionalmente insesgado óptimo que es lineal en $\tilde{\mathbf{Y}}$. Pero debido a que $\tilde{\mathbf{Y}} = \mathbf{F}\mathbf{Y}$ y \mathbf{F} (en este caso) se supone que es conocido, y debido a que \mathbf{F} es invertible (por ser $\boldsymbol{\Omega}$ definida positiva), la clase de estimadores que son lineales en $\tilde{\mathbf{Y}}$ es la misma clase de estimadores que son lineales en \mathbf{Y} . Por lo tanto el estimador MCO de $\boldsymbol{\beta}$ en la Ecuación (18.42) es asimismo el estimador condicionalmente insesgado óptimo entre los estimadores que son lineales en \mathbf{Y} . En otras palabras, bajo los supuestos de MCG, el estimador MCG es ELIO.

El estimador MCG puede expresarse directamente en términos de $\boldsymbol{\Omega}$, por lo que en principio no es necesario calcular la raíz cuadrada de la matriz \mathbf{F} . Debido a que $\tilde{\mathbf{X}} = \mathbf{F}\mathbf{X}$ y $\tilde{\mathbf{Y}} = \mathbf{F}\mathbf{Y}$, $\tilde{\boldsymbol{\beta}}^{MCG} = (\mathbf{X}'\mathbf{F}'\mathbf{F}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{F}'\mathbf{F}\mathbf{Y})$. Pero $\mathbf{F}'\mathbf{F} = \boldsymbol{\Omega}^{-1}$, por lo que.

$$\tilde{\boldsymbol{\beta}}^{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}) \quad (18.43)$$

En la práctica, Ω suele ser desconocida, por lo que el estimador MCG de la Ecuación (18.43) por lo general no se puede calcular y, por lo tanto, a veces se le denomina estimador **MCG infactible**. Sin embargo, si Ω tiene una forma funcional conocida, pero los parámetros de esa función son desconocidos, entonces Ω se puede estimar y se puede calcular una versión factible del estimador MCG.

MCG cuando Ω contiene parámetros desconocidos

Si Ω es una función conocida de algunos parámetros que a su vez pueden ser estimados, entonces se pueden utilizar estos parámetros estimados para calcular un estimador de la matriz de covarianzas. Por ejemplo, consideremos la aplicación de series temporales analizada de acuerdo con la Ecuación (18.41), en la que $\Omega(X)$ no depende de X , $\Omega_{ij} = \sigma_u^2$, $\Omega_{ij} = \rho\sigma_u^2$ para $|i - j| = 1$, y $\Omega_{ij} = 0$ para $|i - j| > 1$. Por tanto, Ω tiene dos parámetros desconocidos, σ_u^2 y ρ . Estos parámetros se pueden estimar a partir de los residuos de una regresión MCO preliminar; en concreto σ_u^2 se puede estimar mediante s_u^2 y ρ puede ser estimado a partir de la correlación muestral entre todos los pares de residuos MCO contiguos. Estos parámetros estimados se pueden utilizar a su vez para calcular un estimador de Ω , $\hat{\Omega}$.

En general, supongamos que se dispone de un estimador $\hat{\Omega}$ de Ω . Entonces el estimador MCG basado en $\hat{\Omega}$ es

$$\hat{\beta}^{MCG} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y) \quad (18.44)$$

El estimador MCG de la Ecuación (18.44) se denomina a veces estimador de **MCG factibles**, ya que se puede calcular si la matriz de covarianzas contiene algunos parámetros desconocidos que pueden ser estimados.

El supuesto de media condicional igual a cero y MCG

Para que el estimador MCO sea consistente, debe cumplirse el primer supuesto de mínimos cuadrados; es decir, $E(u_i|X_i)$ debe ser igual a cero. Por otro lado, el primer supuesto MCG es que $E(u_i|X_1, \dots, X_n) = 0$. En otras palabras, el primer supuesto de MCO es que el error para la i -ésima observación tiene una media condicional igual a cero, dados los valores de los regresores para esa observación, mientras que el primer supuesto MCG es que u_i tiene una media condicional igual a cero dados los valores de las variables explicativas para *todas* las observaciones.

Tal y como se analizó en la Sección 18.1, los supuestos de que $E(u_i|X_i) = 0$ y de que el muestreo sea i.i.d. implican conjuntamente que $E(u_i|X_1, \dots, X_n) = 0$. Por tanto, cuando el muestreo es i.i.d. con lo que MCG es MCP, el primer supuesto MCG está implícito en el primer supuesto de mínimos cuadrados del Concepto clave 18.1.

No obstante, cuando el muestreo no es i.i.d., el primer supuesto MCG no está implícito en el supuesto de que $E(u_i|X_i) = 0$; es decir, el primer supuesto MCG es más fuerte. Aunque la distinción entre estas dos condiciones podría parecer insignificante, puede resultar muy importante en las aplicaciones con datos de series temporales. Esta distinción se estudia en la Sección 15.5 en el contexto de si el regresor es exógeno «pasado y presente» o «estrictamente» exógeno; el supuesto de que $E(u_i|X_1, \dots, X_n) = 0$ corresponde a la exogeneidad estricta. En este caso, analizamos esta distinción a un nivel más general mediante notación matricial. Para ello, nos centramos en el caso en que U es homocedástico, Ω es conocida, y Ω tiene elementos fuera de la diagonal principal distintos de cero.

El papel del primer supuesto de MCG. Para ver el origen de la diferencia entre estos supuestos, resulta útil contrastar los argumentos sobre la consistencia para MCG y MCO.

En primer lugar, se esboza el argumento a favor de la consistencia del estimador MCG de la Ecuación (18.43). Sustituyendo la Ecuación (18.4) en la Ecuación (18.43), se obtiene que $\tilde{\beta}^{MCG} = \beta + (X'\Omega^{-1}X/n)^{-1}(X'\Omega^{-1}U/n)$. Bajo el primer supuesto de MCG, $E(X'\Omega^{-1}U) = E[X'\Omega^{-1}E(U|X)] = \mathbf{0}_n$. Si además la varianza de $X'\Omega^{-1}U/n$ tiende a cero y $X'\Omega^{-1}X/n \xrightarrow{P} \tilde{Q}$, donde \tilde{Q} es alguna matriz invertible, entonces $\tilde{\beta}^{MCG} \xrightarrow{P} \beta$. Fundamentalmente, cuando Ω tiene elementos fuera de la diagonal principal distintos de cero, el término $X'\Omega^{-1}U = \sum_{i=1}^n \sum_{j=1}^n X_i(\Omega^{-1})_{ij}u_j$ incluye los productos de X_i y u_j para diferentes i, j , donde

$(\Omega^{-1})_{ij}$ es el elemento (i, j) de Ω^{-1} . Por tanto, para que $\mathbf{X}'\Omega^{-1}\mathbf{U}$ tenga una media igual a cero, no es suficiente con que $E(u_i|X_i) = 0$; sino que además $E(u_i|X_j)$ debe ser igual a cero para todos los pares i, j que se corresponden con los valores distintos de cero de $(\Omega^{-1})_{ij}$. Dependiendo de la estructura de la covarianza de los errores, solamente algunos del total de los elementos de $(\Omega^{-1})_{ij}$ podrían ser distintos de cero. Por ejemplo, si u_i sigue un modelo autorregresivo de primer orden (como se analizó en la Sección 15.5), los únicos elementos distintos de cero $(\Omega^{-1})_{ij}$ son aquellos para los que $|i - j| \leq 1$. No obstante, en general, todos los elementos de Ω^{-1} pueden ser distintos de cero, por lo que en general, para que $\mathbf{X}'\Omega^{-1}\mathbf{U}/n \xrightarrow{P} \mathbf{0}_{(k+1) \times 1}$ (y por tanto, para que $\tilde{\beta}^{MCG}$ sea consistente) es necesario que $E(\mathbf{U}|\mathbf{X}) = \mathbf{0}_n$; es decir, debe cumplirse el primer supuesto de MCG.

Por otra parte, recordando el argumento de que el estimador MCO es consistente. Reescribimos la Ecación (18.14) como $\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X}/n)^{-1}\frac{1}{2}\sum_{i=1}^n \mathbf{X}_i u_i$. Si $E(u_i|X_i) = 0$, entonces el término $\frac{1}{2}\sum_{i=1}^n \mathbf{X}_i u_i$ tiene media igual a cero, y, si este término tiene una varianza que tiende a cero, converge en probabilidad a cero. Si además $\mathbf{X}'\mathbf{X}/n \xrightarrow{P} Q_x$, entonces $\hat{\beta} \xrightarrow{P} \beta$.

¿Es restrictivo el primer supuesto de MCG? El primer supuesto MCG requiere que los errores de la observación i -ésima estén incorrelacionados con los regresores para todas las demás observaciones. Este supuesto es dudoso en algunas aplicaciones de series temporales. Esta cuestión se analiza en la Sección 15.6 en el contexto de un ejemplo empírico, la relación entre la variación en el precio de un contrato futuros de concentrado de zumo de naranja congelado y las condiciones meteorológicas de Florida. Tal y como se explica aquí, el término de error de la regresión de la variación de los precios sobre las condiciones meteorológicas está probablemente incorrelacionado con los valores actuales y pasados de las condiciones meteorológicas, por lo que se cumpliría el primer supuesto de MCO. Sin embargo, este término de error se encuentra probablemente correlacionado con los valores futuros de las condiciones meteorológicas, por lo que no se cumple el primer supuesto de MCG.

Este ejemplo ilustra un fenómeno general en los datos económicos de series temporales que surge cuando el valor de una variable en la actualidad se establece en parte en base a las expectativas futuras: esas expectativas futuras por lo general implican que el término de error hoy depende de una previsión del regresor para mañana, que a su vez está correlacionada con el valor real del regresor mañana. Por esta razón, el primer supuesto MCG es en realidad mucho más fuerte que el primer supuesto MCO. En consecuencia, en algunas aplicaciones con datos económicos de series temporales el estimador MCG no es consistente aunque el estimador MCO lo sea.

18.7 Variables instrumentales y estimación por el método generalizado de momentos

En esta sección se proporciona una introducción a la teoría de estimación por variables instrumentales (VI) así como a la distribución asintótica de los estimadores VI. Se ha supuesto hasta ahora que los supuestos de la regresión VI de los Conceptos clave 12.3 y 12.4 se cumplen y, además, que los instrumentos son fuertes. Estos supuestos son aplicables a datos de sección cruzada con observaciones i.i.d. Bajo ciertas condiciones, los resultados obtenidos en esta sección son aplicables asimismo a los datos de series temporales, y la extensión a los datos de series temporales se analiza brevemente al final de esta sección. Todos los resultados asintóticos de esta sección se obtienen bajo el supuesto de instrumentos fuertes.

Esta sección comienza con la presentación del modelo de regresión VI, el estimador de mínimos cuadrados en dos etapas (MC2E) y su distribución asintótica en el caso general de heterocedasticidad, todo ello en forma matricial. A continuación se muestra que, en el caso particular de homocedasticidad, el estimador MC2E es asintóticamente eficiente entre la clase de estimadores VI cuyos instrumentos son combinaciones lineales de las variables exógenas. Por otra parte, el estadístico J tiene una distribución asintótica chi-cuadrado en la que los grados de libertad son iguales al número de restricciones sobreidentificadas. Esta sección concluye con un análisis de la estimación VI eficiente y el contraste de sobreidentificación de restricciones con errores heterocedásticos, una situación en la que el estimador VI eficiente se conoce como estimador del método generalizado de momentos (MGM).

El estimador VI en forma matricial

En esta sección, sea \mathbf{X} la matriz $n \times (k + r + 1)$ de regresores en la ecuación de interés, por lo que \mathbf{X} contiene los regresores endógenos incluidos (las X del Concepto clave 12.1) y los regresores exógenos incluidos (las W del Concepto clave 12.1). Es decir, en la notación del Concepto clave 12.1, la fila i -ésima de \mathbf{X} es $\mathbf{X}_i = (1 \ X_{1i} \ X_{2i} \ \dots \ X_{ki} \ W_{1i} \ W_{2i} \ \dots \ W_{ri})$. Por otra parte, sea \mathbf{Z} la matriz $n \times (m + r + 1)$ de todos los regresores exógenos, tanto los incluidos en la ecuación de interés (las W) como los excluidos de la ecuación de interés (los instrumentos). Es decir, en la notación del Concepto clave 12.1, la i -ésima fila de \mathbf{Z} es $\mathbf{Z}_i = (1 \ Z_{1i} \ Z_{2i} \ \dots \ Z_{mi} \ W_{1i} \ W_{2i} \ \dots \ W_{ri})$.

Con esta notación, el modelo de regresión VI del Concepto clave 12.1, expresado en forma matricial, es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \quad (18.45)$$

donde \mathbf{U} es el vector $n \times 1$ de los errores de la ecuación de interés, cuyo i -ésimo elemento es u_i .

La matriz \mathbf{Z} se compone de todos los regresores exógenos, por lo que bajo los supuestos de la regresión VI del Concepto clave 12.4,

$$E(\mathbf{Z}_i u_i) = \mathbf{0} \quad (\text{exogeneidad de instrumentos}) \quad (18.46)$$

Debido a que existen k regresores endógenos incluidos, la regresión de la primera etapa consta de k ecuaciones.

El estimador MC2E. El estimador MC2E es el estimador de variables instrumentales en el que los instrumentos son los valores de predicción de \mathbf{X} basados en la estimación MCO de la regresión de la primera etapa. Sea $\hat{\mathbf{X}}$ esta matriz de valores estimados de manera que la fila i -ésima de $\hat{\mathbf{X}}$ es $(\hat{X}_{1i} \ \hat{X}_{2i} \ \dots \ \hat{X}_{ki} \ W_{1i} \ W_{2i} \ \dots \ W_{ri})$, donde \hat{X}_{1i} es el valor de predicción de X_{1i} sobre \mathbf{Z} , etc. Debido a que las W están contenidas en \mathbf{Z} , el valor estimado de una regresión de W_{1i} sobre \mathbf{Z} es exactamente W_{1i} , etc., por lo que $\hat{\mathbf{X}} = \mathbf{P}_Z \mathbf{X}$, donde $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ [véase la Ecuación (18.27)]. En consecuencia, el estimador MC2E es

$$\hat{\boldsymbol{\beta}}^{MC2E} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{Y} \quad (18.47)$$

Debido a que $\hat{\mathbf{X}} = \mathbf{P}_Z \mathbf{X}$, $\hat{\mathbf{X}}'\hat{\mathbf{X}} = \mathbf{X}'\mathbf{P}_Z \mathbf{X}$, y $\hat{\mathbf{X}}'\mathbf{Y} = \mathbf{X}'\mathbf{P}_Z \mathbf{Y}$, el estimador MC2E se puede reescribir como

$$\hat{\boldsymbol{\beta}}^{MC2E} = (\mathbf{X}'\mathbf{P}_Z \mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z \mathbf{Y} \quad (18.48)$$

Distribución asintótica del estimador MC2E

Sustituyendo la Ecuación (18.45) en la Ecuación (18.48), reordenando, y multiplicando por \sqrt{n} se obtiene la expresión para el estimador MC2E centrado y escalado:

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}^{MC2E} - \boldsymbol{\beta}) &= \left(\frac{\mathbf{X}'\mathbf{P}_Z \mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}'\mathbf{P}_Z \mathbf{U}}{\sqrt{n}} \\ &= \left[\frac{\mathbf{X}'\mathbf{Z}}{n} \left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1} \frac{\mathbf{Z}'\mathbf{X}}{n}\right]^{-1} \left[\frac{\mathbf{X}'\mathbf{Z}}{n} \left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1} \frac{\mathbf{Z}'\mathbf{u}}{\sqrt{n}}\right] \end{aligned} \quad (18.49)$$

donde en la segunda igualdad se utiliza la definición de \mathbf{P}_Z . Bajo los supuestos de la regresión VI, $\mathbf{X}'\mathbf{Z}/n \xrightarrow{P} \mathbf{Q}_{XZ}$ y $\mathbf{Z}'\mathbf{Z}/n \xrightarrow{P} \mathbf{Q}_{ZZ}$, donde $\mathbf{Q}_{XZ} = E(\mathbf{X}_i \mathbf{Z}_i')$ y $\mathbf{Q}_{ZZ} = E(\mathbf{Z}_i \mathbf{Z}_i')$. Además, bajo los supuestos de regresión VI, $\mathbf{Z}_i u_i$ es i.i.d. con media igual a cero [Ecuación (18.46)] y varianza finita y distinta de cero, por lo que su suma, dividida por \sqrt{n} , satisface las condiciones del teorema central del límite y

$$\mathbf{Z}'\mathbf{U}/\sqrt{n} \xrightarrow{d} \boldsymbol{\Psi}_{ZU}, \text{ donde } \boldsymbol{\Psi}_{ZU} \sim N(\mathbf{0}, \mathbf{H}) \text{ y } \mathbf{H} = E(\mathbf{Z}_i \mathbf{Z}_i' u_i^2) \quad (18.50)$$

donde $\boldsymbol{\Psi}_{ZU}$ es $(m + r + 1) \times 1$.

La aplicación de la Ecuación (18.50) y de los límites $\mathbf{X}'\mathbf{Z}/n \xrightarrow{P} \mathbf{Q}_{XZ}$ y $\mathbf{Z}'\mathbf{Z}/n \xrightarrow{P} \mathbf{Q}_{ZZ}$ a la Ecuación (18.49) proporciona el resultado de que, bajo los supuestos de la regresión VI, el estimador MC2E está distribuido asintóticamente normal:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{MC2E} - \boldsymbol{\beta}) \xrightarrow{d} (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{ZU} \sim N(\mathbf{0}, \Sigma^{MC2E}) \quad (18.51)$$

donde

$$\Sigma^{MC2E} = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{H}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1} \quad (18.52)$$

donde \mathbf{H} se define en la Ecuación (18.50).

Errores estándar para MC2E. La fórmula de la Ecuación (18.52) es de proporciones enormes. Sin embargo, proporciona una manera de estimar Σ^{MC2E} mediante la sustitución de los momentos muestrales por los momentos poblacionales. El estimador de la varianza resultante es

$$\hat{\Sigma}^{MC2E} = (\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1}\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{H}}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX}(\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1} \quad (18.53)$$

donde $\hat{\mathbf{Q}}_{XZ} = \mathbf{X}'\mathbf{Z}/n$, $\hat{\mathbf{Q}}_{ZZ} = \mathbf{Z}'\mathbf{Z}/n$, $\hat{\mathbf{Q}}_{ZX} = \mathbf{Z}'\mathbf{X}/n$, y

$$\hat{\mathbf{H}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i' \hat{u}_i^2, \text{ donde } \hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{MC2E} \quad (18.54)$$

por lo que $\hat{\mathbf{U}}$ es el vector de residuos MC2E y donde \hat{u}_i es el i -ésimo elemento de este vector (el residuo MC2E para la observación i -ésima).

Los errores estándar MC2E son las raíces cuadradas de los elementos de la diagonal principal de $\hat{\Sigma}^{MC2E}$.

Propiedades de MC2E con errores homocedásticos

Si los errores son homocedásticos, entonces el estimador MC2E es asintóticamente eficiente entre la clase de los estimadores VI en los cuales los instrumentos son combinaciones lineales de las filas de \mathbf{Z} . Este resultado es el homólogo VI del teorema de Gauss-Markov y constituye una importante justificación para el uso de MC2E.

La distribución MC2E con homocedasticidad. Si los errores son homocedásticos, es decir, si $E(u_i^2 | \mathbf{Z}_i) = \sigma_u^2$, entonces $\mathbf{H} = E(\mathbf{Z}_i \mathbf{Z}_i' u_i^2) = E[E(\mathbf{Z}_i \mathbf{Z}_i' | \mathbf{Z}_i u_i^2)] = E[\mathbf{Z}_i \mathbf{Z}_i' E(u_i^2 | \mathbf{Z}_i)] = \mathbf{Q}_{ZZ} \sigma_u^2$. En este caso, la varianza de la distribución asintótica del estimador MC2E de la Ecuación (18.52) se simplifica a

$$\Sigma^{MC2E} = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}\sigma_u^2 \quad (\text{válido con homocedasticidad}) \quad (18.55)$$

El estimador válido con homocedasticidad de la matriz de varianzas MC2E es

$$\tilde{\Sigma}^{MC2E} = (\hat{\mathbf{Q}}_{XZ}\hat{\mathbf{Q}}_{ZZ}^{-1}\hat{\mathbf{Q}}_{ZX})^{-1}\hat{\sigma}_u^2, \text{ donde } \hat{\sigma}_u^2 = \frac{\hat{\mathbf{U}}' \hat{\mathbf{U}}}{n - k - r - 1} \quad (\text{válido con homocedasticidad}) \quad (18.56)$$

y los errores estándar MC2E válidos con homocedasticidad son las raíces cuadradas de los elementos de la diagonal principal de $\tilde{\Sigma}^{MC2E}$.

La clase de los estimadores VI que utilizan combinaciones lineales de Z. La clase de los estimadores VI que utilizan combinaciones lineales de \mathbf{Z} como instrumentos pueden ser generados de dos formas equivalentes.

La primera forma considera el problema de la estimación como una minimización de una función objetivo cuadrática, que al igual que el estimador MCO se obtiene minimizando la suma de los residuos al cuadrado. Bajo el supuesto de exogeneidad de los instrumentos, los errores $\mathbf{U} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ no están correlacionados con los regresores exógenos; es decir, evaluada en el verdadero valor de $\boldsymbol{\beta}$, la Ecuación (18.46) implica que

$$E[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{Z}] = 0 \quad (18.57)$$

La ecuación (18.57) consiste en un sistema de $m + r + 1$ ecuaciones que incluyen los $k + r + 1$ elementos desconocidos de β . Cuando $m > k$, estas ecuaciones son redundantes, en el sentido de que todas se cumplen en el verdadero valor de β . Cuando estos momentos poblacionales se reemplazan por sus momentos muestrales, el sistema de ecuaciones $(Y - Xb)'Z = 0$ puede resolverse para b cuando existe identificación exacta ($m = k$). Este valor de b es el estimador VI de β . Sin embargo, cuando existe sobreidentificación ($m > k$), el sistema de ecuaciones por lo general no puede cumplirse para el mismo valor de b , debido a la variación muestral —existen más ecuaciones que incógnitas— y en general este sistema no tiene solución.

Una forma de abordar el problema de la estimación de β cuando existe sobreidentificación es encontrar el equilibrio entre la voluntad de satisfacer cada ecuación con la minimización de una forma cuadrática que incluya a todas las ecuaciones. En concreto, sea A una matriz $(m + r + 1) \times (m + r + 1)$ simétrica semidefinida positiva de ponderaciones y sea $\hat{\beta}_A^{VI}$ el estimador que minimiza

$$\min_b (Y - Xb)'ZAZ'(Y - Xb) \quad (18.58)$$

La solución a este problema de minimización se obtiene tomando la derivada de la función objetivo con respecto a b , igualando el resultado a cero, y reordenando. Haciéndolo se obtiene $\hat{\beta}_A^{VI}$, el estimador VI basado en la matriz de ponderaciones A :

$$\hat{\beta}_A^{VI} = (X'ZAZ'X)^{-1}X'ZAZ'Y \quad (18.59)$$

La comparación de las Ecuaciones (18.59) y (18.48) muestra que MC2E es el estimador VI con $A = (Z'Z)^{-1}$. Es decir, MC2E es la solución del problema de minimización de la Ecuación (18.58) con $A = (Z'Z)^{-1}$.

Los cálculos con los que se obtienen las Ecuaciones (18.51) y (18.52), aplicados a $\hat{\beta}_A^{VI}$ muestran que

$$\sqrt{n}(\hat{\beta}_A^{VI} - \hat{\beta}) \xrightarrow{d} N(\mathbf{0}, \Sigma_A^{VI}), \text{ donde}$$

$$\Sigma_A^{VI} = (Q_{XZ}AQ_{ZX})^{-1}Q_{XZ}AHAQ_{ZX}(Q_{XZ}AQ_{ZX})^{-1} \quad (18.60)$$

La segunda manera de generar la clase de estimadores VI que utilizan combinaciones lineales de Z consiste en considerar los estimadores VI en los que los instrumentos son ZB , donde B es una matriz $(m + r + 1) \times (k + r + 1)$ con rango completo de filas. Entonces el sistema de $(k + r + 1)$ ecuaciones, $(Y - Xb)'ZB = 0$, se puede resolver únicamente para los $(k + r + 1)$ elementos desconocidos de b . Resolviendo estas ecuaciones para b se obtiene $\hat{\beta}^{VI} = (B'Z'X)^{-1}(B'Z'Y)$, y sustituyendo $B = AZ'X$ en esta expresión se obtiene la Ecuación (18.59). Por tanto, los dos métodos para la definición de los estimadores VI que son combinaciones lineales de los instrumentos dan lugar a la misma familia de estimadores VI. Es habitual trabajar con el primer método, en el que el estimador VI resuelve el problema cuadrático de minimización de la Ecuación (18.58), y que es el método adoptado aquí.

Eficiencia asintótica de MC2E con homocedasticidad. Si los errores son homocedásticos, entonces $H = Q_{ZZ}\sigma_u^2$ y la expresión para Σ_A^{VI} de la Ecuación (18.60) se convierte en

$$\Sigma_A^{VI} = (Q_{XZ}AQ_{ZX})^{-1}Q_{XZ}AQ_{ZZ}AQ_{ZX}(Q_{XZ}AQ_{ZX})^{-1}\sigma_u^2 \quad (18.61)$$

Para demostrar que MC2E es asintóticamente eficiente entre la clase de los estimadores que son combinaciones lineales de Z cuando los errores son homocedásticos, es necesario demostrar que, con homocedasticidad,

$$c' \Sigma_A^{VI} c \geq c' \Sigma^{MC2E} c \quad (18.62)$$

para todas las matrices A semidefinidas positivas y todos los vectores c $(k + r + 1) \times 1$, donde $\Sigma^{MC2E} = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}\sigma_u^2$ [Ecuación (18.55)]. La Ecuación (18.62), que se demuestra en el Apéndice 18.6, es el mismo criterio de eficiencia que se utiliza en el teorema de Gauss-Markov multivariante del Concepto clave 18.3. En consecuencia, MC2E es el estimador VI eficiente con homocedasticidad, entre la clase de estimadores en los que los instrumentos son combinaciones lineales de Z .

El estadístico J con homocedasticidad. El estadístico J (Concepto clave 12.6) contrasta la hipótesis nula de que todas las restricciones sobreidentificadas se cumplen frente a la alternativa de que todas o algunas de ellas no se cumplen.

La idea del estadístico J es que, si se cumplen las restricciones sobreidentificadas, u_i estará incorrelacionado con los instrumentos y, por lo tanto, una regresión de \mathbf{U} sobre \mathbf{Z} tendrá unos coeficientes de regresión poblacionales que serán todos iguales a cero. En la práctica, \mathbf{U} no es observable, pero puede ser estimado mediante los residuos MC2E $\hat{\mathbf{U}}$, por lo que una regresión de $\hat{\mathbf{U}}$ sobre \mathbf{Z} debería dar lugar a coeficientes estadísticamente no significativos. En consecuencia, el estadístico J MC2E es el estadístico F válido con homocedasticidad para el contraste de la hipótesis de que los coeficientes de \mathbf{Z} son todos ellos iguales a cero, en la regresión de $\hat{\mathbf{U}}$ sobre \mathbf{Z} , multiplicado por $(m + r + 1)$ por lo que el estadístico F se encuentra en su forma asintótica chi-cuadrado.

Se puede obtener una fórmula explícita para el estadístico J mediante la Ecuación (7.13) para el estadístico F válido con homocedasticidad. La regresión sin restricciones es la regresión de $\hat{\mathbf{U}}$ sobre los $m + r + 1$ regresores \mathbf{Z} , y la regresión restringida no tiene regresores. Por tanto, en la notación de la Ecuación (7.13), $SR_{sinrestringir} = \hat{\mathbf{U}}' \mathbf{M}_Z \hat{\mathbf{U}}$ y $SR_{restringida} = \hat{\mathbf{U}}' \hat{\mathbf{U}}$, por lo que $SR_{restringida} - SR_{sinrestringir} = \hat{\mathbf{U}}' \hat{\mathbf{U}} - \hat{\mathbf{U}}' \mathbf{M}_Z \hat{\mathbf{U}} = \hat{\mathbf{U}}' \mathbf{P}_Z \hat{\mathbf{U}}$ y el estadístico J es

$$J = \frac{\hat{\mathbf{U}}' \mathbf{P}_Z \hat{\mathbf{U}}}{\hat{\mathbf{U}}' \mathbf{M}_Z \hat{\mathbf{U}} / (n - m - r - 1)} \quad (18.63)$$

El método para el cálculo del estadístico J , que se describe en el Concepto clave 12.6, implica contrastar solamente la hipótesis de que los coeficientes de los instrumentos excluidos son iguales a cero. A pesar de que estos dos métodos siguen diferentes pasos para su cálculo, dan lugar a idénticos estadísticos J (Ejercicio 18.14).

Se demuestra en el Apéndice 18.6 que, bajo la hipótesis nula de que $E(u_i Z_i) = 0$,

$$J \xrightarrow{d} \chi^2_{m-k} \quad (18.64)$$

Estimación por el método generalizado de momentos en modelos lineales

Si los errores son heterocedásticos, entonces el estimador MC2E ya no es eficiente entre la clase de los estimadores VI que utilizan combinaciones lineales de \mathbf{Z} como instrumentos. El estimador eficiente en este caso se conoce como el estimador eficiente del método generalizado de momentos (MGM). Además, si los errores son heterocedásticos, entonces el estadístico J , tal y como se define en la Ecuación (18.63), ya no tiene una distribución chi-cuadrado. Sin embargo, una formulación alternativa del estadístico J , construida con el estimador eficiente MGM, tiene una distribución chi-cuadrado con $m - k$ grados de libertad.

Estos resultados van en paralelo a los resultados de la estimación del modelo de regresión habitual con regresores exógenos y errores heterocedásticos: si los errores son heterocedásticos, entonces el estimador MCO no es eficiente entre los estimadores que son lineales en \mathbf{Y} (las condiciones de Gauss-Markov no se cumplen) y el estadístico F válido con homocedasticidad ya no tiene una distribución F , incluso en grandes muestras. En el modelo de regresión con regresores exógenos y heterocedasticidad, el estimador eficiente es el de mínimos cuadrados ponderados; en el modelo de regresión VI con heterocedasticidad, el estimador eficiente utiliza una matriz de ponderación diferente a la de MC2E, y el estimador resultante es el estimador de MGM eficiente.

Estimación MGM. La estimación por el **método generalizado de momentos (MGM)** es un método general para la estimación de los parámetros de los modelos lineales o no lineales, en los que los parámetros se eligen de forma que ofrezcan el mejor ajuste para varias ecuaciones, cada una de las cuales iguala un momento muestral a cero. Estas ecuaciones, que en el contexto de MGM se denominan condiciones de momentos, por lo general no pueden todas ellas cumplirse al mismo tiempo. El estimador MGM busca equilibrar la voluntad de satisfacer cada una de las ecuaciones con la minimización de una función objetivo cuadrática.

En el modelo de regresión lineal VI con variables exógenas \mathbf{Z} , la clase de los estimadores MGM está compuesta por todos los estimadores que son soluciones del problema cuadrático de minimización de la Ecuación (18.58). Por lo tanto, la clase de los estimadores de MGM basada en el conjunto completo de los instrumentos \mathbf{Z} con diferentes matrices de ponderaciones \mathbf{A} es la misma que la clase de los estimadores de VI, en la que los instrumentos son combinaciones lineales de \mathbf{Z} . En el modelo de regresión lineal VI, MGM es solamente otro nombre para la clase de los estimadores que ya hemos estado estudiando, es decir, los estimadores que resuelven la Ecuación (18.58).

El estimador MGM asintóticamente eficiente. Entre la clase de los estimadores MGM, el estimador **MGM eficiente** es el estimador de MGM con la matriz de varianzas asintótica más pequeña [donde la menor matriz de varianzas está definida como en la Ecuación (18.62)]. Por tanto, el resultado de la Ecuación (18.62) puede ser reformulado diciendo que MC2E es el estimador MGM eficiente en el modelo de regresión lineal cuando los errores son homocedásticos.

Para motivar la expresión del estimador MGM eficiente cuando los errores son heterocedásticos, recordemos que cuando los errores son homocedásticos, \mathbf{H} [la matriz de varianzas de $\mathbf{Z}_i u_i$; véase la Ecuación (18.50)] es igual a $\mathbf{Q}_{\mathbf{Z}\mathbf{Z}} \sigma_u^2$, y la matriz de ponderaciones asintóticamente eficiente se obtiene haciendo $\mathbf{A} = (\mathbf{Z}'\mathbf{Z})^{-1}$, lo cual da lugar al estimador MC2E. En muestras grandes, utilizar la matriz de ponderaciones $\mathbf{A} = (\mathbf{Z}'\mathbf{Z})^{-1}$ es equivalente a utilizar $\mathbf{A} = (\mathbf{Q}_{\mathbf{Z}\mathbf{Z}} \sigma_u^2)^{-1} = \mathbf{H}^{-1}$. Esta interpretación del estimador MC2E sugiere que, por analogía, el estimador VI eficiente con heterocedasticidad puede obtenerse haciendo $\mathbf{A} = \mathbf{H}^{-1}$ y resolviendo

$$\min_b (\mathbf{Y} - \mathbf{X}\mathbf{b})' \mathbf{Z} \mathbf{H}^{-1} \mathbf{Z}' (\mathbf{Y} - \mathbf{X}\mathbf{b}) \quad (18.65)$$

Esta analogía es correcta: la solución al problema de minimización de la Ecuación (18.65) es el estimador MGM eficiente. Sea $\tilde{\boldsymbol{\beta}}^{MGM.Ef}$ la solución al problema de minimización de la Ecuación (18.65). Por la Ecuación (18.59) este estimador es

$$\tilde{\boldsymbol{\beta}}^{MGM.Ef} = (\mathbf{X}' \mathbf{Z} \mathbf{H}^{-1} \mathbf{Z}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \mathbf{H}^{-1} \mathbf{Z}' \mathbf{Y} \quad (18.66)$$

La distribución asintótica de $\tilde{\boldsymbol{\beta}}^{MGM.Ef}$ se obtiene sustituyendo $\mathbf{A} = \mathbf{H}^{-1}$ en la Ecuación (18.60) y simplificando; por lo que

$$\begin{aligned} \sqrt{n} (\tilde{\boldsymbol{\beta}}^{MGM.Ef} - \boldsymbol{\beta}) &\xrightarrow{d} N(\mathbf{0}, \Sigma^{MGM.Ef}) \\ \text{donde } \Sigma^{MGM.Ef} &= (\mathbf{Q}_{\mathbf{X}\mathbf{Z}} \mathbf{H}^{-1} \mathbf{Q}_{\mathbf{Z}\mathbf{X}})^{-1} \end{aligned} \quad (18.67)$$

El resultado de que $\tilde{\boldsymbol{\beta}}^{MGM.Ef}$ es el estimador MGM eficiente se demuestra mediante la demostración de que $\mathbf{c}' \Sigma_A^{VI} \mathbf{c} \geq \mathbf{c}' \Sigma^{MGM.Ef} \mathbf{c}$ para todos los vectores \mathbf{c} , donde Σ_A^{VI} viene dado por la Ecuación (18.60). La prueba de este resultado se ofrece en el Apéndice 18.6.

Estimación MGM eficiente factible. El estimador MGM definido en la Ecuación (18.66) no es un estimador factible, ya que depende de la matriz de varianzas desconocida \mathbf{H} . Sin embargo, se puede calcular un estimador de MGM eficiente factible sustituyendo \mathbf{H} por un estimador consistente de \mathbf{H} en el problema de minimización de la Ecuación (18.65) o, de manera equivalente, introduciendo un estimador consistente de \mathbf{H} en la fórmula de $\tilde{\boldsymbol{\beta}}^{MGM.Ef}$ de la Ecuación (18.66).

El estimador MGM eficiente puede calcularse en dos etapas. En la primera etapa, se estima $\boldsymbol{\beta}$ utilizando cualquier estimador consistente. Se utiliza este estimador de $\boldsymbol{\beta}$ para calcular los residuos de la ecuación de interés, y posteriormente se utilizan estos residuos para calcular un estimador de \mathbf{H} . En la segunda etapa, se utiliza el estimador de \mathbf{H} para estimar la matriz de ponderaciones óptima \mathbf{H}^{-1} y para calcular el estimador MGM eficiente. Para concretar, en el modelo de regresión lineal VI, es natural utilizar el estimador MC2E en la primera etapa y utilizar los residuos MC2E para estimar \mathbf{H} . Si se utiliza MC2E en la primera etapa, entonces el estimador MGM eficiente factible calculado en la segunda etapa es

$$\hat{\boldsymbol{\beta}}^{MGM.Ef} = (\mathbf{X}' \mathbf{Z} \hat{\mathbf{H}}^{-1} \mathbf{Z}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \hat{\mathbf{H}}^{-1} \mathbf{Z}' \mathbf{Y} \quad (18.68)$$

donde $\hat{\mathbf{H}}$ está dado en la Ecuación (18.54).

Debido a que $\hat{\mathbf{H}} \xrightarrow{p} \mathbf{H}$, $\sqrt{n}(\hat{\boldsymbol{\beta}}^{MGM.Ef} - \tilde{\boldsymbol{\beta}}^{MGM.Ef}) \xrightarrow{p} 0$ (Ejercicio 18.12), y

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{MGM.Ef} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}^{MGM.Ef}) \quad (18.69)$$

donde $\boldsymbol{\Sigma}^{MGM.Ef} = (Q_{XZ}\mathbf{H}^{-1}Q_{ZX})^{-1}$ [Ecuación (18.67)]. Es decir, el estimador en dos etapas $\hat{\boldsymbol{\beta}}^{MGM.Ef}$ factible de la Ecuación (18.68) es, asintóticamente, el estimador MGM eficiente.

El estadístico J heterocedástico-robusto. El estadístico J heterocedástico-robusto, conocido asimismo como el **estadístico J MGM**, es el homólogo del estadístico J basado en MC2E, calculado mediante el estimador MGM eficiente y la función de ponderaciones. Es decir, el estadístico J MGM está dado por

$$J^{MGM} = (\mathbf{Z}'\hat{\mathbf{U}}^{MGM})'\hat{\mathbf{H}}^{-1}(\mathbf{Z}'\hat{\mathbf{U}}^{MGM})/n \quad (18.70)$$

donde $\hat{\mathbf{U}}^{MGM} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{MGM.Ef}$ son los residuos de la ecuación de interés que se estiman mediante MGM eficientes (factibles), y $\hat{\mathbf{H}}^{-1}$ es la matriz de ponderaciones utilizada para calcular $\hat{\boldsymbol{\beta}}^{MGM.Ef}$.

Bajo la hipótesis nula $E(\mathbf{Z}_i u_i) = \mathbf{0}$, $J^{MGM} \xrightarrow{d} \chi^2_{m-k}$ (véase el Apéndice 18.6).

MGM con datos de series temporales. Los resultados de esta sección se han obtenido bajo los supuestos de la regresión VI para datos de sección cruzada. Sin embargo, en muchas aplicaciones estos resultados se amplían para casos de regresión de series temporales VI y MGM. A pesar de que el tratamiento matemático formal de MGM con datos de series temporales queda fuera del alcance de este libro (para un tratamiento de ese tipo, consúltese Hayashi, 2000, Capítulo 6), no obstante se resumen las ideas clave de la estimación MGM con datos de series temporales. Este resumen supone cierta soltura en el manejo del material de los Capítulos 14 y 15. Se supone para este análisis, que las variables son estacionarias.

Resulta útil distinguir entre dos tipos de aplicaciones: aplicaciones en las que el término de error u_t está serialmente correlacionado y aplicaciones en las que u_t está serialmente incorrelacionado. Si el término de error u_t está serialmente correlacionado, entonces la distribución asintótica del estimador MGM sigue teniendo una distribución normal, pero la fórmula para \mathbf{H} de la Ecuación (18.50) ya no es correcta. En su lugar, la expresión correcta de \mathbf{H} depende de las autocovarianzas de $Z_t u_t$ y es análoga a la fórmula dada en la Ecuación (15.14) para la varianza del estimador MCO cuando el término de error está serialmente correlacionado. El estimador MGM eficiente puede continuar construyéndose mediante un estimador consistente de \mathbf{H} ; sin embargo, este estimador consistente debe ser calculado utilizando los métodos HAC estudiados en el Capítulo 15.

Si el término de error u_t no está serialmente correlacionado, entonces la estimación HAC de \mathbf{H} resulta innecesaria y todas las fórmulas presentadas en esta sección son extensibles a las aplicaciones MGM con series temporales. En las aplicaciones modernas a las finanzas y la macroeconomía, es habitual encontrar modelos en los que el término de error representa una perturbación inesperada o impredecible, en cuyo caso el modelo implica que u_t está serialmente incorrelacionado. Por ejemplo, consideremos un modelo con una única variable endógena incluida y sin variables exógenas por lo que la ecuación de interés es $Y_t = \beta_0 + \beta_1 X_t + u_t$. Supongamos que una teoría económica implica que u_t es impredecible dada la información pasada. Por tanto la teoría implica la condición sobre los momentos

$$E(u_t | Y_{t-1}, X_{t-1}, Z_{t-1}, Y_{t-2}, X_{t-2}, Z_{t-2}, \dots) = 0 \quad (18.71)$$

donde Z_{t-1} es el valor retardado de alguna otra variable. La condición sobre los momentos de la Ecuación (18.71) implica que todas las variables retardadas $Y_{t-1}, X_{t-1}, Z_{t-1}, Y_{t-2}, X_{t-2}, Z_{t-2}, \dots$ son candidatas a ser instrumentos válidos (cumplen la condición de exogeneidad). Además, debido a que $u_{t-1} = Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$, la condición de momentos de la Ecuación (18.71) es equivalente a $E(u_t | u_{t-1}, X_{t-1}, Z_{t-1}, u_{t-2}, X_{t-2}, Z_{t-2}, \dots) = 0$. Debido a que u_t está serialmente incorrelacionado, la estimación HAC de \mathbf{H} no es necesaria. Por tanto la teoría de MGM presentada en esta sección, incluyendo la estimación MGM eficiente y el estadístico J MGM, es directamente aplicable a las aplicaciones de series temporales con condiciones de momentos de la forma de la Ecuación (18.71), bajo la hipótesis de que la condición sobre los momentos de la Ecuación (18.71) es, en realidad, correcta.

Resumen

1. El modelo de regresión lineal múltiple en forma matricial es $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$, donde \mathbf{Y} es el vector de observaciones $n \times 1$ sobre la variable dependiente, \mathbf{X} es la matriz $n \times (k+1)$ de n observaciones sobre los $k+1$ regresores (incluyendo una constante), $\boldsymbol{\beta}$ es el vector $k+1$ de parámetros desconocidos, y \mathbf{U} es el vector $n \times 1$ de los términos de error.
2. El estimador MCO es $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Bajo los cuatro primeros supuestos de mínimos cuadrados del Concepto clave 18.1, $\hat{\boldsymbol{\beta}}$ es consistente y está distribuido asintóticamente normal. Si además los errores son homocedásticos, entonces la varianza condicional de $\hat{\boldsymbol{\beta}}$ es $\text{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$.
3. En general las restricciones lineales sobre $\boldsymbol{\beta}$ se pueden escribir como las q ecuaciones $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, y puede utilizarse esta formulación para contrastar la hipótesis conjunta que incluye a varios coeficientes o para construir conjuntos de confianza para los elementos de $\boldsymbol{\beta}$.
4. Cuando los errores de regresión son i.i.d. y tienen una distribución normal, condicionada a \mathbf{X} , $\boldsymbol{\beta}$ tiene una distribución exacta normal y los estadísticos t y F válidos con homocedasticidad, tienen distribuciones exactas t_{t-k-1} y $F_{q,n-k-1}$ respectivamente.
5. El teorema de Gauss-Markov establece que, si los errores son homocedásticos y están condicionalmente incorrelacionados entre observaciones y si $E(u_i|\mathbf{X}) = 0$, el estimador MCO es eficiente entre los estimadores lineales condicionalmente insesgados (MCO es ELIO).
6. Si la matriz de covarianzas del error $\boldsymbol{\Omega}$ no es proporcional a la matriz identidad, y si $\boldsymbol{\Omega}$ es conocida o puede estimarse, entonces el estimador MCG es asintóticamente más eficiente que MCO. Sin embargo, MCG requiere que, en general, u_i esté incorrelacionado con *todas* las observaciones de las variables explicativas, no solo con \mathbf{X}_i como requiere MCO, un supuesto que debe ser evaluado cuidadosamente en las aplicaciones.
7. El estimador MC2E es un miembro de la clase de estimadores MGM del modelo lineal. En MGM, los coeficientes se estiman haciendo la covarianza muestral entre el error de regresión y las variables exógenas tan pequeña como sea posible —en concreto, resolviendo el problema $\min_b [(\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{Z}] \mathbf{A}[\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\mathbf{b})]$, donde \mathbf{A} es una matriz de ponderaciones. El estimador MGM asintóticamente eficiente hace $\mathbf{A} = [E(\mathbf{Z}_i\mathbf{Z}_i'u_i^2)]^{-1}$. Cuando los errores son homocedásticos, el estimador MGM asintóticamente eficiente del modelo de regresión VI lineal es MC2E.

Términos clave

condiciones de Gauss-Markov para regresión múltiple (513)	método generalizado de momentos (MGM) (522)
teorema de Gauss-Markov para regresión múltiple (513)	MGM eficientes (523)
mínimos cuadrados generalizados (MCG) (515)	estadístico J heterocedástico-robusto (524)
MCG infactibles (517)	estadístico J MGM (524)
MCG factibles (517)	vector de medias (532)
	matriz de covarianzas (532)

Revisión de conceptos

- 18.1** Un investigador que estudia la relación entre los ingresos salariales y el género para un grupo de trabajadores especifica el modelo de regresión, $Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + u_i$, donde X_{1i} es una variable binaria que es igual a 1 si la persona i -ésima es una mujer y X_{2i} es una variable binaria que es igual a 1 si la i -ésima persona es un varón. Escriba el modelo en la forma matricial de la Ecuación (18.2) para un conjunto hipotético de $n = 5$ observaciones. Demuestre que las columnas de \mathbf{X} son linealmente dependientes, por lo que \mathbf{X} no tiene rango completo. Explique cómo se reespecificaría el modelo para eliminar la multicolinealidad perfecta.

- 18.2** Se está analizando un modelo de regresión lineal con 500 observaciones y un único regresor. Explique cómo se construiría un intervalo de confianza para β_1 si:
- Los Supuestos #1 al #4 del Concepto clave 18.1 son ciertos, pero se cree que el Supuesto #5 o el #6 podrían no ser ciertos.
 - Los Supuestos #1 al #5 son ciertos, pero se cree que el Supuesto #6 podría no ser cierto (proporcione dos modos de construir el intervalo de confianza).
 - Los Supuestos #1 al #6 son ciertos.
- 18.3** Supóngase que los Supuestos #1 al #5 del Concepto clave 18.1 son ciertos, pero que el Supuesto #6 no lo es. ¿Se cumple el resultado de la Ecuación (18.31)? Explíquelo.
- 18.4** ¿Se puede calcular el estimador ELIO de β si se cumple la Ecuación (18.41) y Ω es desconocida? ¿Qué ocurre si Ω es conocida?
- 18.5** Diseñe un ejemplo de un modelo de regresión que satisfaga el supuesto $E(u_i|X_i) = 0$, pero en el que $E(U|X) \neq \mathbf{0}_n$.

Ejercicios

- 18.1** Considérese la regresión poblacional de las calificaciones en los exámenes sobre la renta y sobre el cuadrado de la renta de la Ecuación (8.1).
- Escriba la regresión de la Ecuación (8.1) en la forma matricial de la Ecuación (18.5). Defina \mathbf{Y} , \mathbf{X} , \mathbf{U} , y β .
 - Explique cómo contrastar la hipótesis nula de que la relación entre las calificaciones en las pruebas y la renta es lineal frente a la alternativa de que es de segundo grado. Exprese la hipótesis nula en la forma de la Ecuación (18.20). ¿Cómo son \mathbf{R} , \mathbf{r} , y q ?
- 18.2** Supóngase que una muestra de $n = 20$ hogares presenta las medias y covarianzas muestrales que aparecen a continuación para una variable dependiente y dos regresores:

	Medias muestrales	Covarianzas muestrales		
		\mathbf{Y}	\mathbf{X}	\mathbf{X}_2
\mathbf{Y}	6,39	0,26	0,22	0,32
\mathbf{X}_1	7,24		0,80	0,28
\mathbf{X}_2	4,00			2,40

- Calcule los estimadores MCO $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$. Calcule s_u^2 . Calcule el R^2 de la regresión.
 - Supóngase que se cumplen los seis supuestos del Concepto clave 18.1. Contraste la hipótesis de que $\beta_1 = 0$ al nivel de significación del 5%.
- 18.3** Sea \mathbf{W} un vector $m \times 1$ con matriz de covarianzas Σ_W , donde Σ_W es finita y definida positiva. Sea \mathbf{c} un vector $m \times 1$ no aleatorio, y sea $Q = \mathbf{c}'\mathbf{W}$.
- Demuestre que $\text{var}(Q) = \mathbf{c}'\Sigma_W\mathbf{c}$.
 - Supóngase que $\mathbf{c} \neq \mathbf{0}_m$. Demuestre que $0 < \text{var}(Q) < \infty$.
- 18.4** Considérese el modelo de regresión del Capítulo 4, $Y_i = \beta_0 + \beta_1 X_i + u_i$, y supóngase que se cumplen los supuestos del Concepto clave 4.3.
- Escriba el modelo en forma matricial dado por las Ecuaciones (18.2) y (18.4).
 - Demuestre que se satisfacen los Supuestos #1 a #4 del Concepto clave 18.1.
 - Utilice la fórmula general para $\hat{\beta}$ de la Ecuación (18.11) para obtener las expresiones de $\hat{\beta}_0$ y $\hat{\beta}_1$ dadas en el Concepto clave 4.2.

- d)** Demuestre que el elemento (1,1) de la matriz $\Sigma_{\hat{\beta}}$ en la Ecuación (18.13) es igual a la expresión para $\sigma_{\hat{\beta}_0}^2$ dada en el Concepto clave 4.4.
- 18.5** Sean P_X y M_X como las que se definen en las Ecuaciones (18.24) y (18.25).
- Demuestre que $P_X M_X = \mathbf{0}_{n \times m}$ y que P_X y M_X son idempotentes.
 - Obtenga las Ecuaciones (18.27) y (18.28).
- 18.6** Considerése el modelo de regresión en forma matricial, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\gamma + \mathbf{U}$, donde \mathbf{X} es una matriz $n \times k_1$ de regresores y \mathbf{W} es una matriz $n \times k_2$ de regresores. Entonces, tal y como se muestra en el Ejercicio 18.17, el estimador MCO, $\hat{\boldsymbol{\beta}}$ se puede expresar mediante
- $$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} (\mathbf{X}' \mathbf{M}_W \mathbf{Y})$$
- Sea ahora $\hat{\beta}_1^{BV}$ el estimador de efectos fijos de «variable binaria» calculado mediante la estimación de la Ecuación (10.11) por MCO y sea $\hat{\beta}_1^{DM}$ el estimador de efectos fijos «en desviaciones respecto de la media» calculado mediante la estimación de la Ecuación (10.14) por MCO, en la que se ha restado a X y a Y las medias muestrales específicas de cada entidad individual. Utilice la expresión de $\hat{\boldsymbol{\beta}}$ dada anteriormente para probar que $\hat{\beta}_1^{BV} = \hat{\beta}_1^{DM}$. [Sugerencia: escriba la Ecuación (10.11) utilizando un conjunto completo de efectos fijos, $D1_i, D2_i, \dots, Dn_i$ y sin término constante. Incluya todos los efectos fijos en \mathbf{W} . Escriba la matriz $\mathbf{M}_W \mathbf{X}$].
- 18.7** Considérese el modelo de regresión $Y_i = \beta_1 X_i + \beta_2 W_i + u_i$, donde por simplicidad se ha omitido el término independiente y se ha supuesto que todas las variables tienen una media igual a cero. Supóngase que X_i se distribuye de forma independiente de (W_i, u_i) , pero W_i y u_i podrían estar correlacionadas y sean $\hat{\beta}_1$ y $\hat{\beta}_2$ los estimadores MCO para este modelo. Demuestre que
- Tanto si W_i y u_i están correlacionados como si no lo están $\hat{\beta}_1 \xrightarrow{p} \beta_1$.
 - Si W_i y u_i están correlacionados, entonces $\hat{\beta}_2$ es inconsistente.
 - Sea $\hat{\beta}_1^r$ el estimador MCO de la regresión de Y sobre X (la regresión restringida que excluye a W). Proporcione las condiciones bajo las cuales $\hat{\beta}_1$ tiene una menor varianza asintótica que $\hat{\beta}_1^r$, teniendo en cuenta la posibilidad de que W_i y u_i estén correlacionados.
- 18.8** Considérese el modelo de regresión $Y_i = \beta_0 + \beta_1 X_i + u_i$, donde $u_i = \tilde{u}_1 + \dots + \tilde{u}_{i-1} + \tilde{u}_i$ para $i = 2, 3, \dots, n$. Supóngase que \tilde{u}_i son i.i.d. con media igual a 0 y varianza igual a 1 y están distribuidos de forma independiente de X_j para todo j .
- Obtenga una expresión para $E(\mathbf{U}\mathbf{U}') = \Omega$.
 - Explique la forma de estimar el modelo mediante MCG, sin invertir de forma explícita la matriz Ω . (Sugerencia: transformar el modelo para que los errores de regresión sean $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n$).
- 18.9** Este ejercicio muestra que el estimador MCO de un subconjunto de coeficientes de regresión es consistente bajo el supuesto de independencia en media condicional enunciado en el Apéndice 7.2. Considérese el modelo de regresión múltiple en forma matricial $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\gamma + \mathbf{U}$, donde \mathbf{X} y \mathbf{W} son, respectivamente, matrices de regresores $n \times k_1$ y $n \times k_2$. Sean \mathbf{X}'_i y \mathbf{W}'_i las expresiones de las filas i -ésimas de \mathbf{X} y \mathbf{W} [como en la Ecuación (18.3)]. Supóngase que (i) $E(u_i | \mathbf{X}_i, \mathbf{W}_i) = \mathbf{W}'_i \boldsymbol{\delta}$, donde $\boldsymbol{\delta}$ es un vector de parámetros desconocidos $k_2 \times 1$; (ii) $(\mathbf{X}_i, \mathbf{W}_i, Y_i)$ son i.i.d.; (iii) $(\mathbf{X}_i, \mathbf{W}_i, u_i)$ tienen momentos de cuarto orden finitos y distintos de cero; y (iv) no existe multicolinealidad perfecta. Estos son los supuestos #1-#4 del Concepto clave 18.1, con el supuesto de independencia en media condicional (i) que sustituye al supuesto habitual de media condicional igual a cero.
- Utilice la expresión para $\hat{\boldsymbol{\beta}}$ dada en el Ejercicio 18.6 para escribir $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (n^{-1} \mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} (n^{-1} \mathbf{X}' \mathbf{M}_W \mathbf{U})$.
 - Demuestre que $n^{-1} \mathbf{X}' \mathbf{M}_W \mathbf{X} \xrightarrow{p} \Sigma_{XX} - \Sigma_{XW} \Sigma_{WW}^{-1} \Sigma_{WX}$, donde $\Sigma_{XX} = E(\mathbf{X}'_i \mathbf{X}_i)$, $\Sigma_{XW} = E(\mathbf{X}'_i \mathbf{W}_i)$, etc. [La matriz $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ si $\mathbf{A}_{n,ij} \xrightarrow{p} \mathbf{A}_{ij}$ para todo i, j , donde $\mathbf{A}_{n,ij}$ y \mathbf{A}_{ij} son los elementos (i,j) de \mathbf{A}_n y \mathbf{A}].
 - Demuestre que los supuestos (i) y (ii) implican que $E(\mathbf{U} | \mathbf{X}, \mathbf{W}) = \mathbf{W}\boldsymbol{\delta}$.
 - Utilice el apartado (c) y la ley de esperanzas iteradas para demostrar que $n^{-1} \mathbf{X}' \mathbf{M}_W \mathbf{U} \xrightarrow{p} \mathbf{0}_{k_1 \times 1}$.
 - Utilice los apartados (a) a (d) para concluir que, bajo las condiciones (i) hasta (iv), $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$.

18.10 Sea \mathbf{C} una matriz simétrica e idempotente.

- a) Demuestre que los valores propios de \mathbf{C} son 0 ó 1. (*Sugerencia:* téngase en cuenta que $\mathbf{C}\mathbf{q} = \gamma\mathbf{q}$ implica que $0 = \mathbf{C}\mathbf{q} - \gamma\mathbf{q} = \mathbf{C}\mathbf{q} - \gamma\mathbf{q} = \gamma\mathbf{C}\mathbf{q} - \gamma\mathbf{q} = \gamma^2\mathbf{q} - \gamma\mathbf{q}$ y resolver para γ).
- b) Demuestre que $\text{traza}(\mathbf{C}) = \text{rango}(\mathbf{C})$.
- c) Sea \mathbf{d} un vector $n \times 1$. Demuestre que $\mathbf{d}'\mathbf{C}\mathbf{d} \geq 0$.

18.11 Supóngase que \mathbf{C} es una matriz $n \times n$ simétrica e idempotente con rango r y sea $\mathbf{V} \sim N(\mathbf{0}_n, \mathbf{I}_n)$.

- a) Demuestre que $\mathbf{C} = \mathbf{A}\mathbf{A}'$, donde \mathbf{A} es $n \times r$ con $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$. (*Sugerencia:* \mathbf{C} es semidefinida positiva y puede expresarse como $\mathbf{Q}\Lambda\mathbf{Q}'$, tal y como se explica en el Apéndice 18.1).
- b) Demuestre que $\mathbf{A}'\mathbf{V} \sim N(\mathbf{0}_r, \mathbf{I}_r)$.
- c) Demuestre que $\mathbf{V}'\mathbf{C}\mathbf{V} \sim \chi_r^2$.

18.12 a) Demuestre que $\hat{\beta}^{MGM.Ef}$ es el estimador de MGM eficiente —es decir, que $\hat{\beta}^{MGM.Ef}$ en la Ecuación (18.66) es la solución a la Ecuación (18.65).

- b) Demuestre que $\sqrt{n}(\hat{\beta}^{MGM.Ef} - \tilde{\beta}^{MGM.Ef}) \xrightarrow{P} 0$.
- c) Demuestre que $J^{MGM} \xrightarrow{d} \chi_{m-k}^2$.

18.13 Considérese el problema de minimización de la suma de los cuadrados de los residuos sujeto a la restricción de que $\mathbf{R}\mathbf{b} = \mathbf{r}$, donde \mathbf{R} es $q \times (k+1)$ con rango q . Sea $\tilde{\beta}$ el valor de \mathbf{b} que resuelve el problema de minimización restringida.

- a) Demuestre que el Lagrangiano para el problema de minimización es $L(\mathbf{b}, \boldsymbol{\gamma}) = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) + \boldsymbol{\gamma}'(\mathbf{R}\mathbf{b} - \mathbf{r})$, donde $\boldsymbol{\gamma}$ es un vector $q \times 1$ de multiplicadores de Lagrange.
- b) Demuestre que $\tilde{\beta} = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})$.
- c) Demuestre que $(\mathbf{Y} - \mathbf{X}\tilde{\beta})'(\mathbf{Y} - \mathbf{X}\tilde{\beta}) - (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = (\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})$.
- d) Demuestre que \tilde{F} en la Ecuación (18.36) es equivalente al estadístico F válido con homocedasticidad de la Ecuación (7.13).

18.14 Considérese el modelo de regresión $\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$. Se partitiona \mathbf{X} como $[\mathbf{X}_1 \ \mathbf{X}_2]$ y β como $[\beta'_1 \ \beta'_2]$, donde \mathbf{X}_1 tiene k_1 columnas y \mathbf{X}_2 tiene k_2 columnas. Supóngase que $\mathbf{X}_2'\mathbf{Y} = \mathbf{0}_{k_2 \times 1}$. Sea $\mathbf{R} = [\mathbf{I}_{k_1} \ \mathbf{0}_{k_1 \times k_2}]$.

- a) Demuestre que $\hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta} = (\mathbf{R}\hat{\beta})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta})$.
- b) Considérese la regresión que se describe en la Ecuación (12.17). Sea $\mathbf{W} = [\mathbf{1} \ \mathbf{W}_1 \ \mathbf{W}_2 \ \dots \ \mathbf{W}_r]$, donde $\mathbf{1}$ es un vector $n \times 1$ de unos, \mathbf{W}_1 es el vector $n \times 1$ cuyo i -ésimo elemento es \mathbf{W}_{1i} , etc. Sea $\hat{\mathbf{U}}^{MC2E}$ el vector de residuos de mínimos cuadrados en dos etapas.
 - i) Demuestre que $\mathbf{W}'\hat{\mathbf{U}}^{MC2E} = 0$.
 - ii) Demuestre que el método para calcular el estadístico J que se describe en el Concepto clave 12.6 (mediante un estadístico F válido con homocedasticidad) y la fórmula de la Ecuación (18.63) dan lugar al mismo valor para el estadístico J . [*Sugerencia:* utilice los resultados de los apartados (a), (b, i), y del Ejercicio 18.13].

18.15 (Consistencia de los errores estándar agrupados). Considérese el modelo de datos de panel $Y_{it} = \beta X_{it} + \alpha_i + u_{it}$, en el que todas las variables son escalares. Supóngase que se cumplen los Supuestos #1, #2, y #4 del Concepto clave 10.3 y el Supuesto #3 reforzado de manera que X_{it} y u_{it} tienen ocho momentos finitos distintos de cero. Sea $\mathbf{M} = \mathbf{I}_T - T^{-1}\mathbf{1}\mathbf{1}'$, donde $\mathbf{1}$ es un vector $T \times 1$ de unos. Sea asimismo $\mathbf{Y}_i = (Y_{i1} \ Y_{i2} \ \dots \ Y_{iT})'$, $\mathbf{X}_i = (X_{i1} \ X_{i2} \ \dots \ X_{iT})'$, $\mathbf{u}_i = (u_{i1} \ u_{i2} \ \dots \ u_{iT})'$, $\tilde{\mathbf{Y}}_i = \mathbf{M}\mathbf{Y}_i$, $\tilde{\mathbf{X}}_i = \mathbf{M}\mathbf{X}_i$, y $\tilde{\mathbf{u}}_i = \mathbf{M}\mathbf{u}_i$. En este problema, para los cálculos asintóticos, supóngase que T está fijado y que $n \rightarrow \infty$.

- a) Demuestre que el estimador de efectos fijos de β en la Sección 10.3 se puede escribir como $\hat{\beta} = (\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{Y}}_i$.
- b) Demuestre que $\hat{\beta} - \beta = (\sum_{i=1}^n \tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}\sum_{i=1}^n \tilde{\mathbf{X}}_i'\mathbf{u}_i$. (*Pista:* \mathbf{M} es idempotente).
- c) Sean $Q_{\tilde{\mathbf{X}}} = T^{-1}E(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)$ y $Q_{\tilde{\mathbf{X}}} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^2$. Demuestre que $\hat{Q}_{\tilde{\mathbf{X}}} \xrightarrow{P} Q_{\tilde{\mathbf{X}}}$.
- d) Sea $\eta_i = \tilde{\mathbf{X}}_i'\mathbf{u}_i/\sqrt{T}$ y $\sigma_\eta^2 = \text{var}(\eta_i)$. Demuestre que $\sqrt{\frac{1}{n}} \sum_{i=1}^n \eta_i \xrightarrow{d} N(0, \sigma_\eta^2)$.

- e) Utilice las respuestas a los apartados (b) hasta (d) para demostrar la Ecuación (10.25); es decir, demuestre que $\sqrt{nT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_{\eta}^2/Q_{\tilde{X}}^2)$.
- f) Sea $\tilde{\sigma}_{\eta, agrupada}^2$ el estimador infactible de la varianza agrupada calculado con los errores verdaderos en lugar de los residuos de manera que $\tilde{\sigma}_{\eta, agrupada}^2 = \frac{1}{nT} \sum_{i=1}^n (\tilde{X}'_i \tilde{\mathbf{u}}_i)^2$. Demuestre que $\tilde{\sigma}_{\eta, agrupada}^2 \xrightarrow{P} \sigma_{\eta}^2$.
- g) Sea $\tilde{\mathbf{u}}_1 = \tilde{\mathbf{Y}}_i - \hat{\beta} \tilde{\mathbf{X}}_i$ y $\hat{\sigma}_{\eta, agrupada}^2 = \frac{n}{n-1} \frac{1}{nT} \sum_{i=1}^n (\tilde{\mathbf{X}}'_i \hat{\mathbf{u}}_i)^2$ [es decir, la Ecuación (10.27) en forma matricial]. Demuestre que $\hat{\sigma}_{\eta, agrupada}^2 \xrightarrow{P} \sigma_{\eta}^2$ [Sugerencia: utilice un argumento similar al usado en la Ecuación (17.16) para demostrar que $\hat{\sigma}_{\eta, agrupada}^2 - \tilde{\sigma}_{\eta, agrupada}^2 \xrightarrow{P} 0$, a continuación, utilice la respuesta al apartado (f)].
- 18.16** Este ejercicio trata del problema de los datos perdidos estudiado en la Sección 9.2. Considérese el modelo de regresión $Y_i = X_i \beta + u_i$, $i = 1, \dots, n$ donde todas las variables son escalares y por conveniencia se omite el término constante/intercepto.
- Supóngase que se satisfacen los supuestos de mínimos cuadrados del Concepto clave 4.3. Demuestre que el estimador de mínimos cuadrados de β es insesgado y consistente.
 - Supóngase ahora que faltan algunas de las observaciones. Sea I_i una variable aleatoria binaria que indica las observaciones que no se han perdido; es decir, $I_i = 1$ si la observación i -ésima está disponible y $I_i = 0$ si la observación i -ésima se ha perdido. Supóngase que $\{I_i, X_i, u_i\}$ son i.i.d.
 - Demuestre que el estimador MCO se puede escribir como

$$\hat{\beta} = \left(\sum_{i=1}^n I_i X_i X'_i \right)^{-1} \left(\sum_{i=1}^n I_i X_i Y_i \right) = \beta + \left(\sum_{i=1}^n I_i X_i X'_i \right)^{-1} \left(\sum_{i=1}^n I_i X_i u_i \right)$$
 - Supóngase que los datos han «desaparecido de forma completamente aleatoria» en el sentido de que $\Pr(I_i = 1 | X_i, u_i) = p$, donde p es una constante. Demuestre que $\hat{\beta}$ es insesgado y consistente.
 - Supóngase que la probabilidad de que la observación i -ésima falte depende de X_i pero no de u_i ; es decir, $\Pr(I_i = 1 | X_i, u_i) = p(X_i)$. Demuestre que $\hat{\beta}$ es insesgado y consistente.
 - Supóngase que la probabilidad de que la observación i -ésima falte depende tanto de X_i como de u_i ; es decir, $\Pr(I_i = 1 | X_i, u_i) = p(X_i, u_i)$. ¿Es $\hat{\beta}$ insesgado? ¿Es consistente? Explíquelo.
 - Supóngase que $\beta = 1$ y que X_i y u_i son variables aleatorias normales estándar mutuamente independientes [por lo que tanto X_i como u_i se distribuyen como $N(0, 1)$]. Supóngase que $I_i = 1$ cuando $Y_i \geq 0$, pero $I_i = 0$ cuando $Y_i < 0$. ¿Es $\hat{\beta}$ insesgado? ¿Es $\hat{\beta}$ consistente? Explíquelo.
- 18.17** Considérese el modelo de regresión en forma matricial $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\gamma + \mathbf{U}$, donde \mathbf{X} y \mathbf{W} son matrices de regresores y $\boldsymbol{\beta}$ y γ son vectores de coeficientes de regresión desconocidos. Sean $\tilde{\mathbf{X}} = \mathbf{M}_W \mathbf{X}$ e $\tilde{\mathbf{Y}} = \mathbf{M}_W \mathbf{Y}$, donde $\mathbf{M}_W = \mathbf{I} - \mathbf{W}(\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}'$.
- Demuestre que los estimadores MCO de $\boldsymbol{\beta}$ y γ pueden escribirse como

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{X} & \mathbf{X}' \mathbf{W} \\ \mathbf{W}' \mathbf{X} & \mathbf{W}' \mathbf{W} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \mathbf{Y} \\ \mathbf{W}' \mathbf{Y} \end{bmatrix}$$
 - Demuestre que

$$\begin{bmatrix} \mathbf{X}' \mathbf{X} & \mathbf{X}' \mathbf{W} \\ \mathbf{W}' \mathbf{X} & \mathbf{W}' \mathbf{W} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} & -(\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \\ -(\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{X} (\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} & (\mathbf{W}' \mathbf{W})^{-1} + (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{X} (\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \end{bmatrix}$$
 (Sugerencia: demuestre que el producto de las dos matrices es igual a la matriz identidad).
 - Demuestre que $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{M}_W \mathbf{X})^{-1} \mathbf{X}' \mathbf{M}_W \mathbf{Y}$.
 - El teorema de Frisch-Waugh (Apéndice 6.2) establece que $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{Y}}$. Utilice el resultado del apartado (c) para probar el teorema de Frisch-Waugh.

APÉNDICE

18.1 Resumen de álgebra matricial

En este apéndice se resumen los conceptos de vectores, matrices, y los elementos del álgebra matricial utilizados en el Capítulo 18. El objetivo de este apéndice es el de revisar algunos conceptos y definiciones de un curso de álgebra lineal, no el de sustituir a dicho curso.

Definiciones de vectores y matrices

Un **vector** es un conjunto de n números o elementos, recogidos ya sea en una columna (**vector columna**), o en una fila (**vector fila**). El vector columna \mathbf{b} n -dimensional y el vector fila n -dimensional \mathbf{c} son

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \text{y} \quad \mathbf{c} = [c_1 \ c_2 \ \cdots \ c_n]$$

donde b_1 es el primer elemento de \mathbf{b} y, en general, b_i es el i -ésimo elemento de \mathbf{b} .

Un símbolo en negrita señala un vector o una matriz.

Una **matriz** es una colección, o un conjunto, de números o elementos en el que los elementos están colocados en columnas y filas. La dimensión de una matriz es $n \times m$, donde n es el número de filas y m es el número de columnas. La matriz \mathbf{A} $n \times m$ es

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

donde a_{ij} es el elemento (i, j) de \mathbf{A} , es decir, a_{ij} es el elemento que aparece en la i -ésima fila y j -ésima columna. Una matriz $n \times m$ se compone de n vectores fila o, alternativamente, de m vectores columna.

Para distinguir los números unidimensionales de los vectores y las matrices, a los números unidimensionales se les denomina **escalares**.

Tipos de matrices

Matrices cuadradas, simétricas y diagonales. Se dice que una matriz es **cuadrada**, si el número de filas es igual al número de columnas. Una matriz cuadrada se dice que es **simétrica** si su elemento (i, j) es igual a su elemento (j, i) . Una matriz **diagonal** es una matriz cuadrada en la que todos los elementos fuera de la diagonal principal son iguales a cero; es decir, si la matriz cuadrada \mathbf{A} es diagonal, entonces $a_{ij} = 0$ para $i \neq j$.

Matrices particulares. Una matriz importante es la **matriz identidad**, \mathbf{I}_n , que es una matriz $n \times n$ diagonal con unos en su diagonal principal. La **matriz nula** es la matriz $n \times m$ con todos los elementos iguales a cero.

La traspuesta. La **traspuesta** de una matriz intercambia las filas y las columnas. Es decir, la traspuesta de una matriz transforma la matriz \mathbf{A} $n \times m$ en la matriz $m \times n$, que se expresa como \mathbf{A}' , donde el elemento (i, j) de \mathbf{A} se convierte en el elemento (j, i) de \mathbf{A}' ; dicho de otro modo, la traspuesta de la matriz \mathbf{A} cambia las filas de \mathbf{A} en las columnas de \mathbf{A}' . Si a_{ij} es el elemento (i, j) de \mathbf{A} , entonces \mathbf{A}' (la traspuesta de \mathbf{A}) es

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}$$

La traspuesta de un vector es un caso particular de la traspuesta de una matriz. Por tanto la trasposición de un vector convierte un vector columna en un vector fila; es decir, si \mathbf{b} es un vector columna $n \times 1$, entonces su traspuesta es el vector fila $1 \times n$

$$\mathbf{b}' = [b_1 \ b_2 \ \cdots \ b_n]$$

La traspuesta de un vector fila es un vector columna.

Elementos de álgebra matricial: suma y multiplicación

Suma de matrices. Dos matrices \mathbf{A} y \mathbf{B} que tienen la misma dimensión (ambas son $n \times m$) se pueden sumar. La suma de dos matrices es la suma de sus elementos; es decir, si $\mathbf{C} = \mathbf{A} + \mathbf{B}$, entonces $c_{ij} = a_{ij} + b_{ij}$. Un caso particular de la suma de matrices es la suma de vectores: si \mathbf{a} y \mathbf{b} son dos vectores columna $n \times 1$, entonces su suma $\mathbf{c} = \mathbf{a} + \mathbf{b}$ es la suma elemento por elemento; es decir, $c_i = a_i + b_i$.

Multiplicación de vectores y matrices. Sean \mathbf{a} y \mathbf{b} dos vectores columna $n \times 1$. Entonces, el producto del traspuesto de \mathbf{a} (que es un vector fila) por \mathbf{b} es $\mathbf{a}'\mathbf{b} = \sum_{i=1}^n a_i b_i$. Aplicando esta definición con $\mathbf{b} = \mathbf{a}$ se obtiene $\mathbf{a}'\mathbf{a} = \sum_{i=1}^n a_i^2$.

Del mismo modo, las matrices \mathbf{A} y \mathbf{B} pueden multiplicarse entre sí si el número de columnas de \mathbf{A} es igual al número de filas de \mathbf{B} . En concreto, supongamos que \mathbf{A} tiene dimensión $n \times m$, y que \mathbf{B} tiene dimensión $m \times r$. Entonces el producto de \mathbf{A} por \mathbf{B} es una matriz $n \times r$, \mathbf{C} ; es decir, $\mathbf{C} = \mathbf{AB}$, donde el elemento (i, j) de \mathbf{C} es $c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$. Dicho de otra manera, el elemento (i, j) de \mathbf{AB} es el resultado de multiplicar el vector fila que es la fila i -ésima de \mathbf{A} por el vector columna que es la columna j -ésima de \mathbf{B} .

El producto de un escalar d por la matriz \mathbf{A} da como resultado el elemento da_{ij} (i, j); es decir, cada elemento de \mathbf{A} se multiplica por el escalar d .

Algunas propiedades útiles de la suma y la multiplicación de matrices. Sean \mathbf{A} y \mathbf{B} matrices. Entonces:

- a) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$;
- b) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$;
- c) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$;
- d) Si \mathbf{A} es $n \times m$, entonces $\mathbf{A}\mathbf{I}_m = \mathbf{A}$ e $\mathbf{I}_n\mathbf{A} = \mathbf{A}$;
- e) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$;
- f) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$, y
- g) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

En general, la multiplicación de matrices no es comutativa, es decir, en general, $\mathbf{AB} \neq \mathbf{BA}$, aunque existen algunos casos particulares en los que la multiplicación de matrices es comutativa; por ejemplo si \mathbf{A} y \mathbf{B} son matrices diagonales $n \times n$, entonces $\mathbf{AB} = \mathbf{BA}$.

Matriz inversa, matriz raíz cuadrada, y otros temas relacionados

La matriz inversa. Sea \mathbf{A} un matriz cuadrada. Suponiendo que existe, la inversa de la matriz \mathbf{A} se define como la matriz para la cual $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$. Si la matriz inversa \mathbf{A}^{-1} existe, entonces se dice que \mathbf{A} es **invertible** o **no singular**. Si tanto \mathbf{A} como \mathbf{B} son invertibles, entonces $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Matrices definidas positivas y semidefinidas positivas. Sea \mathbf{V} una matriz cuadrada $n \times n$. Entonces \mathbf{V} es definida positiva si $\mathbf{c}'\mathbf{V}\mathbf{c} > 0$ para todo vector \mathbf{c} de dimensión $n \times 1$ distinto de cero. Del mismo modo, \mathbf{V} es **semidefinida positiva** si $\mathbf{c}'\mathbf{V}\mathbf{c} \geq 0$ para todo vector \mathbf{c} de dimensión $n \times 1$ distinto de cero. Si \mathbf{V} es definida positiva, entonces es invertible.

Independencia lineal. Los vectores \mathbf{a}_1 y \mathbf{a}_2 ambos de dimensión $n \times 1$ son **linealmente independientes** si no existen escalares c_1 y c_2 distintos de cero tales que $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 = \mathbf{0}_{n \times 1}$. De manera más general, el conjunto de k vectores, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$, son linealmente independientes si no existen k escalares distintos de cero c_1, c_2, \dots, c_k tales que $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \dots + c_k\mathbf{a}_k = \mathbf{0}_{n \times 1}$.

El rango de una matriz. El **rango** de la matriz \mathbf{A} $n \times m$ es el número de columnas linealmente independientes de \mathbf{A} . El rango de \mathbf{A} se expresa como $\text{rango}(\mathbf{A})$. Si el rango de \mathbf{A} es igual al número de columnas de \mathbf{A} , entonces se dice que \mathbf{A} tiene rango completo de columnas (o de filas). Si la matriz \mathbf{A} $n \times m$ tiene rango completo de columnas, entonces no existe un vector \mathbf{c} de dimensión $m \times 1$ distinto de cero tal que $\mathbf{Ac} = \mathbf{0}_{n \times 1}$. Si \mathbf{A} es $n \times n$ con $\text{rango}(\mathbf{A}) = n$, entonces \mathbf{A} es no singular. Si la matriz \mathbf{A} $n \times m$ tiene rango completo de columnas, entonces $\mathbf{A}'\mathbf{A}$ es no singular.

La matriz raíz cuadrada. Sea V una matriz cuadrada $n \times n$ simétrica y definida positiva. La raíz cuadrada de la matriz V se define como una matriz F $n \times n$ tal que $F'F = V$. La raíz cuadrada de una matriz definida positiva existe siempre, pero no es única. La matriz raíz cuadrada tiene la propiedad de que $FV^{-1}F' = I_n$. Además, la matriz raíz cuadrada de una matriz definida positiva es invertible, por lo que $F'^{-1}VF^{-1} = I_n$.

Valores propios y vectores propios. Sea A una matriz $n \times n$. Si el vector q $n \times 1$ y el escalar λ satisfacen $Aq = \lambda q$, donde $q'q = 1$, entonces λ es un **valor propio** de A y q es el **vector propio** de A asociado al valor propio. Una matriz $n \times n$ tiene n valores propios, que no necesariamente toman valores distintos, y n vectores propios.

Si V es una matriz $n \times n$ simétrica y definida positiva, entonces todos los valores propios de V son números reales positivos y todos los vectores propios de V son reales. Además, V se puede expresar en términos de sus valores y vectores propios como $V = Q\Lambda Q'$, donde Λ es una matriz diagonal $n \times n$ cuyos elementos de la diagonal principal son los valores propios de V y Q es una matriz $n \times n$ formada por los vectores propios de V , dispuestos de modo que la columna i -ésima de Q es el vector propio correspondiente al valor propio que es el i -ésimo elemento de la diagonal principal de Λ . Los vectores propios o autovectores son ortonormales, por lo que $Q'Q = I_n$.

Matrices idempotentes. Una matriz C es idempotente si C es cuadrada y $CC = C$. Si C es una matriz idempotente $n \times n$ que es asimismo simétrica, entonces C es semidefinida positiva y C tiene r valores propios que son iguales a 1 y $n - r$ valores propios que son iguales a 0, donde $r = \text{rango}(C)$ (Ejercicio 18.10).

APÉNDICE

18.2 Distribuciones multivariantes

En este apéndice se recogen varias definiciones y propiedades acerca de las distribuciones de los vectores de variables aleatorias. Comenzamos por definir la media y la matriz de covarianzas de la variable aleatoria n -dimensional V . A continuación se presenta la distribución normal multivariante y posteriormente, se resumen algunas características de las distribuciones de las funciones lineales y cuadráticas de variables aleatorias distribuidas conjuntamente normales.

El vector de medias y la matriz de covarianzas

Los momentos primero y segundo de un vector de variables aleatorias $m \times 1$, $V = (V_1 \ V_2 \ \dots \ V_m)'$, están caracterizados por su vector de medias y matriz de covarianzas.

Debido a que V es un vector, el vector de sus medias, es decir, su **vector de medias**, es $E(V) = \mu_V$. El i -ésimo elemento del vector de medias es la media del i -ésimo elemento de V .

La **matriz de covarianzas** de V es la matriz formada por las varianzas $\text{var}(V_i)$, $i = 1, \dots, n$, a lo largo de la diagonal principal y cuyos elementos (i, j) fuera de la diagonal principal son $\text{cov}(V_i, V_j)$. En forma matricial, la matriz de covarianzas Σ_V es

$$\Sigma_V = E[(V - \mu_V)(V - \mu_V)'] = \begin{bmatrix} \text{var}(V_1) & \cdots & \text{cov}(V_1, V_m) \\ \vdots & \ddots & \vdots \\ \text{cov}(V_m, V_1) & \cdots & \text{var}(V_m) \end{bmatrix} \quad (18.72)$$

La distribución normal multivariante

El vector $m \times 1$ de variables aleatorias V tiene una distribución normal multivariante con vector de medias μ_V y matriz de covarianzas Σ_V si tiene una función de densidad de probabilidad conjunta

$$f(V) = \frac{1}{\sqrt{(2\pi)^m \det(\Sigma_V)}} \exp \left[\frac{1}{2} (V - \mu_V)' \Sigma_V^{-1} (V - \mu_V) \right] \quad (18.73)$$

donde $\det(\Sigma_V)$ es el determinante de la matriz Σ_V . La distribución normal multivariante se expresa mediante $N(\mu_V, \Sigma_V)$.

Una propiedad importante de la distribución normal multivariante es que si dos variables aleatorias normalmente distribuidas de forma conjunta no están correlacionadas (o lo que es lo mismo, tienen una matriz de covarianzas diagonal), entonces están distribuidas de forma independiente. Es decir, sean V_1 y V_2 variables aleatorias distribuidas normal-

mente de forma conjunta con dimensiones respectivas $m_1 \times 1$ y $m_2 \times 1$. Entonces si $\text{cov}(\mathbf{V}_1, \mathbf{V}_2) = E[(\mathbf{V}_1 - \boldsymbol{\mu}_{\mathbf{V}_1})(\mathbf{V}_2 - \boldsymbol{\mu}_{\mathbf{V}_2})'] = \mathbf{0}_{m_1 \times m_2}$, \mathbf{V}_1 y \mathbf{V}_2 son independientes.

Si $\{V_i\}$ son $N(0, \sigma_v^2)$ i.i.d., entonces $\Sigma_V = \sigma_v^2 \mathbf{I}_m$, y la distribución normal multivariante se simplifica al producto de m densidades normales univariantes.

Distribuciones de combinaciones lineales y formas cuadráticas de variables aleatorias normales

Las combinaciones lineales de variables aleatorias normales multivariantes se distribuyen del mismo modo normalmente, y algunas formas cuadráticas de variables aleatorias normales multivariantes tienen una distribución chi-cuadrado. Sea \mathbf{V} una variable aleatoria $m \times 1$ distribuida $N(\boldsymbol{\mu}_V, \Sigma_V)$, sean \mathbf{A} y \mathbf{B} matrices no aleatorias $a \times m$ y $b \times m$, y sea \mathbf{d} un vector no aleatorio $a \times 1$. Entonces

$$\mathbf{d} + \mathbf{a}\mathbf{V} \text{ está distribuida } N(\mathbf{d} + \mathbf{A}\boldsymbol{\mu}_V, \mathbf{A}\Sigma_V\mathbf{A}') \quad (18.74)$$

$$\text{cov}(\mathbf{A}\mathbf{V}, \mathbf{B}\mathbf{V}) = \mathbf{A}\Sigma_V\mathbf{B}' \quad (18.75)$$

$$\text{Si } \mathbf{A}\Sigma_V\mathbf{B}' = \mathbf{0}_{a \times b}, \text{ entonces, } \mathbf{A}\mathbf{V} \text{ y } \mathbf{B}\mathbf{V} \text{ están independientemente distribuidas; y} \quad (18.76)$$

$$(\mathbf{V} - \boldsymbol{\mu}_V)' \Sigma_V^{-1} (\mathbf{V} - \boldsymbol{\mu}_V) \text{ se distribuye } \chi_m^2 \quad (18.77)$$

Sea \mathbf{U} una variable aleatoria normal estándar multivariante m -dimensional con distribución $N(\mathbf{0}, \mathbf{I}_m)$. Si \mathbf{C} es simétrica e idempotente, entonces

$$\mathbf{U}' \mathbf{C} \mathbf{U} \text{ tiene una distribución } \chi_r^2, \text{ donde } r = \text{rango}(\mathbf{C}) \quad (18.78)$$

La ecuación (18.78) se demuestra como Ejercicio (18.11).

APÉNDICE

18.3 Obtención de la distribución asintótica de $\hat{\beta}$

En este apéndice se expone la deducción de la distribución asintótica normal de $\sqrt{n}(\hat{\beta} - \boldsymbol{\beta})$ dada en la Ecuación (18.12). Una implicación de este resultado es que $\hat{\beta} \xrightarrow{p} \boldsymbol{\beta}$.

Considérese en primer lugar la matriz «denominador» $\mathbf{X}'\mathbf{X}/n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i'$ de la Ecuación (18.15). El elemento (j, l) de esta matriz es $\frac{1}{n} \sum_{i=1}^n X_{ji} X_{li}$. Por el segundo supuesto del Concepto clave 18.1, \mathbf{X}_i es i.i.d., por lo que $X_{ji} X_{li}$ es i.i.d. Por el tercer supuesto del Concepto clave 18.1, cada elemento de \mathbf{X}_i tiene momentos hasta el orden cuarto, por lo que, por la desigualdad de Cauchy-Schwarz (Apéndice 17.2), $X_{ji} X_{li}$ tiene momentos de segundo orden. Debido a que $X_{ji} X_{li}$ es i.i.d. con momentos de segundo orden, $\frac{1}{n} \sum_{i=1}^n X_{ji} X_{li}$ cumple la ley de los grandes números, por lo que $\frac{1}{n} \sum_{i=1}^n X_{ji} X_{li} \xrightarrow{p} E(X_{ji} X_{li})$. Esto es cierto para todos los elementos de $\mathbf{X}'\mathbf{X}/n$, por lo que $\mathbf{X}'\mathbf{X}/n \xrightarrow{p} E(\mathbf{X}_i \mathbf{X}_i') = \mathbf{Q}_X$.

Consideremos a continuación la matriz «numerador» de la Ecuación (18.15), $\mathbf{X}'\mathbf{U}/\sqrt{n} = \sqrt{\frac{1}{n}} \sum_{i=1}^n \mathbf{V}_i$, donde $\mathbf{V}_i = \mathbf{X}_i u_i$. Por el primer supuesto del Concepto clave 18.1 y por la ley de esperanzas iteradas, $E(\mathbf{V}_i) = E[\mathbf{X}_i E(u_i | \mathbf{X}_i)] = \mathbf{0}_{k+1}$. Por el segundo supuesto de mínimos cuadrados, \mathbf{V}_i es i.i.d. Sea \mathbf{c} un vector finito $k+1$ dimensional. Por la desigualdad de Cauchy-Schwarz, $E[(\mathbf{c}'\mathbf{V}_i)^2] = E[(\mathbf{c}'\mathbf{X}_i u_i)^2] = E[(\mathbf{c}'\mathbf{X}_i)^2 u_i^2] \leq \sqrt{E[(\mathbf{c}'\mathbf{X}_i)^4]} E(u_i^4)$, que es finita por el tercer supuesto de mínimos cuadrados. Esto es cierto para todo vector \mathbf{c} , por lo que $E(\mathbf{V}_i \mathbf{V}_i') = \Sigma_V$ es finita y, suponemos, definida positiva. Por tanto, el teorema central del límite multivariante del Concepto clave 18.2 es aplicable a $\sqrt{\frac{1}{n}} \sum_{i=1}^n \mathbf{V}_i = \frac{1}{\sqrt{n}} \mathbf{X}'\mathbf{U}$; es decir,

$$\frac{1}{\sqrt{n}} \mathbf{X}'\mathbf{U} \xrightarrow{d} N(\mathbf{0}_{k+1}, \Sigma_V) \quad (18.79)$$

El resultado de la Ecuación (18.12) se deduce a partir de las Ecuaciones (18.15) y (18.79), la consistencia de $\mathbf{X}'\mathbf{X}/n$, el cuarto al supuesto de mínimos cuadrados (que garantiza que $(\mathbf{X}'\mathbf{X})^{-1}$ existe), y el teorema de Slutsky.

APÉNDICE

18.4 Obtención de las distribuciones exactas de los estadísticos de contraste MCO con errores normales

Este apéndice expone las demostraciones de las distribuciones bajo la hipótesis nula del estadístico t válido con homocedasticidad de la Ecuación (18.35) y del estadístico F válido con homocedasticidad de la Ecuación (18.37), suponiendo que se cumplen los seis supuestos del Concepto clave 18.1.

Demostración de la Ecuación (18.35)

Si (i) Z tiene una distribución normal estándar, (ii) W tiene una distribución χ_m^2 , y (iii) Z y W están independientemente distribuidas, entonces la variable aleatoria $Z/\sqrt{W/m}$ tiene una distribución t con m grados de libertad (Apéndice 17.1). Para expresar \tilde{t} de esta forma, debe tenerse en cuenta que $\hat{\Sigma}_{\beta} = (s_u^2/\sigma_u^2)\Sigma_{\beta|X}$. Se reescribe la Ecuación (18.34) como

$$\tilde{t} = \frac{(\hat{\beta}_j - \beta_{j,0})/\sqrt{(\Sigma_{\hat{\beta}|X})_{jj}}}{\sqrt{W/(n-k-1)}} \quad (18.80)$$

donde $W = (n-k-1)(s_u^2/\sigma_u^2)$, y sea $Z = (\hat{\beta}_j - \beta_{j,0})/\sqrt{(\Sigma_{\hat{\beta}|X})_{jj}}$ y $m = n-k-1$. Con estas definiciones, $\tilde{t} = Z/\sqrt{W/m}$. Por lo tanto, para probar el resultado de la Ecuación (18.35), es necesario demostrar desde (i) hasta (iii) para estas definiciones de Z , W , y m .

- i) Una implicación de la Ecuación (18.30) es que, bajo la hipótesis nula, $Z = (\hat{\beta}_j - \beta_{j,0})/\sqrt{(\Sigma_{\hat{\beta}|X})_{jj}}$ tiene una distribución exacta normal estándar, lo que demuestra (i).
- ii) A partir de la Ecuación (18.31), W se distribuye como χ_{n-k-1}^2 , lo que demuestra (ii).
- iii) Para demostrar (iii), debe demostrarse que $\hat{\beta}_j$ y s_u^2 se distribuyen de forma independiente. Por las Ecuaciones (18.14) y (18.29), $\hat{\beta} - \beta = (X'X)^{-1}X'U$ y $s_u^2 = (\mathbf{M}_X U)'(\mathbf{M}_X U)/(n-k-1)$. Por tanto $\hat{\beta} - \beta$ y s_u^2 son independientes si $(X'X)^{-1}X'U$ y $\mathbf{M}_X U$ son independientes. Tanto $(X'X)^{-1}X'U$ como $\mathbf{M}_X U$ son combinaciones lineales de U , que tiene una distribución $N(\mathbf{0}_{n \times 1}, \sigma_u^2 I_n)$, condicionada a X . Sin embargo, debido a que $\mathbf{M}_X X(X'X)^{-1} = \mathbf{0}_{n \times (k+1)}$ [Ecuación (18.26)], se deduce que $(X'X)^{-1}X'U$ y $\mathbf{M}_X U$ están independientemente distribuidos [Ecuación (18.76)]. En consecuencia, bajo los seis supuestos del Concepto clave 18.1,

$$\hat{\beta} \text{ y } s_u^2 \text{ están independientemente distribuidas} \quad (18.81)$$

lo que demuestra (iii) y por lo tanto demuestra la Ecuación (18.35).

Demostración de la Ecuación (18.37)

La distribución F_{n_1, n_2} es la distribución de $(W_1/n_1)/(W_2/n_2)$, donde (i) W_1 está distribuida $\chi_{n_1}^2$; (ii) W_2 está distribuida $\chi_{n_2}^2$; y (iii) W_1 y W_2 están independientemente distribuidas (Apéndice 17.1). Para expresar \tilde{F} de esta forma, sea $W_1 = (\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(X'X)^{-1}\mathbf{R}'\sigma_u^2]^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})$ y $W_2 = (n-k-1)/s_u^2/\sigma_u^2$. La sustitución de estas definiciones en la Ecuación (18.36) muestra que $\tilde{F} = (W_1/q)/[W_2/(n-k-1)]$. Por lo tanto, por la definición de la distribución F , \tilde{F} tiene una distribución $F_{q, n-k-1}$ si se cumplen desde (i) hasta (iii) con $n_1 = q$ y $n_2 = n-k-1$.

- i) Bajo la hipótesis nula, $\hat{\mathbf{R}}\beta - \mathbf{r} = \mathbf{R}(\hat{\beta} - \beta)$. Debido a que $\hat{\beta}$ tiene distribución condicional normal en la Ecuación (18.30) y debido a que \mathbf{R} es una matriz no aleatoria, $\mathbf{R}(\hat{\beta} - \beta)$ está distribuida $N(\mathbf{0}_{q \times 1}, \mathbf{R}(X'X)^{-1}\mathbf{R}'\sigma_u^2)$, condicionada a X . Por lo tanto, por la Ecuación (18.77) del Apéndice 18.2, $(\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(X'X)^{-1}\mathbf{R}'\sigma_u^2]^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})$ se distribuye χ_q^2 . Lo que demuestra (i).
- ii) El requisito (ii) se muestra en la Ecuación (18.31).
- iii) Ya se ha demostrado que $\hat{\beta} - \beta$ y s_u^2 se distribuyen de forma independiente [Ecuación (18.81)]. De ello se desprende que $\mathbf{R}\hat{\beta} - \mathbf{r}$ y s_u^2 se distribuyen de forma independiente, lo que a su vez implica que W_1 y W_2 se distribuyen de forma independiente, lo que demuestra (iii) y completa la demostración.

APÉNDICE

18.5 Prueba del teorema de Gauss-Markov para regresión múltiple

En este apéndice se demuestra el teorema de Gauss-Markov (Concepto clave 18.3) para el modelo de regresión múltiple. Sea $\tilde{\beta}$ un estimador lineal condicionalmente insesgado de β por lo que $\tilde{\beta} = A'Y$ y $E(\tilde{\beta}|X) = \beta$, donde A es una matriz $n \times (k+1)$ que puede depender de X y de constantes no aleatorias. Se demuestra que $\text{var}(c'\tilde{\beta}) \leq \text{var}(c'\hat{\beta})$ para todos los vectores c de dimensión $k+1$, y tal que la desigualdad se convierte en igualdad solamente si $\tilde{\beta} = \hat{\beta}$.

Debido a que $\tilde{\beta}$ es lineal, se puede escribir como $\tilde{\beta} = A'Y = A'(X\beta + U) = (A'X)\beta + A'U$. Por la primera condición de Gauss-Markov, $E(U|X) = \mathbf{0}_{n \times 1}$, por lo que $E(\tilde{\beta}|X) = (A'X)\beta$, pero debido a que $\tilde{\beta}$ es condicionalmente insesgado, $E(\tilde{\beta}|X) = \beta = (A'X)\beta$, lo que implica que $A'X = I_{k+1}$. Por tanto $\tilde{\beta} = \beta + A'U$, por lo que $\text{var}(\tilde{\beta}|X) = \text{var}(A'U|X) = E(A'UU'A|X) = A'E(UU')A = \sigma_u^2 A'A$, donde la tercera igualdad se obtiene debido a que A puede depender de X , pero no de U y la última igualdad se deduce de la segunda condición de Gauss-Markov. Es decir, si $\tilde{\beta}$ es lineal e insesgado, entonces bajo las condiciones de Gauss-Markov,

$$A'X = I_{k+1} \quad \text{y} \quad \text{var}(\tilde{\beta}|X) = \sigma_u^2 A'A \quad (18.82)$$

Los resultados de la Ecuación (18.82) son aplicables asimismo a $\hat{\beta}$ con $A = \hat{A} = X(X'X)^{-1}$, donde $(X'X)^{-1}$ existe por la tercera condición de Gauss-Markov.

Sea ahora $A = \hat{A} + D$ por lo que D es la diferencia entre las matrices de ponderaciones A y \hat{A} . Téngase en cuenta que $\hat{A}'A = (X'X)^{-1}X'A = (X'X)^{-1}$ [por la Ecuación (18.82)] y $\hat{A}'\hat{A} = (X'X)^{-1}X'X(X'X)^{-1} = (X'X)^{-1}$, por lo que $\hat{A}'D = \hat{A}'(A - \hat{A}) = \hat{A}'A - \hat{A}'\hat{A} = \mathbf{0}_{(k+1) \times (k+1)}$. Sustituyendo $A = \hat{A} + D$ en la fórmula para la varianza condicional de la Ecuación (18.82) se obtiene

$$\begin{aligned} \text{var}(\tilde{\beta}|X) &= \sigma_u^2 (\hat{A} + D)'(\hat{A} + D) \\ &= \sigma_u^2 [\hat{A}'\hat{A} + \hat{A}'D + D'\hat{A} + D'D] \\ &= \sigma_u^2 (X'X)^{-1} + \sigma_u^2 D'D \end{aligned} \quad (18.83)$$

donde la última igualdad utiliza las propiedades $\hat{A}'\hat{A} = (X'X)^{-1}$ y $\hat{A}'D' = \mathbf{0}_{(k+1) \times (k+1)}$.

Debido a que $\text{var}(\tilde{\beta}|X) = \sigma_u^2 (X'X)^{-1}$, las Ecuaciones (18.82) y (18.83) implican que $\text{var}(\tilde{\beta}|X) - \text{var}(\hat{\beta}|X) = \sigma_u^2 D'D$. La diferencia entre las varianzas de los dos estimadores de las combinaciones lineales $c'\beta$ por tanto es

$$\text{var}(c'\tilde{\beta}|X) - \text{var}(c'\hat{\beta}|X) = \sigma_u^2 c'D'Dc \geq 0 \quad (18.84)$$

La desigualdad de la Ecuación (18.84) se cumple para todas las combinaciones lineales $c'\beta$, y la desigualdad se cumple con igualdad para todo c distinto de cero solo si $D = \mathbf{0}_{n \times (k+1)}$, es decir, si $A = \hat{A}$, o equivalentemente, $\tilde{\beta} = \hat{\beta}$. Por lo tanto $c'\hat{\beta}$ tiene la menor varianza de todos los estimadores lineales condicionalmente insesgados de $c'\beta$; es decir, el estimador MCO es ELIO.

APÉNDICE

18.6 Pruebas de algunos resultados seleccionados de la estimación VI y MGM

La eficiencia de MC2E con homocedasticidad [prueba de la Ecuación (18.62)]

Cuando los errores u_i son homocedásticos, la diferencia entre Σ_A^{VI} [Ecuación (18.61)] y Σ^{MC2E} [Ecuación (18.55)] está dada por

$$\begin{aligned} \Sigma_A^{VI} - \Sigma^{MC2E} &= (Q_{XZ}A Q_{ZX})^{-1} Q_{XZ}A Q_{ZZ}A Q_{ZX} (Q_{XZ}A Q_{ZX})^{-1} \sigma_u^2 - (Q_{XZ}Q_{ZZ}^{-1} Q_{ZX})^{-1} \sigma_u^2 \\ &= (Q_{XZ}A Q_{ZX})^{-1} Q_{XZ}A [Q_{ZZ} - Q_{ZX}(Q_{XZ}Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ}] A Q_{ZX} (Q_{XZ}A Q_{ZX})^{-1} \sigma_u^2 \end{aligned} \quad (18.85)$$

donde el segundo término entre corchetes de la segunda igualdad se deduce de que $(Q_{XZ}A Q_{ZX})^{-1} Q_{XZ}A Q_{ZX} = I_{(k+r+1)}$. Sea F la matriz raíz cuadrada de Q_{ZZ} , por lo que $Q_{ZZ} = F'F$ y $Q_{ZZ}^{-1} = F^{-1}F'^{-1}$ [la última igualdad se deduce de que

$(\mathbf{F}'\mathbf{F})^{-1}$ y $\mathbf{F}^{-1}\mathbf{F}'^{-1}$ y $\mathbf{F}'^{-1} = \mathbf{F}^{-1}$. Por tanto la última expresión de la Ecuación (18.85) se puede reescribir para dar lugar a

$$\begin{aligned}\Sigma_A^{VI} - \Sigma^{MC2E} &= (\mathbf{Q}_{XZ}A\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}A\mathbf{F}'[\mathbf{I} - \mathbf{F}^{-1}\mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}\mathbf{F}^{-1}\mathbf{F}'^{-1}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{F}^{-1}] \\ &\quad \times \mathbf{F}\mathbf{A}\mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}A\mathbf{Q}_{ZX})^{-1}\sigma_u^2\end{aligned}\quad (18.86)$$

donde la segunda expresión entre paréntesis utiliza $\mathbf{F}'\mathbf{F}^{-1} = \mathbf{I}$. Por lo tanto

$$\mathbf{c}'(\Sigma_A^{VI} - \Sigma^{MC2E})\mathbf{c} = \mathbf{d}'[\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{d}\sigma_u^2 \quad (18.87)$$

donde $\mathbf{d} = \mathbf{F}\mathbf{A}\mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}A\mathbf{Q}_{ZX})^{-1}\mathbf{c}$ y $\mathbf{D} = \mathbf{F}^{-1}\mathbf{Q}_{ZX}$. Ahora, $\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ es una matriz simétrica e idempotente (Ejercicio 18.5). Como resultado, $\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ tiene valores propios que son 0 o 1 y $\mathbf{d}'[\mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{d} \geq 0$ (Ejercicio 18.10). Por lo tanto $\mathbf{c}'(\Sigma_A^{VI} - \Sigma^{MC2E})\mathbf{c} \geq 0$, demostrando que MC2E es eficiente con homocedasticidad.

Distribución asintótica del estadístico J con homocedasticidad

El estadístico J está definido en la Ecuación (18.63). Téngase en cuenta que

$$\begin{aligned}\hat{\mathbf{U}} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{MC2E} \\ &= \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{Y} \\ &= (\mathbf{X}\boldsymbol{\beta} + \mathbf{U}) - \mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}) \\ &= \mathbf{U} - \mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{U} \\ &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z]\mathbf{U}\end{aligned}\quad (18.88)$$

Por tanto

$$\begin{aligned}\hat{\mathbf{U}}'\mathbf{P}_Z\hat{\mathbf{U}} &= \mathbf{U}'[\mathbf{I} - \mathbf{P}_Z\mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}']\mathbf{P}_Z[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z]\mathbf{U} \\ &= \mathbf{U}'[\mathbf{P}_Z - \mathbf{P}_Z\mathbf{X}(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z]\mathbf{U}\end{aligned}\quad (18.89)$$

donde la segunda igualdad se deduce mediante la simplificación de la expresión precedente. Debido a que $\mathbf{Z}'\mathbf{Z}$ es simétrica y definida positiva, se puede escribir en términos de su matriz raíz cuadrada, $\mathbf{Z}'\mathbf{Z} = (\mathbf{Z}'\mathbf{Z})^{-1/2}(\mathbf{Z}'\mathbf{Z})^{1/2}$, y esta matriz raíz cuadrada es invertible, por lo que $(\mathbf{Z}'\mathbf{Z})^{-1} = (\mathbf{Z}'\mathbf{Z})^{-1/2}(\mathbf{Z}'\mathbf{Z})^{-1/2}$, donde $(\mathbf{Z}'\mathbf{Z})^{-1/2} = [(\mathbf{Z}'\mathbf{Z})^{1/2}]^{-1}$. Por tanto \mathbf{P}_Z se puede escribir como $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{B}\mathbf{B}'$, donde $\mathbf{B} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2}$. Sustituyendo esta expresión por \mathbf{P}_Z en la última expresión de la Ecuación (18.89) se obtiene

$$\begin{aligned}\hat{\mathbf{U}}'\mathbf{P}_Z\hat{\mathbf{U}} &= \mathbf{U}'[\mathbf{B}\mathbf{B}' - \mathbf{B}\mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{B}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{B}']\mathbf{U} \\ &= \mathbf{U}'\mathbf{B}[\mathbf{I} - \mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{B}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}]\mathbf{B}'\mathbf{U} \\ &= \mathbf{U}'\mathbf{B}\mathbf{M}_{\mathbf{B}'\mathbf{X}}\mathbf{B}'\mathbf{U}\end{aligned}\quad (18.90)$$

donde $\mathbf{M}_{\mathbf{B}'\mathbf{X}} = \mathbf{I} - \mathbf{B}'\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{B}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}$ es una matriz simétrica e idempotente.

La distribución asintótica nula de $\hat{\mathbf{U}}'\mathbf{P}_Z\hat{\mathbf{U}}$ se deduce mediante el cálculo de los límites en probabilidad y en distribución de los distintos términos de la última expresión de la Ecuación (18.90) bajo la hipótesis nula. Bajo la hipótesis nula de que $E(\mathbf{Z}_i u_i) = 0$, $\mathbf{Z}'\mathbf{U}/\sqrt{n}$ tiene una media igual a cero y el teorema central del límite es aplicable, por lo que $\mathbf{Z}'\mathbf{U}/\sqrt{n} \xrightarrow{d} N(0, \mathbf{Q}_{ZZ}\sigma_u^2)$. Además, $\mathbf{Z}'\mathbf{Z}/n \xrightarrow{p} \mathbf{Q}_{ZZ}$ y $\mathbf{X}'\mathbf{Z}/n \xrightarrow{p} \mathbf{Q}_{XZ}$. Por lo tanto $\mathbf{B}'\mathbf{U} = (\mathbf{Z}'\mathbf{Z})^{-1/2}\mathbf{Z}'\mathbf{U} = (\mathbf{Z}'\mathbf{Z}/n)^{-1/2}(\mathbf{Z}'\mathbf{U}/\sqrt{n}) \xrightarrow{d} \sigma_u z$, donde z se distribuye $N(\mathbf{0}_{m+r+1}, \mathbf{I}_{m+r+1})$. Además, $\mathbf{B}'\mathbf{X}/\sqrt{n} = (\mathbf{Z}'\mathbf{Z}/n)^{-1/2}(\mathbf{Z}'\mathbf{X}/n) \xrightarrow{p} \mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZX}$, por lo que $\mathbf{M}_{\mathbf{B}'\mathbf{X}} \xrightarrow{p} \mathbf{I} - \mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZX}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZX})^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1/2} = \mathbf{M}_{\mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZX}}$. Por tanto

$$\hat{\mathbf{U}}'\mathbf{P}_Z\hat{\mathbf{U}} \xrightarrow{d} (z' \mathbf{M}_{\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1/2}\mathbf{Q}_{ZX}} z)\sigma_u^2. \quad (18.91)$$

Bajo la hipótesis nula, el estimador MC2E es consistente y los coeficientes de la regresión de $\hat{\mathbf{U}}$ sobre \mathbf{Z} convergen en probabilidad a cero [una consecuencia de la Ecuación (18.91)], por lo que el denominador de la definición del estadístico J es un estimador consistente de σ_u^2 :

$$\hat{\mathbf{U}}'\mathbf{M}_{\mathbf{Z}}\hat{\mathbf{U}}/(n - m - r - 1) \xrightarrow{p} \sigma_u^2 \quad (18.92)$$

De la definición del estadístico J y de las Ecuaciones (18.91) y (18.92), se deduce que

$$\mathbf{J} = \frac{\hat{\mathbf{U}}' \mathbf{P}_Z \hat{\mathbf{U}}}{\hat{\mathbf{U}}' \mathbf{M}_Z \hat{\mathbf{U}} / (n - m - r - 1)} \xrightarrow{d} \mathbf{z}' \mathbf{M}_{Q_{ZZ}^{-1/2} Q_{ZX}} \mathbf{z} \quad (18.93)$$

Debido a que \mathbf{z} es un vector aleatorio normal estándar y $\mathbf{M}_{Q_{ZZ}^{-1/2} Q_{ZX}}$ es una matriz simétrica e idempotente, \mathbf{J} se distribuye como una variable aleatoria chi-cuadrado con grados de libertad que son iguales al rango de $\mathbf{M}_{Q_{ZZ}^{-1/2} Q_{ZX}}$ [Ecuación (18.78)]. Debido a que $\mathbf{Q}_{ZZ}^{-1/2} \mathbf{Q}_{ZX}$ es $(m + r + 1) \times (k + r + 1)$ y $m > k$, el rango de $\mathbf{M}_{Q_{ZZ}^{-1/2} Q_{ZX}}$ es $m - k$ [Ejercicio (18.5)]. Por lo tanto $\mathbf{J} \xrightarrow{d} \chi^2_{m-k}$, que es el resultado enunciado en la Ecuación (18.64).

La eficiencia del estimador MGM eficiente

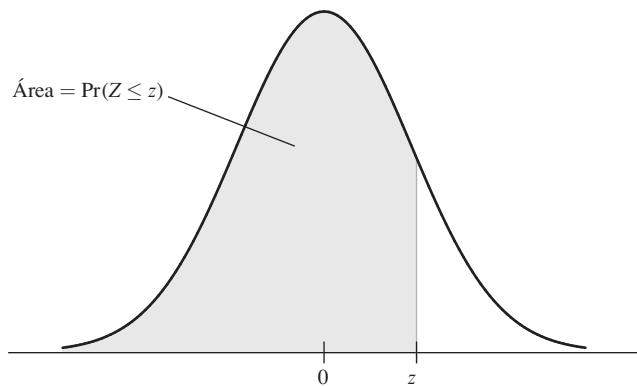
El estimador MGM eficiente infactible, $\tilde{\beta}^{MGM,Ef}$ se define en la Ecuación (18.66). La prueba de que $\tilde{\beta}^{MGM,Ef}$ es eficiente implica demostrar que $\mathbf{c}'(\Sigma_A^{VII} - \Sigma^{MGM,Ef})\mathbf{c} \geq 0$ para todo vector \mathbf{c} . La prueba es muy similar a la prueba de la eficiencia del estimador MC2E en la primera sección de este apéndice, con la única modificación de que \mathbf{H}^{-1} sustituye a $\mathbf{Q}_{ZZ} \sigma_u^2$ en la Ecuación (18.85) y en adelante.

Distribución del estadístico J MGM

El estadístico J MGM está dado en la Ecuación (18.70). La prueba de que, bajo la hipótesis nula, $J^{MGM} \xrightarrow{d} \chi^2_{m-k}$ es muy similar a la prueba correspondiente para el estadístico J MC2E con homocedasticidad.

Apéndice

TABLA 1 La función de distribución normal estándar acumulada, $\Phi(z) = \Pr(Z \leq z)$



Valor del segundo decimal de z

z	0	1	2	3	4	5	6	7	8	9
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611

(Table I continued)

(Tabla 1 continuación)

z	Valor del segundo decimal de z									
	0	1	2	3	4	5	6	7	8	9
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
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	0,8980	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	0,9452	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	0,9616	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

Esta tabla puede utilizarse para calcular $\Pr(Z \leq z)$, donde Z es una variable normal estándar. Por ejemplo, cuando $z = 1,17$, esta probabilidad es de 0,8790, que en la tabla es la casilla de la fila 1.1 y la columna 7.

TABLA 2 Valores críticos para los contrastes bilaterales y unilaterales realizados mediante la distribución *t* de Student

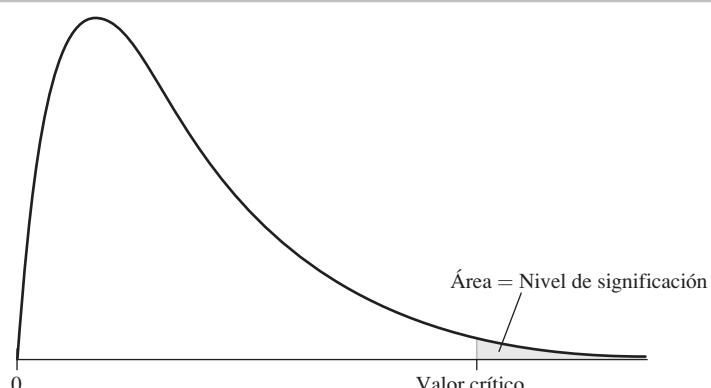
Grados de libertad	Nivel de significación					
	20% (Bilateral) 10% (Unilateral)	10% (Bilateral) 5% (Unilateral)	5% (Bilateral) 2,5% (Unilateral)	2% (Bilateral) 1% (Unilateral)	1% (Bilateral) 0,5% (Unilateral)	
1	3,08	6,31	12,71	31,82	63,66	
2	1,89	2,92	4,30	6,96	9,92	
3	1,64	2,35	3,18	4,54	5,84	
4	1,53	2,13	2,78	3,75	4,60	
5	1,48	2,02	2,57	3,36	4,03	
6	1,44	1,94	2,45	3,14	3,71	
7	1,41	1,89	2,36	3,00	3,50	
8	1,40	1,86	2,31	2,90	3,36	
9	1,38	1,83	2,26	2,82	3,25	
10	1,37	1,81	2,23	2,76	3,17	
11	1,36	1,80	2,20	2,72	3,11	
12	1,36	1,78	2,18	2,68	3,05	
13	1,35	1,77	2,16	2,65	3,01	
14	1,35	1,76	2,14	2,62	2,98	
15	1,34	1,75	2,13	2,60	2,95	
16	1,34	1,75	2,12	2,58	2,92	
17	1,33	1,74	2,11	2,57	2,90	
18	1,33	1,73	2,10	2,55	2,88	
19	1,33	1,73	2,09	2,54	2,86	
20	1,33	1,72	2,09	2,53	2,85	
21	1,32	1,72	2,08	2,52	2,83	
22	1,32	1,72	2,07	2,51	2,82	
23	1,32	1,71	2,07	2,50	2,81	
24	1,32	1,71	2,06	2,49	2,80	
25	1,32	1,71	2,06	2,49	2,79	
26	1,32	1,71	2,06	2,48	2,78	
27	1,31	1,70	2,05	2,47	2,77	
28	1,31	1,70	2,05	2,47	2,76	
29	1,31	1,70	2,05	2,46	2,76	
30	1,31	1,70	2,04	2,46	2,75	
60	1,30	1,67	2,00	2,39	2,66	
90	1,29	1,66	1,99	2,37	2,63	
120	1,29	1,66	1,98	2,36	2,62	
∞	1,28	1,64	1,96	2,33	2,58	

Se muestran los valores para los valores críticos de las hipótesis alternativas bilaterales (\neq) y unilaterales ($>$). El valor crítico para el contraste unilateral ($<$) es el opuesto ($>$) del valor crítico mostrado en la tabla. Por ejemplo, 2,13 es el valor crítico para un contraste bilateral con un nivel de significación del 5% realizado mediante la distribución *t* de Student con 15 grados de libertad.

TABLA 3 Valores críticos para la distribución χ^2

Grados de libertad	Nivel de significación		
	10 %	5 %	1 %
1	2,71	3,84	6,63
2	4,61	5,99	9,21
3	6,25	7,81	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,72
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,81
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	35,17	41,64
24	33,20	36,41	42,98
25	34,38	37,65	44,31
26	35,56	38,89	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89

Esta tabla contiene los percentiles 90, 95 y 99 de la distribución χ^2 . Estos sirven como valores críticos para los contrastes con niveles de significación del 10 %, 5 % y 1 %.

TABLA 4 Valores críticos para la distribución $F_{m,\infty}$ 

Grados de libertad	Nivel de significación		
	10 %	5 %	1 %
1	2,71	3,84	6,63
2	2,30	3,00	4,61
3	2,08	2,60	3,78
4	1,94	2,37	3,32
5	1,85	2,21	3,02
6	1,77	2,10	2,80
7	1,72	2,01	2,64
8	1,67	1,94	2,51
9	1,63	1,88	2,41
10	1,60	1,83	2,32
11	1,57	1,79	2,25
12	1,55	1,75	2,18
13	1,52	1,72	2,13
14	1,50	1,69	2,08
15	1,49	1,67	2,04
16	1,47	1,64	2,00
17	1,46	1,62	1,97
18	1,44	1,60	1,93
19	1,43	1,59	1,90
20	1,42	1,57	1,88
21	1,41	1,56	1,85
22	1,40	1,54	1,83
23	1,39	1,53	1,81
24	1,38	1,52	1,79
25	1,38	1,51	1,77
26	1,37	1,50	1,76
27	1,36	1,49	1,74
28	1,35	1,48	1,72
29	1,35	1,47	1,71
30	1,34	1,46	1,70

Esta tabla contiene los percentiles 90, 95 y 99 de la distribución $F_{m,\infty}$. Estos sirven como valores críticos para los contrastes con niveles de significación del 10 %, 5 % y 1 %.

TABLA 5A Valores críticos para la distribución F_{n_1, n_2} . Nivel de significación del 10 %.

Grados de libertad del denominador	Grados de libertad del numerador (n_1)									
	1	2	3	4	5	6	7	8	9	10
1	39,86	49,50	53,59	55,83	57,24	58,20	58,90	59,44	59,86	60,20
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60

Esta tabla contiene los percentiles 90 de la distribución F_{n_1, n_2} , que sirven como valores críticos para un contraste con un nivel de significación del 10%.

TABLA 5B Valores críticos para la distribución F_{n_1,n_2} . Nivel de Significación del 5 %.

Grados de libertad del denominador (n_2)	Grados de libertad del numerador (n_1)									
	1	2	3	4	5	6	7	8	9	10
1	161,40	199,50	215,70	224,60	230,20	234,00	236,80	238,90	240,50	241,90
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,39	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Esta tabla contiene los percentiles 95 de la distribución F_{n_1,n_2} , que sirven como valores críticos para un contraste con nivel de significación del 5 %.

TABLA 5C Valores críticos para la distribución F_{n_1,n_2} . Nivel de significación del 1%.

Grados de libertad del denominador (n_2)	Grados de libertad del numerador (n_1)									
	1	2	3	4	5	6	7	8	9	10
1	4052,00	4999,00	5403,00	5624,00	5763,00	5859,00	5928,00	5981,00	6022,00	6055,00
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

Esta tabla contiene los percentiles 99 de la distribución F_{n_1,n_2} , que sirven como valores críticos para un contraste con nivel de significación del 1 %.

Bibliografía

- Adda, Jérôme, y Francesca Cornaglia (2006): «Taxes, Cigarette Consumption, and Smoking Intensity», *American Economic Review*, 96(4): 1013-1028.
- Aggarwal, Rajesh K., y Philippe Jorion (2010): «The Performance of Emerging Hedge Funds and Managers», *Journal of Financial Economics*, 96: 238-256.
- Anderson, Theodore W., y Herman Rubin (1950): «Estimators of the Parameters of a Single Equationin a Complete Set of Stochastic Equations», *Annalsof Mathematical Statistics*, 21: 570-582.
- Andrews, Donald W. K. (1991): «Heteroskedasticity and Auto-correlation Consistent Covariance Matrix Estimation», *Econometrica*, 59(3): 817-858.
- Andrews, Donald W. K. (1993): «Tests for Parameter Instability and Structural Change with Unknown Change Point», *Econometrica*, 61(4): 821-856.
- Andrews, Donald W. K. (2003): «Tests For Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum», *Econometrica*, 71: 395-397.
- Angrist, Joshua (1990): «Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records», *American Economic Review*, 80(3): 313-336.
- Angrist, Joshua, y William Evans (1998): «Childrenand Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size», *American Economic Review*, 88(3): 450-477.
- Angrist, Joshua; Kathryn Graddy, y Guido Imbens (2000): «The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish», *Review of Economic Studies*, 67(232): 499-527.
- Angrist, Joshua, y Alan Krueger (1991): «Does Compulsory School Attendance Affect Schooling and Earnings?», *Quarterly Journal of Economics*, 106(4): 979-1014.
- Angrist, Joshua, y Alan B. Krueger (2001): «Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments», *Journal of Economic Perspectives*, 15(4), fall: 69-85.
- Arellano, Manuel (2003): *Panel Data Econometrics*, Oxford: Oxford University Press.
- Ayres, Ian, y John Donohue (2003): «Shooting Down the ‘More Guns Less Crime’ Hypothesis», *Stanford Law Review*, 55: 1193-1312.
- Barendregt, Jan J. (1997): «The Health Care Costs of Smoking», *New England Journal of Medicine*, 337(15): 1052-1057.
- Beck, Thorsten; Ross Levine, y Norman Loayza (2000): «Finance and the Sources of Growth», *Journal of Financial Economics*, 58: 261-300.
- Benartzi, Shlomo, y Richard H. Thaler (2007): «Heuristics and Biases in Retirement Savings Behavior», *Journal of Economic Perspectives*, 21(3): 81-104.
- Bergstrom, Theodore A. (2001): «Free Labor for Costly Journals?», *Journal of Economic Perspectives*, 15(4), fall: 183-198.
- Bertrand, Marianne, y Kevin Hallock (2001): «The Gender Gap in Top Corporate Jobs», *Industrial and Labor Relations Review*, 55(1): 3-21.
- Bertrand, Marianne, y Sendhil Mullainathan (2004): «Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination», *American Economic Review*, 94(4): 991-1013.
- Beshears, John; James J. Choi; David Laibson, y Brigitte C. Madrian (2008): «The Importance of Default Options for Retirement Saving Outcomes: Evidence from the United States», en *Lessons from Pension Reform in the Americas*, editado por Stephen J. Kay y Tapen Sinha, Oxford: Oxford University Press, 59-87.
- Bollersev, Tim (1986): «Generalized Autoregressive Conditional Heteroskedasticity», *Journal of Econometrics*, 31(3): 307-327.
- Bound, John; David A. Jaeger, y Regina M. Baker (1995): «Problems with Instrumental Variables Estimation When the Correlation Between the Instrument and the Endogenous Explanatory Variable Is Weak», *Journal of the American Statistical Association*, 90(430): 443-450.
- Campbell, John Y. (2003): «Consumption-Based Asset Pricing», Cap. 13 en *Handbook of the Economics of Finance*, editado por Milton Harris y Rene Stulz, Amsterdam, Elsevier.

- Campbell, John Y., y Motohiro Yogo (2005): «Efficient Tests of Stock Return Predictability», *Journal of Financial Economics*, 81(1): 27-60.
- Card, David (1990): «The Impact of the Mariel Boatlift on the Miami Labor Market», *Industrial and Labor Relations Review*, 43(2): 245-257.
- Card, David (1999): The Causal Effect of Education on Earnings», Cap. 30, en *The Handbook of Labor Economics*, editado por Orley C. Ashenfelter y David Card, Amsterdam, Elsevier.
- Card, David, y Alan B. Krueger (1994): «Minimum Wages and Employment: A Case Study of the Fast Food Industry», *American Economic Review*, 84(4): 772-793.
- Carhart, Mark M. (1997): «On Persistence in Mutual Fund Performance», *Journal of Finance*, 52(1): 57-82.
- Carpenter, Christopher, y Philip J. Cook (2008): «Cigarette Taxes and Youth Smoking: New Evidence from National, State, and Local Youth Risk Behavior Surveys», *Journal of Health Economics*, 27: 287-299.
- Chaloupka, Frank J.; Michael Grossman, y Henry Saffer (2002): «The Effect of Price on Alcohol Consumption and Alcohol-Related Problems», *Alcohol Research & Health*, 26: 22-34.
- Chaloupka, Frank J., y Kenneth E. Warner (2000): «The Economics of Smoking», Cap. 29, en *The Handbook of Health Economics*, editado por Joseph P. Newhouse y Anthony J. Cuyler, New York, North Holland.
- Chow, Gregory (1960): «Tests of Equality Between Sets of Coefficients in Two Linear Regressions», *Econometrica*, 28(3): 591-605.
- Clements, Michael P. (2004): «Evaluating the Bank of England Density Forecasts of Inflation», *Economic Journal*, 114: 844-866.
- Cochrane, D., y Guy Orcutt (1949): «Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms», *Journal of the American Statistical Association*, 44(245): 32-61.
- Cohen, Alma, y Liran Einav (2003): «The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities», *The Review of Economics and Statistics*, 85(4): 828-843.
- Cook, Philip J., y Michael J. Moore (2000): «Alcohol», Cap. 30, en *The Handbook of Health Economics*, editado por Joseph P. Newhouse y Anthony J. Cuyler, New York, North Holland.
- Cooper, Harris, y Larry V. Hedges (1994): *The Handbook of Research Synthesis*, New York, Russell Sage Foundation.
- Dang, Jennifer N. (2008): «Statistical Analysis of Alcohol-Related Driving Trends, 1982-2005», Technical Report DOT HS 810 942, Washington, DC, U.S. National Highway Traffic Safety Administration.
- Davidson, James E. H.; David F. Hendry; Frank Srba, y Stephen Yeo (1978): «Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom», *Economic Journal*, 88: 661-692.
- Dickey, David A., y Wayne A. Fuller (1979): «Distribution of the Estimators for Autoregressive Time Series with a Unit Root», *Journal of the American Statistical Association*, 74(366): 427-431.
- Diebold, Francis X. (2007): *Elements of Forecasting* (4.^a ed.), Cincinnati, South-Western.
- Ehrenberg, Ronald G.; Dominic J. Brewer; Adam Gamoran, y J. Douglas Willms (2001a): «Class Size and Student Achievement», *Psychological Science in the Public Interest*, 2(1): 1-30.
- Ehrenberg, Ronald G.; Dominic J. Brewer, Adam Gamoran, y J. Douglas Willms (2001b): «Does Class Size Matter?», *Scientific American*, 285(5): 80-85.
- Eicker, F. (1967): «Limit Theorems for Regressions with Unequal and Dependent Errors», *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 59-82, Berkeley, University of California Press.
- Elliott, Graham; Thomas J. Rothenberg, y James H. Stock (1996): «Efficient Tests for an Autoregressive Unit Root», *Econometrica*, 64(4): 813-836.
- Enders, Walter (1995), *Applied Econometric Time Series*, New York, Wiley.
- Engle, Robert F. (1982): «Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation», *Econometrica*, 50(4): 987-1007.
- Engle, Robert F., y Clive W. J. Granger (1987): «Cointegration and Error Correction: Representation, Estimation and Testing», *Econometrica*, 55(2): 251-276.
- Evans, William; Matthew Farrelly, y Edward Montgomery (1999): «Do Workplace Smoking Bans Reduce Smoking?», *American Economic Review*, 89(4): 728-747.
- Foster, Donald (1996): «Primary Culprit: An Analysis of a Novel of Politics», *New York Magazine*, 29(8), febrero 26.
- Fuller, Wayne A. (1976): *Introduction to Statistical Time Series*, New York, Wiley.
- Garvey, Gerald T., y Gordon Hanka (1999): «Capital Structure and Corporate Control: The Effect of Antitakeover Statutes on Firm Leverage», *Journal of Finance*, 54(2): 519-546.
- Gillespie, Richard (1991): *Manufacturing Knowledge: A History of the Hawthorne Experiments*, New York, Cambridge University Press.
- Goering, John, y Ron Wienk (eds.) (1996): *Mortgage Lending, Racial Discrimination, and Federal Policy*, Washington, DC, Urban Institute Press.
- Goyal, Amit, y Ivo Welch (2003): «Predicting the Equity Premium with Dividend Ratios», *Management Science*, 49(5): 639-654.
- Granger, Clive W. J. (1969): «Investigating Causal Relations by Econometric Models and Cross-Spectral Methods», *Econometrica*, 37(3): 424-438.
- Granger, Clive W. J., y A. A. Weiss (1983): «Time Series Analysis of Error-Correction Models», pp. 255-278, en *Studies in Econometrics: Time Series and Multivariate Statistics*, editado por S. Karlin, T. Amemiya, y L. A. Goodman, New York, Academic Press.
- Greene, William H. (2000): *Econometric Analysis* (4.^a ed.), Upper Saddle River, NJ, Prentice Hall.
- Gruber, Jonathan (2001): «Tobacco at the Crossroads: The Past and Future of Smoking Regulation in the United States», *Journal of Economic Perspectives*, 15(2): 193-212.
- Haldrup, Niels, y Michael Jansson (2006): «Improving Size and Power in Unit Root Testing», pp. 252-277, en *Palgrave*

- Handbook of Econometrics*, vol. 1: *Econometric Theory*, editado por Terrence Mills y Kerry Patterson, Basingstoke U. K., Palgrave MacMillan.
- Hamerhoff, Daniel, y Amy Parker (2005): «Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity», *Economics of Education Review*, 24(4): 369-376.
- Hamilton, James D. (1994): *Time Series Analysis*, Princeton, NJ, Princeton University Press.
- Hansen, Bruce (1992): «Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends», *Journal of Econometrics*, 53(1-3): 86-121.
- Hansen, Bruce (2001): «The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity», *Journal of Economic Perspectives*, 15(4), fall: 117-128.
- Hanushek, Eric (1999a): «Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects», *Educational Evaluation and Policy Analysis*, 21: 143-164.
- Hanushek, Eric (1999b): «The Evidence on Class Size», Cap. 7, en *Earning and Learning: How Schools Matter*, editado por S. Mayer y P. Peterson, Washington, DC, Brookings Institution Press.
- Hayashi, Fumio (2000): *Econometrics*, Princeton, NJ, Princeton University Press.
- Heckman, James J. (1974): «Shadow Prices, Market Wages, and Labor Supply», *Econometrica*, 42: 679-694.
- Heckman, James J. (2001): «Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture», *Journal of Political Economy*, 109(4): 673-748.
- Heckman, James J.; Robert J. LaLonde, y Jeffrey A. Smith (1999): «The Economics and Econometrics of Active Labor Market Programs», Cap. 31, en *Handbook of Labor Economics*, editado por Orley Ashenfelter y David Card, Amsterdam, Elsevier.
- Hedges, Larry V., y Ingram Olkin (1985): *Statistical Methods for Meta-analysis*, San Diego, Academic Press.
- Hetland, Lois (2000): «Listening to Music Enhances Spatial-Temporal Reasoning: Evidence for the ‘‘Mozart Effect’’», *Journal of Aesthetic Education*, 34(3-4): 179-238.
- Hoxby, Caroline M. (2000): «The Effects of Class Size on Student Achievement: New Evidence from Population Variation», *Quarterly Journal of Economics*, 115(4): 1239-1285.
- Huber, P. J. (1967): «The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions», *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233, Berkeley, University of California Press.
- Imbens, Guido W., y Joshua D. Angrist (1994): «Identification and Estimation of Local Average Treatment Effects», *Econometrica*, 62: 467-476.
- Johansen, Søren (1988): «Statistical Analysis of Cointegrating Vectors», *Journal of Economic Dynamics and Control*, 12: 231-254.
- Jones, Stephen R. G. (1992): «Was There a Hawthorne Effect?», *American Journal of Sociology*, 98(3): 451-468.
- Kremer, Michael; Edward Miguel, y Rebecca Thornton (2009): «Incentives to Learn», *The Review of Economics and Statistics*, 91: 437-456.
- Krueger, Alan B. (1999): «Experimental Estimates of Education Production Functions», *Quarterly Journal of Economics*, 14(2): 497-562.
- Ladd, Helen (1998): «Evidence on Discrimination in Mortgage Lending», *Journal of Economic Perspectives*, 12(2), spring: 41-62.
- Levitt, Steven D. (1996): «The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation», *Quarterly Journal of Economics*, 111(2): 319-351.
- Levitt, Steven D., y Jack Porter (2001): «How Dangerous Are Drinking Drivers?», *Journal of Political Economy*, 109(6): 1198-1237.
- List, John (2003): «Does Market Experience Eliminate Market Anomalies», *Quarterly Journal of Economics*, 118(1): 41-71.
- Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.
- Maddala, G. S., y In-Moo Kim (1998): *Unit Roots, Cointegration, and Structural Change*, Cambridge, Cambridge University Press.
- Madrian, Brigitte C., y Dennis F. Shea (2001): «The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior», *Quarterly Journal of Economics*, 116(4): 1149-1187.
- Malkiel, Burton G. (2003): *A Random Walk Down Wall Street*, New York, W. W. Norton.
- Manning, Willard G., et al. (1989): «The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?», *Journal of the American Medical Association*, 261(11): 1604-1609.
- Matsudaira, Jordan D. (2008): «Mandatory Summer School and Student Achievement», *Journal of Econometrics*, 142: 829-850.
- McClellan, Mark; Barbara J. McNeil, y Joseph P. Newhouse (1994): «Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?», *Journal of the American Medical Association*, 272(11): 859-866.
- Meyer, Bruce D. (1995): «Natural and Quasi-Experiments in Economics», *Journal of Business and Economic Statistics*, 13(2): 151-161.
- Meyer, Bruce D., W. Kip Viscusi, y David L. Durbin (1995): «Workers' Compensation and Injury Duration: Evidence from a Natural Experiment», *American Economic Review*, 85(3): 322-340.
- Moreira, M. J. (2003): «A Conditional Likelihood Ratio Test for Structural Models», *Econometrica*, 71: 1027-1048.
- Mosteller, Frederick, (1995): «The Tennessee Study of Class Size in the Early School Grades», *The Future of Children: Critical Issues for Children and Youths*, 5(2), summer/fall, 113-127.
- Mosteller, Frederick; Richard Light, y Jason Sachs (1996): «Sustained Inquiry in Education: Lessons from Skill Grouping and Class Size», *Harvard Educational Review*, 66(4), winter, 631-676.

- Mosteller, Frederick, y David L. Wallace (1963): «Inference in an Authorship Problem», *Journal of the American Statistical Association*, 58: 275-309.
- Munnell, Alicia H.; Geoffrey M. B. Tootell; Lynne E. Browne, y James McEneaney (1996): «Mortgage Lending in Boston: Interpreting HMDA Data», *American Economic Review*, 86(1): 25-53.
- Neumark, David, y William Wascher (2000): «Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment», *American Economic Review*, 90(5): 1362-1396.
- Newey, Whitney, y Kenneth West (1987): «A Simple Positive Semi-definite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix», *Econometrica*, 55(3): 703-708.
- Newhouse, Joseph P., et. al. (1993): *Free for All? Lessons from the Rand Health Insurance Experiment*, Cambridge, MA, Harvard University Press.
- Perry, Craig, y Harvey S. Rosen (2004): «The Self-Employed Are Less Likely Than Wage-Earners to Have Health Insurance. So What?», pp. 23-58, en *Public Policy and the Economics of Entrepreneurship*, editado por Douglas Holtz-Eakin y Harvey S. Rosen, Boston, MIT Press.
- Phillips, Peter C. B., y Sam Ouliaris (1990): «Asymptotic Properties of Residual Based Tests for Cointegration», *Econometrica*, 58(1): 165-194.
- Porter, Robert (1983): «A Study of Cartel Stability: The Joint Executive Committee, 1880-1886», *Bell Journal of Economics*, 14(2): 301-314.
- Quandt, Richard (1960): «Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes», *Journal of the American Statistical Association*, 55(290): 324-330.
- Rauscher, Frances; Gordon L. Shaw, y Katherine N. Ky (1993): «Music and Spatial Task Performance», *Nature*, 365(6447): 611.
- Roll, Richard (1984): «Orange Juice and Weather», *American Economic Review*, 74(5): 861-880.
- Rosenzweig, Mark R., y Kenneth I. Wolpin (2000): «Natural “Natural Experiments” in Economics», *Journal of Economic Literature*, 38(4): 827-874.
- Rouse, Cecilia (1995): «Democratization or Diversion? The Effect of Community Colleges on Educational Attainment», *Journal of Business and Economic Statistics*, 12(2): 217-224.
- Ruhm, Christopher J. (1996): «Alcohol Policies and Highway Vehicle Fatalities», *Journal of Health Economics*, 15(4): 435-454.
- Ruud, Paul (2000): *An Introduction to Classical Econometric Theory*, New York, Oxford University Press.
- Shadish, William R.; Thomas D. Cook, y Donald T. Campbell (2002): *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston, Houghton Mifflin.
- Shiller, Robert J. (2005): *Irrational Exuberance* (2.^a ed.), Princeton, NJ, Princeton University Press.
- Sims, Christopher A. (1980): «Macroeconomics and Reality», *Econometrica*, 48(1): 1-48.
- Stock, James H. (1994): «Unit Roots, Structural Breaks, and Trends», cap. 46, en *Handbook of Econometrics*, vol. IV, editado por Robert Engle y Daniel McFadden, Amsterdam, Elsevier.
- Stock, James H., y Francesco Trebbi (2003): «Who Invented Instrumental Variable Regression?», *Journal of Economic Perspectives*, 17: 177-194.
- Stock, James H., y Mark W. Watson (1988): «Variable Trends in Economic Time Series», *Journal of Economic Perspectives*, 2(3): 147-174.
- Stock, James H., y Mark W. Watson (1993): «A Simple Estimator of Cointegrating Vectors in Higher-Order Integrated Systems», *Econometrica*, 61(4): 783-820.
- Stock, James H., y Mark W. Watson (2001): «Vector Autoregressions», *Journal of Economic Perspectives*, 15(4), fall, 101-115.
- Stock, James H., y Motohiro Yogo (2005): «Testing for Weak Instruments in Linear IV Regression», Cap. 5, en *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, editado por Donald W. K. Andrews y James H. Stock, Cambridge, Cambridge University Press.
- Tobin, James (1958): «Estimation of Relationships for Limited Dependent Variables», *Econometrica*, 26(1): 24-36.
- Wagenaar, Alexander C.; Matthew J. Salois, y Kelli A. Komro (2009): «Effects of Beverage Alcohol Price and Tax Levels on Drinking: A Meta-Analysis of 1003 Estimates from 112 Studies», *Addiction*, 104: 179-190.
- Watson, Mark W. (1994): «Vector Autoregressions and Cointegration», cap. 47, en *Handbook of Econometrics*, vol. IV, editado por Robert Engle y Daniel McFadden, Amsterdam, Elsevier.
- White, Halbert (1980): «A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity», *Econometrica*, 48: 827-838.
- Winner, Ellen, y Monica Cooper (2000): «Mute Those Claims: No Evidence (Yet) for a Causal Link Between Arts Study and Academic Achievement», *Journal of Aesthetic Education*, 34(3-4): 11-76.
- Wooldridge, Jeffrey (2002): *Economic Analysis of Cross Section and Panel Data*, Cambridge, MA, MIT Press.
- Wright, Philip G. (1915): «Moore’s Economic Cycles», *Quarterly Journal of Economics*, 29: 631-641.
- Wright, Philip G. (1928): *The Tariff on Animal and Vegetable Oils*, New York, MacMillan.
- Young, Douglas J., y Agnieszka Bielinska-Kwapisz (2006): «Alcohol Prices, Consumption, and Traffic Fatalities», *Southern Economic Journal*, 72: 690-703.

Glosario

Abandono: Véase *deserción*.

Aceptación, región de: Conjunto de valores de un contraste estadístico para los cuales se acepta la hipótesis nula (no se rechaza).

ADL(p, q): Véase *ARD (p, q)*.

AIC: Véase *criterio de información*.

AR(p): Véase *autorregresión*.

ARCH: Véase *heterocedasticidad condicional autorregresiva*.

ARD(p, q): Véase *modelo autorregresivo de retardos distribuidos*.

Asimetría: Medida de forma (asimetría) de una distribución probabilística.

Atípico (outlier): Valor excepcionalmente alto o bajo de una variable aleatoria.

Ausencia de estacionariedad (no estacionariedad): Si la distribución conjunta de una variable de serie temporal y sus retardos cambia en el tiempo.

Autocorrelación: Correlación entre una variable de serie temporal y su retardo. La autocorrelación j -ésima de Y es la correlación entre Y_t e Y_{t-j} .

Autocovarianza: Covarianza entre una variable de serie temporal y su retardo. La covarianza j -ésima de Y es la correlación entre Y_t e Y_{t-j} .

Autorregresión: Modelo de regresión lineal que relaciona una variable de serie temporal consigo misma en el pasado (es decir, con sus retardos). La autorregresión con p retardos como regresores se denomina AR(p).

BIC: Véase *criterio de información*.

BLUE: Véase *estimador lineal insesgado óptimo (ELIO)*.

Causalidad simultánea: Si, además de la relación causal de interés de X hacia Y , existe una relación causal de Y hacia X . La causalidad simultánea genera que X esté correlacionada con el término de error en la regresión poblacional de interés.

Coeficiente de correlación: Véase *correlación*.

Coeficiente de correlación muestral (correlación muestral):

Estimador de la correlación entre dos variables aleatorias.

Coeficiente de determinación: Véase R^2 .

Coeficientes poblacionales: Véase *intercepto y pendientes poblacionales*.

Cointegración: Cuando dos o más variables de series temporales comparten una tendencia estocástica común.

Conjunto de confianza al 95 %: Conjunto de confianza con el 95 % de nivel de confianza. Véase *intervalo de confianza*.

Consistencia: Propiedad de los estimadores consistentes. Véase *estimador consistente*.

Contraste de causalidad de Granger: Procedimiento para contrastar si los valores corrientes y pasados de una serie temporal sirven para predecir valores de otras series temporales.

Contraste de Chow: Contraste para la detección de cambio estructural en una regresión de series temporales con punto de ruptura conocido.

Contraste de Dickey-Fuller: Método de contraste de la presencia de raíz unitaria en un proceso autorregresivo de primer orden [AR(1)].

Contraste de Dickey-Fuller aumentado (ADF): Contraste de raíz unitaria en un modelo AR(p) basado en la regresión.

Contraste de hipótesis: Procedimiento para utilizar la evidencia muestral para ayudar a determinar si una hipótesis específica sobre la población es verdadera o falsa.

Contraste para la diferencia de medias: Procedimiento para contrastar si dos poblaciones tienen la misma media.

Convergencia en distribución: Convergencia de una sucesión de distribuciones a un límite; la definición precisa está dada en la Sección 17.2.

Convergencia en probabilidad: Convergencia de una sucesión de variables aleatorias a un valor determinado; por ejemplo, cuando la media muestral se hace más próxima a la media

poblacional al aumentar el tamaño muestral. Véase Concepto clave 2.6 y Sección 17.2.

Correlación: Medida del grado del grado de evolución, o variación, conjunta de dos variables aleatorias, no se expresa en términos de unidades. La correlación (o coeficiente de correlación) entre X e Y es $\sigma_{XY}/\sigma_X\sigma_Y$ y se expresa como $\text{corr}(X, Y)$.

Correlación serial: Véase *autocorrelación*.

Covarianza: Medida del grado de evolución conjunta de dos variables aleatorias. La covarianza entre X y Y es el valor esperado $E[(X - \mu_X)(Y - \mu_Y)]$ y se expresa como $\text{cov}(X, Y)$ o como σ_{XY} .

Covarianza muestral: Estimador de la covarianza entre dos variables aleatorias.

Criterio de información: Estadístico utilizado para estimar el número de variables retardadas a incluir en un modelo de autorregresión o de retardos distribuidos. Como ejemplo, los más utilizados son el criterio de información de Akaike (AIC) y el criterio de información de Bayes (BIC).

Criterio de Información de Akaike (AIC): Véase *criterio de información*.

Criterio de información de Bayes: Véase *criterio de información*.

Cuasi experimento: Circunstancia en la que la aleatoriedad se introduce mediante variaciones en las circunstancias de los individuos que hacen que parezca *como si* el tratamiento se asignara aleatoriamente.

Cumplimiento parcial: Sucede si algunos participantes rechazan seguir el protocolo de tratamiento en un experimento aleatorio.

Curtosis o kurtosis: Medida acerca de la acumulación de masa probabilística en las colas de una distribución de probabilidad.

Datos de panel: Datos para múltiples individuos en los que a cada individuo le corresponden observaciones para dos o más períodos en el tiempo.

Datos de sección cruzada: Datos recogidos para distintos individuos en un único periodo de tiempo.

Datos de sección cruzada repetida: Colección de conjuntos de datos de sección cruzada en la que cada conjunto de datos de sección cruzada se refiere a un periodo de tiempo distinto.

Datos de series temporales: Datos sobre el mismo individuo para diferentes períodos temporales.

Datos experimentales: Datos obtenidos de un experimento diseñado para evaluar un tratamiento o una política o para investigar efectos causales.

Datos longitudinales: Véase *datos de panel*.

Datos observacionales: Datos procedentes de observaciones, o medidas, comportamientos reales fuera del ámbito del experimento.

Deserción: Pérdida de sujetos de un estudio tras la asignación de un tratamiento o grupo de control.

Desviación típica: Raíz cuadrada de la varianza. La desviación típica de la variable aleatoria Y , expresada como σ_Y , se expresa en las mismas unidades que Y y es una medida de la dispersión de Y en torno a su media.

Desviación típica muestral: Estimador de la desviación típica de una variable aleatoria.

Diagrama de dispersión: Gráfico de dispersión de las n observaciones de X_i e Y_i en el que cada observación está representada por el punto (X_i, Y_i) .

Distribución asintótica: Distribución muestral aproximada de una variable aleatoria calculada para una muestra grande. Por ejemplo, la distribución asintótica de la media muestral es normal.

Distribución asintótica normal: Distribución normal que aproxima la distribución muestral de un estadístico calculado a partir de una muestra grande.

Distribución chi-cuadrado: Distribución de la suma de m variables aleatorias normales estándar independientes al cuadrado. Al parámetro m se le denomina grados de libertad de la distribución chi-cuadrado.

Distribución condicional: Distribución de probabilidad de una variable aleatoria dado que otra variable aleatoria toma un valor particular.

Distribución de Bernouilli: Distribución de probabilidad de una variable aleatoria de Bernouilli.

Distribución de probabilidad: Para una variable aleatoria discreta, lista de todos los valores que la variable aleatoria puede tomar y la probabilidad asociada a que tome cada uno de esos valores.

Distribución de probabilidad acumulada: Función que expresa la probabilidad de que una variable aleatoria sea menor o igual que un valor determinado.

Distribución de probabilidad conjunta: Distribución de probabilidad que determina las probabilidades de los sucesos que implican a dos o más variables aleatorias.

Distribución de probabilidad marginal: Otro nombre para la función de probabilidad de una variable aleatoria Y , que distingue la distribución aislada de Y (distribución marginal) de la distribución conjunta de Y y otra variable aleatoria.

Distribución exacta (para muestras finitas): Distribución probabilística exacta de una variable aleatoria.

Distribución $F_{m,\infty}$: Distribución de una variable aleatoria con distribución chi-cuadrado con m grados de libertad dividida por m .

Distribución $F_{m,n}$: Distribución del cociente de dos variables aleatorias independientes en el que el numerador es una variable aleatoria chi-cuadrado con m grados de libertad dividida por m y el denominador es una variable aleatoria chi-cuadrado con n grados de libertad dividida por n .

Distribución muestral: Distribución de un estadístico para todas las muestras posibles; la distribución surge del cálculo repetido del estadístico a partir de muestras obtenidas aleatoriamente de la misma población.

Distribución normal: Distribución de una variable aleatoria continua con forma de campana comúnmente utilizada.

Distribución normal bivariante: Generalización de la distribución normal que describe la distribución conjunta de dos variables aleatorias.

Distribución normal estándar: Distribución normal con media 0 y varianza igual a 1, se expresa como $N(0, 1)$.

Distribución t: Véase *distribución t de Student*.

Distribución t de Student: La distribución t de Student con m grados de libertad es la distribución del cociente de una variable normal estándar dividida por la raíz cuadrada de una variable independientemente distribuida chi-cuadrado de m grados de libertad dividida por m . A medida que m aumenta, la distribución t de Student converge hacia una distribución normal estándar.

Ecuaciones simultáneas: Véase *causalidad simultánea*.

Efecto causal dinámico: Efecto causal de una variable sobre los valores actuales y futuros de otra variable.

Efecto causal promedio: Media poblacional de los efectos causales individuales en una población heterogénea. Otra posible denominación es efecto promedio de tratamiento.

Efecto del tratamiento: Efecto causal en un experimento o un cuasi experimento. Véase *efecto causal*.

Efecto del tratamiento promedio local (LATE): Estimación del efecto de tratamiento promedio ponderado, por ejemplo, por mínimos cuadrados en dos etapas.

Efecto experimental: Efecto por el que el comportamiento de los sujetos sometidos a un experimento cambia, por el hecho de participar en él.

Efecto Hawthorne: Véase efecto experimental.

Efecto impacto: Efecto contemporáneo, o inmediato, de un cambio unitario en una variable de series temporales X_t , sobre Y_t .

Efecto parcial: Efecto sobre Y del cambio en uno de los regresores manteniendo constante el resto.

Efectos causales: Efecto esperado de una intervención o tratamiento determinado en un experimento aleatorio controlado y realizado bajo condiciones ideales.

Efectos fijos temporales: Véase *efectos temporales*.

Efectos temporales: Variables binarias que indican el periodo temporal en una regresión de datos de panel.

Elasticidad precio de la demanda: Cambio porcentual en la cantidad demandada resultante de un aumento del precio en un 1 %.

Error cuadrático medio de predicción (ECMP): Raíz cuadrada de la media de los cuadrados de los errores de predicción.

Error de especificación de la forma funcional: Cuando la forma de la función de regresión estimada no coincide con la forma funcional de la regresión poblacional; por ejemplo cuando se utiliza una forma funcional lineal y la función de regresión poblacional verdadera es cuadrática.

Error de predicción: Diferencia entre el valor que toma en realidad una variable y su valor estimado.

Error estándar de la regresión (ESR): Estimador de la desviación típica del error de la regresión u .

Error estándar de un estimador: Estimador de la desviación estándar del estimador.

Error estándar heterocedástico-robusto: Error estándar para el estimador de mínimos cuadrados ordinarios que es apropiado tanto si el término de error es homocedástico como si es heterocedástico.

Error estándar válido con homocedasticidad: Error estándar del estimador de mínimos cuadrados ordinarios apropiado solamente si el término de error es homocedástico.

Error tipo I: En contraste de hipótesis, el error cometido cuando la hipótesis nula es verdadera pero se rechaza.

Error tipo II: En contraste de hipótesis, el error cometido cuando la hipótesis nula es falsa pero no se rechaza.

Errores estándar (HAC): Véase errores estándar consistentes a heterocedasticidad y autocorrelación (HAC).

Errores estándar agrupados: Método para el cálculo de errores estándar apropiados para datos de panel.

Errores estándar consistentes a heterocedasticidad y autocorrelación (HAC): Errores estándar de los estimadores de mínimos cuadrados ordinarios que son consistentes tanto si los errores presentan heterocedasticidad y autocorrelación como si no.

Especificación base: Especificación de la regresión que sirve de punto de partida o referencia y que incluye un conjunto de regresores elegidos a partir de una combinación entre el juicio experto, teoría económica y el conocimiento del método de recogida de los datos.

Especificación de la regresión: Descripción de la regresión que incluye el conjunto de regresores y las transformaciones no lineales que han sido aplicadas.

Esperanza: Véase *valor esperado*.

Esperanza condicional: Valor esperado de una variable aleatoria dado un valor determinado tomado por otra variable aleatoria.

Estacionariedad: Cuando la distribución conjunta de una variable de series temporales y sus valores retardados no cambia en el tiempo.

Estadísticamente no significativo: La hipótesis nula (habitualmente, que los coeficientes de la regresión son iguales a cero) no puede rechazarse al nivel de significación establecido.

Estadísticamente significativo: La hipótesis nula (habitualmente, que los coeficientes de la regresión son iguales a cero) se rechaza al nivel de significación establecido.

Estadístico F: Estadístico utilizado para la contrastación de forma conjunta de hipótesis referentes a más de un coeficiente de regresión.

Estadístico F válido con homocedasticidad: Forma del estadístico F válida tan solo si se cumple el supuesto de error de regresión homocedástico.

Estadístico J: Estadístico utilizado para la contrastación de restricciones sobreidentificadas en una regresión de variables instrumentales.

Estadístico t: Estadístico utilizado para la contrastación de hipótesis. Véase Concepto clave 5.1.

Estadístico t heterocedástico-robusto: estadístico *t* construido utilizando un error estándar heterocedástico-robusto.

Estandarización de una variable aleatoria: Operación consistente en sustraer la media y dividir por la desviación típica, esto genera una variable aleatoria con media 0 y desviación típica 1. El valor estandarizado de Y es $(Y - \mu_Y)/\sigma_Y$.

Estimación: Valor numérico calculado para un estimador a partir de los datos de una muestra específica.

Estimador consistente: Estimador que converge en probabilidad al verdadero valor del parámetro a estimar.

Estimador de diferencias: Estimador del efecto causal expresado como la diferencia en los resultados muestrales medios entre el grupo de tratamiento y el de control.

Estimador de diferencias de las diferencias: Media de cambio en Y para los individuos del grupo de tratamiento menos la media de cambio para los individuos del grupo de control.

Estimador de mínimos cuadrados no lineales: Estimador obtenido al minimizar la suma de los residuos al cuadrado cuando la función de regresión no es lineal en los parámetros.

Estimador insesgado: Estimador con sesgo igual a cero.

Estimador lineal insesgado óptimo (ELIO): Estimador que presenta la menor varianza de entre todos los que están calculados a partir de funciones lineales de los valores muestrales de Y . Bajo las condiciones de Gauss-Markov, el estimador de mínimos cuadrados ordinarios es el estimador lineal insesgado óptimo de los coeficientes de la regresión condicionado a los valores de los regresores.

Estimador máximo verosímil (EMV): Estimador de parámetros desconocidos obtenido de maximizar la función de verosimilitud; véase Apéndice 11.2.

Estimador MCO: Véase *estimador de mínimos cuadrados ordinarios*.

Estimador mínimo cuadrático: Estimador obtenido al minimizar la suma de los residuos al cuadrado.

Estimadores: Función de una muestra de datos obtenidos aleatoriamente a partir de una población. Un estimador es un procedimiento para el cálculo de un valor de un parámetro poblacional a partir de datos muestrales, tales como la media poblacional, de forma fundamentada.

Estimadores de mínimos cuadrados ordinarios: Estimadores del intercepto y de la(s) pendiente(s) de la regresión que minimizan la suma de los residuos al cuadrado.

Evaluación de programas: Disciplina de estudio encargada de estimar los efectos de un programa, política, o cualquier otra intervención o «tratamiento».

Exogeneidad estricta: Requisito de que el término de error de la regresión tenga media cero condicionada a los valores presentes, pasados y futuros de los regresores en un modelo de retardos distribuidos.

Experimento aleatorizado controlado: Experimento en el que los participantes se asignan de manera aleatoria al grupo de control, que no recibe tratamiento, o al grupo de tratamiento, que sí lo recibe.

Experimento natural: Véase *cuasi experimento*.

Función de densidad de probabilidad (f.d.p.): Para una variable aleatoria continua, el área bajo la función de densidad probabilística entre dos puntos cualesquiera es la probabilidad de que la variable aleatoria se encuentre entre esos dos puntos.

Función de distribución acumulada: Véase *distribución de probabilidad acumulada*.

Función de regresión lineal: Función de regresión con pendiente constante.

Función de regresión no lineal: Función de regresión con pendiente no constante.

GARCH: Véase *heterocedasticidad condicional autorregresiva generalizada*.

Grupo de control: Grupo que no recibe tratamiento o intervención en un experimento.

Grupo de tratamiento: Grupo que recibe el tratamiento o intervención en un experimento.

Heterocedasticidad: Situación en la que la varianza del término de error de la regresión u_i , condicionado a los regresores, no es constante.

Heterocedasticidad condicional: Varianza, habitualmente de un término de error, dependiente de otras variables.

Heterocedasticidad condicional autorregresiva (ARCH): Modelo de series temporales de heterocedasticidad condicional.

Heterocedasticidad condicional autorregresiva generalizada (GARCH): Modelo de series temporales para la heterocedasticidad condicional.

Hipótesis alternativa: La hipótesis que se supone verdadera si la hipótesis nula es falsa. La hipótesis nula se denomina frecuentemente H_1 .

Hipótesis alternativa bilateral: Si, bajo la hipótesis alternativa, el parámetro de interés no es igual al valor dado por la hipótesis nula.

Hipótesis alternativa unilateral: Si el valor del parámetro de interés se puede encontrar solamente a uno de los lados, solo puede ser mayor (o solo menor) del valor establecido en la hipótesis nula.

Hipótesis conjunta: Hipótesis consistente en dos o más hipótesis, es decir, que incluye más de una restricción acerca de los parámetros del modelo.

Hipótesis nula: Hipótesis a contrastar en un contraste de hipótesis, a menudo se expresa como H_0 .

Homocedasticidad: Si la varianza del término de error u_i , condicionada a los regresores, es constante.

$I(0)$, $I(1)$, y $I(2)$: Véase *orden de integración*.

Idénticamente distribuidas: Si dos o más variables aleatorias tienen la misma función de distribución.

Identificación exacta: Se produce si el número de variables instrumentales iguala el número de regresores endógenos.

Incorrelación serial: Si la variable de serie temporal presenta todas las autocorrelaciones nulas.

Incorrelacionado: Dos variables aleatorias están incorrelacionadas si su correlación es cero.

Independencia: Cuando el conocimiento del valor tomado por una variable aleatoria no proporciona información acerca del valor que puede tomar otra variable aleatoria. Dos variables aleatorias son independientes si su distribución conjunta es el producto de sus distribuciones marginales.

Independencia en media condicional: Esperanza condicional del error de regresión u_i dados los regresores, dependiente de alguno de los regresores pero no de todos.

Independientes e idénticamente distribuidas (i.i.d.): Cuando dos o más variables aleatorias independientes tienen la misma distribución.

Instrumento: Véase *variable instrumental*.

Intercepto: El valor de β_0 en un modelo de regresión lineal.

Intercepto y pendientes poblacionales: Valores verdaderos o poblacionales de β_0 (intercepto) y β_1 (pendiente) en el modelo simple de regresión. En regresión múltiple, existen múltiples coeficientes de pendiente ($\beta_1, \beta_2, \dots, \beta_k$), uno por cada regresor.

Intervalo de confianza (conjunto de confianza): Intervalo (o conjunto de valores) que contiene el verdadero valor de un parámetro poblacional con una probabilidad preestablecida cuando se calcula sobre muestras repetidas.

Intervalo de predicción: Intervalo que contiene el valor futuro de una variable de serie temporal con una probabilidad preestablecida.

Instrumentos débiles: Variables instrumentales que tienen baja correlación con los regresores endógenos.

Ley de esperanzas iteradas: Resultado de teoría de probabilidad que dice que el valor esperado de Y es el valor esperado de su esperanza condicional dado X , es decir, $E(Y) = E[E(Y|X)]$.

Ley de los grandes números: De acuerdo con este resultado de teoría de probabilidad, bajo ciertas condiciones generales, la media muestral se acerca a la media poblacional con probabilidad muy alta si el tamaño muestral es grande.

Logaritmo: Función matemática definida para un argumento positivo; su pendiente siempre es positiva pero tiende a cero. El logaritmo natural es la inversa de la función exponencial; esto es, $X = \ln(e^X)$.

Logaritmo natural: Véase *logaritmo*.

MC2E: Véase *mínimos cuadrados en dos etapas*.

MCG factibles: Versión del estimador de mínimos cuadrados generalizados (MCG) que utiliza un estimador de la varianza condicional de los errores de la regresión y la covarianza entre los regresores y diferentes observaciones.

Media: Valor esperado de una variable aleatoria. La media de Y se expresa como μ_Y .

Media condicional: Media de una distribución condicional. Véase *esperanza condicional*.

Método generalizado de momentos (MGM): Método de estimación de los parámetros mediante ajuste de los momentos muestrales a los momentos poblacionales que son función de parámetros desconocidos. Los estimadores de variables instrumentales son un caso particular importante.

MGM: Véase *método generalizado de momentos*.

Mínimos cuadrados en dos etapas: Estimador de variables instrumentales, descrito en el Concepto clave 12.2.

Mínimos cuadrados generalizados (MCG): Generalización de los mínimos cuadrados ordinarios apropiado cuando los errores de la regresión tienen una forma de heterocedasticidad conocida (en cuyo caso MCG también se conoce como mínimos cuadrados ponderados, MCP) o una forma conocida de correlación serial.

Mínimos cuadrados no lineales: El análogo de mínimos cuadrados ordinarios que se aplica cuando la función de regresión es una función no lineal de los parámetros desconocidos.

Mínimos cuadrados ponderados (MCP): Alternativa a los mínimos cuadrados ordinarios que puede utilizarse cuando el error de regresión es heterodédástico y la forma de la heterocedasticidad es conocida o puede ser estimada.

Mínimos cuadrados ponderados factibles: Versión del estimador de mínimos cuadrados ponderados (MCP) que utiliza un estimador de la varianza condicional de los errores de la regresión.

Modelo autorregresivo de retardos distribuidos (ARD): Modelo de regresión lineal en el que la variable de serie temporal Y_t se expresa en función de los retardos de Y_t y de otra variable X_t . El modelo se denomina ARD(p, q), donde p es el número de retardos de la variable Y_t , y q el número de retardos de la variable X_t .

Modelo de probabilidad lineal: Modelo de regresión en el que la variable dependiente Y es una variable binaria.

Modelo de regresión cuadrático: Función de regresión no lineal que incluye X y X^2 como regresores.

Modelo de regresión cúbico: Función de regresión no lineal que incluye X , X^2 y X^3 como regresores.

Modelo de regresión de efectos fijos temporales e individuales: Regresión de datos de panel que incluye tanto efectos fijos individuales como efectos fijos temporales.

Modelo de regresión log-lineal: Función de regresión no lineal en el que la variable dependiente es $\ln(Y)$ y la independiente es X .

Modelo de regresión múltiple: Extensión del modelo de regresión de variable única que permite que Y pueda depender de k regresores.

Modelo de regresión poblacional múltiple: Modelo de regresión múltiple en el Concepto clave 6.2.

Modelo de regresión polinomial: Regresión no lineal que incluye como regresores X , X^2 , ..., y X^r , donde r es número entero.

Modelo de retardos distribuidos: Modelo de regresión en el que los regresores son variables actuales y retardadas de X .

Modelo lineal-log: Función de regresión no lineal en el que la variable dependiente es Y y la independiente es $\ln(X)$.

Modelo log-log: Función de regresión no lineal en el que la variable dependiente es $\ln(Y)$ y la independiente es $\ln(X)$.

Momentos de una distribución: Valor esperado de una variable aleatoria elevado a diferentes potencias. El r -ésimo momento de la variable aleatoria Y es $E(Y^r)$.

Muestra aleatoria simple: Si los individuos se escogen de manera independiente de una población utilizando un método que asegure que cada individuo presenta la misma probabilidad de ser seleccionado.

Multicolinealidad: Véase *multicolinealidad exacta* y *multicolinealidad imperfecta*.

Multicolinealidad exacta: Si uno de los regresores es combinación lineal exacta del resto de regresores.

Multicolinealidad imperfecta: Si dos o más regresores están altamente correlacionados.

Multiplicador dinámico: El multiplicador dinámico del periodo h -ésimo es el efecto que produce el cambio unitario en la variable X sobre la variale Y_{t+h} .

Multiplicador dinámico acumulativo: Efecto acumulado de una variación unitaria de la variable de serie temporal X so-

bre Y . El multiplicador dinámico acumulativo del periodo h -ésimo es el efecto de un cambio unitario en X_t sobre $Y_t + Y_{t+1} + Y_{t+2} + \dots + Y_{t+h}$.

Multiplicador dinámico acumulativo de largo plazo: Efecto de largo plazo acumulado sobre la variable de serie temporal Y de un cambio en X .

Nivel de confianza: Probabilidad preestablecida de que un intervalo (o conjunto) de confianza contenga el verdadero valor del parámetro.

Nivel de significación: La probabilidad de rechazo preestablecida para un contraste de hipótesis estadístico en el que la hipótesis nula es verdadera.

Número de observación: El único identificativo asignado a cada individuo en un conjunto de datos.

Orden de integración: Número de veces que una variable de serie temporal debe ser diferenciada para convertirla en estacionaria. Una serie temporal integrada de orden d debe ser diferenciada d veces y se expresa como $I(d)$.

Panel equilibrado: Conjunto completo de datos de panel, esto es, en el que se disponen de observaciones de cada variable para cada uno de los períodos.

Panel incompleto: Panel de datos en el que no están disponibles algunos datos.

Parámetros: Constantes que caracterizan una característica o una distribución probabilística o función de regresión de una población.

Paseo aleatorio: Proceso de series temporales en el que el valor de la variable aleatoria es igual a su valor en el periodo anterior más un término de error no predecible.

Paseo aleatorio con constante (deriva): Generalización del paseo aleatorio en el que el cambio en la variable tiene media distinta de cero pero es igualmente impredecible.

Población: Grupo de individuos —tales como personas, compañías o distritos escolares— a estudio.

Potencia: Probabilidad de que un contraste rechace correctamente la hipótesis nula cuando la alternativa es verdadera.

Predicción pseudo fuera-de-la-muestra: Predicción calculada sobre parte de la muestra utilizando un procedimiento *como si* esos datos muestrales no se hubiesen observado todavía.

Primera diferencia: La primera diferencia de una variable de serie temporal Y_t es $Y_t - Y_{t-1}$, se expresa como ΔY_t .

Probabilidad: Proporción de veces que un efecto (o suceso) ocurrirá en el largo plazo.

Punto de ruptura: Instante en el que se produce un cambio discreto en los coeficientes de regresión poblacional en series temporales.

p-valor (significación estadística): Suponiendo cierta la hipótesis nula, es la probabilidad de extraer un valor del estadístico tan adverso hacia el cumplimiento de dicha hipótesis. Asimismo denominado significatividad probabilística mar-

ginal, el p -valor es el menor nivel de significación al cual se rechazaría la hipótesis nula.

R^2 : En una regresión, la proporción de la varianza muestral de la variable dependiente explicada por los regresores.

R^2 ajustado (\bar{R}^2): Versión modificada del R^2 que no aumenta necesariamente al añadir un nuevo regresor a la regresión.

\bar{R}^2 : Véase R^2 ajustado.

Raíz unitaria: Se refiere a una autorregresión cuya raíz más elevada es igual a 1.

Recta de regresión MCO: Recta de regresión con los coeficientes poblacionales reemplazados por los estimadores MCO.

Recta de regresión poblacional: En una regresión simple, la recta de regresión poblacional es $\beta_0 + \beta_1 X_i$, y en una regresión múltiple es $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$.

Región de rechazo: Conjunto de valores de un contraste estadístico para los cuales la hipótesis nula se rechaza.

Regresando: Véase *variable dependiente*.

Regresión de la primera etapa: Regresión de una variable endógena incluida sobre las variables exógenas incluidas, si las hay, y las variables instrumentales en mínimos cuadrados en dos etapas.

Regresión de variables instrumentales (VI): Método de obtención de estimadores consistentes de los coeficientes desconocidos de una función de regresión poblacional cuando el regresor, X , está correlacionado con el término de error u .

Regresión logit: Modelo de regresión no lineal para una variable dependiente binaria en la que la función de regresión poblacional se modeliza utilizando la función de distribución logística acumulada.

Regresión probit: Modelo de regresión no lineal para una variable dependiente binaria en la que la función de regresión poblacional está modelizada de acuerdo a la función de distribución normal estándar acumulada.

Regresión restringida: Regresión en la que los coeficientes están restringidos al cumplimiento de una condición. Por ejemplo, cuando se calcula el estadístico F válido con homocedasticidad, es la regresión con coeficientes restringidos al cumplimiento de la hipótesis nula.

Regresión sin restringir: En el cálculo del estadístico F válido con homocedasticidad, es la regresión que se aplica bajo la hipótesis alternativa de que los coeficientes no están restringidos al cumplimiento de la hipótesis nula.

Regresor: Variable que aparece en el lado derecho de la regresión; variable independiente de la regresión.

Regresor constante: Regresor asociado al intercepto o término constante o independiente de la regresión; este regresor es siempre igual a 1.

Residuos MCO: Diferencia entre Y_i y la recta de regresión mínimocuadrática, se expresa mediante \hat{u}_i en este libro.

Retardos: Valor de una serie temporal en períodos de tiempo anteriores. El j -ésimo retardo de Y_t es Y_{t-j} .

Sesgo: Valor esperado de la diferencia entre el estimador y el parámetro a estimar. Si $\hat{\mu}_y$ es un estimador de μ_y , entonces el sesgo de $\hat{\mu}_y$ es $E(\hat{\mu}_y) - \mu_y$.

Sesgo de selección muestral: Sesgo en el estimador de un coeficiente de regresión que aparece si el proceso de selección está influido por la disponibilidad de los datos y el proceso está correlacionado con la variable dependiente. Este sesgo induce correlación entre uno o más regresores y el término de error.

Sesgo de variable omitida: Sesgo que aparece en un estimador debido a que una variable explicativa de Y , que está correlacionada con un regresor, ha sido omitida de la regresión.

Sesgo por errores de medición en las variables: Sesgo del estimador de una regresión que aparece por la presencia de errores de medida en los regresores.

Sobreidentificación: Si el número de variables instrumentales excede el número de regresores endógenos incluidos.

Subidentificación: Cuando el número de variables instrumentales es menor que el número de regresores endógenos.

Suma de los cuadrados de los residuos (SR): Suma de los cuadrados de los residuos mínimo cuadráticos.

Suma explicada (SE): Suma de las desviaciones al cuadrado de los valores estimados de Y_i , \hat{Y}_i respecto de su media; véase Ecuación (4.14).

Suma total de cuadrados (ST): Suma de las desviaciones al cuadrado de Y_i , respecto de su media.

Supuestos de mínimos cuadrados: Supuestos para el modelo de regresión lineal especificados en el Concepto clave 4.4 (regresión simple) y el Concepto clave 6.4 (regresión múltiple).

Tamaño del contraste: Probabilidad de que un contraste rechace la hipótesis nula de forma incorrecta cuando ésta es cierta.

Tendencia común: Tendencia compartida por dos o más series temporales.

Tendencia determinística: Movimiento persistente en el tiempo a largo plazo de una variable que puede expresarse como una función no estocástica de la variable tiempo.

Tendencia estocástica: Movimiento de una variable en el tiempo persistente a largo plazo pero aleatorio.

Teorema central del límite: Resultado estadístico formal que consiste en que, bajo condiciones determinadas, la distribución muestral de la media muestral estandarizada se approxima por una distribución normal estándar cuando el tamaño muestral es grande.

Teorema de Gauss-Markov: Resultado matemático que establece que, bajo ciertas condiciones, el estimador de mínimos cuadrados, es el estimador óptimo entre los lineales e

insesgados de los coeficientes de la regresión condicionado a los valores de los regresores.

Término constante: Intercepto de la regresión.

Término de error: Diferencia entre Y y la función de regresión poblacional, en este libro se expresa como u .

Término de interacción: Regresor formado por el producto de dos o más regresores, como $X_{1i} \times X_{2i}$.

Trampa de las variables ficticias: Problema causado por la inclusión de un conjunto completo de variables binarias en una regresión junto con un regresor constante (intercepto), lo que conduce a la multicolinealidad perfecta.

***t*-ratio:** Véase *estadístico t*.

Validez externa: Las inferencias y conclusiones de un estudio estadístico son externamente válidas si pueden generalizarse a partir de la población y situación estudiada para otras poblaciones y situaciones.

Validez interna: Si las inferencias sobre los efectos causales de un estudio estadístico son válidas para la población a estudio.

Valor ajustado: Véase *valor de predicción*.

Valor crítico: Valor de un contraste estadístico para el que resulta indiferente rechazar o aceptar la hipótesis nula para un nivel de significación pre establecido.

Valor de predicción: El valor para Y_i estimado por la recta de regresión mínima cuadrática, expresado como \hat{Y}_i en este libro.

Valor esperado: Valor medio de largo plazo de una variable aleatoria si el número de ensayos o repeticiones es elevado. Es la media ponderada por probabilidad de todos los posibles valores que puede tomar una variable aleatoria. El valor esperado de Y se expresa como $E(Y)$ y asimismo se denomina esperanza.

VAR: Véase *vector autorregresivo*.

Variable aleatoria continua: Variable aleatoria que toma valores en un intervalo continuo.

Variable aleatoria de Bernouilli: Variable aleatoria que toma los valores 0 o 1.

Variable aleatoria discreta: Variable aleatoria que toma valores de un conjunto discreto.

Variable binaria: Variable que es 0 o 1. La variable binaria se utiliza para indicar un resultado binario. Por ejemplo, X es una variable binaria (o indicador, o variable ficticia o *dummy*) para el sexo de una persona con valor $X = 0$ si la persona es mujer y con valor $X = 1$ si la persona es hombre.

Variable de control: Regresor que recoge un factor omitido determinado por la variable dependiente.

Variable dependiente limitada: Variable dependiente que toma solamente un número de valores limitado. Por ejemplo, la variable puede ser una variable binaria 0-1 o surgir de uno de los modelos descritos en el Apéndice 11.3.

Variable endógena: Variable que presenta correlación con el término de error.

Variable exógena: Variable incorrelacionada con el término de error de la regresión.

Variable explicativa: Véase *regresor*.

Variable ficticia o dummy: Véase *variable binaria*.

Variable indicador: Véase *variable binaria*.

Variable instrumental: Variable correlacionada con un regresor endógeno (relevancia del instrumento) e incorrelacionada con el término de error de la regresión (exogeneidad del instrumento).

Variable respuesta (Resultados potenciales): Conjunto de resultados que puede obtener un individuo (unidad de tratamiento) tras recibir, o no recibir, un tratamiento experimental.

Variables incluidas endógenas: Regresores que están correlacionados con el término de error (habitualmente en el contexto de regresión de variables instrumentales).

Variables incluidas exógenas: Regresores que están incorrelacionados con el término de error (habitualmente en el contexto de regresión de variables instrumentales).

Varianza: Valor esperado de la diferencia cuadrática entre una variable aleatoria y su media; la varianza de Y se expresa como σ_Y^2 .

Varianza condicional: Varianza de una distribución condicional.

Varianza muestral: Estimador de la varianza de una variable aleatoria.

Vector autorregresivo: Modelo de k variables de series temporales que consiste en k ecuaciones, en las que los regresores de todas las ecuaciones son valores retardados de todas las variables.

Volatilidad agrupada: Si la serie temporal presenta algunos períodos agrupados de varianza elevada y otros períodos agrupados de varianza menor.

Índice

A

Aleatoriedad basada en las covariables, 341
ARCH generalizado (GARCH), 476
 $ARD(p, q)$, 386
Asimetría, 17
Atípico, 18
Autocorrelación, 378
Autocorrelacionada, 261
Autocovarianza j -ésima, 378
Autorregresión, 381

C

Causalidad simultánea, 232
Coeficiente
de autocorrelación, 378
de cointegración, 468
de correlación muestral (correlación muestral), 66
de la pendiente de X_{1i} , 134
de la pendiente de X_{2i} , 134
de X_{1i} , 134
de X_{2i} , 134
que multiplica a D_i , 110
sobre D_i , 110
Coeficientes poblacionales, 79
Cointegración, 468
Condición
de exogeneidad del instrumento, 304
de relevancia del instrumento, 304
Condiciones de Gauss-Markov, 125
para regresión múltiple, 513
Conjunto de confianza, 57
al 95 %, 162
Consistencia, 35
Contraste
de Bonferroni, 177
de causalidad de Granger, 389
de Dickey-Fuller, 399
de hipótesis, 51

de la recepción aleatoria del tratamiento, 342
de sobreidentificación de restricciones, 318
DF-MCG, 465
EG-ADF, 470
para la diferencia entre dos medias, 58
Controlando por X_2 , 134
Convergencia
en distribución, 487
en probabilidad, 35, 485
Correlación, 23
serial, 378
Covarianza, 23
muestral, 66
Criterio de información
de Akaike (AIC), 393
de Bayes (BIC), 392
Cuasi-diferencia, 435
Cuasi-experimento, 352
Cumplimiento parcial, 343
Curtosis, 18

D

Datos
de panel, 8, 249
de sección cruzada, 6
cruzada repetida, 356
de series temporales, 6
experimentales, 5
longitudinales, 8
observacionales, 6
Densidad, 14
Dependencia débil, 389
Deserción o abandono, 343
Desviación típica, 16
muestral, 53
Diagrama de dispersión, 65
Discontinuidad de la regresión, 358
Distribución
asintótica, 34, 487

asintóticamente normal, 37
chi-cuadrado, 28
condicional, 20
de Bernouilli, 13
de probabilidad, 12
acumulada, 12
conjunta, 19
marginal, 20
exacta (muestras finitas), 34
 F , 30
muestral, 32
normal, 26
bivariante, 28
estándar, 26
multivariante, 28
 t , 30
de Student, 28

E

Efecto
causal, 5, 60
dinámico, 421
promedio, 340
del tratamiento, 60
Hawthorne, 343
impacto, 429
parcial, 135
promedio del tratamiento, 340
Efectos fijos
individuales, 255
temporales, 258
Elasticidad, 190
ELIO (Estimador Lineal Insesgado Óptimo), 50
Error
de especificación de la forma funcional, 227
de predicción, 381
estándar consistente a heterocedasticidad y autocorrelación (HAC), 432

estándar de la regresión (ESR), 86
 estándar de \bar{Y} , 54
 estándar de $\hat{\beta}$, 104
 tipo I, 56
 tipo II, 56
Errores estándar
 agrupados, 261
 consistentes a heterocedasticidad y autocorrelación (HAC), 261
 heterocedástico-robustos, 113
 válidos con homocedasticidad, 113
Espacio muestral, 12
Especificación base, 166
Especificaciones alternativas, 166
Esperanza, 15
 condicional, 21
Estacionariedad, 387
Estadístico
 de causalidad de Granger, 389
 de Dickey-Fuller, 399
 aumentado (ADF), 399
 de la razón de verosimilitud de Quandt (QLR), 403
F, 158
 de la primera etapa, 317
 válido con homocedasticidad, 160
J heterocedástico-robusto, 524
J MGM, 524
t, 105
 t (t -radio), 54
Estandarizar variable, (26)
Estimación, 48
 por variables instrumentales del efecto del tratamiento, 343
Estimador, 48
 consistente, 486
 de las diferencias, 341
 con regresores adicionales, 341
 en diferencias, 354
 diferencias con regresores adicionales, 355
 de máxima verosimilitud (EMV), 285
 de mínimos cuadrados, 50
 de varianza de Newey-West, 433
 lineal insesgado óptimo (ELIO), 116
 MCG factible, 437
 MCG infactible, 437
 MCO dinámico (MCOD), 472
 MCP, 493
Estimadores
 de mínimos cuadrados no lineales, 220
 de mínimos cuadrados ordinarios (MCO), 82
 de $\beta_0, \beta_1, \dots, \beta_k$, 137
Evaluación de programas, 339
Exactamente identificado, 311

Exogeneidad, 426
 estricta, 426
Experimento
 aleatorizado controlado, 5
 natural, 352
F
 f.d.p. normal, 499
 bivariante, 500
Forma reducida, 312
Función
 de densidad, 14
 de probabilidad (f.d.p.), 14
 de distribución acumulada (f.d.a.), 13
 de regresión muestral, 82
 de regresión no lineal, 185
 de regresión poblacional, 79, 134
 de verosimilitud, 285
 exponencial, 190
G
Grados de libertad, 54
Grupo
 de control, 5
 de tratamiento, 5
H
Heterocedasticidad
 condicional autorregresiva (ARCH), 476
 y homocedasticidad, 111
Heterocedástico, 136
Hipótesis
 alternativa, 51
 bilateral, 51, 104
 unilateral, 57
 conjunta, 157
 nula, 51, 104
Homocedástico, 136
I
Idénticamente distribuida, 32
Incorrelacionada, 23
Independencia en media condicional, 165
Independiente, 22
Independientemente distribuida, 22
Independientes e idénticamente distribuidas (i.i.d.), 32
Instrumentos débiles, 316
Integrada
 de orden cero [$I(0)$], uno [$I(1)$], o dos [$I(2)$], 464
 de orden, d , $I(d)$, 464

Intercepto
 poblacional, 79
 término independiente, 134
Intervalo
 de confianza, 57
 para β_1 , 108
 de predicción, 390
J
 j -ésimo retardo, 375
L
Leptocútica, 19
Ley
 de esperanzas iteradas, 22
 de los grandes números, 34
Logaritmo natural, 190
Logit, 279
M
 Manteniendo X_2 constante, 134
Matriz de covarianzas, 532
MCG
 factibles, 517
 infactibles, 517
MCP
 factible, 494
 infactible, 493
Media, 15
 condicional, 21
 muestral, 32
Método generalizado de momentos (MGM), 522
MGM eficientes, 523
Mínimos cuadrados
 en dos etapas, 304
 generalizados (MCG), 434, 515
 no lineales, 220
 ponderados (MCP), 117, 492
Modelo
 autorregresivo de orden p -ésimo [$AR(p)$], 382
 de retardos distribuidos (ARD), 386
 clásico de error de medición, 229
 de probabilidad lineal, 277
 de regresión con interacción, 197
 de regresión cuadrático, 183
 de regresión cúbico, 188
 de regresión de efectos fijos, 254
 individuales y temporales, 258
 de regresión lineal con un regresor único, 79
 de regresión múltiple, 134
 poblacional, 135

- de regresión polinomial, 188
de retardos distribuidos, 425
del vector de corrección de error, 469
lineal-log, 191
log-lineal, 192
log-log, 193
- M**
Momento
 de una distribución, 17
 r-ésimo, 19
- Muestreo aleatorio simple, 31
- Multicolinealidad
 imperfecta, 145
 perfecta, 141
- Multiplicador dinámico, 429
 acumulativo, 429
 de largo plazo, 429
- N**
Nivel
 de confianza, 57, 108
 de significación, 56
- Número de observación, 6
- O**
Orden de integración, 464
- P**
Panel
 equilibrado, 250
 incompleto, 250
- Parámetro, 79
 de truncamiento, 432
- Paseo aleatorio, 396
 con deriva, 396
- Pendiente poblacional, 79
- Población, 31
 de interés, 223
 estudiada, 223
- Potencia de un contraste, 56
- Predicción**
 multiperiodo directa, 462
 multiperiodo iterada AR, 461
 multiperiodo iterada VAR, 461
 pseudo fuera de la muestra, 407
- Primer retardo, 375
- Primera diferencia, 375
- Probabilidad, 12
 de cobertura, 58
- Probit, 279
- Promedio
 local del efecto del tratamiento, 362
 muestral, 32
- Proporción correctamente estimada, 286
- Pseudo- R^2 , 287
- Punto de ruptura, 403
p-valor, 105
 probabilidad de significación, 52
- R**
 R^2 ajustado (R^2), 140
 R^2 de la regresión, 85
 R^2 , 139
- Raíz
 del error cuadrático medio de la predicción (RECM), 382
 unitaria, 397
- Recta de regresión
 MCO, 82, 137
 muestral, 82
 poblacional, 79, 134
- Región
 de aceptación, 56
 de rechazo, 56
- Regresión
 de la primera etapa, 313
 de la segunda etapa, 313
 de variables instrumentales (VI), 303
 espuria, 398
 logística, 279
 restringida, 160
 sin restringir, 160
- Regresor, 79
 constante, 135
 de interacción, 197
- Residuo, 82
 MCO, 137
- Restricciones, 157
- Resultados, 12
- S**
Segunda diferencia, 463
Serialmente correlacionada, 261
- Sesgo
 consistencia y eficiencia, 49
 de ecuaciones simultáneas, 232
 de selección muestral, 230
 de variable omitida, 130
 por errores en las variables, 228
- Sobreidentificado, 311
- Subidentificado, 311
- Suceso, 12
- Suma
 de los cuadrados de los residuos (SR), 86
 explicada de cuadrados (SE), 86
 total de cuadrados (ST), 86
- Supuestos
- de la regresión normal homocedástica, 117
de mínimos cuadrados, 87
- T**
Tamaño de un contraste, 56
- Tendencia, 395
 común, 468
 determinística, 395
 estocástica, 395
- Teorema
 central del límite, 36
 de Gauss-Markov, 116
 para regresión múltiple, 513
 de la función continua, 488
 de Slutsky, 488
- Término
 constante, 135
 de corrección de error, 469
 de error, 79
 de interacción, 197
- Trampa de la variable ficticia, 145
- V**
Validez
 externa, 224
 interna, 224
- Valor
 crítico, 56
 de predicción, 82, 137
 esperado, 15
- Variable
 aleatoria continua, 12
 aleatoria de Bernouilli, 13
 aleatoria discreta, 12
 de control, 164
 dependiente, 79
 limitada, 275
 endógena, 304
 exógena, 304
 ficticia o dummy, 109
 independiente, 79
 indicador, 109
- Variables
 exógenas incluidas, 311
 instrumentales (instrumento), 303
 respuesta, 340
- Varianza, 16
 condicional, 22
 muestral, 53
- Vector
 autorregresivo (VAR), 456
 de medias, 532
- Volatilidad agrupada, 475

Valores críticos de muestra para el estadístico t a partir de la distribución normal estándar

	Nivel de significación		
	10 %	5 %	1 %
Contraste bilateral (≠)			
Rechazar si $ t $ es mayor que	1,64	1,96	2,58
Contraste unilateral (>)			
Rechazar si t es mayor que	1,28	1,64	2,33
Contraste unilateral (<)			
Rechazar si t es menor que	1,28	-1,64	-2,33

Valores críticos de muestra para el estadístico F a partir de la distribución $F_{m,\infty}$

Grados de libertad	Rechazar si $F >$ Valor crítico		
	Nivel de significación	10 %	5 %
	1 %		
1	2,71	3,84	6,63
2	2,330	3,00	4,61
3	2,08	2,60	3,78
4	1,94	2,37	3,32
5	1,85	2,21	3,02
6	1,77	2,10	2,80
7	1,72	2,01	2,64
8	1,67	1,94	2,51
9	1,63	1,88	2,41
10	1,60	1,83	2,32
11	1,57	1,79	2,25
12	1,55	1,75	2,18
13	1,52	1,72	2,13
14	1,50	1,69	2,08
15	1,49	1,67	2,04
16	1,47	1,64	2,00
17	1,46	1,62	1,97
18	1,44	1,60	1,93
19	1,43	1,59	1,90
20	1,42	1,57	1,88
21	1,41	1,56	1,85
22	1,40	1,54	1,83
23	1,39	1,53	1,81
24	1,38	1,52	1,79
25	1,38	1,51	1,77
26	1,37	1,50	1,76
27	1,36	1,49	1,74
28	1,35	1,48	1,72
29	1,35	1,47	1,71
30	1,34	1,46	1,70