

A vertical line of small, dark gray dots runs down the left side of the page, starting from the top of the text area and ending near the bottom.

ELEMENTOS DE

# ECONOMETRÍA

APLICACIONES PARA CUBA

Esta edición ha sido posible gracias al programa de "Formación en Economía para Cuba"

---

© Renato Aguilar

© Anicia García

© Pavel Vidal

© sobre la presente edición:

Centro de Estudios de la Economía Cubana de la Universidad de la Habana, 2011

Avenida 41 N° 707 Esquina a 9na. Miramar, Ciudad de la Habana, Cuba

Teléfonos: 202 1391 y 209 0563

[www.uh.cu/centros/ceec/index800.htm](http://www.uh.cu/centros/ceec/index800.htm)

ISBN xxxxxxxxxxxxxxxxxxxxxx

Departamento de Economía  
Facultad de Ciencias Sociales  
Universidad de la República  
Constituyente 1502 / Piso 6  
Montevideo/ URUGUAY

Tel: +(598) 24106449

Fax: +(598) 24106450

E-mail: [pub@decon.edu.uy](mailto:pub@decon.edu.uy)

[www.decon.edu.uy](http://www.decon.edu.uy)

---

Diseño y puesta en página

Rodolfo Fuentes NAO

[info@rodolfofuentes.com](mailto:info@rodolfofuentes.com)

Impreso y encuadernado en ZONALIBRO

San Martín 2437

Tel: +(598-2) 208 7819

Fax: +(598-2) 2089603

[zonalibro@adinet.com.uy](mailto:zonalibro@adinet.com.uy)

D.L. xxxxxxxx

Edición amparada en el Decreto 218/996 (Comisión del Papel)

ELEMENTOS DE  
**ECONOMETRÍA**  
APLICACIONES PARA CUBA

Renato Aguilar  
Anicia García  
Pavel Vidal

NUEVA EDICIÓN, CORREGIDA Y AUMENTADA



Centro de Estudios de la Economía Cubana

**dECON**  
Departamento de Economía  
Facultad de Ciencias Sociales  
Universidad de la República





# PRESENTACIÓN

La segunda edición del libro “Elementos de Econometría. Aplicaciones para Cuba”, representa un esfuerzo de actualización y ampliación de la primera edición publicada en el año 2009. La misma es el fruto del trabajo de un equipo conformado por dos investigadores del Centro de Estudios de la Economía Cubana de la Universidad de la Habana, y un profesor chileno-sueco del Departamento de Economía de la Universidad de Gotemburgo, Suecia y se realiza y publica en el marco del programa de colaboración que el Departamento de Economía de la Facultad de Ciencias sociales de la Universidad de la República (UdelaR) de Uruguay, desarrolla con organismos del Estado cubano, en particular con la Universidad de La Habana, para la “Formación en Economía en Cuba”.

El origen de este programa de formación se remonta al año 1996 y en su recorrido ha implicado actividades docentes, de investigación, seminarios y ha resultado particularmente fructífero en términos de un importante número de publicaciones, tanto libros de texto destinados a estudiantes de grado y posgrado en instituciones cubanas, como libros que recogen los resultados de investigaciones en el área económica.

Bajo el liderazgo y la coordinación de Ruben Tansini (†) este programa, en los últimos años, ha desarrollado una serie de programas de posgrado, otras actividades docentes y de investigación, seminarios y publicaciones con el objetivo de contribuir a la consolidación de una masa crítica que promueva el desarrollo tanto de la enseñanza como de la investigación económica. En particular, Ruben fue el gran impulsor del Diplomado en Métodos Estadísticos y Econométricos dictado en la Facultad de Economía de la Universidad de la Habana, que durante tres generaciones ha contado con el apoyo del Departamento de Economía de la UdelaR. Este Diplomado ha estado dirigido a estudiantes, docentes e investigadores de la Facultad de Economía, docentes de otras universidades cubanas, e investigadores de diversas instituciones. La consolidación de este Diplomado en el ámbito de la Maestría en Economía de la Facultad de Economía de la Universidad de la Habana a partir de este año, con docentes cubanos formados por el programa, hubiera significado una enorme satisfacción para su impulsor.

El esfuerzo realizado en incorporar el análisis econométrico ha permitido mejorar el nivel de las investigaciones económicas, y los participantes en el Diplomado han desarrollado diversos proyectos de investigación que han sido presentados y discutidos en los seminarios de la Facultad de Economía de la Universidad de la Habana y en seminarios del Departamento de Economía en Montevideo.

Este libro cuya segunda edición se pone a consideración del lector tiene como objetivo servir de texto docente en econometría para estudiantes universitarios de grado en Cuba, por lo tanto incorpora ejemplos y ejercicios referidos a la economía cubana como forma de facilitar la enseñanza y aprendizaje de esta materia a los estudiantes cubanos. El mismo está estructurado en seis partes y veinticuatro capítulos y con su publicación esperamos colaborar con la difusión de esta materia en el ámbito de la formación en economía en Cuba.

Rosario Domingo

*Coordinadora del Programa  
de Formación en Economía para Cuba*

# ÍNDICE GENERAL

|                                                                         |           |
|-------------------------------------------------------------------------|-----------|
| Presentación Segunda Edición .....                                      | 7         |
| Capítulo 1. Introducción .....                                          | 15        |
| <b>PARTE UNO. FUNDAMENTOS ESTADÍSTICOS DE LA ECONOMETRÍA .....</b>      | <b>21</b> |
| Capítulo 2. Variables Aleatorias y Distribuciones Probabilísticas ..... | 23        |
| 2.1. Probabilidad .....                                                 | 23        |
| 2.2. Variables Aleatorias y Distribuciones Probabilísticas .....        | 25        |
| 2.3. Una Distribución Frecuente .....                                   | 30        |
| 2.4. Distribuciones Conjuntas .....                                     |           |
| 2.5. Combinado y Creando Nuevas Variables Aleatorias.....               | 34        |
| Preguntas.....                                                          | 37        |
| Ejercicios.....                                                         | 38        |
| Capítulo 3. Inferencia Estadística y Estimación .....                   | 41        |
| 3.1. Inferencia Estadística .....                                       | 41        |
| 3.2. Estimación .....                                                   | 43        |
| Preguntas.....                                                          | 47        |
| Ejercicios.....                                                         | 48        |
| Capítulo 4. Prueba de Hipótesis .....                                   | 51        |
| 4.1. Hipótesis y Estadígrafos .....                                     | 51        |
| 4.2. Implementando la Prueba de Hipótesis .....                         | 52        |
| Preguntas.....                                                          | 57        |
| Ejercicios.....                                                         | 58        |
| <b>PARTE DOS. EL MODELO LINEAL NORMAL DE REGRESIÓN .....</b>            | <b>61</b> |
| Capítulo 5. Presentando el modelo .....                                 | 65        |
| 5.1. El Modelo .....                                                    | 65        |
| 5.2. Los Supuestos del Modelo .....                                     | 68        |
| 5.3. Estimadores Mínimo Cuadráticos Ordinarios (EMCO) .....             | 72        |
| 5.4. Los EMCO en Desviaciones de Media.....                             | 83        |
| Preguntas.....                                                          | 87        |
| Ejercicios.....                                                         | 88        |
| Capítulo 6. Inferencia y Prueba de Hipótesis .....                      | 91        |
| 6.1. Pruebas de Hipótesis e Intervalo de Confianza.....                 | 91        |
| 6.2. Test de Significación de los Regresores.....                       | 95        |
| 6.3. Comparando los $B_k$ con Valores Específicos .....                 | 97        |
| 6.4. Test de Hipótesis Compuestas (o Test-t Generalizado) .....         | 99        |
| 6.5. Evaluando la Bondad de Ajuste.....                                 | 103       |
| 6.6. Aplicando el Análisis de Varianza a la Inferencia .....            | 107       |
| 6.6.1. Test de Significación Global.....                                | 107       |
| 6.6.2. Test de Hipótesis Simultáneas .....                              | 109       |

|                                                                                 |            |
|---------------------------------------------------------------------------------|------------|
| 6.7. Consiguiendo una Predicción .....                                          | 112        |
| Preguntas.....                                                                  | 119        |
| Ejercicios.....                                                                 | 120        |
| Capítulo 7. Extensiones del Modelo de Regresión Lineal Normal.....              | 125        |
| 7.1. Posibles Formas Funcionales.....                                           | 125        |
| 7.1.1. Del Modelo Exponencial al Modelo Log-Lineal .....                        | 126        |
| 7.1.2. Los Modelos Semilogarítmicos .....                                       | 128        |
| 7.1.3. Los Modelos Recíprocos.....                                              | 131        |
| 7.1.4. Los Modelos Polinomiales.....                                            | 133        |
| 7.2. Variables Dicótomicas.....                                                 | 134        |
| 7.3. Cambio de Estructura en el MLNR.....                                       | 144        |
| Preguntas.....                                                                  | 149        |
| Ejercicios.....                                                                 | 150        |
| Capítulo 8. Multicolinealidad .....                                             | 157        |
| 8.1. Consecuencias.....                                                         | 159        |
| 8.2. Indicadores para Detectarla .....                                          | 163        |
| 8.3. Medidas Remediales.....                                                    | 167        |
| Preguntas.....                                                                  | 173        |
| Ejercicios.....                                                                 | 174        |
| <b>PARTE TRES. DISCUTIENDO LOS SUPUESTOS DEL MLNR.....</b>                      | <b>177</b> |
| Capítulo 9. Heteroscedasticidad .....                                           | 181        |
| 9.1. Heteroscedasticidad .....                                                  | 181        |
| 9.2. Consecuencias de la Heteroscedasticidad .....                              | 184        |
| 9.3. Detección de la Heteroscedasticidad.....                                   | 185        |
| 9.4. Medidas Correctivas para la Heteroscedasticidad .....                      | 197        |
| Preguntas.....                                                                  | 205        |
| Ejercicios.....                                                                 | 206        |
| Capítulo 10. Autocorrelación.....                                               | 211        |
| 10.1. Autocorrelación.....                                                      | 211        |
| 10.2. Consecuencias de la Autocorrelación .....                                 | 215        |
| 10.3. Detección de la Autocorrelación.....                                      | 217        |
| 10.3.1. Ejemplo: Detectando Autocorrelación en la Demanda de Importaciones..... | 222        |
| 10.4. Medidas correctivas para la autocorrelación .....                         | 225        |
| Preguntas.....                                                                  | 231        |
| Ejercicios.....                                                                 | 232        |
| Capítulo 11. Normalidad.....                                                    | 235        |
| 11.1. Consecuencias de la no-Normalidad.....                                    | 235        |
| 11.2. Detección de la no-Normalidad .....                                       | 236        |
| 11.3. Medidas Correctivas para la no-Normalidad .....                           | 239        |
| Preguntas.....                                                                  | 243        |
| Ejercicios.....                                                                 | 244        |
| Capítulo 12. Regresores Estocásticos .....                                      | 247        |
| 12.1. Reformulando el Modelo .....                                              | 247        |
| 12.2. Variables Instrumentales .....                                            | 250        |

|                                                                                            |            |
|--------------------------------------------------------------------------------------------|------------|
| 12.3. Pruebas de Endogeneidad.....                                                         | 253        |
| 12.4. Validez de las Variables Instrumentales .....                                        | 254        |
| 12.5. Modelos de Ecuaciones Simultáneas.....                                               | 255        |
| 12.6. Estimaciones Multi-etápicas para Ecuaciones Simultáneas .....                        | 258        |
| Preguntas.....                                                                             | 261        |
| Ejercicios.....                                                                            | 262        |
| <br>Capítulo 13. Errores de Especificación y Selección del Modelo .....                    | 265        |
| 13.1. Características de un Buen Modelo.....                                               | 265        |
| 13.2. Variables Omitidas .....                                                             | 267        |
| 13.3. Inclusión de Variables Irrelevantes .....                                            | 270        |
| 13.4. Forma Funcional Incorrecta .....                                                     | 273        |
| 13.5. Detectando Errores de Especificación .....                                           | 274        |
| 13.6. Comprobando la Estabilidad de los Coeficientes de Regresión.....                     | 282        |
| 13.7. Algunos Consejos sobre Cómo Conducir Trabajos Empíricos y Escribir sus Informes..... | 288        |
| Preguntas.....                                                                             | 293        |
| Ejercicios.....                                                                            | 294        |
| <br><b>PARTE CUATRO. MODELOS NO-LINEALES Y DATOS DE PANEL .....</b>                        | <b>297</b> |
| <br>Capítulo 14. Variable Dependiente Cualitativa o Limitada .....                         | 301        |
| 14.1. Variable Dependiente Binaria .....                                                   | 301        |
| 14.1.1. El modelo lineal de probabilidad .....                                             | 302        |
| 14.1.2. El modelo Probit.....                                                              | 303        |
| 14.1.3. Interpretando un modelo Probit .....                                               | 305        |
| 14.2. La Función Logística .....                                                           | 307        |
| 14.3. Otros Modelos de Variables Discretas .....                                           | 308        |
| 14.3.1. Modelos multinomiales .....                                                        | 308        |
| 14.3.2. Modelos de cuenta .....                                                            | 308        |
| 14.4. El Modelo Tobit .....                                                                | 309        |
| Preguntas.....                                                                             | 311        |
| Ejercicios.....                                                                            | 314        |
| <br>Capítulo 15. Datos de Panel .....                                                      | 317        |
| 15.1. Introducción a los Paneles .....                                                     | 317        |
| 15.2. Modelos de Una Dirección y de Dos Direcciones.....                                   | 319        |
| 15.3. El Modelo de Efectos Fijos.....                                                      | 320        |
| 15.4. El Modelo de Efectos Aleatorios.....                                                 | 323        |
| 15.5. ¿Efectos Fijos o Efectos Aleatorios? .....                                           | 326        |
| 15.6. Consideraciones Finales.....                                                         | 328        |
| Preguntas.....                                                                             | 329        |
| Ejercicios.....                                                                            | 330        |
| <br>Capítulo 16. Modelos de Frontera o Asimétricos.....                                    | 333        |
| 16.1. Estimaciones Simétricas y Estimaciones Asimétricas .....                             | 333        |
| 16.2. Fronteras Estocásticas.....                                                          | 334        |
| 16.3. Estimaciones de Fronteras Estocásticas .....                                         | 335        |
| Preguntas.....                                                                             | 339        |
| Ejercicios.....                                                                            |            |

|                                                                                    |            |
|------------------------------------------------------------------------------------|------------|
| <b>PARTE CINCO. INTRODUCCIÓN A SERIES DE TIEMPO .....</b>                          | <b>341</b> |
| Capítulo 17. Series de Tiempo y sus Propiedades.....                               | 345        |
| 17.1. Procesos Estocásticos Estacionarios y no Estacionarios .....                 | 346        |
| 17.2. Prueba de Raíz Unitaria .....                                                | 355        |
| 17.2.1. Aplicación de la Prueba ADF a Series de Tiempo de la Economía Cubana ..... | 359        |
| Preguntas.....                                                                     | 363        |
| Ejercicios.....                                                                    | 364        |
| Capítulo 18. Análisis Univariante de Series de Tiempo.....                         | 367        |
| 18.1. Componentes de las Series de Tiempo.....                                     | 367        |
| 18.1.1. Métodos de Estimación de Componentes Inobservables .....                   | 370        |
| 18.1.2. Modelos Estructurales Univariantes (opcional) .....                        | 376        |
| 18.2 Metodología Box-Jenkins .....                                                 | 381        |
| 18.2.1. Formulación de los Modelos ARIMA .....                                     | 381        |
| 18.2.2. Las Fases de la Metodología .....                                          | 388        |
| 18.2.3. Fase de Identificación .....                                               | 390        |
| Preguntas.....                                                                     | 399        |
| Ejercicios.....                                                                    | 400        |
| Capítulo 19. Función de Transferencia .....                                        | 403        |
| 19.1. Modelos Multivariantes Dinámicos.....                                        | 403        |
| 19.2. Estimación de una Función de Transferencia .....                             | 409        |
| 19.2.1. Identificación de b, s y p.....                                            | 412        |
| 19.3. Análisis de Intervención .....                                               | 413        |
| Preguntas.....                                                                     | 423        |
| Ejercicios.....                                                                    | 424        |
| Capítulo 20. Cointegración y Mecanismo de Corrección de Errores.....               | 427        |
| 20.1. Definición de Cointegración .....                                            | 427        |
| 20.2. Contraste de Cointegración sobre los Residuos.....                           | 431        |
| 20.2.1. Ejemplo del Contraste de Cointegración .....                               | 433        |
| 20.3. Mecanismo de Corrección de Errores .....                                     | 436        |
| 20.3.1. Estimación Bietápica del MCE. Ejemplo con Datos Cubanos .....              | 439        |
| Preguntas.....                                                                     | 443        |
| Ejercicios.....                                                                    | 444        |
| Capítulo 21. Vectores Autorregresivos.....                                         | 447        |
| 21.1. Formulación del VAR.....                                                     | 447        |
| 21.2. Identificación del sistema.....                                              | 451        |
| 21.3. Utilidades del VAR.....                                                      | 454        |
| 21.3.1. Causalidad de Granger.....                                                 | 458        |
| 21.4. Un VAR con los precios de Cuba.....                                          | 461        |
| Preguntas.....                                                                     | 467        |
| Ejercicios.....                                                                    | 468        |
| <b>PARTE SEIS. DOS MODELOS ECONOMÉTRICOS PARA CUBA.....</b>                        | <b>473</b> |
| Capítulo 22. Modelo para la Predicción del PIB Cubano .....                        | 477        |

|                                                                              |     |
|------------------------------------------------------------------------------|-----|
| 22.1. Datos .....                                                            | 478 |
| 22.2. Estimación del Modelo.....                                             | 479 |
| 22.3. Validación del Modelo .....                                            | 482 |
| 22.4. Predicción para 2007 .....                                             | 485 |
| Capítulo 23. Efectividad Conjunta de los Factores de Producción .....        |     |
| en la Agricultura Cañera Cubana .....                                        |     |
| 487 .....                                                                    |     |
| 23.1. Introducción.....                                                      | 487 |
| 23.2. El Marco y el Modelo Teórico .....                                     | 488 |
| .....                                                                        |     |
| 23.3. Los Datos, el Modelo Estimado y los Resultados.....                    | 491 |
| 23.4. Conclusiones y Recomendaciones.....                                    | 495 |
| 23.5. Anexos del Capítulo .....                                              | 497 |
| <b>APÉNDICES</b>                                                             |     |
| A. Apéndice Metodológico .....                                               | 509 |
| 1. Propiedades de las Sumatorias.....                                        | 509 |
| 2. Álgebra de Matrices.....                                                  | 510 |
| 2.1. Definiciones Básicas .....                                              | 510 |
| 2.2. Operaciones con Matrices .....                                          | 511 |
| 2.3. Determinantes e Inversas.....                                           |     |
| 515 .....                                                                    |     |
| 2.4. Producto de Kronecker.....                                              | 517 |
| 2.5. Algunas Matrices Importantes para la Econometría de Datos de Panel..... | 518 |
| B. Tablas Estadísticas.....                                                  | 521 |
| Área Bajo la Distribución Normal Estándar.....                               | 521 |
| Valores Críticos para la Distribución-t .....                                | 522 |
| Valores Críticos para la Distribución-F .....                                | 524 |
| Valores del estadístico Durbin-Watson .....                                  | 528 |
| Valores Críticos para la Distribución- $\chi^2$ .....                        | 531 |
| Valores Críticos del estadístico de Dickey-Fuller .....                      | 532 |
| REFERENCIAS BIBLIOGRÁFICAS .....                                             | 535 |
| .....                                                                        |     |
| ACRÓNIMOS .....                                                              | 540 |
| GLOSARIO .....                                                               | 543 |

CAPÍTULO I





# INTRODUCCIÓN

Este es un texto sobre Econometría. Entonces, resulta conveniente comenzar con una breve discusión sobre qué es la Econometría. En un sentido muy amplio la Econometría es lo que permite a los economistas conocer y explorar la realidad, o mundo real, que se nos aparece como un conjunto muy grande y caótico de pequeñas unidades de información. La econometría nos permite ordenar y sistematizar esta información e identificar sus regularidades y los patrones de comportamiento de los agentes económicos. Podemos decir que la Econometría es el instrumento, basado en la teoría estadística de las regresiones y de la estimación, que da a los economistas la posibilidad de ver el bosque y no solamente los árboles.

Tratemos de ser un poco más precisos. Tenemos dos tipos de información sobre la economía. Por una parte, tenemos información teórica. Esta consiste en un cierto número de proposiciones e hipótesis formuladas por los teóricos, las que en su conjunto llamamos Teoría Económica. Estas proposiciones e hipótesis identifican cuáles son las variables importantes y relevantes de la economía, así como las relaciones que guardan entre sí. Es decir, la información teórica nos describe el posible funcionamiento de las economías y el comportamiento de los agentes económicos. Por otra parte, tenemos información empírica que refleja las conductas y resultados reales de la economía y de las acciones de sus agentes. La Econometría usa esta información empírica para determinar en qué medida la economía se comporta, en la realidad, como sugiere la teoría.

En este sentido, la Econometría tiene dos objetivos: la predicción y la prueba de hipótesis. Este último objetivo, la prueba de hipótesis, resulta usualmente dominante y es el que recibe mayor atención.

Consideramos, en primer lugar, las predicciones. Dada la evidencia empírica contenida en una muestra y una hipótesis teórica, tratamos de estimar la conducta probable de los agentes o de la economía para un individuo, observación o período fuera de la muestra. La experiencia sugiere que las predicciones econométricas, en general, no son buenas, especialmente en el mediano y largo plazos. Esta limitación tiene básicamente dos causas. Una de ellas es que nuestras predicciones están basadas en una teoría que es débil y difusa, especialmente en lo que se refiere a las especificaciones funcionales a usar en nuestros modelos. La segunda razón es que raramente podemos tener la información que realmente necesitamos para nuestros modelos y debemos conformarnos con información secundaria o que ha sido compilada con otros propósitos.

Consideremos ahora la prueba o test de hipótesis. Con las pruebas de hipótesis la Econometría trata de determinar si la evidencia empírica sugiere o no el rechazo de las hipótesis sugeridas por la teoría. La medida y extensión en que esta evidencia sugiere o no el rechazo de las hipótesis es medida en términos probabilísticos. Discutiremos este proceso de prueba de hipótesis más detalladamente más adelante. Sin embargo, debemos recalcar que este procedimiento es un elemento esencial del método científico. Una proposición teórica, o hipótesis, sólo tiene sentido en la medida en que la evidencia empírica disponible no puede sugerir su rechazo, al menos dentro de una calificación probabilística razonable. Estas ideas resultarán más claras en el capítulo correspondiente.

Estas consideraciones nos llevan al problema de la evidencia empírica y los datos. La Econometría enfrenta el problema fundamental que la Teoría Económica no es una ciencia experimental. Es decir, las pruebas de hipótesis no pueden ser hechas a partir de datos obtenidos de experimentos cuidadosamente diseñados y controlados. Por el contrario, debemos proceder con datos que surgieron naturalmente o que fueron recolectados con otros propósitos, los que raramente responden con precisión a las necesidades de nuestras pruebas de hipótesis. Recientemente, una línea de investigación que incluye experimentos se ha hecho más y más popular. Sin embargo, sus resultados realzan las dificultades que enfrenta el diseño de experimentos válidos. También, muchas veces la interpretación de sus resultados ha sido controversial.

Pese a todas estas dificultades, resulta indudable que los estudios empíricos juegan hoy un papel protagónico en el ámbito de trabajo de los economistas. Por una parte, resulta impensable el desarrollo de la teoría económica, o la expansión de nuestra comprensión de las economías nacionales, sin una actividad de estudio e investigación fuertemente centrada en los estudios empíricos. Por otra parte, el diseño, implementación y control de la política económica tendrá una esperanza razonable de éxito sólo si tiene una sólida base en estudios empíricos.

La Econometría es relativamente joven como disciplina científica. De hecho el nombre de esta disciplina aparece sólo en 1926 y es una contribución de Ragnar Frisch, distinguido investigador noruego, que también fundó la Sociedad

Econometría (Econometric Society). Esta Sociedad Econométrica es hoy, posiblemente, la asociación profesional más influyente en el ámbito de la Economía en el mundo. Frisch tomó este nombre de la Biometría, la que ya había conseguido notables éxitos por la aplicación de métodos estadísticos a problemas de la biología. Si bien el fundamento básico de la Econometría está en la teoría de regresiones y el método de mínimos cuadrados, que ya era conocido en el siglo XVIII, son descubrimientos más recientes en teoría de probabilidad, especialmente en el siglo XX, lo que permite el sorprendente desarrollo de esta disciplina. Entre estas contribuciones hay que mencionar los trabajos de F. Borel (1871-1956), Andrey N. Kolmogorov (1903-1987) y Andrey A. Markov (1865-1922).

Un importante precursor es Oskar R. Lange (1904-1965) de Polonia, quien, en la introducción de su libro<sup>1</sup> sobre el tema, señala que la Econometría resulta apropiada en el contexto de una economía centralmente planificada. Lange nos dice que “...un método de investigación econométrica que puede, hasta cierto punto, ser empleado en el manejo de una economía socialista...”. Esta importante observación de Oskar Lange resulta hoy especialmente apropiada para Cuba. Durante las dos últimas décadas la economía cubana ha explorado numerosos senderos y experiencias previamente inéditos. Resulta fácil observar que la estructura misma de la economía cubana ha cambiado significativamente. Se hace, entonces, imprescindible la realización de muchos estudios empíricos que develen los parámetros básicos de la nueva situación. Esta es una base imprescindible para el diseño e implementación de políticas que resulten apropiadas.

Tenemos indicaciones de estudios econométricos antes de 1959 en Cuba, fundamentalmente en el Banco Nacional de Cuba. También hay experiencias posteriores y se ha hecho cada vez más frecuente en la isla la aparición de estudios económicos con un fuerte contenido econométrico. Aparentemente, la Econometría ha llegado a Cuba para quedarse. De hecho, es fácil observar una fuerte actividad y preocupación por temas econométricos en los principales centros de investigación económica del país.

Ya a comienzos de la década de los ochenta se comenzaron a introducir temas de Econometría en los programas de la Universidad de La Habana, estrechamente vinculados a los programas de Estadística. También durante la década de los noventa hubo alguna actividad sobre el tema en diversos programas de postgrado.

Este libro es una consecuencia muy directa del programa de colaboración de Cuba con el Departamento de Economía de la Universidad de la República (Uruguay), que se ha desarrollado ya durante un cierto número de años. En el marco de este programa se han impartido numerosas charlas y cursos sobre tópicos de Econometría, en el ámbito de diversos ministerios e instituciones oficiales cubanas. La evaluación de estas experiencias resaltó la necesidad de dar un

<sup>1</sup> Ver Oskar Lange (1962).

apoyo más directo a la formación de los economistas cubanos en estos temas. Si bien en la actualidad hay disponible un número considerable de textos de Econometría en idioma español, ellos no se adaptaban completamente a las necesidades de los estudios en Cuba. Por una parte, la ejemplificación y los ejercicios propuestos están referidos a realidades económicas muy diferentes y que, a menudo, resultan irrelevantes. Por otra parte, estos textos suponen una formación y experiencias previas que no corresponden a la realidad cubana. De esta manera, concluimos que sería importante el producir un libro de texto que supliera estas dificultades.

Este es, entonces, un libro de texto básicamente dirigido a estudiantes cubanos que se orientan a temas econométricos. Esperamos que este texto pueda servir de base a futuros cursos de Econometría en las universidades y otros centros de formación profesional y académica. También es uno de nuestros objetivos el que este libro pueda constituirse en un texto básico de referencia para economistas cubanos. De allí nuestro interés en desarrollar, a veces muy detalladamente, ejemplos tomados directamente de la economía cubana.

No es posible escribir un texto de Econometría perfectamente autocontenido. De esta manera es necesario suponer un cierto nivel básico de conocimientos previos en los lectores de este texto. Estos conocimientos básicos son los que usualmente consiguen los estudiantes después de dos o tres años en una Facultad de Ciencias Económicas. Más específicamente, suponemos que los lectores tienen los conocimientos y la experiencia necesaria para manejar, con relativa soltura, cuestiones de Álgebra Clásica, que conocen los rudimentos del Cálculo Diferencial y, finalmente, que manejan y entienden las reglas del Álgebra de Matrices. En el terreno de las Estadísticas, suponemos que los lectores manejan las ideas de la Estadística Descriptiva. Es decir, los lectores manejan los elementos básicos que permiten describir una distribución empírica en forma paramétrica y no-paramétrica. También es importante que los lectores tengan conocimientos relativamente sólidos sobre Teoría de las Probabilidades y las distribuciones de probabilidad asociadas a variables aleatorias. Resulta fácil suponer que algunos lectores tendrán lagunas en estos conocimientos básicos o, simplemente, habrán olvidado lo que alguna vez aprendieron sobre estos temas. Sin embargo, algunos de estos temas resultan esenciales para una discusión válida de los temas econométricos. Por este motivo incluimos un cierto número de capítulos, especialmente en la Parte Primera, y un apéndice, que pueden ser considerados como un recordatorio sobre estos temas.

Los temas econométricos son complejos y suelen tener un tratamiento de relativa sofisticación técnica. Sin embargo, un importante objetivo para nosotros es el de presentar un texto relativamente simple que lleve, lo más rápidamente posible, a las ideas centrales del tema en discusión. Por este motivo adoptamos un enfoque doctrinario, en el que evitamos, salvo contadas excepciones, las demostraciones de teoremas y engorrosas deducciones formales. En cambio, nos centramos en la presentación de los principales resultados de la Teoría Econométrica y una discusión detallada de las consecuencias de estos resultados para los estudios empíricos.

El libro está organizado en seis partes y varios apéndices. La Parte Uno, con tres capítulos, “Fundamentos Estadísticos de la Econometría”, discute las ideas estadísticas fundamentales en las que se basa la Econometría. Estos capítulos son estrictamente una revisión de ideas estadísticas básicas. El propósito es el de rellenar posibles lagunas de los lectores sobre estos temas, así como fijar una nomenclatura y vocabulario básico. Naturalmente, lectores y estudiantes con una buena formación estadística pueden prescindir de parte o de toda esta primera parte.

La Parte Dos, con cuatro capítulos, “El Modelo Lineal Normal de Regresión”, presenta lo que es posiblemente el eje central de las discusiones econométricas: el método de estimación por mínimos cuadrados. Muchos de estos temas están usualmente incluidos en las secciones sobre regresiones lineales en cursos de estadísticas. Sin embargo, el tratamiento detallado del tema, como lo hacemos aquí, queda justificado por un número de especificidades y peculiaridades que son relevantes para la discusión econométrica.

La Parte Tres, con cuatro capítulos, “Discutiendo los Supuestos Básicos”, discute la validez de los supuestos en que se basa la estimación mínimo-cuadrática del modelo lineal normal de regresión, así como las consecuencias de su incumplimiento. Esta es una cuestión central para la Econometría, pues la teoría y estimación del MLNR fue desarrollada en el contexto de experimentos controlados, lo que casi nunca sucede en el marco de estudios económicos. Una consecuencia de esta situación es que rara vez podemos confiar en la validez de los supuestos de nuestras estimaciones mínimo-cuadráticas.

Estas tres primeras partes constituyen, usualmente, el contenido de un primer curso de Econometría o “Econometría I”, como es llamado en numerosas universidades y centros de formación. Estos son los contenidos mínimos que debe manejar quien pretenda utilizar la Econometría como un instrumento para estudiar o comprender la realidad económica.

La Parte Cuatro, “Variables Cualitativas, Paneles y Fronteras”, reúne un grupo de modelos para los que resulta un tanto difícil encontrar una ubicación cierta en otras partes del libro. No por eso estos modelos son menos importantes o relevantes para la formación en econometría. Tampoco resulta fácil ubicar estos temas en un esquema de cursos en Facultades de Economía. Algunos de estos contenidos son frecuentemente tratados en cursos de Econometría I y otros en cursos de Econometría II.

La Parte Cinco, “Introducción a Series de Tiempo”, incluye una serie de tópicos en que se exploran las consecuencias que pueden tener las peculiares propiedades estadísticas que a menudo observamos en series de tiempo, dificultando su tratamiento en el marco del modelo lineal normal de regresión. Estos temas son especialmente interesantes para aquellos que deben realizar estudios empíricos basados en series de tiempo. Estas técnicas resultan especialmente relevantes en estudios macroeconómicos. Estos tópicos corresponden, claramente, a segundo curso de Econometría o “Econometría II”.

Finalmente, la Parte Seis, “Tres Modelos Econométricos para Cuba”, presenta versiones sintetizadas de tres importantes aplicaciones para Cuba que han sido realizadas en el último tiempo.

Esta es la segunda edición de este libro. Inmediatamente después de publicada la primera edición pudimos observar una cierta cantidad de errores y deficiencias. La primera edición fue preparada con cierta premura, lo que llevó a que un cierto número de problemas se filtraran. La buena acogida del libro y estos problemas nos llevaron, entonces, a planear esta segunda edición. Algunos errores menores en el texto han sido corregidos. Las secciones de preguntas y ejercicios que acompañan a la mayoría de los capítulos han sido ampliadas. Hemos introducido tres nuevos capítulos que presentan tópicos sobre los cuales hay todavía poca experiencia en Cuba. Sin embargo, estos temas son importantes, especialmente en sus aplicaciones microeconómicas. Esperamos que esta segunda edición haya podido corregir los problemas de la primera. Sin embargo, nuevamente pedimos a los lectores que nos comuniquen posibles errores que continúen presentes en el libro. Estas observaciones, pueden ser dirigidas a [anicia@rectuh.cu](mailto:anicia@rectuh.cu) y a [pavel@ceec.uh.cu](mailto:pavel@ceec.uh.cu), así como otros comentarios sobre el libro. 

• PARTE UNO

• VARIABLES ALEATORIAS  
• Y DISTRIBUCIONES  
• PROBABILÍSTICAS

CAPÍTULO 2





# VARIABLES ALEATORIAS Y DISTRIBUCIONES PROBABILÍSTICAS

En este capítulo trataremos, someramente, un cierto número de conceptos estadísticos básicos que son esenciales para una discusión de las ideas econométricas que queremos presentar. Suponemos que el lector está familiarizado con los conceptos básicos de la teoría estadística. Sin embargo, hemos incluido una revisión de ciertos elementos estadísticos que son esenciales. Esta revisión pretende resolver problemas de nomenclatura y notación, así como un mínimo de rigor cuando usemos estos elementos posteriormente.

## 2.1 Probabilidad

El concepto de probabilidad es central para la Teoría Estadística. En cierto sentido la Estadística es, simplemente, una larga discusión sobre probabilidades. Comencemos, entonces, con una discusión sobre este concepto y por su definición. La probabilidad es una medida de la frecuencia con que ocurre, o puede ocurrir, un evento. Utilizaremos un número real entre 0 y 1 para medir la probabilidad de un evento. Por ejemplo, si lanzamos un dado hay seis eventos posibles: la aparición de cada una de las seis caras del dado. De esta manera, la aparición de una cara determinada, el 3 por ejemplo, ocurre con probabilidad igual a  $1/6$ .

Nótese que también podemos considerar eventos compuestos. Por ejemplo, si lanzamos dos dados tenemos 36 eventos posibles, las 36 combinaciones posibles de las seis caras de cada dado. Sin embargo, podemos definir nuevos eventos compuestos de estos eventos. Por ejemplo, el evento de obtener 11 puntos que se compone de un evento donde el primer dado aparece con un 5 y el segundo con un 6 y el evento donde el primer dado aparece con un 6 y el segundo con un 5. El considerar eventos compuestos ha llevado utilizar la notación de Teoría de Con-

juntos para el análisis de Probabilidades. En este contexto podemos decir que si  $A$  denota un evento, simple o compuesto, tendremos que

$$0 \leq P(A) \leq 1$$

Nótese que si la probabilidad de que suceda  $A$  es  $P(A)$ , entonces la probabilidad de que no suceda  $A$  es igual a  $1-P(A)$ . Un evento tal que  $P(A)=0$  es un evento imposible y un evento tal que  $P(A)=1$  es un evento cierto. Si tenemos un experimento, como el de los dados, con  $n$  eventos posibles e independientes, denotados con  $A_i, i=1, \dots, n$ , podemos escribir

$$\sum_{i=1}^n P(A_i) = P(A_1) + P(A_2) + \dots + P(A_n) = 1$$

Supongamos ahora que conocemos la probabilidad con que ocurren dos eventos  $A$  y  $B$ . Es decir, conocemos los números  $P(A)$  y  $P(B)$ . A partir de estos dos eventos podemos definir otros eventos compuestos. Por ejemplo, el evento  $C$  en que suceden ambos eventos,  $A$  y  $B$ . Este nuevo evento lo denotaremos por  $C = A \cap B$ . Por ejemplo, considerando el experimento de lanzar dos dados  $A$  puede ser el evento de obtener un 3 en el primer dado y  $B$  puede ser el evento de obtener un 4 en el segundo dado.  $C$  puede ser el evento de obtener un 3 en el primer dado y un 4 en el segundo dado.

Ahora podemos definir el evento compuesto  $D$  en que sucede  $A$  o  $B$  (o ambos). Este evento puede describirse como  $D = A \cup B$ . Es decir, siguiendo con el ejemplo anterior este sería el evento en que obtenemos un 3 en el primer dado o un 4 en el segundo dado. En este caso tendremos que la probabilidad de  $D$  está dada por la fórmula:

$$P(D) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Nótese que cuando dos eventos son mutuamente excluyentes, en el sentido de que no pueden ocurrir simultáneamente, tenemos que  $P(A \cap B) = 0$ . En este caso también podemos ver que  $P(A \cup B) = P(A) + P(B)$ .

Otro concepto importante es el de **probabilidad condicionada**. En este caso nos preguntamos por la probabilidad de que ocurra un evento  $A$  dado a que ocurre otro evento  $B$ . Esto lo denotamos por la fórmula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

En este caso leeremos probabilidad de que ocurra  $A$  dado que ocurre  $B$ , o bien probabilidad de la ocurrencia de  $A$  condicionada a la ocurrencia de  $B$ . Con frecuencia consideraremos eventos que no están relacionados entre sí. Es decir, la ocurrencia de un evento no afecta para nada la ocurrencia de otro evento. Estos eventos se llaman **eventos independientes**. Consideremos dos eventos independientes  $A$  y  $B$ . Naturalmente la ocurrencia de  $A$  condicionada a la ocurrencia de  $B$  es simplemente igual a la probabilidad de  $A$  ya que  $B$  no puede influir en su

ocurrencia. Es decir tenemos que  $P(A|B) = P(A)$ . Este resultado aplicado a la fórmula anterior nos lleva a concluir que para eventos independientes tenemos que:

$$P(A \cap B) = P(A)P(B)$$

Un ejemplo de independencia es el lanzamiento de dos dados. La cara que aparece en un dado es independiente de la cara que aparece en el otro dado. Así, tenemos que la probabilidad de que aparezca una cara determinada en un dado es igual a  $1/6$ . La probabilidad de que aparezca un 5 en el primer dado y un 2 en el segundo dado es, entonces, igual a  $1/36 = (1/6)(1/6)$ . Podemos preguntarnos cuál es la probabilidad de obtener 11 puntos. Esto ocurre con cualquiera de dos eventos, el evento de un 5 y un 6 y el evento de un 6 y un 5. Cada uno de estos eventos ocurre con probabilidad  $1/36$ . Entonces la probabilidad de que ocurra uno de estos eventos es  $P(X=11) = (1/36) + (1/36) = 1/18$ , ya que son eventos mutuamente excluyentes.

## 2.2 Variables Aleatorias y Distribuciones Probabilísticas

Continuemos esta revisión de las ideas estadísticas básicas, que son necesarias para nuestra discusión econométrica, por una presentación de la idea de variable aleatoria. Con frecuencia describimos la realidad y, en particular, la realidad económica, con ayuda de variables. Es decir, cantidades variables a las que les atribuimos valores para indicar una determinada característica o estado de la realidad. Por ejemplo, podemos decir que la temperatura máxima de ayer fue de 32 grados centígrados, la tasa de cambio fue de 29 pesos por dólar, las exportaciones alcanzaron 4100 millones de dólares en 2006, etc.

Entre estas variables hay un tipo que nos resulta especialmente interesante: las variables aleatorias. En el caso de estas variables los valores que asumen no están predeterminados. Es decir, en el caso de las variables aleatorias no conocemos sus valores antes de que estos ocurran. Con frecuencia, un valor asumido por una variable lo llamaremos una realización de dicha variable. Entonces, diremos que no conocemos los valores de una variable aleatoria hasta que estos son realizados. Las variables no aleatorias son también llamadas determinísticas o números fijos. En forma sistemática usaremos letras mayúsculas para denotar una variable aleatoria, y la correspondiente letra minúscula para indicar una de sus realizaciones. Por ejemplo, diremos que la temperatura máxima diaria de La Habana es  $X$  y que ayer observamos una temperatura máxima de  $x$  grados.

Aunque no conocemos los valores que asume una variable aleatoria antes de su realización, sí conocemos un rango de valores posibles. Más aún, con frecuencia conocemos la probabilidad de que una variable aleatoria asuma un determinado valor, o un valor dentro un determinado intervalo. Por ejemplo, podemos decir que la temperatura máxima diaria de La Habana es un número mayor que 5 y menor que

50. También podemos decir que la probabilidad de que la temperatura máxima de mañana alcance un valor entre 28 y 35 grados es del 90 por ciento.

Esta es una característica de las variables aleatorias o estocásticas. Cada variable aleatoria tiene asociada una función de distribución de probabilidades, la que indica la probabilidad con que sus diversos valores son realizados. Formalmente decimos que para una variable aleatoria  $X$  existe una función  $P(X=x)=F(x)$ . Esta expresión la leemos diciendo que el valor  $x$  de la variable aleatoria  $X$  ocurre con una probabilidad igual a  $F(x)$ .

Hay dos tipos de variables aleatorias: **variables discretas** y **variables continuas**. Una variable discreta es una variable que puede asumir solamente un número finito de valores, o tomados de un conjunto de valores que pueden ser enumerados usando los números enteros. Variables continuas son variables que pueden asumir como valores todos los números reales dentro de un rango o intervalo de números. Naturalmente, este intervalo puede ir de menos infinito a más infinito, incluyendo todos los números reales posibles. Por ejemplo, el número de días soleados en La Habana durante un determinado mes es una variable aleatoria discreta, la temperatura media en La Habana, durante un determinado mes, es una variable aleatoria continua.

Comencemos discutiendo el caso de una variable aleatoria discreta. Para ello utilizaremos un experimento intelectual clásico: el lanzamiento de dos dados. Es fácil ver que este experimento, lanzar dos dados, da lugar a 36 eventos posibles, porque cada una de las seis caras del primer dado se combina con cada una de las seis caras del segundo dado. Definiremos nuestra variable aleatoria  $X$  como la suma de los valores de las caras de ambos dados. De esta manera solo hay 11 valores posibles para nuestra variable, el menor de los cuales es 2 y el mayor es 12. Como no podemos predecir el valor que tendrá nuestra variable cada vez que lancemos los dados, ésta es claramente una variable aleatoria discreta.

La tabla siguiente muestra, en la tercera columna, cuantas diferentes combinaciones de dados producen los diferentes valores que puede asumir la variable  $X$ . Es decir, esta columna nos dice cuantos eventos están asociados a un determinado valor de la variable. Dividiendo por el número total de eventos (36) obtenemos la probabilidad de que la variable  $X$  asuma cada uno de los 11 valores que son posibles. La última columna de esta tabla representa la función o distribución de probabilidad asociada a la variable aleatoria  $X$ : valor acumulado de dos dados. Estas funciones o distribuciones de probabilidad tienen algunas importantes propiedades, las que se deducen directamente de la definición de probabilidad.

Podemos usar esta función para responder diversas preguntas sobre nuestra variable aleatoria. Por ejemplo, el valor con la más alta probabilidad es 7, con una probabilidad igual a  $1/6$ . Es decir, tenemos que  $P(X=7)=0.17$ . Este valor lo llamaremos la **moda**.

Esta función, que llamamos la distribución probabilidades de la variable  $X$ , tiene varios **momentos** asociados; es decir parámetros o números asociados que describen características de la distribución. Primeramente, consideremos la **media**, la que definimos con la siguiente fórmula:

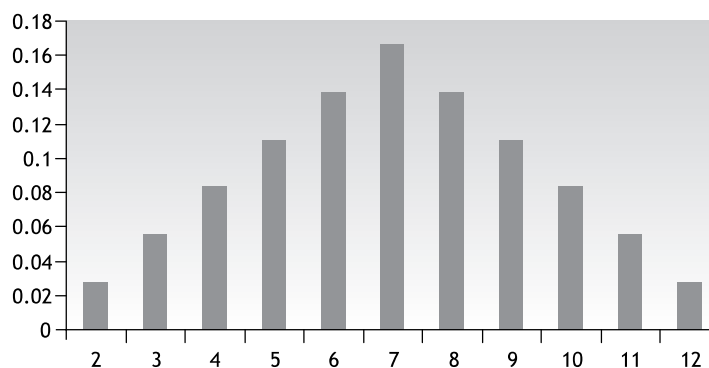
$$E[X] = \mu = \sum_{i=1}^N x_i f(x_i)$$

donde  $N$  es el número de valores posibles para la variable  $X$ . Los distintos valores que la variable  $X$  puede asumir están indicados por el subíndice  $i$ .

Tabla 2.1: Distribución de probabilidad de dos dados

| Evento | $x$ | Frecuencia Absoluta | $P(X=x)$ | Frecuencia Acumulada |
|--------|-----|---------------------|----------|----------------------|
| 1      | 2   | 1                   | 0.03     | 0.03                 |
| 2      | 3   | 2                   | 0.06     | 0.08                 |
| 3      | 4   | 3                   | 0.08     | 0.17                 |
| 4      | 5   | 4                   | 0.11     | 0.28                 |
| 5      | 6   | 5                   | 0.14     | 0.42                 |
| 6      | 7   | 6                   | 0.17     | 0.58                 |
| 7      | 8   | 5                   | 0.14     | 0.72                 |
| 8      | 9   | 4                   | 0.11     | 0.83                 |
| 9      | 10  | 3                   | 0.08     | 0.92                 |
| 10     | 11  | 2                   | 0.06     | 0.97                 |
| 11     | 12  | 1                   | 0.03     | 1.00                 |
| Total  |     | 36                  | 1.00     |                      |

Figura 2.1: Distribución del experimento de dados



La media es también llamada el primer momento central de la distribución. Es importante comparar este momento distribucional con la **media muestral**. Consideremos una muestra de  $n$  observaciones de la variable  $x$ . Es decir, consideremos la colección de  $n$  valores que realmente obtenemos cuando lanzamos dos dados  $n$  veces. Entonces, la media muestral es

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Nótese el diferente significado de  $n$  y  $N$ . En el primer caso  $N$  se refiere al número de valores diferentes que puede asumir la variable  $x$ . En el segundo caso simplemente representa el tamaño de la muestra y puede asumir cualquier valor entero positivo. En el primer caso estaremos hablando de una media **poblacional**, o **valor esperado**, o **esperanza matemática**, o media teórica. En el segundo caso hablamos de media **muestral** o media empírica. Es importante tener siempre presente si nos referimos a valores poblacionales o teóricos, o nos referimos a valores muestrales o empíricos. De hecho, en el futuro nuestra discusión estará cambiando continuamente entre un nivel poblacional y un nivel muestral.

Otro momento o parámetro de esta distribución que es importante es el segundo momento en torno a la media o **varianza**. Este momento está dado por la fórmula:

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 f(x_i)$$

donde  $\mu$  es la media de la distribución, de acuerdo con la fórmula anterior. Consideremos la versión muestral de este momento. La siguiente fórmula presenta este parámetro en su versión muestral o empírica:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

En forma semejante podemos definir una secuencia de momentos para estas distribuciones. Las siguientes fórmulas nos dan los momentos de orden  $r$  en torno a cero y en torno a la media, respectivamente.

$$\mu^r = \sum_{i=1}^N x_i^r f(x_i), \quad \sigma^r = \sum_{i=1}^N (x_i - \mu)^r f(x_i)$$

Observando estas fórmulas resulta claro por qué decimos que la media es el primer momento en torno al cero y la varianza el segundo momento en torno a la media. Las fórmulas correspondientes para el caso muestral pueden ser fácilmente deducidas. El lector puede deducir fácilmente que el primer momento en torno a la media es siempre igual a cero. También resulta fácil deducir la siguiente fórmula abreviada:

$$\sigma^2 = E[X^2] - \mu^2$$

Consideremos ahora el caso de una variable continua; es decir, una variable aleatoria que puede asumir cualquier valor dentro de un rango de números reales. El problema que presentan las variables aleatorias se deriva del hecho de que en cada intervalo de números reales existen infinitos valores. En consecuencia, la probabilidad de que una variable aleatoria continua asuma un valor dado es, por definición, igual a cero. Esto es, podemos escribir que  $P(x = a) = 0$  es siempre cierto para cualquier distribución continua y para cualquier valor  $a$ . Esto nos lleva a considerar una definición alternativa de la función de distribución. Con el objeto de dar un contenido a esta función la definiremos de la siguiente manera:

$$F(a) = P(x \leq a)$$

Es decir, nuestra función registra ahora la probabilidad de que la variable aleatoria  $X$  asuma un valor menor o igual que un valor dado  $a$ . Esta función la llamaremos función de distribución de probabilidades de la variable aleatoria  $X$  o simplemente función de probabilidad. Algunas veces esta función se denomina función de distribución acumulada. Esta función es derivable, su primera derivada existe, y podemos escribir:

$$f(x) = F'(x)$$

Esta nueva función, la derivada de la función de distribución, la llamaremos **función de densidad**. Esto nos permite representar la distribución de probabilidades asociadas a una variable aleatoria como una integral:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$$

En este caso la media y la varianza están dadas por las siguientes fórmulas:

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Al igual que en el caso anterior, también en este caso podemos definir momentos de orden superior.

Podemos usar variables aleatorias simples para construir nuevas variables aleatorias. Es fácil de ver que si  $X$  y  $Y$  son variables aleatorias, entonces una combinación lineal de estas dos variables es también una variable aleatoria. Es decir, la variable  $Z = \alpha X + \beta Y$ , donde  $\alpha$  y  $\beta$  son constantes arbitrarias, es también una variable aleatoria. Naturalmente, este resultado puede ser generalizado para un número arbitrario de variables aleatorias. También este resultado puede ser generalizado a una función arbitraria  $g$ . Es decir, la variable  $Z=g(X)$ , donde  $X$  es una variable aleatoria, es también una variable aleatoria.

También podemos deducir algunos de los momentos de la distribución de  $Z$ . En el caso de variables discretas la media de una combinación lineal de variables aleatorias es la combinación lineal de la media de estas variables. En otras palabras, si tenemos  $n$  variables aleatorias  $X_1, X_2, \dots, X_n$  la variable

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n$$

es también una variable aleatoria. La media muestral de esta variable  $Z$  está dada por la fórmula:

$$\bar{x} = \alpha_1 \bar{x}_1 + \alpha_2 \bar{x}_2 + \dots + \alpha_n \bar{x}_n$$

La correspondiente fórmula para la esperanza matemática de  $Z$ , o media poblacional, es:

$$\mu = \alpha_1 \mu_1 + \alpha_2 \mu_2 + \dots + \alpha_n \mu_n$$

Sin embargo, este resultado no puede ser generalizado a funciones arbitrarias  $g$ . La fórmula correspondiente para la varianza está dada por:

$$\sigma^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 + \sum_{i \neq j} \alpha_i \alpha_j \text{cov}[X_i X_j]$$

El término  $\text{cov}[X_i X_j]$ , la **covarianza**, es un nuevo indicador, cuyo sentido quedará claro más adelante.

## 2.3 Una Distribución Frecuente

Una distribución continua extremadamente importante es la llamada **Distribución Normal**. En realidad se trata de una familia de distribuciones que dependen de dos parámetros, la media y la varianza. La función de densidad de esta distribución está dada por la siguiente fórmula:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\beta)^2}{2\sigma^2}}$$

Esta fórmula depende de los parámetros  $\beta$  y  $\sigma^2$ . Es fácil demostrar que el primer parámetro es igual a la media de esta distribución y el segundo parámetro es igual a su varianza. De esta manera, cada vez que especificamos valores para estos parámetros, especificamos una nueva distribución perteneciente a esta familia de distribuciones normales. Cuando queremos expresar que la variable aleatoria  $Y$  se distribuye normalmente con media igual a  $\beta$  y varianza igual a  $\sigma^2$ , escribimos la siguiente expresión:

$$Y \sim N(\beta, \sigma^2)$$

Un miembro especialmente importante de esta familia es la **Distribución Normal Estándar**. Esta distribución se obtiene cuando la media es igual a cero y la varianza es igual a uno. En consecuencia, la correspondiente función de densidad es:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Estas distribuciones son simétricas en torno a la media, tal como lo muestra la Figura 2.2. La curva que representa esta función de densidad es a veces llamada **Campana de Gauss**, debido a su forma y a que fue primeramente introducida por el matemático alemán Carl Friedrich Gauss (1777-1855).

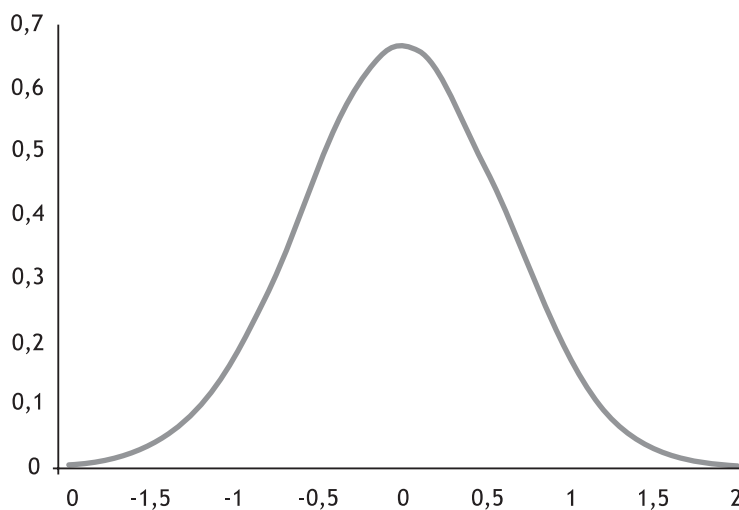
Supongamos que tenemos una variable aleatoria  $Y$  normalmente distribuida, con media igual a  $\beta$  y varianza igual a  $\sigma^2$ . Consideremos ahora la siguiente fórmula, llamada **Fórmula de Estandarización**:



$$Z = \frac{Y - \beta}{\sigma}$$

Es posible demostrar que la nueva variable aleatoria  $Z$  definida por esta fórmula también se distribuye normalmente, pero con media igual a cero y varianza igual a uno. Es decir, la variable  $Z$  sigue una distribución normal estándar.

Figura 2.2: La distribución normal



Este resultado tiene una importante aplicación que ilustraremos con la ayuda de un ejemplo. Supongamos que tenemos una variable aleatoria  $Y$  que se distribuye normalmente con media igual 4 y varianza igual a 4. Queremos averiguar la probabilidad de que esta variable asuma un valor que esté entre -1 y 5. Recordemos que en el Apéndice B tenemos una tabla que se llama “Área bajo la Distribución Normal Estándar”. En esta tabla se registra el valor bajo la función de densidad normal estándar para un número de valores. Observe que la tabla registra el área entre cero y un valor dado  $z$ . Es decir, para una realización igual a  $z$  de la variable  $Z$ , esta tabla nos entrega el valor de la probabilidad  $P(0 \leq Z \leq z)$ .

Nuestro primer paso será utilizar la fórmula de estandarización para encontrar los valores estandarizados que corresponden a -1 y a 5. Aplicando la fórmula de estandarización obtenemos:

$$z_1 = \frac{5-4}{2} = 0,5, \quad z_2 = \frac{-1-4}{2} = -2,5$$

Buscando en la primera columna de la tabla el valor 0,5 vemos que la probabilidad de un valor de  $z$  entre 0 y 0,5 es igual a 0,1915. El valor de -2,5 no está en la tabla que contiene sólo valores positivos de  $z$ . Si observamos el valor correspondiente a 2,5 obtenemos que la probabilidad de un  $z$  entre 0 y 2,5 es igual a 0,4938.

Ya que esta distribución es simétrica en torno a cero, la probabilidad entre -2,5 y 0 también será igual a 0,4938. Por lo tanto, componiendo ambos resultados, vemos que

$$P(-2,5 \leq z \leq 0,5) = P(-2,5 \leq z \leq 0) + P(0 < z \leq 0,5) = 0,4938 + 0,1915 = 0,6853$$

Este resultado puede ser trasladado a la distribución de la variable no estandarizada y podemos afirmar que  $P(-1 \leq y \leq 5) = 0,6853$ . Naturalmente, no podemos computar tablas para todas las distribuciones normales. Pero utilizando sólo la distribución normal estándar podemos calcular probabilidades correspondientes a otras distribuciones normales con razonable precisión.

Nótese que estos resultados pueden ser fácilmente conseguidos con ayuda de una computadora. Todos los programas especializados en aplicaciones estadísticas o econométricas tienen algún procedimiento apropiado para resolver el problema anterior, y aún con mayor precisión. Un programa de cálculo de enorme difusión como Excel también cuenta con una función que permite calcular un buen número de distribuciones.

## 2.4 Distribuciones Conjuntas

Consideremos dos variables aleatorias  $X$  y  $Y$ . Muchas veces nuestro interés se centra, antes que en realizaciones individuales de estas variables, en la realización de pares ordenados de realizaciones de estas variables,  $(x, y)$ , en que  $x$  es una realización de la variable  $X$  y  $y$  es una realización de la variable  $Y$ . El siguiente podría ser un buen ejemplo. Supongamos que estamos reclutando personal para una oficina de estadísticas agrícolas. Para ello disponemos de una larga lista de posibles candidatos. Para estos potenciales candidatos hemos cuantificado dos características. En primer lugar, un indicador que refleja sus conocimientos estadísticos y su experiencia en este tipo de cuestiones. En segundo lugar, otro indicador que refleja conocimientos y experiencia en cuestiones agrícolas. Naturalmente tenemos interés en conocer la probabilidad de conseguir trabajadores adecuados. Para resolver el problema debemos preocuparnos por la realización de distintas combinaciones de estas características. Distintas combinaciones de características van a tener, en principio, distintas probabilidades de realización. De esta manera podemos definir la distribución o función de probabilidad conjunta de  $X, Y$  con la siguiente fórmula:

$$F(x, y) = P(X \leq x, Y \leq y)$$

Nuestro ejemplo anterior, que se refiere a una distribución conjunta discreta, incluye entonces dos variables estocásticas. La primera de ellas es la variable  $X$ , que es igual a cero si el candidato no tiene conocimientos estadísticos, y es igual a uno si tiene estos conocimientos. La segunda es la variable  $Y$  que es igual a cero para los candidatos que no tienen experiencia agrícola, e igual a uno para los que sí tienen esta experiencia. Este ejemplo, con variables aleatorias discretas con sólo dos valores, puede ser fácilmente generalizado a variables con más valores

posibles, y aun para variables continuas. La siguiente tabla presenta la función de distribución de las probabilidades asociadas a este ejemplo:

|                      | $Y=0$            | $Y=1$            | Marginal<br>para $X$ |
|----------------------|------------------|------------------|----------------------|
| $X=0$                | $F(0,0)$<br>0,15 | $F(0,1)$<br>0,45 | $f(0)$<br>0,60       |
| $X=1$                | 0,35<br>$F(1,0)$ | 0,05<br>$F(1,1)$ | $f(1)$<br>0,40       |
| Marginal<br>para $Y$ | $f(0)$<br>0,50   | $f(1)$<br>0,50   |                      |

Cada una de las celdas de esta tabla representa un valor de la función  $F(X,Y)$ . Esta tabla podemos leerla de la siguiente manera. Si escogemos un candidato de nuestra muestra al azar, la probabilidad de que sea una persona con experiencia agrícola ( $X=1$ ) y con conocimientos estadísticos ( $Y=1$ ) es igual al 5 por ciento ( $F(1,1)=0,05$ ). Es posible dar interpretaciones semejantes a las restante tres celdas. Nótese que la suma de las cuatro celdas es igual a uno, ya que esta tabla representa una función de probabilidades.

Una distribución conjunta tiene asociada una **distribución marginal** correspondiente a cada variable. En nuestro ejemplo hay una distribución marginal asociada a la variable  $X$  y otra asociada a la variable  $Y$ . La distribución marginal de  $X$  nos dice cual es la probabilidad de realización de los valores de  $X$ , cualesquiera que sean los valores asumidos por  $Y$ . Esta distribución puede ser obtenida sumando, para cada valor de  $X$ , todos los valores de probabilidad asociados a  $Y$ . Una definición semejante puede ser dada para distribución marginal de  $Y$ . Las fórmulas correspondientes son:

$$f(x) = \sum_y F(x, y), \quad f(y) = \sum_x F(x, y)$$

La última columna de la tabla anterior presenta la distribución marginal de la variable  $X$ . La última fila de la tabla anterior presenta la distribución marginal de la variable  $Y$ . Por ejemplo, la probabilidad de encontrar un candidato con conocimientos estadísticos, no importa cuál sea su experiencia agrícola es igual al 40 por ciento. Nótese que ambas distribuciones marginales suman uno, ya que son funciones de probabilidad.

$$f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f(y)}$$

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f(x)}$$

También podemos considerar la distribución condicional de la variable  $X$  dado que la variable  $Y$  asume un valor dado  $y$ . Las siguientes fórmulas nos indican como se calculan estas distribuciones con ayuda de la distribución conjunta de ambas variables y las correspondientes distribuciones marginales.

Esto es, la distribución de  $X$  condicionada a  $Y$  la obtenemos dividiendo la distribución conjunta de  $X$  e  $Y$ , por la distribución marginal de  $Y$ . Volviendo a nuestro ejemplo anterior la probabilidad de que un candidato tenga experiencia agrícola, bajo la condición de que posea conocimientos estadísticos, es  $P(X=1|Y=1)=0,05/0,50=0,10$ . También podemos decir que, dado que sabemos que un determinado candidato tiene conocimientos estadísticos, la probabilidad de que tenga experiencia agrícola es igual al 10 por ciento.

Este ejemplo nos permite también discutir la idea de **independencia estadística**. Decimos que dos variables aleatorias son independientes cuando los valores que asume una de ellas no influyen los valores que asume la otra. Para las variables independientes tenemos que la distribución conjunta de ellas es simplemente igual al producto de sus distribuciones marginales. Formalmente, si tenemos  $n$  variables aleatorias  $X_1, \dots, X_n$  podemos escribir:

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n)$$

En estas condiciones, la distribución de una variable aleatoria  $X$  condicionada a otra variable aleatoria  $Y$  se convierte en:

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x)$$

Este resultado recalca el hecho que, para variables independientes, los valores asumidos por  $X$  no dependen de los valores asumidos por  $Y$ .

Cuando consideramos dos variables aleatorias un estadígrafo importante es la **covarianza**. La covarianza entre dos variables aleatorias,  $X$  e  $Y$ , es una medida del grado de asociación lineal entre ellas. La fórmula correspondiente es:

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ \text{cov}(X, Y) &= E[XY] - E[X]E[Y] \end{aligned}$$

La varianza es un caso particular de covarianza, simplemente cuando consideramos la covarianza de una variable consigo misma. Una aplicación importante de la covarianza es el **coeficiente de correlación** definido como:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$$

El coeficiente de correlación es siempre un número entre -1 y 1. En este sentido es más útil que la covarianza para comparar el grado de asociación que existe entre pares de variables aleatorias. Más aún. Las variables independientes tienen una covarianza igual a cero y, en consecuencia, un coeficiente de correlación que también es igual a cero. Sin embargo, la conclusión inversa no es necesariamente cierta.

## 2.5 Combinado y Creando Nuevas Variables Aleatorias

Si tenemos un cierto número de variables aleatorias, podemos combinarlas con el propósito de crear otras variables aleatorias. Para discutir esta idea supongamos que tenemos  $n$  variables aleatorias  $X_1, X_2, \dots, X_n$ . Supongamos que tenemos también  $n$  constantes  $c_1, c_2, \dots, c_n$ . Entonces podemos presentar los siguientes resultados:

$$E[c_1 X_1 + c_2 X_2 + \dots + c_n X_n] = c_1 E[X_1] + c_2 E[X_2] + \dots + c_n E[X_n]$$

Es decir, la media o esperanza matemática de una combinación lineal de variables aleatorias es igual a la combinación lineal de las medias de las variables. Existe también una fórmula para la varianza.

$$\begin{aligned} \text{var}[c_1 X_1 + c_2 X_2 + \dots + c_n X_n] &= c_1^2 \text{var}[X_1] + c_2^2 \text{var}[X_2] + \dots + \\ &+ c_n^2 \text{var}[X_n] + \sum_{i \neq j} \sum c_i c_j \text{cov}(X_i, X_j) \end{aligned}$$

Nótese que si las variables son independientes, entonces sus covarianzas serán iguales a cero y la fórmula anterior se reduce a:

$$\text{var}[c_1 X_1 + c_2 X_2 + \dots + c_n X_n] = c_1^2 \text{var}[X_1] + c_2^2 \text{var}[X_2] + \dots + c_n^2 \text{var}[X_n]$$

Cuando las constantes son todas iguales a 1, podemos decir que la varianza de una suma de variables aleatorias es igual a la suma de las varianzas correspondientes.

Supongamos ahora que cada una de nuestras variables se distribuye normalmente con una media y una varianza determinada. Es decir,

$$X_1 \sim N(\beta_1, \sigma_1^2), X_2 \sim N(\beta_2, \sigma_2^2), \dots, X_n \sim N(\beta_n, \sigma_n^2)$$

Entonces, una variable  $Z$  igual a una combinación lineal de variables normales, también se distribuirá normalmente. Es decir,

$$Z = c_1 X_1 + c_2 X_2 + \dots + c_n X_n \sim N(E[Z], \text{var}(Z))$$

A partir de la distribución normal podemos crear un cierto número de otras distribuciones que son importantes y que tendrán importantes aplicaciones más adelante. Por ejemplo, supongamos que tenemos  $m$  variables independientes  $Z_1, Z_2, \dots, Z_m$ , todas ellas distribuidas siguiendo una distribución normal estándar. Es decir,  $Z_i \sim N(0, 1), i = 1, \dots, m$ . Entonces una variable  $V$  definida como la suma de los cuadrados de estas variables sigue una distribución llamada **chi-cuadrado**, con  $m$  grados de libertad. Formalmente diremos que

$$V = Z_1^2 + Z_2^2 + \dots + Z_m^2 \sim \chi_{(m)}^2$$

Nótese que esta nueva distribución depende de un parámetro  $m$ . Es decir se trata de una familia de distribuciones, donde obtenemos una nueva distribución cada vez que especificamos un valor para este parámetro  $m$ , llamado grados de

libertad. La media de esta distribución es igual a estos grados de libertad  $m$ , y la varianza es igual a  $2m$ .

Supongamos ahora que tenemos dos variables aleatorias independientes, cada una de ellas distribuida de acuerdo a una distribución chi-cuadrado con  $m_1$  y  $m_2$  grados de libertad, respectivamente. Entonces podemos crear una nueva variable  $F$  que se distribuye de acuerdo a una distribución llamada **distribución  $F$**  (Fischer). La siguiente fórmula nos indica como construir esta variable y nos da una notación para esta nueva distribución.

$$F = \frac{V_1 / m_1}{V_2 / m_2} \sim F_{(m_1, m_2)}$$

Decimos que  $F$  es una variable distribuida  $F$  con  $m_1$  grados de libertad en el numerador y  $m_2$  grados de libertad en el denominador.

Finalmente, supongamos que tenemos dos variables independientes; una de ellas,  $Z$ , distribuida de acuerdo a una normal estándar, y la otra,  $V$ , siguiendo una distribución chi-cuadrado con  $m$  grados de libertad. Es decir,  $Z \sim N(0,1)$  and  $V \sim \chi^2_{(m)}$ . Entonces podemos crear otra variable,  $t$ , que sigue una distribución  **$t$**  o **Student**. La fórmula siguiente nos dice cómo crear esta variable y su distribución.

$$t = \frac{Z}{\sqrt{V/m}} \sim t_{(m)}$$

Decimos que  $t$  se distribuye Student con  $m$  grados de libertad. Nuevamente estamos frente a una familia de distribuciones dependientes de un parámetro  $m$ . Especificando este parámetro obtenemos miembros individuales de esta familia. Esta distribución, tiene media cero y varianza igual a  $m/(m-2)$ . Esta distribución es simétrica en torno a cero y se aproxima más y más a una distribución normal estándar a medida que el número de grados de libertad  $m$  aumenta.

El Apéndice B presenta estas distribuciones tabuladas para una selección de valores importantes. Es importante entrenarse adecuadamente en el uso de estas tablas. Todos los programas estadísticos y econométricos pueden entregarnos los valores de estas tablas, y muchos más, con una elevada precisión. Estos resultados pueden ser encontrados en Excel, abriendo “funciones”, y después eligiendo “funciones estadísticas”. Excel también entrega extensos recursos de ayuda para entender como se manejan estas funciones.

# *preguntas*

Trate de responder a las siguientes preguntas con ayuda de este capítulo.

1. ¿Qué es una variable aleatoria?
2. Dé ejemplos de variables aleatorias discretas y continuas.
3. ¿Qué es una distribución de probabilidad?
4. ¿Cómo se llama el primer momento en torno a cero?
5. ¿Cómo se llama el segundo momento en torno a la media?
6. ¿Qué es una distribución marginal?
7. ¿Qué significa que dos variables aleatorias sean dependientes?
8. ¿Cuál es el valor de la covarianza de dos variables independientes?
9. Si tenemos dos variables aleatorias distribuidas normalmente, ¿cómo se distribuye la suma de estas variables?



# ejercicios

- 2.1 Tenemos una variable aleatoria  $X$  que se distribuye de acuerdo a una distribución normal estándar. Calcule las siguientes probabilidades:
- $P[0 \leq X \leq 1,2]$
  - $P[0,3 \leq X \leq 1,2]$
  - $P[-0,8 \leq X \leq 1,2]$
- 2.2 Una variable aleatoria se distribuye normalmente con media  $\mu=1,2$  y varianza  $\sigma^2=0,81$ . Calcule las siguientes probabilidades:
- $P[0 \leq X \leq 3,2]$
  - $P[0,3 \leq X \leq 2,2]$
  - $P[-1,8 \leq X \leq 3,4]$
- 2.3 Supongamos que una variable aleatoria  $V$  se distribuye chi-cuadrado con 8 grados de libertad ¿Cuál es el valor  $v$  de la variable tal que  $P[V \geq v] \geq 0,05$ ?
- 2.4 Tenemos tres variables aleatorias  $X$ ,  $Y$  y  $Z$ . Sus medias son 5, 2 y -1 respectivamente. La siguiente tabla nos da sus varianzas y covarianzas:

|     | $X$   | $Y$  | $Z$  |
|-----|-------|------|------|
| $X$ | 16,00 | 1,65 | 0,00 |
| $Y$ | 1,65  | 6,25 | 0,00 |
| $X$ | 0,00  | 0,00 | 2,25 |

Los valores de la diagonal principal son las varianzas y los valores en la intersección de cada fila y columna indican las covarianzas de las variables correspondientes.

- ¿Qué variables son independientes y cuales son dependientes?
  - Si definimos  $W$  como  $W=2X - 3Y$ , calcule la media y la varianza de  $W$ .
  - Si se define  $H$  como  $H=3Y + 2,5Z$ , calcule la media y la varianza de  $H$ .
  - Supongamos que  $Z$  se distribuye normalmente. ¿Qué distribución tiene la variable aleatoria  $J = \frac{Z-2}{1,5}$ ?
  - Evalúe la siguiente proposición probabilística:  $P[-1,1 \leq J \leq 0,5]$
- 2.5 Quince muchachos estudiantes de economía, después de un accidente, se encuentran en una balsa salvavidas con una capacidad máxima de 1300 kilos. Si la distribución del peso de los estudiantes varones de economía es normal con una media de 80 kilos y con una desviación estándar de 8 kilos, encuentre:



- a. La probabilidad que tiene un estudiante de pesar menos de 85 kilos.
- b. La probabilidad de que todos los estudiantes pesen menos de 85 kilos.
- c. La probabilidad de que la balsa esté sobrecargada.
- d. ¿Cuál es el número máximo de estudiantes que pueden estar en la balsa si la probabilidad de sobrecarga no debe ser mayor que 0,001?

CAPÍTULO 3



# INFERENCIA ESTADÍSTICA Y ESTIMACIÓN

En este capítulo vamos a ver dos temas que son centrales para discutir las cuestiones econométricas que tratamos más adelante: Inferencia Estadística y Estimación. Estos temas se refieren a como la Estadística puede informarnos sobre una realidad compleja y que, usualmente, sólo podemos observar en forma parcial y fragmentada. Para ello, utilizaremos dos conceptos básicos: el de población y el de muestra.

## 3.1 Inferencia Estadística

Consideremos una población, ya sea de individuos, empresas u otros elementos estadísticos. Queremos conocer y caracterizar esta población pero no podemos acceder a ella; ya sea porque es imposible, por ejemplo, en el caso de una población infinita; o impráctico, por ejemplo, porque la población es muy grande o muy dispersa geográficamente. A pesar de esto podemos decir muchas cosas sobre esta población, usando un procedimiento llamado *inferencia estadística*. Este procedimiento consiste en deducir o inferir hechos y resultados sobre una población a partir de la información contenida en muestras extraídas de la población en estudio. Es decir, utilizando colecciones limitadas de elementos extraídos de la población.

Nuestros resultados estarán referidos, casi invariablemente, a *muestras aleatorias*. Diremos que una muestra de tamaño  $n$  es aleatoria cuando cualquier combinación de  $n$  elementos de una población tienen la misma probabilidad de ser extraída como muestra. En otras palabras, cada elemento en la población tiene la misma probabilidad de ser incluido en una muestra determinada.

Cada vez que consideramos resultados obtenidos sobre la base de muestras debemos considerar el problema de la variabilidad muestral. Diferentes muestras darán lugar a resultados diferentes. De esta manera, las variables que observamos en nuestras muestras son variables aleatorias, pues los valores que asumen dependen de la muestra que obtuvimos.

Supongamos que tenemos una variable  $X$  definida sobre una población. Nos interesa conocer ciertos parámetros que caracterizan la distribución de esta variable sobre la población. Estos son los **parámetros poblacionales** y son números fijos (no aleatorios), usualmente desconocidos o inobservables. Un claro ejemplo son los momentos de la distribución de esta variable y, en particular, la media  $E(X) = \mu$  y la varianza  $\text{var}(X) = \sigma^2$ . Los correspondientes parámetros para la muestra son la media muestral y la varianza muestral. La Teoría Estadística, en el siguiente teorema, nos entrega un muy importante resultado, que relaciona la media muestral y la media poblacional.

**Teorema.** Si extraemos muestras aleatorias de tamaño  $n$  de una población infinita con media  $\mu$  y varianza  $\sigma^2$ , tenemos el siguiente resultado para la distribución de la media muestral sobre las distintas muestras:

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

Este teorema tiene una importante consecuencia: si la variable  $X$  está normalmente distribuida, entonces la media muestral para muestras de tamaño  $n$  estará también normalmente distribuida. Formalmente,

$$X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \sigma^2/n\right)$$

Este resultado puede ser generalizado como un resultado clásico:

**Teorema Central del Límite.** Si muestras **grandes** son extraídas aleatoriamente de una población con media  $\mu$  y varianza  $\sigma^2$ , entonces la distribución de la media muestral será aproximadamente normal, con la media y varianza indicadas anteriormente, *sin importar la forma de la distribución original de la población*.

Este resultado requiere dos comentarios. Por una parte, está claro que mientras más grande sean las muestras mejor es la aproximación a una distribución normal. Es decir, el límite de la distribución de la media muestral es la distribución normal cuando  $n$ , el tamaño de la muestra, tiende a infinito. Por otra parte, este teorema tiene algún grado de ambigüedad. No resulta obvio qué es una muestra “grande” y qué es una muestra “pequeña”. ¿Cuál es el tamaño mínimo de muestra que debemos exigir para que la aproximación sea razonablemente buena?, es una pregunta que no puede ser respondida en forma precisa y varía de población en población.

### 3.2 Estimación

Supongamos que tenemos una población con un parámetro  $\beta$  que nos resulta desconocido, porque es inobservable o, simplemente, no podemos observar. Sin embargo, disponemos de una muestra extraída de esta población. Queremos utilizar la información contenida en esta muestra para construir una conjetura adecuada del valor que puede tener el parámetro poblacional. Un procedimiento como este se llama **estimación**. Diremos que estamos estimando el parámetro poblacional  $\beta$ . Comencemos por un par de definiciones básicas.

Una fórmula usada para estimar un parámetro poblacional, en función de la información contenida en una muestra, es llamada un **estimador**. En general, si el parámetro poblacional en el que estamos interesados lo denominamos  $\beta$ , el estimador será denominado por  $\hat{\beta}$ , o bien por  $b$ . El valor que asume el estimador para una muestra en particular se llama un **estimado**. Como el valor que asume el estimador, o estimado, depende de la muestra que se ha extraído, y las muestras son aleatorias, el estimador es, entonces, también una variable aleatoria. Es decir, el estimador tiene asociada una función de distribución de probabilidades con una media y una varianza.

En general, dado un parámetro poblacional, existen varios posibles estimadores. Es decir, podemos utilizar diferentes fórmulas para estimar su valor a partir de la información contenida en una muestra. Naturalmente, algunos estimadores serán mejores o preferibles a otros. En consecuencia, podemos preguntarnos cuáles son las propiedades que son deseables para un estimador.

En primer lugar, deseamos que nuestros estimadores sean **insesgados**. Con esto queremos decir que si  $\hat{\beta}$  es un estimador de un parámetro poblacional  $\beta$ , entonces tenemos que  $E(\hat{\beta}) = \beta$ . Esto podemos interpretarlo como que el estimador apunta en la dirección correcta, en el sentido de que los posibles errores de estimación son positivos o negativos con igual probabilidad. No hay un error de estimación sistemático. En nuestra discusión anterior habíamos visto que la esperanza matemática de la media muestral es igual a la media poblacional. En consecuencia, la media muestral es un estimador insesgado de la media poblacional. Más precisamente diremos que la media muestral,  $\bar{X}$ , es un estimador puntual insesgado de la media poblacional. Por otra parte, si la varianza de la muestra esta dada por la fórmula:

$$v^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

entonces tenemos que:

$$E(v^2) = \left( \frac{n-1}{n} \right) \sigma^2 \neq \sigma^2$$

Es decir, la varianza muestral es un estimador puntual sesgado de la varianza poblacional, pues su esperanza matemática no es igual al parámetro poblacional. Sin embargo, si definimos

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

éste sí es un estimador puntual insesgado de la varianza poblacional.

Otra propiedad importante de los estimadores es la **eficiencia**. Supongamos que tenemos dos estimadores insesgados  $\hat{\beta}_1$  y  $\hat{\beta}_2$  de un parámetro poblacional  $\beta$ . diremos que  $\hat{\beta}_1$  es más eficiente que  $\hat{\beta}_2$  si y sólo si  $\text{var}(\hat{\beta}_1) < \text{var}(\hat{\beta}_2)$ . Con esto queremos significar que  $\hat{\beta}_1$  estima al parámetro  $\beta$  con mayor precisión que  $\hat{\beta}_2$ . Diremos que estas propiedades de los estimadores son **propiedades de pequeña muestra**, porque son independientes del tamaño de la muestra. Es decir, estas propiedades son válidas no importa el tamaño de la muestra utilizada en su construcción, por más pequeñas que estas sean.

También podemos considerar **propiedades de gran muestra o propiedades asintóticas**. Supongamos que tenemos un estimador cuya varianza decrece a medida que el tamaño de la muestra aumenta, aproximándose más y más a cero. Entonces un estimador insesgado se aproximará más y más al parámetro poblacional hasta finalmente converger en este. Decimos que el estimador colapsa en el valor poblacional. Formalmente,

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\beta}) = 0 \text{ y } \lim_{n \rightarrow \infty} \hat{\beta} = \beta$$

En este caso diremos que nuestro estimador es **consistente**. Nótese que hay un cierto grado de ambigüedad en esta definición. No resulta claro qué queremos decir con gran muestra. Es decir, a partir de qué tamaño de muestra estas propiedades resultan relevantes. En general, este tamaño mínimo depende de la población, del parámetro que está siendo estimado y del estimador. No podemos dar una indicación general sobre el tamaño de muestra a partir del cual los resultados asintóticos resultan relevantes.

Disponemos de una variedad de métodos de estimación. Es decir, de métodos para construir estimadores de parámetros poblacionales. En este lugar sólo mencionaremos tres métodos de estimación que usaremos más adelante en el texto de este libro. En primer lugar, podemos mencionar el método de los **mínimos cuadrados**, probablemente el más influyente método de estimación. Este método y sus detalles constituyen el tema principal de la Parte Dos de este libro, así que postergamos su discusión hasta entonces.

En segundo lugar, tenemos el **método de los momentos**. La idea básica de este método es simplemente estimar los momentos poblacionales de una distribución utilizando los momentos muestrales correspondientes. Desgraciadamente este método no siempre nos entrega estimadores con las propiedades deseadas. Por

ejemplo, nuestra discusión anterior mostró que la media muestral es un estimador insesgado de la media poblacional, pero la varianza muestral es un estimador sesgado de la varianza poblacional.

Finalmente mencionaremos los **estimadores máximo-verosímiles**. En este caso suponemos que tenemos una muestra aleatoria en que cada elemento es seleccionado en forma independiente, con una probabilidad dada por la función de densidad  $f(x_i, \theta)$ , en que el índice  $i$  indica a cada uno de los  $n$  elementos que componen la muestra, y que incluye un parámetro  $\theta$  que queremos estimar. La distribución conjunta de las observaciones muestrales es la probabilidad de extraer estos  $n$  elementos; es decir, de extraer precisamente la muestra que estamos observando. Dado que estos elementos son independientes, esta distribución está dada por el producto de las distribuciones marginales. Formalmente,

$$L = f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta)$$

Esta función es llamada la **función de verosimilitud** y nos indica la probabilidad de que una determinada muestra sea escogida a partir de la población en estudio. Intuitivamente podemos pensar que si observamos una determinada muestra, esta debe, entonces, ser aquella con la mayor probabilidad de ocurrencia o verosimilitud. El método máximo-verosímil consiste entonces en encontrar un valor  $\hat{\theta}$  que maximice el valor de la función de verosimilitud  $L$ . Estos estimadores tienen todas las propiedades deseadas en grandes muestras. Sin embargo, casi siempre sus propiedades de pequeñas muestras son desconocidas. Más aún, no siempre es posible encontrar una solución analítica para el problema de maximizar la función de verosimilitud y es necesario recurrir a complejos y tediosos procedimientos numéricos. La aparición de programas de computador eficientes ha contribuido a la popularización de este método de estimación.

Además de las estimaciones puntuales podemos considerar **estimaciones de intervalo**. Discutiremos esta idea con ayuda de un ejemplo. Supongamos que tenemos una variable aleatoria  $X$  con media  $\mu$  y varianza  $\sigma^2$ . Consideremos una muestra de tamaño  $n$ , que nos permite estimar la media y varianza poblacionales por  $\bar{X}$  y  $s^2$ . Si  $n$  es suficientemente grande o si  $X$  está normalmente distribuida, podemos suponer que  $\bar{X}$  también sigue una distribución normal. Esto es,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Consultando la tabla correspondiente a la distribución normal estándar, podemos escribir que, para un valor arbitrario igual a 0,95,

$$\Pr(-1,96 < Z < 1,96) = 0,95$$

$$\Pr\left(-1,96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1,96\right) = 0,95$$

$$\Pr\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Ahora bien, supongamos que utilizando los 80 valores observados para  $X$  en la muestra obtuvimos que  $\bar{X}=574$  y  $s=94$ . Entonces, las fórmulas anteriores se convierten en

$$\Pr(574 - 20,6 < \mu < 574 + 20,6) = 0,95$$

$$\Pr(553,4 < \mu < 594,6) = 0,95$$

Este resultado lo podemos leer de dos maneras alternativas. Podemos decir que la probabilidad de que la media poblacional esté entre 553,4 y 594,6 es igual al 95 por ciento. Alternativamente podemos decir que para el 95 por ciento de las muestras de tamaño 80 la media muestral estará dentro de este **intervalo de confianza**.

Nótese que el intervalo de confianza está centrado en nuestro estimador, la media muestral. La amplitud de este intervalo depende de la probabilidad que elijamos en forma arbitraria para construirlo. Pero la amplitud del intervalo también depende críticamente de la varianza del estimador. Es decir, para una misma probabilidad, o **nivel de confianza**, un estimador más preciso nos dará un intervalo más estrecho que otro estimador con mayor varianza.





# *preguntas*

Trate de responder a las siguientes preguntas con ayuda de este capítulo.

1. ¿Cuál es el objetivo de la inferencia estadística?
2. ¿Cuál es el objetivo de una estimación?
3. ¿Qué es un estimador?
4. ¿Cuándo un estimador es insesgado?
5. ¿Cuándo un estimador es eficiente?
6. ¿Qué es una estimación Mínimo-Cuadrática?
7. ¿Qué es una estimación Máximo Verosímil?
8. ¿Qué es el Método de los Momentos?



# *ejercicios*



- 3.1 Hemos tomado una muestra de 81 turistas en el Aeropuerto Internacional José Martí y les hemos preguntado por su edad. El resultado es que esta muestra tiene una edad media de 42,5 años con una varianza igual a 36.
- a. ¿En cuánto podemos estimar la edad media de los turistas que llegan a La Habana?
  - b. ¿Cuál es la probabilidad de que el siguiente turista tenga entre 60 y 65 años de edad?
  - c. Construya un intervalo de confianza de 95%. Interprete este resultado.



CAPÍTULO 4



# PRUEBA DE HIPÓTESIS

La prueba de hipótesis, también llamada a veces “test” de hipótesis, es central a la actividad científica. La ciencia puede ser descrita como una gran colección de proposiciones teóricas, o hipótesis, que describen como funciona la realidad o como es ésta realmente. La actividad central de los científicos es acumular evidencia empírica que les permita examinar si la realidad es o funciona realmente como estas hipótesis teóricas sugieren. En este sentido, el objetivo fundamental de la econometría es probar las hipótesis económicas, sobre la base de la información contenida en muestras. Estas hipótesis están fundamentalmente resumidas en los signos y valores de determinados parámetros asociados a los modelos con que representamos la realidad económica.

## 4.1 Hipótesis y Estadígrafos

Las hipótesis deben tener dos características básicas. En primer lugar, las hipótesis deben ser lógicamente consistentes y coherentes. Es decir, no deben llevarnos a contradicciones lógicas, en que afirmamos que una determinada proposición es cierta y falsa al mismo tiempo. De aquí la tendencia de la economía a formalizar en ecuaciones sus razonamientos. En segundo lugar, las hipótesis no pueden ser tautológicas. Es decir, debe existir alguna posibilidad de que sean falsas y deben poder contrastarse con la evidencia empírica. Por ejemplo, si digo “los círculos son redondos”, esto es naturalmente cierto, sin ningún lugar a duda o posibilidades de falsedad, pero no aporta nada al conocimiento. En consecuencia no es una hipótesis válida; es simplemente una tautología. Tampoco puedo decir “el número de visitantes a Varadero en este año depende de los números que ganarán en la Lotería Nacional de España el próximo año”. Esta no es una hipótesis válida

porque no existe evidencia empírica que permita comprobarla, ya que la lotería del próximo año aún no se ha jugado y no podemos saber sus resultados.

La Estadística ha desarrollado un formato formal para realizar pruebas de hipótesis basadas en la información contenida en muestras. Este procedimiento, directamente vinculado a los tópicos anteriores, será el tema central de este capítulo. Una prueba de hipótesis considera cuatro elementos esenciales: una hipótesis nula, una hipótesis alternativa, un estadígrafo de prueba y una región de rechazo.

La **hipótesis nula**,  $H_0$ , es lo que crearemos o supondremos como cierto, hasta que no encontremos evidencia empírica que sugiera lo contrario. La **hipótesis alternativa**,  $H_1$ , es lo contrario de la hipótesis nula. Es decir, si la hipótesis nula es cierta, entonces la hipótesis alternativa es falsa y viceversa. La formulación de una hipótesis nula adecuada y, en consecuencia, también la formulación de una correcta hipótesis alternativa, es crítico para una prueba de hipótesis exitosa. En general, la hipótesis sostenida por nuestra teoría será la hipótesis alternativa y esperamos que la información contenida en la muestra sugiera el rechazo de la hipótesis nula. Debemos insistir en este punto. La evidencia contenida en la muestra no puede rechazar la hipótesis nula pero sí sugerir su rechazo, dentro de cierto marco que discutiremos más adelante. Si la evidencia de la muestra falla en sugerir el rechazo de la hipótesis nula, esto no “prueba” que la hipótesis alternativa sea cierta. Simplemente sugiere que es muy probable que lo sea.

El **estadígrafo de prueba** es un estimador o fórmula que resume la información contenida en la muestra. Este estadígrafo debe tener una distribución conocida, bajo la condición de que la hipótesis nula sea cierta. Ya que conocemos la distribución del estadígrafo de prueba, podemos elegir, en forma relativamente arbitraria, un nivel de probabilidad que nos permite, para una muestra dada y el correspondiente valor del estadígrafo o indicador de prueba, determinar una **región de rechazo**. Si el valor del estadígrafo queda en esta región, decimos que la evidencia empírica sugiere el rechazo de la hipótesis nula. Alternativamente, podemos definir un intervalo en que, para valores del estadígrafo de prueba dentro del mismo, no podemos rechazar la hipótesis nula.

## 4.2 Implementando la Prueba de Hipótesis

En consecuencia la prueba de hipótesis consiste, esencialmente, en construir un estadígrafo, que resume la información contenida en la muestra, y que tiene una distribución conocida bajo el supuesto que la hipótesis nula es cierta. Elegimos un nivel de probabilidad que nos permite definir una región de rechazo. Con esta información debemos decidir si rechazamos o no la hipótesis nula. Esta decisión puede ser correcta o incorrecta.

Nuestra decisión es **correcta** si:

- La hipótesis nula es falsa y decidimos rechazarla.
- La hipótesis nula es cierta y decidimos no rechazarla.

Nuestra decisión es *incorrecta* si:

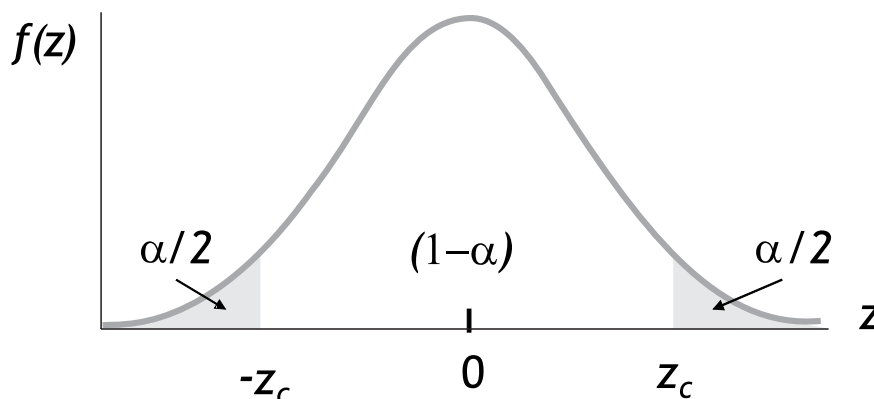
- La hipótesis nula es cierta y decidimos rechazarla. En este caso cometemos un **error tipo I**.
- La hipótesis nula es falsa y decidimos no rechazarla. En este caso cometemos un **error tipo II**.

En general, pensamos que el error tipo II es peor que el error tipo I, porque puede orientar nuestra reflexión teórica por senderos equivocados, generando visiones equivocadas de la realidad. Por este motivo la probabilidad de **no** cometer un error del tipo II es llamada la potencia de la prueba. Una prueba con una mayor potencia es menos probable que cometa un error del tipo II.

Podemos considerar tres formas de definir la región de rechazo. Su relevancia depende de la forma que asuma la hipótesis nula y de la distribución del estadígrafo de prueba:

**Prueba de dos colas.** La región de rechazo incluye ambas colas de la distribución. Es decir, rechazamos la hipótesis nula si el valor del estadígrafo cae en cualquiera de las colas de la distribución. Este tipo de prueba usualmente está asociado a hipótesis nulas que se presentan como igualdades y distribuciones simétricas.

Figura 4.1: Distribución normal estándar prueba de dos colas

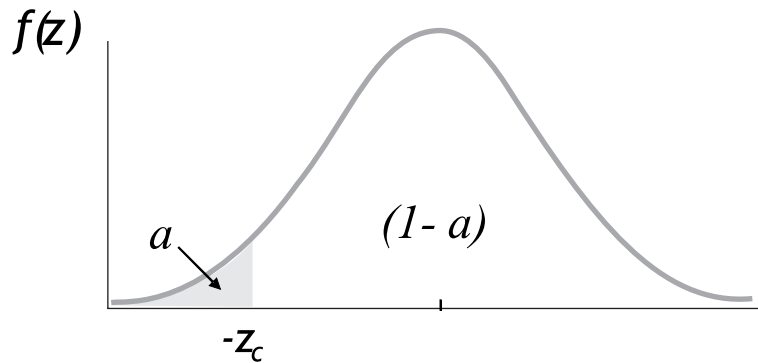


En la Figura 4.1 mostramos una prueba de dos colas. La región sombreada muestra una región de rechazo correspondiente a una probabilidad igual a  $\alpha$  por ciento. Observe que esta probabilidad se distribuye igualmente, con  $\alpha/2$  por ciento en cada cola. Los valores  $z_c$  y  $-z_c$  son los límites del área de rechazo y se los conoce como **valores críticos** de la distribución. El valor  $\alpha$  se lo denomina nivel de significación estadística y el valor  $(1-\alpha)$  es el nivel de confianza. Nótese que los valores críticos determinan un intervalo de confianza con probabilidad  $(1-\alpha)$ .

**Prueba de cola izquierda.** En este caso la región de rechazo está constituida por la cola izquierda de la distribución. Es decir, rechazamos la hipótesis nula si el

valor del estadígrafo se ubica en la cola izquierda de la distribución. Este tipo de prueba está usualmente asociado a distribuciones simétricas y a hipótesis nulas formuladas como desigualdades.

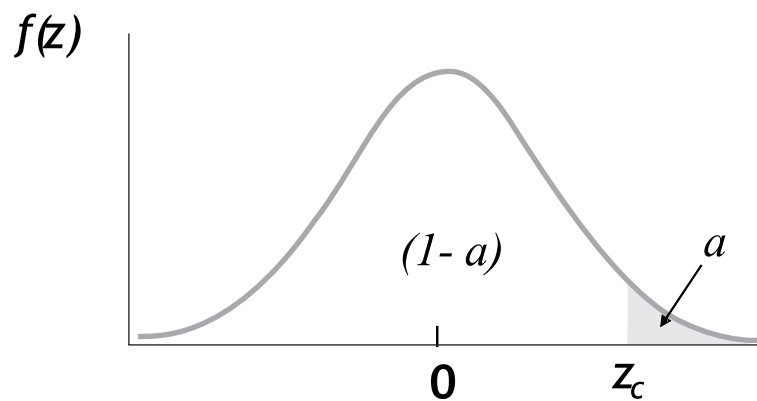
Figura 4.2: Distribución normal estándar  
prueba de cola izquierda



En la figura 4.2 tenemos una prueba de cola izquierda. En este caso la región sombreada en la izquierda de la distribución representa una región de rechazo correspondiente a una probabilidad de  $\alpha$  por ciento.

**Prueba de cola derecha.** En este caso la región de rechazo está constituida por la cola derecha de la distribución. Es decir, rechazamos la hipótesis nula si el valor del estadígrafo se ubica en la cola derecha de la distribución. Este tipo de prueba está usualmente asociado a distribuciones simétricas y a hipótesis nulas formuladas como desigualdades, pero también a distribuciones asimétricas y con soporte positivo.

Figura 4.3: Distribución normal estándar  
prueba de cola derecha





En la figura 4.3 tenemos una prueba de cola derecha. En este caso la región sombreada en la derecha de la distribución representa una región de rechazo correspondiente a una probabilidad de  $\alpha$  por ciento. Nótese que en todas estas figuras, la región no sombreada representa una región de no rechazo de  $(1-\alpha)$  por ciento.

Estas regiones de rechazo pueden también ser descritas con ayuda de proposiciones de probabilidad:

Prueba de dos colas:

$$P(z < -z_c) = P(z > z_c) = \alpha/2$$

$$P(-z_c \leq z \leq z_c) = 1 - \alpha$$

Prueba de cola izquierda:

$$P(z < -z_c) = \alpha$$

$$P(z \geq -z_c) = 1 - \alpha$$

Prueba de cola derecha

$$P(z > z_c) = \alpha$$

$$P(z \leq z_c) = 1 - \alpha$$

Resumiendo lo anterior podemos proponer la siguiente receta para una prueba de hipótesis:

1. Determine la hipótesis nula y la hipótesis alternativa.
2. Especifique el estadígrafo de prueba y la distribución que tendría suponiendo que la hipótesis nula es cierta.
3. Seleccione un nivel de significación  $\alpha$  y determine la región de rechazo de la hipótesis nula. No hay una tradición teórica sólida que nos oriente en la selección de  $\alpha$  pero resulta convencional en econometría el usar valores iguales a 1 por ciento, 5 por ciento o 10 por ciento.
4. Calcule el valor muestral del estadígrafo de prueba.
5. Decida si rechaza o no la hipótesis nula. 



# *preguntas*

Trate de responder a las siguientes preguntas con ayuda de este capítulo.

1. ¿Qué es una hipótesis nula?
2. ¿Qué es un estadígrafo de prueba?
3. ¿Qué es el nivel de significación estadística?
4. ¿Qué es una región de rechazo?
5. ¿Qué es un error de tipo I?
6. ¿Qué es un error de tipo II?

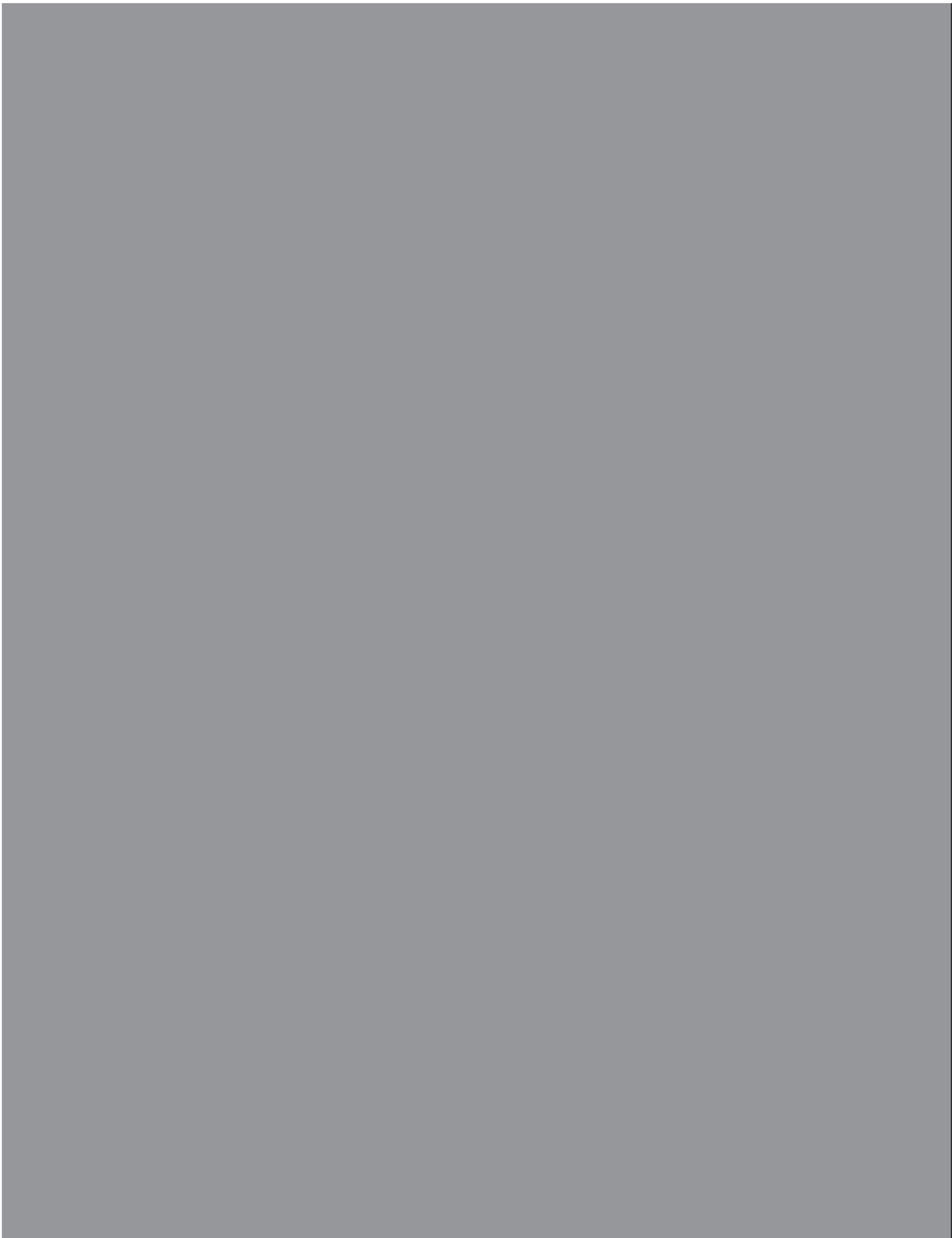


# *ejercicios*



- 4.1 Alguien afirma que un cierto producto cuesta 2,50 pesos en los mercados agropecuarios. Sin embargo, la observación casual permite observar una diversidad de precios, cercanos a 2,50 pesos, pero en ningún caso igual a esta cantidad. Una encuesta en que se observaron 25 precios dio como resultado una media de 2,52 pesos, con una desviación estándar igual a 0,10 pesos.
- a. ¿A la luz de estos resultados podemos aceptar la afirmación anterior?
  - b. ¿Cuál es la probabilidad de cometer un error tipo II?
  - c. ¿Cuál es la potencia de esta prueba?





• PARTE DOS

• EL MODELO  
• LINEAL NORMAL  
• DE  
• REGRESIÓN





En esta parte del libro discutiremos el modelo más utilizado en los estudios econométricos, denominado *modelo lineal normal de regresión*. Este modelo incluye una variable explicada (o dependiente) y, por lo general, varias variables explicativas (o independientes). En este libro obviaremos la explicación del modelo con sólo dos variables (la explicada y sólo una explicativa) y desarrollaremos directamente el modelo multivariado. Para ello nos apoyaremos en el Álgebra Matricial, que facilita el planteo y tratamiento de este modelo. Para recordar las principales operaciones con matrices, se puede consultar el apéndice metodológico correspondiente. Sin embargo, y con el propósito de facilitar la comprensión de algunos tópicos relevantes, en general presentaremos gráficos y otros aspectos ilustrativos con sólo dos variables.

Al modelo lineal normal de regresión también se le conoce como modelo de regresión lineal general o clásico. Si bien se trata de un modelo con supuestos restrictivos, constituye uno de los pilares básicos de la econometría, cuyo dominio es necesario para poder abordar situaciones más generales. De hecho, después de que discutamos el modelo clásico, flexibilizaremos algunos de los supuestos asociados a este modelo y reflexionaremos acerca del impacto de este relajamiento en los estimadores. Este será nuestro objetivo en la Parte Tres.

Esta Parte Dos está compuesta por cuatro capítulos que abordan la presentación del modelo y la estimación de los parámetros por mínimos cuadrados, incluyendo las propiedades de los estimadores obtenidos; los problemas de inferencia estadística a partir de estos estimadores y las pruebas de hipótesis correspondientes; las posibles extensiones del modelo lineal clásico y el problema de la multicolinealidad.

CAPÍTULO 5



## PRESENTANDO EL MODELO

En este capítulo presentamos el **Modelo Lineal Normal de Regresión (MLNR)** y el método adecuado de estimación.

### 5.1 El Modelo

Un elemento esencial del modelo lineal normal de regresión es la variable explicada (o dependiente), que denotaremos generalmente, pero no exclusivamente, como  $Y$ . Esta variable representa el fenómeno objeto de estudio, del cual nos interesa indagar sus fuentes de variación.

El otro componente del modelo son las variables explicativas (también llamadas dependientes o regresores), denotadas como  $X_j$ , donde el subíndice  $j$  representa a cada variable de este tipo con  $j = 2, \dots, k$ . Las  $X_j$  expresan  $k-1$  elementos conocidos que, supuestamente, influyen en el comportamiento de la variable objetivo.

La teoría económica nos sugiere que existe una relación funcional entre esta variable explicada  $Y$  y las variables explicativas  $X_j$ . Es decir, podemos escribir:

$$Y = f(\mathbf{x})$$

donde  $\mathbf{x} = (x_2, x_3, \dots, x_k)$  es un vector que agrupa las variables explicativas. Muy a menudo podemos suponer que una forma lineal representa, o es adecuada para representar esta relación. Entonces escribimos

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k$$

Este es nuestro *modelo teórico*. Sin embargo, en general no observamos toda la realidad sino que sólo una muestra de ella. Por ejemplo, un cierto número de individuos o agentes económicos seleccionados aleatoriamente de la población; o bien un cierto número de períodos de un proceso que es básicamente continuo y se extiende más allá de los períodos observados.<sup>1</sup> Entonces escribimos

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik}, \text{ para } i = 1, \dots, n$$

donde el subíndice  $i$  indica a las observaciones de la muestra (de tamaño  $n$ ), ya sean individuos o períodos. En general, cuando queramos destacar explícitamente que nos referimos a series de tiempo usaremos un índice  $t$  en lugar de  $i$ .

Ya que los valores  $Y_i$  son extraídos de una muestra que suponemos aleatoria, es inevitable considerar que tienen un carácter estocástico. Es decir, se trata de variables aleatorias. Con el objeto de considerar explícitamente este carácter aleatorio, lo supondremos representado por un término aditivo aleatorio  $\varepsilon_i$  y supondremos que las variables explicativas  $x_j$  son números fijos, no aleatorios. Entonces, podemos escribir

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + \varepsilon_i, \text{ para } i = 1, \dots, n \quad (5.1)$$

Este es el *modelo econométrico*, donde distinguimos una parte aleatoria,  $\varepsilon_i$ , y una parte determinística o sistemática,  $\beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik}$ .

Al carácter estocástico de  $Y_i$  contribuyen también otros factores, entre ellos: variables que realmente afectan el comportamiento de  $Y_i$  y no se están incluyendo en el modelo, por escaso desarrollo de la teoría económica (por desconocimiento); errores de medida en las variables del modelo, que pueden estar presentes; ciertos factores que escapan al control de los agentes económicos y que asimismo pueden influir en el comportamiento de determinadas variables económicas, como por ejemplo, las condiciones meteorológicas o ciertos eventos en países con importantes relaciones económicas con el nuestro; e igualmente debemos considerar que las relaciones económicas son, en definitiva, relaciones entre personas o grupos de personas, cada uno con su impronta particular y con elementos impredecibles. Todos estos elementos están incluidos en la parte aleatoria del modelo econométrico.

$\varepsilon_i$  es una variable aleatoria no observable, que también denominamos parte no-sistemática de  $Y_i$ , perturbación aleatoria, término de error o error estocástico.

<sup>1</sup> Recordar nuestra discusión en la Parte Uno sobre cortes de sección y series de tiempo.

Este modelo contiene, además, los coeficientes  $\beta_j$  que expresan la contribución marginal que realiza cada variable  $X_j$  a la explicación de la variable  $Y$ . El coeficiente  $\beta_1$  es una constante que expresa el nivel de la variable  $Y$  común a todas las observaciones, aun sin contribución de sus fundamentos explicativos. Se denomina también *componente autónoma* de  $Y$ .

Las  $Y_i$  y las  $\varepsilon_i$  pueden interpretarse como vectores columna  $\mathbf{Y}$  y  $\boldsymbol{\varepsilon}$ , de dimensión  $(n \times 1)$ . Los  $\beta_j$  forman también un vector columna  $\boldsymbol{\beta}$ , de dimensión  $(k \times 1)$  cuando incluimos a  $\beta_1$  y ampliamos, en consecuencia, el recorrido de  $j=1, 2, \dots, k$ . Las  $X_{ij}$  forman una matriz  $\mathbf{X}$  de dimensión  $(n \times [k-1])$ . A esta matriz le agregamos un vector columna adicional de “unos”, lo que equivale a incluir una variable  $X_1 = (1, 1, \dots, 1)$  que multiplicará a la constante  $\beta_1$ , con lo que obtenemos una matriz  $\mathbf{X}$  de dimensión  $(n \times k)$ . En forma de vectores y matrices desplegadas, podemos describir el modelo como:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{22} & X_{23} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (5.2)$$

Podemos reemplazar los vectores y la matriz por su notación compacta, y tenemos:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.3)$$

Esta ecuación expresa que los cambios en la variable  $Y$  se pueden explicar por un conjunto de otras variables, cada una de las cuales contribuye de cierta manera a esa variación. La contribución total de esas variables es  $\mathbf{X}\boldsymbol{\beta}$ . Sin embargo, en esa expresión reconocemos también que existe una parte de la variación de  $Y$  que no puede ser explicada por causas contenidas en las  $X$  o que sencillamente no sabemos cómo explicar.

La ecuación (5.3) puede representar cualquier modelo económico; por ejemplo, una función de demanda o una función de producción. Disponemos de información sobre  $Y$  y  $X$ . Las  $\varepsilon$  y las  $\beta$  son valores teóricos no observables. La tarea que tenemos por delante es encontrar los valores más probables de  $\varepsilon$  y  $\beta$ , a partir de lo que conocemos de  $Y$  y  $X$  sobre la base de la información contenida en una muestra.

Por ejemplo, si queremos saber en qué medida reaccionan las importaciones de bienes y servicios de Cuba al crecimiento del Producto Interno Bruto (PIB), debemos considerar una función de demanda de importación para nuestro contexto.

La teoría económica sugiere que una función típica de este tipo relaciona la demanda de importaciones con el ingreso o producto del país y con los precios relativos de las importaciones, que son sus factores explicativos más importantes. Si asumimos una forma funcional lineal para esta relación y expresamos las variables involucradas en términos de tasas de variación, tendremos que el correspondiente modelo teórico puede ser escrito como:

$$\dot{m}_t = \beta_0 + \beta_1 \dot{y}_t + \beta_2 \dot{p}_t$$

donde  $\dot{m}$  es la tasa de crecimiento de las importaciones reales de bienes y servicios;  $\dot{y}$ , la tasa de crecimiento del PIB en términos reales;  $\dot{p}$ , la tasa de crecimiento de los precios relativos de las importaciones, expresados en este caso como precios de las importaciones entre precios de las exportaciones. El subíndice  $t$  indica que estamos trabajando con series de tiempo.

El parámetro de interés en este ejemplo es  $\beta_1$ , que expresa en cuantos puntos porcentuales varía la tasa de crecimiento de las importaciones reales por cada punto porcentual de incremento del PIB real.  $\beta_2$  indica la variación de la demanda de importaciones reales por cada punto porcentual de variación de los precios relativos. Esperamos que  $\beta_1$  tenga signo positivo y  $\beta_2$ , negativo. Estos  $\beta$  son los valores teóricos de los parámetros del modelo económico.

## 5.2 Los Supuestos del Modelo

Los parámetros  $\beta$  de nuestro modelo no sólo son desconocidos sino que son inobservables, ya que no podemos acceder a toda la población. En el capítulo 3 discutimos algunos métodos que nos permiten estimar parámetros poblacionales sobre la base de la información contenida en una muestra. En este caso encontraremos los estimadores de este MLNR utilizando el método de los Mínimos Cuadrados Ordinarios. Sin embargo, para obtener estimadores con propiedades adecuadas debemos introducir un cierto número de supuestos sobre este modelo y sus variables. En los párrafos siguientes discutimos estos supuestos.

- A. Las variables explicativas deberán ser **no estocásticas**. Los modelos de regresión y uno de sus métodos de estimación, el de los Mínimos Cuadrados Ordinarios, surgen vinculados a la solución de problemas de las Ciencias Naturales, donde este supuesto puede reflejar bien la realidad. Sin embargo, para las Ciencias Económicas este es un supuesto bastante restrictivo.

Por ejemplo, en nuestro estudio sobre la demanda de importaciones podremos decir que este supuesto sería válido si las variables tasa de variación del PIB real y de los precios relativos de las importaciones no recibieran la influencia de choques aleatorios. Evidentemente esto es bastante improbable.

Con todo, aun cuando esta condición no se cumpla, ello no tendrá mayores consecuencias para la estimación de los parámetros de interés, siempre que no

exista correlación entre las variables explicativas y los términos de error. Volvemos a esta cuestión en el Capítulo 12 de este libro.

**B. Toman valores que son fijos en muestras repetidas.**

Este supuesto también resulta fuerte y difícil de verificar en el área de las Ciencias Económicas, donde es virtualmente imposible repetir experimentos en la práctica, y deberemos conformarnos con los resultados disponibles del desempeño real.

Siguiendo con nuestro ejemplo, difícilmente hubiéramos podido repetir durante los años 90 del siglo pasado las condiciones de crecimiento de la economía cubana de los años 80, y mucho menos de conjunto con ellas, las condiciones de precios para nuestras importaciones.

**C. Son tales que  $(1/n) \sum_{i=1}^n x_{ij}^2 \rightarrow Q_j$ , donde  $Q_j$  son constantes finitas para  $j=2,3,\dots,k$ .**

Esto significa que esas variables deberán tener una varianza finita.

**D. No existe una relación lineal exacta entre los valores muestrales de dos o más de las variables explicativas.**

Es decir, ninguna de las variables explicativas se puede obtener como una combinación lineal de otras variables explicativas incluidas en el modelo. Cuando esto ocurre, decimos que existe multicolinealidad perfecta y la variable en cuestión no está aportando información adicional a la explicación del fenómeno bajo observación.

En términos prácticos esta condición es imprescindible para poder estimar los parámetros del modelo y poder deslindar los efectos individuales de cada variable explicativa sobre la variabilidad de la variable objeto de estudio. Cuando esta condición no se cumple, en realidad, se trata más bien de un problema de la muestra que del modelo de regresión lineal clásico.

En nuestro ejemplo de la función de demanda de importaciones, digamos que cada regresor (tasa de variación del PIB real y de los precios de las importaciones) debe aportar lo suyo a la explicación de la variabilidad de las importaciones reales, y estos efectos no deben solaparse.

**E.  $E(\varepsilon_i) = 0$  para todo  $i=1,\dots,n$ ; es decir, el valor esperado o esperanza de cada  $\varepsilon_i$  deberá ser cero.**

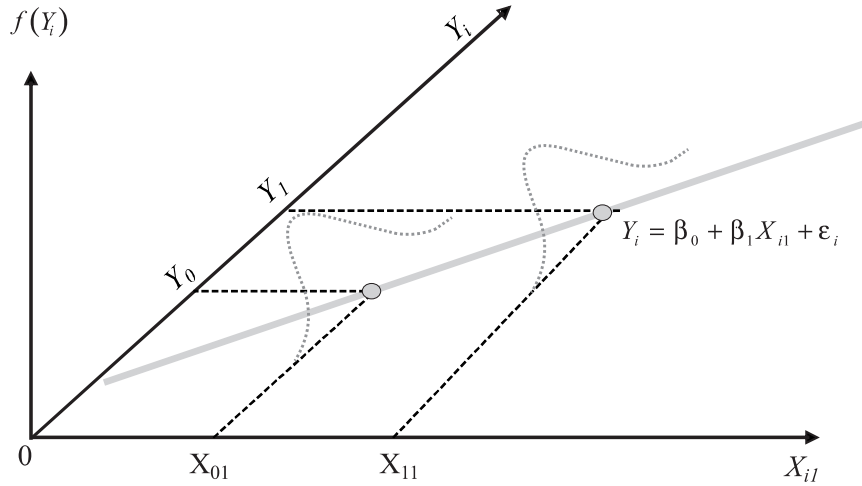
Si el modelo de regresión está bien planteado, debemos tener la misma probabilidad de que aparezcan errores positivos como negativos. No deberá aparecer una tendencia sistemática en la perturbación en ninguno de los dos sentidos.

Para ilustrar este supuesto vamos a tomar como ejemplo el modelo lineal normal de regresión con sólo dos variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

Vamos a asumir que los  $\beta_j$ , con  $j=0,1$  son positivos. En la figura 5.1 observamos que a cada valor específico de  $X$ , por ejemplo  $X_0$ , le pueden corresponder múltiples valores de  $Y$ , según la función de distribución del error  $\varepsilon_i$  que aparece en la figura. Si  $E(\varepsilon_i)=0$ , entonces podemos esperar que el valor más probable de  $Y$  que corresponda a  $X_0$  sea  $Y_0$ .

Figura 5.1: Distribución de los errores con homoscedasticidad



- F.  $Var(\varepsilon_i) = E[\varepsilon_i - E(\varepsilon_i)]^2 = E[\varepsilon_i^2] = \sigma^2$ ; es decir, la **varianza** de los  $\varepsilon_i$  deberá ser **idéntica** para todos los individuos.

En este caso decimos que los errores son **homoscedásticos**, lo que significa que tienen la misma varianza. La figura 5.1 nos sirve también para ilustrar esta condición.

- G.  $Covar(\varepsilon_i, \varepsilon_j) = E[(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))] = E[\varepsilon_i \cdot \varepsilon_j] = E(\varepsilon_i)E(\varepsilon_j) = 0$ ,

para todo  $i \neq j$ ; es decir, la **varianza conjunta** o **covarianza** de los términos de error correspondientes a dos individuos diferentes deberá ser **cero**.

Esto significa que estamos tratando con perturbaciones para distintos individuos con distribuciones independientes unas de otras. En este caso decimos que los errores **no están autocorrelacionados**. En términos prácticos esto quiere decir que el valor del error de una determinada observación no estará afectado por el valor del error en otra observación.

En nuestro ejemplo sobre la demanda de importaciones, esta condición será válida si el comportamiento del término de error de la función de demanda en un año no influye en el error correspondiente al año próximo.



H. Los  $\varepsilon_i$  están distribuidos normalmente.

Asumimos que las perturbaciones tienen una distribución normal. Este supuesto es el que nos permitirá después, hacer inferencia estadística a partir de los resultados de la regresión. Se justifica a partir del Teorema Central del Límite (TCL), que postula que si una variable aleatoria es el resultado del efecto compuesto de muchas otras variables aleatorias, entonces ella tendrá una distribución que se acercará a la distribución normal, aun cuando sus componentes no se distribuyan de esa forma. En el caso que nos ocupa, el término de error se compone de un sinnúmero de factores no explícitos, por lo que asumimos que reproduce este resultado del TCL.

Los supuestos E, F y H nos permiten aseverar que los términos de error se distribuyen normalmente con media cero y varianza  $\sigma^2$ , lo que se denota como:

$$\varepsilon_i \sim N(0, \sigma^2) \quad (5.4)$$

Los supuestos F y G están relacionados con la varianza y covarianza de los errores, por lo que vamos a desarrollar la expresión analítica, la matriz de varianzas y covarianzas de los errores, para ilustrar las implicaciones que tienen estas condiciones en la forma o características de esta matriz.

La matriz de varianzas y covarianzas se obtiene a partir de calcular el valor esperado del producto de los vectores de error, tal cual desarrollamos seguidamente:

$$E(\varepsilon \varepsilon') = E \left( \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix} \right) = E \left( \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1 \cdot \varepsilon_2 & \cdots & \varepsilon_1 \cdot \varepsilon_n \\ \varepsilon_2 \cdot \varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2 \cdot \varepsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n \cdot \varepsilon_1 & \varepsilon_n \cdot \varepsilon_2 & \cdots & \varepsilon_n^2 \end{bmatrix} \right) = \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1 \cdot \varepsilon_2) & \cdots & E(\varepsilon_1 \cdot \varepsilon_n) \\ E(\varepsilon_2 \cdot \varepsilon_1) & E(\varepsilon_2^2) & \cdots & E(\varepsilon_2 \cdot \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_n \cdot \varepsilon_1) & E(\varepsilon_n \cdot \varepsilon_2) & \cdots & E(\varepsilon_n^2) \end{pmatrix}$$

Esta matriz es una matriz simétrica, pues los elementos por encima de la diagonal son un reflejo exacto de los elementos por debajo de la diagonal.

El supuesto F significa que  $E(\varepsilon_1^2) = E(\varepsilon_2^2) = \dots = E(\varepsilon_n^2) = \sigma^2$ , donde  $\sigma^2$  es un escalar constante para todas las observaciones  $i$ , la varianza común. Por eso representamos en la figura 5.1 distribuciones con igual dispersión (amplitud) de los valores de  $Y$  alrededor de su valor esperado (que obtenemos por la línea de regresión), cualquiera sea el valor específico de  $X$ .

Al cumplirse el supuesto de igual varianza de los errores u homoscedasticidad, tendremos una matriz de varianzas y covarianzas con la diagonal ocupada por un mismo elemento, la varianza común.

El supuesto G implica que los términos  $E(\varepsilon_i \cdot \varepsilon_j) = 0$  para todo  $i \neq j$ , por lo tanto sólo tendremos ceros por encima y por debajo de la diagonal de la matriz de varianzas y covarianzas de los términos de error. Es decir, la matriz de varianzas y covarianzas puede expresarse como el producto de un escalar, la varianza común, por la matriz identidad  $\mathbf{I}$ .

$$E(\boldsymbol{\varepsilon} \times \boldsymbol{\varepsilon}') = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix} = \sigma^2 \cdot \mathbf{I}$$

Para resumir las condiciones o supuestos del término de error y sus implicaciones para la variable dependiente, debemos enfatizar que las perturbaciones o errores poseen distribuciones normales idénticas e independientes, con media cero y varianza  $\sigma^2$ . Dado que el carácter estocástico de  $Y_i$  se deriva de su componente  $\varepsilon_i$ , todos estos supuestos se transmiten a nuestra variable independiente. Entonces, tendremos un sistema de supuestos paralelos para  $Y_i$ :

E'.  $E(Y_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik}$  para todo  $i = 1, \dots, n$ , dado que  $E(\varepsilon_i) = 0$ ; es decir, el **valor esperado** de cada  $Y_i$  coincidirá con la parte sistemática del modelo, lo que implícitamente asume que hemos recogido en esa parte sistemática todas las influencias relevantes sobre la variable  $Y$ .

F'.  $Var(Y_i) = \sigma^2$  para todo  $i = 1, \dots, n$ ; es decir, la **varianza** de las  $Y_i$  deberá ser **idéntica** para todas las observaciones y es la misma que la de la perturbación estocástica.

G'.  $Covar(Y_i, Y_j) = 0$ , para todo  $i \neq j$ ; es decir, la varianza conjunta o **covarianza** de las variables dependientes correspondientes a dos individuos diferentes deberá ser **cero**.

H'. Y, finalmente, las  $Y_i$  también estarán distribuidas normalmente.

Es decir, las  $Y_i$  distribuyen normal (con media  $\sum_{j=1}^k X_{ij} \beta_j = \mathbf{X}_i \boldsymbol{\beta}$  y varianza  $\sigma^2$ ) y son independientes.

### 5.3 Estimadores Mínimo Cuadráticos Ordinarios (EMCO)

Ahora vamos a explicar el método que aplicaremos para estimar los parámetros desconocidos de nuestra función de demanda de importaciones para Cuba, así como para estimar los términos de error de esa función. Es importante enfatizar que para este trabajo nos vamos a estar moviendo en dos niveles distintos: la población, desconocida e inaccesible en toda su dimensión, y la muestra conocida.

El modelo general (5.3) representa la función de regresión en la población. Se trata del modelo teórico que captura la verdadera relación que existe entre la  $Y$

y las  $X$ . Si aplicamos el operador esperanza matemática a esta relación, y en virtud de los supuestos B y E del modelo de regresión lineal básico, tendremos que

$$E(Y) = E(X\beta) + E(\varepsilon) = X\beta$$

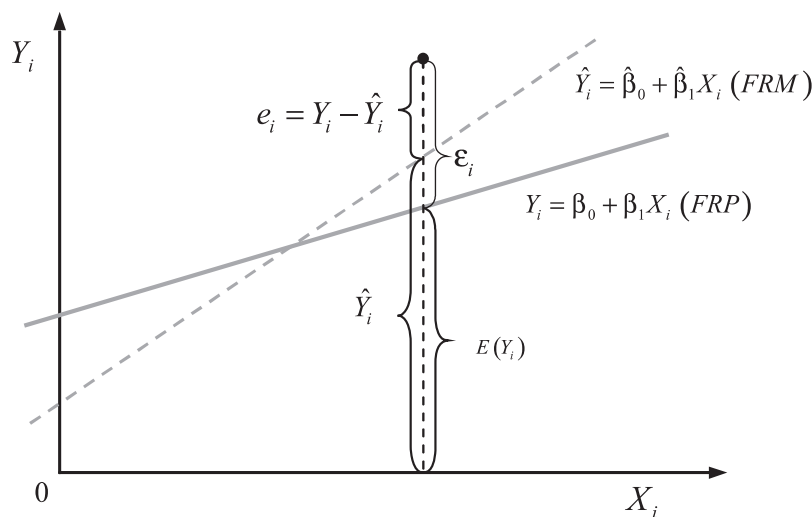
Nuestro propósito es estimar esta función y para ello sólo tenemos a nuestra disposición un conjunto de observaciones que hemos extraído (o más bien surgidas) de la población. A partir de esa muestra de observaciones deberemos conseguir un acercamiento a los parámetros de interés. La función de regresión correspondiente a la muestra, o función de regresión muestral, se define como:

$$Y = X\hat{\beta} + e \quad (5.5)$$

donde  $\hat{\beta}$  es el vector de estimadores de los parámetros  $\beta$  en la población, y  $e$  es el término de error muestral, a través del que estimamos  $\varepsilon$ .

Para ilustrar esta aproximación al problema vamos a representar en un plano ambas funciones de regresión, poblacional y muestral, para el caso de una regresión con dos variables (véase figura 5.2).

Figura 5.2: Función de regresión poblacional y muestral



En la figura 5.2 hemos representado la función de regresión poblacional (FRP) con la línea continua, mientras que la función de regresión muestral (FRM) la hemos dibujado con una línea discontinua. Los parámetros característicos de ambas rectas (interceptos,  $\beta_0$  y  $\hat{\beta}_0$ , y pendientes,  $\beta_1$  y  $\hat{\beta}_1$ ) son diferentes, indicando claramente que no es posible a través de la muestra llegar al conocimiento preciso del valor de estos parámetros en la población. Sin embargo, disponemos de métodos que nos permiten obtener estimadores con muy buenas propiedades para los parámetros de la FRP. Algunos de estos métodos ya los presentamos en el Capítulo

3 de este libro, entre ellos el de los Mínimos Cuadrados Ordinarios (MCO)<sup>2</sup> y el de Máxima Verosimilitud (MV).<sup>3</sup>

La idea central del método de los MCO es minimizar la suma de los residuos o errores de una variable aleatoria alrededor de su media. Precisamos elevar estos residuos al cuadrado, para obtener un resultado de la suma diferente de cero. En el caso de nuestra función de regresión muestral, se trata de minimizar la suma de los errores cuadráticos de esta función. Nótese que estos errores no son otra cosa que la diferencia entre el valor real de  $Y_i$  y su media condicionada a los valores de  $X_{ij}$ , en este caso  $\mathbf{X}\hat{\beta}$ .

La aplicación del método de los MCO para la estimación de los parámetros de la FRM significa encontrar la línea de regresión muestral que mejor se ajuste a las observaciones disponibles. Encontrar una línea o recta quiere decir hallar valores para los parámetros característicos de este lugar geométrico: intercepto y pendiente. La medida del ajuste de las observaciones a la línea de regresión muestral la obtenemos precisamente a través de los valores del término de error en la FRM,  $\mathbf{e}$ . Se trata entonces de encontrar valores tales para los  $\hat{\beta}_j$  que minimicen la suma de los errores de la regresión muestral al cuadrado, tal y como planteamos en (5.6).

$$\min_{\hat{\beta}} \left( \sum_{i=1}^n e_i^2 \right) = \min_{\hat{\beta}} (\mathbf{e}' \cdot \mathbf{e}) \quad (5.6)$$

No conocemos a priori los  $\mathbf{e}$ , por lo que debemos plantear esta expresión (5.6) sólo en función de magnitudes conocidas, a saber, las observaciones de la muestra ( $\mathbf{Y}$  y  $\mathbf{X}$ ) y de los  $\hat{\beta}$  que necesitamos estimar. Para resolver esto, despejamos  $\mathbf{e}$  de la ecuación (5.5) y obtenemos:

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\beta} \quad (5.7)$$

Sustituimos (5.7) en (5.6) y llegamos a:

$$\min_{\hat{\beta}} (\mathbf{e}' \cdot \mathbf{e}) = \min_{\hat{\beta}} \left( [\mathbf{Y} - \mathbf{X}\hat{\beta}]' \cdot [\mathbf{Y} - \mathbf{X}\hat{\beta}] \right) \quad (5.8)$$

A partir de la aplicación de las reglas de la multiplicación de matrices —que pueden consultar en el apéndice metodológico A—, podrán comprobar que el desarrollo de (5.8) conduce a la siguiente expresión:

$$\min_{\hat{\beta}} (\mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}) \quad (5.9)$$

<sup>2</sup> Sus siglas en inglés son OLS (Ordinary Least Squares).

<sup>3</sup> Sus siglas en inglés son ML (Maximum Likelihood).

La ecuación (5.9) es nuestro problema de optimización. Para resolverlo tenemos que calcular la primera derivada del argumento de la función de minimización con respecto a  $\hat{\beta}$ , igualarla a cero, despejar  $\hat{\beta}$ , calcular la segunda derivada de la función, sustituir  $\hat{\beta}$  en la segunda derivada y verificar que esa segunda derivada evaluada en el valor obtenido para  $\hat{\beta}$  sea mayor que cero, lo que garantizará que efectivamente estamos en presencia de un mínimo de la función original (la suma de los errores al cuadrado).

El desarrollo del procedimiento antes descrito nos conduce a la expresión para  $\hat{\beta}$ :

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (5.10)$$

El vector  $\hat{\beta}$  obtenido por este método es una función de los datos observados  $\mathbf{Y}$  y  $\mathbf{X}$ ; así, es precisamente un estimador del vector  $\beta$  poblacional. Dado que  $\mathbf{Y}$  es una variable aleatoria con distribución normal, entonces  $\hat{\beta}$  ostentará las mismas características.

Para la mejor comprensión de este resultado, vamos a escribir las matrices que aparecen en el cálculo de  $\hat{\beta}$  de forma detallada:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} & \cdots & \sum X_{ik} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} & \cdots & \sum X_{i2}X_{ik} \\ \sum X_{i3} & \sum X_{i3}X_{i2} & \sum X_{i3}^2 & \cdots & \sum X_{i3}X_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ik} & \sum X_{ik}X_{i2} & \sum X_{ik}X_{i3} & \cdots & \sum X_{ik}^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & X_{32} & \cdots & X_{n2} \\ X_{13} & X_{23} & X_{33} & \cdots & X_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & X_{3k} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{i2} \\ \sum Y_i X_{i3} \\ \vdots \\ \sum Y_i X_{ik} \end{bmatrix}$$

En las sumatorias anteriores los recorridos son comunes, con  $i$  desde 1 hasta  $n$ . Los hemos omitido para aligerar la notación.

El vector  $\hat{\beta}$  está compuesto por  $k$  componentes. Su primer elemento  $\hat{\beta}_1$  es el efecto promedio de todas las variables aleatorias que no fueron incluidas en el modelo y que tienen una influencia sobre la variable dependiente; se trata de un parámetro de posición del hiperplano correspondiente a la función de regresión muestral; en el caso particular de un modelo con sólo dos variables, es el

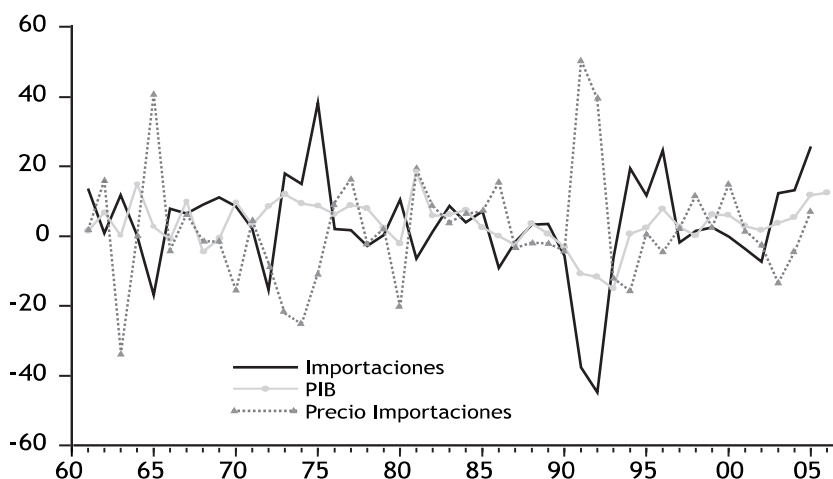
intercepto en el eje de las ordenadas. Los elementos  $\hat{\beta}_j$ , para  $j = 2, \dots, k$ , son los gradientes del hiperplano de regresión con respecto a la variable  $X_j$ ; ellos miden el cambio en la variable dependiente ante cambios en  $X_j$ . Para el modelo biva-riado representan la pendiente de la línea de regresión con respecto a  $X$ .

Vamos a ilustrar la estimación de  $\beta$  en el ejemplo de nuestra función de deman-da de importaciones para Cuba. Para ello debemos convertir el modelo económico teórico que presentamos al finalizar el epígrafe 5.1 en un modelo econométrico, re-saltando el carácter de estimado de los parámetros y del residuo, que se denotan a través de  $\hat{\beta}_j$ , con  $j = 0, 1, 2$ , y  $e_t$ , tal y como se expresa seguidamente:

$$\dot{m}_t = \hat{\beta}_0 + \hat{\beta}_1 \dot{y}_t + \hat{\beta}_2 \dot{p}_t + e_t$$

Para realizar la estimación vamos a utilizar series de datos sobre las importa-ciones de bienes y servicios, el PIB y los términos de intercambio entre los años 1960 y 2005, producidas por de la Oficina Nacional de Estadísticas de Cuba (ONE).<sup>4</sup> Antes de estimar, hemos recalculado convenientemente las series para expresar las importaciones y el producto a precios constantes de 1997 (en términos reales). Después hemos obtenido las tasas de variación de un año a otro (véase Figura 5.3).

**Figura 5.3: Tasa de variación de las importaciones y del PBI cubano (a precios de 1997) y de los precios relativos de las importaciones, 1961-2005**



Fuente: Elaboración propia a partir de ONE

Aplicaremos los MCO para estimar los parámetros de nuestra función de impor-taciones. Todas las operaciones con matrices las hemos realizado utilizando el

<sup>4</sup> Ver datos en el apéndice B.

programa Excel de Microsoft Office, que por lo general está disponible en cualquier computadora y es fácil de usar.

En nuestro caso particular de la función de demanda de importaciones, con dos variables explicativas —las tasas de cambio del PIB real y de los precios de las importaciones—, la matriz  $\mathbf{X}'\mathbf{X}$  viene dada por:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i3}X_{i2} & \sum X_{i3}^2 \end{bmatrix} = \begin{bmatrix} 45 & 161,35146 & 84,26889 \\ 161,35146 & 2428,21514 & -475,12628 \\ 84,26889 & -475,12628 & 11451,06287 \end{bmatrix}$$

La inversa de esta matriz viene dada por:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0,03021 & -0,00207 & -0,00031 \\ -0,00207 & 0,00056 & 0,00004 \\ -0,00031 & 0,00004 & 0,00009 \end{bmatrix}$$

El vector  $\mathbf{X}'\mathbf{Y}$  responde a la siguiente expresión, con los resultados que mostramos en este caso específico:

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{i2} \\ \sum Y_i X_{i3} \end{bmatrix} = \begin{bmatrix} 132,22125 \\ 2284,22320 \\ -6462,65152 \end{bmatrix}$$

Finalmente obtenemos el vector  $\hat{\beta}$  a partir de la multiplicación de la inversa de  $\mathbf{X}'\mathbf{X}$  y del vector  $\mathbf{X}'\mathbf{Y}$ , según (5.10). Este vector contiene los valores estimados para los  $\hat{\beta}_j$ , con  $j = 1, 2, 3$ :

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 0,03021 & -0,00207 & -0,00031 \\ -0,00207 & 0,00056 & 0,00004 \\ -0,00031 & 0,00004 & 0,00009 \end{bmatrix} \begin{bmatrix} 132,22125 \\ 2284,22320 \\ -6462,65152 \end{bmatrix} = \begin{bmatrix} 1,26270 \\ 0,75064 \\ -0,54252 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

Al sustituir estos resultados en la función de regresión muestral, obtenemos la siguiente expresión:

$$\dot{m}_t = \hat{\beta}_0 + \hat{\beta}_1 \dot{y}_t + \hat{\beta}_2 \dot{p}_t + e_t = 1,26270 + 0,75064 \dot{y}_t - 0,54252 \dot{p}_t + e_t$$

Los resultados obtenidos nos indican que:

- ✓ las importaciones reales presentarían una tasa de crecimiento autónoma de 1,26270%,
- ✓ por cada punto porcentual de crecimiento del producto interno bruto real, las importaciones reales se incrementan en 0,75064%,
- ✓ a cada aumento de los precios relativos de las importaciones, equivalente al 1%, corresponde una disminución en la importaciones de 0,54252%.

El sentido de estas variaciones corresponde con el previsto por la teoría económica, sin embargo, ¿cómo podemos asegurar que los parámetros estimados nos dan una referencia adecuada acerca de los parámetros teóricos?, ¿cuándo podemos inferir los parámetros reales a partir de estos estimados? Para poder garantizar esta relación entre parámetros reales y estimados precisamos métodos que produzcan estimadores con ciertas propiedades deseables: insesgadez y eficiencia.<sup>5</sup> Más adelante, en este mismo epígrafe, veremos que el método de estimación de los MCO nos asegura estas propiedades en los estimadores de  $\beta$ , pero esto será así siempre y cuando se compruebe que tanto las variables explicativas como la perturbación aleatoria cumplen los supuestos que discutimos en el epígrafe 5.2.

Entonces, antes de que evaluemos los resultados del modelo de regresión (es decir, antes de lanzarnos a sacar conclusiones acerca de la posible magnitud de la variación de la variable objetivo y del sentido de dicha variación), será preciso que verifiquemos que el conjunto de información que brinda la muestra disponible no brinda suficiente evidencia como para rechazar el cumplimiento de los supuestos antes aludidos. Esta comprobación la postergaremos hasta los capítulos de cierre de la Parte Dos de este libro y su Parte Tres, que se dedicará precisamente a la discusión de los supuestos del MLNR.

Otra posible interrogante respecto de los resultados a que arribamos es sobre su exactitud. ¿Podríamos garantizar que la reacción de las importaciones reales cubanas a un incremento de un punto porcentual del PIB real no fuese 0,8% ó 0,73%?

Es obvio que, dado el carácter de variables aleatorias de estos estimadores, los resultados que hemos presentado no están completos. Hemos obtenido un juego de valores, de los muchos posibles, para los  $\hat{\beta}_j$ .<sup>6</sup> Para completar este ejercicio deberíamos poder calcular varianzas para estos estimadores y ofrecer la respuesta que buscamos en términos de un intervalo de confianza y no de una estimación puntual. Estas varianzas también nos permitirían hacer pruebas de hipótesis sobre los posibles coeficientes de regresión. Una parte de estas cuestiones las vamos a examinar en este capítulo cinco y la otra, en el seis.

¿Por qué sostenemos que los estimadores obtenidos por el método de los MCO son estimadores apropiados para los parámetros de la regresión poblacional? Ellos ostentan propiedades que los hacen buenos representantes de  $\beta$  porque:

- i. Son funciones **lineales** de  $Y$ .

La ecuación (5.10) expresa claramente que obtenemos el vector de estimadores a partir de la multiplicación de una matriz  $C = (X'X)^{-1} X'$  por el vector de observaciones para la variable dependiente. La matriz  $C$  se calcula a partir de la

<sup>5</sup> Véase Capítulo 3.

<sup>6</sup> Eso sí, dada la muestra disponible, este es el juego de valores que minimiza el error cuadrático medio.



matriz  $\mathbf{X}$  y transformaciones de ésta que son valores fijos, de modo  $\mathbf{C}$  está compuesta por constantes que multiplican al vector  $\mathbf{Y}$ . Así,  $\hat{\boldsymbol{\beta}}$  es una función lineal de  $\mathbf{Y}$ .

ii. Son **insesgados**.

Esta es una de las propiedades más importantes para considerar a los  $\hat{\boldsymbol{\beta}}$  representantes adecuados de  $\boldsymbol{\beta}$ , pues significa que  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ; es decir, el valor esperado de  $\hat{\boldsymbol{\beta}}$  es justamente el valor del parámetro en la población  $\boldsymbol{\beta}$ .

Para demostrar que esto es realmente así, vamos a sustituir  $\mathbf{Y}$  según la ecuación de regresión poblacional (5.3) en la fórmula (5.10) para  $\hat{\boldsymbol{\beta}}$ , arribando a  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$ . Aplicamos las reglas de la operación con matrices para simplificar esa expresión y obtenemos:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \quad (5.11)$$

Ahora vamos a aplicar el operador esperanza matemática a la expresión (5.11), para determinar cuál es el valor esperado de  $\hat{\boldsymbol{\beta}}$ :

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'] \cdot E(\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + E[\mathbf{C}] \cdot E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + E[\mathbf{C}] \cdot 0 \\ &= \boldsymbol{\beta} \end{aligned}$$

Para esta demostración nos hemos basado en que el valor esperado de una constante (en este caso  $\boldsymbol{\beta}$ ) es igual a la constante y en la propiedad E del término de error en el modelo lineal normal de regresión. Hemos obtenido que la esperanza del estimador de la función de regresión muestral  $\hat{\boldsymbol{\beta}}$  es igual al vector de parámetros en la función de regresión poblacional  $\boldsymbol{\beta}$ , por lo que podemos sostener que el estimador  $\hat{\boldsymbol{\beta}}$  de  $\boldsymbol{\beta}$  es un estimador insesgado.

iii. Son **consistentes**.

Esta propiedad es la correlativa a la insesgadez en el caso de muestra grande y significa que la distribución muestral de los estimadores  $\hat{\boldsymbol{\beta}}_j$  colapsa en un único valor, que es precisamente el valor del parámetro poblacional  $\beta_j$ , cuando el tamaño de la muestra ( $n$ ) tiende a infinito. Decimos en ese caso que el estimador

$\hat{\beta}_j$  converge en probabilidad a  $\beta_j$ . Expresamos esta propiedad como  $\text{plim}(\hat{\beta}_j) = \beta_j$ , para todo  $j$ .

iv. Son MELI.<sup>7</sup>

Los  $\hat{\beta}_j$  que obtenemos por el método de los MCO son los mejores estimadores lineales insesgados de  $\beta_j$ . Esto quiere decir que además de ser insesgados son eficientes, pues son los estimadores de menor varianza entre los estimadores insesgados de  $\beta_j$ . Como sabemos, la varianza es una medida de dispersión alrededor del valor esperado del estimador, así que con esta propiedad garantizamos que los valores que estimemos van a estar lo más cerca posible de  $\beta_j$ .

Para demostrar esta condición debemos encontrar una expresión para la matriz de varianzas y covarianzas de  $\hat{\beta}$ . Sabemos que, por definición, esta matriz será igual a:

$$\text{var}(\hat{\beta}) = E \left[ (\hat{\beta} - E(\hat{\beta})) \cdot (\hat{\beta} - E(\hat{\beta}))' \right]$$

Pero por la propiedad ii del vector de estimadores  $\hat{\beta}$ , sabemos que  $E(\hat{\beta}) = \beta$ , por lo que la expresión anterior se puede simplificar como:

$$\text{var}(\hat{\beta}) = E \left[ (\hat{\beta} - \beta) \cdot (\hat{\beta} - \beta)' \right] \quad (5.12)$$

Por otra parte, de la ecuación (5.11) se puede obtener una expresión para  $(\hat{\beta} - \beta)$  en función de la matriz  $\mathbf{X}$  y el vector de las perturbaciones  $\varepsilon$ :

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon \quad (5.13)$$

Sustituimos (5.13) en (5.12) y trabajamos en la nueva expresión obtenida para su simplificación: aplicamos las propiedades de la transposición de matrices y el operador esperanza matemática, así como los supuestos del modelo normal lineal de regresión referidos a la homoscedasticidad y no-autocorrelación. El resultado es la siguiente formulación para la matriz de varianzas y covarianzas de los estimadores  $\hat{\beta}_j$ :

$$\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (5.14)$$

<sup>7</sup> El acrónimo equivalente en idioma inglés es BLUE, *best linear unbiased estimator*.

A partir de esa expresión es posible demostrar que ésta es la mínima varianza entre los estimadores insesgados de  $\beta$ . Esta demostración podrá encontrarla el lector inquieto en variados textos de Econometría.<sup>8</sup>

v. Están distribuidos normalmente.

Esta propiedad se deriva de la normalidad de la perturbación  $\varepsilon$ , que se traspa-  
sa a  $Y$  y a su vez a  $\hat{\beta}$ , ya que éstos últimos son una función lineal de  $Y$ , según la  
propiedad i. En este caso los dos momentos que caracterizan la distribución nor-  
mal, su media y su varianza, ya los hemos obtenido de la propiedad ii y en la  
ecuación (5.14), respectivamente. Así, podemos expresar que:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}) \quad (5.15)$$

Queremos hacer notar que hasta este punto hemos desarrollado el método de  
los MCO, hemos llegado a estimar los parámetros de interés con el vector  $\hat{\beta}$ ,  
hemos discutido algunas de sus propiedades y no hemos precisado recurrir a la  
distribución de probabilidad de los errores (que hemos supuesto normal). El méto-  
do de los MCO no requiere que se conozca la distribución de  $\varepsilon$ , es válido para  
cualquier otra distribución continua de la perturbación que no sea normal.

Sin embargo, para estimar por el método de Máxima Verosimilitud sí se precisa  
conocer la distribución de las perturbaciones. Una consecuencia de suponer nor-  
malidad en los  $\varepsilon_i$  es que el estimador MCO de  $\hat{\beta}$ , también será máximo verosímil.<sup>9</sup>

Este supuesto de normalidad nos permite **hacer inferencia estadística**, es de-  
cir, podemos hacer test de hipótesis sobre los parámetros poblacionales.

Hasta este punto hemos conseguido estimadores  $\hat{\beta}_j$  para los  $\beta_j$ , a partir de  
 $Y$  y  $X$ , que son los valores conocidos de la muestra. Sin embargo, para poder rea-  
lizar los test de hipótesis, antes aludidos necesitamos conocer la matriz de varian-  
zas y covarianzas de los  $\hat{\beta}_j$ . Para ello precisaremos, según la ecuación (5.14), de  
un estimador para  $\sigma^2$ , la varianza común de las perturbaciones  $\varepsilon_i$ , que es un es-  
calar desconocido.

¿Cómo conseguir este estimador? Vamos a partir de las funciones de regresión  
poblacional (5.3) y muestral (5.5). Como ya sabemos calcular  $\hat{\beta}$ , podemos encon-  
trar un estimado para  $\sigma^2$  a partir de despejar en la función de regresión muestral:

<sup>8</sup> Véase Dresdner y Vázquez (2004, pp. 78-81), Pindyck y Rubinfeld (2001, p. 63-68), Spanos (1986).

<sup>9</sup> El lector acucioso puede comprobar por sí mismo esta aseveración, a partir de maximizar para  $\beta$  y  $\sigma^2$  el logaritmo de la función de verosimilitud, que en este caso es:

$$l = \ln L = C - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta'X_i)^2$$

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (5.16)$$

En este resultado sustituimos  $\mathbf{Y}$  por la función de regresión poblacional:

$$\mathbf{e} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Acá reemplazamos  $\hat{\boldsymbol{\beta}}$  por su equivalente según (5.11) y desarrollamos la expresión resultante:

$$\begin{aligned} \mathbf{e} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X} \left[ \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \right] \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \\ &= \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} \\ &= \left[ \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \boldsymbol{\varepsilon} \end{aligned}$$

A partir de esta formulación para  $\mathbf{e}$  y de la ecuación (5.16) podemos llegar a encontrar un estimador insesgado para  $\sigma^2$ , que responde a la expresión:

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n-k} \quad (5.17)$$

Así, a partir de la varianza de la muestra específica estimamos la varianza poblacional.

Vamos a estimar la varianza muestral para nuestro ejemplo de la función de importaciones para Cuba, para mostrar cómo proceder y también porque necesitaremos recurrir a esta estimación tan pronto como en el siguiente capítulo.

Ya tenemos algunos de los ingredientes necesarios para este cálculo: los vectores  $\hat{\boldsymbol{\beta}}$  y  $\mathbf{X}'\mathbf{Y}$ . Requerimos, además  $\mathbf{Y}'\mathbf{Y}$  y  $n-k$ . Seguidamente nuestros insumos.

$$\hat{\boldsymbol{\beta}}' = [1,25270 \ 0,75064 \ -0,54252] \text{ y } \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 132,22125 \\ 2284,22320 \\ -6462,65152 \end{bmatrix}, \text{ de donde } \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = 5387,69448$$

Disponemos de 45 observaciones, entre los años 1961 y 2005, y estimamos tres coeficientes de regresión, las dos elasticidades y la constante, de modo que  $n-k = 45-3 = 42$ .

Disponemos de 45 observaciones, entre los años 1961 y 2005, y estimamos tres coeficientes de regresión, las dos elasticidades y la constante, de modo que  $n-k = 45-3 = 42$ .

Entonces, la varianza muestral es

$$\hat{\sigma}^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n-k} = \frac{9151,99831 - 5387,69448}{43} = 89,62628$$

y el desvío estándar correspondiente será  $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{89,62628} = 9,46712$ .

Con estos avances estamos en condiciones de pasar a abordar las cuestiones relativas a la evaluación del modelo econométrico obtenido, tales como la realización de test de hipótesis sobre si los regresores son realmente importantes o no, sobre los posibles valores que ellos deben tomar, sobre las posibles relaciones entre ellos que prescribe la teoría económica y sobre la significación global del modelo. A estos asuntos dedicaremos el próximo capítulo.

Pero antes de adentrarnos en esos temas, vamos a presentarles otra forma de expresar los datos de las variables involucradas en el MLNR que resultará de utilidad tanto para la estimación del modelo mismo, como para el trabajo que tenemos por delante con el Álgebra de Matrices en los capítulos subsiguientes.

## 5.4 Los EMCO en desviaciones de media

Como ya habrán podido advertir, en Estadística y en particular en Econometría frecuentemente se expresan los datos en desviaciones de media; es decir, como la diferencia entre cada observación y el promedio de la muestra para la variable en cuestión, por ejemplo  $(X_{ij} - \bar{X}_j)$  para cada variable explicativa  $X_j$  y  $(Y_i - \bar{Y})$  para la variable dependiente  $Y$ . Cuando trabajamos con los datos de esta forma, se aligera la presentación de las matrices de información para estimar el vector de coeficientes de regresión.

Asimismo, cuando expresamos el modelo con las variables en desviaciones de media, se facilita la interpretación del modelo de regresión. Al utilizar  $(Y_i - \bar{Y})$  y  $(X_{ij} - \bar{X}_j)$  en lugar de  $Y_i$  y  $X_{ij}$ , estamos trabajando directamente con la variabilidad de ambos tipos de variables con relación a sus promedios. Esta forma de trabajar es más afín con la esencia del MLNR, a saber: explicar la variabilidad de la variable regresada, endógena o dependiente a partir de la variabilidad de las variables regresoras, explicativas o independientes. Así, el modelo expresado en desviaciones de media nos informa directamente qué parte del cambio en la variable dependiente se explica por el modelo postulado por la teoría económica.

Vamos a continuar trabajando con matrices y vectores. ¿Cómo expresar  $\mathbf{X}$ ,  $\mathbf{Y}$  y  $\boldsymbol{\varepsilon}$  en desviaciones de medias? En el caso de la matriz  $\mathbf{X}$  sustraemos a cada vector columna que la compone (que contiene las  $n$  observaciones correspondientes a la variable  $X_j$  y que representamos como  $\mathbf{X}_j$  con  $j = 1, \dots, k$ ) el promedio de la variable explicativa correspondiente

$$\bar{X}_j = 1/n \sum_{i=1}^n X_{ij}$$

A cada elemento del vector  $\mathbf{Y}$  igualmente le deducimos  $\bar{Y}$ . Y de la misma forma procedemos con el vector  $\boldsymbol{\varepsilon}$ , restándole  $\bar{\varepsilon}$ . Así, transformamos la información de la muestra y denotamos la matriz y los vectores resultantes con letras minúsculas, y en el caso del término de error con el símbolo  $\tilde{\varepsilon}$ .

$$\mathbf{x} = \begin{bmatrix} X_{12} - \bar{X}_2 & X_{13} - \bar{X}_3 & \cdots & X_{1k} - \bar{X}_k \\ X_{22} - \bar{X}_2 & X_{23} - \bar{X}_3 & \cdots & X_{2k} - \bar{X}_k \\ \vdots & \vdots & \ddots & \vdots \\ X_{n2} - \bar{X}_2 & X_{n3} - \bar{X}_3 & \cdots & X_{nk} - \bar{X}_k \end{bmatrix} = \begin{bmatrix} x_{12} & x_{13} & \cdots & x_{1k} \\ x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix} = [\mathbf{x}_2 \quad \mathbf{x}_3 \quad \cdots \quad \mathbf{x}_k]$$

$$\mathbf{y} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 - \bar{\varepsilon} \\ \varepsilon_2 - \bar{\varepsilon} \\ \vdots \\ \varepsilon_n - \bar{\varepsilon} \end{bmatrix} = \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{bmatrix}$$

Adviertan que eliminamos la columna correspondiente a  $\mathbf{X}_1$  en la matriz  $\mathbf{x}$ , ya que estaría compuesta por ceros. Esto equivale a haber eliminado la constante del modelo de regresión.

Si a nuestro MLNR (5.1) le aplicamos el operador de esperanza obtenemos la siguiente expresión:

$$E(Y_i) = \beta_1 + \beta_2 E(X_{i2}) + \beta_3 E(X_{i3}) + \cdots + \beta_k E(X_{ik}) + E(\varepsilon_i) \quad (5.18)$$

Si restamos (5.18) de (5.1), llegamos a:

$$Y_i - E(Y_i) = \beta_1 + \beta_2 [X_{i2} - E(X_{i2})] + \beta_3 [X_{i3} - E(X_{i3})] + \cdots + \beta_k [X_{ik} - E(X_{ik})] + [\varepsilon_i - E(\varepsilon_i)]$$

Esa expresión puede ser escrita de forma compacta a partir de  $\mathbf{x}$ ,  $\mathbf{y}$  y  $\tilde{\boldsymbol{\varepsilon}}$  como:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}} \quad (5.19)$$

El modelo (5.19) es la función de regresión poblacional. La función de regresión muestral correspondiente será:

$$\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \tilde{\mathbf{e}}$$

donde  $\tilde{\mathbf{e}} = \mathbf{e} - \bar{\mathbf{e}}$ . Ya que  $\bar{\mathbf{e}} = \mathbf{0}$ <sup>10</sup>, el modelo se puede reescribir como

$$\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \mathbf{e} \quad (5.20)$$

<sup>10</sup> Recuerden que estamos obteniendo  $\hat{\boldsymbol{\beta}}$  por MCO.

El método de los MCO lo que persigue es minimizar la suma de residuos al cuadrado. Si aplicamos los MCO al modelo en (5.20)<sup>11</sup> llegaremos a la siguiente solución para los EMCO:

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y} \quad (5.21)$$

Aparentemente, esta solución es igual a la que obtuvimos en (5.10). Sin embargo, recuerden que hemos eliminado la columna correspondiente a  $\mathbf{X}_1 = [1 \ 1 \ 1 \ \dots \ 1]$  de la matriz de datos de las variables explicativas  $\mathbf{X}$ , por lo tanto, a través de (5.21) obtendremos los coeficientes de pendientes parciales  $\beta_j$ , para  $j = 2, 3, \dots, k$ . La constante o coeficiente de posición podremos recuperarlo a partir de (5.18).

$$\hat{\beta}_1 = E(Y_i) - \hat{\beta}_2 E(X_{i2}) - \hat{\beta}_3 E(X_{i3}) - \dots - \hat{\beta}_k E(X_{ik}) = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 - \dots - \hat{\beta}_k \bar{X}_k = \bar{Y} - \hat{\beta} \bar{\mathbf{X}}$$

Así, podremos trabajar con una matriz más pequeña en el momento de estimar  $\hat{\beta}$ , lo que facilita el trabajo, sobre todo cuando no se dispone de medios de cómputo. Esta variante también nos facilitará algunas de las demostraciones que haremos en capítulos posteriores.

<sup>11</sup> Esta es otra buena oportunidad para reactivar los conocimientos de cálculo diferencial y álgebra de matrices. Anímense y comprueben lo que les decimos.





# *preguntas*

Trate de responder las siguientes preguntas para que compruebe si ha comprendido los contenidos básicos de este capítulo. Puede re-consultar el texto si no recuerdan bien o tienen alguna duda.

1. Plantee un modelo de regresión lineal normal (MRLN) en forma matricial. Indique cuáles son sus elementos o componentes fundamentales.
2. ¿Por qué incluir un término de error estocástico en el modelo de regresión?
3. ¿Por qué debemos imponer ciertos supuestos a las variables y perturbación estocástica de un MRLN?
4. ¿Cuáles son los supuestos asociados a las variables explicativas del MRLN? ¿cuáles al término de error estocástico?
5. ¿Qué consecuencias tienen estos supuestos para la variable dependiente o endógena del MRLN?
6. ¿Qué diferencias existen entre el MRLN poblacional y el correspondiente a una muestra de observaciones de las variables del modelo?
7. ¿Qué métodos de estimación podemos emplear para conseguir los parámetros de un MRLN?
8. ¿En qué se basa el método de los mínimos cuadrados ordinarios?
9. ¿Cuál es la fórmula para estimar los coeficientes de un MRLN?
10. ¿Por qué decimos que  $\hat{\beta}$  es una variable aleatoria? Caracterice su distribución.
11. ¿Por qué decimos que  $\hat{\beta}$  es un buen representante de  $\beta$ ? ¿Cuáles son las propiedades de los estimadores  $\hat{\beta}$ ?



# *ejercicios*



- 5.1 Plantee los modelos teórico y econométrico que correspondan a las situaciones que se describen seguidamente.
- a. Un vendedor de carne de cerdo que opera en un Mercado Agropecuario cubano, desea conocer cuál podría ser la respuesta de la demanda de este producto ante cambios en su precio (elasticidad-precio de la demanda). Esta información le serviría de respaldo para determinar el nivel de precios que maximice sus ganancias. Ayúdele a esclarecer cuáles variables deben estar presentes en la función de demanda de la carne de cerdo, a plantear los modelos correspondientes y a diseñar el experimento a través del cual puede llegar a estimar el parámetro que desea conocer.
  - b. El director comercial de una cadena hotelera cubana desea conocer cómo reacciona la demanda de viajes en cierto país de América ante los gastos en publicidad de su entidad. Esto le permitiría realmente saber si está en condiciones de recortar los gastos de promoción de su producto y que no se afecten las ventas de la cadena. Ayúdele a plantear los modelos teórico y econométrico que podrían contribuir a esclarecer si tiene o no reservas para reducir sus gastos de publicidad.
  - c. La directora de recursos humanos de una planta de producción de fármacos desea fundamentar adecuadamente los gastos en capacitación del personal de su entidad a partir de su posible contribución al crecimiento de la producción de su planta. Ayúdela a encontrar un modelo teórico apropiado para el fin que persigue, así como a plantear el modelo econométrico y a cómo conseguir los datos de las variables que necesitará.
  - d. El presidente de una cooperativa para la producción de cítricos de provincia Habana desea saber si precisa comprar algún fertilizante adicional al asignado a su entidad, aprovechando la apertura de las nuevas tiendas para la venta de insumos agropecuarios. Ayúdele a resolver este problema, planteando las funciones que requiere estimar y los modelos correspondientes.
- 5.2 La tabla B.2 del disco acompañante contiene datos para el producto interno bruto cubano y sus principales factores de producción para el período 1980-2000. Según la teoría económica convencional, existen diversas posibles funciones de producción para explicar la variación del producto ( $Q$ ) a partir de cambios en sus factores ( $L$ , trabajo;  $K$ , capital físico y  $H$ , capital humano). Algunas de ellas pueden ser, explicar:
- a. las primeras diferencias del producto a partir de las primeras diferencias de los factores productivos;

- b. los niveles del producto a partir de los niveles de aplicación de factores de producción;
- c. los logaritmos de los niveles del producto a partir de los logaritmos de los niveles de aplicación de factores de producción.

Plantee los modelos correspondientes y estime sus coeficientes. Interprete los y discuta su coincidencia o no con los postulados teóricos. En caso de discrepancias, cuáles podrían ser las fuentes de las mismas en nuestro contexto particular.

CAPÍTULO 6



## INFERENCIA Y PRUEBA DE HIPÓTESIS

En este capítulo discutiremos los conceptos relacionados con las pruebas antes aludidas. Como ya habíamos comentado, para estimar los regresores a través del método de los MCO no se requiere considerar supuestos especiales sobre la distribución probabilística de los errores del modelo. Es suficiente asumir que dicha distribución tiene la característica de poseer media cero y varianza constante. Pero como queremos hacer inferencia estadística sobre los valores de los parámetros en la población a partir de los resultados muestrales, debemos entonces incorporar un supuesto sobre la distribución de estos errores. La idea básica es que, si queremos conocer con qué probabilidad es posible observar un determinado valor del parámetro, debemos conocer la forma de su distribución.

### 6.1 Pruebas de Hipótesis e Intervalo de Confianza

El supuesto H relativo al componente de error del modelo impone que su distribución es *normal*, por lo tanto, los  $Y_i$  también se distribuyen de esa forma. Asimismo, sabemos que la función de regresión es *lineal en los parámetros*, o sea, que  $\hat{\beta}_j$  es una combinación lineal de los valores de  $Y_i$ . De acuerdo a un teorema básico de la estadística, cualquier combinación lineal de los  $Y_i$  se distribuirá igualmente normal, de modo que los  $\hat{\beta}_j$  tienen esa distribución.

En el capítulo 5 habíamos llegado también a una respuesta acerca de la media y la varianza del vector  $\hat{\beta}$ : demostramos que  $\hat{\beta}$  es un estimador insesgado de  $\beta$ , por lo que su media coincide con  $\beta$ , y su varianza es igual a  $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$  (véase la ecuación

(5.14)). Así, podemos decir que los parámetros estimados  $\hat{\beta}$  distribuyen en forma normal, con media  $\beta$  y varianza  $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ , lo que denotamos con la expresión  $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ .

Nuestra variable aleatoria normal  $\hat{\beta}$  puede ser estandarizada, tal cual se muestra seguidamente:

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2(\mathbf{X}'\mathbf{X})^{-1}}} \sim N(0,1) \quad (6.1)$$

A partir de (6.1) podríamos construir un intervalo de confianza para la estimación de  $\hat{\beta}$  y realizar pruebas de hipótesis utilizando directamente la distribución normal estándar.

Sin embargo, el valor de la varianza común  $\sigma^2$  no es conocido, sólo disponemos de un estimador para el mismo (véase la ecuación (5.17)). La solución a esta dificultad es crear una nueva variable Chi-Cuadrado, con la cual transformaremos nuestra distribución normal estándar en una distribución t de Student, para poder utilizar el estimador  $\hat{\sigma}^2$  en lugar de  $\sigma^2$ .

En la construcción de esta variable Chi-Cuadrado vamos a poner en práctica los conocimientos que adquirimos en el capítulo 2 de este libro. Partimos del supuesto de que  $\varepsilon_i \sim N(0, \sigma^2)$ . Estandarizamos los  $\varepsilon_i$ : en este caso la media es cero, por lo que sólo debemos dividir entre  $\sigma$  para este proceder. Si elevamos al cuadrado y sumamos los  $n$  elementos del vector  $\varepsilon$  estandarizado, estamos construyendo una variable  $\chi^2$  con  $n$  grados de libertad; es decir,

$$\sum_{i=1}^n \left( \frac{\varepsilon_i}{\sigma} \right)^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{\sigma^2} \sim \chi_{(n)}^2 \quad (6.2)$$

En la expresión (6.2) no conocemos los  $\varepsilon_i$ , por lo que los sustituimos por sus estimadores, los residuos muestrales  $e_i$ . A diferencia de los errores, los residuos muestrales no son independientes, dado que dependen de los  $\hat{\beta}_j$ , por lo que sólo se precisan  $n-k$  observaciones de  $Y_i$  y  $X_{ij}$  para obtener los  $n$  residuos. Así, nuestros grados de libertad se reducen a  $n-k$ . Sabemos también que:

$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}$  (véase la expresión (5.17)). Entonces, podemos plantear que:

$$\frac{\sum_{i=1}^n e_i^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-k)}^2 \quad (6.3)$$

En el capítulo 2 de este libro también explicamos que al dividir una variable aleatoria normal estándar entre la raíz cuadrada de otra variable aleatoria que distribuya Chi-Cuadrado —dividida a su vez entre sus grados de libertad—, se obtiene una variable aleatoria con una distribución de probabilidades del tipo t-Student, con tantos grados de libertad como la variable  $\chi^2$  en el denominador de la fracción.

Aplicamos este procedimiento a nuestras variables, la normal estándar de (6.1) y la Chi-Cuadrado de (6.3), y obtenemos una variable que distribuirá t-Student con  $n-k$  grados de libertad:

$$\frac{\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}}}}{\sqrt{\frac{(n-k)\hat{\sigma}^2/\sigma^2}{n-k}}} = \frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}}} \sim t_{(n-k)} \quad (6.4)$$

Hemos conseguido una expresión para la variable resultante que nos permite computarla a partir de valores que conocemos (la matriz  $\mathbf{X}$ ) o que podemos estimar (el vector  $\hat{\beta}$  y el escalar  $\hat{\sigma}^2$ , calculados a partir de  $\mathbf{X}$  y  $\mathbf{Y}$ ). El conocimiento de la distribución t-Student nos permitirá realizar las pruebas de hipótesis que necesitamos. La notación de tipo matricial se puede expresar también en términos de cada regresor como

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{(n-k)} \quad (6.5)$$

donde  $a_{jj}$  es el elemento de la diagonal de la matriz  $(\mathbf{X}'\mathbf{X})^{-1}$  que corresponde para obtener la varianza de  $\hat{\beta}_j$ . El denominador de esa fracción es la desviación estándar de  $\hat{\beta}_j$ , que denotaremos con  $se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 a_{jj}}$ .

Con esta variable aleatoria  $t$  podemos construir un intervalo de confianza para realizar las pruebas de hipótesis, con un error de tipo I o nivel de significación que denotaremos  $\alpha$ . A este nivel de significación y a los grados de libertad de que dispongamos, corresponderá un valor crítico para el estadístico  $t_{(n-k)}^\alpha$ , que nombramos  $t_c$  y que se encuentra debidamente tabulado (véase nuestro apéndice C de tablas estadísticas). La proposición de probabilidad que sostiene este planteamiento es la siguiente:

$$P(t < -t_c) = P(t > t_c) = \alpha/2$$

$$P(-t_c \leq t \leq t_c) = 1 - \alpha$$

Sustituyendo  $t$  por su expresión en (6.5) llegamos a:

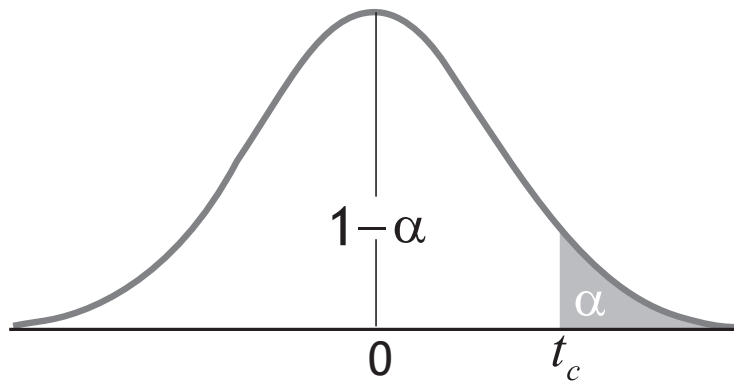
$$P\left[-t_c \leq \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \leq t_c\right] = 1 - \alpha$$

$$P\left[\hat{\beta}_j - t_c se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_c se(\hat{\beta}_j)\right] = 1 - \alpha$$

El intervalo alrededor de un determinado  $\beta_j$  —el valor hipotético del regresor en la población—, que construimos a partir de  $\hat{\beta}_j \pm t_c se(\hat{\beta}_j)$ , lo denominamos intervalo de confianza bilateral  $(1 - \alpha) \times 100\%$  para  $\beta_j$ .

Esta proposición de probabilidad para el intervalo de confianza la hemos representado gráficamente, para su mejor comprensión, en la figura 6.1.

**Figura 6.1:** Intervalo de confianza  $(1 - \alpha) \times 100\%$  para  $\beta_j$ , test bilateral



Área sombreada= Región de rechazo para test unilateral

Si nuestro  $\beta_j$  hipotético se encuentra comprendido en el intervalo de confianza que hemos construido a partir del  $\hat{\beta}_j$  estimado, podremos decir que este último no difiere estadísticamente de  $\beta_j$ . Con una probabilidad de  $(1 - \alpha) \times 100\%$ , el valor verdadero del regresor en la población será  $\beta_j$ . Así, la evidencia empírica sustenta el valor previsto de  $\beta_j$ .

Otra forma de verificar si la evidencia empírica no contradice nuestra hipótesis sobre  $\beta_j$  es sencillamente calcular el valor de:



$$t = \left| \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \right|$$

y después ver qué probabilidad se asocia a ese valor, dado que sabemos distribuye  $t_{(n-k)}$ . Llamaremos  $p$  a esa probabilidad y la regla de decisión es la siguiente: si  $p \leq \alpha/2$ , entonces se rechaza la hipótesis de que  $\beta_j$  y  $\hat{\beta}_j$  son estadísticamente iguales; de lo contrario, no se rechaza ese planteamiento. Habrán notado que hemos comparado  $p$  con  $\alpha/2$ ; es porque en este caso estamos testeando una hipótesis bilateral, tal y como lo hemos representado en la figura 6.1. El valor verdadero del regresor se ubica en un intervalo alrededor de  $\beta_j$ , puede ser mayor, pero también menor.

Estos tests t-Student podemos aplicarlos para investigar cualquier combinación lineal de los coeficientes de regresión. Esta posibilidad incluye desde verificar si cada uno de ellos por separado puede asumir un determinado valor o cae en un cierto intervalo, hasta verificar si la combinación de ellos cumple una determinada condición lineal. Vamos a iniciar nuestra discusión por el caso más sencillo, que también es uno de los más relevantes.

## 6.2 Test de Significación de los Regresores

Entre las múltiples pruebas de hipótesis interesantes, una bien importante es la de significación de los regresores. Testear la significación de los regresores quiere decir verificar si los coeficientes estimados obtenidos a través del procedimiento de los MCO son estadísticamente distintos de cero con un cierto nivel de confianza. Esta es una prueba crítica ya que la hipótesis básica detrás del modelo de regresión lineal planteado es que la variación de cada una de las  $X_j$  tiene un efecto sobre  $Y$ .

En este caso el valor hipotético de  $\beta_j$  es cero, esta es la hipótesis nula; la alternativa correspondiente será que el regresor es diferente de cero. El planteamiento es el siguiente:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

No rechazar  $H_0$  significará en este caso que la  $X_j$  correspondiente no tiene efecto sobre  $Y$ .

Todos los programas econométricos incluyen estos tests de significación entre sus resultados estándar. Sin embargo, a partir del programa Excel de Microsoft Office se pueden calcular fácilmente los valores del estadístico  $t$  para cada regre-

sor con  $\beta_j = 0$ . En este caso  $t$  se computa directamente como el regresor entre su desvío estándar estimado a partir de la muestra disponible, o sea  $\hat{\beta}_j / se(\hat{\beta}_j)$ . Una vez obtenido el valor de  $t$ , encontramos la probabilidad asociada al mismo. Excel también trae incorporada la función de distribución probabilística  $t$  de Student.<sup>1</sup> Si esta probabilidad es menor que el error de tipo I con que trabajamos, entonces rechazamos  $H_0$  y diremos que la variable  $X_j$  es significativa.

Vamos a aplicar este test de significación de los regresores a nuestro ejemplo de la función de importaciones cubana.

Ya en el capítulo 5 habíamos estimado el vector  $\hat{\beta}$ , con

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 1,26270 \\ 0,75064 \\ -0,54252 \end{bmatrix}$$

Ahora debemos estimar los desvíos estándar de los  $\hat{\beta}_j$ . Por nuestro trabajo previo en el capítulo 5 ya disponemos de una estimación para el desvío estándar del error, que asciende a 9,46712. También tenemos la matriz

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0,03021 & -0,00207 & -0,00031 \\ -0,00207 & 0,00056 & 0,00004 \\ -0,00031 & 0,00004 & 0,00009 \end{bmatrix}$$

Sólo nos resta seleccionar los elementos necesarios de su diagonal para obtener  $se(\hat{\beta}_j)$ .

Comencemos por calcular el valor del estadístico  $t$  correspondiente a  $\hat{\beta}_1$ , que es el regresor correspondiente a la variable  $\dot{y}$  (la tasa de crecimiento del Producto Interno Bruto en términos reales). En este caso tenemos:

$$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 a_{11}}} = \frac{0,75064}{\sqrt{89,62628 \cdot 0,00056}} = \frac{0,75064}{0,22338} = 3,36039$$

A este valor de  $t$  corresponde una probabilidad de 0,0017, muy por debajo del valor que prefijamos para  $\alpha/2 = 2,5\% = 0,025$ , por lo que se rechaza la hipótesis de que  $\beta_1 = 0$ .

Otra forma de verificar dicha hipótesis es comparar ese valor de  $t$  con el valor

<sup>1</sup> Igualmente trae otros tipos de distribuciones de amplia aplicación en el trabajo econométrico como la normal y la Chi-cuadrado.

crítico correspondiente que aparece en la tabla del apéndice estadístico  $t_c = t_{(n-k)}^{\alpha/2} = t_{42}^{0.025} = 2,021$ . Habrán notado que el valor de  $t$  que estamos utilizando como valor crítico corresponde a 40 grados de libertad; es porque en la tabla del apéndice no aparece el valor correspondiente a 42 grados, entonces hemos escogido el que más se le acerca. El valor que obtuvimos para  $t = 3,36039$  es mayor que este valor crítico; así, este resultado nos conduce igualmente al rechazo de la hipótesis de que  $\beta_1$  sea igual a cero, ¡no podía ser de otra manera!

En el caso de  $\hat{\beta}_2$  —el regresor correspondiente a la variable  $\dot{p}$ , que representa la tasa de crecimiento de los precios relativos de las importaciones—, obtenemos el siguiente resultado después de aplicar idéntico procedimiento:

$$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 a_{jj}}} = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 a_{22}}} = \frac{-0,54252}{\sqrt{89,62628 \cdot 0,00009}} = \frac{-0,54252}{0,09040} = -6,00112$$

A este valor de  $t$  corresponde una probabilidad de 0,0000004, también inferior al valor de referencia de 0,025. Igualmente podemos corroborar que el valor absoluto de  $t$ ,  $|t| = 6,00112$ , es mayor que su valor crítico.

Entonces podemos concluir que en este ejemplo ambos regresores son significativamente distintos de cero y que, por lo tanto, las tasas de variación del PIB y de los precios relativos de las importaciones cubanas sí resultan relevantes para explicar la tasa de variación de las importaciones del país. Esta información puede ser de utilidad para las autoridades de planificación, ya que les permitiría predecir cuál sería la demanda futura de importaciones dado el crecimiento previsto de PIB y el comportamiento previsible de los precios relativos de las importaciones. Sin embargo, antes de utilizar nuestro modelo de esta forma, deberemos cerciorarnos de que cumple los supuestos que discutimos en el capítulo 5. Al finalizar la Parte Dos de este libro y en su Parte Tres proveeremos las herramientas requeridas para cerciorarnos de que esto sea así.

### 6.3 Comparando los $\beta_k$ con Valores Específicos

Otro test interesante es el de verificar si se puede considerar que el regresor estimado es menor (o también mayor) que un cierto valor. Se trata en este caso de tests unilaterales y se procede de la misma manera que en el caso que ya examinamos del test bilateral, sólo que trabajaremos con  $\alpha$  y no con  $\alpha/2$  para encontrar el valor crítico de  $t$ , así como para comparar la probabilidad correspondiente.

Para nuestro caso de estudio podríamos considerar si el signo del coeficiente de regresión correspondiente a la tasa de variación de los precios de las importaciones es negativo, tal cual predice la teoría económica. El planteamiento de las hipótesis nula y alternativa en ese caso es el siguiente:

$$H_0 : \beta_2 \geq 0$$

$$H_1 : \beta_2 < 0$$

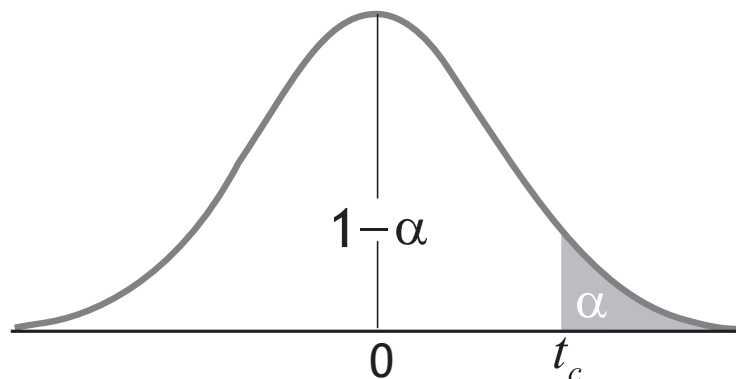
La proposición de probabilidad correspondiente es:

$$P(t < t_c) = \alpha$$

$$P(t \geq t_c) = 1 - \alpha$$

Ilustramos gráficamente esta situación en la figura 6.2.

**Figura 6.2: Test-t unilateral para  $\beta_j \geq 0$**



Área sombreada= Región de rechazo para test unilateral

El valor de  $t$  en este caso ya lo habíamos determinado en el test anterior y es de  $-6,00611$ . Observen que en esta situación no trabajamos con el valor absoluto, porque en realidad el signo es lo que importa. Busquemos el valor crítico de  $t$  en tabla del apéndice C para  $\alpha = 0,05$  y  $n - k = 42$ . Tampoco aparece, por lo que vamos a interpolar linealmente de nuevo entre los valores de  $t$  correspondientes a 40 y 50 grados de libertad. Se obtiene un valor de  $t_c = 1,6824$ . Nuestro valor para  $t$  es mucho menor que el valor crítico, por lo que estamos en zona de rechazo de que  $\beta_2$  sea mayor o igual que cero. Mas en el ejercicio anterior ya descartamos la posibilidad de que  $\beta_2$  fuese igual a 0 con un 95% de confianza. Así, podemos concluir  $\beta_2$  es estrictamente negativo, por lo que afirmamos que el resultado empírico no contradice lo planteado por la teoría económica: que a un incremento en los precios relativos de las importaciones corresponderá una disminución de la demanda por importaciones, y viceversa.

## 6.4 Test de Hipótesis Compuestas (o Test-t Generalizado)

Ya habíamos adelantado que estos tests-t nos posibilitan indagar sobre hipótesis compuestas, que permiten testear sobre una condición que combine linealmente varios regresores a la vez. Podemos plantear en este caso la siguiente hipótesis bilateral:

$$H_0 : \mathbf{t}'\boldsymbol{\beta} = M$$

$$H_0 : \mathbf{t}'\boldsymbol{\beta} \neq M$$

donde  $\mathbf{t}$  es un vector columna de constantes y  $M$  un determinado valor.

En este caso  $\mathbf{t}'\hat{\boldsymbol{\beta}} = \sum_{j=1}^k t_j \hat{\beta}_j = t_1 \hat{\beta}_1 + t_2 \hat{\beta}_2 + \dots + t_k \hat{\beta}_k$ , donde los  $t_j$  pueden tomar cualquier valor.

Para contrastar este tipo de hipótesis se usa la *prueba t generalizada*. Para poder realizar esta prueba necesitamos conocer cuál es la distribución de  $\mathbf{t}'\hat{\boldsymbol{\beta}}$ . Sabemos que el vector  $\hat{\boldsymbol{\beta}}$  distribuye normal y que  $\mathbf{t}$  es un vector de constantes; entonces  $\mathbf{t}'\hat{\boldsymbol{\beta}}$  distribuirá igualmente normal y podemos determinar los parámetros característicos de su distribución, media y varianza, a partir de:

$$E(\mathbf{t}'\hat{\boldsymbol{\beta}}) = \mathbf{t}' \cdot E(\hat{\boldsymbol{\beta}}) = \mathbf{t}'\boldsymbol{\beta}$$

$$\text{var}(\mathbf{t}'\hat{\boldsymbol{\beta}}) = \mathbf{t}' \cdot \text{var}(\hat{\boldsymbol{\beta}}) \cdot \mathbf{t} = \mathbf{t}'\sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{t} = \sigma^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{t}$$

Esta variable normal la podemos estandarizar como ya sabemos:

$$\frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\sigma^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{t}}} \sim N(0,1) \quad (6.6)$$

De nuevo por no conocer el verdadero valor de  $\sigma^2$  debemos construir una variable —al igual que hicimos en el caso de las pruebas de hipótesis sencillas—, con distribución conocida y donde podamos utilizar el estimado de la varianza común de los  $e_i$ ,  $\hat{\sigma}^2$ . Esta variable es:

$$\frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{t}}} \sim t_{(n-k)} \quad (6.7)$$

A partir de esa variable se puede realizar la inferencia estadística. Al inicio planteamos una hipótesis bilateral, pero también se pueden contrastar hipótesis unilaterales, y se procede de forma análoga a como explicamos para el caso de los test de hipótesis sencillas.

Una situación de la práctica donde este tipo de hipótesis compuesta es relevante se refiere a la estimación de una función de producción tipo Cobb-Douglas. Esta función responde al siguiente planteamiento:

$$Q = A \cdot K^{a_K} \cdot L^{a_L} \cdot H^{a_H}$$

donde  $Q$  es la variable que representa la producción,  $K$  el capital físico,  $L$  la fuerza de trabajo y  $H$  el capital humano.  $A$  es la eficiencia técnica, que estamos suponiendo constante en este planteamiento. Los parámetros de esta función  $a_K$ ,  $a_L$  y  $a_H$  son las elasticidades respectivas.

Esta función teórica no es lineal, sin embargo, podemos transformarla fácilmente en una función lineal con tan sólo aplicar logaritmos y utilizar convenientemente sus propiedades:

$$\begin{aligned} \ln Q &= \ln(A \cdot K^{a_K} \cdot L^{a_L} \cdot H^{a_H}) \\ &= \ln A + a_K \ln K + a_L \ln L + a_H \ln H \end{aligned}$$

El resultado de esta transformación es una función lineal en los regresores, que es precisamente lo que necesitamos para poder aplicar nuestro procedimiento de estimación, para lo que debemos trabajar con las variables en logaritmos.

En este caso nos interesará saber qué tipo de rendimientos a escala tenemos:

- ✓ si constantes, como para las industrias maduras, en cuyo caso la suma de los regresores deberá ser igual a la unidad,  $a_K + a_L + a_H = 1$ ;
- ✓ si crecientes, como para las industrias jóvenes, con la suma mayor que 1,  $a_K + a_L + a_H > 1$
- ✓ o decrecientes, como para las industrias envejecidas, con la suma menor que uno,  $a_K + a_L + a_H < 1$

Para este tipo de verificación el vector  $\mathbf{t}$  será  $\mathbf{t}' = [1 \ 1 \ 1]$ , para representar la parte izquierda de la hipótesis que deseamos investigar:

$$\mathbf{t}'\boldsymbol{\beta} = [1 \ 1 \ 1] \begin{bmatrix} a_K \\ a_L \\ a_H \end{bmatrix} = a_K + a_L + a_H$$

Si queremos verificar si es posible asumir rendimientos constantes a escala, las pruebas de hipótesis correspondientes son:

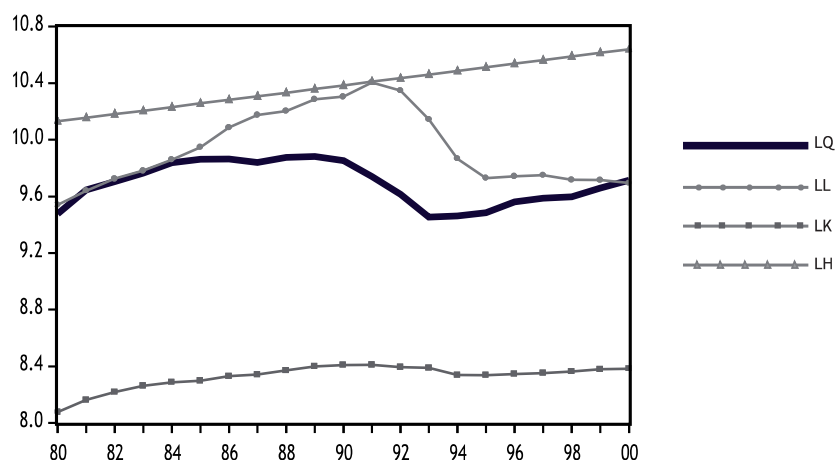
$$H_0 : \mathbf{t}'\boldsymbol{\beta} = 1$$

$$H_1 : \mathbf{t}'\boldsymbol{\beta} \neq 1$$

Para ilustrar cómo realizar esta prueba de hipótesis de tipo compuesto para los regresores en el caso de una función de producción, vamos a utilizar los datos correspondientes a la producción y algunos de sus factores en Cuba correspondientes al período 1980-2000, tomados del trabajo empírico realizado por Fernández de Bulnes (2001). Esta información se refiere al PIB ( $Q$ ) y a la existencia de capital físico ( $K$ ),<sup>2</sup> medidos ambos a precios de 1981; a la fuerza de trabajo ( $L$ ), a través del nivel de ocupados en el sector estatal civil de la economía; y al capital humano ( $H$ ), como la cantidad de años de escolaridad de la fuerza de trabajo.

Con estos insumos calculamos las variables de nuestra función de producción en términos logarítmicos y las representamos anteponiendo una  $L$  a la notación de cada variable, para después estimar las elasticidades correspondientes a cada factor y verificar si podemos asumir rendimientos constantes a escala en las condiciones de la economía cubana entre esos años. El recorrido de las variables de esta función de producción se muestra en la figura 6.3.

**Figura 6.3: Logaritmos de las variables para una función de producción macroeconómica en Cuba, 1980-2000**



*Fuente: Elaboración propia a partir de Fernández de Bulnes (2001)*

Como resultado intermedio de esta estimación, vamos a retener  $(X'X)^{-1}$ , que es una matriz importante para el cálculo de los estimadores de  $\beta$  y también para conseguir el valor del estadístico  $t$  correspondiente a la prueba generalizada, según indica (6.7). Obtenemos los siguientes resultados:

- ✓ La función de producción estimada:

<sup>2</sup> Fue estimado por Fernández de Bulnes (2001) a partir de los datos anuales de inversión aportados por la ONE, las tasas de depreciación fijadas por el Ministerio de Finanzas y Precios (MFP) y aplicando el método del inventario permanente.

$$LQ = C + a_K LK + a_L LL + a_H LH$$

$$LQ = 2,187148 - 0,500365 LK + 3,714342 LL - 1,777154 LH$$

✓ La matriz  $(\mathbf{X}'\mathbf{X})^{-1}$ :

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 740,51442 & 34,29401 & -201,96471 & 57,82832 \\ 34,29401 & 4,48889 & -18,48601 & 7,22578 \\ -201,96471 & -18,48601 & 90,15417 & -35,15491 \\ 57,82832 & 7,22578 & -35,15491 & 15,70712 \end{bmatrix}$$

✓ La varianza de la regresión:  $\sigma^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{0,14128652}{17} = 0,008311$

✓ El valor del estadístico  $t$  para la prueba generalizada:

$$t = \frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{t}}} = \frac{1,436822675 - 1}{0,381585421} = 1,144757$$

Para el cálculo del denominador del estadístico  $t$  se utilizan las filas y columnas de la matriz  $(\mathbf{X}'\mathbf{X})^{-1}$  correspondientes a los regresores involucrados en la restricción  $\mathbf{t}'\boldsymbol{\beta} = M$ , en este caso particular se suprimen las primeras fila y columna, dado que no se incluye la constante en la restricción  $a_K + a_L + a_H = 1$ .

A este valor de  $t$  corresponde una probabilidad de 0,268172, superior al nivel de referencia de 0,025 (nos mantenemos trabajando con  $\alpha = 5\%$ ). Igualmente, podemos verificar que ese valor absoluto de  $t$  es menor que su valor crítico (tenemos 21 observaciones y estimamos cuatro regresores, incluyendo la constante, por lo que los grados de libertad se reducen a 17, de modo que  $t_c = t_{(n-k)}^{\alpha/2} = t_{17}^{0,025} = 2,110$ ). Por ambas vías concluimos que la información de la muestra no contradice la hipótesis de rendimientos constantes a escala; por supuesto, siempre que se trate de un modelo correctamente especificado a través de un MLNR, es decir, que cumpla los supuestos asociados a un MLNR, lo que todavía está por verificar.

También podemos enfrentar en la práctica situaciones en que precisemos realizar test de hipótesis compuestas pero a partir de relaciones no lineales entre los regresores. Un ejemplo de esta situación lo encontramos en los modelos dinámicos, donde aparecen *rezagos*, *retardos* o *lags* de las variables como variables explicativas en el modelo de regresión.

Un caso particular de los modelos dinámicos lo constituye el modelo de corrección de error (MCE). El planteamiento correspondiente a este tipo de mecanismo, en el caso de incluir una sola variable explicativa y de que intervengan sólo un



retardo de  $Y$  y de  $X$ , es el siguiente:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \quad (6.8)$$

En este planteamiento, siempre resultan de interés los multiplicadores de largo plazo, que se obtienen a partir de suponer que la economía ha arribado a un equilibrio, a un estado estacionario, por lo que consideramos que las variables se mantienen estáticas, de modo que  $Y_t = Y_{t-1}$  y  $X_t = X_{t-1}$ . Al sustituir en (6.8) y despejar  $Y_t$ , obtenemos:

$$Y_t = \frac{\alpha_0}{(1-\alpha_1)} + \frac{\beta_0 + \beta_1}{(1-\alpha_1)} X_t + u_t \quad (6.9)$$

En la ecuación (6.9), la constante y el coeficiente correspondiente a  $X_t$  son los coeficientes correspondientes al equilibrio de largo plazo

$$\theta_0 = \frac{\alpha_0}{(1-\alpha_1)} \text{ y } \theta_1 = \frac{\beta_0 + \beta_1}{(1-\alpha_1)}$$

Noten que estos multiplicadores se obtienen a partir de relaciones no lineales entre los coeficientes de la regresión dinámica, por lo que para someterlos a pruebas de hipótesis específicas ya no podremos trabajar más con test tipo  $t$ . En estos casos se recurre al Teorema de Slutsky, que permite construir la varianza de ese tipo de relación no lineal. Igualmente se profundizará más en estos tópicos en la Parte Cuatro de este texto, dedicada especialmente al tratamiento de series de tiempo.

## 6.5 Evaluando la Bondad de Ajuste

Otro indicador que se utiliza con frecuencia con el propósito de evaluar los modelos de regresión estimados es el **coeficiente de determinación**, que se denota como  $R^2$ . Este coeficiente también se nombra como **bondad de ajuste**, y es un índice que permite expresar la capacidad que tiene un determinado modelo de regresión de explicar la variación de la variable dependiente a partir de las variaciones de las explicativas. Este índice se construye a partir de descomponer la suma de los cuadrados de la diferencia entre el valor observado de la variable dependiente y su promedio o esperanza matemática en la muestra disponible.

El mencionado desglose se basa directamente en el modelo de regresión, que postula que el valor observado de  $Y$  se descompone en una parte sistemática, explicada a través de la regresión, y una parte no sistemática, que es el término de error, tal y como se expresa en:

$$Y_i = \hat{Y}_i + e_i \quad (6.10)$$

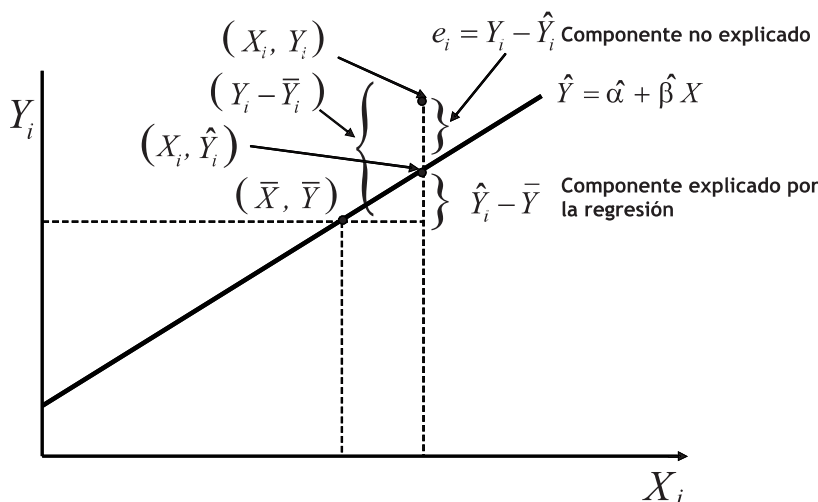
A la expresión (6.10) podemos deducirle el valor del promedio de  $Y$ ,  $\bar{Y}$ , y re-

agrupamos convenientemente para obtener:

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + e_i = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \quad (6.11)$$

que es la descomposición que representamos gráficamente para el caso de una regresión bivariada en la figura 6.4. Noten que el punto  $(\bar{X}, \bar{Y})$  siempre pertenecerá a la línea de regresión estimada por MCO, lo que constituye una característica de este método de estimación.

**Figura 6.4: El coeficiente de determinación**



Fuente: Elaboración propia

Elevamos al cuadrado la expresión (6.11) y sumamos a través de todas las observaciones disponibles ( $i=1, \dots, n$ ) para arribar al siguiente resultado:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n e_i^2 \quad (6.12)$$

El primer componente de la parte derecha de la expresión (6.12) es la suma de las diferencias entre el valor explicado por la regresión y el valor promedio al cuadrado, por lo que lo denominamos *suma de cuadrados explicada (SCE)*.<sup>3</sup> El segundo miembro, la suma de residuos de la regresión al cuadrado, por lo que se le nombra *suma de residuos al cuadrado (SRC)*;<sup>4</sup> es justamente la parte de la variación de la variable dependiente que no logra ser explicada por la regresión. La sumatoria de la izquierda de la expresión (6.12) es la *suma total de cuadrados*

<sup>3</sup> Las siglas en inglés de este componente de la suma de cuadrados son SSE. Sin embargo, también lo encontramos denotado como SCR (SSR en inglés), por suma de cuadrados de la regresión.

<sup>4</sup> Las siglas en inglés de este componente de la suma de cuadrados son SSR. Sin embargo, también se denota como SCE (SSE en inglés), por suma de cuadrados del error, que son justo las siglas de “su contrario”. Obviamente, hay que estar muy concentrados cuando estudiemos estos tópicos en diferentes bibliografías, por esta cuestión de las mismas notaciones para elementos opuestos.

( $STC$ ), pues resume las dos sumas de cuadrados anteriores, la explicada y la correspondiente a los residuos. Así, (6.12) puede ser también escrita como:

$$STC = SCE + SRC \quad (6.13)$$

Al dividir la expresión (6.13) entre la  $STC$ , se obtiene la relación:

$$1 = \frac{SCE}{STC} + \frac{SRC}{STC} = R^2 + \frac{SRC}{STC} \quad (6.14)$$

El índice  $SCE/STC$  es precisamente el coeficiente de bondad de ajuste o coeficiente de determinación, ya que al relacionar la suma de cuadrados explicada con la suma total de cuadrados, expresa la proporción de la variabilidad total de  $Y$  que es explicada por la regresión. Este coeficiente lo denotamos con  $R^2$  y es un mero número, que se ubica en el intervalo  $[0,1]$ . Mientras más cercano a la unidad, mejor ajuste se consigue con el modelo de regresión; mientras más cercano a cero, peor será el ajuste.

Como ya dijimos al comenzar a tratar este tópico, este coeficiente es muy popular. Sin embargo, es bueno que noten que la base de este índice es pura álgebra, que no tiene ninguna función de distribución asociada y, por lo tanto, no es posible hacer inferencia estadística a partir del mismo: es decir, sólo nos describe cuán bueno es el ajuste en el caso de nuestro modelo de regresión y para la muestra disponible. La bondad de ajuste no es precisamente un sinónimo de la calidad del modelo de regresión; de hecho, muchas regresiones en las que se obtiene un altísimo  $R^2$  finalmente han sido regresiones sin sentido o espurias.<sup>5</sup> Por lo tanto, no es bueno dejarse entusiasmar por un alto  $R^2$  sin haber sometido antes el modelo econométrico a las pruebas de diagnóstico o incorrecta especificación.

Otro asunto relativo al  $R^2$  es que está directamente relacionado con el número de variables explicativas en la regresión: mientras más variables incorporemos a la explicación de  $Y$ , mayor será el coeficiente de determinación.

Para entender esta relación es conveniente partir de la expresión (6.14). Al despejar en ella  $R^2$ , obtenemos:

$$R^2 = 1 - (SRC/STC) \quad (6.15)$$

¿Cómo reaccionan  $SRC$  y  $STC$  ante un aumento de las variables explicativas? A medida que incorporemos más variables al modelo, su residuo deberá disminuir,<sup>6</sup> por lo que también lo hará la  $SRC$ ;  $STC$  no se afecta por un incremento de  $k$ . Entonces, la fracción  $SRC/STC$  disminuirá ante un aumento de la cantidad de  $X_j$ , por lo que  $R^2$  tenderá a aumentar.

<sup>5</sup> En la Parte Cuatro de este texto se abordará en detalle esta cuestión de las regresiones espurias.

<sup>6</sup> En el caso de que las variables incorporadas adicionalmente al modelo no contribuyan absolutamente a explicar las variaciones de  $Y$ , sencillamente se mantendrán los residuos y la  $SRC$ .

Así, para poder comparar los  $R^2$  de dos modelos, éstos tendrían que tener el mismo número de variables. Por esta razón se ha desarrollado otro coeficiente de determinación, nombrado  $R^2$  ajustado o corregido y denotado con el símbolo  $\bar{R}^2$ . El “ajuste” se realiza a los fines de la comparación de modelos con diferente número de variables explicativas y para ello se utilizan los grados de libertad correspondientes a las sumas de cuadrados STC y SRC: en el primer caso se divide entre  $n-1$  (para obtener esta suma recurrimos a la media muestral estimada,  $\bar{Y}$ ) y en el segundo, entre  $n-k$  (para obtener esta suma utilizamos los residuos de la regresión muestral, estimados a partir de los  $k$  componentes del vector  $\hat{\beta}$ ). La expresión (6.15) se modifica tal y como mostramos seguidamente:

$$\bar{R}^2 = 1 - \frac{SRC/n-k}{STC/n-1} = 1 - \frac{\hat{\sigma}^2}{s_Y^2} = 1 - \frac{SRC}{STC} \frac{n-1}{n-k} \quad (6.16)$$

Noten que la división del numerador de la fracción del primer término a la derecha de (6.16) no es otra cosa que la varianza de los residuos en la muestra y que el cociente del denominador corresponde a la varianza muestral de  $Y$ .

La relación entre  $R^2$  y  $\bar{R}^2$  la obtenemos combinando las fórmulas (6.15) y (6.16):

$$\bar{R}^2 = 1 - \frac{SRC}{STC} \frac{n-1}{n-k} = 1 - (1 - R^2) \frac{n-1}{n-k} \quad (6.17)$$

donde se evidencia que el incremento de  $k$ , aun cuando tiende a aumentar la fracción  $(n-1)/(n-k)$ , al provocar también un incremento del  $R^2$ , disminuye el factor  $(1 - R^2)$ , lo que consigue más que compensar el efecto inicialmente comentado. Es posible demostrar que  $\bar{R}^2 < R^2$  para  $k > 1$ .<sup>7</sup>

Según Gujarati (2004, pp. 210-211), algunos autores recomiendan trabajar con el  $\bar{R}^2$  en lugar del  $R^2$ ; sin embargo, otros argumentan que el  $\bar{R}^2$  es sólo una de las posibles variantes de modificación del  $R^2$ . No obstante, en la mayoría de los paquetes econométricos podrán encontrar el cómputo de  $\bar{R}^2$  junto a  $R^2$ , así como también otras medidas para la bondad de ajuste del modelo de regresión, entre ellas los criterios de información de Akaike (AIC) y de Schwartz (SIC), utilizados para escoger entre dos modelos alternativos.

Los coeficientes de determinación (ajustado y no) son criterios que miden la bondad del ajuste *dentro de la muestra*: expresan la exactitud del pronóstico de

<sup>7</sup> En el caso de  $k=1$  ambos coeficientes son iguales. Desarrollando la expresión (6.17) y agrupando términos convenientemente llegamos a  $\bar{R}^2 = \frac{n-1}{n-k}(R^2 - 1) + R^2$ , donde podemos comprobar que  $\bar{R}^2 < R^2$  si  $k > 1$ .

$Y$  a partir de los valores de  $X_j$  disponibles en la muestra y a través del modelo de regresión (de los  $\hat{\beta}_j$ ). No tenemos certeza de que un modelo con altos coeficientes de determinación logre buenos pronósticos *fuera de la muestra*. Los criterios de Akaike (AIC) y Schwartz (SIC) son válidos para el desempeño del pronóstico tanto dentro como fuera de la muestra.

El AIC se define como:

$$AIC = \ln\left(\frac{\sum e_i^2}{n}\right) + \frac{2k}{n} \quad (6.18)$$

Noten que a mayor tamaño de muestra, menor será el AIC, por lo que obviamente es del mayor interés que el AIC sea pequeño, y cuanto más, mejor. Observen también que a mayor  $k$ , el segundo término de la suma tiende a compensar la contracción que se origina por la disminución de la SRC en el primer término, de modo que este es un criterio que ya viene ajustado por la influencia de un incremento en el número de variables explicativas y que es, por lo tanto, un indicador que permite comparar modelos de diferente “dimensión”.

El SIC se define como:

$$SIC = \frac{k}{n} \ln(n) + \ln\left(\frac{SRC}{n}\right) \quad (6.19)$$

Noten que también se impone una “penalidad”<sup>8</sup> a la incorporación de un mayor número de variables explicativas, a través del primer término de la suma, y que este criterio es mejor mientras más pequeño.

Estos criterios, AIC y SIC, entre otros similares, son ampliamente utilizados para la comparación de modelos, sobre todo en el caso de determinar el número óptimo de rezagos a incorporar en los modelos dinámicos. La utilización de estos criterios con este propósito la trataremos más ampliamente en la Parte Cuatro de este texto.

## 6.6 Aplicando el Análisis de Varianza a la Inferencia

### 6.6.1 Test de significación global

Otra aplicación interesante de la descomposición de la STC que desarrollamos en (6.13) es el análisis de varianza<sup>9</sup> y merece una atención particular, ya que permite una forma adicional de examinar el problema de la inferencia estadística.

<sup>8</sup> En el sentido de que se compensa la disminución del criterio que origina un incremento de  $k$  en el segundo término de la suma, incorporando el primer término que se mueve en sentido contrario (por eso penaliza).

<sup>9</sup> ANOVA, por sus siglas en inglés.

Recién repasamos que a cada componente de la suma de cuadrados de (6.13) corresponden ciertos grados de libertad (gl), que no son más que el número de observaciones independientes sobre las que descansa su cálculo. En nuestro caso tenemos que:

- ✓ La STC tiene  $n-1$  gl, pues perdemos un gl al calcular la media muestral  $\bar{Y}$ ;
- ✓ La SRC tiene  $n-k$  gl, pues perdemos  $k$  gl al calcular los  $\hat{\beta}_j$  (con  $j=1,\dots,k$ ) necesarios para estimar los residuos;
- ✓ La SCE tiene  $k-1$ , pues se trata de la variación de  $Y$  explicada por las  $k-1$  variables explicativas  $X_2, X_3, \dots, X_k$  y es una función de ellas (excluimos  $X_1 = 1$ ).

A partir de las sumas de cuadrados y sus gl construimos una tabla para el ANOVA, tal y como mostramos en la tabla 6.1.

Bajo los supuestos de distribución normal de los residuos y de que todos los  $\beta_j$  sean cero, se tiene que la razón entre las sumas medias de cuadrados correspondientes a SCE y SRC sigue una distribución  $F$  con  $k-1$  y  $n-k$  grados de libertad en su numerador y denominador, respectivamente. Esta razón es de la mayor importancia, pues nos sirve para realizar el *test de significación global* de la regresión.

Tabla 6.1: Análisis de varianza con  $k$  variables

| Origen de la variación | Suma de cuadrados | gl    | Suma media de cuadrados * |
|------------------------|-------------------|-------|---------------------------|
| Debido a la regresión  | SCE               | $k-1$ | $\frac{SCE}{k-1}$         |
| Debido al residuo      | SRC               | $n-k$ | $\frac{SRC}{n-k}$         |
| Total                  | STC               | $n-1$ | $\frac{STC}{n-1}$         |

\* La obtenemos dividiendo la suma de cuadrados entre sus gl.

Nuevamente insistimos en que los resultados de este test —al igual que del resto de los test de evaluación del modelo de regresión—, deben ser atendidos una vez que tengamos un modelo de regresión correctamente especificado. Por ejemplo, no tiene sentido atender a los resultados de este test  $F$ , si no estamos seguros de que la evidencia empírica no contradice el supuesto de normalidad en los residuos.

El test de significación global de la regresión se basa en las siguientes hipótesis:

$$H_0 : \beta_j = 0 \text{ para todo } j = 2, 3, \dots, k$$

$$H_1 : \beta_j \neq 0 \text{ para algún } j = 2, 3, \dots, k$$

Note que de no poder rechazar  $H_0$ , nuestro modelo de regresión “se viene abajo” (nos quedamos sin modelo), pues ninguna de las variables que hemos

incluido en él nos sirven para explicar las variaciones de  $Y$ .

El estadístico de prueba correspondiente es:

$$F = \frac{SCE/k-1}{SRC/n-k} \sim F_{(k-1, n-k)}$$

Y la regla de rechazo, como siempre, se basa en comparar  $F$  con su valor crítico en tablas (dados los gl y el nivel de significación): si  $F > F_c = F_{(k-1, n-k)}^\alpha$ , entonces rechazamos  $H_0$ . También se puede calcular la probabilidad asociada a  $F$  y comparar con el nivel de significación, si  $p < \alpha$ , igualmente rechazamos  $H_0$ . Por supuesto, nos contentaremos si conseguimos un alto valor para  $F$  y una  $p$  bien pequeña con nuestro modelo. La mayoría de los paquetes econométricos computan la  $F$  global de la regresión estimada y también la  $p$ .

Vamos a replicar el ANOVA y el test de significación para nuestra función de importaciones para la economía cubana, presentada en el capítulo 5.

**Tabla 6.2: Análisis de varianza para la función de importaciones cubanas**

| Origen de la variación                  | Suma de cuadrados | gl | Suma media de cuadrados * |
|-----------------------------------------|-------------------|----|---------------------------|
| Debido a la regresión                   | 4999,19537        | 2  | 2499,59768                |
| Debido al residuo                       | 3764,30383        | 42 | 89,62628                  |
| Total                                   | 8763,4992         | 44 | 199,17044                 |
| Estadístico F para significación global |                   |    | 27,88912                  |
| Probabilidad asociada                   |                   |    | 0,00000002                |
| Valor crítico de F para                 |                   |    |                           |
| 5% de significación                     |                   |    | 3,21994                   |
| 1% de significación                     |                   |    | 5,14914                   |

Como pueden apreciar en la tabla 6.2, obtenemos un alto valor para  $F$  y una pequeña probabilidad asociada, que permiten rechazar la hipótesis nula de que no tenemos modelo de regresión aun al 1% de significación. Los valores críticos de  $F$  que corresponderían en este caso son muy inferiores al valor de  $F$ , lo que corrobora el rechazo de  $H_0$ . Los valores críticos que mostramos en la tabla 6.2 los hemos conseguido a través de la función inversa de  $F$  en Excel. Si hubiéramos buscado en la tabla del apéndice C, hubiéramos escogido el valor correspondiente a 2 y 40 grados de libertad para el numerador y el denominador del estadístico, respectivamente, que es el que más se acerca a nuestro ejemplo. Las lecturas en la tabla serían de 3,23 y 5,18 para 5 y 1% de significación, respectivamente.

### 6.6.2 Test de hipótesis simultáneas

Así, este test F nos permitió comprobar una hipótesis compuesta, donde los regresores debían cumplir de manera simultánea determinadas restricciones: en

este caso particular, todos debían ser cero. Con frecuencia se pueden presentar en la práctica situaciones en las que interese comprobar que varias restricciones se cumplan de forma conjunta. Por ejemplo, en una función de producción podríamos querer probar que las elasticidades capital y trabajo del producto suman 1, y que la elasticidad capital humano es igual a 1. En ese caso tendríamos la siguiente hipótesis compuesta:

$$H_0 : a_K + a_L = 1 \text{ y } a_H = 1$$

$$H_1 : a_K + a_L \neq 1 \text{ o } a_H \neq 1$$

Esta hipótesis compuesta también la podemos escribir apoyándonos en el álgebra matricial como:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

donde  $\mathbf{R}$  es una matriz de dimensión  $m \times k$ , con  $m$  representando el número de restricciones que se imponen y  $k$  la cantidad de regresores involucrados en las restricciones. Esta matriz contiene los coeficientes correspondientes a los regresores en las restricciones.  $\mathbf{r}$  es un vector columna de dimensión  $m \times 1$ , que contiene las partes derechas de dichas restricciones. En el caso particular de la restricción que queremos verificar para la función de producción, la matriz  $\mathbf{R}$  y los vectores  $\boldsymbol{\beta}$  y  $\mathbf{r}$  adoptan la siguiente forma:

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} a_K \\ a_L \\ a_H \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Existen dos alternativas para realizar este tipo de prueba. La primera es aplicando el álgebra de matrices. Por esta vía se llega a desarrollar una variable que, bajo la hipótesis nula, distribuye  $F$  con  $m$  grados de libertad en el numerador y  $n-k$  en el denominador:

$$F = \frac{n-k}{m} \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})' (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}') (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})}{\mathbf{e}'\mathbf{e}} \sim F_{(m, n-k)}$$

A partir de computar  $F$ , procedemos igual que habitualmente: calculamos la probabilidad  $p$  asociada y la comparamos con el  $\alpha$  previsto, o comparamos  $F$  directamente con el valor crítico correspondiente a los grados de libertad y la significación. Las reglas de rechazo ya las conocemos bien.

La otra alternativa de realizar esta prueba es utilizando las sumas de los residuos al cuadrado de los modelos. Estas sumas se calculan para el caso del modelo restringido, según la hipótesis que se desea verificar, y para el modelo sin restricción alguna. La idea es comparar ambas sumas de residuos y determinar si son estadísticamente diferentes o no. Si no lo fuesen, entonces no rechazaríamos la pertinencia del modelo restringido.



Por esta vía se llega a plantear una variable que distribuye  $F$  de la siguiente manera:

$$F = \frac{n-k}{m} \frac{(SRRC - SRC)}{SRC} \sim F_{(m, n-k)} \quad (6.20)$$

donde  $SRRC$  es la *suma de residuos restringidos al cuadrado*, y no es más que la  $SRC$  correspondiente al modelo restringido, y  $SRC$  es la correspondiente al modelo sin restricción. Para la  $SRRC$  también se utilizan otras notaciones, por ejemplo  $SSR_R$ , así como para la  $SRC$ , que se denota como  $SSR_U$ .<sup>10</sup>

Con este resultado es posible verificar múltiples hipótesis, siguiendo el siguiente procedimiento:

- ✓ Primero estimamos el modelo sin restricciones y recuperamos la  $SRC$ . Digamos que este modelo es  $\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$  y que

$$SRC = \mathbf{e}'\mathbf{e} = \sum_{i=1}^n e_i^2$$

- ✓ Después estimamos el modelo sujeto a la hipótesis nula, es decir, teniendo en cuenta las restricciones, y recuperamos la  $SRRC$ . Si nuestra hipótesis nula es  $H_0 : \boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ , estimamos entonces  $\mathbf{Y} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}$  y recuperamos

$$SRRC = \tilde{\mathbf{e}}'\tilde{\mathbf{e}} = \sum_{i=1}^n \tilde{e}_i^2$$

- ✓ Finalmente, calculamos el valor de  $F$  según (6.20) y lo comparamos con el valor tabulado u obtenido a través de la función inversa de la distribución  $F$  correspondiente a nuestros grados de libertad, para decidir si rechazamos o no  $H_0$ .

Vamos a aplicar este procedimiento en el caso de nuestra función de producción para verificar la hipótesis que adelantamos al inicio de esta explicación. Ya estimamos el modelo no restringido (es decir, sin condicionamiento alguno a sus regresores) y disponemos de la  $SRC$ , que necesitamos antes para realizar el test acerca de la hipótesis de rendimientos constantes a escala, con los siguientes resultados:

$$LQ = 2,187148 - 0,500365 LK + 3,714342 LL - 1,777154 LH$$

$$SRC = 0,141287$$

Ahora debemos estimar el modelo restringido, según

<sup>10</sup> Por las siglas en inglés de suma de residuos al cuadrado,  $SSR$ , con la “ $R$ ” de *restricted* y la “ $U$ ” de *unrestricted*.

$$H_0 : \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_K \\ a_L \\ a_H \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Esta restricción significa que  $a_K + a_L = 1$  y  $a_H = 1$ , de modo que tenemos que plantear nuestro modelo restringido siguiendo  $H_0$ , como:

$$\begin{aligned} \ln Q &= \ln A + a_K \ln K + a_L \ln L + a_H \ln H \\ \ln Q &= \ln A + a_K \ln K + (1 - a_K) \ln L + 1 \cdot \ln H \\ \ln Q &= \ln A + a_K \ln K + \ln L - a_K \ln L + \ln H \\ \ln Q - \ln L - \ln H &= \ln A + a_K (\ln K - \ln L) \end{aligned}$$

Así, tendríamos que estimar el modelo en función de dos variables que hemos llamado  $D$  y  $E$ , con  $D = \ln Q - \ln L - \ln H$  y  $E = \ln K - \ln L$ , y obtendríamos dos regresores: la constante y  $a_K$ . El resultado de esta estimación es  $D = -9,64233 + 0,38650 E$  y la  $SRRC = 1,558799$ . Entonces, el valor de  $F$  es:

$$F = \frac{n-k}{m} \frac{(SRRC - SRC)}{SRC} = \frac{17}{2} \frac{(1,558799 - 0,141287)}{0,141287} = 85,279583 \sim F_{(2,17)}$$

El valor crítico que corresponde a un  $\alpha = 5\%$  es  $F_{(m,n-k)}^{\alpha} = F_{(2,17)}^{0,05} = 3,59$ . Nuestra  $F$  es francamente mayor, por lo que se ubica en la zona de rechazo de  $H_0$ , y concluimos que la evidencia empírica no sostiene el conjunto de restricciones impuesto.

## 6.7 Consiguiendo una Predicción

Finalmente, no podemos dejar de tratar el tema de la predicción en este capítulo. En muchas ocasiones el propósito del trabajo empírico es conseguir modelos que apoyen el trabajo de pronóstico. Por ejemplo, las autoridades del Ministerio de Economía y Planificación seguramente estarán interesadas en conocer qué efecto puede tener un determinado incremento porcentual del PIB en la demanda de importaciones del país. Para saberlo sería suficiente con sustituir el crecimiento previsto del PIB en la función de importaciones y calcular cuál será la variación que corresponda a las importaciones, si los demás factores que influyen se mantienen constantes.

La exactitud de una predicción dependerá de varios factores. Supongamos que el modelo real vigente en la población se expresa por la regresión  $Y = X\beta + \varepsilon$ . Supongamos que estamos trabajando con una serie de tiempo con  $t = 1, \dots, T$  y que queremos conocer cuál será el futuro valor de  $Y$ ,  $q$  períodos después del extremo

superior del intervalo temporal que abarca la muestra disponible, es decir,  $Y_{T+q}$ . Para ello necesitamos conocer el valor que tomarán nuestras variables explicativas en el momento  $T+q$ , representado por el vector  $\mathbf{X}_{T+q}$ , así como los valores de los parámetros  $\beta$  de la regresión. En la medida en que tengamos un conocimiento certero del comportamiento futuro de las variables que influyen en  $Y$  y que consigamos una buena estimación para los parámetros de la regresión poblacional, obtendremos un pronóstico de calidad para  $Y_{T+q}$ .

Al valor pronosticado o predicción vamos a denotarlo con  $\hat{Y}_{T+q}$ , y lo calculamos como  $\hat{Y}_{T+q} = \mathbf{X}_{T+q}\hat{\beta}$ . Nos interesa la exactitud de esta previsión, por lo que examinamos el *error de pronóstico o de predicción*, en pos de sus fuentes. Este error de pronóstico, denotado con  $e_{T+q}^P$ , es la diferencia entre el valor que debería tener la variable si conociéramos la verdadera regresión poblacional y la predicción,

$$e_{T+q}^P = Y_{T+q} - \hat{Y}_{T+q} \quad (6.21)$$

A fin de investigar los orígenes del error de pronóstico, lo descomponemos para acercarnos a nuestro propósito:

$$e_{T+q}^P = Y_{T+q} - \hat{Y}_{T+q} = \mathbf{X}_{T+q}\beta + \varepsilon_{T+q} - \mathbf{X}_{T+q}\hat{\beta} = \mathbf{X}_{T+q}(\beta - \hat{\beta}) + \varepsilon_{T+q} \quad (6.22)$$

El análisis de la expresión (6.22) nos permite determinar que son tres las fuentes del error de pronóstico. La primera viene dada por el error que se produce al tener que estimar a partir de la información contenida en la muestra disponible; es el error de estimación, dado por  $\beta - \hat{\beta}$ . La segunda está asociada a la perturbación aleatoria o choque estocástico correspondiente al momento  $T+q$ , es decir, viene dada por la parte no sistemática de  $Y$ . Y la tercera vendrá dada por la propia precisión del valor de  $\mathbf{X}_{T+q}$ . Si las variables explicativas del modelo de regresión son variables que responden a decisiones de política económica, que pueden ser controladas, no habrá dificultades por esta fuente; pero si no es este el caso y se trata de variables que a su vez debemos pronosticar, los errores de ese pronóstico influirán en  $e_{T+q}^P$ .

Justamente por la incertidumbre que envuelve al valor del pronóstico, no debemos conformarnos con una predicción puntual del tipo  $\hat{Y}_{T+q}$ , sino que debemos construir un intervalo de confianza alrededor de ésta para ofrecer un pronóstico menos rígido y que, por lo tanto, tenga mayor oportunidad de que se cumpla. No debemos perder nunca de vista el carácter probabilístico de nuestros resultados.

Para la construcción del intervalo de confianza necesitamos conocer la media, la varianza y la distribución del error de predicción. Comencemos por determinar cuál es su valor esperado:

$$\begin{aligned}
 E(e_{T+q}^P) &= E(Y_{T+q} - \hat{Y}_{T+q}) \\
 &= E[\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \varepsilon_{T+q}] \\
 &= E[\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] + E(\varepsilon_{T+q}) \\
 &= \mathbf{X}_{T+q}E(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + E(\varepsilon_{T+q}) \\
 &= 0
 \end{aligned}$$

ya que el vector estimado para los coeficientes de regresión es insesgado y el valor esperado de la perturbación aleatoria es cero.

Conseguir una expresión para la varianza del error de predicción es un poquito más trabajoso, pero es una buena oportunidad para ejercitar el Álgebra Matricial.

$$\begin{aligned}
 \text{var}(e_{T+q}^P) &= E[(e_{T+q}^P - E(e_{T+q}^P))(e_{T+q}^P - E(e_{T+q}^P))'] \\
 &= E(e_{T+q}^P \cdot e_{T+q}^P)' \\
 &= E[(\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \varepsilon_{T+q}) \cdot (\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \varepsilon_{T+q})'] \\
 &= E[(\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \varepsilon_{T+q}) \cdot ((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}_{T+q}' + \varepsilon_{T+q}')] \\
 &= E[(\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}_{T+q}') + (\mathbf{X}_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \varepsilon_{T+q}') + (\varepsilon_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}_{T+q}') + (\varepsilon_{T+q} \varepsilon_{T+q}')] \\
 &= \mathbf{X}_{T+q} E((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})') \mathbf{X}_{T+q}' + \mathbf{X}_{T+q} E((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \varepsilon_{T+q}') + E(\varepsilon_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}_{T+q}') + E(\varepsilon_{T+q} \varepsilon_{T+q}') \\
 &= \mathbf{X}_{T+q} E((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})') \mathbf{X}_{T+q}' + \mathbf{X}_{T+q} E((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \varepsilon_{T+q}') + E(\varepsilon_{T+q}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}_{T+q}') + E(\varepsilon_{T+q} \varepsilon_{T+q}') \\
 &= \mathbf{X}_{T+q} \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \sigma^2 \\
 &= \sigma^2 [\mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \mathbf{1}]
 \end{aligned}$$

Por los supuestos del MLNR sabemos que  $Y$  distribuye normal y ello condiciona que así lo hará también  $\hat{Y}_{T+q}$  y el error de predicción  $e_{T+q}^P$ , de modo que podemos escribir que:

$$e_{T+q}^P \sim N(0, \sigma^2 [\mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \mathbf{1}]) \quad (6.23)$$

Como no conocemos  $\sigma^2$ , debemos trabajar con su estimador muestral insesgado  $\hat{\sigma}^2$  y para eso debemos realizar transformaciones sucesivas, como ya hemos ejercitado durante este capítulo, que nos conducen a:

$$\frac{e_{T+q}^P}{\sqrt{\hat{\sigma}^2 [\mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \mathbf{1}]}} \sim t_{n-k} \quad (6.24)$$

A partir de estos resultados ya estamos en condiciones de construir un intervalo de confianza para el pronóstico o predicción:

$$\hat{Y}_{T+q} - t_{n-k}^{\alpha/2} \sqrt{\hat{\sigma}^2 \left[ \mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \mathbf{I} \right]} \leq Y_{T+q} \leq \hat{Y}_{T+q} + t_{n-k}^{\alpha/2} \sqrt{\hat{\sigma}^2 \left[ \mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{T+q}' + \mathbf{I} \right]} \quad (6.25)$$

Vamos a escribir este intervalo de confianza para el caso de la regresión bivariada, para hacer más evidentes algunas cuestiones, más allá del amasijo de vectores y matrices.<sup>11</sup>

$$\hat{Y}_{T+q} - t_{n-k}^{\alpha/2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{\left( X_{T+q} - \bar{X} \right)^2}{\sum_{i=1}^n \left( X_i - \bar{X} \right)^2} \right]} \leq Y_{T+q} \leq \hat{Y}_{T+q} + t_{n-k}^{\alpha/2} \sqrt{\hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{\left( X_{T+q} - \bar{X} \right)^2}{\sum_{i=1}^n \left( X_i - \bar{X} \right)^2} \right]} \quad (6.26)$$

En la notación anterior podemos distinguir fácilmente los factores de los que depende la amplitud del intervalo de confianza del pronóstico:

- ✓ De los grados de libertad disponibles. Mayores tamaños de muestra ( $n$ ) y modelos relativamente sencillos (menor  $k$ ), mejorarán la exactitud de la predicción, contrayendo los límites del intervalo.
- ✓ De la varianza del término de error de la regresión estimada. A mayor  $\hat{\sigma}^2$ , mayor amplitud del intervalo.
- ✓ Directamente del tamaño de la muestra. Mayores tamaños de muestra disminuyen la amplitud del intervalo.
- ✓ De la distancia entre el valor de la variable explicativa que se utiliza para construir el pronóstico y la media muestral,  $(X_{T+q} - \bar{X})$ . Mientras más se aleja  $X_{T+q}$  del promedio de la muestra con la que se estimó el modelo de regresión, mayor será la amplitud del intervalo.
- ✓ De la variabilidad de la variable explicativa en la muestra disponible para la estimación,  $\sum_{i=1}^n (X_i - \bar{X})^2$ . A mayor variabilidad, menor amplitud del intervalo.

Regresemos a la situación de interés de pronóstico que describimos al inicio de este tópico y, con los conocimientos que tenemos, elaboremos una predicción para la tasa de variación de las importaciones de la economía cubana correspondiente al año 2006.

Para ello vamos a utilizar la ecuación de regresión estimada para nuestra muestra para conseguir el predictor que buscamos.

$$\hat{m}_{2006} = 1,26270 + 0,75064 \dot{y}_{2006} - 0,54252 \dot{p}_{2006} \quad (6.27)$$

<sup>11</sup> En el caso del modelo de regresión bivariado,  $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$  con  $i=1, \dots, n$ .

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \text{ y } (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} - \frac{\bar{X}^2}{\sum x_i^2} & \frac{\bar{X}}{\sum x_i^2} \\ \frac{\bar{X}}{\sum x_i^2} & \frac{-1}{\sum x_i^2} \end{bmatrix}, \text{ con } x_i = X_i - \bar{X}$$

Con estas “pistas” el estudiante inquieto podrá comprobar que (6.26) es un caso particular de (6.25)

Necesitamos saber cuál será la variación del PIB ( $\dot{y}_{2006}$ ) y de los precios de las importaciones ( $\dot{p}_{2006}$ ) del 2006 con relación al 2005, para calcular  $\hat{m}_{2006}$ . Vamos a sustraernos del hecho de que este ejercicio se está realizando mucho después del año 2006. Imaginemos que estamos en la etapa de elaboración del plan anual para el 2006. En el caso de  $\dot{y}_{2006}$ , podemos apelar al crecimiento planificado: 10% vs. el año 2005 (Rodríguez, 2005). Para los precios de las importaciones podemos trabajar con dos escenarios: el que denominaremos A, donde suponemos que no varíen con relación al 2005 (0% de crecimiento); y el B, en el que mantienen el mismo crecimiento que en el 2005 (7,3%). Así, tendremos dos variantes para el vector  $\mathbf{X}_{2006}$ ,  $\mathbf{X}_{2006}^A = [1 \ 10 \ 0]$  y  $\mathbf{X}_{2006}^B = [1 \ 10 \ 7,3]$ , y asimismo dos pronósticos para  $\hat{m}_{2006}$ ,

$$\hat{m}_{2006}^A = [1 \ 10 \ 0] \begin{bmatrix} 1,26270 \\ 0,75064 \\ -0,54252 \end{bmatrix} = 8,76912$$

$$\text{y } \hat{m}_{2006}^B = [1 \ 10 \ 7,3] \begin{bmatrix} 1,26270 \\ 0,75064 \\ -0,54252 \end{bmatrix} = 4,83331$$

Esto quiere decir que en el escenario A debemos esperar un crecimiento de las importaciones de aproximadamente un 8,8% para el año 2006, y en el B, de un 4,8%, por los efectos de mantenerse el aumento previo de los precios relativos de las importaciones en ese año. Sin embargo, el plan del 2006 había previsto una disminución del 1,5% en términos corrientes, que incluye el efecto de los precios, por lo que probablemente en términos reales la contracción que se pretende es mucho mayor. A la luz de nuestros pronósticos puntuales este plan de importaciones parecería optimista, sin embargo, debemos recordar que todavía no sabemos si estamos ante un buen modelo de regresión, por lo que no estamos en condiciones todavía de evaluar y mucho menos de hacer pronósticos. Este ejercicio sólo tiene el propósito de enseñar a utilizar las herramientas hasta aquí aprendidas.

Vamos a construir los intervalos de confianza para esas predicciones a los efectos de examinar en qué rango pudiéramos encontrar el valor esperado del crecimiento de las importaciones del 2006. Del trabajo anterior tenemos prácticamente todos los ingredientes necesarios:  $\hat{m}_{2006}$  y  $\mathbf{X}_{2006}$ , en sus dos variantes,  $(\mathbf{X}'\mathbf{X})^{-1}$ ,  $\hat{\sigma}^2$ ; trabajaremos con  $\alpha=5\%$  y con el valor de  $t$  que provee Excel,  $t_{(n-k)}^{\alpha/2} = t_{42}^{0,025} = 2,018$ .

En el recuadro siguiente mostramos algunos resultados intermedios que constituyen “indicios” para replicar por cuenta propia la construcción del intervalo de confianza. Noten las grandes varianzas de los errores de pronóstico, que son consecuencia de la gran varianza del término de error de la regresión.

| Cálculos intermedios                                              | variante A | variante B |
|-------------------------------------------------------------------|------------|------------|
| $\mathbf{X}_{T+q} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_{T+q}$ | 0,04453    | 0,05042    |
| Varianza del error de pronóstico                                  | 93,61718   | 94,14489   |
| Desvío estándar del error de pronóstico                           | 9,6750     | 9,70283    |

En la tabla 6.3 presentamos los intervalos de confianza obtenidos.

**Tabla 6.3: Intervalos de confianza para el pronóstico de la variación de las importaciones reales según el modelo, en porcentaje**

| Variantes | Pronóstico | Intervalo de confianza |                 |
|-----------|------------|------------------------|-----------------|
|           |            | Límite inferior        | Límite superior |
| A         | 8,8        | -10,8                  | 28,3            |
| B         | 4,8        | -14,7                  | 24,4            |

*Elaboración propia.*

Evidentemente, el pronóstico a través de este modelo (6.27) es bastante inexacto, los intervalos de confianza obtenidos son demasiado amplios y, según este resultado, es posible esperar cualquier comportamiento de las importaciones: lo mismo que crezcan como que se contraigan. Ello tiene mucho que ver con que hemos escogido  $\alpha=5\%$ .

En los ejercicios reales de predicción por lo general se utilizan niveles mayores de  $\alpha$  que conduzcan a menores intervalos para el pronóstico.





# *preguntas*

Intente dar respuesta a las siguientes preguntas. Así sabrá si ha captado los principales mensajes de este capítulo.

1. ¿Por qué es importante dar la respuesta de nuestras estimaciones en forma de intervalos de confianza? Explique cómo se obtiene ese intervalo para  $\hat{\beta}$ .
2. ¿En qué consiste el test de significación de los regresores?
3. ¿Qué significa no rechazar la  $H_0$  de un test de significación para un coeficiente de regresión?
4. Hay test de dos colas y de una cola. Explique de qué se trata en cada caso.
5. ¿Cómo podemos testear hipótesis compuestas sobre los coeficientes de regresión estimados?
6. ¿Qué son el  $R^2$  y el  $R^2$  ajustado? ¿Por qué precisamos de “un ajuste”?
7. ¿Podemos utilizar el  $R^2$  cuando queremos evaluar el ajuste fuera de la muestra? ¿Qué otras opciones tendríamos en ese caso?
8. ¿Cómo podemos testear hipótesis simultáneas sobre los coeficientes de regresión estimados?
9. ¿Qué es el test de significación global de un modelo de regresión?
10. ¿Qué significa no rechazar la  $H_0$  de un test de significación global de un modelo de regresión?
11. ¿Qué es una predicción? ¿Por qué son importantes los pronósticos?
12. ¿Cuáles son los factores que influyen en la calidad de una predicción?



# ejercicios

6.1 Considere el modelo lineal  $Y_t = B_0 + B_1X_{1t} + B_2X_{2t} + u_t$ .

Hemos obtenido los siguientes estimados MCO

| <i>Coefficiente</i> | <i>Estimado</i> | <i>Error estándar</i> |
|---------------------|-----------------|-----------------------|
| $B_0$               | 1.289           | 0.215                 |
| $B_1$               | 0.235           | 0.056                 |
| $B_2$               | 2.453           | 0.965                 |
| ESM*                | 2.313           |                       |
| $R^2$               | 0.932           |                       |
| Observaciones       | 20              |                       |

a. Use el enfoque de intervalo de confianza para testear la siguiente hipótesis:

$H_0: B_1 = 0$  versus  $H_1: B_1 \neq 0$  (seleccione un nivel de significación del 5%)

b. Use el enfoque de test de significación para evaluar

$H_0: B_2 = 3$  versus  $H_1: B_2 < 3$  (seleccione un nivel de significación del 5%)

c. Encuentre el valor de p correspondiente al valor del estadístico en b) e interprete el resultado.

d. Encuentre un valor de predicción para  $Y_t$ , cuando  $X_1 = 10$  y  $X_2 = 0.5$ .

e. Evalúe la significación global de la regresión estimada. Plantee la hipótesis nula correspondiente, el estadístico de prueba, el valor de prueba, el valor crítico y concluya convenientemente.

f. Explique el significado de un error de tipo II en la evaluación de una hipótesis y calcule la probabilidad de cometer tal error en el inciso a) con un nivel de significación del 5% y cuando el coeficiente poblacional es igual a 0.2 (utilice la tabla de la distribución normal estándar si lo necesita).

6.2 Considere el siguiente modelo  $Y_t = b_1 + b_2X_{2t} + b_3X_{3t} + u_t$ .

Cuando lo estimamos usando MCO a partir de una muestra de 20 observaciones, obtenemos los siguientes resultados:

\* Error estándar del modelo o regresión

| <i>Variable</i> | <i>Coeficiente</i> | <i>Varianza</i> |
|-----------------|--------------------|-----------------|
| $X_1$           | 0.96587            | 0.21812         |
| $X_2$           | 0.69914            | 0.048526        |
| $X_3$           | 1.7769             | 0.037120        |

$$\hat{\sigma}^2 = 2.5193 \quad R^2 = 0.9466$$

- Calcule la suma total de cuadrados (STC), la suma de cuadrados explicada (SCE) y la suma de residuos al cuadrado (SRC).
- Calcule el  $R^2$  ajustado.
- ¿Son los parámetros estimados significativos a un nivel de significación del 1%? ¿Son significativos al nivel de significación del 5%?
- Encuentre intervalos de confianza del 95% para los coeficientes  $b_2$  y  $b_3$ .
- Evalúe la hipótesis  $b_2 \geq 1$  (considere un nivel de significación del 5%).
- Suponga que la covarianza de  $b_2$  y  $b_3$  es de -0.031223. Evalúe la hipótesis nula de que  $b_2 + b_3 = 2$  (a un nivel de significación del 10%).
- Examine la hipótesis nula conjunta:

$$H_0 : B_2 = 0, \quad B_3 = 0.$$

- 6.3 Estimamos el modelo  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \varepsilon$  utilizando Mínimos Cuadrados Ordinarios. Conseguimos la siguiente tabla ANOVA (note que faltan algunos datos en esa tabla).

| <i>Fuente de la variación</i> | <i>gl</i> | <i>Suma de Cuadrados</i> |
|-------------------------------|-----------|--------------------------|
| Debido a la regresión         | 3         | 820.70204                |
| No-explicada                  | 96        | -                        |
| Total                         | -         | 3556.75835               |

El resumen de resultados de los MCO es el siguiente: (observe que aquí también faltan algunos datos)

| <i>Variable</i> | <i>Coeficiente</i> | <i>Error estándar</i> | <i>razón-t</i> | <i>P&gt; t </i> |
|-----------------|--------------------|-----------------------|----------------|-----------------|
| Constante       | 19.9251            | 44.53348              | 0.447          | 0.6556          |
| $x_1$           | 0.2418             | 0.11317               | -              | 0.0352          |
| $x_1$ -cuadrado | -0.0008            | 0.00056               | -1.429         | 0.1805          |
| $x_2$           | 0.2699             | -                     | 0.144          | 0.8859          |

- ¿Cuántas observaciones tuvimos disponibles para estimar este modelo?
  - Calcule la suma de cuadrados no explicada.
  - Calcule el test- $F$  de significación global del modelo. ¿Cuántos grados de libertad usaría usted con este índice?
  - Calcule el  $R^2$  y el  $R^2$  ajustado.
  - ¿Cuál es el valor- $t$  correspondiente al estimado del parámetro que afecta a  $x_1$  y cuál el error estándar del parámetro que afecta a  $x_2$ ?
  - ¿Qué diría usted acerca de la significación de los parámetros?
  - ¿Sería razonable la hipótesis de un efecto no lineal de  $x_1$ ? ¿Por qué?
  - Estime la varianza del modelo.
- 6.4 Para el modelo planteado y estimado en el ejercicio 5.6.2 c), evalúe:
- Si los coeficientes estimados son estadísticamente diferentes de cero, con un nivel de significación del 5%.
  - La significación global de la función de producción estimada, con un nivel de confiabilidad del 95%.
  - La hipótesis de que la suma de los coeficientes de regresión sea igual a 1.5.
  - Prediga cuál sería el valor más probable del PIB cubano para:
    - $L = 4300, K = 16500, H = 40$
    - $L = 4400, K = 30000, H = 45$

Expresé sus respuestas en forma de intervalo de confianza, para un nivel de significación del 5%. Compare la amplitud de estos intervalos y explique la diferencia entre ellos.
- 6.5 El Centro de Estudios de la Economía Cubana (ver Vidal, 2006) ha desarrollado un modelo econométrico a los efectos de predecir el crecimiento económico del país, a partir de los datos contenidos en la tabla B.4 del disco acompañante. Siguiendo la idea de Vidal (2006), vamos a estimar

preliminarmente un modelo con la siguiente especificación:

$$LPIB_t = \alpha_0 + \alpha_1 LX_t + \alpha_2 LFBK_t,$$

donde  $LPIB_t$  es el logaritmo del PIB,  $LX_t$  el de las exportaciones y  $LFBK_t$  el de la formación bruta de capital fijo.

Se le pide que:

- a. Estime el modelo para el período que comprende la tabla.
- b. Suponiendo que se trate de un modelo bien especificado, evalúe su bondad de ajuste.
- c. Examine también su capacidad predictiva.
- d. Elabore una predicción puntual considerando un crecimiento de las exportaciones del 10% y del 5% para la formación de capital fijo con relación al año 2007.
- e. Exprese la respuesta anterior a través de un intervalo de confianza, al 5% de nivel de significación.
- f. Evalúe si se puede sustentar un crecimiento del PIB cubano para el año 2008 equivalente a un 10%, a partir de los resultados anteriores.

CAPÍTULO 7



## EXTENSIONES DEL MODELO DE REGRESIÓN LINEAL NORMAL

En este capítulo vamos a abordar algunas extensiones del modelo lineal clásico de regresión. Primero nos referiremos a las posibles formas funcionales a las que podemos aplicar los MCO para estimar sus parámetros relevantes. Sobre esta cuestión existe gran confusión, debido al epíteto “lineal” en la nomenclatura del Modelo Lineal Normal de Regresión. Seguidamente explicaremos algunos tipos de variables especiales que son de suma utilidad para modelar efectos cualitativos y estacionales, las variables dicotómicas o mudas. Finalmente, aplicaremos también las variables mudas a la modelación de los cambios de estructura en el modelo de regresión.

### 7.1 Posibles Formas Funcionales

Una de las características del método de los MCO es que los estimadores de  $\beta$ , los  $\hat{\beta}_j$ , son funciones lineales de  $Y$ . Este rasgo y el hecho de que en nuestro modelo la variable dependiente se expresa como una combinación lineal de las explicativas, le confieren el adjetivo de lineal al MLNR. Esto frecuentemente se interpreta como que sólo estamos en capacidad de estimar polinomios de primer grado con el herramental que hemos presentado en el capítulo 5. Esta comprensión del supuesto de linealidad es errónea.

La linealidad a que aludimos se refiere a los  $\beta$  y no a las variables del MLNR. En nuestros modelos pueden aparecer variables con exponentes diferentes de uno o expresadas en términos de otras funciones no lineales. Lo importante es que los coeficientes o parámetros que las enlazan para formar una función sí sean lineales. Por ejemplo, el modelo (7.1) es no-lineal en las variables, pero sí lo es en sus

parámetros  $B_1$ ,  $B_2$ , y  $B_3$ . A eso nos referimos.

$$\ln Y_i = B_1 + B_2 X_i + B_3 X_i^2 \quad (7.1)$$

En este epígrafe presentaremos algunas formas funcionales muy útiles para el trabajo y la investigación en Economía, que pueden ser estimadas por MCO y que no son precisamente lineales. Algunas de ellas ni siquiera lo son en los parámetros, pero aplicando nuestros conocimientos de Matemática podemos transformarlas en funciones lineales en los parámetros y con ello conseguimos aplicar el procedimiento de estimación estudiado.

### 7.1.1 Del modelo exponencial al modelo log-lineal

En el capítulo 6 de este texto ya tuvimos un encuentro con un modelo exponencial, del tipo (7.2): una de las especificaciones funcionales más importantes y utilizada para relacionar la producción con sus factores es la Cobb-Douglas, así denominada en honor a sus creadores e introducida en 1928. Ya presentamos una función de ese tipo para explicar los test t-generalizados.

$$Y_i = B_1 X_{2i}^{B_2} X_{3i}^{B_3} e^{\varepsilon_i} \quad (7.2)$$

Esta función exponencial, no lineal ni en las variables ni en los parámetros, puede ser fácilmente transformada en una función lineal. Basta aplicar una transformación logarítmica. Esta transformación, tomando logaritmos en ambos lados de la ecuación (7.2) nos lleva al siguiente modelo:

$$\ln Y_i = \ln B_1 + B_2 \ln X_{2i} + B_3 \ln X_{3i} + \varepsilon_i$$

Como habrán advertido, hemos convertido la función exponencial en una función lineal en los parámetros  $A$ ,  $B_2$  y  $B_3$ , con  $A = \ln B_1$ .

$$\ln Y_i = A + B_2 \ln X_{2i} + B_3 \ln X_{3i} + \varepsilon_i \quad (7.3)$$

Los parámetros del modelo (7.3) pueden ser estimados por el método de MCO. Por su forma, a este tipo de modelos los denominamos modelos log-log o doble log, pues aparecen los logaritmos tanto en la parte izquierda como en la derecha del modelo; es decir, tanto la variable dependiente como las explicativas están expresadas en logaritmos. También se les conoce con el nombre de modelos log-lineales.

Los coeficientes  $B_2$  y  $B_3$ , correspondientes a las pendientes parciales del modelo log-lineal, tienen una interpretación muy interesante. Por ejemplo, noten que  $B_2$  es la derivada del logaritmo de  $Y$  con respecto al logaritmo de  $X_2$ . ¿Cómo se expresa esa derivada? Podemos asegurarles que es igual a:<sup>1</sup>

<sup>1</sup> Esta es también una buena ocasión para ejercitar los conocimientos de Cálculo Diferencial.



$$\frac{d \ln Y}{d \ln X_2} = \frac{dY/Y}{dX_2/X_2} = \frac{dY}{dX} \frac{X_2}{Y} = B_2 = \frac{\text{cambio relativo en } Y}{\text{cambio relativo en } X_2} = \dot{E}_{YX_2}$$

Y ese resultado es justamente la elasticidad instantánea de  $Y$  con respecto a  $X_2$ ,  $\dot{E}_{YX_2}$ . Así, en este modelo el coeficiente de pendiente mide directamente la elasticidad instantánea. Sabemos que este coeficiente expresa en cuántos puntos porcentuales varía  $Y$  por cada punto porcentual de incremento de  $X$ . Es un parámetro muy útil en Economía. Por eso este tipo de modelo log-lineal es muy popular en el trabajo empírico.

Volviendo a nuestro ejemplo de la función de producción estimada como:

$$LQ = 2,187 - 0,500 LK + 3,714 LL - 1,777 LH$$

tendríamos que el coeficiente asociado a la variable fuerza de trabajo<sup>2</sup> indicaría que por cada punto porcentual que se incremente la cantidad de trabajadores, el producto interno bruto aumentará en aproximadamente 3,7%. Por supuesto, siempre que se trate de un modelo bien especificado, cosa que no hemos demostrado aún.

Si nuestra función en (7.2) fuese una función de demanda, con  $Y$  como la cantidad demandada de un cierto producto,  $X_2$  representado el precio del propio producto y  $X_3$ , el precio de un producto sustitutivo,<sup>3</sup> entonces el coeficiente  $B_2$  sería la elasticidad-precio de la demanda y nos diría en cuántos puntos porcentuales disminuye la cantidad demandada ante un incremento del 1% del precio del producto.

Como es obvio, este modelo log-lineal supone que el coeficiente de elasticidad se mantiene constante en el tiempo, de ahí que reciba también el nombre de modelo de elasticidad constante. Esto quiere decir, en términos prácticos, que sea cual fuere el nivel de la variable explicativa  $X_2$ , cuando ella se incrementa en un punto porcentual, ello siempre desencadenará una variación en  $Y$  de  $B_2$  puntos porcentuales.

En un modelo lineal, del tipo  $Y_i = B_1 + B_2 X_{2i} + B_3 X_{3i} + \varepsilon_i$ , la elasticidad de  $Y$  con respecto a, por ejemplo,  $X_2$  no se mantiene constante, sino que varía a lo largo de todo su recorrido, tomando un valor distinto para cada nivel de  $(X_2, Y)$ , asumiendo el resto de las explicativas constantes. Para este tipo de función, la elasticidad instantánea con respecto a  $X_2$  respondería a la expresión

<sup>2</sup> Observen que este es el único de los coeficientes estimados que tiene un signo coherente con la teoría económica y por eso es el único que interpretamos.

<sup>3</sup> Lógicamente, con  $B_2 < 0$  y  $B_3 > 0$ .

$$\frac{dY}{dX_2} \frac{X_2}{Y} = B_2 \frac{X_2}{Y}, \text{ mientras que la elasticidad media la obtendríamos como}$$

$$E_{YX_2} = B_2 \frac{\bar{X}_2}{\bar{Y}}.$$

La teoría económica, por lo general, nos informa acerca de cuáles variables se relacionan y cuál es el sentido de la relación, pero no sugiere cuál es la forma funcional correcta que adoptan esas relaciones. Para escoger entre un modelo lineal o log-lineal, el diagrama de dispersión de los datos de la muestra puede ofrecer una pista importante.

Por otra parte, aunque con los MCO obtenemos estimadores insesgados para  $\alpha$  y  $B_2$  del modelo (7.3), a través de  $\hat{\alpha}$  y  $\hat{B}_2$ , el parámetro  $B_1$  de nuestro modelo original (7.2) tendría que ser estimado como  $e^{\hat{\alpha}}$ , que es un estimador sesgado. Sin embargo, por lo general ese parámetro de posición o intercepto de la función no resulta de interés teórico, por lo que no debemos preocuparnos por el sesgo en ese coeficiente.

El modelo log-lineal que hemos presentado para dos variables explicativas se puede generalizar para muchas más variables. Cada coeficiente de pendiente parcial mide en este caso la elasticidad parcial con respecto a la variable en cuestión. Un ejemplo lo tenemos en la función de producción antes aludida, cuyos coeficientes miden la elasticidad-capital físico, la elasticidad-trabajo y la elasticidad-capital humano del producto, respectivamente.

### 7.1.2 Los modelos semilogarítmicos

Los modelos semilogarítmicos, como su denominación nos indica, presentan el logaritmo en alguna de sus componentes, bien en la variable dependiente o bien en las explicativas. En el primer caso se trata de los modelos log-lin; en el segundo, de los lin-log.

#### Modelos log-lin

Estos modelos son también muy populares en Economía, pues sirven para estimar la tasa de crecimiento instantánea de cualquier variable. Parten de la idea del cálculo del interés compuesto:

$$Y_t = Y_0 (1+r)^t \quad (7.4)$$

donde el subíndice  $t$  nos indica que estamos tratando con una serie de tiempo,  $Y_0$  es el monto del depósito inicial,  $r$  es la tasa de crecimiento compuesta de  $Y$  (expresada en tanto por uno) y  $Y_t$  es la magnitud del depósito después de transcurridos  $t$  períodos de tiempo. Adviertan que  $r$  se mantiene constante durante el período.

Este planteamiento también se puede aplicar para el cálculo de las tasas de crecimiento instantánea y compuesta de cualquier variable económica. Sólo es

preciso asimilar  $Y_0$  como el valor de la variable al inicio del período de estudio y  $r$  como la tasa de crecimiento geométrico promedio de la variable durante el período  $[0, t]$ . Si aplicamos logaritmos a la expresión (7.4), llegamos a  $\ln Y_t = \ln Y_0 + t \ln(1+r)$ ; y si, además, hacemos  $B_1 = \ln Y_0$ ,  $B_2 = \ln(1+r)$  y agregamos un término de error estocástico  $u_t$ , obtenemos la ecuación:

$$\ln Y_t = B_1 + B_2 t + u_t \quad (7.5)$$

que es lineal en los parámetros  $B_1$  y  $B_2$ , por lo que podemos considerarla un modelo de regresión lineal, sólo que con dos variables muy específicas, el tiempo  $t$  como explicativa y  $\ln Y_t$  como regresada o dependiente.

El modelo (7.5) pertenece a la clase de modelos semilog porque sólo una de sus variables se expresa en términos logarítmicos. En este caso, la variable dependiente es un logaritmo y la explicativa se mantiene lineal, por lo que adopta el nombre de log-lin.

Si derivamos el logaritmo de  $Y$  con respecto al tiempo en (7.5), llegamos a:

$$\frac{d \ln Y}{dt} = \frac{dY/Y}{dt} = B_2 = \frac{\text{cambio relativo en } Y}{\text{cambio absoluto en } t}$$

así, el coeficiente  $B_2$  —que es la pendiente del modelo (7.5)— expresa el cambio relativo en la variable  $Y$  por cada incremento absoluto del tiempo en una unidad. Entonces, si multiplicamos  $B_2$  por 100, obtendremos el cambio porcentual en  $Y$  que “origina” un incremento de una unidad en el decursar del tiempo. En este caso estamos ante la tasa de crecimiento instantánea y no la promedio del período. Para recuperar nuestra tasa promedio debemos volver a  $B_2 = \ln(1+r)$ , aplicar la función exponencial con base  $e$  (que como sabemos es la función inversa del logaritmo neperiano) y despejar, para obtener la tasa media  $r = e^{B_2} - 1$ . Igualmente, el valor de  $Y_0$  lo recuperamos como  $e^{B_1}$ .

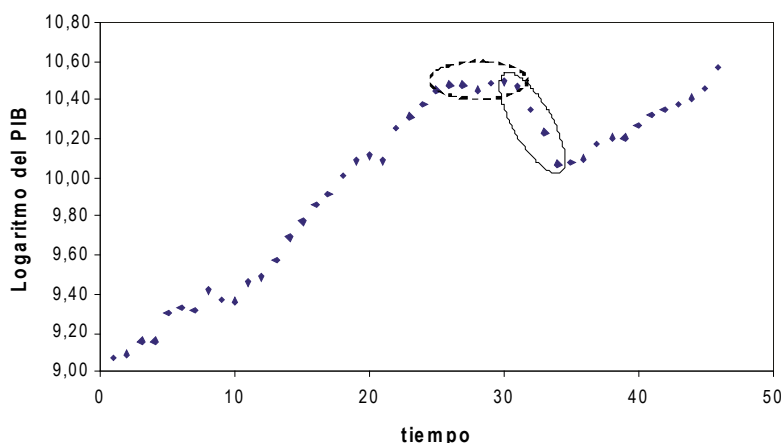
Debemos advertir que, si bien  $\hat{B}_1$  y  $\hat{B}_2$  son insesgados, las estimaciones de  $r$  y  $Y_0$  sí lo serán, por lo que los resultados de este tipo de modelos deben ser tomados con cautela.

Apliquemos lo aprendido para determinar el ritmo de incremento del PIB cubano en el período 1960-2005. La serie temporal de esa variable a precios constantes de 1997 y que hemos denotado desde el capítulo 5 con  $Y_t$ , la mostramos en la tabla B.1. del CD acompañante.

Antes de lanzarnos a estimar la tasa de crecimiento media, recomendamos hacer un gráfico de dispersión de esta serie, para saber si será pertinente realizar tal tarea. Sabemos que en un período temporal tan extenso han ocurrido sucesos

que ponen en duda suponer que el crecimiento económico en Cuba se haya mantenido constante. Mostramos tal gráfico en la figura 7.1.

**Figura 7.1: Recorrido del PIB real cubano, 1960-2005 en términos logarítmicos**



*Fuente: Elaboración propia a partir de ONE (Varios años)*

Evidentemente, asumir que la tasa de crecimiento de la economía cubana se mantuvo constante en el período 1960-2005 no es adecuado. Se observan cambios en la tasa promedio durante el período. El más fuerte, sin dudas, asociado al derrumbe del campo socialista y que hemos resaltado con una elipse continua. En esa figura también podemos apreciar que en la segunda mitad de los años 80 ya no crecimos con el mismo ímpetu anterior (destacado con elipse discontinua). Así, es más apropiado estimar un ritmo o tasa de crecimiento para el período entre 1960 y 1985. El resultado de este ejercicio se muestra seguidamente:

$$\begin{array}{rcl} \ln Y_t = & 8,9234 & + \quad 0,0587t \\ se & (0,0284) & (0,0018) \\ t & (314,0992)^* & (31,9042)^* \end{array}$$

Como se hace habitualmente en los ejercicios empíricos, hemos reportado inmediatamente debajo de la ecuación de regresión estimada y entre paréntesis los valores del error estándar de los coeficientes, así como el valor del estadístico  $t$  que les corresponde. Hemos colocado un asterisco como superíndice de los valores de  $t$  para indicar que los coeficientes estimados son significativamente distintos de cero. Los valores de la probabilidad  $p$  en cada caso son menores que  $\alpha = 1\%$ .

¿Cómo interpretar estos coeficientes de regresión?

$B_2 = 0,0587$  nos revela que en un período de un año el PIB real cubano creció a un ritmo de aproximadamente 5,9%. La tasa de crecimiento promedio anual para el período 1960-2005 se obtiene a partir del procedimiento antes explicado:

$r = e^{B_2} - 1 = e^{0,0587} - 1 = 0,0604$ , así la tasa que buscamos es de algo más de un 6%.

$B_1 = 8,9234$ , el intercepto de la regresión, es el logaritmo del valor de  $Y_0$ . En nuestro ejemplo el momento “cero” corresponde al año 1959; así, el PIB cubano en 1959 ascendió aproximadamente a 7505,8 millones de pesos de 1997.

### Modelos lin-log

Otro tipo de modelo semilogarítmico es aquel que relaciona una variable dependiente lineal con variables explicativas en logaritmos. Por esta razón se le denomina lin-log, al revés del caso anterior. Su expresión analítica con una sola variable explicativa es:

$$Y_t = B_1 + B_2 \ln X_t + u_t \quad (7.6)$$

Este planteamiento permite encontrar la variación absoluta de la variable  $Y$  ante un cambio relativo de la variable  $X$ , a través del coeficiente de pendiente  $B_2$ . Para verificar esto no tienen más que calcular la derivada de  $Y$  con respecto al  $\ln X$ , tal y como mostramos a continuación:

$$\frac{dY}{d \ln X} = \frac{dY}{dX/X} = B_2 = \frac{\text{cambio absoluto en } Y}{\text{cambio relativo en } X}$$

Por lo que el cambio absoluto en  $Y$  podrá expresarse a partir de  $dY = B_2 \frac{dX}{X}$ , es decir, multiplicando la pendiente de **¡Error! No se encuentra el origen de la referencia.** por el cambio relativo en  $X$ . Deberán tener cuidado con la forma de expresar el cambio relativo para obtener el efecto absoluto en  $Y$ , trabajando siempre en tanto por uno.

Estos modelos lin-log encuentran una aplicación práctica en los modelos de gasto de Engel, así nombrados en consideración a su creador, el estadístico alemán Ernst Engel (1821-1896). Al estudiar los gastos de consumo de las familias, Engel observó que el gasto que se dedica a la compra de alimentos tiende a incrementarse en progresión aritmética, mientras que el gasto total de consumo aumenta en progresión geométrica. Es decir, el gasto en alimentos se incrementa más lentamente que el gasto total, por lo que este tipo de modelo es el adecuado para relacionar los crecimientos de ambas variables.

### 7.1.3 Los Modelos recíprocos

Los modelos recíprocos responden al siguiente planteamiento:

$$Y_i = B_1 + B_2 \left( \frac{1}{X_i} \right) + u_i \quad (7.7)$$

Noten que este modelo no es lineal en sus variables, ya que en su parte

derecha tenemos el recíproco de  $X$ ,  $1/X$ , como variable explicativa; sin embargo, el modelo es lineal en sus coeficientes  $B_1$  y  $B_2$ , por lo que estamos ante un modelo de regresión lineal. Y si definimos  $Z=1/X$  y lo reescribimos como  $Y_i = B_1 + B_2 Z_i + u_i$ , entonces es también lineal en las variables.

En este modelo, cuando  $X$  crece indefinidamente, su recíproco tiende a cero y  $Y$  tiende a  $B_1$  como un límite asintótico.

Existen disímiles situaciones en Economía que pueden ser representadas a través de un modelo como (7.7). Por ejemplo, la función de costos totales unitarios responde a esa forma funcional. Esta afirmación podemos comprobarla fácilmente. La función de costos totales se define como:

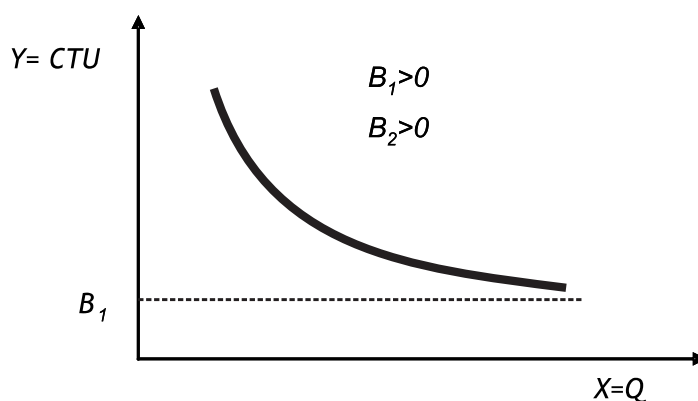
$$CT = CF + CV = CF + Q \cdot CVU$$

donde  $CT$  son los costos totales,  $CF$  los costos fijos,  $CV$  los costos variables,  $Q$  la producción y  $CVU$  los costos variables unitarios. Si dividimos esa función entre el nivel de producción  $Q$ , obtenemos la función de costos totales unitarios  $CTU$ :

$$CTU = \frac{CF}{Q} + CVU$$

Reparen que en esa función los costos fijos y los costos variables unitarios son constantes, es decir, son sus parámetros. La variable explicativa es precisamente el recíproco o inverso de la producción  $1/Q$ . Así, la función de costos totales unitarios es una función recíproca, con  $B_1 = CVU$  como asíntota horizontal. Como  $B_1$  y  $B_2$  en este caso son magnitudes mayores que cero, esta función se representa gráficamente como mostramos en la figura 7.2.

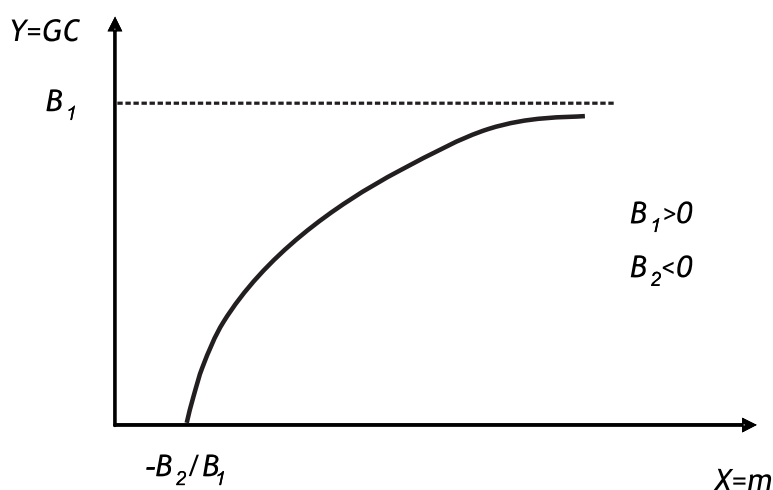
**Figura 7.2: Modelo recíproco con  $B_1, B_2 > 0$**   
Función de costos totales unitarios



Fuente: Elaboración propia

Otra relación económica que puede ser representada por una función recíproca es la conocida curva de gastos de consumo de Engel. En sus estudios sobre el consumo, Engel encontró que existe un límite máximo al que tiende esta variable a medida que se incrementa el ingreso; es decir, existe un nivel de saciedad para el consumo. La forma de la curva de gastos de consumo de Engel se muestra en la figura 7.3, donde la variable dependiente, gastos de consumo, se representa con  $GC$  y la variable explicativa, ingresos, se representa con  $m$ .

**Figura 7.3: Modelo recíproco con  $B_1 > 0$  y  $B_2 < 0$ , función de gastos de consumo de Engel**



*Fuente: Elaboración propia*

Noten que en la función de Engel, a medida que se aumenta el ingreso, los gastos de consumo se incrementan, pero cada vez en menor proporción a medida que nos alejamos del nivel de ingreso que corresponde a  $GC = 0$ . Ese nivel es igual a  $-B_2/B_1$ <sup>4</sup> y constituye un umbral crítico de ingreso. Cuando el ingreso tiende a infinito, según esa función el gasto de consumo tiende a  $B_1$ , que es su asíntota horizontal.

#### 7.1.4 Los Modelos polinomiales

Los polinomios<sup>5</sup> son funciones que también se utilizan para modelar las relaciones entre algunas variables económicas. Ellos se describen con la forma funcional  $Y_i = B_1 + B_2X_i + B_3X_i^2 + \dots + B_gX_i^g$ , con  $g$  como el exponente mayor, que se denomina grado del polinomio.

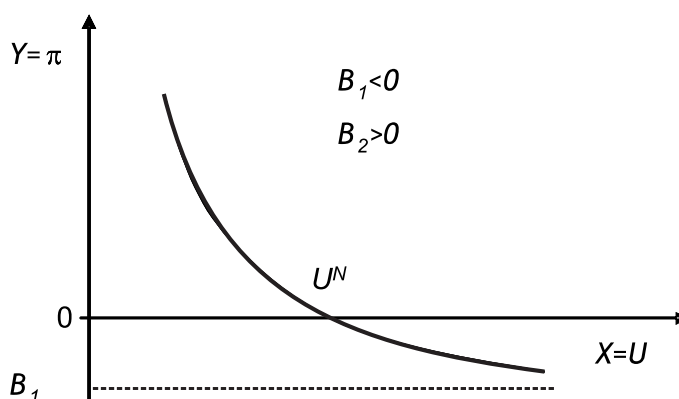
<sup>4</sup> Sólo deben despejar  $m$  en  $GC = B_1 + B_2(1/m)$  con  $GC = 0$ .

<sup>5</sup> En este tópico nos referiremos a los polinomios con grado mayor que uno.

Algunos investigadores se preocupan porque el polinomio incorpora como variables explicativas a diferentes potencias de la misma variable y, en su criterio, esto puede conducir a problemas de multicolinealidad.<sup>6</sup> Sin embargo, debemos acotar que esto no necesariamente es así, ya que  $X^2, X^3, \dots, X^g$  son funciones no-lineales de  $X$ .

Un ejemplo de relación económica que se describe a través de un polinomio de segundo grado es la relación entre los ingresos ( $I$ ) y la edad ( $E$ ) de cada individuo, que denominaremos función de ingresos personales y se formula como  $I_i = B_1 + B_2 E_i + B_3 E_i^2$ . Representamos esta función en la figura 7.4. Noten que en esta función  $B_2 > 0$  y  $B_3 < 0$ .

**Figura 7.4: Modelo recíproco con  $B_1 < 0$  y  $B_2 > 0$ ,  
Curva de Phillips**



Fuente: Elaboración propia

El parámetro  $B_2 > 0$  sugiere que a medida que aumenta la edad los ingresos progresan. Sin embargo, al incluir el parámetro  $B_3 < 0$  estamos apuntando a que la tasa de progreso es cada vez menor y que a partir de cierta edad pasa a ser negativa.

## 7.2 Variables Dicotómicas

En Economía no siempre las variables son de tipo cuantitativo —como el PIB, los costos, los salarios, el desempleo. Muy frecuentemente nos interesa investigar si ciertas características tienen influencia o no en una determinada magnitud eco-

<sup>6</sup> Nos referimos a la violación o no cumplimiento del supuesto D del MLNR.



nómica. Por ejemplo, puede interesar saber si el color de la piel, el género o la región geográfica de residencia tienen influencia en los niveles de salario que se perciben; o si una determinada variedad de planta presenta mejores rendimientos que otra; o si la temporada (o estación) del año tiene influencia en las ventas o la oferta; o si una determinada medida de política económica implementada en un cierto año, desencadenó realmente comportamientos diferentes al período anterior a su aplicación. En estas situaciones se pueden utilizar variables dicotómicas para modelar estos efectos.

Las variables dicotómicas son variables especiales, que toman sólo dos valores: cero o uno. Así, resultan sumamente útiles para modelar características cualitativas de los fenómenos, asociando el valor uno de la dicotómica con la presencia de la característica de interés y cero con la ausencia de la característica.

A las variables dicotómicas se les conoce también como variables mudas,<sup>7</sup> binarias, indicador, categoriales o de intervención.

Un modelo de regresión puede contener sólo variables dicotómicas como explicativas. Este tipo de modelos se denomina modelo de análisis de varianza (modelo ANOVA). En el caso de que el modelo combine variables explicativas cuantitativas y dicotómicas, lo denominamos modelo de análisis de covarianza (modelo ANCOVA).

Vamos a ilustrar esta cuestión de las variables dicotómicas a través de un ejemplo. En los Anuarios Estadísticos de Cuba ustedes podrán encontrar información acerca del salario medio mensual de los trabajadores de las entidades en el sector estatal y mixto por provincias (ver tabla B.3 del CD acompañante). Supongamos que quisiéramos saber si existen diferencias significativas entre los promedios salariales correspondientes a las tres grandes regiones geográficas del país; a saber, Occidente (de Pinar del Río a Matanzas, incluyendo el municipio especial Isla de la Juventud), Centro (de Villa Clara hasta Camagüey) y Oriente (de Las Tunas en adelante, hasta Guantánamo). Vamos a plantear un modelo que explique la variable salarios en función de la región a la que pertenece cada provincia. Así, la variable dependiente serán los salarios, que denotaremos con  $S_i$ , con  $i$  el subíndice que identifica a cada provincia; y las variables explicativas serán de tipo cualitativo, en este caso las regiones. Tenemos tres categorías para nuestra variable cualitativa, a tenor con las tres regiones. Sin embargo, como vamos a plantear un modelo de regresión con intercepto, vamos a incorporar sólo dos categorías.<sup>8</sup> Llamaremos  $D_{1i}$  a la variable dicotómica o muda que expresa la pertenencia de la provincia  $i$  a la región occidental y  $D_{2i}$ , la correspondiente a la región central. El modelo queda especificado de la siguiente manera:

$$S_i = \beta_3 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i \quad (7.8)$$

con  $u_i$  como el término de error de la regresión.  $D_{1i}$  asume el valor 1 si la provin-

<sup>7</sup> Dummy, por su notación en inglés.

<sup>8</sup> ¿Por qué?

cia  $i$  se encuentra en la zona occidental y  $D_{2i}$  será igual a la unidad si la provincia  $i$  pertenece a la zona central. Advertan que hemos denotado a las variables dicotómicas con la letra  $D$  más un subíndice que expresa la categoría que corresponde. Esta es una práctica habitual en el manejo de este tipo de variables.

El modelo (7.8) es similar a cualquier otro modelo de regresión múltiple que hayamos visto antes, con la diferencia de que las variables explicativas son todas dicotómicas.

¿Cómo interpretar el modelo (7.8)? Asumiendo que el término de error  $u_i$  cumpla con los supuestos que exige un MLNR, podemos interpretar a  $S_i = \beta_3 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$  como la media de  $S_i$  condicionada a los valores que asuman las variables dicotómicas:

- ✓  $E(S_i | D_{1i} = 1, D_{2i} = 0) = \beta_3 + \beta_1$ , será el salario medio en la región occidental;
- ✓  $E(S_i | D_{1i} = 0, D_{2i} = 1) = \beta_3 + \beta_2$ , el salario medio en la región central y
- ✓  $E(S_i | D_{1i} = 0, D_{2i} = 0) = \beta_3$ , el salario medio en la región oriental.

Noten que el intercepto de la función de regresión corresponde a la media de la región para la que no incluimos variable dicotómica y que las medias de las restantes regiones se obtienen como la suma de la media de la región oriental más los coeficientes de regresión que corresponden a las respectivas variables dicotómicas. Así, estos coeficientes aportan la diferencia entre las medias de las regiones Occidente y Centro con respecto al Oriente, que de esta forma se constituye en punto de referencia. ¿Serán estas diferencias entre regiones significativas? El haber planteado el modelo de regresión (7.8) ayuda mucho a ese análisis, ya que si demostramos que cumple los supuestos de un MLNR (es decir, que se trata de un modelo bien especificado), entonces el test  $t$  de significación de los regresores de las variables mudas nos puede dar la respuesta a nuestra inquietud.

Si bien estas variables dicotómicas, binarias, mudas o dummies se pueden incorporar a los modelos de regresión con relativa facilidad, hay que tener precaución al hacer esto.

Un error frecuente es añadir tantas variables mudas como categorías se quieran distinguir en el estudio empírico. Si especificamos el modelo con un intercepto, esta práctica conduce a un modelo que no se puede estimar. Volviendo a nuestro ejemplo anterior de los salarios promedio por provincias, si pretendemos diferenciar las observaciones en la muestra atendiendo a la región geográfica, no debemos añadir tres dummies (una para Occidente, otra para el Centro y otra más para Oriente). Si lo hiciéramos, no podríamos estimar el modelo, porque estaríamos ante un caso de multicolinealidad perfecta: la combinación de las tres dummies es una variable idéntica al el intercepto del modelo de regresión. Así, si la variable cualitativa tiene  $m$  categorías, sólo hay que agregar  $m - 1$  variables dicotómicas. No respetar esta regla nos lleva a la situación antes descrita, que se conoce con el término acuñado como trampa de la variable dicotómica.

La categoría a la cual no se asigna variable dicotómica la denominamos categoría de base, de comparación, *benchmark*, de control, de referencia u omitida, pues todas los contrastes se hacen con respecto a esa categoría. El valor de la intersección, que denotamos con  $\beta_3$ , representa el valor esperado o media condicionada de esa categoría de comparación. En nuestro ejemplo, habíamos dejado sin dummy a la región oriental, por lo que esta es nuestra región de control o referencia. Ya discutimos antes que los resultados de los regresores correspondientes a las dummies incluidas lo que expresan es la diferencia entre los valores esperados, o medias de la categoría de que se trate (región occidental o central), con respecto al promedio del Oriente del país. Por esto los coeficientes correspondientes a las variables dicotómicas se denominan coeficientes de intercepto diferencial.

El investigador siempre deberá escoger cuál categoría utilizará como referencia o control. Sugerimos que esta selección se realice teniendo en cuenta la respuesta que busca con el MLNR postulado. En nuestro ejemplo anterior, escogimos a la región oriental como la de referencia, pues queremos investigar precisamente los posibles diferenciales de salario del resto de las regiones con la oriental.

Los resultados de la regresión basada en el modelo (7.8), y con los datos de los salarios promedios para las provincias cubanas en el año 2005 (véase la tabla B.3 del CD acompañante), se plasman a continuación:

$$\begin{array}{rcccl} S_i & = & 319,2 & + & 7,8 D_1 + 12,2 D_2 \\ se & & (4,5898) & & (6,4910) \quad (6,4910) \\ t & & (69,5448)^* & & (1,2017) \quad (1,8795) \end{array}$$

Noten que ninguno de los coeficientes de regresión de las dummies se presenta significativo. Si este modelo estuviese bien especificado, podríamos decir entonces que no existen diferencias significativas entre los salarios medios mensuales por regiones en la Cuba del 2005, por la no-significación de  $\beta_1$  y  $\beta_2$ .

A partir de esos resultados, los salarios medios por regiones son:  $\beta_3 = 319,2$  para Oriente,  $\beta_3 + \beta_1 = 319,2 + 7,8 = 327,0$  para Occidente y  $\beta_3 + \beta_2 = 319,2 + 12,2 = 331,4$  para la zona central, todo medido en pesos corrientes. Sin embargo, estas diferencias entre regiones no resultan significativas estadísticamente.

Una especificación alternativa a (7.8) es el modelo:

$$S_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \quad (7.9)$$

Fíjense que hemos eliminado  $\beta_3$  como intercepto de la regresión, hemos incorporado una nueva dummy  $D_{3i}$ , igual a uno si la provincia está en la zona oriental, y hemos utilizado el parámetro  $\beta_3$  como su coeficiente de pendiente.

Si estimamos (7.9) por MCO, obtenemos los siguientes resultados:

$$\begin{array}{rcccl}
S_i = & 327,0 D_1 + & 331,4 D_2 + & 319,2 D_3 \\
se & (4,5898) & (4,5898) & (4,5898) \\
t & (71,2442)^* & (72,2029)^* & (69,5448)^*
\end{array}$$

Observen que en este caso los coeficientes de regresión nos dan directamente el valor del salario promedio para cada región, tal y como habíamos adelantado. Noten también que todos estos coeficientes son altamente significativos desde el punto de vista estadístico. La cuestión es entonces investigar si podemos suponer que esos regresores son estadísticamente iguales; es decir:

$$\begin{aligned}
H_0 : \beta_1 &= \beta_2 = \beta_3 \\
H_1 : \beta_1 &\neq \beta_2, \text{ o } \beta_1 \neq \beta_3, \text{ o } \beta_2 \neq \beta_3
\end{aligned}$$

Esta puede ser una buena ocasión para ejercitar los contenidos del capítulo 6 en su tópico de test de hipótesis simultáneas. Esperamos que consigan no rechazar  $H_0$ .

Este modelo de nuestro ejemplo podría complicarse más, introduciendo otras variables cualitativas, como puede ser el género. Vamos a incorporar entonces una nueva variable a (7.8) que denotaremos  $D_H$  y que será igual a uno para el caso de los hombres, de modo que la especificación se modifica como se muestra a continuación:

$$S_i = \beta_3 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_H D_{Hi} + u_i \quad (7.10)$$

Cuando se incorpora más de una variable cualitativa al modelo de regresión, es muy importante tener claridad sobre cuál es el grupo de referencia con que se está trabajando, para poder interpretar adecuadamente los resultados. Por ejemplo, en el modelo (7.10) la categoría de control o de referencia serán las mujeres de la zona oriental.

Es posible también combinar variables cualitativas y cuantitativas como variables explicativas en el MLNR, tal y como mostramos en el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + u_i \quad (7.11)$$

con  $X_i$  como variable cuantitativa y  $D_i$  como binaria o dicotómica.

Por ejemplo, supongamos que en (7.11) la variable dependiente  $Y_i$  es el rendimiento agrícola de la producción de caña en la finca  $i$ ;  $X_i$ , la cantidad de fertilizante aplicado y  $D_i$ , una variable cualitativa que representa el tipo de variedad de caña que se ha plantado, con sólo dos variedades disponibles, la “tradicional” y una recién desarrollada en el Instituto de Investigaciones de la Caña de Azúcar.

Si quisiéramos evaluar si el tipo de variedad utilizado tiene un efecto beneficioso sobre el rendimiento agrícola, pudiéramos utilizar el modelo (7.11), con  $D_i = 1$  para la nueva variedad. Naturalmente, si dejáramos la especificación del modelo tal cual se muestra en (7.11), obtendríamos resultados distorsionados al estimar la regresión. De hecho, para que el planteamiento en (7.11) tenga sentido, tendríamos que asumir que se trata de fincas con igual disponibilidad de

fuerza de trabajo, con idéntica calificación y experiencia, así como con similares condiciones para el resto de los factores que pueden influir en los rendimientos agrícolas (tamaño de la finca, calidad del suelo, disponibilidad de agua, entre otros). Por lo tanto, si queremos conducir adecuadamente este estudio, tenemos que incluir además de las variables sobre las que deseamos investigar, otras variables que llamaremos de control o controles. Ellas nos permiten controlar por las diferencias atribuibles a otros factores que influyen en la variable dependiente, más allá de los factores que constituyen el motivo de nuestra investigación.

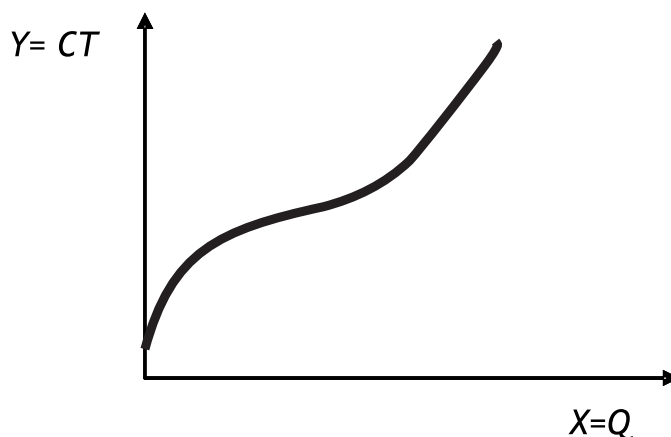
Vamos a modificar la especificación de (7.11), adicionando otras variables de control ( $X_{ji}$ , para  $j = 3, \dots, m$ ).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 D_i + \sum_{j=3}^m \beta_j X_{ji} + u_i \quad (7.12)$$

En el caso de que logremos estimar  $\hat{\beta}_3$  en (7.12) y éste sea significativamente distinto de cero y positivo, entonces no rechazamos la hipótesis de que la nueva variedad influye provechosamente en los resultados productivos y tendremos dos funciones para el rendimiento cañero: una con constante  $\hat{\beta}_1$ , correspondiente a la variedad tradicional, y otra con constante  $\hat{\beta}_1 + \hat{\beta}_3$ , para la nueva variedad.

Para simplificar la representación gráfica de este ejemplo, sólo hemos incluido dos dimensiones (observar la figura 7.5).

**Figura 7.5: Modelo polinomial,  
la función de costos totales**



*Fuente: Elaboración propia*

La ventaja de estimar de forma conjunta ambas funciones es obvia: aunque se incrementan en uno los parámetros a estimar, mantenemos el tamaño de la muestra original, de modo que salimos ganando en términos de grados de libertad.

Obviamente, la situación de referencia en este análisis es la estimación de dos modelos distintos, uno para cada variedad.

En los ejemplos hasta aquí examinados hemos incorporado únicamente variables dicotómicas de intercepto. Estas variables las denominamos también forma aditiva.

Sin embargo, también podemos incorporar a nuestro MLNR variables dicotómicas de pendiente. Para ilustrar este caso vamos a continuar con la discusión del ejemplo de la introducción de una nueva variedad a la producción de caña. En (7.11) modelamos el efecto de esta nueva variedad como un desplazamiento de la función de producción cañera: con independencia del nivel de fertilizante aplicado, para la nueva variedad se logra un incremento del rendimiento siempre constante y equivalente a  $\beta_3$ ; la pendiente de ambas funciones se mantiene invariable, y es igual a  $\beta_2$ .

Pero el efecto benéfico de la nueva variedad podría expresarse también en que consiga un rendimiento incremental proporcional al nivel de fertilizante aplicado. Es decir, que el efecto se refleje como una rotación de la función de producción alrededor de su intercepto  $\beta_1$ .

Este efecto se modela a través de una dummy de pendiente, de la siguiente forma:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i X_i + u_i \quad (7.13)$$

Fíjense que en (7.13) la dummy incluida aparece multiplicando a la variable cuantitativa. Este tipo de dummy recibe también el nombre de dummy interactiva o forma multiplicativa.

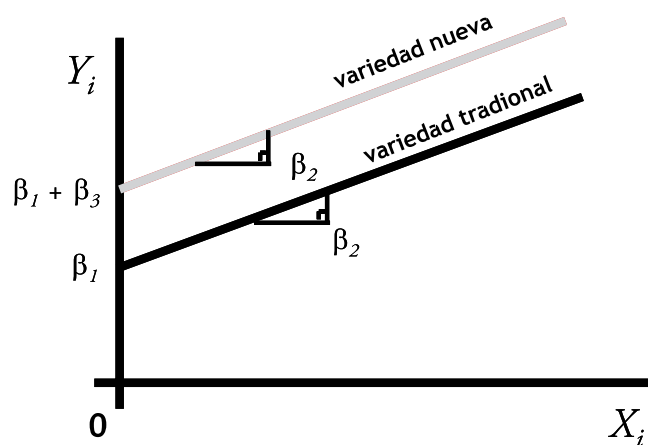
En este caso, si lográramos verificar que  $\hat{\beta}_3$  es significativamente distinto de cero, entonces la función de producción correspondiente a la nueva variedad es  $Y_i = \beta_1 + (\beta_2 + \beta_3) X_i + u_i$ , mientras que para la variedad tradicional será  $Y_i = \beta_1 + \beta_2 X_i + u_i$ .

Igualmente, la utilización de la dummy de pendiente o interactiva nos permite resumir dos modelos distintos en un solo modelo, con la consecuente ganancia de grados de libertad.

La representación gráfica de un modelo con dummy de pendiente la pueden observar en la figura 7.6.

Antes de cerrar este tópico, discutamos una aplicación más de las dummies de intercepto y pendiente que ilustra muy bien cuándo modelar de una u otra forma y que puede ser de interés en nuestro contexto. En un cierto país las autoridades del Ministerio del Trabajo están preocupadas por confirmar si existe discriminación de género en los salarios, en cuyo caso desean proponer e implementar un plan de acción para solucionar esta condición a favor de la mujer trabajadora.

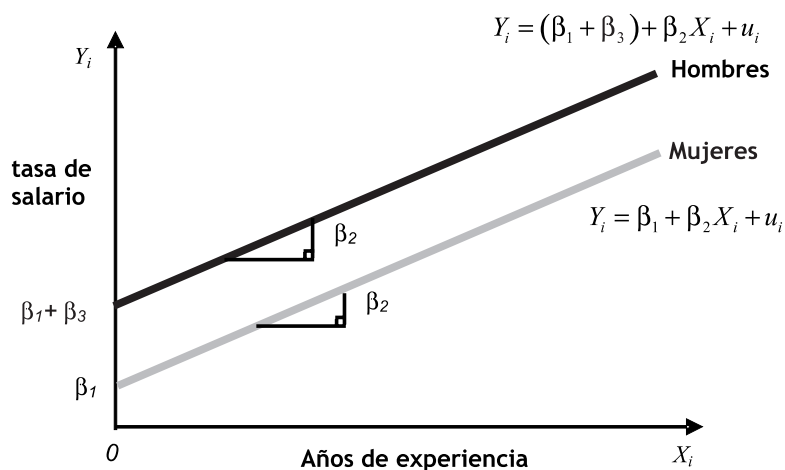
**Figura 7.6: Variable dicótoma de intercepto, el efecto de una nueva variedad de caña sobre el rendimiento agrícola**



Fuente: Elaboración propia

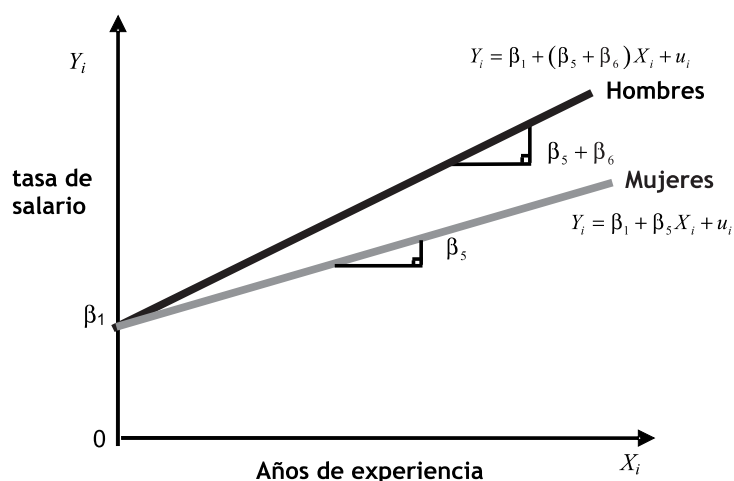
Existen dos opiniones al respecto. Unos especialistas plantean que el problema reside en que hombres y mujeres acceden al mercado de trabajo con un diferencial de salario inicial, que se mantiene a lo largo de la vida laboral de ambos géneros; es decir, asumen que la diferencia salarial entre los géneros se puede modelar a través de una dummy de intercepto (véase la figura 7.7). Ellos proponen resolver este problema de discriminación ofreciendo un salario inicial mayor a las mujeres.

**Figura 7.7: La discriminación salarial de la mujer, modelación con variable dicotómica de intercepto**



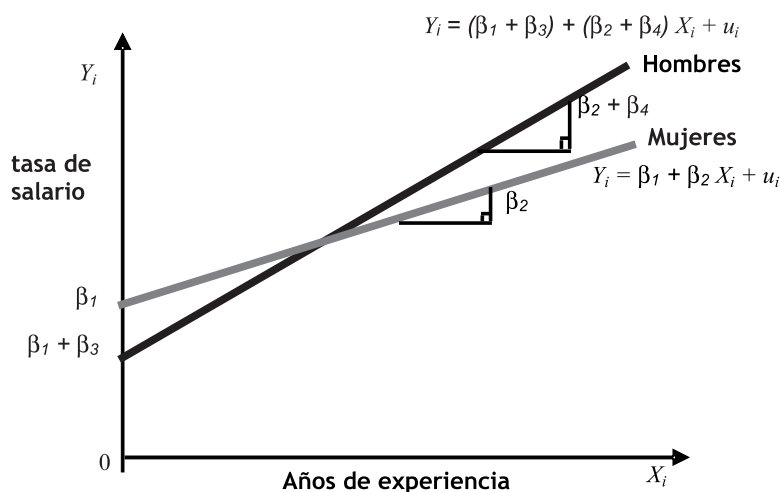
Sin embargo, otros expertos opinan que el problema de la discriminación salarial de la mujer no estriba en que su salario inicial sea menor, sino en que la tasa de progreso de los salarios —a medida que transcurre la trayectoria laboral y ganan experiencia—, es mayor en el caso de los hombres, por lo que la diferencia salarial la modelan con una dummy de pendiente (véase la figura 7.8).

**Figura 7.8 La discriminación salarial de la mujer, modelación con variable dicotómica de pendiente**



De confirmarse esta última hipótesis, no tendría sentido el plan de acción afirmativa propuesto para la mujer sobre la base de garantizarle un salario inicial superior. Esta ventaja se borrará en algún momento de la trayectoria laboral y a partir de ese momento crecerá el diferencial a favor de los hombres en el transcurso de su vida activa (véase la figura 7.9).

**Figura 7.9 La discriminación salarial de la mujer, un plan de acción afirmativa inefectivo**





El núcleo del modelo de la tasa salarial que utilizamos para representar la situación descrita es  $Y_i = \beta_1 + \beta_2 X_i + u_i$ , con  $Y_i$  representando la tasa de salario y  $X_i$ , los años de experiencia laboral. Cada grupo de especialistas diseña su modelo específico en función del tipo de dummy:

- ✓ El primer grupo sostiene que la discriminación se produce en el nivel de salario inicial, por lo tanto, el modelo quedaría formalizado introduciendo una dummy de intercepto  $D_i$ , como  $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + u_i$ , con  $D_i = 1$  para los hombres. Así, la categoría de base son las mujeres.
- ✓ El segundo grupo afirma que la discriminación se produce en la tasa de progresión de los salarios. Así, el modelo se modifica incorporando una dummy de pendiente y el planteamiento es  $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i X_i + u_i$ , con  $D_i$  igualmente 1 para los hombres.

Sin dudas, es de la mayor importancia verificar cuál de las hipótesis de partida es válida, para poder diseñar un plan de acción afirmativa a favor de la mujer que sea efectivo. Por lo tanto, plantearemos un modelo que incorpore ambos tipos de dummy,

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 D_i X_i + u_i \quad (7.14)$$

y someteremos a tests de significación las diferencias de intercepto y las diferencias de pendiente (véase el recuadro), para encontrar la respuesta que buscamos.

#### Testeando efectos cualitativos

Modelo  $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 D_i X_i + u_i$ , con  $D_3$  como dummy de intercepto y  $D_4$  como dummy de pendiente.

##### 1. Test de las diferencias de intercepto

$$H_0 : \beta_3 \leq 0 \text{ vs. } H_1 : \beta_3 > 0, \text{ con } \frac{\hat{\beta}_3 - 0}{se(\hat{\beta}_3)} \sim t_{(n-4)}$$

##### 2. Test de las diferencias de pendiente

$$H_0 : \beta_4 \leq 0 \text{ vs. } H_1 : \beta_4 > 0, \text{ con } \frac{\hat{\beta}_4 - 0}{se(\hat{\beta}_4)} \sim t_{(n-4)}$$

Por último, las variables binarias también se han mostrado efectivas para la modelación de los cambios que ocurren en las variables por el efecto de las estaciones del año. Estas variaciones estacionales son de consideración en los casos de la oferta agrícola, de las visitas de turistas internacionales y

de las ventas de algunos artículos de consumo, entre los numerosos ejemplos posibles. Es decir, muchas series de tiempo en Economía exhiben patrones estacionales.

Para atender a estas diferencias estacionales se utilizan variables dicotómicas: si las observaciones disponibles son de frecuencia mensual, se introducen 11 dummies; 3, cuando las observaciones son trimestrales y una dummy para datos semestrales. Esta modelación constituye la base para aplicar métodos de remoción o eliminación de los componentes estacionales; es decir, métodos de desestacionalización o de ajuste estacional. En la Parte Cuatro del libro retomaremos el tema de la estacionalidad.

El tipo de dummy a incorporar para esta modelación —si de intercepto o de pendiente— es una cuestión empírica; es decir, dependerá del comportamiento de los datos.

Como habrán apreciado en la exposición de este tópico, las variables dicotómicas tienen disímiles empleos y son muy útiles para el trabajo de modelación. Sin embargo, su servicio no se reduce a los casos previamente descritos. En el próximo epígrafe examinaremos otras importantes aplicaciones de las variables binarias.

### 7.3 Cambio de Estructura en el MLNR

Cuando estamos trabajando con datos de series de tiempo, y sobre todo en el caso de muestras que abarquen largos períodos, debemos investigar exhaustivamente si los datos sostienen un único modelo de regresión, con parámetros constantes durante toda la etapa analizada. Este supuesto podría ser fuerte para tan largo período de tiempo como 20 años, por ejemplo.

Múltiples son los factores que pueden incidir en que los parámetros de un modelo varíen en el tiempo. Según Gujarati (2004, p. 263), acontecimientos de índole externa (como lo fue el derrumbe del otrora campo socialista, que se inició con la caída del muro de Berlín en 1989 y terminó con el desplome del sistema en la URSS en 1991) o interna, así como cambios introducidos en las políticas económicas de un país, pueden desencadenar procesos de cambio estructural en un MLNR.

Cambio estructural significa, en este contexto, que los valores de los parámetros del modelo de regresión —a saber, coeficientes de regresión y varianza del término de error— no permanecen constantes a lo largo del período o entre individuos.<sup>9</sup>

¿Cómo detectar la presencia de cambios estructurales en los coeficientes de un modelo de regresión?

---

<sup>9</sup> Véase Gujarati (2004, p. 263). Para el desarrollo de este tópico hemos seguido las explicaciones de este autor en el texto previamente citado pp. 263-269.

Supongamos que hemos postulado nuestro MLNR como

$$Y_t = \beta_1 + \beta_2 X_t + u_t, \text{ con } t = 1, \dots, T \quad (7.15)$$

Sin embargo, sospechamos que a partir de un cierto momento en el período de análisis, que denotaremos con  $\tau$ , puede haber ocurrido un cambio en los coeficientes de regresión (lo mismo en  $\beta_1$  que en  $\beta_2$ ) por acontecimientos de diversa índole. Es decir, es probable que tengamos dos modelos, uno para cada subperíodo  $t = 1, \dots, \tau$  y  $t = \tau + 1, \dots, T$ , en lugar de un único modelo para  $t = 1, \dots, T$ . Planteamos entonces los modelos correspondientes a cada intervalo.

$$Y_t = \alpha_1 + \alpha_2 X_t + u_{1t}, \text{ con } t = 1, \dots, \tau \quad (7.16)$$

$$Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}, \text{ con } t = \tau + 1, \dots, T \quad (7.17)$$

Para verificar si ha ocurrido un cambio estructural en  $\beta_1$  y  $\beta_2$  a partir de  $\tau$  se realiza la prueba de Chow. Este test se construye en base a las regresiones (7.15)-(7.17) y parte de las siguientes consideraciones:

1. Las varianzas de los términos de error de (7.16) y (7.17) son iguales; es decir,  $u_{1t} \sim N(0, \sigma^2)$  y  $u_{2t} \sim N(0, \sigma^2)$ .
2. Los términos de error  $u_{1t}$  y  $u_{2t}$  están independientemente distribuidos.

¿En qué consiste el procedimiento de esta prueba?

1. Estimamos la regresión (7.15), que sería el modelo válido de no haber inestabilidad en los coeficientes de regresión. De esta estimación recuperamos la *SRC* correspondiente, que denotaremos *SRRC*. ¿Por qué es esta la suma de residuos restringidos? Porque al estimar el modelo así especificado estamos imponiendo dos restricciones,  $\beta_1 = \alpha_1 = \gamma_1$  y  $\beta_2 = \alpha_2 = \gamma_2$ . Los gl son  $T - k$ , con  $T$  el tamaño de la muestra (la longitud del período en este caso) y  $k$  el número de coeficientes estimados, dos para este modelo.
2. Estimamos las regresiones (7.16) y (7.17), y recuperamos sus *SRC* que denotamos con  $SRC_1$  y  $SRC_2$ , con  $gl_1 = \tau - k$  y  $gl_2 = T - \tau - k$ , respectivamente.
3. Como hemos supuesto que los conjuntos de muestras que dan lugar a las regresiones (7.16) y (7.17) son independientes, podemos sumar  $SRC_1$  y  $SRC_2$  para obtener la suma de residuos al cuadrado del modelo sin restricciones  $SRC = SRC_1 + SRC_2$ . Los grados de libertad se obtienen también a través de la suma,  $gl = gl_1 + gl_2 = \tau - k + T - \tau - k = T - 2k$ .
4. Con estos insumos construimos la variable  $F = \frac{T - 2k}{k} \frac{(SRRC - SRC)}{SRC}$ .
5. La idea que subyace a la prueba de Chow es que, si no existe cambio estructural en los regresores ( $H_0: \beta_1 = \alpha_1 = \gamma_1$  y  $\beta_2 = \alpha_2 = \gamma_2$ ), entonces

$SRRC$  y  $SRC$  deberían ser estadísticamente iguales. Bajo esta hipótesis,  $F \sim F_{(k, T-2k)}$ .

6. Así, no se rechaza  $H_0$  si el valor  $F$  que calculamos en 4. fuese menor que su valor crítico  $F_{(k, T-2k)}^\alpha$ , o si la probabilidad asociada a  $F$  fuese mayor que el nivel de significación  $\alpha$ .

Como habrán advertido, la prueba de Chow no es más que un test de hipótesis simultáneas (en este caso  $\beta_1 = \alpha_1 = \gamma_1$  y  $\beta_2 = \alpha_2 = \gamma_2$ ) como los que examinamos en el tópico correspondiente del capítulo 6 de este libro.

Si rechazamos la  $H_0$  del test de Chow, diremos que el modelo de regresión (7.15) presenta un cambio estructural para  $t = \tau$ , y  $\tau$  es el punto de ruptura.

Recuerden que esta prueba de Chow tiene un supuesto implícito a satisfacer: debemos comprobar que, efectivamente, las varianzas de los errores de las regresiones (7.15) y (7.16) son las mismas; este aspecto lo discutiremos en el Capítulo 9 de este texto.

De constatarse la validez de un cambio estructural a partir del momento  $\tau$ , ¿cómo incluir esta consideración en el modelo de regresión?

Para ello introducimos una variable dummy o binaria, que denotaremos con  $D_\tau$ . Para que  $D_\tau$  nos permita modelar el cambio en los coeficientes de regresión a partir del momento  $\tau$ , la definimos de la siguiente manera:

$$D_\tau = \begin{cases} D_\tau = 0, & \text{para } t = 1, \dots, \tau - 1 \\ D_\tau = 1, & \text{para } t = \tau, \dots, T \end{cases}$$

El modelo original se modifica a partir de introducir la dummy, lo mismo como dummy de intercepto (si tenemos motivos para suponer que ese es el coeficiente que se modifica), que como dummy de pendiente, si es  $\beta_2$  quien debe variar a partir de  $\tau$ , o introducimos las dos posibilidades de conjunto, si puede haber ocurrido lo mismo un desplazamiento que una rotación en la función o ambos movimientos. El nuevo modelo quedará especificado como sigue:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 D_\tau + \beta_4 D_\tau X_t + u_t$$

donde  $\beta_3$  o  $\beta_4$ , o ambos, deben ser estadísticamente distintos de cero para cierto nivel de significación.

El cambio estructural no sólo está asociado al trabajo con series de tiempo. Hemos partido de este tipo de información para ejemplificar y que se comprenda mejor esta cuestión. Sin embargo, en el trabajo con series de corte transversal también pueden ocurrir cambios estructurales. En este caso el cambio se definiría

como la presencia de coeficientes de regresión diferentes para diferentes grupos de individuos (personas, o empresas o países). Los procedimientos de verificación de hipótesis son similares a los ya considerados.

Para cerrar este tópico, ilustraremos lo explicado a través de nuestro ejemplo de la función de importaciones para Cuba entre 1961 y 2005. En el capítulo 5 habíamos postulado una única función para el período bajo estudio. Sin embargo, a la luz de lo acontecido después del desmoronamiento del campo socialista a partir de 1989, este supuesto parece bastante restrictivo. Por eso, nos interesará investigar si el suceso del derrumbe del socialismo en Europa del Este tuvo un impacto en la especificación de nuestra función de demanda de importaciones. Realicemos entonces un test de Chow fijando un posible punto de ruptura en 1989; es decir, con  $\tau = 1989$ .

Del trabajo desarrollado con este modelo en los capítulos previos, ya tenemos una estimación del modelo restringido y, por tanto, la *SRRC*. Necesitamos estimar el modelo para los dos períodos 1961-1989 y 1990-2005, para conseguir sus *SRC*. Vamos a ordenar estos resultados empíricos en la tabla 7.1.

**Tabla 7.1: Resultados del test de Chow, función de demanda de importaciones cubanas y punto de ruptura en 1989**

| Modelo                                                                           | Período   | Resultados de la estimación                                   | Suma de residuos al cuadrado | Grados de libertad   |
|----------------------------------------------------------------------------------|-----------|---------------------------------------------------------------|------------------------------|----------------------|
| $\dot{m}_t = \beta_0 + \beta_1 \dot{y}_t + \beta_2 \dot{p}_t + \varepsilon_t$    | 1961-2005 | $\hat{\dot{m}}_t = 1,263 + 0,751 \dot{y}_t - 0,543 \dot{p}_t$ | $SRRC = 3764,3$              | $gl = T - k = 42$    |
| $\dot{m}_t = \alpha_0 + \alpha_1 \dot{y}_t + \alpha_2 \dot{p}_t + \varepsilon_t$ | 1961-1989 | $\hat{\dot{m}}_t = 4,369 + 0,045 \dot{y}_t - 0,409 \dot{p}_t$ | $SRC_1 = 1950,4$             | $gl_1 = 29 - 9 = 26$ |
| $\dot{m}_t = \gamma_0 + \gamma_1 \dot{y}_t + \gamma_2 \dot{p}_t + \varepsilon_t$ | 1990-2005 | $\hat{\dot{m}}_t = 1,795 + 1,471 \dot{y}_t - 0,584 \dot{p}_t$ | $SRC_2 = 826,8$              | $gl_2 = 16 - 3 = 13$ |
| $SRC = 2777,2$                                                                   | $k = 3$   | $gl_1 + gl_2 = T - 2k = 39$                                   | $F = 4,621$                  | $p = 0,007$          |

Fuente: Elaboración propia

Noten que la mera comparación de los coeficientes de regresión de los modelos estimados entre sí da fundamento para sostener la posibilidad de un cambio estructural en 1989. Por ejemplo, los interceptos y los coeficientes de regresión correspondientes a la tasa de crecimiento del producto de los modelos para 1961-1989 y 1990-2005 parecieran ser bastante diferentes. Esta sospecha la confirmamos definitivamente a partir del valor  $F$  calculado y la probabilidad a él asociada, bastante menor que  $\alpha = 1\%$ .

¿Qué hacer para reespecificar nuestra función de demanda de importaciones?

Construimos una variable binaria  $D_{89}$ , con  $D_{89} = 0$  en el período 1961-1989 y  $D_{89} = 1$  entre 1990-2005. A partir de esa dummy reespecificamos el modelo, permitiendo la posibilidad de cambios de estructura en los interceptos y las pendientes:

$$\dot{m}_t = \beta_0 + \beta_1 \dot{y}_t + \beta_2 \dot{p}_t + D_{89} + \beta_1 D_{89} \dot{y}_t + \beta_2 D_{89} \dot{p}_t + \varepsilon_t \quad (7.18)$$

Presentamos los resultados de la estimación del modelo (7.18), tal y como ya acostumbramos a hacer:

$$\begin{array}{rcccccc} \hat{m}_t & = & 4,369 & + & 0,045\dot{y}_t & - & 0,409\dot{p}_t & - & 2,573D_{89} & + & 1,425D_{89}\dot{y}_t & - & 0,175D_{89}\dot{p}_t \\ se & & (2,175) & & (0,294) & & (0,107) & & (3,123) & & (0,435) & & (0,170) \\ t & & (2,008) & & (0,155) & & (-3,831)^* & & (-0,824) & & (3,278)^* & & (-1,031) \end{array}$$

Noten que sólo se mantienen significativos en este modelo los regresores correspondientes a  $\dot{p}_t$  y a  $D_{89}\dot{y}_t$ . Esto nos estaría indicando que la elasticidad-precio de las importaciones puede considerarse invariable durante el período 1961-2005 (al menos la evidencia empírica no rechaza esta suposición) y que la tasa de crecimiento del producto interno bruto no constituía un factor explicativo del crecimiento de las importaciones en el período anterior a 1989, pero sí lo es después de esa fecha.

Con estos resultados confirmamos la importancia de la verificación de los supuestos que subyacen al MLNR antes de su evaluación. Por ejemplo, según la estimación inicial habíamos llegado a una elasticidad-producto para la demanda de importaciones de 0,751% en el período 1961-2005. Sin embargo, ahora sabemos que este estimador no es válido.

En el siguiente capítulo continuaremos aportándoles herramientas para desarrollar el diagnóstico del MLNR. En particular trataremos el problema de la multicolinealidad, que resulta bastante molesto en el trabajo empírico, sobre todo con series de tiempo.

# *preguntas*

Intente dar respuesta a las siguientes preguntas. Así sabrá si ha captado los principales mensajes de este capítulo.

1. ¿Qué significa la linealidad en el contexto de un MRLN?
2. De entre los siguientes modelos, diga cuáles cumplen la condición de linealidad en los parámetros. Para los que no cumplan la condición, indique si puede llegar a satisfacerla a través de transformaciones matemáticas.

a.  $Y_i = \alpha X_i^\beta Z_i^\gamma e^{\varepsilon_i}$

b.  $Y_i = \frac{1}{\alpha + \beta X_i + \varepsilon_i}$

c.  $Y_i = e^{\alpha + \beta X_i + \gamma Z_i + \varepsilon_i}$

d.  $C_i = \alpha_0 + \alpha_1 Y_i + \alpha_2 Y_i^2 + \alpha_3 Y_i^3 + \varepsilon_i$

e.  $\ln Y_i = \alpha + \beta \frac{1}{X_i} + \gamma \frac{1}{Z_i} + \varepsilon_i$

f.  $e^{Y_i} = \alpha X_i^\beta Z_i^\gamma e^{\varepsilon_i}$

g.  $Y_i = \frac{1}{1 + e^{\alpha + \beta X_i + \gamma Z_i + \varepsilon_i}}$

h.  $Z_i = \alpha_0 + \alpha_1 W_i + \alpha_2^2 W_i + \alpha_3^3 W + \varepsilon_i$

3. ¿Qué es una variable dicotómica, binaria, muda o dummy? ¿Qué utilidad tiene para la modelación econométrica?
4. ¿Qué reglas debemos seguir para evitar la “trampa de la muda”?
5. ¿Cuáles tipos de variables mudas podemos incorporar a la modelación? ¿Cómo interpretar los coeficientes de regresión en cada caso?
6. ¿Por qué son necesarias las variables de control en un modelo econométrico?
7. ¿Qué es un test de Chow y cuál es su utilidad?
8. ¿Cómo modelar un cambio en la estructura de coeficientes de un modelo econométrico utilizando variables dicotómicas?



# ejercicios

- 7.1 Después de recolectar datos sobre 2929 profesores de diferentes universidades, pudimos estimar el siguiente modelo (errores estándar entre paréntesis):

$$\hat{S}_i = 36721 + \frac{817}{(259)} M_i + \frac{426}{(456)} A_i + \frac{406}{(24)} R_i + \frac{3539}{(458)} T_i$$

$$\bar{R}^2 = 0.77$$

donde

---

$S_i$  = el salario del  $i$ -ésimo profesor universitario,

$M_i$  = una variable muda igual a 1 si el  $i$ -ésimo profesor es hombre y 0 en otro caso,

$A_i$  = una variable muda igual a 1 si el  $i$ -ésimo profesor es inmigrante y 0 en otro caso,

$R_i$  = años de experiencia del  $i$ -ésimo profesor,

$T_i$  = una variable muda igual a 1 si el  $i$ -ésimo profesor enseña economía y 0 en otro caso.

---

Responda las siguientes preguntas, dando argumentos estadísticos para sustentar sus opiniones:

- ¿Ganan más los profesores que las profesoras?
  - ¿Es mayor el salario de los profesores inmigrantes?
  - ¿Es importante la experiencia para el salario de un profesor?
  - ¿Es Economía una mejor materia (al menos en términos de salario)?
- 7.2 En un artículo de 1986, Frederick Schut y Peter VanBergeijk discutían acerca de la posibilidad de discriminación de precios en la industria farmacéutica. Ellos suponían que mientras mayor fuera la capacidad de pago de un país, medida en términos de su PIB per cápita, menor sería (en valor absoluto) la elasticidad-precio de la demanda de fármacos y mayor el precio que un discriminador de precios podría cargar o cobrar. Así, ellos esperaban un coeficiente positivo al regresar el nivel de precios de los fármacos con respecto al PIB per cápita. Adicionalmente, ellos esperaban que los precios fueran mayores si se permitieran patentes farmacéuticas y menores si existieran controles de precios, si se promoviera la competencia o si la industria farmacéutica del país fuera relativamente grande. Ellos estimaron el siguiente modelo para una muestra de 32 países:

$$\hat{P}_i = 38.22 + 1.43GDPN_i - 0.6CVN_i + 7.31PP_i - 15.63DPC_i - 11.38IPC_i,$$

(s)                      (0.21)                      (0.22)                      (6.12)                      (6.93)                      (7.16)

$$n = 32 \quad \bar{R}^2 = 0.775 \quad F = 22.35,$$

donde



|         |                                                                                                                             |
|---------|-----------------------------------------------------------------------------------------------------------------------------|
| $P_i$   | El nivel de precios de los fármacos en el $i$ -ésimo país                                                                   |
| $GDN_i$ | El PIB per cápita en el $i$ -ésimo país.                                                                                    |
| $CVN_i$ | El volumen de consumo per cápita de productos farmacéuticos en el $i$ -ésimo país.                                          |
| $PP_i$  | Una variable muda igual a 1 si se reconocen patentes en el $i$ -ésimo país e igual a cero en otro caso                      |
| $DPC_i$ | Una variable muda igual a 1 si se aplican controles estrictos de precios en el $i$ -ésimo país e igual a cero en otro caso. |
| $IPC_i$ | Una variable muda igual a 1 si el $i$ -ésimo país promueve la competencia de precios e igual a cero en otro caso.           |

Todas las variables relevantes están expresadas como porcentajes del valor que corresponde a los Estados Unidos.

- Discuta la significación de los estimados de los coeficientes de pendiente al 5 por ciento y 1 por ciento de nivel de significación.
- Discuta la significación global de este modelo a los mismos niveles de significación anteriores.
- ¿Piensa Ud. que estos resultados apoyan las hipótesis discutidas por Schut y VanBergeik? ¿Por qué?
- ¿Piensa Ud. que el efecto de los controles de precios es más fuerte en valor absoluto que el efecto de políticas de fomento de la competencia en precios? Evalúe sus hipótesis. (Suponga que la covarianza entre los estimados de los parámetros es igual a 3.17).

7.3 El siguiente modelo fue estimado para una muestra de 9746 individuos:

$$y_i = 2.485 + 0.0875 EDUC_i + 0.035 AGE_i - 0.00085 AGE_i^2$$

(t)      (6.583)   (2.875)                      (3.458)                      (-2.015)

donde  $y_i$  representa el logaritmo del salario del individuo  $i$ ,  $EDUC_i$  es el número de años que asistió a la escuela,  $AGE_i$  es la edad,  $AGE_i^2$  es la edad al cuadrado y las razones- $t$  se han colocado entre paréntesis. También sabemos que  $STC=7486.24$  y  $SRC=5372.48$ .

El mismo modelo fue estimado para los 4923 hombres incluidos en la muestra como:

$$y_i = 2.782 + 0.0905 EDUC_i + 0.039 AGE_i - 0.00023 AGE_i^2$$

(t)      (5.438)   (2.543)                      (2.985)                      (-0.875)

con  $STC=3028.45$  y  $SCE= 1104.23$ .

Para las mujeres incluidas en la muestra obtuvimos:

$$y_i = 2.943 + 0.0798 EDUC_i + 0.031 AGE_i - 0.00071 AGE_i^2$$

(t)      (3.857)   (2.632)                      (2.045)                      (-1.985)

con  $SRC=1800.68$  y  $SCE=945.23$ .

- Calcule el  $R^2$  ajustado para cada uno de estos tres modelos.
- Estime la varianza de cada modelo.
- ¿Es razonable la hipótesis de no-linealidad del efecto de la edad sobre y para cada uno de estos modelos? ¿Por qué?
- ¿Tenemos suficiente evidencia empírica para rechazar la hipótesis de que ambos, hombres y mujeres, tengan los mismos parámetros? ¿Por qué?

- 7.4 El crecimiento del trigo en áreas de baja precipitación depende críticamente de la lluvia caída durante cada uno de las tres fases: germinación, crecimiento y florecimiento. La precipitación en el año  $t$  durante cada uno de estas fases vegetativas se denota como  $GERM_t$ ,  $GROW_t$  y  $F_t$ , respectivamente. Estas variables de precipitación se expresan como proporciones relativas al promedio de lluvia caída para cada fase en el período muestral. Por ejemplo,  $GROW_t < 1$  significa que la lluvia caída en el año  $t$  durante la fase de crecimiento fue menor que la precipitación promedio para esa fase en la muestra. Nuestra muestra incluye 39 observaciones anuales. Así,

$$\sum_{t=1}^{39} GROW_t / 39 = 1$$

El siguiente modelo fue estimado para nuestra muestra de 39 observaciones anuales:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 GERM_t + \beta_3 GERM_t^2 + \beta_4 GROW_t + \beta_5 GROW_t^2 + \beta_6 F_t + \beta_7 F_t^2 + \varepsilon_t$$

donde  $Y_t$  es el rendimiento en toneladas por hectárea y  $t$  es una tendencia temporal igual a  $1, 2, \dots, 39$ , utilizada como proxy para el cambio tecnológico. La estimación MCO de esta ecuación, con los errores estándar entre paréntesis es:

$$\begin{aligned} Y_t = & -0.689 + 0.0218t + 1.850 GERM_t - 0.824 GERM_t^2 \\ & \quad (0.503) \quad (0.0038) \quad (0.616) \quad (0.290) \\ & + 0.703 GROW_t - 0.339 GROW_t^2 + 0.064 F_t - 0.079 F_t^2 \\ & \quad (0.610) \quad (0.273) \quad (0.442) \quad (0.203) \end{aligned}$$

La siguiente es una tabla de las covarianzas entre los estimados de los coeficientes que afectan a las variables de primer y segundo grados.

| Variables          | Covarianzas  |
|--------------------|--------------|
| $GERM_t, GERM_t^2$ | -0.175099403 |
| $GROW_t, GROW_t^2$ | -0.163901387 |
| $F_t, F_t^2$       | -0.087896733 |

- Discuta la significación de los estimados de este modelo.

- b. Note que se ha asumido una relación no-lineal para la mayoría de las variables. ¿Es este un supuesto razonable? ¿Por qué?
  - c. La variable  $t$  es un proxy para el cambio tecnológico. Ofrezca una interpretación para el estimado del parámetro que afecta a esta variable.
  - d. Un resultado importante de este modelo son los efectos marginales de  $GERM$ ,  $GROW$ , y  $F$ . Estime esos efectos marginales y el promedio del período.
  - e. Esperamos que los efectos marginales de las variables  $GERM$ ,  $GROW$ , y  $F$  sean positivos. Evalúe esta hipótesis.
- 7.5 La tabla B.5 del CD acompañante contiene información sobre la demanda de energía en países de la OCDE para el período 1960-1982. Las variables son:  $Y$  = demanda final,  $X_2$  = PIB real,  $X_3$  = Precio real de la energía. Estime el siguiente modelo lineal:

$$Y_t = B_1 + B_2 X_{2t} + B_3 X_{3t} + u_t$$

- a. Discuta los coeficientes de regresión. Calcule los errores estándar usando la fórmula. Calcule el  $R^2$  ajustado, discuta los resultados comparando con el  $R^2$ .
- b. Calcule los valores del estadígrafo  $t$ . Encuentre el valor- $p$  usando la función de Excel para la distribución  $t$ -Student.
- c. Comente la significación global de la regresión ( $F$ -test).
- d. Calcule las elasticidades precio e ingreso a los valores medios de las variables.
- e. Estime el modelo log-lineal

$$\ln(Y_t) = B_1 + B_2 X_{2t} + B_3 X_{3t} + u_t$$

Repita los comentarios anteriores.

- a. Estime las elasticidades en d) y comente.
- b. Compare los  $R^2$  y comente si es un criterio válido para elegir un modelo.
- c. Realice el test de normalidad de Jarque-Bera para ambas estimaciones.

- 7.6 La tabla B.6 del disco acompañante contiene información sobre las características de 1483 individuos. Esta información fue tomada de una encuesta de hogares sueca. La siguiente tabla presenta las variables incluidas en esa base de datos.

| Variable    | Significado                                     |
|-------------|-------------------------------------------------|
| <i>S</i>    | Sexo: 1=Hombre; 0=Mujer                         |
| <i>Edad</i> | Edad en años                                    |
| <i>ED</i>   | Educación en años                               |
| <i>w</i>    | Salario por años                                |
| <i>St</i>   | Estado civil: 1=Casado o conviviente; 2=Soltero |
| <i>EX</i>   | Experiencia de trabajo en años                  |

- a. La especificación más frecuente para analizar la formación de salarios es la ecuación de Mincer:

$$\ln w_i = b_0 + b_1 ED_i + b_2 EX_i + b_3 (EX_i)^2 + \varepsilon_i$$

Note la especificación no lineal para *EX* (forma de U invertida). Un enfoque teórico sostiene que lo relevante para la determinación del salario es el número de años de educación. Un enfoque alternativo sostiene que lo relevante son los niveles educacionales alcanzados. Este enfoque puede ser analizado usando variables mudas para cada nivel: primario (hasta 9 años), secundario (10-12 años) y universitario o post secundario (más de 12 años). Note que la muestra incluye individuos mayores de 65 años, los que estarían retirados. Hay que eliminar a estos individuos.

- b. Estime la tasa de retorno de la educación usando la especificación anterior.
- c. Explore la posibilidad de una relación no lineal con la educación, usando la siguiente especificación:

$$\ln w_i = b_0 + b_1 ED_i + b_2 (ED_i)^2 + b_3 EX_i + b_4 (EX_i)^2 + \varepsilon_i$$

- d. Reconsidere la primera especificación incluyendo variables mudas por nivel de educación en lugar de años de educación. Es decir, estimar:

$$\ln w_i = b_0 + b_1 D_{2i} + b_2 D_{3i} + b_3 EX_i + b_4 (EX_i)^2 + \varepsilon_i$$

Las variables mudas  $D_2$  y  $D_3$  toman los valores 1 para los niveles de educación cumplidos secundario y universitario, respectivamente. ¿Por qué no incluimos una variable muda para el nivel primario terminado?

- e. Calcule el retorno de la educación como el porcentaje de cambio de  $w$  (y no de  $\ln w$ ) cuando una variable muda pasa de 0 a 1. Usar la fórmula

$$retorno = \frac{E[w | D_i = 1] - E[w | D_i = 0]}{E[w | D_i = 0]}$$

- f. ¿Difieren los retornos de la educación para hombres y mujeres? Use la siguiente especificación:

$$\ln w_i = b_0 + b_1 D_{2i} + b_2 D_{3i} + b_3 S_i \times D_{2i} + b_4 S_i \times D_{3i} + b_5 EX_i + b_6 (EX_i)^2 + \varepsilon_i$$

- g. Agregue una variable reflejando el sexo ( $S_i$ ), como en la especificación siguiente. ¿Porqué los rendimientos cambian tanto?

$$\ln w_i = b_0 + b_1 D_{2i} + b_2 D_{3i} + b_3 S_i \times D_{2i} + b_4 S_i \times D_{3i} + b_5 S_i + b_6 EX_i + b_7 (EX_i)^2 + \varepsilon_i$$

Interprete los resultados.

CAPÍTULO 8



## MULTICOLINEALIDAD

En este capítulo vamos a abordar un supuesto básico del Modelo Lineal Normal de Regresión (MLNR) que consiste en que no puede existir una relación lineal exacta entre las variables explicativas. Se trata del supuesto de no-colinealidad, también llamado de no-multicolinealidad para el caso de modelos con más de dos variables explicativas y donde exista más de una relación lineal entre las mismas. Cuando se presenta una relación de tal naturaleza entre los regresores, decimos que estamos en presencia de colinealidad o multicolinealidad.

Es un supuesto que tiene que ver con la construcción o diseño mismos del MLNR. Como ya explicamos en el capítulo 5, este modelo supone que cada una de las variables independientes aporta una parte específica de la explicación a la variación de  $Y$ . Si existen relaciones lineales entre las variables, ellas prácticamente se “solapan” y estaríamos ante una situación de información redundante. Este es un fenómeno que aparece comúnmente en la práctica económica y puede ser bastante dañino para la estimación certera de los coeficientes de regresión, como podrán comprobar en el primer epígrafe de este capítulo.

Una relación lineal entre las variables independientes la describimos a partir del siguiente planteamiento. Consideremos las  $k$  variables explicativas, posiblemente incluyendo la columna de unos para el intercepto, y una muestra de tamaño  $n$ . Agrupemos las  $n$  observaciones de estas variables en los vectores  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k$ , que forman las columnas de la matriz  $\mathbf{X}$  de dimensión  $n \times k$ . Supongamos que existe un vector de coeficientes  $\mathbf{c} = (c_1, c_2, \dots, c_k)$ , donde al menos uno de ellos es distinto de cero y tal que

$$\mathbf{X}\mathbf{c} = \sum_{j=1}^k \mathbf{X}_j c_j = \mathbf{0} \quad (8.1)$$

entonces diremos que los vectores  $\mathbf{X}_j$  son linealmente dependientes. Supongamos que el  $r$ -ésimo coeficiente es diferente de cero ( $c_r \neq 0$ ). Entonces podemos escribir:

$$\mathbf{X}_r = \sum_{j \neq r} \frac{c_j}{c_r} \mathbf{X}_j \quad (8.2)$$

Es decir, hay al menos una variable (posiblemente más de una) que es una combinación lineal de las otras variables. Es posible demostrar que en este caso la matriz  $\mathbf{X}'\mathbf{X}$  es singular y, en consecuencia  $(\mathbf{X}'\mathbf{X})^{-1}$  no existe y no podemos encontrar  $\hat{\beta}$  de acuerdo a la fórmula (5.10). Este es el caso de multicolinealidad perfecta y queda excluido por el supuesto C.

Sin embargo, la expresión en (8.2) puede ser aproximada, aunque no exacta. En ese caso el determinante de  $\mathbf{X}'\mathbf{X}$  estará muy próximo a cero. Entonces resulta difícil calcular  $(\mathbf{X}'\mathbf{X})^{-1}$  con precisión y habrá problemas para determinar  $\hat{\beta}$ . Este es el caso de multicolinealidad parcial o aproximada.

Entre los factores que condicionan la multicolinealidad podemos mencionar:

- ✓ El método empleado para la recolección de la información. Con frecuencia sucede que sólo tenemos acceso a muestras en un intervalo restringido de entre los posibles valores que pueden tomar las variables explicativas en la población. Para ilustrar más podemos decir que, para estimar la función de demanda de importaciones que presentamos en el capítulo 5 sólo contamos con información para el período de 1960-2005 en el que el PIB cubano se movió entre 8,8 y 39,2 miles de millones de pesos. Si dispusiéramos de datos para años anteriores a 1960, muy probablemente tendríamos una mayor variabilidad del PIB, pero no tenemos forma de recuperar esas observaciones. Durante los años a los que se restringe nuestra muestra, pudiera ocurrir que casualmente exista una relación aproximadamente lineal entre el PIB real y los precios relativos de las importaciones, que afectaría la estimación que pretendemos de las elasticidades respectivas.
- ✓ Las restricciones que afectan al modelo o a la población objeto de estudio. Muchas veces sucede que ciertas combinaciones de variables independientes no aparecen en una determinada muestra y sólo ocurren combinaciones que llevan implícita una relación lineal entre ellas. Por ejemplo, en los datos utilizados para estimar la función de producción promedio —al menos hasta mediados de los 80— por lo general sólo encontramos combinaciones de los factores donde todos tienen una magnitud alta o baja de forma simultánea. No se observan combinaciones con alto capital físico y poca fuerza de trabajo. Así, prácticamente todos los factores se mueven al unísono y con seguridad tendremos un problema de multicolinealidad.
- ✓ La especificación del modelo. A veces se incorporan términos polinomiales a un MLNR y esto nos trae problemas de multicolinealidad —si bien no pre-



cisamente perfecta—, sobre todo cuando la variable en cuestión se mueve en un intervalo reducido. También en ocasiones cometemos errores al construir el modelo. Por ejemplo, incluimos una constante al unísono con un tipo de cambio que en determinado período se manejó como fijo. En ese caso tendremos un conflicto de colinealidad entre la constante y el tipo de cambio. O cuando incluimos simultáneamente la constante y variables dummies para todas las categorías de una variable cualitativa, como ya explicamos en un capítulo previo. O cuando no atendemos convenientemente las identidades o definiciones macroeconómicas en los modelos e incorporamos variables redundantes, por ejemplo, cuando introducimos en un modelo el saldo del comercio exterior de bienes y al mismo tiempo las importaciones, estamos incorporando información repetida, ya que el saldo se construye precisamente a partir de las importaciones.

- ✓ La sobredeterminación del modelo. Esto ocurre cuando la cantidad de variables explicativas en el modelo supera al número de observaciones; es decir, tenemos demasiadas variables y pocas observaciones. Así,  $k > n$ .
- ✓ La presencia de una tendencia común en las variables explicativas del modelo, cuando trabajamos con datos de series de tiempo. Esta es igualmente una situación bastante frecuente. Por ejemplo, volviendo a los datos de la función de producción, tenemos que todos los factores crecieron entre 1960 y mediados de la década del 80. Si lo hicieron aproximadamente a la misma velocidad, este será un factor que introduce el problema de multicolinealidad.

En este capítulo abordaremos cuáles son las principales consecuencias del fenómeno de la multicolinealidad para el análisis de regresión, cuáles son los procedimientos fundamentales para detectar este problema y, finalmente, veremos si es posible tomar alguna medida para circunvalar este problema y cuáles son las sugerencias básicas al respecto. Demostraremos que ambas situaciones causan múltiples trastornos y hacen impracticable o infructuoso el análisis de regresión.

## 8.1 Consecuencias

Podemos asociar la colinealidad o multicolinealidad con el significado de los coeficientes de pendiente en un MLNR. Sabemos que esos coeficientes nos informan sobre la variación que ocurre en la variable dependiente  $Y$  cuando se incrementa en una unidad la variable independiente correspondiente, digamos  $X_j$ , manteniendo constantes el resto de las variables explicativas. Pero justamente esto último es imposible en el caso de la existencia de colinealidad o multicolinealidad. Por (8.2) sabemos que no es posible incrementar  $X_j$  sin que varíe alguna de las variables de la derecha de esa expresión (aquellas para las que los coeficientes  $c_m/c_j$ , con  $m = 1, 2, \dots, j-1, j+1, \dots, k$ , sean distintos de cero).

El MLNR podría ser estimado en caso de que lo transformemos. Supongamos que sólo existe la relación lineal que expresamos en (8.2), podemos utilizar esa expresión

para sustituir en el modelo original a la variable  $X_j$  y obtendríamos un modelo con sólo  $k-1$  variables y donde los coeficientes de las mismas se expresan como:

$$\left( \beta_m - \beta_j \frac{c_m}{c_j} \right), \text{ con } m = 1, 2, \dots, j-1, j+1, \dots, k$$

Es decir, no podemos obtener las influencias individuales de cada variable explicativa, sino conjunta con otras variables con las que está relacionada. Por otra parte, no tenemos una solución única para cada  $\beta$ .

Asimismo, sabemos que la varianza de los coeficientes de regresión estimados es  $\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ . Siendo la matriz  $(\mathbf{X}'\mathbf{X})$  una matriz singular, tendremos que esas varianzas serán infinitas.

En resumen, en condiciones de multicolinealidad perfecta no podremos hacer inferencia estadística. Sin embargo, la situación más complicada se presenta en el caso de que tengamos multicolinealidad aproximada, porque no llegamos a una situación tan drástica como la anterior en que no podemos estimar el modelo. Con multicolinealidad aproximada sí obtendremos una solución para  $\hat{\beta}$ , pero ¿serán estos resultados adecuados?

Mientras la colinealidad no sea perfecta, los EMCO se mantendrán MELI. Sin embargo, debemos acotar que:

- ✓ La insesgadez es una propiedad en muestras repetidas. No se puede decir lo mismo acerca de las propiedades de los estimados para una muestra específica, que puede ser justamente la muestra disponible para la que se presenta la situación de multicolinealidad.
- ✓ El hecho de que los estimadores posean varianza mínima no significa que el valor numérico de la varianza será pequeño. Como demostraremos será el caso en presencia de esta situación.
- ✓ La multicolinealidad es esencialmente un fenómeno de la muestra disponible, no necesariamente del modelo.<sup>1</sup> En este caso no podremos aislar el efecto individual de las variables explicativas sobre la variable dependiente.

Para esclarecer acerca de las consecuencias de la colinealidad o multicolinealidad, vamos a examinar cómo se refleja la relación entre variables explicativas en la estimación de los coeficientes de regresión y de la varianza del modelo. Para facilitar este análisis, trabajaremos con el MLNR en desviaciones de media que presentamos al finalizar el capítulo 5, ecuación (5.19). Con ese mismo propósito consideraremos el caso de sólo dos variables explicativas,  $\mathbf{X}_2$  y  $\mathbf{X}_3$ . El EMCO de  $\beta$  para el modelo en desviaciones de media lo obtuvimos anteriormente en (5.21)

<sup>1</sup> A no ser que hayamos cometido algún error como los anteriormente referidos en el diseño del modelo, por lo que debemos ser exhaustivos en el estudio de la teoría que subyace al fenómeno objeto de estudio, y también cuidadosos de no cometer deslices como los que explicamos antes.

como  $\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$ .

En el caso de sólo dos variables explicativas, la matriz  $\mathbf{x}$  contiene dos vectores columna con los datos correspondientes a  $\mathbf{X}_2$  y  $\mathbf{X}_3$  en desviaciones de media; es decir,  $\mathbf{x}_2 = \mathbf{X}_2 - \bar{X}_2$  y  $\mathbf{x}_3 = \mathbf{X}_3 - \bar{X}_3$ . De modo que  $\mathbf{x} = (\mathbf{x}_2 \ \mathbf{x}_3)$ . ¿Cómo será  $(\mathbf{x}'\mathbf{x})$ ?

$$(\mathbf{x}'\mathbf{x}) = \begin{bmatrix} \mathbf{x}_2' \\ \mathbf{x}_3' \end{bmatrix} \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_2'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_3 \\ \mathbf{x}_3'\mathbf{x}_2 & \mathbf{x}_3'\mathbf{x}_3 \end{bmatrix}$$

¿Y cómo será  $(\mathbf{x}'\mathbf{x})^{-1}$ ? Aplicando los conocimientos de Álgebra Matricial, podrán verificar que tiene la siguiente forma:

$$(\mathbf{x}'\mathbf{x})^{-1} = \frac{\begin{bmatrix} \mathbf{x}_3'\mathbf{x}_3 & -\mathbf{x}_2'\mathbf{x}_3 \\ -\mathbf{x}_3'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_2 \end{bmatrix}}{\begin{vmatrix} \mathbf{x}_2'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_3 \\ \mathbf{x}_3'\mathbf{x}_2 & \mathbf{x}_3'\mathbf{x}_3 \end{vmatrix}}} = \frac{1}{\mathbf{x}_2'\mathbf{x}_2\mathbf{x}_3'\mathbf{x}_3 - (\mathbf{x}_2'\mathbf{x}_3)^2} \begin{bmatrix} \mathbf{x}_3'\mathbf{x}_3 & -\mathbf{x}_2'\mathbf{x}_3 \\ -\mathbf{x}_3'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_2 \end{bmatrix}$$

Sabemos por (5.14) que la matriz de varianzas se define como  $\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ . Por ejemplo, en el caso del coeficiente de regresión  $\beta_2$ , el estimador de su varianza lo obtenemos como:

$$\text{var}(\beta_2) = \frac{\sigma^2 \mathbf{x}_3' \mathbf{x}_3}{\mathbf{x}_2' \mathbf{x}_2 \mathbf{x}_3' \mathbf{x}_3 - (\mathbf{x}_2' \mathbf{x}_3)^2}$$

Utilizamos nuestros conocimientos de álgebra para transformar convenientemente la expresión anterior y obtenemos una equivalente, que nos permite entender cuál sería la influencia de la relación lineal entre  $\mathbf{X}_2$  y  $\mathbf{X}_3$ .

$$\text{var}(\beta_2) = \frac{\sigma^2 \mathbf{x}_3' \mathbf{x}_3}{\mathbf{x}_2' \mathbf{x}_2 \mathbf{x}_3' \mathbf{x}_3 \left[ 1 - \frac{(\mathbf{x}_2' \mathbf{x}_3)^2}{\mathbf{x}_2' \mathbf{x}_2 \mathbf{x}_3' \mathbf{x}_3} \right]} = \frac{\sigma^2 \mathbf{x}_3' \mathbf{x}_3}{\mathbf{x}_2' \mathbf{x}_2 \mathbf{x}_3' \mathbf{x}_3 [1 - r_{23}^2]} = \frac{\sigma^2}{\mathbf{x}_2' \mathbf{x}_2 [1 - r_{23}^2]}$$

Noten que en el denominador de la varianza del estimador del coeficiente de regresión  $\beta_2$  aparece el coeficiente de correlación entre las variables  $\mathbf{X}_2$  y  $\mathbf{X}_3$ , que denotamos con  $r_{23}^2$ . Sabemos que  $r_{23}^2$  se mueve entre cero y uno. ¿Qué ocurre cuando  $r_{23}^2$  sea igual a uno o se acerque a uno?

En el primer caso, estamos en presencia de colinealidad perfecta. Adviertan cómo podemos corroborar las consecuencias antes descritas. El segundo caso es

precisamente el que nos interesa, el de colinealidad aproximada o menos que perfecta. En esta situación el denominador de  $var(\hat{\beta}_2)$  tiende a cero, por lo que la varianza será muy grande (mayor, mientras más se acerque a uno el coeficiente de correlación). ¿Cuáles son las repercusiones de esa gran varianza?

- ✓ Si la varianza es grande, la desviación estándar de los EMCO será también grande, y no sólo para  $\hat{\beta}_2$ , sino también para  $\hat{\beta}_3$ . Esto constituye una dificultad mayor, pues no podremos estimar con precisión nuestros parámetros de interés.
- ✓ Como consecuencia de los grandes errores estándar, tendremos asimismo amplios intervalos de confianza para los estimados.
- ✓ En asociación a los grandes errores estándar tendremos también razones-t muy pequeñas. Por lo tanto, no lograremos rechazar que los coeficientes de regresión sean iguales a cero. Es decir, las razones-t sugerirán erróneamente baja significación de las variables incluidas en la especificación del modelo.
- ✓ Sin embargo, junto a estas poco significativas razones-t encontraremos un alto  $R^2$  y un estadístico F que nos permitirá rechazar la hipótesis nula del test de significación global (que todos los coeficientes de regresión sean simultáneamente iguales a cero), lo cual es contradictorio con el aspecto anterior.
- ✓ Los estimadores MCO y sus errores estándar se vuelven muy sensibles a pequeños cambios en los datos; es decir, tienden a volverse inestables o volátiles.
- ✓ Se presentarán signos incorrectos (en relación con los postulados teóricos) en los coeficientes de regresión, ya que tendremos coeficientes estimados cercanos a cero (lo mismo por la derecha que por la izquierda).
- ✓ Y encontraremos dificultades para determinar las contribuciones individuales de las variables explicativas a la suma de cuadrados explicada (SCE) o  $R^2$ .

A la fracción

$$\frac{1}{[1 - r_{23}^2]} = FIV \quad (8.3)$$

la denominamos factor de inflación de varianza, y la denotamos con  $FIV$ . Fíjense que mientras más se acerque  $r_{23}^2$  a uno, mayor será  $FIV$  y la varianza del coeficiente de regresión.

Como habrán podido apreciar, la presencia de alta colinealidad entre variables explicativas de un MRLN constituye una dificultad mayor. Por lo tanto, cada vez que apliquemos un MLNR para resolver algún problema, tendremos que verificar si se dan relaciones de ese tipo entre las variables explicativas de la especificación

que estamos utilizando. En el epígrafe siguiente desarrollaremos cómo detectar la colinealidad o multicolinealidad y a evaluar su grado.

## 8.2 Indicadores para Detectarla

Para la detección de la multicolinealidad no existen pruebas o tests estadísticos, sino una serie de indicadores, que se derivan de la propia discusión acerca de qué es la multicolinealidad y sus consecuencias.

En el abordaje de estos temas vimos que en este fenómeno su grado (intensidad o magnitud) es muy importante: si es perfecta (coeficiente de correlación lineal igual a uno), no podremos estimar el MLNR; en la medida que el coeficiente de correlación lineal se aleja de uno, tendremos menos consecuencias adversas. Lo ideal es que ese coeficiente sea cero, pero ya explicamos que en el caso de las variables con que usualmente trabajamos en Economía esto es bastante difícil. Así, nos interesa no sólo si existe correlación entre las variables explicativas, sino también la magnitud de esa correlación para saber si los datos disponibles serán útiles para la estimación del MLNR.

Un indicador que nos da información acerca de la presencia de multicolinealidad en nuestros datos y su magnitud son los coeficientes de correlación lineal entre las variables explicativas del MLNR. Estos coeficientes constituyen resultados estándar de cualquier paquete estadístico, especialmente econométrico, y pueden ser calculados también en una hoja de Excel. Basta con pedir en la rutina de Análisis de Datos del menú Herramientas, la opción de Coeficiente de Correlación, para obtener la matriz de correlación entre los datos que indiquemos.

Esta matriz es una matriz con unos ocupando la diagonal principal<sup>2</sup> y simétrica con respecto a ella, por lo que los programas por lo general sólo devuelven como respuesta la parte de la matriz por debajo de la diagonal. Si en el rango de entrada de datos para el análisis de correlación incluimos en primer lugar a la variable dependiente, lo deseable es que los coeficientes de correlación de la primera columna<sup>3</sup> sean altos y que el resto de los coeficientes de la matriz sean cercanos a cero.

Como ya explicamos, en el caso de series de tiempo es usual la presencia de colinealidad por la existencia de tendencias comunes en los datos de la muestra. Cuando trabajamos con datos de esta naturaleza, se puede considerar alto un coeficiente de correlación cuando rebasa el 80% (o el nivel 0,8). En el caso de datos de corte transversal, se puede considerar alto un coeficiente de correlación que sobrepase el 60%.

<sup>2</sup> La diagonal contiene los coeficientes de correlación de cada variable con respecto a sí misma, que evidentemente son igual a uno.

<sup>3</sup> Son los coeficientes de correlación entre la variable dependiente y las explicativas o independientes.

Así, antes de estimar un MLNR, debemos pedir junto con las estadísticas básicas de las mismas, la matriz de correlación.

Como ejemplo, usemos la muestra disponible en la Tabla B.2 del CD acompañante, para la producción y sus factores que utilizamos para la estimación de la función de producción macroeconómica en Cuba correspondiente al período de 1980 al 2000, la matriz de correlación que obtuvimos con Excel la mostramos en la tabla 8.1.

**Tabla 8.1: Matriz de correlaciones entre variables de la función de producción macroeconómica para Cuba, 1980-2000**

|    | LQ     | LK     | LL    | LH |
|----|--------|--------|-------|----|
| LQ | 1      |        |       |    |
| LK | 0,472  | 1      |       |    |
| LL | 0,209  | 0,633  | 1     |    |
| LH | -0,354 | -0,015 | 0,714 | 1  |

*Fuente: Elaboración propia*

Aparentemente, los coeficientes de correlación simple entre los logaritmos del capital físico y de la fuerza de trabajo, así como entre los logaritmos de ésta última y el capital humano, no son tan altos —dado que se trata de datos de corte longitudinal—, por lo que no deberíamos preocuparnos de que esté presente un alto grado de correlación. Seguidamente veremos que esta apreciación puede ser errónea.

Los coeficientes de correlación entre parejas de regresoras pueden ser bajos y aun así puede existir alta colinealidad entre grupos de más de dos variables explicativas. Por lo tanto, no nos podemos conformar con esta indagación para dictaminar si tenemos un fenómeno grave de colinealidad o multicolinealidad. Se trata de una condición suficiente, pero no necesaria para la presencia de alta colinealidad.

Otro índice de la presencia de un problema de alta colinealidad es la obtención de un elevado coeficiente de bondad de ajuste para la regresión estimada, de conjunto con pocas razones- $t$  significativas. En este caso tendremos una prueba  $F$  para la validez del modelo de regresión que es significativa, pero la mayoría de las variables, por separado, no rebasan el valor crítico de  $t$ .

Nuevamente en nuestro ejemplo de función de producción macroeconómica para Cuba, vamos a completar el reporte de los resultados de la estimación preliminar que realizamos en el capítulo 6, para valorar la situación que se presenta en cuanto a la multicolinealidad.

$$\begin{aligned}
 LQ &= 2,187148 - 0,500365 LK + 3,714342 LL - 1,777154 LH \\
 t &\quad (0,88163) \quad (-2,59055)^* \quad (4,29105)^* \quad (-4,91871)^* \\
 R^2 &= 0,68484
 \end{aligned}$$

Evidentemente, esta situación no se presenta en el caso que nos ocupa.

Tenemos que todos los factores de producción incluidos en esta especificación resultan significativos, aun cuando la bondad de ajuste no resulta tan alta.

Dado que las correlaciones apareadas pueden ser engañosas, se recomienda también examinar los coeficientes de correlación parcial. En un modelo con alta bondad de ajuste, si los coeficientes de correlación parcial entre la variable dependiente y cada regresora —que por definición suponen que el resto de las regresoras se mantienen constantes—<sup>4</sup> son bajos, ello es indicativo de la posible presencia de alta correlación entre las regresoras y que al menos alguna de ellas sea redundante o superflua.

Otra investigación posible es la estimación de regresiones auxiliares o subsidiarias a la regresión principal que estamos examinando. Si existe multicolinealidad entre las regresoras, es porque al menos una de ellas puede ser expresada como una combinación lineal (exacta o aproximada) de al menos otra de ellas, tal y como planteamos en (8.2). Por lo tanto, en esta indagación planteamos regresiones auxiliares que relacionan cada regresora con el resto de ellas, las estimamos y calculamos el  $R^2$  correspondiente, que denotamos con  $R_j^2$ , con  $j$  como el subíndice correspondiente a la regresora que colocamos como variable independiente en la regresión auxiliar. El test  $F$  de significación global de esas regresiones auxiliares nos dará información sobre si la variable  $j$  es colineal con las demás: en el caso de que la  $F$  calculada supere el valor crítico (según  $gl$  y nivel de significación) o que la probabilidad  $p$  asociada sea menor que el nivel de significación escogido.

Klein ha propuesto una regla práctica para el examen de las regresiones auxiliares: en lugar de probar todos los  $R_j^2$  por separado,<sup>5</sup> se sugiere que la multicolinealidad puede ser un problema a tomar en consideración sólo si alguna de las  $R_j^2$  obtenidas resulta mayor que el  $R^2$  correspondiente a la regresión principal.<sup>6</sup>

Esta prueba, como las otras, puede arrojar resultados distorsionados en el caso de que la multicolinealidad comprenda sólo un número reducido de las regresoras.

Otro análisis posible es el del factor de inflación de varianza definido en (8.3). Podemos expresar la varianza del estimador del coeficiente de regresión  $\beta_j$  en función del  $FIV$  como:

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\mathbf{x}_j' \mathbf{x}_j [1 - R_j^2]} = \frac{\sigma^2}{\mathbf{x}_j' \mathbf{x}_j} FIV_j$$

<sup>4</sup> Es decir, “depuramos” la posible influencia de las otras regresoras a través de la que se incluye en el análisis.

<sup>5</sup> Recuerden que, según (6.20), el test  $F$  de significación global es equivalente a un test del  $R^2$ .

<sup>6</sup> La que se obtiene de regesar  $Y$  en función de  $X_j$ , con  $j = 1, \dots, k$ .

Así,  $\text{var}(\hat{\beta}_j)$  es directamente proporcional a  $FIV_j$ . Por ello algunos autores utilizan el  $FIV_j$  como un indicador de la multicolinealidad. Una regla práctica es que si  $FIV_j > 10$  (que equivale a  $R_j^2 > 0,90$ ), se considera que la variable  $X_j$  es altamente colineal con el resto.

Noten, sin embargo, que este no es el único elemento que influye en  $\text{var}(\hat{\beta}_j)$ . Para tener pequeñas varianzas de los coeficientes de regresión, como deseamos y necesitamos, también es importante que la variable en cuestión posea suficiente variabilidad (alta  $\mathbf{x}_j' \mathbf{x}_j = \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$ ) y que la varianza del término de error  $\sigma^2$  sea pequeña. Es decir, bajas  $\sigma^2$  y altas  $(\mathbf{x}_j' \mathbf{x}_j)$  pueden contrarrestar un alto  $FIV_j$ . De modo que un alto  $FIV_j$  no tiene necesariamente que ocasionar altas  $\text{var}(\hat{\beta}_j)$  y  $\text{se}(\hat{\beta}_j)$ .

De vuelta a nuestro ejemplo de la función de producción, vamos a concentrar en la tabla 8.2 algunos indicadores que nos servirán para llegar a alguna conclusión acerca de los problemas de multicolinealidad entre los factores de producción incorporados, a partir del examen de los  $R_j^2$  y los  $FIV_j$ .

**Tabla 8.2: Indicadores de multicolinealidad para la función de producción macroeconómica en Cuba, 1980-2000**

|                                                                                                                                                                                                                                                                      | $R_j^2$ | $F_j$     | $p$     | $FIV_j$ |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-----------|---------|---------|
| LK                                                                                                                                                                                                                                                                   | 0,90096 | 77,32083  | 0,00000 | 10,1    |
| LL                                                                                                                                                                                                                                                                   | 0,94504 | 146,16300 | 0,00000 | 18,2    |
| LH                                                                                                                                                                                                                                                                   | 0,91912 | 96,59956  | 0,00000 | 12,4    |
| LQ                                                                                                                                                                                                                                                                   | 0,68484 | 12,31351  | 0,00016 | -       |
| *: Los reglones correspondientes a los logaritmos de los factores de producción contienen la información de las regresiones auxiliares; mientras que el correspondiente al logaritmo del producto, la vinculada a la regresión principal (la función de producción). |         |           |         |         |

Fuente: Elaboración propia

Podrán constatar que, mientras que los métodos de examinar los coeficientes de regresión simples o de inspeccionar el  $R^2$  de la regresión principal de conjunto con las razones- $t$  no nos daban motivo de inquietud acerca del posible grado de multicolinealidad entre los factores de producción, el examen de los  $R_j^2$  de las regresiones auxiliares, sus test F de significación global, así como los factores de inflación de la varianza de los coeficientes de regresión que se obtienen sí son preocupantes. Adviertan que el  $R^2$  global es bastante menor que los  $R_j^2$ , por lo que la regla de Klein nos indica que los problemas son de temer.



Estos no son los únicos indicadores para detectar la existencia y medir el grado de multicolinealidad, se han diseñado otros tales como los valores propios y el índice de condición, que pueden ser consultados en Gujarati (2004, pp. 347-348).

Como habrán podido comprobar a través del ejemplo desarrollado, todos estos métodos tienen sus pros y contras, por lo que sugerimos examinarlos todos exhaustivamente, para cerciorarnos de que la multicolinealidad no constituya un problema de consideración para nuestro trabajo empírico.

Con todo, la multicolinealidad es un problema relativo a la muestra disponible, sobre la cual no tenemos control; en especial si se trata de datos que no provienen de experimentos cuidadosamente diseñados, como es el caso de la mayoría de los datos disponibles en Economía. La pregunta clave será entonces si podemos convivir con la multicolinealidad.

La respuesta a esta cuestión dependerá de los resultados obtenidos en la regresión y del objetivo del estudio empírico:

- ✓ Si obtenemos razones  $t$  altas para los parámetros estimados, el problema no es tan grave; en cambio, si las razones  $t$  están cercanas a su valor crítico, entonces este debe ser un problema a considerar.
- ✓ Si lo que pretendemos es conseguir una herramienta para pronosticar y podemos suponer que en el futuro esa interrelación entre regresoras se mantendrá, entonces la multicolinealidad no es tan dañina.
- ✓ Sin embargo, si el propósito del trabajo empírico es conseguir estimados confiables de los parámetros de interés, ya sabemos que el problema puede ser serio.

### 8.3 Medidas Remediales

¿Qué hacer en presencia de multicolinealidad? No existe un remedio infalible para circunvalar este problema. Algunos autores aconsejan no hacer nada al respecto, dado que se trata de un problema asociado a deficiencia de datos. Otros sí recomiendan actuar siguiendo ciertas reglas prácticas.<sup>7</sup>

Uno de los procedimientos recomendados consiste en eliminar variables del MLNR. Pero este remedio puede ser peor que el problema mismo de multicolinealidad, porque al eliminar una variable de un modelo económicamente viable incurrimos en un error de incorrecta especificación peor que la multicolinealidad, que es no estar incluyendo todas las variables que contribuyen a la explicación de la variabilidad de  $Y$ . Es decir, tendremos variables omitidas (justo las que eliminamos), y ello nos conduce a estimados sesgados de los parámetros de la regresión. Sobre este asunto profundizaremos en el capítulo 13 de la Parte Tres de este texto.

<sup>7</sup> Véase Gujarati (2004, pp. 349-356) y Dresdner y Vázquez (2004, pp. 166-168).

La sugerencia práctica más eficiente para resolver el problema de multicolinealidad es conseguir más datos o una nueva muestra. Lamentablemente, esto la mayoría de las veces no es factible.

También se recomienda en estos casos repensar el modelo a estimar. Buscar otra explicación en la teoría económica para  $Y$  a partir de regresoras que no presenten este problema.

Es posible también, si se dispone de estimaciones obtenidas en investigaciones empíricas anteriores que no presentaban esta deficiencia o donde era menos grave, basarse en esos resultados para la nueva estimación. Es el procedimiento que denominamos utilizar información *a priori* sobre algunos parámetros de interés. A partir de esa información, lo que hacemos es estimar un MLNR restringido a la información incorporada.

Para el caso de nuestra función de producción, por ejemplo, si en un estudio anterior pudimos comprobar la presencia de rendimientos constantes a escala, esta información se puede incorporar a la nueva estimación para circunvalar el problema de multicolinealidad de  $LL$  con el resto de las regresoras, de la siguiente forma:

$$\begin{aligned}
 \ln Q &= \ln A + a_K \ln K + a_L \ln L + a_H \ln H, \\
 \text{si } a_K + a_L + a_H &= 1, \text{ entonces } a_L = 1 - a_K - a_H \\
 \ln Q &= \ln A + a_K \ln K + (1 - a_K - a_H) \ln L + a_H \ln H \\
 \ln Q &= \ln A + a_K \ln K + \ln L - a_K \ln L - a_H \ln L + a_H \ln H \\
 \ln Q - \ln L &= \ln A + a_K (\ln K - \ln L) + a_H (\ln H - \ln L) \\
 \text{si } Y_Q &= \ln Q - \ln L, X_K = \ln K - \ln L \text{ y } X_H = \ln H - \ln L, \text{ entonces} \\
 Y_Q &= \ln A + a_K X_K + a_H X_H \tag{8.4}
 \end{aligned}$$

Así, estimamos el modelo restringido (8.4), recuperamos  $a_K$  y  $a_H$  de la nueva estimación —en la que, con suerte,  $X_K$  y  $X_H$  no son colineales— y conseguimos el estimado de  $a_L$  a partir de  $\hat{a}_L = 1 - \hat{a}_K - \hat{a}_H$ .

Sin embargo, esta vía es también bastante controvertida, porque con el trabajo empírico por lo general pretendemos validar esas predicciones *a priori* de la teoría económica y no precisamente imponerlas.

Otra vía bastante frecuente es la transformación de las variables regresoras para obtener resultados donde se reduzca la colinealidad. Por ejemplo, entre las transformaciones reductoras de colinealidad podemos mencionar el trabajo con series de datos en términos per cápita o en términos reales, lo que significa operar con variables obtenidas a través de una razón, dividiendo el MLNR entre una determinada variable —como puede ser la población, para obtener datos en términos de per cápita, o entre un índice de precios, para obtener datos en términos reales. También se puede trabajar con las variables en primeras diferencias,

en lugar de operar con las variables en niveles absolutos.

Noten que la transformación de la función de producción que desarrollamos anteriormente, introduciendo información *a priori*, es equivalente a haber trabajado con los logaritmos del producto por trabajador, así como de la dotación de factores (en este caso capital físico y humano) por trabajador. Puede perfectamente ser una transformación que mitigue la multicolinealidad.

El trabajo con series en primeras diferencias significa estimar el modelo con  $\Delta Y_t = Y_t - Y_{t-1}$  y con  $\Delta X_{ij} = X_{ij} - X_{t-1,j}$ , en lugar de  $Y_t$  y  $X_{ij}$ . Esta transformación tiene una ventaja adicional, como mostraremos en la Parte Cinco de este libro sobre la Econometría de Series de Tiempo. Esa ventaja consiste en que, además de mitigar la multicolinealidad, permite convertir series no estacionarias en estacionarias.

Adviertan que en la función de demanda de importaciones para la economía cubana que estimamos en el capítulo 5, trabajamos las variables en diferencias relativas con respecto al año anterior. En las tablas 8.3 y 8.4 mostramos algunos de los indicadores que estudiamos para la detección de multicolinealidad y su grado correspondientes a esta función de demanda. Adviertan que los problemas no son de consideración, en contraste con los que detectamos para el ejemplo de la función de producción: una correlación simple de apenas 17% entre variables regresoras, un test  $F$  no significativo y un  $FIV$  bastante menor que 10.

**Tabla 8.3: Matriz de correlaciones entre variables de la función de demanda de importaciones para Cuba, 1961-2005**

|             | $\dot{m}_t$ | $\dot{y}_t$ | $\dot{p}_t$ |
|-------------|-------------|-------------|-------------|
| $\dot{m}_t$ | 1           |             |             |
| $\dot{y}_t$ | 0,44960     | 1           |             |
| $\dot{p}_t$ | -0,67451    | -0,17007    | 1           |

Fuente: Elaboración propia

Fíjense, además, cómo se cumple la regla o criterio de Klein: el  $R^2$  global es bastante mayor que el  $R_j^2$  (véase tabla 8.4).

**Tabla 8.4: Indicadores de multicolinealidad para la función de demanda de importaciones para Cuba, 1961-2005\***

|                                                                                                                                                                                             | $R_j^2$ | $F_j$    | $p$     | $FIV_j$ |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|----------|---------|---------|
| $\dot{y}_t$                                                                                                                                                                                 | 0,02892 | 1,28073  | 0,26404 | 1,03    |
| $\dot{m}_t$                                                                                                                                                                                 | 0,57046 | 27,88918 | 0,00000 | -       |
| *: El primer renglón contiene la información de la regresión auxiliar entre las dos regresoras; el segundo, la vinculada a la regresión principal (la función de demanda de importaciones). |         |          |         |         |

Fuente: Elaboración propia

Como ya mencionamos antes, otra transformación posible es trabajar con las variables expresadas como razones. Apliquemos esa transformación al caso de la función de producción: vamos a trabajar con la producción por trabajador y con la dotación de capital físico y humano por trabajador. Observemos el impacto de esta transformación en los indicadores de multicolinealidad (tabla 8.5).

**Tabla 8.5: Indicadores de multicolinealidad para la función de producción macroeconómica en Cuba con variables por trabajador, 1980-2000\***

|                                                                                                                                                                                                                                           | $R_j^2$ | $F_j$    | $p$     | $FIV_j$ |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|----------|---------|---------|
| L(K/L)                                                                                                                                                                                                                                    | 0,47842 | 16,51042 | 0,00073 | 1,9     |
| L(Q/L)                                                                                                                                                                                                                                    | 0,47640 | 7,73365  | 0,00409 | -       |
| *: El primer renglón contiene la información de la regresión auxiliar entre las dos regresoras (los capitales físico y humano por trabajador); el segundo, la vinculada a la regresión principal (la función de producción transformada). |         |          |         |         |

Fuente: Elaboración propia

Adviertan cómo mejoró el factor de inflación de varianza con relación a los resultados mostrados en la tabla 8.2. También la comparación entre el  $R^2$  de la regresión principal y la auxiliar mejora; pero continúa siendo mayor el de la regresión auxiliar, por lo que el criterio de Klein indica que subsisten problemas.

Estas transformaciones de variables con el objetivo de disminuir la multicolinealidad pueden, no obstante, tener efectos secundarios no deseados con relación a otros supuestos del MLNR.

Cuando trabajamos con series en diferencias, en realidad lo que hacemos es plantear el MLNR para los momentos  $t$  y  $t-1$  y deducir uno de otro:

$$\begin{aligned}
 Y_t &= \beta_1 + \beta_2 X_{t,2} + \cdots + \beta_k X_{t,k} + \varepsilon_t \\
 Y_{t-1} &= \beta_1 + \beta_2 X_{t-1,2} + \cdots + \beta_k X_{t-1,k} + \varepsilon_{t-1} \\
 Y_t - Y_{t-1} &= \beta_2 (X_{t,2} - X_{t-1,2}) + \cdots + \beta_k (X_{t,k} - X_{t-1,k}) + u_t, \text{ con } u_t = \varepsilon_t - \varepsilon_{t-1}
 \end{aligned} \tag{8.5}$$

Fíjense que el término de error  $u_t$  del modelo (8.5), obtenido como la

diferencia entre las perturbaciones del modelo original en dos momentos contiguos en el tiempo, puede no satisfacer uno de los supuestos del MLNR: el supuesto G, referido a la no autocorrelación entre los términos de error. Este resultado es altamente probable si el término de error original cumple el supuesto G. Adviertan que, por ejemplo,  $u_t = \varepsilon_t - \varepsilon_{t-1}$  y  $u_{t+1} = \varepsilon_{t+1} - \varepsilon_t$ , por lo que la perturbación del modelo original  $\varepsilon_t$  interviene en la construcción de  $u$  en dos momentos distintos. Esto puede acarrear probablemente correlación entre  $u_t$  y  $u_{t+1}$ ; es decir, autocorrelación serial.

La transformación diferencias nos hace perder una observación, con lo que perdemos también un grado de libertad, lo que no es deseable en muestras pequeñas.

Y finalmente esta transformación tiene una interpretación adecuada en series de tiempo o longitudinales, no así en datos de corte transversal.

En el caso de que trabajemos con variables obtenidas a través de una razón, el nuevo término de error —resultante de dividir el término original entre la variable común para la transformación— podría incumplir uno de los supuestos del MLNR: la homoscedasticidad. Si el término original es homoscedástico, el nuevo con seguridad no lo será.

Así, con ambas transformaciones mejoramos en cuanto a la multicolinealidad, pero podemos incurrir en errores mayores, que nos conducen a problemas todavía más serios. De modo que debemos ser cuidadosos al aplicar estas transformaciones y verificar que el modelo definitivo mantenga los supuestos que subyacen a un MLNR.

Una variante a la introducción de información a priori para mitigar la multicolinealidad, es la combinación de datos de series de tiempo con datos de corte transversal. De lo que se trata es de incorporar resultados de trabajos empíricos anteriores realizados con datos de corte transversal, donde los problemas de multicolinealidad no sean de relevancia, en la estimación del modelo con datos longitudinales.

Gujarati (2004, p.351) nos brinda un ejemplo bastante esclarecedor en tal sentido: supongamos que necesitamos conocer las elasticidades ingreso y precio de la demanda de automóviles; para ello requerimos estimar la correspondiente función de demanda; los factores explicativos en este caso son los ingresos de los individuos y el precio de los automóviles, pero frecuentemente estas dos variables son altamente colineales en la información de series longitudinales, de modo que las elasticidades ingreso y precio de la demanda de automóviles que estaremos serán muy imprecisas y en cada caso no reflejarán bien ese parámetro, porque ambas estarán “contaminadas” por la influencia de la otra variable explicativa. Si dispusiéramos de información de corte transversal (obtenida, por ejemplo, a través de las Encuestas de la Situación Económica de los Hogares, o de encuestas sobre grupos de consumidores) acerca de la elasticidad-ingreso, ésta podría ser incorporada a nuestra función de demanda para obtener un estimado de la elasticidad-precio.

Algunos autores mencionan otros métodos para mitigar el problema de la multicolinealidad como el análisis de componentes principales y el análisis factorial.<sup>8</sup> El análisis de componentes principales es un método de las técnicas estadísticas multivariadas cuyo objetivo es transformar un conjunto de variables dado (por ejemplo, nuestras regresoras) en un conjunto de variables menor, que reciben el nombre de componentes principales, y que logran explicar una parte importante de la variabilidad de la variable dependiente.<sup>9</sup> De esta forma, podemos concentrar el análisis de regresión en las regresoras que se revelen como componentes principales. Es un método exploratorio que puede ser utilizado para aceptar o excluir variables en la regresión.

El propósito del análisis factorial es diseñar nuevas variables que resuman toda la información contenida en las variables originales. Es muy útil cuando existe una gran cantidad de variables explicativas y, por supuesto, al reducir la cantidad de variables con pérdida mínima de información, puede ser aplicado para mitigar el problema que nos ocupa: la multicolinealidad.

Por lo general, los software estadísticos avanzados incluyen estas técnicas del análisis de componentes principales y del análisis factorial en sus menús de opciones.

Con este acercamiento al problema de la multicolinealidad cerramos la Parte Dos de este texto. En la Parte Tres vamos a abordar las reiteradamente mencionadas pruebas de incorrecta especificación o pruebas de diagnóstico del MLNR, y procederemos tal cual hemos hecho en este capítulo 8: profundizar sobre del resto de los supuestos del MLNR (más allá de la no colinealidad), sobre los efectos que tiene su no cumplimiento para la estimación de los parámetros de interés y de cómo remediar, de la mejor manera posible, los problemas de especificación presentes en los modelos estimados supuestamente como MLNR. 

<sup>8</sup> Véase Gujarati (2004, p. 355).

<sup>9</sup> Véase Alfonso (2007, pp. 127-134).

# *preguntas*

Pruebe dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo.

1. ¿Cuáles pueden ser las causas de la multicolinealidad?
2. ¿Qué tipos de multicolinealidad se pueden presentar? ¿Cuáles son las consecuencias en cada caso para la estimación del modelo de regresión?
3. ¿Qué es el factor de inflación de la varianza? ¿Para qué nos sirve ese indicador?
4. Diga si conoce algún test para detectar la multicolinealidad. En caso afirmativo, diga cuáles son sus hipótesis nula y alternativa, y cuál el estadígrafo de prueba. De lo contrario, explique por qué no existen tests al respecto.
5. ¿Cómo podemos detectar la multicolinealidad? Explique para cada uno de los tipos de multicolinealidad.
6. Si Ud. descubre que hay presencia de multicolinealidad en sus datos, ¿qué posibilidades tiene de solucionar este problema?
7. Discuta las medidas para la corrección del problema de la multicolinealidad



# ejercicios

8.1 La tabla B.7 del CD acompañante contiene datos antropométricos de 20 mujeres sanas entre 25 y 34 años de edad sobre las siguientes características individuales:  $Y$  = grasa corporal,  $X_1$  = espesor de la piel en los tríceps,  $X_2$  = perímetro de las caderas,  $X_3$  = perímetro del antebrazo.

- a. Estime los siguientes modelos de regresión y disponga los resultados en una tabla.

$$Y_i = \beta_0 + \beta_1 X_{1i} + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

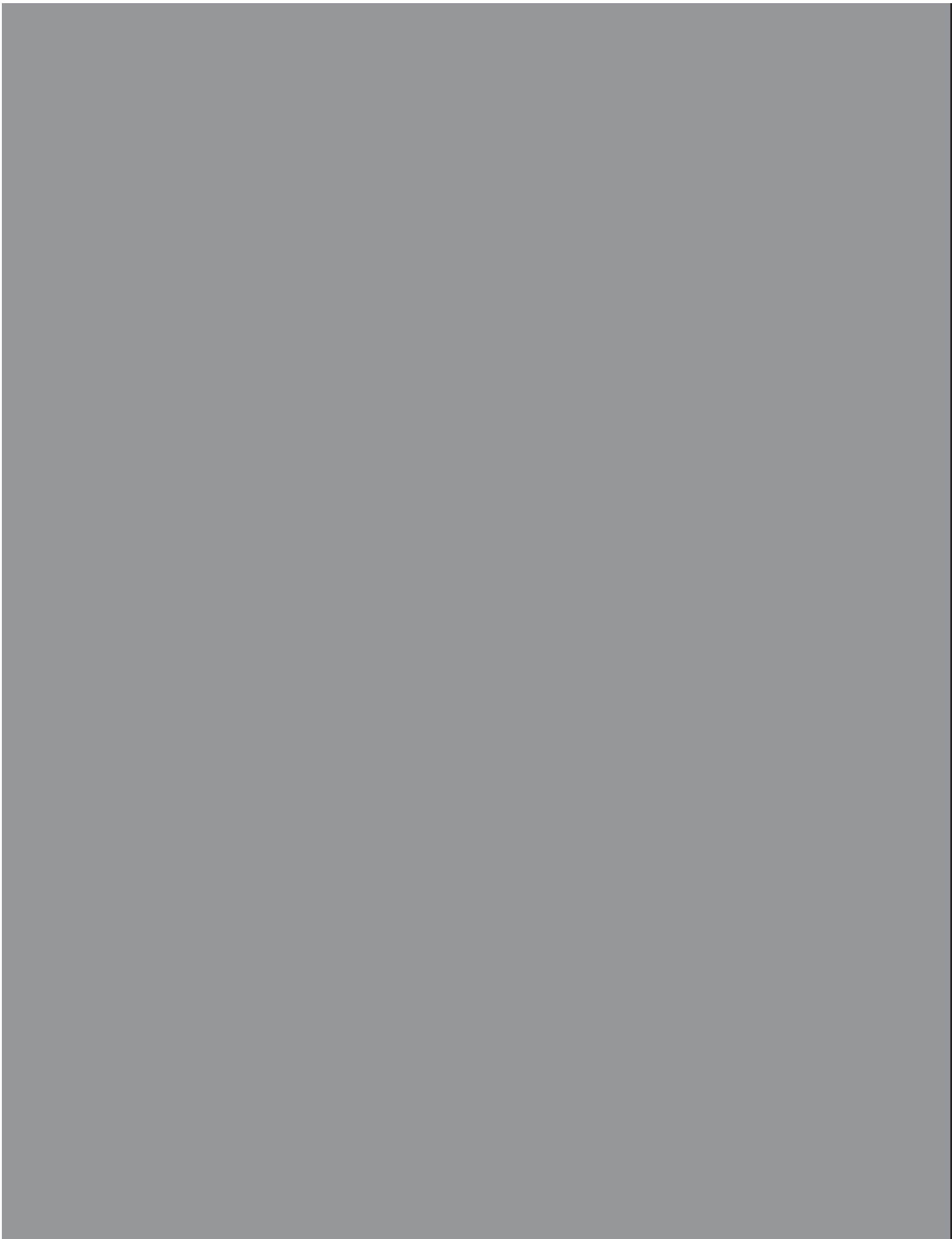
- b. Comente las estimaciones del coeficiente de  $X_1$  en las tres estimaciones.
- c. Comente la significación de los coeficientes de la tercera regresión. Comente la significación global de la misma regresión.
- d. Calcule la matriz de correlaciones de las variables. Comente sus resultados.
- e. Calcule el factor de inflación de varianza del coeficiente de  $X_1$ .

8.2 En la tabla B.8 del CD acompañante presentamos datos de la producción de caña de azúcar en Cuba ( $Q_t$ , medida en millones de toneladas) y de algunos de los factores que influyen en su variabilidad para el período 1975-2001: área a zafra ( $T_t$ , en miles de hectáreas), fertilizantes aplicados en términos de sustancia activa ( $F_t$ , en miles de toneladas), precipitaciones anuales ( $P_t$ , en milímetros de lluvia caída); e inversiones ( $INV_t$ , en millones de pesos).

- a. Estime una función de producción tipo Cobb-Douglas.
- b. Investigue acerca de la estabilidad de los coeficientes de regresión estimados, a partir de hacer estimaciones para diferentes períodos muestrales. Compare los coeficientes estimados para las submuestras con los estimados previamente. Comente los resultados.
- c. Calcule la matriz de correlaciones de las variables. Comente los resultados.
- d. Aplique el método sugerido por Klein para investigar problemas de multicolinealidad. Comente sus resultados.
- e. Aplique una de las medidas remediales que se sugieren para mitigar problemas de multicolinealidad: reespecifique la función a partir de todas las variables por unidad de área a zafra.
- f. Realice las anteriores investigaciones al nuevo modelo estimado y comente los resultados.



- 8.3 Para el mismo conjunto de datos con el que estimamos la función de producción macroeconómica de Cuba para el período 1980-2000 (ver tabla B.2 del CD acompañante), realice las siguientes investigaciones:
- Reespecifique dicha función a partir de expresar todas las variables involucradas en términos relativos al capital físico de que dispone la economía. Investigue si esta transformación disminuyó los problemas de multicolinealidad. Aplique los distintos procedimientos presentados en este capítulo.
  - Reespecifique la función de producción a partir de trabajar con las variables en primeras diferencias de sus logaritmos. Investigue igualmente los resultados obtenidos.



• PARTE TRES

• DISCUTIENDO  
• LOS SUPUESTOS  
• DEL MLNR



Al iniciar la Parte Dos de este libro habíamos comentado que, si bien el Modelo Lineal Normal de Regresión es uno de los pilares básicos de la econometría, se trata de un modelo restrictivo, dados los supuestos que entraña, que con frecuencia no son representativos de la realidad económica.

Esta parte del libro la dedicamos a flexibilizar los supuestos asociados al MLNR y discutir acerca del posible impacto de este relajamiento en los estimadores de los parámetros de interés. Seguidamente desarrollamos cómo detectar contradicciones entre la realidad y el modelo postulado: se trata de evaluar si efectivamente un MLNR puede haber dado origen a los datos que conocemos sobre el fenómeno que intentamos modelar. Es decir, investigamos si los datos de que disponemos para las variables dependiente y explicativas, y su interconexión a través de un MLNR, no contradicen los supuestos subyacentes a ese tipo de modelo. Para hacerlo, aplicamos procedimientos *ad-hoc* y variados tests estadísticos, algunos de ellos especialmente diseñados con ese propósito y otros no tan específicos, pero que dan pistas sobre posibles desviaciones entre la realidad expresada en los datos y los supuestos que corresponden a un MRLN.

En el caso de que los datos no sostengan un determinado supuesto, damos indicaciones acerca de posibles procedimientos para remediar esta situación — asumiendo que ese fuera el único error de especificación en el modelo— y, finalmente, ofrecemos algunos consejos para la selección del modelo adecuado, así como para la conducción de trabajos empíricos y la redacción de los informes correspondientes.

Esta Parte Tres está compuesta por cinco capítulos, que abordan la presencia de errores no-esféricos en el modelo (heteroscedasticidad y autocorrelación), la no-normalidad del término de error o componente aleatorio del modelo, el posible carácter estocástico de las variables explicativas, así como otros errores de especificación del modelo de regresión y criterios para la selección del modelo.

CAPÍTULO 9



# HETEROSCEDASTICIDAD

En este capítulo relajamos uno de los supuestos del Modelo Lineal Normal de Regresión: que la varianza de la variable dependiente o de la perturbación del modelo sea constante para todas las observaciones. Ya en el Capítulo 5 de este texto referimos que cuando esta condición se cumple, decimos que la variable dependiente o el término de error del modelo son homoscedásticos; es decir, que tienen igual varianza para todas las observaciones  $Var(\varepsilon_i) = E[\varepsilon_i^2] = \sigma^2$ . Esto implica que  $Var[Y_i] = \sigma^2$ .

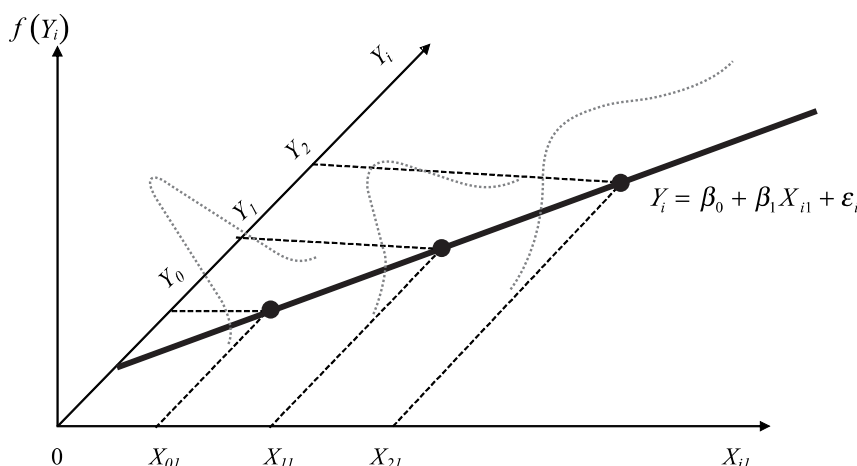
## 9.1 Heteroscedasticidad

Cuando las varianzas para cada observación no son iguales, estamos ante el error no esférico de heteroscedasticidad:  $Var(\varepsilon_i) = \sigma_i^2$ . Noten cómo representamos la diferencia entre las varianzas de la perturbación correspondiente a cada observación de la muestra a través del subíndice  $i$ . Así, la matriz de varianzas y covarianzas del término de error ya no va a tener más un valor común en su diagonal, como se supone en un MRLN, sino que los elementos de la diagonal serán diferentes entre sí.

Incluimos dos representaciones gráficas que ilustran cómo pueden comportarse las varianzas del término de error estocástico (que es también la varianza de la variable explicada) en un modelo con dos variables: la figura 5.1 del Capítulo 5, correspondiente a un patrón homoscedástico en la varianza, y la 9.1, para el caso que nos preocupa ahora, bajo heteroscedasticidad.

En la figura 5.1 podemos ver que todas las observaciones tienen la misma varianza; la dispersión en torno al valor medio sobre la línea de regresión es igual en todos los casos. Noten que en la figura 9.1 hemos representado varianzas que crecen a medida que crece la variable explicativa  $X$ . En esta situación, como ya explicamos, los elementos de la diagonal principal de la matriz de varianzas y covarianzas no serán iguales, sino diferentes para cada elemento de la muestra.

**Figura 9.1: Distribución de errores con heteroscedasticidad**



Diversas razones pueden provocar que la varianza del término de error de una regresión se comporte de tal manera. A continuación consideramos posibles fuentes de heterogeneidad de la varianza.

*Modelos de aprendizaje.* Suponemos que los individuos, con el tiempo y la experiencia acumulada, pueden disminuir el número de errores en su tarea o quehacer habitual y, por lo tanto, la variabilidad de su desempeño se puede reducir.

*Efecto de escala.* Esta situación es típica en estudios sobre los ingresos de las familias. Por lo general, encontramos que a mayores ingresos de las familias o los individuos, existe una mayor posibilidad de elección entre variantes de consumo y ahorro, por lo tanto, estas magnitudes presentarán mayor variabilidad; es decir, su varianza aumentará a medida que progresen los ingresos. De igual manera, este comportamiento que se observa en los hogares es válido también para las empresas: a medida que una empresa tiene mayores ingresos, tiene mayores posibilidades de elegir entre distribuir dividendos a sus accionistas o retenerlos para invertir (siempre que se desenvuelva en un contexto que le permita esa decisión).

*Existencia de grupos heterogéneos al interior de una muestra.* Este caso es muy común también en las situaciones previamente reseñadas. Cuando realizamos estudios sobre los ingresos familiares podemos tener distintos grupos de observaciones, además de hogares diferenciados por su nivel de ingresos (por ejemplo altos, medios y bajos), podemos tener hogares rurales y urbanos, hogares donde predominen los



ancianos o los niños, donde el jefe de familia tenga mayor o menor instrucción, donde predominen mujeres u hombres. En fin, pueden presentarse disímiles clasificaciones, y estas diferencias pueden tener una influencia en la variabilidad del término de error. Asimismo, en los estudios sobre las empresas al interior de una industria o sobre las industrias que componen un determinado sector económico, podemos encontrar una importante heterogeneidad, con la consecuencia antes mencionada.

*Mejora en las técnicas de recolección de información.* A medida que estas técnicas mejoran, el error de medida debe tender a disminuir y, consecuentemente, disminuirá la variabilidad.

*Presencia de factores atípicos.* La heteroscedasticidad puede ser consecuencia de la existencia de observaciones atípicas en la muestra: se trata de observaciones que son muy diferentes del resto, bien porque son muy pequeñas o muy grandes comparadas con estas, y muy probablemente esto es así porque dichas observaciones han sido extraídas de una población diferente de la que dio origen al resto. Cuando tenemos este tipo de observación en una muestra, y sobre todo si se trata de una muestra pequeña, el incluirla o no en la regresión modifica sustancialmente sus resultados. Ya en el Capítulo 7 de este texto señalamos como utilizar las variables dicotómicas, lo que nos da una pista de cómo lidiar con este problema.<sup>1</sup>

*Modelo de regresión incorrectamente especificado.* Cuando postulamos un MLRN para relacionar nuestras variables de interés y esta suposición no es acertada, podemos encontrar entonces que los errores de la regresión estimada a partir del modelo erróneo no son homoscedásticos, como resultado de la mala especificación. Esto ocurre, por ejemplo, cuando:

No incluimos todas las variables relevantes en el modelo. Recuerden que toda variable relevante que no incluyamos explícitamente en la parte sistemática de la regresión estará en este caso incorporada al término de error, que en tal situación puede presentar comportamientos distintos de lo que se espera para un modelo bien especificado.

La forma funcional que hemos postulado no es la correcta. Por ejemplo, cuando trabajamos con un modelo lineal, siendo un doble log el correcto.

Se transforman incorrectamente los datos de la regresión. Por ejemplo, cuando se trabaja con las primeras diferencias o con razones, sin que ello sea necesario.

A partir de las explicaciones previas es fácil colegir que esta situación de varianzas diferentes para cada individuo de la muestra o por grupos de individuos, suele presentarse con mayor frecuencia en datos de corte transversal.

Existen distintos tipos de heteroscedasticidad. La comúnmente denominada heteroscedasticidad *de bloque*, que se refiere a que dos o más grupos al interior de una muestra tienen errores con varianzas diferentes. Para determinar si estamos en presencia de ese tipo de heteroscedasticidad, podemos dividir la muestra en los grupos o bloques antes aludidos, estimar la regresión para cada bloque y

<sup>1</sup> ¿Tendría Ud. alguna sugerencia sobre cómo modelar esta situación?

comparar entre sí las varianzas de los términos de error encontrados en cada caso.

Otro tipo de heteroscedasticidad es la **proporcional**, así llamada porque la varianza está relacionada a una o más variables explicativas. Por ejemplo, podemos suponer que las varianzas individuales se relacionan con una de las variables explicativas de la siguiente forma:  $\sigma_i^2 = X_i \sigma^2$ , donde  $\sigma^2$  es una constante y  $X_i$  es una variable explicativa. Por supuesto, la relación entre la varianza y las variables explicativas puede ser mucho más compleja, tal como  $\sigma_i^2 = f(X_i, \sigma^2)$ , donde  $f$  es una función.

En un modelo de regresión no podemos descartar *a priori* que no exista heteroscedasticidad. Debemos evaluar si los residuos de la regresión no contradicen este supuesto o que, al menos, este problema no sea lo suficientemente importante como para distorsionar de forma relevante los parámetros estadísticos de nuestro interés y la inferencia estadística sobre ellos. En caso contrario, habría necesariamente que tomar medidas para corregir esta situación.

## 9.2 Consecuencias de la Heteroscedasticidad

En presencia de heteroscedasticidad los estimadores de los coeficientes de regresión son todavía lineales, así como insesgados en pequeña muestra y consistentes en muestras grandes. Este problema no afecta estas propiedades, que no requieren de imponer condiciones especiales a la varianza de las perturbaciones para que se cumplan.

Sin embargo, en situación de heteroscedasticidad la varianza del estimador MCO de  $\beta$  no será la que obtuvimos en (5.14), pues ya no tendremos una varianza común a las perturbaciones de todas las observaciones, sino diferenciada.

En este caso la matriz de varianzas y covarianzas adopta la siguiente forma:

$$E(\varepsilon \cdot \varepsilon') = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} = \Omega \quad (9.1)$$

¿Cómo será bajo estas condiciones la varianza del estimador  $\hat{\beta}$  obtenido por MCO? Se puede demostrar que,

$$\begin{aligned} Var(\hat{\beta}) &= E[(\hat{\beta} - \beta) \cdot (\hat{\beta} - \beta)'] \\ &= E\left[\left((X'X)^{-1} X' \varepsilon\right) \cdot \left((X'X)^{-1} X' \varepsilon\right)'\right] \\ Var(\hat{\beta}) &= (X'X)^{-1} X' \Omega X (X'X)^{-1} \end{aligned}$$

Pero ahora  $\Omega \neq \sigma^2 I$  y no podemos simplificar la expresión anterior como lo hicimos en el Capítulo 6 para el caso homoscedástico. Esta varianza ya no es la menor posible,<sup>2</sup> de modo que el estimador MCO pierde su eficiencia y, por lo tanto, no es adecuado hacer inferencias en base a los resultados de la estimación MCO. El estimador MCO de la varianza es un estimador sesgado cuando estamos en presencia de heteroscedasticidad y desconocemos el sentido del sesgo (si positivo o negativo), por lo que no sabremos cuán precisa es la estimación de los parámetros  $\beta_j$  y, en definitiva, no podremos confiar en los test  $t$  y  $F$ . El resultado anterior muestra también que la fórmula estándar del método MCO no es la correcta para estimar la varianza de los estimadores de los parámetros. White ha introducido una fórmula alternativa que mejora esta estimación. Los errores estándares de los estimadores obtenidos por esta fórmula de White son, muchas veces, llamados estimadores robustos.

Resumiendo, si estimamos un modelo de regresión con perturbaciones heteroscedásticas a través de MCO, los estimadores de  $\beta$  se mantienen insesgados y consistentes, pero pierden eficiencia. Esto puede no ser tan grave, sobre todo cuando nuestras estimaciones están respaldadas por razones  $t$  elevadas. Con todo, si se ubican en valores cercanos a dos, entonces estaremos en aprietos y debemos conseguir la forma de mitigar este problema, para asegurarnos de que realmente tenemos coeficientes de regresión adecuados. En el epígrafe siguiente trataremos el tema de la detección de este desvío de los supuestos del modelo clásico.

### 9.3 Detección de la Heteroscedasticidad

Para empezar, es importante que analicemos si la naturaleza del problema que abordamos da motivos para suponer la presencia de heteroscedasticidad, tal cual examinamos en el primer epígrafe de este capítulo. Por ejemplo, ya discutimos que en los estudios sobre los ingresos familiares o sobre la productividad de las firmas al interior de una industria, la presencia de heteroscedasticidad es altamente probable.

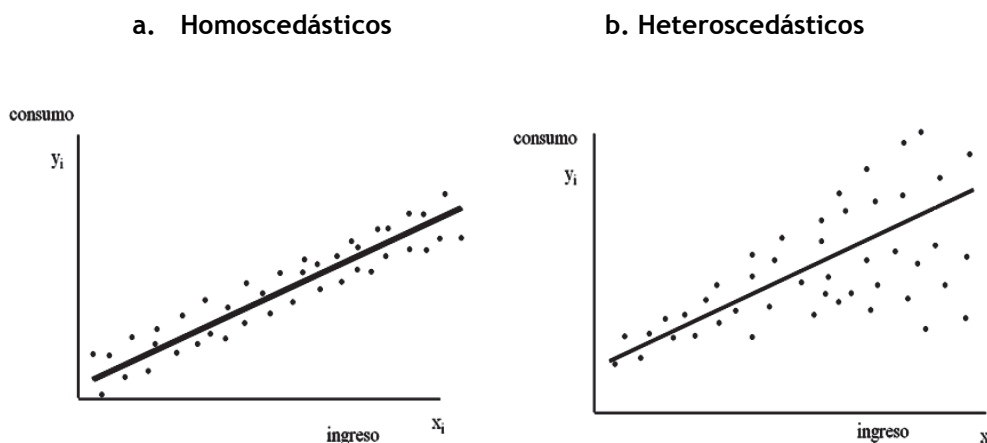
En el caso de que no tengamos pistas sobre la presencia o ausencia de este fenómeno, podemos recurrir a ciertos procedimientos para su detección. Estos procedimientos descansan, en todos los casos, en la estimación previa del MLRN y el análisis posterior del comportamiento de sus residuos o de sus residuos al cuadrado, como buenos estimadores de la varianza de cada observación.

Una forma sencilla de revelar la posible presencia de heteroscedasticidad en un modelo de regresión es graficando la función de regresión y los valores reales de la variable observada. Si estos valores se encuentran ubicados dentro de un

<sup>2</sup> Más adelante en este mismo capítulo demostramos que podemos conseguir otro estimador insesgado para  $\beta$  con menor varianza.

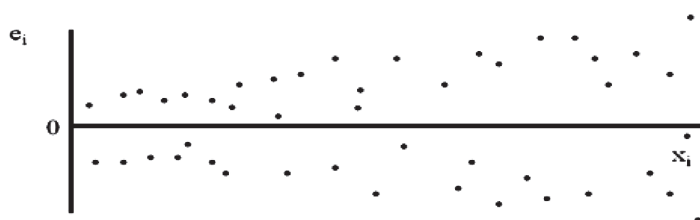
intervalo de amplitud constante a ambos lados de la función de regresión —tal y como se muestra en la figura 9.2, panel a.—, es probable que estemos en presencia de un patrón de residuos homoscedásticos. En cambio, si los valores reales se ubican en un intervalo alrededor de la función de regresión que crece (o decrece) en amplitud, entonces es posible que estemos ante un caso de residuos heteroscedásticos —ver figura 9.2, panel b). En general, un patrón sistemático de dispersión de las observaciones alrededor de la línea de regresión, es indicativo de heteroscedasticidad.

Figura 9.2: Patrones de errores



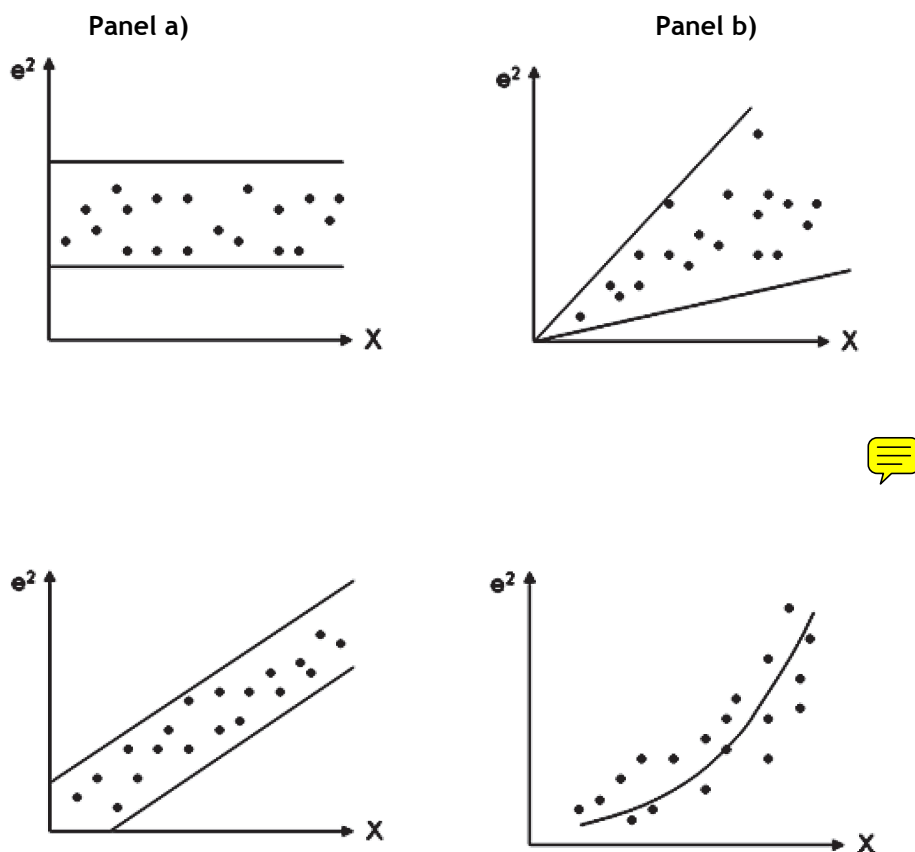
Otro procedimiento relativamente sencillo consiste en graficar directamente los residuales de una estimación mínimo-cuadrática contra las observaciones de la variable explicativa, tal y como mostramos en la figura 9.3. Cuando tenemos varias variables explicativas, podemos graficar contra el valor estimado de la variable dependiente ( $\hat{Y}$ ). Igualmente, el patrón de comportamiento nos indicará en qué caso nos encontramos: en el caso homoscedástico, estos errores deberán encontrarse la mayoría al interior de un intervalo constante alrededor del eje de las abscisas, donde ubicamos la variable independiente o el estimado de la dependiente; en caso contrario, tendremos un probable problema de heteroscedasticidad.

Figura 9.3: Residuos de la regresión



También, en lugar de graficar directamente los residuos, podemos graficar los residuos al cuadrado conjuntamente con cada variable explicativa o con el valor estimado de la media condicionada de  $Y$ . En la figura 9.4 mostramos posibles patrones a observar. Por ejemplo, en el panel a) no encontramos una relación sistemática entre los residuos al cuadrado y la variable independiente, así que no hay contradicción aparente con el supuesto de homoscedasticidad. El panel b) representa una relación más que proporcional entre ambas magnitudes; el c), una relación lineal y en crecimiento y el d), una cuadrática.

**Figura 9.4: Residuos de la regresión al cuadrado, posibles patrones**



Ilustremos la presencia de heteroscedasticidad con un ejemplo. Para ello hemos seleccionado una muestra del Anuario Estadístico de Cuba de 1988, que contiene datos para la producción bruta (a precios de 1981) de las 21 ramas que componen el sector industrial cubano, así como información sobre el promedio de trabajadores y de los fondos básicos<sup>3</sup> de cada una de ellas para el año 1987. Con estos in-

<sup>3</sup> Los fondos básicos comprenden lo que en Occidente se conoce como capital físico.

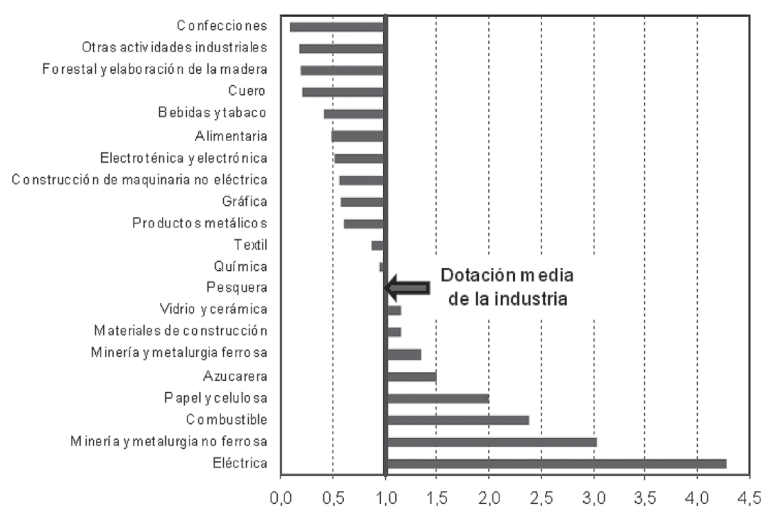
gredientes estimamos una función de producción promedio para el sector de la manufactura cubana, a partir del modelo:

$$LPB_i = \beta_1 + \beta_2 LL_i + \beta_3 LFB_i + \varepsilon_i \quad (9.2)$$

donde la “L” que antecede a las notaciones en este caso indica que trabajamos en términos logarítmicos, por lo que estamos asumiendo un modelo Cobb-Douglas y los  $\beta$ 's pueden ser interpretados como elasticidades.

La simple inspección de los datos disponibles para esas variables nos ayuda a comprender que estas industrias son heterogéneas en cuanto a su *composición orgánica*,<sup>4</sup> que representaremos simplificadaamente a través de la dotación de fondos básicos por trabajador en la figura 9.5.

**Figura 9.5: Dotación de medios básicos por trabajador en las ramas de la industria cubana, para el año 1987**



Fuente: Elaboración propia a partir de CEE, Anuario Estadístico de Cuba de 1988.

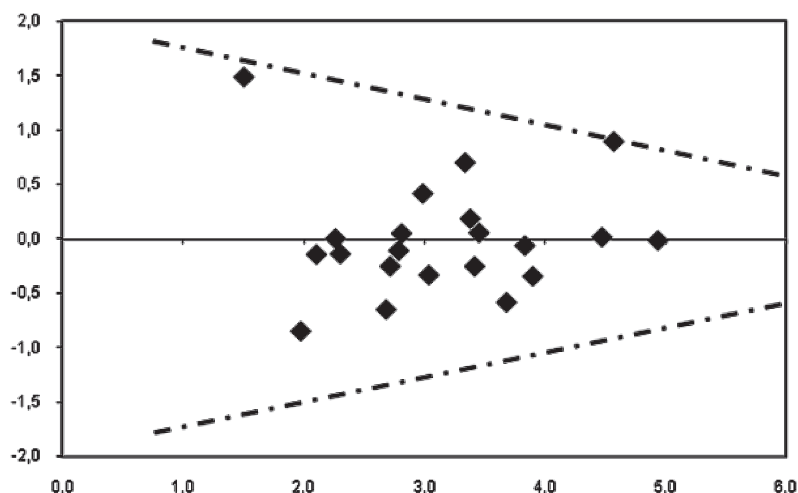
Como pueden observar, tenemos dentro del conjunto industrias que tienen una dotación de fondos por trabajador que rebasa con creces el promedio de toda la manufactura, entre las que destacan la eléctrica, la minería y la metalurgia no ferrosa y el combustible. Otro grupo está cercano a la media, como la pesquera, la química y la industria del vidrio y la cerámica. Y finalmente, están aquellas con muy baja dotación relativa, como la industria del cuero, la forestal y elaboradora de la madera, así como las confecciones. A partir de esta heterogeneidad entre individuos de esta muestra, la probabilidad de que aparezca el fenómeno de la heteroscedasticidad es elevada.

Otra evidencia sobre esto podemos encontrarla, como ya explicamos, en los gráficos de los residuos de la regresión estimada por MCO con respecto a alguna de

<sup>4</sup> Se trata de la relación entre capital constante y capital variable.

las variables explicativas de la función, como puede ser el promedio de trabajadores, como se muestra en la figura 9.6. En ella observamos unos residuos que parecen disminuir a medida que se incrementa el promedio de trabajadores en las industrias específicas, indicando un probable patrón de heteroscedasticidad.

**Figura 9.6:** Residuos del modelo (9.2) contra el logaritmo del promedio de trabajadores (variable  $LL_i$ )



Fuente: Elaboración propia

Además de estos procedimientos gráficos, también se han desarrollado algunos tests para investigar si la muestra disponible contradice el supuesto de homoscedasticidad de la varianza del error estocástico del MRLN. Comenzaremos esta discusión por el test de Goldfeld y Quandt. Este test nos permite comparar las varianzas de dos submuestras, sobre la base de la hipótesis nula que sostiene la igualdad de ambas. En consecuencia, su aplicabilidad en el caso de heteroscedasticidad de bloque es natural, aunque veremos que también podemos utilizarlo en el caso de heteroscedasticidad proporcional.

Para realizar este test seguimos el siguiente procedimiento:

- ✓ En el caso de heteroscedasticidad de bloque tenemos dos submuestras naturales sugeridas por la naturaleza del problema.
- ✓ En el caso de heteroscedasticidad proporcional, ordenamos la muestra de forma ascendente según el valor de la variable explicativa que asumimos relacionada con la varianza del término de error. Formamos una submuestra para los valores más altos de la variable y otra para los valores más bajos, pero previamente excluimos un cierto número  $c$  de observaciones centrales. Usualmente elegimos  $c = n/6$ , donde  $n$  es el tamaño de la muestra.
- ✓ Estimamos la regresión original para cada una de las submuestras aplicando MCO. Así, obtenemos las sumas de residuos al cuadrado para ambos grupos que nos permiten estimar las varianzas respectivas de cada submuestra

$(\hat{\sigma}_1^2 \text{ y } \hat{\sigma}_2^2)$ .

- ✓ Si se cumple el supuesto de homoscedasticidad, ambas varianzas no deben diferenciarse significativamente. Entonces, nuestra hipótesis nula es:  $H_0 : \sigma_1^2 = \sigma_2^2$ . Si no rechazamos que estas varianzas sean iguales, podemos decir que nuestra muestra no contradice el supuesto de homoscedasticidad.

Ahora tenemos que encontrar un estadígrafo de prueba y deducir su distribución de probabilidades. Para este problema es crítica la condición de normalidad del término de error de la regresión implícita en los supuestos del MLNR:  $\varepsilon_i \sim N(0, \sigma^2)$ . Esta distribución la podemos estandarizar dividiendo entre la desviación estándar del término de error:  $\frac{\varepsilon_i}{\sigma} \sim N(0, 1)$ . Si elevamos al cuadrado cada error estandarizado y los sumamos todos, obtendremos una variable que distribuye Chi-cuadrado con  $n$  grados de libertad.

$$\frac{\sum_{i=1}^n \varepsilon_i^2}{\sigma^2} \sim \chi_n^2 \quad (9.3)$$

Además, sabemos que el estimador insesgado de la varianza del error es:

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\sum_{i=1}^n e_i^2}{n-k}$$

de acuerdo a la ecuación (5.17), de modo que  $\sum_{i=1}^n e_i^2 = \hat{\sigma}^2 (n-k)$ .

Así, si reemplazamos en (9.3), podemos llegar a una nueva variable  $\hat{\sigma}^2 (n-k) / \sigma^2 \sim \chi_{n-k}^2$ , útil para la prueba de hipótesis que perseguimos. Para las dos submuestras tendremos correspondientemente que:

$$\frac{\hat{\sigma}_1^2 (n_1 - k)}{\sigma_1^2} \sim \chi_{n_1 - k}^2 \text{ y } \frac{\hat{\sigma}_2^2 (n_2 - k)}{\sigma_2^2} \sim \chi_{n_2 - k}^2$$

donde  $n_1$  y  $n_2$  son los tamaños de ambas submuestras. Si dividimos estas dos variables Chi-cuadrado entre sí, obtenemos una variable que distribuye  $F$ , con  $n_1 - k$  grados de libertad en el numerador y  $n_2 - k$  grados de libertad en el denominador. Es decir,

$$\frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2} \sim F_{(n_1 - k, n_2 - k)}$$

Noten que, de cumplirse la hipótesis nula de homoscedasticidad, tendríamos que



las varianzas reales de ambas submuestras serían iguales entre sí e iguales a la varianza común poblacional,  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . Entonces, nuestro estadístico de prueba será:

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{(n_1-k, n_2-k)} \quad (9.4)$$

El valor de la variable en (9.4) se compara con el valor crítico de  $F$  que corresponda al nivel de significación escogido y los grados de libertad del caso, para responder al test. Si el valor fuera mayor que el valor crítico, rechazaríamos la hipótesis nula de homoscedasticidad. De lo contrario, podremos decir que no hay evidencias en contra de ese supuesto.

Vamos a aplicar el test de Goldfeld-Quandt al ejemplo anteriormente referido de nuestra función de producción para el sector industrial cubano. Como en realidad disponemos de pocas observaciones, excluirémos una única observación central y así los dos grupos quedan compensados con diez elementos. Utilizaremos como variable de referencia para ordenar las observaciones, el promedio de trabajadores. Seguidamente mostramos los resultados que nos interesan de la estimación del modelo (9.2) para las dos submuestras a los efectos de realizar esta prueba en la tabla 9.1.

**Tabla 9.1: Resultados de las regresiones auxiliares para el test de Goldfeld-Quandt, función de producción del sector industrial cubano, 1987**

|                                       | Submuestra A | Submuestra B |
|---------------------------------------|--------------|--------------|
| SRC                                   | 3,029738     | 1,248674     |
| $N$                                   | 10           | 10           |
| $K$                                   | 3            | 3            |
| $n-k$                                 | 7            | 7            |
| $\hat{\sigma}_m^2$ , con $m = 1, 2$   | 0,43282      | 0,17838      |
| $\hat{\sigma}_1^2 / \hat{\sigma}_2^2$ | 2,42636      |              |
| $F_{(7,7)}^{0,05}$                    | 3,78704      |              |

Fuente: Elaboración propia.

Noten que el valor calculado no rebasa el valor crítico, por lo que no tenemos evidencias en contra de que se cumpla el supuesto de homoscedastidad, de acuerdo a este test.

Seguidamente abordaremos un conjunto de tests que se basan en la construcción de regresiones auxiliares con el cuadrado de los residuos estimados de la regresión,  $e_i^2$ , como variable dependiente. Las variables explicativas y formas funcionales de la relación son las que cambian de una prueba a otra. Algunas de ellas las pueden encontrar directamente en variados programas econométricos, pero en caso de no disponer de estos paquetes computacionales especializados, podrían ser construidas por ustedes mismos en una simple planilla de Excel.

Una de estas pruebas es el test de Park, quien postula que la varianza del término de error de cada observación está relacionada con alguna de las variables explicativas del modelo a través de la forma funcional:

$$\sigma_i^2 = \sigma^2 X_{ij}^{B_2} e^{v_i}$$

Observemos que en esta especificación la letra  $e$  significa la base de la función exponencial y no los residuos mínimo-cuadráticos. Estamos suponiendo que la variable  $j$  es la que influye la varianza. Tomando logaritmos en ambos lados de esta ecuación obtenemos:

$$\ln \sigma_i^2 = \ln \sigma^2 + B_2 X_{ij} + v_i$$

Entonces, el test se basa en la regresión auxiliar

$$\ln e_i^2 = B_1 + B_2 \ln X_{ij} + v_i \quad (9.5)$$

donde los  $e_i$  se refieren a los residuos mínimo-cuadráticos. Para realizar el test damos los siguientes pasos:

- ✓ Estimamos la regresión original o principal por MCO.
- ✓ Recuperamos los residuos  $e_i$  de esa regresión.
- ✓ Corremos la regresión (9.5) para cada  $X_j$  o para el estimador  $\hat{Y}_i$  si queremos considerar varias variables explicativas simultáneamente.
- ✓ Si las perturbaciones son homoscedásticas, deberíamos estimar el coeficiente  $B_2$  de (9.5) como estadísticamente igual a cero, lo que equivale a decir que el logaritmo de la varianza estimada de cada observación es igual a una constante más un término aleatorio. La hipótesis nula, que sostiene la homoscedasticidad, es entonces  $H_0 : B_2 = 0$ . Para ello aplicamos un test  $t$  convencional.
- ✓ Si no podemos rechazar la hipótesis nula, entonces nuestro término de error no contradice el supuesto de homoscedasticidad y a partir de  $B_1$  podemos conseguir un estimado de la varianza común homoscedástica  $\sigma^2$ . En caso contrario, el término de error es heteroscedástico.

Realizamos el test de Park a nuestro ejemplo, a ver si logramos detectar algún indicio de heteroscedasticidad. Estimamos el modelo (9.2) por MCO, una selección de los resultados de los resultados correspondientes pueden verla en la tabla 9.2.

Recuperamos los residuos de esa regresión y corremos la regresión auxiliar del test de Park, utilizando como variable explicativa el logaritmo del promedio de trabajadores ( $LL$ ). Los resultados que nos interesan los recogimos en la tabla 9.3.

Tabla 9.2: Resultados del modelo de regresión (9.2)

| Variables                 | Coeficientes | Errores estándar | Razones-t   | Probabilidad |
|---------------------------|--------------|------------------|-------------|--------------|
| LL                        | 0.524702     | 0.170333         | 3.080453    | 0.0064       |
| LFB                       | 0.203638     | 0.130275         | 1.563147    | 0.1354       |
| C                         | 2.954211     | 0.636330         | 4.642578    | 0.0002       |
| Estadístico $F$           |              |                  | 12.78725    | 0.000350     |
| Estadístico Durbin-Watson |              |                  | 1.520355    |              |
| $R^2$                     | 0.586914     |                  | $\bar{R}^2$ | 0.541016     |
| $n$                       | 21           |                  |             |              |

Fuente: Elaboración propia.

Tabla 9.3: Resultados de test de Park (regresión auxiliar (9.5)), aplicado al modelo (9.2)

| Variables       | Coeficientes | Errores estándar | Razones-t | Probabilidad |
|-----------------|--------------|------------------|-----------|--------------|
| LL              | -0,725435    | 0,865977         | -0,837706 | 0,4126       |
| C               | -1,397547    | 2,828166         | -0,494153 | 0,6269       |
| Estadístico $F$ |              |                  | 0,701752  | 0,412610     |

Fuente: Elaboración propia.

Como pueden apreciar, el coeficiente de regresión correspondiente a  $LL$  no resulta significativamente distinto de cero. Tampoco es significativo el test  $F$  de significación global. Por lo tanto, esta prueba no arroja evidencia en contra del supuesto de homoscedasticidad y la regresión auxiliar utilizada no es válida. Sin embargo, recordemos las advertencias que hicimos al concluir el test de Goldfeld y Quandt: podemos seguir indagando, todavía nos quedan más pruebas disponibles.

Otra de ellas es la de Glejser, quien sugiere regresar los valores absolutos de los residuos sobre la variable independiente  $X_j$  que se supone asociada a la varianza. Esta prueba es similar a la de Park, en el sentido de que asume una determinada relación funcional entre la varianza y alguna variable explicativa y realiza tests de hipótesis sobre la significación de esa relación. Tenemos varias formas funcionales propuestas para la conexión entre valor absoluto de los residuos y variables explicativas, entre ellas por ejemplo:

$$|e_i| = B_1 + B_2 X_i + v_i \quad (a)$$

$$|e_i| = B_1 + B_2 \sqrt{X_i} + v_i \quad (b) \quad (9.6)$$

$$|e_i| = B_1 + B_2 \frac{1}{X_i} + v_i \quad (c)$$

Sin embargo, algunos autores advierten que el término de error de las

regresiones en (9.6) tiene algunos problemas, tales como que su valor esperado no es precisamente cero, que presenta autocorrelación serial<sup>5</sup> y que puede padecer también del fenómeno de heteroscedasticidad. En ambos casos debemos recordar que tenemos una pérdida de eficiencia en los estimadores de la regresión auxiliar. Noten que en las dos pruebas que hemos reseñado se supone un determinado patrón de relación entre la varianza de los errores de la regresión MCO y las variables independientes de esa regresión.

Vamos a utilizar la última forma funcional (c) en (9.6) para aplicar el test de Glejser en el caso de nuestra función de producción para el sector industrial cubano. ¿Por qué escogimos precisamente la última forma funcional y no las otras propuestas en (9.6)? El gráfico 9.6 puede darles una pista al respecto. Es decir, utilizaremos como variable dependiente el valor absoluto del residuo del modelo (9.2) estimado por MCO, y como variable explicativa el recíproco de una de las variables explicativas, en particular del promedio de trabajadores. Los resultados de la regresión auxiliar están en la tabla 9.4.

**Tabla 9.4: Resultados de test de Glejser (regresión auxiliar (9.6), panel c), aplicado al modelo (9.2)**



| Variables     | Coeficientes | Errores estándar | Razones-t | Probabilidad |
|---------------|--------------|------------------|-----------|--------------|
| $1/LL$        | 1,647435     | 0,692082         | 2,380405  | 0,0279       |
| C             | -0,209136    | 0,250634         | -0,834428 | 0,4144       |
| Estadístico F |              |                  | 5,666327  | 0,0279       |

Fuente: Elaboración propia.

Noten que este test sí arroja evidencia contraria a la homoscedasticidad: obtenemos una razón-t superior a dos y su probabilidad asociada es menor que 5%. Además, el test de significación global de la regresión auxiliar también resulta significativo, por lo que este tipo de relación parece ser válida.

Uno de los test más conocidos para la detección de heteroscedasticidad es la prueba general<sup>6</sup> de White, disponible en la mayoría de los programas econométricos como una opción estándar. Para realizarlo utilizamos el siguiente procedimiento:

- Estimamos la regresión original o principal por MCO y recuperamos sus residuos  $e_i$ .
- A partir de esos residuos estimamos una regresión auxiliar que involucra como variables independientes a las explicativas del modelo original de forma lineal y también al cuadrado, así como combinaciones de a dos de esas variables. En muchos programas econométricos se pueden estimar dos variantes de esta especificación: una con las variables explicativas y sus

<sup>5</sup> Este tipo de error no esférico lo trataremos en el siguiente capítulo.

<sup>6</sup> Más adelante en este capítulo veremos por qué el calificativo de “general”.

cuadrados únicamente, y otra que incorpora adicionalmente sus combinaciones de a dos o términos cruzados, como se les denomina comúnmente.

Por ejemplo, en el caso de una regresión con sólo dos variables explicativas, la regresión auxiliar incluyendo los términos cruzados adopta la siguiente forma:

$$e_i^2 = A_1 + A_2X_{2i} + A_3X_{3i} + A_4X_{2i}^2 + A_5X_{3i}^2 + A_6X_{2i}X_{3i} + v_i$$

- Multiplicamos el  $R^2$  de la regresión auxiliar por el tamaño de la muestra,  $n$ , para construir una variable que, bajo la hipótesis nula de que no hay heteroscedasticidad, distribuye asintóticamente Chi-cuadrado, con tantos grados de libertad como regresores tengamos en el lado derecho de esa regresión (excluimos la constante):

$$nR^2 \sim \chi_{gl}^2$$

Noten que la hipótesis nula implícita en esta prueba es que los coeficientes de regresión de la regresión auxiliar son todos cero. Como es evidente, este test consume muchos grados de libertad. En nuestro ejemplo son cinco, y este número crece en la medida en que tengamos más variables explicativas. Así, debemos tener cautela en el caso de que nuestro modelo original tenga muchas variables explicativas y relativamente pocas observaciones.

- Si el valor de  $nR^2$  excede al valor de Chi-cuadrado crítico al nivel de significación con que estemos trabajando ( $\chi_{\alpha, gl}^2$ ), entonces habrá heteroscedasticidad. De lo contrario, el término de error de esta muestra no contradice la hipótesis de homoscedasticidad, lo que en nuestro ejemplo equivale a decir que los coeficientes de la regresión auxiliar cumplen la condición  $H_0 : A_2 = A_3 = A_4 = A_5 = A_6 = 0$ .

Apliquemos la prueba general de White a nuestro ejemplo, a ver si se sostiene la evidencia de heteroscedasticidad. En el caso que nos ocupa la regresión auxiliar correspondiente sería:

$$e_i^2 = A_1 + A_2LL_i + A_3LFB_i + A_4LL_i^2 + A_5LFB_i^2 + A_6LL_i \cdot LFB_i + v_i$$

Resumimos los resultados del test de White apoyados en EViews en la tabla 9.5, sin y con términos cruzados.

Como podrán observar, los resultados de las dos pruebas indican rechazo de la hipótesis nula de homoscedasticidad. Fíjense que hemos incluido el valor de la probabilidad asociada al estadístico Chi-cuadrado, lo que facilita nuestras conclusiones. También notarán que aparece, además de la prueba  $\chi^2$ , una prueba  $F$  de significación global de la regresión auxiliar de la prueba de White. Para ambas pruebas la probabilidad es inferior a un 5%, por lo que ambos estadísticos están en franca zona de rechazo. Así, esta evidencia es contraria al supuesto de homoscedasticidad. En particular vemos que los coeficientes del logaritmo del promedio de

trabajadores y del promedio de trabajadores al cuadrado son significativos, en la regresión sin términos cruzados al 5% de significación. Cuando incluimos el cruce de  $LL$  y  $LFB$  ( $LL \cdot LFB$ ), esta variable también resulta significativa.

**Tabla 9.5: Resultados del test general de White al modelo (9.2)**

| Sin términos cruzados |              |                  |           |              |
|-----------------------|--------------|------------------|-----------|--------------|
| Estadístico $F$       | 4,087578     | Probabilidad     | 0,018057  |              |
| Chi-cuadrado          | 1,061370     | Probabilidad     | 0,031266  |              |
| Variables             | Coeficientes | Errores estándar | Razones-t | Probabilidad |
| $C$                   | 1,410139     | 2,336457         | 0,603537  | 0,5546       |
| $LL$                  | -2,443328    | 0,675279         | -3,618251 | 0,0023       |
| $LL^2$                | 0,358825     | 0,109069         | 3,289887  | 0,0046       |
| $LFB$                 | 1,027166     | 0,799470         | 1,284809  | 0,2171       |
| $LFB^2$               | -0,092706    | 0,071180         | -1,302417 | 0,2112       |
| Estadístico $F$       |              |                  | 4,087578  | 0,018057     |
| $R^2$                 | 0,505414     | $\bar{R}^2$      | 0,381768  |              |
| Con términos cruzados |              |                  |           |              |
| Estadístico $F$       | 5,804452     | Probabilidad     | 0,003544  |              |
| Chi-cuadrado          | 1,384453     | Probabilidad     | 0,016628  |              |
| Variables             | Coeficientes | Errores estándar | Razones-t | Probabilidad |
| $C$                   | -0,529730    | 2,137115         | -0,247872 | 0,8076       |
| $LL$                  | -1,900242    | 0,615343         | -3,088104 | 0,0075       |
| $LL^2$                | 0,693661     | 0,159046         | 4,361375  | 0,0006       |
| $LL \cdot LFB$        | -0,448570    | 0,172364         | -2,602456 | 0,0200       |
| $LFB$                 | 1,289612     | 0,692719         | 1,861668  | 0,0824       |
| $LFB^2$               | 0,009793     | 0,072625         | 0,134840  | 0,8945       |
| Estadístico $F$       |              |                  | 5,804452  | 0,003544     |
| $R^2$                 | 0,659263     | $\bar{R}^2$      | 0,545684  |              |

Fuente: Elaboración propia.

Finalmente, queremos enfatizar que estamos realizando todas estas pruebas ***asumiendo que nuestro modelo original no contiene otros errores de especificación***. Puede suceder que el estadístico de prueba desarrollado por White sea significativo y la causa básica no sea precisamente un residuo heteroscedástico. Si, por ejemplo, en nuestro modelo hubiésemos omitido variables

relevantes para explicar el fenómeno que investigamos, estas variables estarán incorporadas al residuo, que dejaría de comportarse aleatoriamente y pudiera parecer que presenta un patrón de relación sistemática con alguna de las variables o con el propio estimado de la variable dependiente  $\hat{Y}$ . Entonces, la prueba de White tiene un alcance más general que la mera verificación de heteroscedasticidad. Con ella podemos detectar también otros errores de especificación del modelo de regresión.

## 9.4 Medidas Correctivas para la Heteroscedasticidad

¿Qué hacer ante la presencia de heteroscedasticidad? Ya vimos previamente que la heteroscedasticidad no hace mella a las propiedades de falta de sesgo y consistencia de los EMCO, pero sí afecta su eficiencia, lo que enturbia la posible inferencia estadística y, justamente ese es un objetivo básico del análisis económico al que no queremos renunciar. Por lo tanto, debemos aplicar medidas correctivas o de mitigación de este problema.

Esas medidas dependerán de cuánta información dispongamos. Supongamos que este es el único error de especificación de nuestro modelo de regresión y que conocemos la matriz de varianzas y covarianzas del término de error en la población,  $\Omega$ . En este caso tenemos disponible otro método de estimación que nos permite obtener estimadores que sí son eficientes. Se trata del método de los mínimos cuadrados generalizados (MCG).

La idea central de este método consiste en conseguir una matriz  $P$  tal que cumpla la condición:

$$P\Omega P' = I \quad (9.7)$$

Una vez obtenida esa matriz, transformamos el modelo de regresión premultiplicando por  $P$ :

$$\begin{aligned} Y &= X\beta + \varepsilon \\ PY &= PX\beta + P\varepsilon \\ Y^* &= X^*\beta + v \end{aligned} \quad (9.8)$$

Noten que con esta transformación obtenemos nuevas variables ( $Y^*$  y  $X^*$ ) y un nuevo término de error ( $v$ ), pero se mantienen los parámetros de interés de la regresión.<sup>7</sup>

Con una matriz de varianzas y covarianzas como la que mostramos en (9.1), la matriz  $P$  que necesitamos para la transformación será:

<sup>7</sup> Con excepción de la constante, que rara vez es de interés y si lo fuera, podríamos recuperarla.

$$\mathbf{P} = \begin{pmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n \end{pmatrix} = \mathbf{P}'$$

ustedes mismos podrán verificar que  $\mathbf{P}\mathbf{\Omega}\mathbf{P}' = \mathbf{I}$ .

Vamos a explorar si el término de error de esta nueva regresión tiene las propiedades necesarias para estimar por mínimos cuadrados. En primer lugar, su valor esperado será:

$$E(\mathbf{v}) = E(\mathbf{P}\boldsymbol{\varepsilon}) = \mathbf{P}E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

¿Cuál será su varianza?

$$E(\mathbf{v}\mathbf{v}') = E[(\mathbf{P}\boldsymbol{\varepsilon})(\mathbf{P}\boldsymbol{\varepsilon})'] = E(\mathbf{P}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{P}') = \mathbf{P}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{P}' = \mathbf{P}\mathbf{\Omega}\mathbf{P}' = \mathbf{I}$$

La matriz de varianzas y covarianzas de este nuevo término de error es una matriz identidad, ocupada por la unidad en su diagonal, de modo que la transformación referida indujo homoscedasticidad.

Así, tenemos un error con valor esperado cero y varianza constante, de modo que podemos aplicar mínimos cuadrados a la regresión (9.8) para obtener estimadores insesgados y también eficientes:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{MCG}} &= (\mathbf{X}^*\mathbf{X}^*)^{-1} \mathbf{X}^*\mathbf{Y}^* \\ &= ((\mathbf{P}\mathbf{X})'(\mathbf{P}\mathbf{X}))^{-1} (\mathbf{P}\mathbf{X})'(\mathbf{P}\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{Y} \end{aligned}$$

Tampoco les será difícil comprobar que, si se cumple la condición (9.7), entonces  $\mathbf{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$ , por lo que podemos reescribir la expresión anterior como:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y} \quad (9.9)$$

Como habrán podido apreciar, el método de los MCG consiste en aplicar MCO sobre variables transformadas de forma tal que para el nuevo modelo transformado se cumplen los supuestos requeridos, en particular el de homoscedasticidad. Así, los estimadores MCG son los mejores estimadores lineales insesgados de  $\boldsymbol{\beta}$  bajo condiciones de heteroscedasticidad.

Como ya referimos al inicio del capítulo, los procedimientos a aplicar dependerán de la información disponible y, lamentablemente, rara vez tenemos información sobre  $\mathbf{\Omega}$ . Por lo tanto, tendremos que hacer supuestos razonables para poder hacer transformaciones que induzcan homoscedasticidad en el modelo. Estas estimaciones, basadas en supuestos sobre la varianza del modelo, se denominan *Mínimos Cuadrados Generalizados Factibles (MCGF)*. Entre los supuestos más recurridos están que la varianza del error sea:



- **Proporcional al nivel de alguna de las variables explicativas.** Es decir, asumimos que la varianza de cada observación contiene un componente común (el factor de proporcionalidad, que denotamos con  $\sigma^2$ ) y otro que corresponde a la variable que genera la heteroscedasticidad,  $X_{ij}$ , lo que expresamos como  $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 X_{ij}$ . Así, la matriz de varianzas y covarianzas será:

$$\mathbf{\Omega} = \begin{pmatrix} \sigma^2 X_{1j} & 0 & \dots & 0 \\ 0 & \sigma^2 X_{2j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 X_{nj} \end{pmatrix}$$

En este caso, la matriz  $\mathbf{P}$  para la transformación pertinente será:

$$\mathbf{P} = \begin{pmatrix} 1/\sqrt{X_{1j}} & 0 & \dots & 0 \\ 0 & 1/\sqrt{X_{2j}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sqrt{X_{nj}} \end{pmatrix} = \mathbf{P}'$$

Esta transformación se conoce como la transformación raíz cuadrada, ya que para obtener la nueva regresión con error homoscedástico premultiplicamos el modelo por la matriz  $\mathbf{P}$  anterior, que es lo mismo que dividir las variables correspondientes a cada observación  $i$  entre la raíz cuadrada de la variable  $X_{ij}$ .

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

$$\frac{Y_i}{\sqrt{X_{ij}}} = \beta_1 \frac{1}{\sqrt{X_{ij}}} + \beta_2 \frac{X_{i2}}{\sqrt{X_{ij}}} + \dots + \beta_j \frac{X_{ij}}{\sqrt{X_{ij}}} + \dots + \beta_k \frac{X_{ik}}{\sqrt{X_{ij}}} + \frac{\varepsilon_i}{\sqrt{X_{ij}}}$$

$$\frac{Y_i}{\sqrt{X_{ij}}} = \beta_1 \frac{1}{\sqrt{X_{ij}}} + \beta_2 \frac{X_{i2}}{\sqrt{X_{ij}}} + \dots + \beta_j \sqrt{X_{ij}} + \dots + \beta_k \frac{X_{ik}}{\sqrt{X_{ij}}} + \frac{\varepsilon_i}{\sqrt{X_{ij}}}$$

$$Y_i^* = \beta_1 X_{i1}^* + \beta_2 X_{i2}^* + \dots + \beta_j X_{ij}^* + \dots + \beta_k X_{ik}^* + v_i$$

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{v}$$

Verifiquemos que, efectivamente, la varianza del nuevo error  $v_i$  sea una constante.

$$Var(v_i) = Var\left(\frac{\varepsilon_i}{\sqrt{X_{ij}}}\right)^2 = \frac{1}{X_{ij}} Var(\varepsilon_i) = \frac{1}{X_{ij}} \sigma^2 X_{ij} = \sigma^2$$

Como pueden apreciar, la varianza de  $v_i$  cumple la condición que se exige

para una estimación usando MCO.

El método aplicado en esta situación, y en las que describimos más adelante, constituye un caso particular del método de los MCG, que denominamos mínimos cuadrados ponderados (MCP). Los ponderadores son precisamente los elementos de la diagonal de la matriz  $\mathbf{P}$ , que en este caso son las correspondientes  $1/\sqrt{X_{ij}}$ .

- **Proporcional al cuadrado de alguna de las variables explicativas.** En este caso asumimos también que la varianza de cada observación tiene un componente común (el factor de proporcionalidad,  $\sigma^2$ ) y su otro componente es el cuadrado de la variable,  $X_j^2$ , que genera la heteroscedasticidad. En términos matemáticos, expresamos esta relación como  $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 X_{ij}^2$ . La matriz de varianzas y covarianzas será como sigue:

$$\mathbf{\Omega} = \begin{pmatrix} \sigma^2 X_{1j}^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 X_{2j}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 X_{nj}^2 \end{pmatrix}$$

Para transformar el modelo en uno con término de error homoscedástico, necesitamos una matriz que consiga eliminar la heteroscedasticidad antes descrita. La matriz  $\mathbf{P}$  siguiente cumple ese requisito:

$$\mathbf{P} = \begin{pmatrix} 1/X_{1j} & 0 & \cdots & 0 \\ 0 & 1/X_{2j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/X_{nj} \end{pmatrix} = \mathbf{P}'$$

Siguiendo el procedimiento de los MCP, transformamos la regresión multiplicando por la matriz  $\mathbf{P}$ , que es lo mismo que dividir las variables correspondientes a cada observación  $i$  entre el nivel de la variable  $X_{ij}$ .

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

$$\frac{Y_i}{X_{ij}} = \beta_1 \frac{1}{X_{ij}} + \beta_2 \frac{X_{i2}}{X_{ij}} + \cdots + \beta_j \frac{X_{ij}}{X_{ij}} + \cdots + \beta_k \frac{X_{ik}}{X_{ij}} + \frac{\varepsilon_i}{X_{ij}}$$

$$\frac{Y_i}{X_{ij}} = \beta_1 \frac{1}{X_{ij}} + \beta_2 \frac{X_{i2}}{X_{ij}} + \cdots + \beta_j + \cdots + \beta_k \frac{X_{ik}}{X_{ij}} + \frac{\varepsilon_i}{X_{ij}}$$

$$Y_i^* = \beta_1 X_{i1}^* + \beta_2 X_{i2}^* + \cdots + \beta_j X_{ij}^* + \cdots + \beta_k X_{ik}^* + v_i$$

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{v}$$

¿Cuál será la varianza del nuevo error  $v_i$ ?

$$Var(v_i) = Var\left(\frac{\varepsilon_i}{X_{ij}}\right)^2 = \frac{1}{X_{ij}^2} Var(\varepsilon_i) = \frac{1}{X_{ij}^2} \sigma^2 X_{ij}^2 = \sigma^2$$

Igualmente, comprobamos que la varianza de  $v_i$  es una constante,  $\sigma^2$ .

- **Hay dos submuestras con varianzas diferentes (heteroscedasticidad de bloque).** En este caso suponemos que la muestra puede ser dividida en dos submuestras con  $n_1$  y con  $n_2$  observaciones respectivamente. Suponemos que  $Var(\varepsilon_i) = \sigma_1^2$  para  $i = 1, \dots, n_1$  y  $Var(\varepsilon_i) = \sigma_2^2$  para  $i = n_1 + 1, \dots, n$ . La matriz  $\mathbf{P}$  adecuada es, en este caso:

$$\mathbf{P} = \begin{pmatrix} 1/\hat{\sigma}_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 1/\hat{\sigma}_1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/\hat{\sigma}_2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1/\hat{\sigma}_2 \end{pmatrix} = \mathbf{P}'$$

Esto es, una matriz diagonal que contiene el estimador mínimo-cuadrático de la desviación estándar de la primera submuestra, y el mismo estimador para la segunda submuestra. La regresión a estimar después de la transformación recomendada será:

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

$$\frac{Y_i}{\hat{\sigma}_1} = \beta_1 \frac{1}{\hat{\sigma}_1} + \beta_2 \frac{X_{i2}}{\hat{\sigma}_1} + \cdots + \beta_k \frac{X_{ik}}{\hat{\sigma}_1} + \frac{\varepsilon_i}{\hat{\sigma}_1}, \quad i = 1, \dots, n_1,$$

$$\frac{Y_i}{\hat{\sigma}_2} = \beta_1 \frac{1}{\hat{\sigma}_2} + \beta_2 \frac{X_{i2}}{\hat{\sigma}_2} + \cdots + \beta_k \frac{X_{ik}}{\hat{\sigma}_2} + \frac{\varepsilon_i}{\hat{\sigma}_2}, \quad i = n_1 + 1, \dots, n$$

$$Y_i^* = \beta_1 X_{i1}^* + \beta_2 X_{i2}^* + \cdots + \beta_j X_{ij}^* + \cdots + \beta_k X_{ik}^* + v_i$$

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{v}$$

Fácilmente podemos ver que el modelo transformado es homoscedástico.

Otra posibilidad a nuestra disposición para mitigar la presencia de heteroscedasticidad, es transformar las variables aplicándoles logaritmos. La transformación logarítmica logra reducir la variabilidad.

Finalmente, White desarrolló un método de estimación de una varianza consistente con heteroscedasticidad, que nos permite realizar tests de significación adecuados a los EMCO en el caso de grandes muestras. Este método está disponible en los paquetes computacionales econométricos, por lo que resulta sencillo de aplicar.

La idea central del procedimiento de White consiste en estimar el modelo de regresión sospechoso de heteroscedasticidad por MCO y utilizar los  $e_i^2$ , en lugar de los verdaderos valores de  $\sigma_i^2$  que desconocemos, para estimar la varianza de los coeficientes de regresión. A los errores estándar así obtenidos los denominamos errores estándar robustos de White.

Entonces, este procedimiento calcula las varianzas de los coeficientes de regresión estimados para un modelo como el que presentamos en (5.1) según la fórmula:

$$Var(\hat{\beta}_s) = \frac{\sum_{i=1}^n \hat{W}_{is}^2 e_i^2}{\left( \sum_{i=1}^n \hat{W}_{is}^2 \right)^2} \quad (9.10)$$

donde, como ya sabemos,  $e_i$  son los residuos estimados del modelo (5.1) y  $\hat{W}_{is}$ , los residuos estimados de la regresión auxiliar de  $X_s$  sobre el resto de las variables explicativas:

$$X_{is} = \alpha_1 + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \cdots + \alpha_{s-1} X_{i,s-1} + \alpha_{s+1} X_{i,s+1} + \cdots + \alpha_k X_{ik} + W_{is}, \text{ para } i = 1, \dots, n$$

Es obvio que en el caso de sólo dos variables (una explicativa y una dependiente,  $X$  y  $Y$ ), (9.10) se calcula directamente con  $X_i$ .

En cualquier caso, si no contaran con un programa econométrico que tuviese incorporada esta facilidad de estimar varianzas de los estimadores consistentes con heteroscedasticidad, o aun si no tuviesen un programa econométrico especializado disponible, con estas indicaciones ustedes bien pueden arreglárselas para hacer este tipo de estimación de la varianza de los  $\hat{\beta}_j$ .

Aunque este procedimiento de estimación de varianzas consistentes con heteroscedasticidad es válido sólo para muestras grandes, aplicamos el estimador de White, con propósito ilustrativo, a nuestro ejemplo de la función de producción para la industria cubana y compararemos sus resultados con los obtenidos a partir de la estimación por MCO. Las estimaciones las resumimos en la tabla 9.6.

¿Cuáles son los hallazgos más relevantes? En primer lugar, no hay diferencia en los coeficientes de regresión estimados, ya que ambos procedimientos aplican MCO a estos fines. La distinción entre ellos radica en que el procedimiento de White toma en cuenta la posible heteroscedasticidad del término de error para la estimación de la varianza y el error estándar de los coeficientes de regresión estimados, por lo que los  $se(\hat{\beta}_j)$  difieren de los que obtenemos cuando obviamos la heteroscedasticidad. Recuerden que en la EMCO que obvia la heteroscedasticidad obtendremos varianzas sesgadas para los  $\hat{\beta}_j$  y no sabemos si las estamos sobreestimando o subestimando.

**Tabla 9.6: Estimaciones del modelo (9.2) con MCO y MCO con errores y varianzas consistentes con heteroscedasticidad (EVCHS), según el método de White**

|                                                   | LL       | LFB      | C        |
|---------------------------------------------------|----------|----------|----------|
| Coeficientes de regresión                         | 0,524702 | 0,203638 | 2,954211 |
| Errores estándar de los coeficientes de regresión |          |          |          |
| EMCO                                              | 0,170333 | 0,130275 | 0,636330 |
| EMCO + EVCHS (White)                              | 0,258903 | 0,113843 | 0,583053 |
| Razones-t                                         |          |          |          |
| EMCO                                              | 3,080453 | 1,563147 | 4,642578 |
| EMCO + EVCHS (White)                              | 2,026634 | 1,788771 | 5,066796 |
| Probabilidad                                      |          |          |          |
| EMCO                                              | 0,0064   | 0,1354   | 0,0002   |
| EMCO + EVCHS (White)                              | 0,0578   | 0,0905   | 0,0001   |

Fuente: Elaboración propia

En el caso que nos ocupa habíamos detectado que, al parecer, es la variable *LL* la que genera el problema de heteroscedasticidad.<sup>8</sup> Esta variable, que resultaba significativa al 1% en la estimación MCO sencilla, con la corrección de la varianza del estimador consistente con heteroscedasticidad ya no lo es más ni siquiera al 5% de significación: la razón-t asociada a su coeficiente de regresión disminuye de algo más que tres a dos, y sólo podríamos considerarla estadísticamente distinta de cero para niveles de significación mejores que 5,78%. Por otra parte, la variable *LFB*, que resultaba no significativa en la estimación MCO sencilla, en la variante de White sí lo sería al 10%.

Como ya advertimos al aplicar el procedimiento de White a nuestro ejemplo, sus resultados sólo son válidos asintóticamente, lo que no es precisamente el caso de la muestra que nos ocupa. No obstante, con esta presentación hemos querido enfatizar en los efectos que podría tener no tomar en cuenta la verdadera especificación de la varianza del término de error en la inferencia a partir de EMCO.

En este epígrafe hemos reseñado un conjunto de variantes para circunvalar el problema de la heteroscedasticidad. ¿Cuál de ellas debemos aplicar? La respuesta no es trivial. Tendremos ineludiblemente que transformar las variables si estamos ante un caso severo. Pero si logramos estimar los regresores del modelo original por MCO con suficiente significación (altas razones-t), no necesariamente debemos intentar corregir este error de especificación. En definitiva, sabemos que aún bajo esta condición, los EMCO son insesgados y consistentes. Si fuera absolutamente necesario intervenir, entonces los gráficos de los residuos al cuadrado contra cada una de las variables explicativas o contra  $\hat{Y}_i$  nos pueden dar pistas sobre la relación entre ellos y, por lo tanto, sobre la transformación a realizar. Además, por lo general tenemos disponible el método de White para conseguir estimadores de las

<sup>8</sup> Comparen los resultados que obtuvimos previamente con los tests de Glejser y de White.

varianzas consistentes con heteroscedasticidad, que también nos pueden ayudar a comprobar si vale la pena atender el problema. 

# *preguntas*

Trate de responder las preguntas a continuación. Con ello comprobará si ha captado e incorporado los contenidos de este capítulo.

1. ¿Qué es heteroscedasticidad? Contraste la matriz  $E(\varepsilon \cdot \varepsilon')$  en el caso heteroscedástico con la correspondiente al supuesto de homoscedasticidad.
2. Mencione las principales causas de la heteroscedasticidad en un modelo de regresión.
3. ¿Qué tipos de heteroscedasticidad encontramos y en qué se diferencian entre sí?
4. ¿Cómo podemos detectar la presencia de heteroscedasticidad en la perturbación estocástica de un modelo de regresión? Explique según los tipos de heteroscedasticidad que mencionó en la pregunta tres.
5. Suponga que hemos estimado un modelo de regresión a través del método de MCO y detectamos que los residuos se comportan con un patrón sistemático que depende de una de las variables explicativas del modelo. ¿Podremos confiar en los estimados obtenidos para  $\hat{\beta}$  y  $\sigma^2$ ?
6. ¿Cuáles medidas correctivas podemos aplicar en el caso de que detectemos heteroscedasticidad en la perturbación estocástica del modelo de regresión?
7. Explique en qué consiste el método desarrollado por White para conseguir estimadores de las varianzas consistentes con heteroscedasticidad.



# ejercicios

9.1 Una bien conocida hipótesis sugiere que existe una fuerte relación positiva entre ingreso y gasto público en educación. Es decir, que el esfuerzo de un país en educación se relaciona con el ingreso. Hemos recolectado datos acerca de los ingresos per cápita y del gasto público en educación de 34 países para el año 1980. Con ellos, estimamos el siguiente modelo:

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

donde  $Y_i$  es el gasto público en educación del país  $i$ , y  $X_i$  representa el ingreso per cápita del mismo país. La siguiente tabla presenta los resultados de la estimación MCO:

## ANOVA

|           | gl | Suma de Cuadrados (SC) | Suma media (SM) | Estadístico-F |
|-----------|----|------------------------|-----------------|---------------|
| Regresión | 1  | 3.68385                | 3.68385         | 199.589       |
| Residuos  | 34 | 0.59063                | 0.01737         |               |
| Total     | 35 | 4.27449                | Prob(>F)        | 0.000         |

## Estimados

|                    | Coefficientes | Errores estándar |
|--------------------|---------------|------------------|
| Intercepto         | -0.12457      | 0.04852          |
| Ingreso per cápita | 0.07317       | 0.00517          |

- ¿Son significativos los coeficientes estimados al 1% de nivel de significación? ¿Lo son con un nivel de significación del 5%? ¿Qué puede Ud. decir de la hipótesis detrás de este modelo?
- Calcule el  $R^2$ .
- ¿Qué puede Ud. decir de este modelo a partir de la observación del estadístico F?

No obstante, este enfoque es frecuentemente criticado, arguyendo que los países ricos y pobres se comportan de manera diferente en lo que respecta a la asignación de sus gastos en educación. Por lo tanto, decidimos estimar por separado modelos para países ricos y pobres. Las siguientes tablas muestran nuestros resultados:



**Países pobres****ANOVA**

|           | gl | SC     | SM       | Estadístico-F |
|-----------|----|--------|----------|---------------|
| Regresión | 1  | 0.1422 | 0.1422   | 18.5882       |
| Residuos  | 16 | 0.1224 | 0.0077   |               |
| Total     | 17 | 0.2647 | Prob(>F) | 0.000         |

**Estimados**

|                    | Coeficientes | Errores estándar |
|--------------------|--------------|------------------|
| Intercepto         | -0.0381      | 0.0549           |
| Ingreso per cápita | 0.0505       | 0.0121           |

**Países ricos****ANOVA**

|           | gl | SC     | SM       | Estadístico-F |
|-----------|----|--------|----------|---------------|
| Regresión | 1  | 0.3539 | 0.3539   | 12.9604       |
| Residuos  | 16 | 0.4369 | 0.0273   |               |
| Total     | 17 | 0.7908 | Prob(>F) | 0.003         |

**Estimados**

|                    | Coeficientes | Errores estándar |
|--------------------|--------------|------------------|
| Intercepto         | -0.1399      | 0.2817           |
| Ingreso per cápita | 0.0751       | 0.0223           |

- d. Discuta la significación de los parámetros de ambos modelos al 1% y 5% de nivel de significación.
- e. Compare los  $R^2$  de ambos modelos.
- f. Use el test de Goldberg-Quandt para investigar si estos datos sugieren que las varianzas son iguales o diferentes en ambas submuestras.
- g. Bajo el supuesto de diferentes varianzas, evalúe la hipótesis de que los países ricos se comportan diferente a los países pobres (test de Chow o test-F).
- h. ¿Cuáles resultados en c) sugerirían heteroscedasticidad para nuestro primer modelo? ¿Cómo lo estimaría Ud., entonces?
- i. Se sugiere que los errores de este modelo están fuertemente relacionados con el tamaño del ingreso per cápita. Es decir, podemos asumir que para

cada país tenemos que  $\sigma_i^2 = X_i \sigma^2$ , donde  $\sigma^2$  es la varianza general subyacente. ¿Cómo afecta esta situación nuestros estimados? ¿Cómo estimaría Ud. el modelo en este caso?

9.2 Utilizando la función de producción tipo Cobb-Douglas que estimó en el ejercicio 8.5.2 para la producción de caña en Cuba (ver Tabla B.8 en el CD acompañante):

- Investigue si este modelo cumple el supuesto de homoscedasticidad. Utilice el test de White.
- Si no se cumple ese supuesto, ante qué tipo de heteroscedasticidad nos encontramos en este caso.
- ¿Cómo estimaría Ud. este modelo para tomar en cuenta los efectos de la heteroscedasticidad? Realice la estimación correspondiente.
- Compare las tablas de resultados de la estimación de la función de producción a través de MCO y la obtenida en el inciso c).

9.3 La siguiente ecuación estima el consumo per cápita de carne de res en los Estados Unidos para los años 1960-1987 (errores estándar entre paréntesis) utilizando mínimos cuadrados generalizados para corregir heteroscedasticidad:

$$\hat{B}_t = -193.3 + 35.2 \ln Y_t - 0.38 PB_t + 0.10 PP_t - 5.7 D_t.$$

(14.1)                      (0.10)                      (0.09)                      (3.9)

$$\bar{R}^2 = 0.857, \quad n = 28, \quad \hat{\rho} = 0.82$$

Donde:

|             |   |                                                                                                                                                                                                                     |
|-------------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\hat{B}_t$ | : | El consumo anual per cápita de carne de res en los Estados Unidos en el año $t$ , medido en libras.                                                                                                                 |
| $\ln Y_t$   | : | El logaritmo del ingreso real disponible per cápita anual en los Estados Unidos, medido en dólares de 1982.                                                                                                         |
| $PB_t$      | : | El precio minorista promedio real anualizado de la carne de res, medido en centavos de dólar por libra.                                                                                                             |
| $PP_t$      | : | El precio minorista promedio real anualizado de la carne de cerdo, medido en centavos de dólar por libra.                                                                                                           |
| $D_t$       | : | Variable muda igual a 1 para los años posteriores a 1981, 0 en otro caso; captura el incremento del conocimiento de los consumidores acerca de los daños que puede provocar para la salud el consumo de carne roja. |

- ¿Qué significa haber estimado utilizando MCG? ¿Cuál podría ser la fuente de la heteroscedasticidad en este problema específico?
- Evalúe la significación de cada coeficiente de pendiente al 5 y 10 por ciento de nivel de significación.
- ¿Es la carne de cerdo un competidor serio de la carne de res? ¿Por qué?

- d. En 1987 el consumo de carne de res fue de 103.3 libras, mientras que el log del ingreso disponible fue igual a 9.30. El precio de la carne de res fue de 61.1 centavos por libra y de 27.3 centavos para la de cerdo. Estime el residuo de la ecuación para este año.
  - e. Suponga que en 1988 el log del ingreso disponible ascendió a 9.42; el precio de la carne de res a 59.7 centavos por libra, mientras el precio del cerdo se incrementó a 28 centavos por libra. Estime el consumo de carne de res para ese año.
  - f. Estime la elasticidad-precio de la carne de res cuando su precio es de 61 centavos por libra y el consumo per cápita de 100 libras al año.
  - g. Discuta el efecto del ingreso en el consumo de carne de res.
- 9.4 Utilizando los datos correspondientes a la encuesta de hogares aludida en el ejercicio 7.6, estime la especificación del modelo de salarios indicada en el inciso a) de dicho ejercicio para dos submuestras: una correspondiente a hombres y otra a mujeres.
- a. Investigue si es posible suponer que los modelos estimados para cada submuestra comparten la misma varianza.
  - b. En caso de que rechace la hipótesis correspondiente al inciso anterior, entonces indague sobre cuál podría ser la causa de esta diferencia entre varianzas por submuestras.
  - c. En caso de que rechace la hipótesis correspondiente al inciso a), indique de qué manera corregiría Ud. ese problema.

CAPÍTULO 10



# AUTOCORRELACIÓN

En este capítulo flexibilizamos otro de los supuestos del Modelo Lineal Normal de Regresión: que los errores del modelo no estén correlacionados. En el capítulo cinco de este texto referimos que la covarianza de los términos de error correspondientes a dos observaciones diferentes debía ser cero:

$$\text{Covar}(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i \cdot \varepsilon_j] = 0, \text{ para todo } i \neq j$$

Ello significa que el valor del error de una determinada observación no está afectado por el valor del error en otra observación.

## 10.1 Autocorrelación

Cuando la covarianza es distinta de cero diremos que estamos ante el error no esférico de **autocorrelación**:

$$\text{Covar}(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i \cdot \varepsilon_j] \neq 0, \text{ para algún } i \neq j$$

La violación de tal supuesto implica que las observaciones no serán más independientes unas de otras.

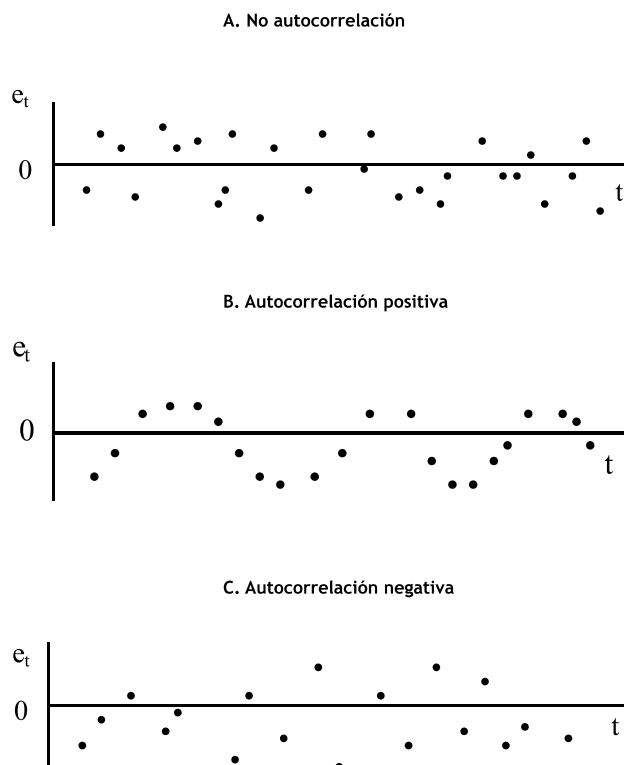
Cuando la autocorrelación ocurre en modelos de series de tiempo también se le suele denominar **correlación serial**. Por su parte, en los modelos con datos de corte transversal, se le nombra **correlación espacial**.

La autocorrelación es más común en los modelos de series de tiempo que en los de corte transversal. Los datos de corte transversal usualmente se recopilan

sobre la base de una muestra aleatoria de unidades (individuos, empresas, países, etc.) en un momento determinado del tiempo, por lo que no existe a priori una sospecha de que estén relacionadas unas con otras. De hecho, en el corte transversal el orden de los datos no tiene ninguna importancia. En cambio, en las series de tiempo el orden de los datos sí importa, precisamente porque esperamos que los datos de hoy tengan que ver con los de ayer, y ayuden, a su vez, a explicar las observaciones futuras. La autocorrelación es un fenómeno casi natural en las series de tiempo.

En la figura 10.1 se muestran algunos patrones de autocorrelación que se pueden observar al graficar el término de error en el tiempo. En la figura B se aprecia un patrón cíclico que produce una autocorrelación positiva. En determinados períodos los sucesivos errores tienden a crecer y en otros a decrecer. Hay rachas de errores siempre mayores que el anterior y rachas con errores menores que el anterior. Esto es, el error tiene un patrón sistemático en el cual depende positivamente del error anterior. En la figura C sucede lo opuesto, en la mayoría de los casos los errores positivos van seguido de errores negativos; casi siempre, a un aumento del error le sucede una disminución del mismo, esto es, están correlacionados negativamente. En la figura A se muestra una serie de errores que no presentan un patrón sistemático, sino que son aleatorios.

Figura 10.1 Término de error en el tiempo



Otra forma sencilla de ver la autocorrelación es por la cantidad de veces que los errores cruzan la línea de la ordenada cero. En la figura B la autocorrelación

positiva conduce a que los errores crucen muy pocas veces la línea cero. En la figura C sucede lo contrario, los errores cruzan muchas veces la línea cero debido a la autocorrelación negativa. El patrón de la figura A, donde no hay autocorrelación, se ubica en un punto intermedio, ni muchas ni pocas veces se cruza el eje cero. Por supuesto, todo esto es sólo para tener una intuición sobre la autocorrelación y poder usar el examen gráfico como una primera impresión sobre el cumplimiento de dicho supuesto, más adelante presentaremos pruebas estadísticas formales para contrastar autocorrelación.

¿Cuáles son las causas que pueden provocar autocorrelación en los residuos de un modelo? Pueden existir varias, algunas de ellas son las siguientes:

- *Inercia en las variables:* Ello ocurre debido a la importancia de los hábitos o determinadas reglas en el comportamiento y las decisiones de los agentes económicos. Los ciclos económicos también provocan inercia en las variables; en momentos de expansión los indicadores económicos tienden a moverse en este mismo sentido, mientras que en recesión, la mayoría de ellos tiende a acompañar la dinámica contractiva de la economía.
- *Rezagos:* Las empresas e individuos acostumbran a corregir sus decisiones sobre la base de la nueva información que reciben y sus expectativas con respecto al curso de la economía y la política económica. Existe por tanto un proceso gradual de aprendizaje que puede marcar un patrón de autocorrelación en las series. La incorporación paulatina de las nuevas tecnologías y la rigidez de algunos mercados también inducen a que los valores presentes de las variables dependan de sus observaciones pasadas.
- *Manipulación de los datos:* Existen diferentes métodos, que veremos en la parte cinco del libro, para transformar las series de tiempo, por ejemplo, para obtener datos desestacionalizados o la tendencia-ciclo. Estos y otros métodos que se usan para alisar o transformar las series de tiempo pueden provocar autocorrelación en la serie resultante.
- *Sesgo de especificación:* Esta se diferencia de las anteriores en que no es una autocorrelación pura, sino que aparece en los residuos de un modelo mal especificado. El problema de especificación puede deberse a la omisión de variables relevantes, cuya ausencia en el modelo es asumida por el término de error, provocando en éste un patrón sistemático. La otra causa del sesgo de especificación es usar una forma funcional incorrecta en el modelo.

Veamos dos ejemplos. Supongamos que el modelo verdadero para la demanda de importaciones es:

$$\dot{m}_t = \beta_0 + \beta_1 \dot{y}_t + \beta_2 \dot{p}_t + \varepsilon_t$$

pero estimamos el modelo excluyendo los precios:

$$\dot{m}_t = \beta_0 + \beta_1 \dot{y}_t + u_t$$

En consecuencia, el término de error será:

$$u_t = \beta_2 \dot{p}_t + \varepsilon_t$$

Así, en la medida que los precios afectan la demanda de importaciones, el término de error  $u_t$  presentará un patrón sistemático.

El otro ejemplo, ahora de sesgo por una incorrecta especificación, lo construimos suponiendo que el modelo verdadero para la función de producción industrial es:

$$LPB_t = \beta_1 + \beta_2 LL_t + \beta_3 LFB_t + \beta_4 LFB_t^2 + \varepsilon_t$$

pero estimamos un modelo con una especificación diferente:

$$LPB_t = \beta_1 + \beta_2 LL_t + \beta_3 LFB_t + u_t$$

Como resultado, el término de error ostentará un patrón sistemático dado por:

$$u_t = \beta_4 LFB_t^2 + \varepsilon_t$$

Ambos sesgos conducen a una falsa autocorrelación, es una autocorrelación que no estaba en el modelo verdadero y que estamos generando artificialmente. Sobre esta última causa de autocorrelación llamaremos la atención más adelante en este capítulo puesto que tiene consecuencias más graves sobre la modelización econométrica que el resto.

Existen diferentes **órdenes de autocorrelación** en dependencia de hasta cuántos períodos hacia atrás están correlacionados los residuos. Cuando hablamos de autocorrelación de primer orden nos referimos a un proceso generado de la siguiente forma:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

donde  $\rho$  es el **coeficiente de autocorrelación** ( $|\rho| \leq 1$ ) y  $u_t \sim N(0, \sigma^2)$ . Tal proceso también se conoce como autorregresivo de orden 1 o AR(1), en él los errores hoy están correlacionados con los errores un período hacia atrás. Igualmente, pueden existir en los errores autocorrelación de segundo orden:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t$$

En esta correlación serial estamos asumiendo que también influyen sobre el presente los acontecimientos de hace dos períodos atrás. Asimismo, se pudiera seguir extendiendo el proceso hasta el orden  $k$ , es decir, considerando hasta  $k$  rezagos de  $\varepsilon_t$ :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_k \varepsilon_{t-k} + u_t$$

El coeficiente de autocorrelación es simplemente una variación del coeficiente de correlación que conocemos de la estadística descriptiva y vimos en el capítulo dos. Sabemos que la covarianza se utiliza para medir el grado de asociación lineal entre dos series, y que al dividir la covarianza entre la varianza llegamos al coeficiente de correlación. Ahora bien, con el objetivo de medir la relación que los



errores tienen con ellos mismos en períodos anteriores, se usa la medida de autocovarianza. Tendremos varios valores para la autocovarianza, para así medir la relación entre los errores y ellos mismos con un rezago, con dos rezagos, con tres rezagos, y así sucesivamente. De esta manera, se forma una **función de autocovarianza** que depende del rezago  $k$ . De la misma forma que en la estadística descriptiva, ahora dividimos la autocovarianza entre la varianza del proceso para llegar al coeficiente de autocorrelación, el cuál también depende del rezago, u orden,  $k$ . Así, obtenemos la **función de autocorrelación (FAC)** que denotamos como  $\rho_k$  y que mide la correlación entre  $\varepsilon_t$  y  $\varepsilon_{t-k}$ :

$$\rho_k = \frac{\text{autocovarianza rezago } k}{\text{varianza}} = \frac{\sum_{t=k+1}^n (\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t-k} - \bar{\varepsilon})}{\sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^2}$$

donde  $n$  representa la cantidad de observaciones,  $\bar{\varepsilon}$  es el valor medio de los errores,  $|\rho_k| \leq 1$  y  $\rho_0 = 1$

En la práctica, como no podemos observar  $\varepsilon_t$ , tendremos una FAC estimada, es decir computamos  $\hat{\rho}_k$  con la estimación que hacemos de  $\varepsilon_t$  ( $e_t$ ) una vez que corremos el modelo de regresión con los datos disponibles de las variables involucradas.

## 10.2 Consecuencias de la Autocorrelación

Cuando el término de error presenta autocorrelación la matriz de varianzas y covarianzas no es una matriz diagonal, sino que tiene elementos no nulos fuera de la diagonal principal:

$$E(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}') = \begin{pmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma^2 \end{pmatrix} = \Omega$$

Las consecuencias sobre los estimadores MCO son similares a la violación del supuesto de homocedasticidad visto en el capítulo anterior. Los estimadores MCO de  $\boldsymbol{\beta}$  van a seguir siendo insesgados, sin embargo, los estimadores de la varianza de  $\boldsymbol{\beta}$  sí serán sesgados. De esta forma se afectan los procedimientos convencionales de inferencia estadística.

Cuando hay autocorrelación, la varianza de  $\boldsymbol{\beta}$  no será más:

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

sino que tendrá una expresión diferente que indica que la formula convencional puede subestimar la varianza de  $\beta$  en presencia de autocorrelación:

$$Var(\hat{\beta}) = (X'X)^{-1} X' \Omega X (X'X)^{-1}$$

Adicionalmente, la autocorrelación también afecta al estimador de la varianza de los residuos:

$$\hat{\sigma}^2 = \frac{e'e}{n-k}$$

Tal expresión subestima la verdadera varianza de los residuos  $\sigma^2$ , con lo cual, se sobreestima el  $R^2$ .

La subestimación de las varianzas de  $\beta$  y de los residuos, llevan a inflar el valor calculado de la razón  $t$  y a sobreestimar la significación estadística de  $\beta$ . En general, las pruebas de significación  $t$  y  $F$  dejan de ser validas.

Esto es lo esencial en cuanto a las consecuencias sobre los procedimientos de inferencia estadística. Con respecto a las propiedades de los estimadores MCO, si bien la autocorrelación no afecta la insesgadez, ella sí afecta la propiedad de mínima varianza. Los estimadores MCO dejan de ser MELI cuando hay autocorrelación, pues existen otros estimadores lineales insesgados con menor varianza: los estimadores mínimos cuadrados generalizado que vimos en el capítulo anterior.

Al igual que con la heteroscedasticidad, debemos obtener una matriz  $P$  tal que cumpla la condición  $P\Omega P' = I$ . Una vez definida dicha matriz, transformamos el modelo de regresión premultiplicando por  $P$ :  $PY = PX\beta + P\varepsilon$ , del cual se deriva la formulación de los estimadores MCG:  $\hat{\beta}_{MCG} = (X'P'PX)^{-1} X'P'PY$ .

La matriz  $P$  tendrá información sobre la autocorrelación. El punto clave es que esta información es considerada en la formulación de los MCG pero no está incluida en la formulación de los MCO. Por esta razón los estimadores MCG tendrán menor varianza. Los estimadores MCG hacen un mejor uso de la información disponible y en consecuencia son más precisos. Sobre la aplicación de los estimadores MCG ante la presencia de autocorrelación volveremos en la sección siguiente 10.3.

Hasta ahora hemos analizado las consecuencias de la autocorrelación pura, es decir, la autocorrelación que está en el modelo verdadero. Pero señalamos en la sección 10.1 que el sesgo de especificación, ya sea por variables omitidas o por una equivocada forma funcional, lleva a generar una falsa autocorrelación. Si la autocorrelación que observamos tiene este origen, entonces los estimadores MCO de  $\beta$  sí serán sesgados. En el capítulo 13 analizaremos en mayor detalle el tema de los errores de especificación, así como, las pruebas estadísticas y medidas correctivas asociadas al mismo. Por ahora, es suficiente con conocer que cuando observamos autocorrelación debemos manejar como una posibilidad que el origen

sea un sesgo de especificación, con lo cual, las implicaciones son más graves sobre los estimadores.

### 10.3 Detección de la Autocorrelación

Como en la heteroscedasticidad, el principal problema es que no conocemos la verdadera autocorrelación de  $\varepsilon_t$ . Por tanto, debemos trabajar con los valores estimados de  $\varepsilon_t$  ( $e_t$ ) y a partir de ellos computar el coeficiente de autocorrelación  $\hat{\rho}_k$  y desarrollar diferentes pruebas estadísticas. No hay un único método para detectar autocorrelación sino una gama de diferentes procedimientos y pruebas estadísticas, en particular ahora estudiaremos tres métodos:

- Método gráfico
- Prueba de Durbin-Watson
- Correlograma y prueba de Ljung y Box

Para el método gráfico podemos colocar los residuos en función del tiempo, tal y como hicimos en la figura 10.1. Este gráfico puede ser construido también con los residuos al cuadrado o los residuos estandarizados.

Otra variante es graficar los residuos contra ellos mismos rezagados un período para así tener una impresión visual sobre la autocorrelación de primer orden. En la figura 10.2 se muestra dicho método. Cuando la autocorrelación es positiva, la mayoría de los residuos se ubicarán en el primer y tercer cuadrante como en la figura 10.2.A. Cuando la autocorrelación es negativa, la mayoría de los residuos se ubicarán en el segundo y cuarto cuadrante como en la figura 10.2.B.

Como siempre, el método gráfico es una primera aproximación que debemos contrastar con pruebas estadísticas formales. Una de ellas es la **prueba Durbin-Watson**.

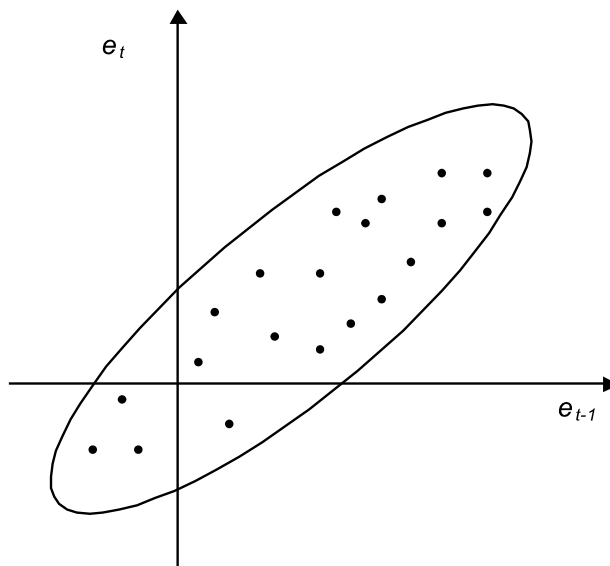
La prueba propone calcular el siguiente estadístico en base a los residuos estimados  $e_t$ :

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

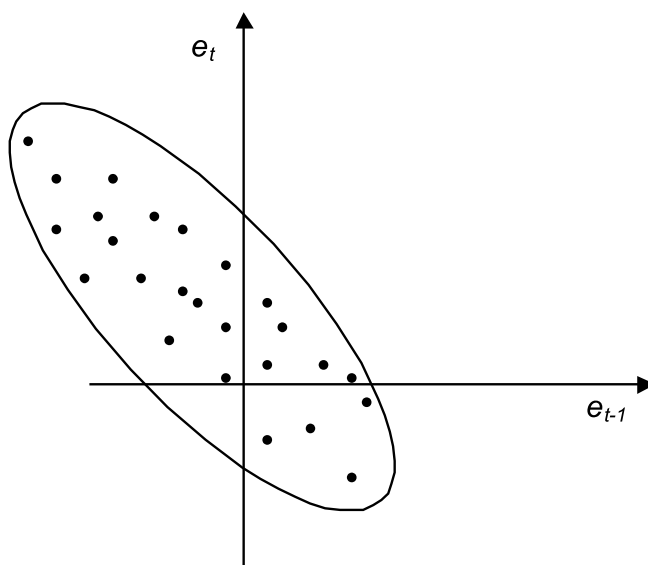
donde  $n$  representa la cantidad de observaciones.

Figura 10.2: Término de error vs. término de error rezagado

## A. Autocorrelación positiva



## B. Autocorrelación negativa



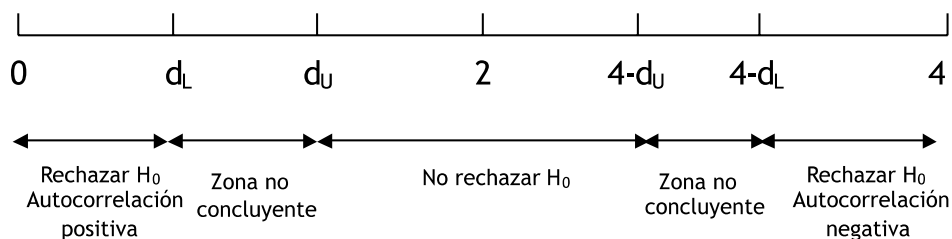
El estadístico  $d$  se apoya en un proceso autorregresivo de orden 1, AR(1):  $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ , donde  $|\rho| \leq 1$  y  $u_t \sim N(0, \sigma^2)$ . En base a este modelo se definen las siguientes hipótesis:

$H_0: \rho = 0$ , no hay autocorrelación de primer orden

$H_1: \rho \neq 0$  sí hay autocorrelación de primer orden

No existe un único valor crítico para comparar el valor calculado de  $d$ . Las tablas con la distribución de probabilidades de este estadístico tienen un límite inferior  $d_L$  y un límite superior  $d_U$ . En el apéndice C de este libro se encuentra la tabla con los valores críticos de Durbin-Watson para un 5%, 2,5% y 1% de significación, los cuales dependen de la cantidad de variables explicativas del modelo ( $k$ ) y la cantidad de observaciones ( $n$ ). Con estos valores críticos se forman las zonas de rechazo, de no rechazo y una zona en la que la prueba no es concluyente sobre la presencia de autocorrelación. En la figura 10.3 se resumen las reglas de decisión del contraste Durbin-Watson.

**Figura 10.3: Prueba de Durbin-Watson**



El estadístico  $d$  de Durbin-Watson se puede aproximar por el coeficiente de autocorrelación usando la expresión  $d \approx 2(1 - \hat{\rho})$ . De ahí se deduce que:

- Cuando  $\hat{\rho} = 0$  entonces  $d \approx 2$  y decimos que no hay autocorrelación.
- Cuando  $\hat{\rho} = 1$  entonces  $d \approx 0$  y decimos que hay perfecta autocorrelación positiva.
- Cuando  $\hat{\rho} = -1$  entonces  $d \approx 4$  y decimos que hay perfecta autocorrelación negativa.

Por tanto, mientras más se acerque a 2 el valor calculado del estadístico Durbin-Watson mayor evidencia tendremos de no autocorrelación. Cuando  $d$  se acerca a cero la evidencia será a favor de la presencia de autocorrelación positiva, mientras que la cercanía de  $d$  a 4 indica existencia de autocorrelación negativa. Los valores calculados de  $d$  que se ubiquen entre  $d_L$  y  $d_U$  o entre  $4-d_U$  y  $4-d_L$  no permiten arribar a ninguna decisión.

La prueba de Durbin-Watson es la más usadas en econometría para contrastar la autocorrelación, no obstante, ella se apoya en un número de supuestos que no deben pasarse por alto:

- El modelo de regresión de donde se toman los  $e_t$  debe incluir un intercepto.

- Las variables explicativas  $X$  son no estocásticas.
- El término de error es generado por un proceso AR(1):  $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ . De hecho, la prueba sólo permite contrastar la autocorrelación de primer orden.

La prueba no es aplicable a modelos de regresión que incluyan rezagos de la variable dependiente, es decir, se quedan fuera de este contraste los modelos autorregresivos del tipo:

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + c_0 x_t + \varepsilon_t$$

El tercer método que vamos a examinar es el correlograma y la prueba de Ljung y Box. El **correlograma** recoge la estructura de dependencia temporal de una serie, en este caso lo vamos a usar para evaluar la correlación serial de los residuos; en el capítulo 15 lo utilizaremos en el proceso de estimación de los modelos ARIMA. El correlograma está compuesto por la función de autocorrelación (FAC) y la función de autocorrelación parcial. La función de autocorrelación ya la vimos en la sección 10.1.

La **función de autocorrelación parcial (FACP)**, al igual que la FAC, mide la correlación entre  $e_t$  y  $e_{t-k}$  pero después de eliminar el efecto de las correlaciones intermedias. Por ejemplo, si existe una correlación alta entre  $e_t$  y  $e_{t-1}$  también la habrá entre  $e_t$  y  $e_{t-2}$  debido a la correlación entre  $e_{t-1}$  y  $e_{t-2}$ . La FACP descuenta  $\hat{\rho}_1$  a la hora de computar  $\hat{\rho}_2$ ; descuenta el efecto de  $\hat{\rho}_1$  y  $\hat{\rho}_2$  para calcular  $\hat{\rho}_3$ ; y así para cada  $\hat{\rho}_k$ . Tanto la FAC y la FACP indicarán una correlación alta positiva (negativa) entre  $e_t$  y  $e_{t-k}$  en la medida que el coeficiente de autocorrelación se acerque a 1 positivo (negativo).

En la figura 10.4 mostramos el correlograma con la FAC y la FACP en la manera más común que lo presentan los programas estadísticos (según las siglas en inglés son las funciones AC y PAC). Se presenta el correlograma hasta el rezago 12. En la figura 10.4.A se ejemplifica un correlograma que indica la presencia de autocorrelación en el proceso. En la figura 10.4.B se muestra un correlograma que nos indica que los residuos no presentan correlación serial, es un correlograma “limpio” de autocorrelación.

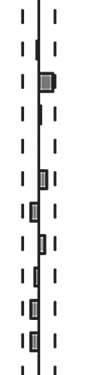
Como hemos reiterado, en la práctica sólo podemos disponer de estimaciones de las FAC y la FACP del modelo verdadero que estamos estudiando. El correlograma que nos presenta el paquete estadístico es una estimación del correlograma verdadero, computado con las observaciones disponibles  $n$ , pudiéndose estimar hasta un máximo  $n/4$  valores de la FAC y la FACP.

Como los valores de  $\rho_k$  son estimados, debemos verificar si son o no significativamente distintos de cero. En la figura 10.4 se marcan las bandas de confianza con líneas discontinuas. Si  $\hat{\rho}_k$  se encuentra dentro de las bandas de confianza entonces no hay autocorrelación de orden  $k$  con un 5% de nivel de significación.

**Figura 10.4: Correlograma**  
**A. Con correlaciones significativas**

| Autocorrelation                                                                   | Partial Correlation                                                               | AC | PAC   | Q-Stat | Prob   |       |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|----|-------|--------|--------|-------|
|  |  | 1  | 0.671 | 0.671  | 488.10 | 0.000 |
|                                                                                   |                                                                                   | 2  | 0.353 | -0.177 | 623.20 | 0.000 |
|                                                                                   |                                                                                   | 3  | 0.088 | -0.135 | 631.62 | 0.000 |
|                                                                                   |                                                                                   | 4  | 0.077 | 0.239  | 638.06 | 0.000 |
|                                                                                   |                                                                                   | 5  | 0.053 | -0.106 | 641.06 | 0.000 |
|                                                                                   |                                                                                   | 6  | 0.039 | -0.024 | 642.75 | 0.000 |
|                                                                                   |                                                                                   | 7  | 0.033 | 0.100  | 643.94 | 0.000 |
|                                                                                   |                                                                                   | 8  | 0.039 | -0.033 | 645.61 | 0.000 |
|                                                                                   |                                                                                   | 9  | 0.034 | -0.014 | 646.86 | 0.000 |
|                                                                                   |                                                                                   | 10 | 0.003 | -0.011 | 646.87 | 0.000 |
|                                                                                   |                                                                                   | 11 | 0.001 | 0.031  | 646.87 | 0.000 |
|                                                                                   |                                                                                   | 12 | 0.015 | 0.014  | 647.13 | 0.000 |

**B. Sin correlaciones significativas**

| Autocorrelation                                                                     | Partial Correlation                                                                 | AC        | PAC    | Q-Stat | Prob  |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------|--------|--------|-------|
|  |  | 1 -0.007  | -0.007 | 0.0103 | 0.919 |
|                                                                                     |                                                                                     | 2 -0.023  | -0.023 | 0.1182 | 0.943 |
|                                                                                     |                                                                                     | 3 0.129   | 0.128  | 3.5294 | 0.317 |
|                                                                                     |                                                                                     | 4 0.022   | 0.023  | 3.6280 | 0.459 |
|                                                                                     |                                                                                     | 5 -0.009  | -0.003 | 3.6442 | 0.602 |
|                                                                                     |                                                                                     | 6 0.095   | 0.081  | 5.5357 | 0.477 |
|                                                                                     |                                                                                     | 7 -0.053  | -0.060 | 6.1357 | 0.524 |
|                                                                                     |                                                                                     | 8 0.037   | 0.042  | 6.4239 | 0.600 |
|                                                                                     |                                                                                     | 9 -0.010  | -0.035 | 6.4472 | 0.694 |
|                                                                                     |                                                                                     | 10 -0.086 | -0.077 | 8.0191 | 0.627 |
|                                                                                     |                                                                                     | 11 -0.052 | -0.060 | 8.5890 | 0.660 |
|                                                                                     |                                                                                     | 12 0.006  | -0.005 | 8.5959 | 0.737 |

Para contrastar autocorrelación Ljung y Box (1978) desarrollaron el **estadístico Q**:

$$Q = n(n+2) \sum_{j=1}^k \rho_j^2 / (n-j)$$

donde  $Q$  se distribuye asintóticamente  $\chi^2$  con  $k$  grados de libertad. La hipótesis nula del contraste plantea que todos los coeficientes de autocorrelación son ceros

hasta el rezago  $k$  ( $\rho_j = 0$ ,  $j = 1, \dots, k$ ). Si se rechaza la hipótesis nula, entonces al menos un coeficiente de autocorrelación es distinto de cero hasta el rezago  $k$  ( $\exists \rho_j \neq 0$ ,  $j = 1, \dots, k$ ). La prueba se repite para cada rezago verificando si hay o no autocorrelación hasta el orden  $k$ . En la figura 10.4 se añaden al correlograma en las dos últimas columnas el valor calculado del estadístico  $Q$  y la probabilidad asociada. Vemos que en la figura 10.4.A se rechaza la hipótesis nula, verificando la presencia de autocorrelación en el proceso; y en la figura 10.4.B no se rechaza la hipótesis nula, indicando que no hay autocorrelación hasta el rezago 12.

En resumen, estos son tres de los métodos más aplicados para contrastar la autocorrelación en los residuos. El análisis gráfico da una primera idea que debe ser verificada por las pruebas estadísticas. De las dos pruebas presentadas, la prueba de Ljung y Box tiene la ventaja sobre la prueba de Durbin-Watson que permiten contrastar autocorrelación de diferentes órdenes.

### 10.3.1 Ejemplo: Detectando autocorrelación en la demanda de importaciones

Continuamos con un ejemplo apoyados nuevamente en el modelo de demanda de importaciones cubanas. Para no complicar la explicación y concentrarnos en el objetivo de este capítulo, obviamos otros problemas detectados en el modelo a lo largo del libro y nos concentramos en el supuesto de no autocorrelación. Los datos se encuentran en la tabla B.1 del CD acompañante. Desde el capítulo 5 habíamos estimado el siguiente modelo:

$$\dot{m}_t = 1,26270 + 0,75064 \dot{y}_t - 0,54252 \dot{p}_t + e_t$$

Ahora, para verificar el supuesto de no autocorrelación empezamos por la inspección gráfica de los residuos  $e_t$ . En la figura 10.5 se muestran los residuos en el tiempo y los residuos contra ellos mismos rezagados un período.

En la figura 10.5.A no se aprecia un patrón sistemático. No hay un comportamiento cíclico, no hay rachas de errores siempre mayores que el anterior y rachas con errores menores que el anterior; tampoco se aprecia una autocorrelación negativa. Por la cantidad de veces que los residuos cruzan el eje cero tampoco se identifica claramente que tengan autocorrelación. Parecen residuos bastante aleatorios si los comparamos con los gráficos que mostramos en las figuras 10.1.B y 10.1.C.

En la figura 10.5.B tenemos a los residuos en función de ellos mismos rezagados un período. Recordemos que cuando la autocorrelación es positiva, la mayoría de los residuos se ubicarán en el primer y tercer cuadrante, y cuando es negativa, la mayoría de los residuos se ubicarán en el segundo y cuarto cuadrante. En la figura 10.5.B no se distinguen ninguno de estos casos, los residuos casi se reparten equitativamente entre los cuatro cuadrantes. En resumen, el método gráfico sugiere que el término de error de la demanda de importaciones cubanas no presenta autocorrelación.

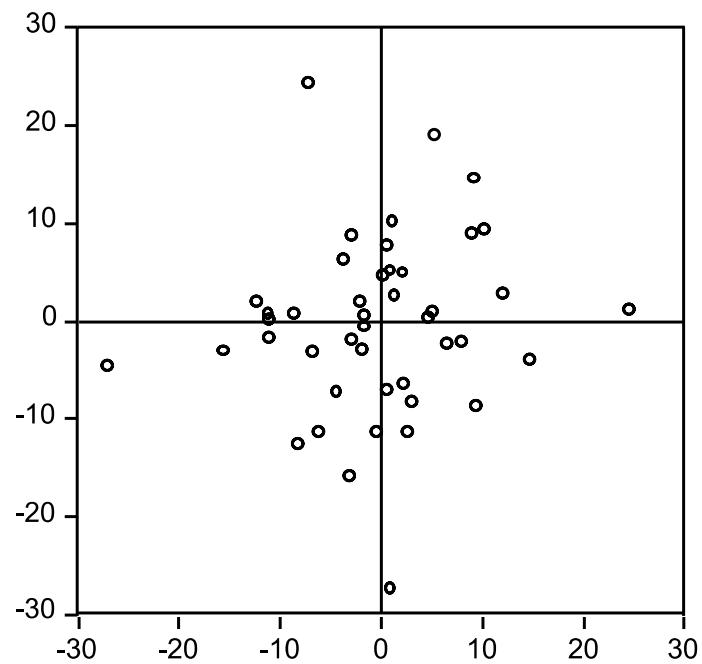


Figura 10.5: Gráfico de los residuos

## A. Residuos en el tiempo



## B. Residuos vs. residuos rezagados



Para corroborar nuestra apreciación gráfica aplicamos a continuación la prueba de Durbin-Watson. Calculamos el estadístico  $d$  en base a los residuos del modelo:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{6098.415}{3764.299} = 1.6200$$

En el apéndice C del libro buscamos los valores críticos del estadístico de Durbin-Watson para  $k=2$  (dos variables explicativas) y  $n=45$  (cantidad de observaciones). Con un 5% de significación tenemos que  $d_L = 1.430$  y  $d_U = 1.615$ . Recordando las zonas de rechazo, no concluyente y no rechazo de la prueba que ilustramos en la figura 10.3, vemos que por muy poco el valor calculado del estadístico  $d$  cae en la zona de no rechazo:  $1,620 > 1,615$ . La prueba indica que no hay autocorrelación de primer orden en el término de error.

Proseguimos con el correlograma y la prueba de Ljung y Box para contrastar autocorrelación de mayores órdenes. En la figura 10.6 se muestra el correlograma con la FAC, la FACP y la prueba  $Q$ .

Figura 10.6: Correlograma de los residuos de la demanda de importaciones

| Autocorrelation                                                                     | Partial Correlation                                                                 | AC | PAC    | Q-Stat | Prob   |       |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----|--------|--------|--------|-------|
|  |  | 1  | 0.122  | 0.122  | 0.7122 | 0.399 |
|  |  | 2  | 0.056  | 0.042  | 0.8662 | 0.648 |
|  |  | 3  | -0.380 | -0.398 | 8.1433 | 0.043 |
|  |  | 4  | -0.196 | -0.123 | 10.128 | 0.038 |
|  |  | 5  | -0.195 | -0.128 | 12.130 | 0.033 |
|  |  | 6  | -0.003 | -0.123 | 12.131 | 0.059 |
|  |  | 7  | 0.076  | -0.021 | 12.454 | 0.087 |
|  |  | 8  | 0.156  | 0.017  | 13.850 | 0.086 |
|  |  | 9  | 0.149  | 0.058  | 15.151 | 0.087 |
|  |  | 10 | 0.058  | 0.024  | 15.356 | 0.120 |
|  |  | 11 | -0.138 | -0.125 | 16.546 | 0.122 |
|  |  | 12 | -0.115 | -0.001 | 17.391 | 0.135 |

La estimación del coeficiente de autocorrelación en el rezago 1,  $\rho_1 = 0.122$ , no se sale de las bandas de confianza con líneas discontinuas. El estadístico  $Q$  no rechaza la hipótesis nula, esto es, el coeficiente de autocorrelación en el rezago 1 no es significativamente distinto de cero. Ello corrobora nuestra inspección gráfica y coincide con el resultado de la prueba de Durbin-Watson.

El coeficiente estimado en el rezago 2 tampoco sobrepasa las bandas de confianza, el estadístico  $Q$  rechaza la hipótesis nula, que en este caso es  $\rho_1 = \rho_2 = 0$ . En el rezago 3 la situación cambia, el coeficiente de autocorrelación

sí sobrepasa las bandas de confianza y la prueba  $Q$  rechaza la hipótesis nula  $\rho_1 = \rho_2 = \rho_3 = 0$  al 5% de significación, indicando que al menos un coeficiente de autocorrelación, hasta el rezago 3, es estadísticamente significativo. En los rezagos 4 y 5 el estadístico  $Q$  repite este resultado.

En balance, el término de error de la demanda de importaciones no presenta autocorrelación de primer orden pero sí de órdenes mayores; todo parece indicar que hay autocorrelación de orden 3.

#### 10.4 Medidas Correctivas para la Autocorrelación

Proseguimos el capítulo evaluando las medidas correctivas que debemos implementar cuando el término de error del modelo que estamos estimando tiene autocorrelación.

Si el término de error de nuestro modelo de regresión no cumple el supuesto de no autocorrelación, lo primero es inspeccionar si se trata de una autocorrelación pura o de una autocorrelación falsa inducida por un sesgo de especificación. Se puede probar añadir otras variables explicativas o modificar la forma funcional del modelo y ver si ello corrige la autocorrelación. En el capítulo 13 se discuten con mayor amplitud diferentes pruebas para detectar variables omitidas y formas funcionales incorrectas. Las mismas conviene aplicarse una vez que encontramos autocorrelación en los residuos, ya que sabemos que el incumplimiento de este supuesto puede estar conectado con sesgos de especificación.

De tratarse de una autocorrelación pura se tienen tres opciones: 1) Estimar un modelo transformado, de manera tal que tome en consideración la autocorrelación. Consistiría en aplicar los estimadores Mínimos Cuadrados Generalizados a los problemas de autocorrelación. 2) Estimar las varianzas de  $\beta$  corregidas por autocorrelación. Tal procedimiento es conocido como errores robustos Newey-West, el cual es una extensión del procedimiento de errores estándar corregidos por heteroscedasticidad de White que estudiamos en el capítulo 9. 3) Modificar la especificación del modelo incluyendo rezagos de la variable dependiente, de las variables independientes y/o términos de medias móviles.

La tercera opción la estudiaremos en detalle en la parte cinco cuando entremos a analizar propiamente las técnicas de series de tiempo. En esa parte del libro aplicaremos los modelos ARIMA, los modelos dinámicos multivariantes, entre otros, para considerar explícitamente los rezagos, inercias y, en general, relaciones dinámicas entre las series de tiempo. Las otras dos opciones las describimos a continuación.

La idea de los MCG es incorporar la información sobre la autocorrelación a los estimadores. En la práctica ello se traduce en inicialmente transformar los datos y posteriormente aplicar los estimadores MCO a los datos transformados.

Por ejemplo, si la autocorrelación que afecta al término de error es de primer orden, entonces procedemos de la siguiente forma.

Tendríamos un modelo de regresión y una expresión para la autocorrelación de primer orden en los residuos:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

donde  $u_t \sim N(0, \sigma^2)$

Sustituimos la expresión para los residuos en el modelo de regresión:

$$Y_t = \beta_0 + \beta_1 X_t + \rho \varepsilon_{t-1} + u_t \quad (10.1)$$

Por otro lado, tenemos la expresión de  $\varepsilon_t$  rezagado un período:

$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$$

$$\varepsilon_{t-1} = Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$$

que sustituimos en (10.1):

$$Y_t = \beta_0 + \beta_1 X_t + \rho(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + u_t$$

$$Y_t = \beta_0 + \beta_1 X_t + \rho Y_{t-1} - \rho \beta_0 - \rho \beta_1 X_{t-1} + u_t$$

$$Y_t - \rho Y_{t-1} = (1 - \rho) \beta_0 + (X_t - \rho X_{t-1}) \beta_1 + u_t$$

De esta forma llegamos al modelo transformado:

$$Y_t^* = \beta_0^* + X_t^* \beta_1 + u_t$$

donde:

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$X_t^* = X_t - \rho X_{t-1}$$

$$\beta_0^* = (1 - \rho) \beta_0$$

En este modelo transformado los residuos no presentarán autocorrelación. Los estimadores MCO aplicados a los datos transformados continúan siendo insesgados y además tienen mínima varianza.

Se observa que la transformación que tenemos que implementar sobre  $Y_t$  y  $X_t$  es restar una proporción  $\rho$  del valor de la variable en el período anterior. A ello se le nombra **cuasi diferenciación**.

Uno de los inconvenientes de esta transformación es que se pierde la primera observación dado que no tiene antecedentes para efectuar la diferenciación. Para evitar tal pérdida se aplica en la primera observación la siguiente transformación:

$$Y_1\sqrt{1-\rho^2} \text{ y } X_1\sqrt{1-\rho^2}$$

En términos matriciales estas transformaciones definirían la matriz  $\mathbf{P}$ :

$$\mathbf{P} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}$$

que debemos multiplicar por el modelo, de manera tal que queda  $\mathbf{PY} = \mathbf{PX}\beta + \mathbf{P}\varepsilon$ . Así, los datos del modelo transformado serán:

$$Y_t^* = \begin{bmatrix} \sqrt{1-\rho^2}Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ \vdots \\ Y_T - \rho Y_{T-1} \end{bmatrix} \quad X_t^* = \begin{bmatrix} \sqrt{1-\rho^2}X_1 \\ X_2 - \rho X_1 \\ X_3 - \rho X_2 \\ \vdots \\ X_T - \rho X_{T-1} \end{bmatrix}$$

Nótese que la aplicación de los MCG requiere conocer  $\rho$ , algo que en la práctica casi nunca sucede. Si  $\rho$  se conoce entonces los MCG se implementan directamente, pero si no, habrá que proceder con diferentes aproximaciones.

Una salida es considerar un punto extremo en las transformaciones asumiendo  $\rho = \pm 1$ , esto es, asumir una autocorrelación perfecta. Así, arribamos a un **modelo en primeras diferencias**:

$$Y_t - Y_{t-1} = \beta_1 (X_t - X_{t-1}) + u_t \\ \Delta Y_t = \beta_1 \Delta X_t + u_t$$

donde  $\Delta$  es el operador en diferencias. En los modelos de series de tiempo veremos que la primera diferencia, además de servir para corregir por autocorrelación se usa para lidiar con la no estacionariedad.

Otro camino es estimar  $\rho$ . Para ello inicialmente utilizar los MCO en el modelo sin transformar. A continuación tenemos dos formas de aproximarnos al coeficiente de autocorrelación  $\rho$ . Una de ellas es mediante el valor calculado del estadístico de Durbin-Watson:

$$\hat{\rho} \approx 1 - \frac{d}{2}$$

Se debe tener cuidado con esta relación en la medida que ella se cumple para muestras grandes.

Otra forma es estimar  $\rho$  a partir de los residuos del modelo MCO:

$$e_t = \hat{\rho}e_{t-1} + u_t$$

Estos procedimientos, en los cuales se aplican los MCG utilizando el valor estimado de  $\rho$  se le conocen en la literatura como **MCG factibles**.

En caso de que usemos el modelo en primeras diferencias, la estimación de  $\rho$  nos sirve para verificar su validez. La transformación de primeras diferencias es más adecuada en la medida que estimamos que efectivamente  $|\rho|$  se acerca a 1.

La segunda opción que habíamos mencionado para corregir la autocorrelación es la estimación de los **errores robustos Newey-West**, lo cual es una extensión del procedimiento de errores estándar corregidos por heteroscedasticidad de White. No se desarrolla aquí el procedimiento por la complejidad del mismo. Los paquetes estadísticos usualmente traen esta opción incorporada.

El procedimiento de Newey-West corrige en los errores estándar tanto los problemas de autocorrelación como los de heteroscedasticidad, a diferencia del procedimiento de White que solamente corrige la heteroscedasticidad. Un punto a destacar es que la estimación de errores robustos Newey-West también requiere de muestras grandes.

Ejemplificamos ahora el procedimiento de corrección de Newey-West con la demanda de importaciones cubanas. Ya habíamos identificado en la sección 10.3.1 que el término de error no cumplía el supuesto de no autocorrelación. En la tabla 10.1 ilustramos las transformaciones que sufre el modelo una vez que aplicamos la corrección de Newey-West.

**Tabla 10.1: Ejemplo del procedimiento de errores robustos Newey-West**

**Variable dependiente:  $\dot{m}_t$**

| Variables explicativas                            | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---------------------------------------------------|-------------|----------------|-----------------|--------------|
| <i>Estimación sin corrección</i>                  |             |                |                 |              |
| Intercepto                                        | 1.262701    | 1.645596       | 0.767321        | 0.4472       |
| $\dot{y}_t$                                       | 0.750641    | 0.223379       | 3.360391        | 0.0017       |
| $\dot{p}_t$                                       | -0.542519   | 0.090403       | -6.001137       | 0.0000       |
| <i>Estimación con errores estándar corregidos</i> |             |                |                 |              |
| Intercepto                                        | 1.262701    | 1.511760       | 0.835252        | 0.4083       |
| $\dot{y}_t$                                       | 0.750641    | 0.267254       | 2.808717        | 0.0075       |
| $\dot{p}_t$                                       | -0.542519   | 0.081621       | -6.646829       | 0.0000       |

Como vemos en la tabla 10.1, la corrección sólo es para los errores estándar, el valor estimado de los coeficientes no se modifica. Ello es coherente con que ante la presencia de autocorrelación y heteroscedasticidad en el término de error los estimadores MCO de  $\beta$  van a seguir siendo insesgados, sin embargo, los estimadores de la varianza de  $\beta$  sí serán sesgados. En este sentido, la estimación

de errores robustos Newey-West corrige el sesgo sobre la varianza de  $\beta$  con el objetivo de eliminar las afectaciones que ello produce en la inferencia estadística.

Por supuesto, junto con los errores estándar cambian también el estadístico  $t$  y la probabilidad asociada. Es decir, la corrección de Newey-West podría cambiar las conclusiones a las que habíamos arribado sobre la significación de los parámetros. En el ejemplo de la demanda de importaciones este no es el caso. Las modificaciones que sufren los errores estándar después de aplicar la corrección del procedimiento Newey-West no son tan drásticas como para cambiar la significación de los coeficientes. Tanto el ingreso como los precios continúan siendo significativos.





# *preguntas*

Intente dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo

1. Explique qué es autocorrelación.
2. ¿Cuáles son las causas del incumplimiento de este supuesto?
3. ¿Cómo se formula la función de autocorrelación?
4. ¿Cuáles son las consecuencias del incumplimiento del supuesto sobre los estimadores MCO?
5. Explique la prueba Durbin-Watson
6. ¿Qué son los errores robustos Newey-West?

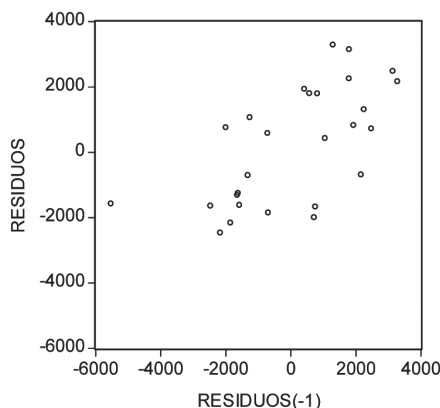


# ejercicios

10.1 Se estimó un modelo con los datos del PIB y exportaciones de la tabla B.4 del CD acompañante. El modelo de regresión estimado por MCO es el siguiente:

$$PIB_t = 19524.6 + 2.414 * X_t$$

- a. Haga una valoración de la autocorrelación a partir del siguiente gráfico con los residuos y los residuos rezagados un período.



- b. Compruebe su apreciación gráfica con la prueba de Durbin-Watson. El valor calculado del estadístico para el modelo estimado es de 0.6269. En el apéndice C de este libro se encuentra la tabla con los valores críticos de Durbin-Watson.
- c. Si encuentra autocorrelación en los residuos indique el orden de la misma.
- d. Analice ahora la autocorrelación a partir del correlograma y la prueba Q siguientes:

| Autocorrelation |                                                                                     | Partial Correlation |                                                                                     | AC | PAC    | Q-Stat | Prob   |       |
|-----------------|-------------------------------------------------------------------------------------|---------------------|-------------------------------------------------------------------------------------|----|--------|--------|--------|-------|
|                 |  |                     |  | 1  | 0.553  | 0.553  | 9.5173 | 0.002 |
|                 |  |                     |  | 2  | 0.331  | 0.037  | 13.067 | 0.001 |
|                 |  |                     |  | 3  | 0.128  | -0.099 | 13.622 | 0.003 |
|                 |  |                     |  | 4  | -0.176 | -0.320 | 14.706 | 0.005 |
|                 |  |                     |  | 5  | -0.229 | -0.004 | 16.621 | 0.005 |
|                 |  |                     |  | 6  | -0.307 | -0.118 | 20.222 | 0.003 |
|                 |  |                     |  | 7  | -0.251 | 0.037  | 22.740 | 0.002 |
|                 |  |                     |  | 8  | -0.305 | -0.278 | 26.651 | 0.001 |
|                 |  |                     |  | 9  | -0.313 | -0.141 | 30.992 | 0.000 |
|                 |  |                     |  | 10 | -0.208 | -0.041 | 33.013 | 0.000 |
|                 |  |                     |  | 11 | -0.186 | -0.060 | 34.718 | 0.000 |
|                 |  |                     |  | 12 | -0.044 | -0.016 | 34.822 | 0.000 |

10.2 Estime el modelo del ejercicio anterior pero añadiendo como variable explicativa la Formación Bruta de Capital Fijo que también aparece en la tabla B.4 del CD acompañante.

- a. Haga una valoración de la autocorrelación a partir de gráfico con los residuos y los residuos rezagados un período.

- b. Compruebe su apreciación gráfica con la prueba de Durbin-Watson. En el apéndice C de este libro se encuentra la tabla con los valores críticos de Durbin-Watson.
  - c. Analice la autocorrelación a partir del correlograma y la prueba Q.
  - d. Aplique alguna de las medidas correctivas para autocorrelación estudiadas en el capítulo.
- 10.3 Estime el modelo del ejercicio anterior pero con las variables en primeras diferencias.
- a. Analice la autocorrelación a partir del correlograma y la prueba Q.
  - b. Aplique el procedimiento conocido como errores robustos Newey-West y compare los resultados

CAPÍTULO I I



Este capítulo lo dedicaremos a examinar qué consecuencias tiene el que los errores aleatorios de una regresión sigan una distribución diferente de la normal, a detectar posibles desvíos de este supuesto y a discutir medidas correctivas ante divergencias al respecto. Ya en el Capítulo 6 habíamos comentado que el supuesto de normalidad no era absolutamente necesario para la estimación por MCO. Este método de estimación no se apoya en hipótesis alguna acerca de la distribución del término de error. Para lo que sí resulta relevante este supuesto es para el trabajo de inferencia estadística, que consiste en la realización de tests de hipótesis sobre los parámetros de interés. Este supuesto se asienta en el hecho de que el término aleatorio del MRLN resume el efecto de un sinnúmero de variables independientes que no hemos incorporado explícitamente al modelo de regresión y en que, según el teorema central del límite (TCL), la distribución de la suma de un gran número de variables aleatorias independientes tiende justamente a ser normal a medida que se incrementa el número de variables en la sumatoria. Aun en los casos de que el número de variables que se suma no sea muy grande, o si éstas no fueran estrictamente independientes, la suma de esas variables pudiera tener una distribución normal. Así, el TCL y sus múltiples formas nos brindan la justificación teórica para asumir la normalidad de  $\varepsilon_i$ .

### 11.1 Consecuencias de la no-Normalidad

Dado un modelo como el presentado en (5.1), si asumimos que  $\varepsilon_i \sim N(0, \sigma^2)$ , entonces  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$ . Ya examinamos en el Capítulo 7 de este texto que, bajo este supuesto podemos derivar las distribuciones de probabilidad de los

estimadores MCO de los parámetros, que también serán normales, así como la del estimador de la varianza común del modelo, que responderá a una Chi-cuadrado, emparentada también con la distribución normal.

La normal es una distribución relativamente sencilla: se expresa totalmente a través de sus dos primeros momentos (media y varianza), sus propiedades teóricas han sido ampliamente estudiadas y su función de distribución está perfectamente tabulada. Además, en la realidad son muchos los fenómenos que parecen regirse por tal distribución. En el caso de muestras grandes, no es absolutamente imprescindible el supuesto de normalidad. Siempre podremos recurrir a la validez asintótica de una distribución normal para  $\hat{\beta}$ . Así, aun sin conocer la distribución específica del término de error, podremos hacer las pruebas estadísticas necesarias, que tendrán validez asintótica.

Sin embargo, en el caso de que dispongamos de muestras pequeñas, el supuesto de normalidad es especialmente crítico. Para poder hacer tests de hipótesis bajo esta circunstancia y para una distribución de  $\varepsilon_i$  diferente de la normal, necesitaremos conocer la forma exacta de la distribución de ese término para poder derivar los estimadores máximo verosímiles de los coeficientes de regresión y sus correspondientes intervalos de confianza.

Así, para poder hacer inferencia estadística a partir de los resultados estándar de la EMCO y en caso de disponer de una muestra pequeña (como suele ocurrir generalmente en el caso de las series cubanas) precisamos comprobar que la evidencia empírica no sea contraria al supuesto de una distribución normal para los errores del modelo. En el próximo epígrafe les proporcionamos herramientas para esta comprobación.

## 11.2 Detección de la no-Normalidad

Para examinar la normalidad de la distribución de una variable aleatoria tenemos varios tests disponibles. La hipótesis nula de estas pruebas es que la distribución a que obedece la variable objeto de estudio es normal y la alternativa, una distribución diferente de la normal. Dependiendo de si la hipótesis alternativa está expresada en forma paramétrica o no, estas pruebas clasificarán en paramétricas y no paramétricas.

Entre los tests no paramétricos están el de Kolmogorov-Smirnov y el de Shapiro-Wilk.<sup>1</sup> Sin embargo, en este texto vamos a concentrarnos particularmente en los tests paramétricos, que son los que podrán encontrar incorporados a los paquetes computacionales econométricos de mayor circulación.

<sup>1</sup> Para mayor detalle sobre estas pruebas no paramétricas, pueden consultar a Spanos (1986, p. 452), quien a su vez refiere el artículo de Durbin (1973).

Uno de ellos es la prueba de Jarque y Bera, incluida en numerosos paquetes econométricos. Otro test de este tipo es el diseñado por Doornik y Hansen (1994). Ambas pruebas son tests de simetría-curtosis, pues se apoyan en los momentos centrales tercero (o coeficiente de simetría,  $S$ ) y cuarto (coeficiente de curtosis o apuntamiento,  $K$ ) de la distribución de variable en cuestión, digamos  $Y_i$ . Estos coeficientes se obtienen como:

$$S = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_i - \bar{Y}}{\hat{\sigma}} \right)^3$$

$$K = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_i - \bar{Y}}{\hat{\sigma}} \right)^4$$

La hipótesis alternativa en estas pruebas es que la variable obedece a una distribución de la llamada familia Pearson de distribuciones diferente de la normal. La familia Pearson incluye un gran número de distribuciones interesantes como la propia normal, la gamma (de la cual la Chi-cuadrado es un caso particular), la beta y la t-Student.

La hipótesis nula del test Jarque-Bera propone que los momentos tercero y cuarto posean las características propias de una distribución normal: a saber, que el coeficiente de simetría sea igual a cero y que el coeficiente de curtosis sea igual a tres. En otras palabras la hipótesis nula sostiene que la serie examinada se distribuye normalmente.

$$H_0 : S = 0 \text{ y } K = 3$$

$$H_1 : S \neq 0 \text{ ó } K \neq 3$$

Estos autores construyeron un estadístico a partir de los momentos estimados para probar la conjunción de ambas condiciones. El indicador de Jarque-Bera ( $JB$ ) responde a la siguiente fórmula:

$$JB = \frac{n-k}{6} \left( S^2 + \frac{(K-3)^2}{4} \right) \quad (11.1)$$

donde  $n$  es el tamaño muestral (o cantidad de observaciones); y  $k$ , el número de coeficientes estimados para generar la serie (se aplica al término de error estimado,  $e_i$ ).

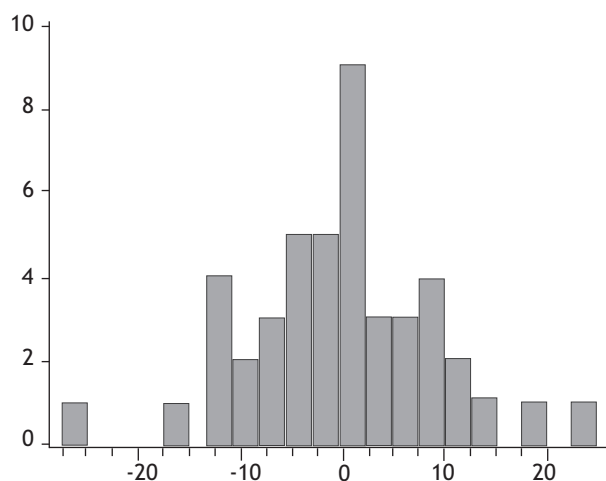
Sin embargo, debemos enfatizar que este test tiene validez asintótica, o sea, que requerimos grandes muestras, y que es sensible a observaciones atípicas. Por lo tanto, ante la detección de no normalidad debemos verificar si tenemos valores atípicos (*outliers*) en nuestra muestra. Si efectivamente hay *outliers* y podemos justificar su presencia, los modelamos y con ello resolvemos el problema (ver

Capítulo 16). En caso contrario, debemos implementar medidas para corregir este desvío de los supuestos clásicos.

Obviamente, como no conocemos los verdaderos valores del término de error en la población, vamos a realizar nuestras investigaciones sobre normalidad a sus mejores representantes: los residuos de la regresión estimada. Se trata de aplicar los tests antes aludidos a estos residuos, para detectar si no aportan evidencia contraria de que provengan de una variable con distribución normal.

Vamos a aplicar el test de Jarque-Bera a los residuos obtenidos al estimar nuestra función de demanda de importaciones para la economía cubana (ver el Capítulo 5). Seguidamente mostramos los resultados que reporta el test en la figura 11.1.

**Figura 11.1: Resultados del test de normalidad a los residuos de la función de demanda de importaciones cubanas, 1961-2005**



*Fuente: Elaboración propia*

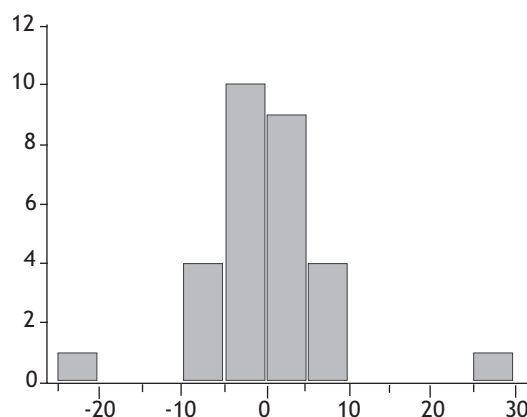
Tal cual pueden observar, la probabilidad asociada al estadístico Jarque-Bera es superior a un 5%, por lo que no rechazamos la hipótesis de normalidad.

Sin embargo, si para este mismo ejemplo modificamos el período muestral y estimamos para el período 1960-1989, los resultados del test de normalidad son dramáticamente distintos (ver la figura 11.2).

Noten que la probabilidad asociada al estadístico de prueba es prácticamente cero, por lo que rechazamos la hipótesis de normalidad. Como ya hemos advertido antes, recuerden que al realizar cada test estamos asumiendo que el modelo no tiene ningún otro error de especificación, por lo que no debemos precipitarnos en aplicar medidas para remediar el presunto problema de especificación asociado a cada prueba que realicemos. Además, no debemos olvidar que este es un resultado asintótico aplicado a una pequeña muestra.



**Figura 11.2: Resultados del test de normalidad a los residuos de la función de demanda de importaciones cubanas, 1961-1989.**



*Fuente: Elaboración propia*

### 11.3 Medidas Correctivas para la no-Normalidad

Si al aplicar cualquiera de los tests disponibles a los residuos estimados obtenemos un rechazo de la hipótesis de normalidad, no podremos hacer la necesaria inferencia estadística. En este caso tenemos varias líneas de acción, en dependencia de la información disponible:

- Si sabemos cuál es la distribución de probabilidades de los residuos, podremos estimar  $\hat{\beta}$  a través del método de máxima verosimilitud y tratar de derivar su distribución en pequeña muestra.
- Si no tenemos idea de cuál puede ser la distribución de los errores y tenemos una muestra grande, podemos estimar por MCO y valernos de las propiedades asintóticas.
- Finalmente queda una última salida, que consiste en encontrar una transformación de las variables (lo mismo  $Y$  que  $X$ , o ambas) que induzca normalidad. Por lo general este tipo de transformaciones también inducen linealidad y homoscedasticidad.

Una transformación muy utilizada que tiene esos efectos es la denominada Box-Cox. Ella convierte las variables según la siguiente expresión general:<sup>2</sup>

<sup>2</sup> Ver Cassoni (1991, p. 32). Vamos a utilizar el subíndice  $t$  en la fórmula para transformar las variables, pues en nuestro ejemplo estamos trabajando con una serie de tiempo. Si la muestra fuese una sección cruzada, el subíndice sería  $i$ .

$$Z_t^* = \frac{Z_t^\lambda - 1}{\lambda}$$

donde  $Z_t$  es la variable original;  $Z_t^*$ , la variable transformada y  $\lambda$ , el parámetro de la transformación que toma valores en el intervalo  $[-1, 1]$ . Según el valor que tome este parámetro, tendremos diferentes transformaciones posibles. Por ejemplo:

- ✓ Si  $\lambda = 1$ , entonces  $Z_t^* = Z_t - 1$ , lo que equivale prácticamente a no transformar las variables.
- ✓ Si  $\lambda = -1$ ,  $Z_t^* = 1 - \frac{1}{Z_t}$ , que equivale a trabajar con los recíprocos de las variables.
- ✓ Si  $\lambda = 0,5$ ,  $Z_t^* = 2(\sqrt{Z_t} - 1)$ , se trata de la transformación raíz cuadrada.
- ✓ Se puede demostrar que si  $\lambda \rightarrow 0$  entonces  $Z_t^* \rightarrow \ln(Z_t)$ ; se trata de la transformación logarítmica.

Entre las transformaciones abordadas, la más interesante para la modelación econométrica es la logarítmica por variadas razones. En primer lugar, su interpretación económica es directa: por ejemplo, ya sabemos que cuando trabajamos todas las variables del modelo en términos de logaritmos, los coeficientes de regresión pueden ser interpretados como elasticidades; o cuando estimamos una función de tendencia para el logaritmo de una determinada magnitud económica, el coeficiente de regresión de la tendencia temporal se interpreta como una tasa de crecimiento.

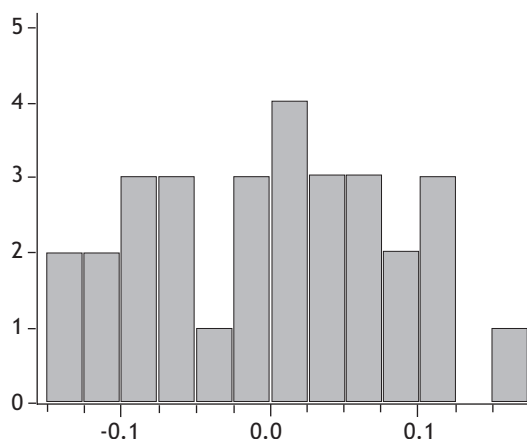
Otra razón a favor de esta transformación reside en el hecho de que algunas variables económicas ostentan distribuciones con asimetría positiva (como la log-normal, la Chi-cuadrado, la gamma). En estos casos las variables sólo toman valores positivos. Si aplicamos logaritmos, esto tendrá el efecto de hacer su distribución más simétrica, pues los logaritmos de números positivos pueden ser negativos.

Vamos a reespecificar nuestra función de demanda de importaciones y a trabajar con los logaritmos del PIB, de la variable *proxy* para los precios de las importaciones y de las propias importaciones, en lugar de operar con las primeras diferencias de estas variables, como habíamos venido haciendo en los capítulos anteriores. El modelo quedaría especificado como sigue:

$$LM_t = \hat{\alpha}_0 + \hat{\alpha}_1 LY_t + \hat{\alpha}_2 LP_t + v_t$$

donde la “L” que antecede a las variables significa que están en logaritmos. Los resultados del test de Jarque-Bera para los residuos de la regresión estimada para el período 1960-1989, que antes resultaron problemáticos, se muestran en la figura 11.3.

**Figura 11.3: Resultados del test de normalidad a los residuos de la función de demanda de importaciones cubanas, especificada en logaritmos, 1960-1989.**



*Fuente: Elaboración propia*

Noten como obtenemos una probabilidad para el estadístico que no permite rechazar el supuesto de normalidad. La transformación logarítmica tuvo el efecto deseado de inducir la normalidad.



*Fuente: Elaboración propia*



# *preguntas*

Trate de responder a las siguientes preguntas con ayuda de este capítulo.

1. ¿Es imprescindible el supuesto de normalidad de los residuos de un modelo de regresión para su estimación por el método de los MCO? Fundamente su respuesta.
2. ¿Por qué precisamos de algún supuesto acerca de la distribución de los términos de error en una regresión?
3. ¿En qué situación no precisaríamos de tal supuesto?
4. ¿Cuáles tests tenemos disponibles para el examen de la normalidad del término de error de una regresión?
5. En caso de detectar no-normalidad de los residuos de una regresión, ¿cuáles opciones tenemos para corregir esta desviación de los supuestos del MRLN?
6. ¿Qué efectos suele tener la aplicación de la transformación logarítmica a las variables de una regresión sobre sus residuos?



# ejercicios

- 11.1 En la tabla B.9 del CD acompañante se incluyen las importaciones cubanas de arroz consumo ( $MQ$ , medidas en miles de toneladas) y algunos de los factores que, según la teoría económica convencional, influyen en el comportamiento de las mismas para los años entre 1980 y 2006: el precio de las importaciones, aproximado por el valor unitario de las importaciones mundiales según datos de FAOSTAT ( $VUM\_W$ , en dólares por tonelada); la producción nacional de arroz ( $PNQ$ , en miles de toneladas) y el ingreso nacional, aproximado por el producto interno bruto ( $PIB$ , en millones de pesos de 1997). No se ha incorporado el precio interno de la producción nacional por haberse mantenido fijo para el acopio en la mayor parte de este período.
- Estime la función de demanda de importaciones a partir de los factores anteriores.
  - Evalúe la normalidad de los residuos de la regresión anterior.
  - En caso de que los residuos no sean normales, realice una transformación inductora de la normalidad a las variables. Reestime el modelo y compruebe si esta transformación surtió efecto.
  - Entre los factores que, según la teoría económica, influyen en las cantidades importadas está el precio doméstico del arroz. En este ejercicio no lo hemos incluido en el modelo. ¿Cree Ud. que esta decisión afectará los resultados de la estimación y en particular la condición de normalidad de sus residuos?
  - En la tabla B.9 hemos incluido también la población media para cada año del período ( $POB$ ). ¿Opina Ud. que esta variable podría ser importante para la explicación de la importación de arroz en Cuba?
  - Incluya la variable referida a la población previamente logaritmizada en la función de demanda de importaciones. Investigue sobre la normalidad de los residuos. Comente sobre los resultados.
- 11.2 Se le solicita estimar con las variables expresadas como tasas de variación, la función de demanda de importaciones cubanas que presentamos en el Capítulo 5 para el período completo, 1960-2005, y para dos subperíodos, de 1960 a 1989 y de 1990 a 2005.
- Compare los estimados obtenidos para los coeficientes de los modelos de regresión en cada uno de los períodos.
  - Investigue la normalidad de los residuos para cada uno de los modelos estimados.
  - En caso de presentarse problemas con la normalidad en alguno de los períodos analizados, comente cuáles podrían ser las causas y muestre alguna evidencia.

- d. Trabaje con otra especificación para este modelo que evite los problemas de no normalidad de los residuos que se presentan.
- 11.3 En el ejercicio 7.6 le indicamos estimar diferentes especificaciones del modelo de Mincer para la formación de salarios.
- a. Investigue para cada una de estas especificaciones acerca de la normalidad de los residuos de cada modelo.
  - b. En caso de presentarse residuos que no sigan la distribución normal, ¿qué consecuencias tendría esto para los estimados obtenidos y su evaluación? ¿qué haría Ud. para mitigar este problema?

CAPÍTULO 12





## REGRESORES ESTOCÁSTICOS

Uno de los supuestos básicos en que se funda el método de estimación por mínimos cuadrados, es el que sostiene que los valores observados para las variables explicativas son números fijos. Es decir, suponemos que las variables explicativas, son variables no aleatorias. Este supuesto es perfectamente razonable y aceptable en el contexto de experimentos controlados, la principal herramienta de investigación empírica en las ciencias naturales. Sin embargo, los experimentos controlados son raramente posibles en la investigación económica. En realidad debemos conformarnos con la información que existe o los datos que nos proporcionan procedimientos muestrales. De esta manera, con muchísima frecuencia debemos manejarnos con variables explicativas que son aleatorias. En este capítulo discutiremos las condiciones en que el carácter aleatorio de las variables explicativas afecta el resultado de nuestras estimaciones, cómo detectar eventuales problemas y las medidas para solucionar los mismos que son posibles. También haremos énfasis en un caso que ha tenido especial relevancia en la literatura: el caso de los modelos de ecuaciones simultáneas.

### 12.1 Reformulando el Modelo

Debido a que la mayoría de las variables económicas tienen un carácter aleatorio, debemos reformular nuestro sistema de supuestos, para acomodar estas nuevas variables explicativas aleatorias. Nuestros nuevos supuestos son los siguientes:

1. El modelo lineal  $y_i = \beta_1 + \beta_2 x_i + e_i$  está correctamente especificado.
2. Nuestros datos han sido obtenidos mediante procedimientos de muestreo aleatorio. Esto es lo opuesto a los experimentos controlados, donde los

valores de las variables explicativas pueden ser elegidos por el investigador. En nuestro caso no tenemos control sobre los valores que asumirán las variables explicativas; las que son, entonces, variables aleatorias.

3. Supondremos que la esperanza matemática de los errores estocásticos es igual a cero, condicionada a los valores asumidos por las variables explicativas. Es decir, suponemos que  $E(e_i | x_i) = 0$  en lugar de simplemente  $E(e_i) = 0$ .
4. Hay al menos dos valores diferentes para  $x_i$ .
5. También supondremos que la varianza de los errores estocásticos, condicionada a los valores asumidos por las variables explicativas, es constante. Es decir, suponemos que  $\text{var}(e_i | x_i) = \sigma^2$  en lugar de simplemente  $\text{var}(e_i) = \sigma^2$ .
6. Finalmente supondremos que la distribución condicional de los errores estocásticos es normal con media igual a cero y varianza constante entre las observaciones. Es decir, suponemos que  $e_i | x_i \sim N(0, \sigma^2)$ .

Nuestro tercer supuesto, una media condicional igual a cero para los errores estocásticos, refleja varios factores. Por ejemplo, que no se han omitido variables importantes y que se ha usado una forma funcional adecuada. Más importante aún, este supuesto refleja el hecho de que no hay razones para pensar que los errores estocásticos están relacionados con las variables explicativas. En realidad podríamos demostrar que si tenemos que  $E(e_i | x_i) = 0$ , entonces  $x_i$  y  $e_i$  no están correlacionados y, en consecuencia, tenemos que  $\text{cov}(e_i, x_i) = 0$ . Inversamente, es fácil de ver que cuando  $x_i$  y  $e_i$  están correlacionados, tenemos que  $\text{cov}(e_i, x_i) \neq 0$ .

Bajo estos supuestos los estimadores de un modelo con variables explicativas aleatorias tienen las siguientes propiedades de pequeñas muestras o de muestras finitas:

1. Si se cumplen los cuatro primeros supuestos, los estimadores mínimo-cuadráticos son insesgados.
2. Si se cumplen los primeros cinco supuestos, los estimadores mínimo-cuadráticos son los mejores estimadores lineales insesgados.
3. Si se cumplen los seis supuestos, la distribución de los estimadores mínimo-cuadráticos es normal condicionada a los valores observados para  $x_i$ . Los errores estándar de estos estimadores pueden ser calculados con las fórmulas habituales. En consecuencia, los procedimientos habituales de prueba de hipótesis continúan siendo válidos.

A juzgar por estos resultados, pareciera que las variables explicativas estocásticas no son causa de problema alguno, pues los estimadores mínimo-cuadráticos tienen las propiedades esperadas. Sin embargo, estos resultados dependen críticamente de que las variables explicativas no estén correlacionadas con los errores estocásticos. Es decir, del supuesto que implica que  $\text{cov}(e_i, x_i) = 0$ . Sin embargo, cuando esto no es cierto y tenemos, por el contrario, que  $\text{cov}(e_i, x_i) \neq 0$  veremos que se presentan serios problemas. En este caso diremos que la variable  $x_i$  es **endógena**. Más adelante veremos cómo podemos tratar este problema de **endogeneidad**.

También es importante considerar las propiedades asintóticas o de gran muestra de los estimadores mínimo-cuadráticos cuando los regresores son aleatorios. Para comenzar esta discusión, recordemos las propiedades de los estimadores mínimo-cuadráticos cuando la muestra es muy grande, o cuando el tamaño muestral tiende a infinito. En primer lugar, los estimadores mínimo-cuadráticos son insesgados. En segundo lugar, la varianza de estos estimadores converge a cero cuando el tamaño de la muestra crece indefinidamente. Esto significa que, a medida que el tamaño de la muestra crece, la distribución de los estimadores colapsa alrededor del valor poblacional o verdadero de los parámetros. En este caso decimos que los estimadores mínimo-cuadráticos son **consistentes**, siendo esta una propiedad deseable.

Para discutir el caso de regresores aleatorios, bajo condiciones asintóticas o de gran muestra, substituiremos nuestro anterior tercer supuesto,  $E(e_i | x_i) = 0$ , por un nuevo supuesto que sostiene que  $\text{cov}(x_i, e_i) = 0$ . Nótese que este es un supuesto más débil que el anterior, ya que:

$$E(e_i | x_i) = 0 \Rightarrow \text{cov}(x_i, e_i) = 0$$

pero no necesariamente al revés. Es decir, las propiedades de pequeña muestra implican las propiedades de gran muestra pero no inversamente.

Bajo estas nuevas condiciones, lo que podemos afirmar es lo siguiente:

1. Para modelos con regresores aleatorios que se ajustan a los cuatro primeros supuestos (incluyendo la nueva versión del tercero), los estimadores mínimo-cuadráticos son consistentes.
2. Para modelos que se ajustan a los cinco supuestos anteriores, los estimadores mínimo-cuadráticos tienen una distribución que es aproximadamente normal. Esta aproximación depende de si los errores estocásticos del modelo están distribuidos normalmente o no. En consecuencia, si la muestra es suficientemente grande, los procedimientos usuales de inferencia, los estadígrafos de prueba y los intervalos de confianza son válidos.
3. Si nuestro nuevo tercer supuesto no se cumple, entonces los estimadores mínimo-cuadráticos son **inconsistentes**. Es decir, no convergen a los valores verdaderos de los parámetros, aún para muestras muy grandes. Más

aún, ninguno de los procedimientos habituales de prueba de hipótesis, ni los intervalos de confianza, son válidos bajo estas circunstancias.

Una vez más el punto crítico es la existencia o no de correlación entre los regresores aleatorios y los errores estocásticos del modelo. Si esta correlación no existe y la covarianza entre los regresores aleatorios y los errores estocásticos es igual a cero, los estimadores mínimo-cuadráticos conservan todas sus buenas propiedades. Sin embargo, si los regresores aleatorios y los errores estocásticos están correlacionados, con una covarianza distinta de cero, los estimadores mínimo-cuadráticos resultan sesgados e inconsistentes. También los procedimientos de prueba de hipótesis y los intervalos de confianza se hacen inválidos. A continuación presentaremos una posible salida a este problema: la estimación con **variables instrumentales**.

## 12.2 Variables Instrumentales

Nuestro enfoque para resolver el problema ocasionado por regresores aleatorios correlacionados con los errores estocásticos del modelo, se basa en una aplicación del método de estimación de los momentos. La idea básica de dicho método es utilizar un momento muestral para estimar un momento poblacional. Hemos discutido esta idea en capítulos anteriores.

Consideremos el modelo lineal

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

Si los regresores,  $x_i$ , son números fijos, no-aleatorios, o aleatorios pero no correlacionados con los errores  $e_i$ , podemos escribir que los siguientes momentos son iguales a cero:

$$\begin{aligned} E(e_i) &= 0 \Rightarrow E(y_i - \beta_0 - \beta_1 x_i) = 0 \\ E(x_i e_i) &= 0 \Rightarrow E[x_i (y_i - \beta_0 - \beta_1 x_i)] = 0 \end{aligned}$$

Estos dos momentos poblacionales pueden ser estimados por los correspondientes momentos muestrales. De esta manera podemos obtener las siguientes dos ecuaciones con dos incógnitas:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i) &= 0 \\ \frac{1}{n} \sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) &= 0 \end{aligned}$$

Pero estas dos ecuaciones son idénticas a las llamadas ecuaciones normales que encontramos en la deducción de los estimadores mínimo-cuadráticos. En consecuencia, los valores de  $b_0$  y  $b_1$  que satisfacen estas ecuaciones, son los estimadores mínimo-cuadráticos. En otras palabras tenemos que los estimadores mínimo-

cuadráticos y los estimadores de momentos son idénticos para el modelo lineal estándar.

Por otra parte, cuando tenemos regresores aleatorios podemos enfrentar una dificultad. Si es el caso que  $E(x_i e_i) \neq 0$ , el método de los momentos no nos permite escribir la segunda ecuación. Sin embargo, supongamos que tenemos otra variable,  $z_i$ , que llamaremos una **variable instrumental**, que satisface la siguiente condición:

$$E(z_i e_i) = 0 \Rightarrow E[z_i (y_i - \beta_0 - \beta_1 x_i)]$$

Entonces, siguiendo con el método de los momentos podemos escribir las siguientes ecuaciones,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i) &= 0 \\ \frac{1}{n} \sum_{i=1}^n z_i (y_i - b_0 - b_1 x_i) &= 0 \end{aligned}$$

Los valores de  $b_0$  y  $b_1$  que satisfacen a estas ecuaciones son nuestros **estimadores de variables instrumentales**.

Naturalmente, la variable instrumental,  $z$ , debe cumplir con ciertas condiciones. En primer lugar,  $z$  no debe tener influencia sobre  $y$ . En segundo lugar, la variable instrumental y los errores estocásticos deben tener covarianza cero. Esto es, debemos tener  $\text{cov}(z, e) = 0$ . Finalmente, la variable instrumental debe estar fuertemente correlacionada con el regresor aleatorio. Es decir, debiera ser  $\text{cov}(z, x) \neq 0$ . Sin embargo, muchas veces debemos contentarnos con variables instrumentales que, al menos, no están débilmente correlacionadas con el regresor.

Para grandes muestras, el estimador de variables instrumentales tiene una distribución que es aproximadamente normal,

$$\hat{b}_1 \sim N\left(\beta_1, \frac{\sigma_{IV}^2}{\sum (x_i - \bar{x})^2 r_{zx}^2}\right)$$

Donde el subíndice *IV* indica “variable instrumental”. Por su parte, la varianza del modelo puede ser calculada de acuerdo a la siguiente fórmula

$$\hat{\sigma}_{IV} = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}{n - 2}$$

Usualmente tenemos a nuestra disposición más variables instrumentales o **instrumentos** que los que necesitamos. En este caso podemos extender nuestro enfoque para incorporar estos instrumentos adicionales y escribir, por ejemplo, cuando tenemos dos instrumentos  $z$  y  $w$ :

$$E(e_i) = E(y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = m_1 = 0$$

$$E(z_i e_i) = E z_i (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum z_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = m_2 = 0$$

$$E(w_i e_i) = E w_i (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum w_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = m_3 = 0$$

Los estimadores se construyen buscando valores que minimicen la siguiente suma de cuadrados:

$$S(\hat{\beta}_0, \hat{\beta}_1) = m_1^2 + m_2^2 + m_3^2$$

Usualmente necesitamos algoritmos de métodos numéricos para esta optimización. Sin embargo, es posible diseñar un simple procedimiento de estimación por **mínimos cuadrados en dos etapas**:

- **Primera Etapa.** Consideramos un modelo en que explicamos el regresor aleatorio,  $x$ , usando las variables instrumentales que hemos elegido. Si tenemos más de un regresor aleatorio con problemas (con covarianza distinta de cero con los errores del modelo), consideramos un modelo para cada una de ellas. Estimamos los parámetros de este modelo usando mínimos cuadrados. Usamos los estimadores mínimo-cuadráticos de este modelo auxiliar para obtener una estimación nuestro regresor aleatorio, es decir, creamos una nueva variable,  $\hat{x}_i$ .
- **Segunda etapa.** Ahora estimamos, usando mínimos cuadrados, el siguiente modelo:

$$y_i = \beta_0 + \beta_1 \hat{x}_i + v_i$$

La fórmula apropiada para la varianza con grandes muestras es,

$$\text{var}(\hat{\beta}_1) = \frac{\sigma_{IV}^2}{\sum (\hat{x}_i - \bar{x})^2}$$

Por su parte, la varianza del modelo puede ser estimada a partir de los residuales del modelo anterior; es decir, podemos usar la fórmula anterior. De esta manera, podemos estimar la varianza del estimador como:

$$\widehat{\text{var}}(\hat{\beta}_1) = \frac{\hat{\sigma}_{IV}^2}{\sum (\hat{x}_i - \bar{x})^2}$$

Nuevamente, los estimadores que obtenemos con ayuda de variables instrumentales tienen una distribución aproximadamente normal para grandes muestras y, en consecuencia, podemos usar este estimador de la varianza del estimador de los parámetros para construir pruebas de hipótesis e intervalos de confianza basados en la distribución de Student. Es decir, nuestros procedimientos habituales de inferencia son válidos.

### 12.3 Pruebas de Endogeneidad

Naturalmente, antes de comenzar la búsqueda de variables instrumentales y poner en marcha estos procedimientos, es necesario diagnosticar si efectivamente tenemos un problema de endogeneidad. Recordemos que la presencia de regresores aleatorios no necesariamente implica que tenemos problemas con los estimadores mínimo cuadráticos. El punto crítico es determinar si podemos o no suponer que la covarianza entre el regresor aleatorio y los errores estocásticos del modelo es igual a cero ( $\text{cov}(x_i, e_i) = 0$ ). Si este supuesto no es sostenible, entonces tenemos un problema de sesgo que es necesario corregir. Más aún, si la correlación entre el regresor aleatorio y los errores del modelo es muy fuerte, el problema puede llegar a ser muy grave.

Lo que necesitamos, entonces, es una prueba donde la hipótesis nula sostenga la ausencia de endogeneidad. Es decir, una prueba cuya hipótesis nula sea  $H_0 : \text{cov}(x, e) = 0$ , con  $H_1 : \text{cov}(x, e) \neq 0$  como hipótesis alternativa. Un enfoque posible para esta prueba de hipótesis es comparar el comportamiento de los estimadores mínimo-cuadráticos y los estimadores de variables instrumentales. Entonces, si la hipótesis nula es cierta, tanto los estimadores mínimo-cuadráticos como los estimadores de variables instrumentales son consistentes. En consecuencia, podemos afirmar que la expresión  $h = (\hat{\beta}_{MCO} - \hat{\beta}_{IV})$  tiende a cero cuando el tamaño de la muestra crece. Por otra parte, si la hipótesis nula no es cierta esta expresión tiende a una constante distinta de cero. Es decir, tenemos que  $h = (\hat{\beta}_{MCO} - \hat{\beta}_{IV}) \rightarrow c \neq 0$ . Una implementación de este enfoque, la que ha adquirido una gran influencia, es la prueba de endogeneidad de Hausman.

#### Prueba de Endogeneidad de Hausman

Esta prueba apunta a saber si, en la regresión

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

la variable  $x$  está o no correlacionada con los errores estocásticos del modelo,  $e$ . Supongamos, que tenemos dos variables instrumentales,  $z_1$  y  $z_2$ , para la variable  $x$ . En general, necesitaremos al menos una variable instrumental por cada regresor aleatorio cuya endogeneidad queremos probar. La prueba se realiza en las siguientes etapas:

1. Estimar, usando mínimos cuadrados, la regresión  $x_i = \alpha_0 + \alpha_1 z_{1i} + \alpha_2 z_{2i} + v_i$
2. Usar estas estimaciones para obtener los residuales  $\hat{v}_i = x_i - \hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i}$
3. Si existe más de un regresor aleatorio que sea sospechoso de endogeneidad, repetir este procedimiento para cada una de estas variables, incluyendo

todas las variables instrumentales y todas las variables explicativas que no son sospechosas de endogeneidad.

4. Incluir los residuales estimados en la etapa 2 en la siguiente regresión artificial

$$y_i = \beta_0 + \beta_1 x_i + \delta \hat{v}_i + e_i$$

5. Estimar esta regresión artificial usando mínimos cuadrados. La prueba “ $t$ ” habitual para el parámetro  $\delta$  se interpreta en el sentido que la hipótesis nula,  $H_0 : \delta = 0$ , implica no correlación entre  $x$  y  $e$ . En consecuencia, la evidencia empírica sugiere la ausencia de endogeneidad. La hipótesis alternativa,  $H_1 : \delta \neq 0$ , implica correlación entre  $x$  y  $e$ . En consecuencia, la evidencia empírica sugiere que hay endogeneidad.
6. Cuando estamos probando más de un regresor aleatorio, la prueba conveniente es una prueba “ $F$ ” de significación conjunta de los coeficientes de los residuales incluidos.

## 12.4 Validez de las Variables Instrumentales

Otro problema asociado al anterior, es la validez de las variables instrumentales escogidas. Hemos dicho anteriormente que estas variables deben estar fuertemente correlacionadas con el regresor aleatorio que instrumentan o, al menos, no deben estar débilmente correlacionadas con éste. Es posible construir una prueba para esto.

Supongamos que tenemos el siguiente modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + e$$

donde las  $k$  primeras variables son exógenas, no correlacionadas con el término de error  $e$ . Supongamos que la variable  $x_{k+1}$  es endógena y que  $z$  es un posible instrumento. Para construir nuestra prueba estimamos, usando mínimos cuadrados, la siguiente regresión:

$$x_{k+1} = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_k x_k + \theta z + v$$

Si el indicador  $t$  correspondiente a  $\theta$  es menor que 3,3, diremos que  $z$  es un instrumento débil. Este procedimiento puede ser generalizado a más variables instrumentales. En este caso necesitamos una prueba  $F$  de la significación conjunta de todos los coeficientes correspondientes a los instrumentos. Si el valor del indicador  $F$  es menor que 10, diremos que los instrumentos son débiles.

También podemos generalizar el procedimiento para más variables endógenas, repitiéndolo para cada una de ellas.



## 12.5 Modelos de Ecuaciones Simultáneas

Los modelos de ecuaciones simultáneas son un área especialmente interesante para la aplicación de estas ideas sobre endogeneidad de los regresores aleatorios. Los modelos de ecuaciones simultáneas son especialmente importantes por su uso en macroeconomía. Para estos modelos, la cuestión de la endogeneidad de los regresores aleatorios endógenos no necesita ser probada, sino que está implícita en la construcción de los mismos.

Para discutir estas ideas usaremos, como ejemplo, lo que podemos llamar el modelo keynesiano mínimo, compuesto de dos ecuaciones:

$$\begin{aligned}C_t &= \beta_1 + \beta_2 Y_t + e_t \\ Y_t &= C_t + I_t\end{aligned}$$

Donde  $C$  representa el consumo,  $Y$  el ingreso nacional, e  $I$  las inversiones. Nótese que este modelo no tiene sector público ni sector externo. Para comenzar podemos hacer varias observaciones sobre estas ecuaciones. En primer lugar, el carácter de la primera ecuación es distinto del carácter de la segunda ecuación. Diremos que la primera ecuación es **conductual** pues refleja la conducta de agentes económicos e individuos. Por este motivo marcamos su carácter estocástico agregando un término aleatorio  $e$ . La segunda ecuación es una **identidad**, es decir es siempre cierta. En realidad es una definición del ingreso nacional, y como tal no admite un término aleatorio. Los modelos de ecuaciones simultáneas deben tener al menos una ecuación conductual y varias o ninguna identidad.

En segundo lugar, este modelo tiene dos clases de variables. Por una parte variables endógenas<sup>1</sup>, que se determinan dentro del modelo ( $Y$  y  $C$ , en este caso), y variables exógenas que toman sus valores fuera del modelo ( $I$ , en este caso). En general, hay una variable endógena por cada ecuación y, usualmente, podemos encontrarlas en el lado izquierdo de las ecuaciones. El resto de las variables las declaramos exógenas. En este caso tenemos dos variables endógenas,  $Y$  y  $C$ , y una variable exógena,  $I$ . Sin embargo, algunas veces nuestra definición, de qué es endógeno y qué es exógeno es, en este contexto, más arbitraria.

Finalmente, el problema con estos modelos surge del hecho de que, muchas veces, las variables endógenas aparecen tanto en el lado izquierdo como el lado derecho de las ecuaciones. La consecuencia de este hecho es que no podemos estimar las ecuaciones de este modelo con mínimos cuadrados. Para entender este problema consideremos la primera ecuación. Claramente, la variable  $C$  es aleatoria e incorpora el elemento estocástico  $e$ . Pero esta variable está también en el lado derecho de la segunda ecuación, con lo que la variable  $Y$  es ahora claramente aleatoria incluyendo el elemento estocástico  $e$ . Pero, entonces, la primera ecuación tiene a  $Y$  como un regresor aleatorio, claramente correlacionado con el com-

<sup>1</sup> Note que ahora estamos usando la palabra endógena en un contexto ligeramente diferente.

ponente estocástico,  $e$ . En consecuencia, los estimadores mínimo-cuadráticos serán sesgados. Este sesgo se llama **sesgo de simultaneidad**.

La forma en que presentamos anteriormente las ecuaciones del modelo keynesiano mínimo se llama una **forma estructural**. Esta forma expresa nuestras hipótesis sobre el funcionamiento de la economía. Por ejemplo, la primera ecuación señala claramente que destinamos una proporción  $\beta_2$  del ingreso al consumo, y que debemos considerar un consumo básico igual a  $\beta_1$ . Por esto tenemos interés en estimar estos parámetros, pues reflejan nuestras hipótesis básicas sobre el funcionamiento de la economía.

Sin embargo, es posible reescribir estas ecuaciones de tal manera que tengamos sólo una variable endógena en el lado izquierdo de cada una de ellas, y sólo variables exógenas en el lado derecho. Por ejemplo, podemos escribir, después de algunas manipulaciones algebraicas, que:

$$C_t = \frac{\beta_1}{(1-\beta_2)} + \frac{\beta_2}{(1-\beta_2)} I_t + \frac{1}{(1-\beta_2)} e_t$$

$$C_t = \pi_{11} + \pi_{21} I_t + v_t$$

Llamaremos a esta ecuación la forma reducida de la ecuación para el consumo,  $C$ . En el caso del ingreso,  $Y$ , podemos escribir:

$$Y_t = \frac{\beta_1}{(1-\beta_2)} + \frac{1}{(1-\beta_2)} I_t + \frac{1}{(1-\beta_2)} e_t$$

$$Y_t = \pi_{12} + \pi_{22} I_t + v_t$$

Esta es la forma reducida de la ecuación del ingreso,  $Y$ .

Aparentemente, el problema de simultaneidad desaparece con la forma reducida y podemos estimar esta forma usando mínimos cuadrados. Una vez que tenemos los estimados  $\hat{\pi}$  podemos usar las fórmulas originales para encontrar estimadores de los parámetros originales  $\hat{\beta}$ . Sin embargo, esta solución no es tan simple, pues si bien siempre podemos encontrar la forma reducida a partir de la forma estructural, no siempre es posible volver a la forma estructural a partir de la forma reducida. Por otra parte, aún si existen estas fórmulas que permiten recuperar los parámetros estructurales, en general, no son lineales, dificultando los procedimientos de inferencia. Incluso podemos encontrar situaciones en que las fórmulas que explican los coeficientes estructurales en términos de los coeficientes reducidos no tienen soluciones únicas.

La posibilidad de recuperar los coeficientes o parámetros estructurales, a partir de los coeficientes o parámetros de la forma reducida, es conocida como el problema de **identificación**. Para poder estimar los parámetros de una ecuación estructural, en un modelo de ecuaciones simultáneas, esta ecuación debe estar identificada. En realidad, dado un modelo con varias ecuaciones simultáneas, cuando consideramos el problema de identificación de alguna de sus ecuaciones, podemos encontrar tres diferentes situaciones.

1. Diremos que una ecuación está **sub-identificada** o simplemente **no identificada** si sus parámetros estructurales no pueden ser expresados en términos de los parámetros de la forma reducida del modelo. En otras palabras, cuando no es posible encontrar fórmulas que permitan recuperar los parámetros estructurales a partir de los parámetros de la forma reducida.
2. Diremos que una ecuación está **exactamente identificada** o simplemente **identificada**, si los parámetros estructurales pueden ser unívocamente expresados en términos de los parámetros de la forma reducida. Es decir, es este caso las fórmulas que expresan los parámetros estructurales en términos de los parámetros de la forma reducida son únicas y, en consecuencia, el problema de recuperar los parámetros estructurales tiene una única solución.
3. Diremos que una ecuación está **sobre-identificada** si las fórmulas que expresan los parámetros estructurales en términos de los parámetros de la forma reducida existen pero no son únicas. En otras palabras, esto sucede cuando el problema de recuperar los parámetros estructurales a partir de los parámetros de la fórmula reducida no tiene una solución única.

Afortunadamente, podemos utilizar una simple regla para resolver el problema de identificación de una ecuación. Antes que nada, observemos que el problema de identificación se refiere a cada ecuación y no al modelo en general. Identifiquemos (o fallamos de identificar) cada una de las ecuaciones, con ayuda de la regla que discutiremos en seguida. Naturalmente, si una de las ecuaciones es una identidad, como la segunda ecuación de nuestro modelo keynesiano en el ejemplo que encabeza este capítulo, no tenemos nada que identificar. Simplemente porque no hay parámetros estructurales que recuperar a partir de los parámetros de la forma reducida. Consideremos ahora una ecuación que no incluye ninguna variable endógena en el lado derecho. Entonces, esta ecuación explica una variable endógena usando solamente variables exógenas. Pero, entonces, la forma estructural es idéntica con la forma reducida de esta misma ecuación. Dado que los parámetros estructurales son idénticos a los parámetros de la forma reducida, esta ecuación está, por definición, exactamente identificada. De hecho esta ecuación puede ser estimada directamente, en forma independiente del resto del modelo pues no tiene problemas de simultaneidad. Pero estos son casos más bien excepcionales. Para las demás ecuaciones utilizaremos la regla que describimos a continuación.

Consideremos un modelo de ecuaciones simultáneas con  $M$  ecuaciones y  $M$  variables endógenas. Ahora consideremos cada una de las ecuaciones del modelo y su identificación. Diremos que:

- Una ecuación está exactamente identificada si  $M-1$  de las variables del modelo han sido excluidas de esta ecuación.
- Una ecuación está sub-identificada si menos de  $M-1$  de las variables del modelo han sido excluidas de esta ecuación.

- Una ecuación está sobre-identificada si más de  $M-1$  de las variables del modelo han sido excluidas de esta ecuación.

Es importante notar que podemos decidir la identificación de una ecuación simplemente contando las variables del modelo y las variables excluidas (o incluidas) en cada ecuación. En primer lugar, hacemos un cuidadoso censo de las variables incluidas en nuestro modelo. En nuestro ejemplo del modelo keynesiano tenemos 2 variables endógenas,  $Y$  e  $C$ , y una variable exógena,  $I$ . Es decir, para nuestro ejemplo,  $M=2$  y en total 3 variables. Consideremos ahora la primera ecuación del modelo keynesiano; es decir, la ecuación de consumo. Es fácil ver que en esta ecuación utilizamos solamente 2 variables,  $C$  y  $Y$ . Entonces, de un total de 3 variables hemos utilizado 2. En consecuencia, en esta ecuación hay  $3-2=1$  variable excluida. Por otra parte  $M-1=2-1=1$ . Es decir, esta ecuación está exactamente identificada. La segunda ecuación es una identidad y no hay nada que identificar.

Este problema de identificación es importante pues para ecuaciones simultáneas exactamente identificadas existe un mecanismo simple y elegante de corrección del sesgo de simultaneidad usando estimaciones de variables instrumentales.

## 12.6 Estimaciones Multi-etápicas para Ecuaciones Simultáneas

El problema de los modelos de ecuaciones simultáneas es el sesgo de simultaneidad. Este problema es básicamente un problema de regresores estocásticos correlacionados con los errores estocásticos del modelo. En consecuencia los estimadores mínimo-cuadráticos resultan sesgados. Pero, en estas condiciones, el enfoque de estimación adecuado es uno de variables instrumentales.

Continuando con esta línea de pensamiento podemos diseñar un procedimiento de estimación de dos etapas que simplemente reproduce nuestra discusión anterior cuando tenemos varias variables instrumentales. Este procedimiento se llama **Mínimos Cuadrados de Dos Etapas**. El aspecto más interesante de este procedimiento es que utiliza como instrumentos, todas las variables exógenas del modelo. Naturalmente, es posible aumentar el número de variables instrumentales con otras ajenas al modelo. Sin embargo nos limitaremos en este momento sólo al uso de las variables exógenas del modelo como instrumentos, para explicar como funciona este método de estimación.

Para explicar el método de Mínimos Cuadrados de Dos Etapas utilizaremos como ejemplo un modelo de dos ecuaciones simultáneas, con la siguiente forma estructural:

$$y_{1t} = \beta_1 + \beta_2 y_{2t} + \beta_3 x_{1t} + \varepsilon_{1t}$$

$$y_{2t} = \alpha_1 + \alpha_2 y_{1t} + \alpha_3 x_{2t} + \varepsilon_{2t}$$

Sugerimos que es un buen ejercicio para el lector el investigar si estas dos ecuaciones pueden ser identificadas o no. La correspondiente forma reducida de este modelo es la siguiente:

$$y_{1t} = \pi_{11} + \pi_{12}x_{1t} + \pi_{13}x_{2t} + v_{1t}$$

$$y_{2t} = \pi_{21} + \pi_{22}x_{1t} + \pi_{23}x_{2t} + v_{2t}$$

Omitimos las fórmulas que relacionan los parámetros de la forma reducida con los de la forma estructural, pues no resultan relevantes para esta discusión. Utilizaremos  $x_1$  y  $x_2$  como variables instrumentales para  $y_1$  y  $y_2$  que aparecen como regresores estocásticos en las ecuaciones estructurales.

En una *Primera Etapa*, y siguiendo el procedimiento que esbozamos anteriormente cuando tenemos más de un instrumento para cada regresor estocástico, estimamos las ecuaciones de la forma reducida, utilizando Mínimos Cuadrados. Podemos usar estas estimaciones para conseguir valores estimados de las variables endógenas,  $\hat{y}$ , y los correspondientes residuales,  $\hat{v}$ . Es decir,

$$\hat{y}_{1t} = \hat{\pi}_{11} + \hat{\pi}_{12}x_{1t} + \hat{\pi}_{13}x_{2t}$$

$$\hat{y}_{2t} = \hat{\pi}_{21} + \hat{\pi}_{22}x_{1t} + \hat{\pi}_{23}x_{2t}$$

También podemos escribir:

$$y_{1t} = \hat{y}_{1t} + \hat{v}_{1t}$$

$$y_{2t} = \hat{y}_{2t} + \hat{v}_{2t}$$

En la *Segunda Etapa* utilizamos las relaciones anteriores para sustituir los regresores estocásticos en la forma estructural y obtenemos un nuevo modelo:

$$y_{1t} = \beta_1 + \beta_2\hat{y}_{2t} + \beta_3x_{1t} + u_{1t}$$

$$y_{2t} = \alpha_1 + \alpha_2\hat{y}_{1t} + \alpha_3x_{2t} + u_{2t}$$

$$u_{1t} = \beta_2\hat{v}_{2t} + \varepsilon_{1t} \quad \text{y} \quad u_{2t} = \alpha_2\hat{v}_{1t} + \varepsilon_{2t}$$

Ahora podemos estimar este nuevo modelo usando Mínimos Cuadrados. Los estimadores obtenidos de esta manera son consistentes, aunque no los más eficientes. En general, estos estimadores resultan suficientemente eficientes y podemos tener confianza en nuestra inferencia. Si estos estimadores no resultaran suficientemente eficientes, podemos pensar en un procedimiento de estimación con **Mínimos Cuadrados de Tres Etapas**.

Para estos Mínimos Cuadrados de Tres Etapas, utilizamos los estimadores obtenidos con los Mínimos Cuadrados de Dos Etapas para obtener nuevos estimados de las variables endógenas. Estos estimados reemplazan, nuevamente, a los regresores que son variables endógenas en la forma estructural, y se repita la segunda etapa. Los nuevos estimadores de tres etapas son consistentes y más eficientes que los de dos etapas.

Naturalmente podemos continuar este proceso obteniendo estimadores más y más eficientes. De hecho muchos paquetes de programas ofrecen **Mínimos Cuadrados Iterativos**, en que nuevas etapas van siendo agregadas hasta que se cumple un cierto criterio de convergencia. Sin embargo, en la práctica, rara vez son necesarias más de dos o tres etapas.



# *preguntas*

Pruebe dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo.

1. ¿Cuándo la presencia de regresores estocásticos no afecta las estimaciones mínimo-cuadráticas?
2. Si tenemos que la covarianza de un determinado regresor aleatorio y los errores estocásticos del modelo es diferente de cero, ¿qué consecuencias podemos esperar para los estimadores?
3. ¿Qué es una variable instrumental?
4. ¿Cuáles son las características deseables para una variable instrumental?
5. ¿Qué es la prueba de Hausman?
6. Si tenemos un modelo de ecuaciones simultáneas, ¿por qué no podemos estimar sus ecuaciones usando Mínimos Cuadrados Ordinarios?
7. ¿Qué es una forma reducida?
8. ¿Qué es el problema de identificación?
9. ¿Cómo podemos estimar modelos de ecuaciones simultáneas?



# *ejercicios*



12.1 Considere el siguiente modelo de ecuaciones simultáneas,

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 z + \varepsilon_1$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 y + \varepsilon_2$$

- Indique, para cada una de estas ecuaciones si está identificada, sobre identificada o sub-identificada.
- Encuentre la forma reducida de estas ecuaciones.
- ¿Cómo estimaría este modelo?

12.2 Considere el siguiente modelo macroeconómico,

$$C_t = \alpha_1 + \alpha_2 Y_t + e_{t1}$$

$$I_t = \beta_1 + \beta_2 r_t + e_{t2}$$

$$Y_t = C_t + I_t + G_t$$

Donde  $Y_t$  es el PIB,  $C_t$  es el consumo,  $I_t$  es la inversión y, finalmente,  $r_t$  es la tasa de interés.

- ¿Cuáles son las variables endógenas y cuales las exógenas?
- ¿Qué ecuaciones son conductuales y cuáles son identidades?
- ¿Qué ecuaciones están identificadas y cuáles no?
- Determine la forma reducida de este modelo.





CAPÍTULO 13



## ERRORES DE ESPECIFICACIÓN Y SELECCIÓN DEL MODELO

Hasta aquí hemos examinado variados problemas que puede presentar un modelo de regresión, tales como: heteroscedasticidad, autocorrelación y no-normalidad de los residuos, regresores no estocásticos y multicolinealidad entre las variables explicativas. Sin embargo, las dificultades no se reducen sólo a las previamente mencionadas. Esta lista es mucho mayor y precisamente a complementarla dedicaremos este capítulo. En particular vamos a abordar los denominados errores de especificación, entre los que incluimos: la omisión de variables relevantes, la inclusión de variables superfluas, la utilización de una forma funcional inadecuada en los modelos y profundizaremos aún más en la posible inestabilidad en los regresores. Igual que hemos procedido en los capítulos precedentes, indicamos cuáles son las consecuencias de estos errores de especificación, cómo detectarlos y qué hacer para mitigarlos. También trataremos los criterios para la selección de modelos y, finalmente, daremos algunas guías útiles para la conducción de estudios empíricos y la redacción de los correspondientes informes.

### 13.1 Características de un Buen Modelo

¿Cuándo cometemos un error en la especificación de un modelo? Cuando en lugar de estimar el modelo correcto, el que verdaderamente genera los datos observados en la práctica, estimamos un modelo diferente. La teoría económica nos debe proporcionar argumentos sobre el modelo correcto, pero sabemos que en muchos casos apenas nos dice cuáles son las variables que deben estar relacionadas y cuáles los signos (el sentido) de esa relación. Así, como la teoría nos deja de informar sobre aspectos que son relevantes cuando de estimar un modelo

econométrico se trata —por ejemplo, la forma funcional correcta—, debemos buscar estas respuestas en la otra parte del problema: los datos.

Debemos enfatizar que difícilmente sabremos cuál es el modelo correcto o verdadero. Nuestra pretensión es mucho más modesta: encontrar un modelo “razonablemente” correcto, que represente relativamente bien la realidad. Esta es una cuestión fundamentalmente práctica o empírica: debemos preguntarle a los datos, a la muestra disponible, si efectivamente ellos pudieron ser generados a partir de un modelo con los rasgos del que postulamos.

Decimos que un determinado modelo es bueno, cuando posee las siguientes características:

- **Parquedad:** Reconocemos que un modelo nunca puede capturar completamente la realidad; de hecho, por definición, sólo representa sus aspectos más esenciales. En el trabajo de modelación debemos atender al principio de “la navaja de Occam”<sup>1</sup> o de parsimonia, que aconseja mantener las cosas tan simplemente como sea posible (naturalmente, siempre que la representación capte lo esencial del fenómeno que estudiamos). Así, lo correcto es partir de un modelo relativamente sencillo e ir complicándolo en la medida que los datos lo exijan. Mientras menos variables y no linealidades, ¡mejor! Por lo general, cuando conocemos poco sobre un proceso, la tendencia que debemos resistir es introducir muchas complicaciones matemáticas para su explicación.
- **Identificabilidad:** Dada una muestra, sólo debe existir un estimado posible por parámetro. Es decir, nuestro procedimiento o algoritmo de estimación debe ser tal que nos permita obtener los estimados de los parámetros de interés y que estos sean únicos.
- **Bondad de ajuste:** Con un buen modelo debemos obtener una buena bondad de ajuste. Ya sabemos que la bondad de ajuste expresa cuánto de la variabilidad de la variable dependiente es explicada por las independientes. Esta característica se explora a partir del  $R^2$  y del  $\bar{R}^2$ , y mejor aún, a partir del test  $F$  de significación global de la regresión. También tenemos otros criterios disponibles para esta evaluación, como el de Akaike y el de Schwartz.
- **Consistencia teórica:** Debemos evitar medida sin teoría. Aun cuando reconocemos que a veces la teoría no nos aporta todos los elementos necesarios para la modelación, construir modelos sin una teoría económica de soporte no es una buena práctica. Algunos investigadores aplican el proceder de estimar todos los modelos posibles, a partir de las variables disponibles, y compararlos entre sí a través del  $R^2$  para seleccionar el mejor según

<sup>1</sup> Occam’s razor, en inglés.

bondad de ajuste. Esto puede ser altamente peligroso, pues nos puede conducir directamente a estimar regresiones espurias.

- **Poder predictivo:** Un buen modelo debe poder predecir su propia muestra y, muy especialmente, fuera de la muestra. Uno de los objetivos de la modelación econométrica es justamente poder hacer pronósticos.

Un buen argumento teórico abre muchas discusiones e induce estudios para verificar si la realidad sostiene la teoría en cuestión. Un buen argumento empírico cierra la discusión: si no logramos ratificar los postulados teóricos, la teoría es la que se equivoca, la realidad no puede estar equivocada. Sin embargo, debemos cerciorarnos de que efectivamente estamos lidiando con una falla de la teoría y no con una de especificación del modelo, que nos conduciría a estimadores sesgados, distorsionados con relación a los postulados teóricos. Seguidamente abordaremos errores de especificación que con frecuencia solemos cometer en los estudios empíricos.

## 13.2 Variables Omitidas

Esta situación se presenta cuando dejamos de incluir una variable importante en el modelo de regresión. En este caso decimos que hemos “subespecificado” el modelo. Por lo general, incurrimos en este lapsus por no haber sido suficientemente exhaustivos en nuestro análisis teórico o, lo que es más común, por no disponer de datos para todas las variables que, de acuerdo a la teoría económica, son relevantes. Una variable importante o relevante tendrá una alta posibilidad de ejercer una influencia estadística significativa en el contexto que se analiza.

Por ejemplo, en la especificación de una función de demanda consideramos la cantidad demandada del producto en cuestión, como variable dependiente, y entre las variables independientes tenemos el precio de ese producto, pero también los precios de los productos complementarios y sustitutos. Supongamos que queremos estimar la función de demanda de café. Así, tendremos entre las variables explicativas el precio del propio café y también los precios de:

- el azúcar (¡al menos en Cuba queremos café bien dulce...!);
- la energía, necesaria para poderlo preparar;
- el té, producto sustituto del café.

Sin embargo, es posible que no tengamos datos para algunas de esas variables. Bajo esas circunstancias ¿valdría la pena estimar el modelo?, ¿cuán grave puede ser estimarlo con esa falta de información?

Ilustraremos las posibles fallas a partir de un ejemplo hipotético. Supongamos que el modelo de regresión real (verdadero) es  $Y_t = B_1 + B_2X_{2t} + B_3X_{3t} + u_t$ , y en su lugar estimamos  $Y_t = a_1 + a_2X_{2t} + v_t$ . Estamos omitiendo una variable relevante y esa omisión provoca los siguientes problemas:

- Si la variable omitida  $X_3$  está relacionada con  $X_2$ , entonces  $\hat{a}_1$  y  $\hat{a}_2$  son sesgados. O sea  $E(\hat{a}_1) \neq B_1$  y  $E(\hat{a}_2) \neq B_2$ .

Podemos demostrar que, en realidad, los valores esperados del intercepto y la pendiente de la regresión estimada al omitir  $X_3$  son:

$$E(\hat{a}_1) = B_1 + B_3(\bar{X}_3 - b_{32}\bar{X}_2)$$

$$E(\hat{a}_2) = B_2 + B_3b_{32}$$

Donde  $b_{32}$  es el coeficiente de la regresión de  $X_3$  en  $X_2$ ,  $X_{3i} = b_{31} + b_{32}X_{2i} + w_i$ . Noten que, dependiendo de los signos de los coeficientes de regresión  $B_3$  y  $b_{32}$ , el valor esperado de  $\hat{a}_2$  será mayor o menor que la verdadera influencia de  $X_2$  sobre  $Y$ .

- Los coeficientes  $\hat{a}_1$  y  $\hat{a}_2$  también serán inconsistentes en grandes muestras.
- Si  $X_3$  no estuviera correlacionada con  $X_2$ ,  $b_{32}$  sería cero y  $\hat{a}_2$  sería insesgado, pero aun así  $\hat{a}_1$  sigue siendo sesgado. La varianza del error estimada  $(\hat{\sigma}_v^2)$  es un estimador sesgado de la verdadera varianza  $\sigma_u^2$ .
- La varianza estimada de  $\hat{a}_2$  es un estimador sesgado de la varianza de  $B_2$ , aun si  $b_{32}$  fuera cero. Tendrá un sesgo positivo, tal y como se deduce de la fórmula correspondiente:

$$E[\text{var}(a_2)] = \text{var}(B_2) + \frac{B_3^2 \sum x_{3i}^2}{(n-2) \sum x_{2i}^2}$$

- El intervalo de confianza usual y los procedimientos para tests de hipótesis no serán confiables. En este caso el intervalo de confianza será más amplio y existirá la tendencia a no poder rechazar la hipótesis nula con más frecuencia que la debida.

Así, la omisión de una variable relevante constituye un error que puede ser grave cuando la variable omitida está correlacionada con las variables incluidas. Vamos a mostrar los efectos perniciosos que puede tener este tipo de error en la regresión a través de la estimación de los modelos del ejemplo anterior. En la tabla 13.1 mostramos la información correspondiente a la estimación de: el modelo correctamente especificado, el modelo con una variable omitida y el modelo de relación entre la variable que permanece en el modelo incorrectamente especificado y la variable omitida.

**Tabla 13.1: Consecuencias de la omisión de variables,  
ejemplo práctico**

| <b>Modelo correcto</b>                                    |                                                                                  |
|-----------------------------------------------------------|----------------------------------------------------------------------------------|
| • Estimación                                              | $\hat{Y}_t = -859.92 + 0.6470X_{2t} - 23.195X_{3t}$                              |
| • Errores estándar                                        | (11.96) (0.0745) (4.2704)                                                        |
| • Razones-t                                               | (-7.6806) (8.6846) (-5.4316)                                                     |
| • Otros indicadores                                       | $\hat{\sigma}^2 = 184.05$ ; $R^2 = 0.9776$ ; $\bar{R}^2 = 0.9750$ ; $d = 1.36$   |
| <b>Modelo con variable omitida</b>                        |                                                                                  |
| • Estimación                                              | $\hat{Y}_t = -261.009 + 0.2452X_{2t}$                                            |
| • Errores estándar                                        | (31.327) (0.0148)                                                                |
| • Razones-t                                               | (-8.334) (16.5676)                                                               |
| • Otros indicadores                                       | $\hat{\sigma}^2 = 475.48$ ; $R^2 = 0.9388$ ; $\bar{R}^2 = 0.9354$ ; $d = 0.5951$ |
| <b>Relación entre <math>X_3</math> y <math>X_2</math></b> |                                                                                  |
| • Estimación                                              | $\hat{X}_{3t} = -25.817 + 0.0173X_{2t}$                                          |
| • Razones-t                                               | (-23.999) (34.177)                                                               |
| • Otros indicadores                                       | $R^2 = 0.9848$                                                                   |

*Fuente: Elaboración propia*

Fíjense como en el modelo incorrectamente especificado el parámetro de  $X_{2t}$  indica un efecto sobre  $Y_t$  menor que el real. Así se manifiesta el sesgo en el coeficiente de regresión, porque existe una fuerte relación entre la variable omitida  $X_{3t}$  y  $X_{2t}$  (vean los resultados más abajo en la misma tabla). En este caso el coeficiente estimado para  $X_{2t}$  absorbe parte del efecto de la variable  $X_{3t}$  y, como el coeficiente  $B_3$  es negativo y  $b_{32}$  positivo, el resultado es un coeficiente de regresión  $\hat{a}_2$  menor que  $B_2$ . Dado que  $E(\hat{a}_2) = B_2 + B_3b_{32}$ , en este caso tendremos que  $B_2 + B_3b_{32} = 0.6470 + (-23.195)(0.0173) = 0.2452$ . También se observa una varianza del término de error mucho mayor en el modelo carente de  $X_{3t}$ , que en el correcto.

Asimismo, obtenemos una mejor bondad de ajuste en el modelo correcto que en el modelo mal especificado: recuerden que debemos comparar los  $\bar{R}^2$ , pues se trata de modelos con diferente número de variables. Igualmente vemos un cambio

significativo en el estadístico de Durbin-Watson, que corrobora lo que habíamos planteado en el capítulo décimo, acerca de que este índice también nos puede señalar otros problemas de especificación, además de la detección de autocorrelación.

Resumiendo, la omisión de variables relevantes puede tener consecuencias muy serias. Por lo tanto, es muy importante que en el desarrollo de un modelo para el análisis empírico, prestemos atención a la teoría apropiada que subyace al fenómeno bajo estudio, de modo que nos aseguremos que todas las variables relevantes sean incluidas.

### 13.3 Inclusión de Variables Irrelevantes

En este epígrafe examinaremos el caso opuesto a la omisión de variables. Veremos qué efectos puede tener la inclusión de variables que no resultan relevantes al fenómeno en cuestión. Aunque este problema pertenece más bien a una época pasada, cuando se prestaba mucha atención a la inclusión de más variables en los modelos por su aporte al incremento del  $R^2$ , es interesante examinar los efectos de este proceder, por si persisten algunas tentaciones en este sentido. A la práctica anteriormente aludida de construir modelos la llamamos enfoque “vertedero”<sup>2</sup>, pues tiende a incluir todo tipo de variables, estén o no teóricamente justificadas. Este proceder es muy usual en el caso de que la teoría subyacente no esté bien desarrollada. Cuando incluimos variables no relevantes, decimos que estamos “sobreajustando” o “sobreespecificando” el modelo.

Supongamos que el modelo regresión real (verdadero) es  $Y_t = B_1 + B_2 X_{2t} + u_t$ , y en su lugar estimamos  $Y_t = a_1 + a_2 X_{2t} + a_3 X_{3t} + v_t$ , donde  $X_3$  es una variable no relevante. Las consecuencias de dicha sobreespecificación son las siguientes:

- Los estimadores MCO del modelo incorrecto son insesgados. En el ejemplo anterior tendremos  $E(\hat{a}_1) = B_1$ ,  $E(\hat{a}_2) = B_2$  y  $E(\hat{a}_3) = 0$ .
- El estimador de la varianza  $\sigma^2$  está correctamente estimado.
- El intervalo de confianza estándar y los procedimientos de prueba de hipótesis permanecen válidos.
- Sin embargo, el estimador de los parámetros del modelo incorrecto será ineficiente. Su varianza será generalmente mayor que aquella de los estimadores del modelo correcto. Entonces, fallaremos en rechazar la

<sup>2</sup> *Kitchen-sink*, en inglés.



hipótesis nula más frecuentemente que lo necesario. Este es un problema potencialmente serio pues nos puede llevar a un error Tipo II, que puede conducirnos a confiar en una hipótesis falsa.

Por lo visto, al parecer, la inclusión de variables irrelevantes es un error menos serio que excluir las relevantes. Así, podríamos vernos tentados a incluir variables adicionales, aun cuando no estemos seguros sobre su importancia, en el afán de no omitir aspectos sobre los que no estamos absolutamente seguros. No obstante, debemos enfatizar que al proceder de esa manera:

- Los estimadores de los parámetros del modelo incorrecto son ineficientes.
- Podemos estar añadiendo multicolinealidad.
- Perdemos grados de libertad al incorporar más variables, lo que puede ser crítico cuando nuestra muestra sea pequeña.

Mostramos los problemas asociados a la inclusión de variables no relevantes en un modelo de regresión, mediante la estimación de los modelos antes aludidos. Veamos qué sucede con los estimados de los parámetros de interés. Los resultados los hemos agrupado en la tabla 13.2.

Noten que las estimaciones de las constantes y de los coeficientes de regresión correspondientes a la variable  $X_2$  difieren en menos de un punto porcentual entre las dos regresiones. Sin embargo, los errores estándar son mayores y las razones-t menores para la regresión que contiene la variable irrelevante, sobre todo en el caso del regresor  $X_2$ , para el que el error estándar en la regresión con la variable irrelevante prácticamente duplica al obtenido con el modelo correcto. Fíjense también en que los coeficientes de la regresión que incluye la variable irrelevante estimada con una observación menos en la muestra, difieren en alrededor de un 4% respecto de los de la regresión para el período completo en el caso de la constante y el correspondiente a  $X_2$ , pero el correspondiente a  $X_3$  es radicalmente diferente en nivel y signo. Esta inestabilidad en el coeficiente de  $X_3$  se debe a la relación lineal que existe entre esa variable, innecesariamente incorporada a la regresión, y la variable  $X_2$ , que introduce los problemas característicos de la existencia de multicolinealidad aproximada. Recuerden que la multicolinealidad es un problema esencialmente muestral, por lo que el agregar (o también eliminar) una observación, puede introducir variaciones importantes en los coeficientes de regresión estimados. De hecho, este procedimiento de eliminar una observación del inicio o del final de la muestra se utiliza también como indicador para evaluar la presencia de multicolinealidad.<sup>3</sup>

<sup>3</sup> Repasen el Capítulo 8 de este texto.

**Tabla 13.2: Consecuencias de la inclusión de variable irrelevante, ejemplo práctico**

| <b>Modelo correcto</b>                                         |                                                    |
|----------------------------------------------------------------|----------------------------------------------------|
| • Estimación                                                   | $\hat{Y}_t = 2.7996 + 1.1270X_{2t}$                |
| • Errores estándar                                             | (1.8573) (0.0438)                                  |
| • Razones-t                                                    | (1.5073) (25.7029)                                 |
| • Otros indicadores                                            | $R^2 = 0.9323$ ; $\bar{R}^2 = 0.9309$              |
| <b>Modelo con variable irrelevante</b>                         |                                                    |
| • Estimación                                                   | $\hat{Y}_t = 2.8127 + 1.1256X_{2t} + 0.0018X_{3t}$ |
| • Errores estándar                                             | (2.2076) (0.1296) (0.1616)                         |
| • Razones-t                                                    | (1.2741) (8.6870) (0.0113)                         |
| • Otros indicadores                                            | $R^2 = 0.9323$ ; $\bar{R}^2 = 0.9294$              |
| <b>Modelo con variable irrelevante y una observación menos</b> |                                                    |
| • Estimación                                                   | $\hat{Y}_t = 2.7129 + 1.1692X_{2t} - 0.0657X_{3t}$ |
| • Errores estándar                                             | (2.1157) (0.1256) (0.1576)                         |
| • Razones-t                                                    | (1.2822) (9.3071) (-0.4170)                        |
| • Otros indicadores                                            | $R^2 = 0.9377$ ; $\bar{R}^2 = 0.9350$              |
| <b>Relación entre <math>X_3</math> y <math>X_2</math></b>      |                                                    |
| • Estimación                                                   | $\hat{X}_{3t} = -5.5276 + 0.7147X_{2t}$            |
| • Razones-t                                                    | (-3.2617)(17.8632)                                 |
| • Otros indicadores                                            | $R^2 = 0.8692$                                     |

Fuente: Elaboración propia

Una situación bastante característica de la introducción de variables irrelevantes es la modelación de eventos extraordinarios o de cambios en la estructura de los modelos de regresión a través de variables mudas. Ya examinamos en el capítulo siete de este texto que estas variables deben ser incorporadas a los modelos con cuidado, porque son altamente colineales con el intercepto o

constante de la regresión y ellas mismas tienen escasa variabilidad, todo lo que hará que nuestras estimaciones sean imprecisas. De hecho, debemos tener la precaución de introducir una variable muda menos que categorías tenga el aspecto cualitativo que queremos modelar, para evitar caer en “la trampa de la muda”. Así, se recomienda utilizar como máximo una cantidad de variables de intervención equivalente a un 5% del número de observaciones en la muestra disponible.

Una vez más, les recordamos que muchas medidas correctivas o paliativas en Econometría aportan beneficios, pero también tienen sus costos. Se trata de una cuestión de costos y beneficios. Muy a menudo lo que ganamos en una dirección lo perdemos en otra. Entender esto es esencial para proceder con sensatez en Econometría.

### 13.4 Forma Funcional Incorrecta

Otro error de especificación en que podemos incurrir consiste en plantear una forma funcional inadecuada para la relación entre las variables. Desafortunadamente, la teoría económica pocas veces nos instruye con precisión acerca de la forma funcional correcta. Sin embargo, la utilización de la forma funcional adecuada es tan importante como incluir todas las variables relevantes, porque en caso de equivocación, estaremos obteniendo estimadores sesgados para nuestros parámetros de interés: coeficientes de regresión y varianza del término de error. Además, la varianza será heteroscedástica, por lo que estaremos violando también el supuesto de homoscedasticidad.

Un ejemplo sobre posibles formas funcionales para describir un determinado proceso podemos esbozarlo usando nuestra función de demanda de importaciones para Cuba. En el epígrafe 5.1 la hemos planteado como una función lineal en las tasas de variación de las importaciones, del producto interno bruto y de los precios relativos de las importaciones:  $\dot{m}_t = \beta_1 \dot{y}_t + \beta_2 \dot{p}_t + \beta_3$ . Sin embargo, podríamos concebir también este modelo utilizando una función log-lineal o doble-log, como se muestra seguidamente:

$$LM_t = \alpha_1 LY_t + \alpha_2 LP_t + \alpha_3$$

donde  $M$  es el nivel de importaciones,  $Y$  el del producto y  $P$  el de los precios relativos de las importaciones; la  $L$  que antecede a cada variable, como ya hemos convenido anteriormente, representa que están expresadas en términos de logaritmos; y el subíndice  $t$  indica que tenemos series de tiempo.

En la tabla 13.3 hemos resumido los resultados de la estimación de ambos modelos. Noten las similitudes y también las diferencias entre los coeficientes de regresión obtenidos en una y otra variante de la función de demanda de importaciones cubanas. ¿Cuál será la forma funcional “correcta” para esta función de demanda de importaciones?

Tabla 13.3: Función de demanda de importaciones en Cuba, dos posibles formas funcionales

| Modelo lineal en las tasas de variación |                                                               |
|-----------------------------------------|---------------------------------------------------------------|
| • Estimación                            | $\hat{m}_t = 1,26270 + 0,75064 \dot{y}_t - 0,54252 \dot{p}_t$ |
| • Razones-t                             | (0.7673) (3.3604) (-6.0011)                                   |
| • Otros indicadores                     | $R^2 = 0.5705$ ; $\bar{R}^2 = 0.5500$                         |
| Modelo loglineal                        |                                                               |
| • Estimación                            | $LM_t = 3,17510 + 0,52201 LY_t - 0,73632 LP_t$                |
| • Razones-t                             | (6.3822) (10.5216) (-11.5635)                                 |
| • Otros indicadores                     | $R^2 = 0.8494$ ; $\bar{R}^2 = 0.8422$                         |

Fuente: Elaboración propia

Fíjense como los resultados del modelo loglineal van aventajando a los del modelo lineal en las tasas de variación, pues las razones-t y la bondad de ajuste son mejores. Sin embargo, todavía no estamos en condiciones de responder a la pregunta previa, pues no tenemos aun los conocimientos necesarios para evaluar si la función de demanda de importaciones estimada a partir de los logaritmos de las variables en niveles corresponde a un modelo correctamente especificado.

13.5 Detectando Errores de Especificación

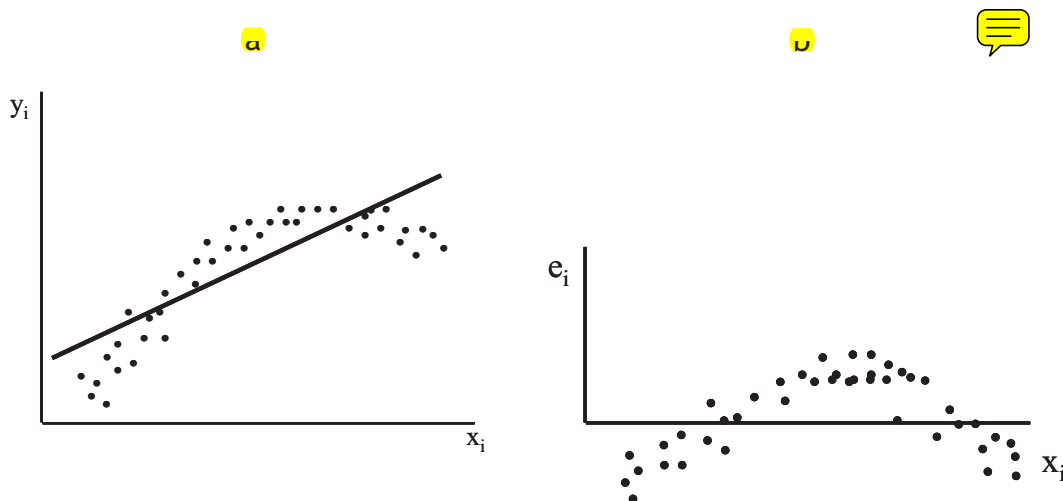
En general, los errores de especificación que hemos examinado en esta parte del libro tienen su origen en debilidades de la teoría misma o de la formación teórica de los que emprendemos el ejercicio de modelación. Igualmente puede suceder que no dispongamos del tipo de datos adecuado para poder someter a prueba un modelo teórico correcto, o que la teoría no nos proporcione información sobre la forma funcional correcta. El remedio a estos errores es bien directo: utilizar la especificación correcta, ¿pero cómo hacerlo?, ¿cómo detectar que estamos incurriendo en un error de especificación de los anteriormente abordados?

Un buen punto de partida –tal y como hicimos también cuando investigábamos la presencia de errores no esféricos–, es el examen de los residuos. Por eso es que todos los paquetes computacionales econométricos traen incorporada la opción de graficar los residuos. En el caso de que hayamos cometido algún error de especificación como la omisión de una variable relevante o la aplicación de una forma

funcional inadecuada, encontraremos ciertos patrones de comportamiento en los residuos del modelo estimado, impropios de un error estocástico.

Por ejemplo, en la figura 13.1 mostramos dos paneles. El (a) representa el gráfico de la variable dependiente contra la independiente y en él hemos dibujado la regresión estimada como función lineal. Observen como los verdaderos valores de  $Y$  se ubican primero por debajo, después por encima y, finalmente, de nuevo por debajo de la línea de regresión estimada. Es obvio que la forma funcional que mejor va a describir esta dependencia entre  $y_i$  y  $x_i$  no es precisamente una recta, sino una forma cuadrática. En el panel (b) hemos graficado el patrón de residuos estimados para la recta de regresión que correspondería a esta situación.

**Figura 13.1:** Gráficos de residuos en caso de una incorrecta especificación de la forma funcional: una recta en lugar de una parábola.



*Fuente: Elaboración propia*

Como ya mencionamos anteriormente, el estadístico de Durbin-Watson es también un buen indicador de problemas de incorrecta especificación. Siempre que obtengan una  $d$  pequeña, sospechen de la presencia de algún problema.

Muchas de las pruebas para detectar los errores de especificación se basan en regresiones auxiliares que utilizan como variable independiente los residuos del modelo de regresión postulado. Para llegar a esa regresión auxiliar se parte del modelo correcto, del modelo postulado y de cómo deberían comportarse los residuos del modelo postulado si fuese correcto. Vamos a reconstruir el razonamiento para llegar a esa regresión.

Supongamos que el modelo correcto es  $Y_i = \bar{\mu}_i + \xi_i$ , donde  $\bar{\mu}_i$  es la media condicionada real de  $Y_i$  y  $\xi_i$ , la parte no sistemática real.

Hemos modelado el proceso a través de  $Y_t = \mu_{t(\beta)} + \varepsilon_t$ , donde  $\mu_{t(\beta)}$  es la media modelada (en nuestro MRLN,  $\mu_{t(\beta)} = X_t\beta$ ) y  $\varepsilon_t$ , la innovación del modelo.

A partir del modelo postulado y los datos, estimamos  $\hat{\beta}$  y conseguimos también los residuos estimados,  $e_t = Y_t - X_t\hat{\beta}$

Estos residuos estimados podemos descomponerlos como  $e_t = \bar{\mu}_t + \varepsilon_t - X_t\hat{\beta} = \bar{\mu}_t + \varepsilon_t - \mu_{t(\hat{\beta})}$

Si sumamos y restamos en la parte derecha de la expresión anterior  $\mu_{t(\beta)}$  y agrupamos sus términos convenientemente, podemos llegar a un interesante equivalente para los residuos estimados:

$$\begin{aligned} e_t &= \bar{\mu}_t + \varepsilon_t - \mu_{t(\hat{\beta})} \\ &= \bar{\mu}_t + \varepsilon_t - \mu_{t(\hat{\beta})} + \mu_{t(\beta)} - \mu_{t(\beta)} \\ e_t &= \left[ \bar{\mu}_t - \mu_{t(\beta)} \right] + \left\{ \mu_{t(\beta)} - \mu_{t(\hat{\beta})} \right\} + \varepsilon_t = EEM + EE + \varepsilon_t \end{aligned} \quad (13.1)$$

El término entre corchetes en la expresión (13.1) representa la diferencia entre la media real del proceso y la media modelada, y lo denominamos error de especificación de la media (*EEM*). Si nuestro modelo es correcto, ese error no debe diferenciarse significativamente de cero.

El segundo término entre llaves es la diferencia entre la media modelada y la media estimada con la muestra disponible. Así, representa el error que surge por la diferencia entre los parámetros reales  $\beta$  y los parámetros estimados  $\hat{\beta}$ , y lo denominamos error de estimación (*EE*). Este error puede ser expresado como  $EE = X_t\beta - X_t\hat{\beta} = X_t(\beta - \hat{\beta}) = X_t\gamma$ . Debe ser mínimo, justamente porque estimamos con MCO.

Entonces, en la expresión (13.1) parametrizamos el *EEM*, en función de un conjunto de variables  $W_t$ , que son las que podrían estar faltando en la especificación de la media condicional y que, por lo tanto, estarían incluidas en el error estimado y supuestamente estocástico del modelo postulado,  $e_t$ . Por eso, muchas pruebas de diagnóstico de incorrecta especificación se basan en la regresión auxiliar:

$$e_t = W_t\alpha + X_t\gamma + \varepsilon_t \quad (13.2)$$

Esta regresión auxiliar tiene un equivalente, con  $Y_i$  como variable independiente:<sup>4</sup>

$$Y_i = W_i\alpha + X_i\beta + \varepsilon_i \quad (13.3)$$

Si al sustituir  $W_i$  en (13.2) por una variable específica y someter a prueba la hipótesis  $\alpha = 0$ , esta es rechazada, eso quiere decir que, efectivamente, esa variable está incluida en  $e_i$ , que dejaría de comportarse de forma no sistemática, como debe ser en un MRLN.

Así, las regresiones (13.2) o (13.3) se utilizan para el test de variables relevantes omitidas en la regresión. Si llegamos a comprobar que  $\alpha$  es significativamente distinto de cero, a través de la prueba t que bien conocemos, entonces podemos decir que la variable  $W_i$  está faltando en la especificación de la media condicionada de  $Y_i$ .

Por ejemplo, vamos a aplicar esta prueba a nuestra función de demanda de importaciones para Cuba con el propósito de detectar posibles variables omitidas. En este test tenemos que postular candidatos a haber sido pasados por alto en la explicación de la tasa de variación de las importaciones. Hemos incluido en el modelo la variación del ingreso del país y de los términos de intercambio, que incorporan la variación combinada de precios de importaciones y exportaciones. No tenemos otros aspirantes a incluir en el modelo, por lo que vamos a probar con algunos retardos o rezagos de las variables explicativas y también de la dependiente. Como ya hemos explicado antes, los retardos no son más que observaciones correspondientes a períodos anteriores al contemporáneo. Incluir este tipo de variables en el modelo nos aleja de un MRLN y nos lleva directo a un modelo dinámico. En la Parte Cuatro de este texto se abordarán la pertinencia y las particularidades de este tipo de modelos.

Por ahora vamos a conformarnos con una sencilla explicación para entender el significado de incorporar variables retardadas a la explicación de un fenómeno. En el caso que nos ocupa, si introduyéramos en el modelo, por ejemplo, rezagos de la tasa de variación de los precios relativos de las importaciones, ello querría decir que la variación contemporánea de las importaciones dependerá también de cuánto hayan variado sus precios relativos en períodos precedentes. ¿Existe una razón económica para suponer que ocurra un vínculo retardado entre nuestras variables? Probablemente el contexto económico anterior a 1989 sea un buen ejemplo para este tipo de relación, cuando el plan de cada quinquenio reflejaba los cambios en los precios del mercado mundial correspondientes al lustro anterior y también las variaciones del año anterior, para el caso de que los bienes fueran adquiridos en el área de moneda libremente convertible.

<sup>4</sup> Noten que  $e_i = W_i\alpha + X_i\gamma + \varepsilon_i = W_i\alpha + X_i\beta - X_i\hat{\beta} + \varepsilon_i$ , por lo que  $Y_i = X_i\hat{\beta} + e_i = W_i\alpha + X_i\beta + \varepsilon_i$ .

Asimismo, la inclusión de la propia variable dependiente retardada en un modelo lo califica como autorregresivo. Este tipo de modelos es muy poderoso y rico en información, pues al incorporar los rezagos de la propia variable dependiente estamos introduciendo toda la historia anterior del fenómeno para su explicación.

En la tabla 13.4 hemos resumido los resultados de las pruebas de omisión de variables para la función de demanda de importaciones cubanas con las variables expresadas como tasas de variación. Fíjense que las variables rezagadas las hemos expresado con un subíndice  $t-r$ , con  $r$  el orden del retardo o rezago en cuestión. Vamos a incluir hasta cinco rezagos, por la particularidad de la planificación quinquenal.

Sólo hemos reportado los resultados para las pruebas con los rezagos de la variable dependiente, donde obtenemos un test-t favorable correspondiente al tercer retardo. Para los retardos de las explicativas no obtenemos resultados significativos.

Otro posible EEM es la inclusión de variables irrelevantes. En este caso nos guiamos por el test de significación individual en la regresión. No obstante, en cuanto a la eliminación de variables de una regresión atendiendo a su significación, es importante recalcar que si contamos con una teoría sólida sobre el proceso que modelamos, no debemos excluir variables teóricamente relevantes, aun cuando sus coeficientes sean estimados con un bajo nivel de significación, según los test t.

Mas cuando se trate de variables de control, que hemos incluido en la regresión para evitar el sesgo de variables omitidas, podríamos eliminarlas si ellas resultaran no significativas, siempre que este proceder no afecte drásticamente los estimadores puntuales del resto de los coeficientes de regresión. ¿Cuál será el umbral que nos indique si ha ocurrido o no un cambio drástico en los estimadores? Consideramos que el cambio es drástico, si el valor absoluto de la diferencia entre los estimadores antes y después de la exclusión de la variable de control rebasa el 68,27% del error estándar del estimador en la regresión que incluye todas las variables.

Un buen consejo asociado a esta cuestión de las variables irrelevantes, es evitar construir los modelos de forma iterativa. A este proceder lo encontramos en la literatura también con otras denominaciones como regresiones paso a paso, enfoque ascendente, búsqueda exhaustiva de datos, regresión al tanteo, exhumación de datos, rastreo de datos y trituración de datos. Nuevamente reafirmamos que en el trabajo empírico siempre debemos partir de la teoría y evitar la medida sin teoría que la sustente.

El posible error en la forma funcional también hay que investigarlo. En primer lugar, debemos verificar si aun siendo no lineal un modelo, es posible convertirlo a uno lineal aplicando alguna transformación matemática, con lo que podríamos circunvalar esta dificultad.



**Tabla 13.4: Función de demanda de importaciones cubanas en tasas de variación, test de variables omitidas**

|                                                                                                        | Valor                                                                                                                                                                         | Probabilidad |
|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| <b>Variables omitidas:</b> $\dot{m}_{t-1}, \dot{m}_{t-2}, \dot{m}_{t-3}, \dot{m}_{t-4}, \dot{m}_{t-5}$ |                                                                                                                                                                               |              |
| Estadístico-F                                                                                          | 3.318618                                                                                                                                                                      | 0.015854     |
| Razón de verosimilitud                                                                                 | 16.70982                                                                                                                                                                      | 0.005084     |
| <b>Regresión auxiliar</b>                                                                              |                                                                                                                                                                               |              |
| • Estimación                                                                                           | $\hat{m}_t = 0,6307 \dot{y}_t - 0,6089 \dot{p}_t + 1,8850 + 0,1007 \dot{m}_{t-1} + 0,2255 \dot{m}_{t-2} - 0,3300 \dot{m}_{t-3} - 0,0644 \dot{m}_{t-4} + 0,0265 \dot{m}_{t-5}$ |              |
| • Errores estándar                                                                                     | (0.2743) (0.1048) (1.5717)<br>(0.1314) (0.1210) (0.1123) (0.1215) (0.1095)                                                                                                    |              |
| • Razones-t                                                                                            | (2.2990) (-5.812) (1.1993)<br>(0.7663) (1.8639) (-2.940) (-0.530) (0.2422)                                                                                                    |              |
| • Probabilidad                                                                                         | (0.0282) (0.0000) (0.2392)<br>(0.4491) (0.0715) (0.0061) (0.6000) (0.8102)                                                                                                    |              |

Fuente: Elaboración propia

En caso de que no sea posible, tenemos disponibles dos enfoques. Uno que parte de postular una determinada función no lineal para realizar la prueba. Nuestras hipótesis nula y alternativa son:

$$H_0 : Y_t = X_t \beta + e_t$$

$$H_1 : Y_t = g(X_t) + v_t, \text{ siendo } g(X_t) \text{ una función no lineal}$$

Para desarrollar esta prueba es posible aplicar la ya conocida transformación Box-Cox, que plantea que:

$$g(X_t, \lambda) = \lambda^{-1} (X_t^\lambda - i) \delta$$

donde  $\delta \neq 0$  es un vector de números reales cualquiera.

Ya sabemos que cuando  $\lambda = 1$ ,  $g(X_t, \lambda)$  es lineal. Aplicamos entonces la expansión de Taylor alrededor de  $\lambda = 1$  y obtenemos la siguiente aproximación:

$$g(X_t, \lambda) \cong X_t \beta + (\lambda - 1) \delta [X_t \ln(X_t) - X_t - i]$$

Así, podemos definir  $\alpha = \lambda - 1$  y  $W_t = X_t \ln(X_t) - X_t - i$  y construir una prueba de linealidad como:

$$H_0 : Y_t = X_t\beta + e_t$$

$$H_1 : Y_t = X_t\beta + W_t\alpha + e_t$$

Fíjense que la regresión bajo  $H_1$  es similar a (13.3). Esta prueba es equivalente a:

$H_0 : \alpha = 0$ , que significa que la función es lineal (porque  $\lambda=1$ )

$H_1 : \alpha \neq 0$ , que significa que no lo es.

El estadístico a utilizar para este test puede ser el de Wald (prueba tipo  $F$ ), o el de multiplicadores de Lagrange o el de razón de verosimilitud.

Si rechazáramos  $H_0$ , podríamos probar otras formas funcionales derivadas de la transformación Box-Cox, como pueden ser:

$H_0 : \alpha = -1$ , que significa que la función es logarítmica (porque  $\lambda=0$ );

$H_0 : \alpha = 0.5$ , que significa que la función es raíz cuadrada (porque  $\lambda=0.5$ );

$H_0 : \alpha = -2$ , que significa que la función es raíz el recíproco (porque  $\lambda=-1$ ).

El otro enfoque es plantear como hipótesis opcional a la linealidad, una aproximación polinómica. Esta prueba fue diseñada por Ramsey y la conocemos como el test RESET.<sup>5</sup> Las hipótesis nula y alternativa son en este caso:

$$H_0 : e_t = X_t b + \varepsilon_t$$

$$H_1 : e_t = X_t b + W_t \alpha + \varepsilon_t$$

De nuevo en  $H_1$  la regresión (13.2), o también su equivalente (13.3):

$$H_0 : Y_t = X_t\beta + \varepsilon_t$$

$$H_1 : Y_t = X_t\beta + W_t\alpha + \varepsilon_t$$

Noten que en el caso de la prueba RESET, la variable  $W_t$  adopta una forma particular:  $W_{kt} = (X_t\hat{\beta})^k = \hat{Y}_t^k$ . Por lo general se hacen pruebas para  $k=2$ ,  $k=3$  o  $k=4$ . Por ejemplo, para  $k=3$  la regresión auxiliar sería  $e_t = X_t b + \hat{Y}_t^2 \alpha_2 + \hat{Y}_t^3 \alpha_3 + \varepsilon_t$ . Esta prueba nos permite comprobar si efectivamente nuestro modelo es lineal o si es más adecuado un polinomio de mayor grado. Este test sirve para investigar sobre la forma funcional, pero no únicamente, pues la aproximación polinómica en que se sustenta también puede capturar otros problemas de especificación.

<sup>5</sup> RESET es un acrónimo en idioma inglés que significa *Regression Specification Error Test*. Estos comentarios sobre la prueba RESET los hemos tomado de Cassoni (1991, p. 36).

Vamos a ilustrar la aplicación de la prueba RESET para el caso de nuestras funciones de demanda de importaciones en sus dos variantes, estimada a partir de las tasas de variación y de los logaritmos de las variables en niveles. Los resultados los recogemos en la tabla 13.5.<sup>6</sup>

**Tabla 13.5: Función de demanda de importaciones en Cuba, test RESET de Ramsey para dos posibles formas funcionales**

|                                     | Valor    | Probabilidad |
|-------------------------------------|----------|--------------|
| <b>Lineal en tasas de variación</b> |          |              |
| • $k=2$                             |          |              |
| Estadístico F                       | 4,071645 | 0,050185     |
| Razón de verosimilitud              | 4,260657 | 0,039005     |
| • $k=3$                             |          |              |
| Estadístico F                       | 2,040636 | 0,143258     |
| Razón de verosimilitud              | 4,372000 | 0,112365     |
| <b>Loglineal</b>                    |          |              |
| • $k=2$                             |          |              |
| Estadístico F                       | 5,601771 | 0,022746     |
| Razón de verosimilitud              | 5,762992 | 0,016367     |
| • $k=3$                             |          |              |
| Estadístico F                       | 4,727181 | 0,014358     |
| Razón de verosimilitud              | 9,547686 | 0,008448     |

*Fuente: Elaboración propia*

Noten que en el modelo en tasas de variación rechazamos la hipótesis de linealidad para  $k=2$  con el estadístico razón de verosimilitud y  $\alpha=5\%$ ; para el F estamos asimismo muy cerca del rechazo. Sin embargo, dados los resultados de las investigaciones a las que hemos venido sometiendo a este modelo, no podríamos afirmar categóricamente que este hallazgo pueda atribuirse a la no-linealidad en  $\beta$  de la media condicionada de  $\dot{m}_t$ . Hasta este punto hemos logrado verificar, por

<sup>6</sup> Estos resultados los hemos obtenido con el software EViews. Para correr el test de Ramsey, este programa pide el “number of fitted terms”, es decir, el número de potencias de  $Y_t$  que se van a introducir en la regresión auxiliar. Entonces, cuando decimos que sólo introduciremos una potencia, esto equivale a  $k=2$ ; si decimos que dos potencias,  $k$  será tres; y así sucesivamente.

ejemplo, que esta regresión presenta: un cambio estructural hacia 1989, evidencia a favor de autocorrelación en los residuos, observaciones atípicas y heteroscedasticidad tipo ARCH. Además, tampoco hemos verificado si hemos omitido alguna variable importante en esta especificación de la función de demanda de importaciones. Así, este resultado en la prueba RESET puede estar originado por las fallas ya detectadas.

Para el modelo loglineal tenemos un franco rechazo de la linealidad o también de la no-existencia de otros EEM con  $\alpha = 5\%$ . Recuerden, sin embargo, que no estamos seguros de que esta regresión no sea espuria. Como ya les vaticinamos, al cierre de la Parte Cuatro de este texto deberán estar en forma para poder llegar a una conclusión al respecto.

### 13.6 Comprobando la Estabilidad de los Coeficientes de Regresión

En un MLNR los parámetros de la regresión se mantienen invariables entre individuos, si estamos trabajando con series de corte transversal, o en el tiempo. Esta condición debemos investigarla también en nuestros casos de estudio. Podemos incurrir en un error asumiendo constante un coeficiente de regresión que en realidad no lo es.

Ya en el Capítulo 7 de este texto abordamos el test de Chow para detectar cambios de estructura en los parámetros de un modelo de regresión, y también explicamos de qué forma podemos modelar dichos cambios a partir de la utilización de variables binarias o dicotómicas.

Para aplicar el test de Chow precisamos tener una idea acerca de en qué momento del tiempo es posible que haya ocurrido el cambio estructural en el modelo. Esta prueba nos ayuda a conseguir evidencia a favor o en contra de esa hipótesis. Sin embargo, puede darse el caso de que no sepamos cuándo ocurre el punto de inflexión. En este caso tenemos a nuestra disposición el método de los *mínimos cuadrados recursivos* (MCR).

La idea en que se basa este método es relativamente intuitiva y sencilla.

Vamos a explicarlo a partir de nuestro ejemplo de la función de demanda de importaciones cubanas. Para la modelación de ese proceso tenemos una muestra que abarca 45 años, de 1961 al 2005.<sup>7</sup> Nuestro modelo tiene tres coeficientes de regresión ( $k = 3$ ), por lo que necesitaremos para conseguir estimarlos, al menos tres observaciones de la muestra. Estimamos entonces un primer *set* de coeficientes para el período 1961-1963. Después volvemos a estimar los coeficientes

<sup>7</sup> Recuerden que hemos perdido el año 1960 porque estamos trabajando con las tasas de variación de nuestras variables.

incorporando una observación más, es decir, para el lapso 1961-1964. Seguidamente incorporamos un dato más a la estimación, consiguiendo estimados para el quinquenio 1961-1965. Y así sucesivamente, hasta que agotamos todas las observaciones de la muestra. Es decir, vamos a calcular un juego de coeficientes para cada observación de la muestra posible entre  $k$  y  $T$ , siendo  $T$  la cantidad de observaciones disponibles.

La rutina de estimación de los coeficientes recursivos es la siguiente:

$$\tilde{\beta}^r = (\mathbf{X}^{r'} \mathbf{X}^r)^{-1} \mathbf{X}^{r'} \mathbf{Y}^r, \text{ con } r = \{k, k+1, k+2, \dots, T\}$$

Como ven, es la misma habitual de estimación de  $\hat{\beta}$ , lo único que cambia son los insumos para cada estimación:

$$\mathbf{X}^r = \begin{bmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{22} & X_{23} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{r2} & X_{r3} & \cdots & X_{rk} \end{bmatrix} \quad \mathbf{Y}^r = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_r \end{bmatrix}, \text{ con } r = \{k, k+1, k+2, \dots, T\}$$

Los coeficientes obtenidos de esta forma son graficados con respecto al tiempo, tomando como abscisa el extremo superior del intervalo temporal para el que fueron estimados. Este gráfico nos dará información útil sobre la estabilidad o desvarío de nuestros coeficientes de regresión.

Otro gráfico interesante es el de los residuos conseguidos con cada una de esas estimaciones para el período contiguo al extremo de cada intervalo. Estos residuos los llamamos residuos recursivos y, como habrán notado, los estamos definiendo como el error de predicción a un paso. Los utilizamos para realizar variadas pruebas de diagnóstico sobre la estabilidad estructural de los modelos de regresión. Los paquetes computacionales econométricos por lo general traen incluida la rutina de cálculo de los residuos recursivos y las pruebas asociadas con el cambio estructural a partir de éstos. El proceder para obtener este tipo de residuos lo formalizamos de la siguiente manera:

$$\omega_t = Y_t - X_t \tilde{\beta}^{t-1}, \text{ con } t = \{k+1, k+2, \dots, T\}$$

El solo examen del gráfico de los residuos recursivos ya nos da pistas sobre los momentos en que los parámetros son inestables: cuando se desbordan del intervalo alrededor de ellos, equivalente a dos errores estándar por encima y por debajo.

Otras pruebas a partir de los residuos recursivos son las denominadas CUSUM y CUSUMQ.<sup>8</sup> La prueba CUSUM construye un estadístico  $C_t$  a partir de la suma acumulada de los residuos recursivos estandarizados:<sup>9</sup>

$$C_t = \sum_{m=k+1}^t \frac{\omega_m}{s}, \text{ que obtenemos para cada } t=\{k+1, \dots, T\}$$

La idea es que si los coeficientes de regresión son estables, el valor esperado de esta suma debe ser cero y cuando lo grafiquemos con respecto al tiempo, se moverá aleatoriamente alrededor del eje de las abscisas. Pero si los coeficientes cambian, los residuos se harán cada vez grandes y la suma de ellos se alejará de cero, en un sentido u otro. Bajo la hipótesis nula de que no hay cambio estructural, este estadístico tiene una distribución  $N(0, t-k)$ .

En el caso del test CUSUMQ lo que hacemos es utilizar un estadístico  $CQ_t$  que construye la suma acumulada los residuos recursivos al cuadrado hasta el momento  $t$ , y la divide entre idéntica suma pero hasta el final del período muestral, como se indica:

$$CQ_t = \frac{\sum_{m=k+1}^t \omega_m^2}{\sum_{m=k+1}^T \omega_m^2}, \text{ para } t = \{k+1, k+2, \dots, T\}$$

Este estadístico, a diferencia de  $C_t$ , siempre será positivo y bajo  $H_0$  se distribuye como una beta, con media  $(t-k)/(T-k)$ .

Ambos estadísticos se grafican, así como los intervalos de confianza correspondientes a un 5% de significación. Un movimiento de  $C_t$  o de  $CQ_t$  fuera de los límites de estos intervalos, nos sugiere inestabilidad en los parámetros de la regresión. En realidad, aun cuando no desborden las líneas críticas, el mero hecho de una deriva de  $C_t$  lejos del eje de abscisas, ya nos da que sospechar acerca de un cambio estructural.

Todas estas pruebas nos dan indicios acerca de cuándo podemos fechar los posibles cambios estructurales en un modelo de regresión, en el caso de que nuestro conocimiento sobre el contexto tenga sus limitaciones.

Veamos la aplicación de estas pruebas a nuestro ejemplo de la función de demanda

<sup>8</sup> Se trata de los acrónimos en idioma inglés para las sumas acumuladas de residuos recursivos y de éstos al cuadrado.

<sup>9</sup> O sea, divididos entre el desvío estándar del error del modelo estimado originalmente, sobre el que estamos investigando la estabilidad.

de importaciones en Cuba. Primero vamos a presentar los resultados para la función estimada a partir de las tasas de variación. Los hemos agrupado en la figura 13.2.

En el panel (a) hemos representado los coeficientes recursivos. El coeficiente  $C(1)$  se corresponde con nuestro  $\beta_1$  en el modelo, el  $C(2)$  con  $\beta_2$  y el tercer coeficiente con la constante  $\beta_3$ . Podrán apreciar como muestran períodos de inestabilidad alrededor de los años 1972, 1975 y 1992. Esto se confirma en el panel (b) de los residuos recursivos, que justo esos años tienen movimientos fuera de los intervalos de confianza. Fíjense también como, aún cuando no desbordan las líneas críticas, los estadísticos CUSUM (panel c) y CUSUMQ (panel d) muestran desviaciones en esos períodos.

La inestabilidad de los parámetros de esta regresión entre finales de los 80 y el primer lustro de la década del 90 no nos sorprende: se trata del período en que la economía cubana sufre su más dramático *shock* externo, relacionado con la caída del socialismo en Europa del Este. El choque negativo justo a posteriori de 1970 podría asociarse a la realización de la más grande zafra azucarera de la historia de Cuba justo ese año, que tuvo consecuencias depresivas para el resto de la economía y para el mismo sector azucarero en los años subsiguientes. Finalmente tenemos el *shock* positivo de 1975. Este puede tener relación con el *boom* de precios internacionales del azúcar, que mejoraron sustancialmente la capacidad de compras externas de Cuba en ese período.

En la figura 13.3 estamos mostrando los mismos indicadores para la regresión estimada en base a los logaritmos de las variables.

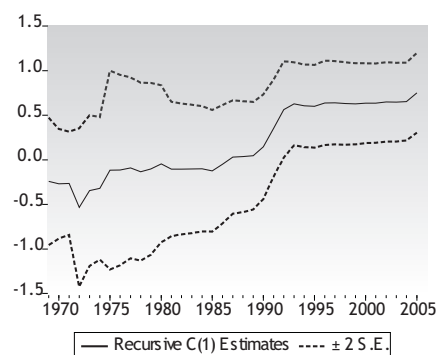
Noten cómo muchas de las observaciones que hiciéramos para el caso anterior coinciden. Acá los coeficientes  $C(k)$  se corresponden con nuestras  $\alpha_k$ . Fíjense como los residuos recursivos sobrepasan el límite inferior del intervalo de confianza entre los años 1992-1994, en que la economía cubana enfrenta un importante choque externo. Estas circunstancias se reflejan más evidentemente en los estadísticos CUSUM y CUSUMQ que en el modelo en tasas de variación. En el panel de CUSUMQ tenemos una salida del estadístico por debajo de los límites del intervalo de confianza.

Así, gracias a los coeficientes y residuos recursivos, y a los estadísticos CUSUM y CUSUMQ, podemos confirmar y también detectar cambios en los parámetros de la regresión que no resultaban tan conocidos o que no recordábamos tanto.

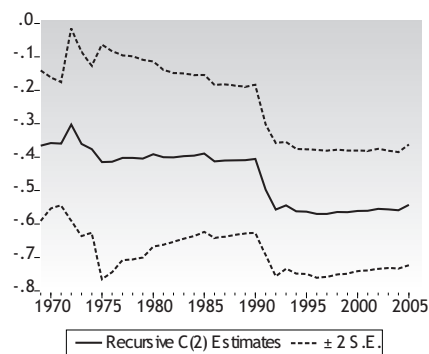
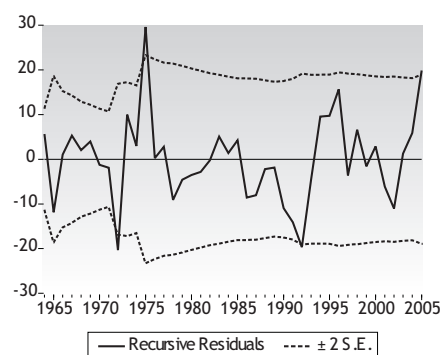
**Figura 13.2: Resultados de las pruebas de estabilidad de los parámetros de la función de demanda de importaciones en Cuba, estimada a partir del modelo**

$$\dot{m}_t = \beta_1 \dot{y}_t + \beta_2 \dot{p}_t + \beta_3$$

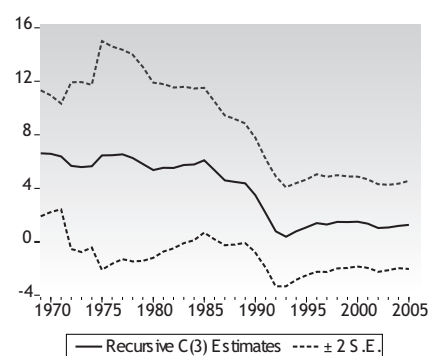
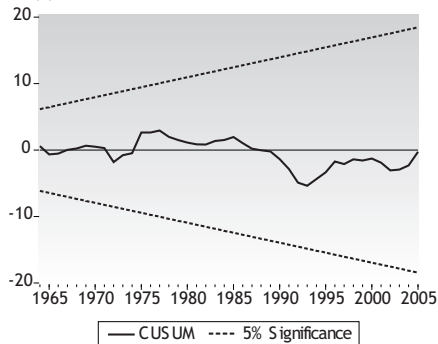
(a) Coeficientes recursivos



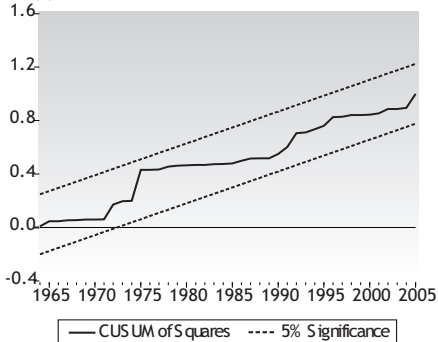
(b) Residuos recursivos



(c) CUSUM



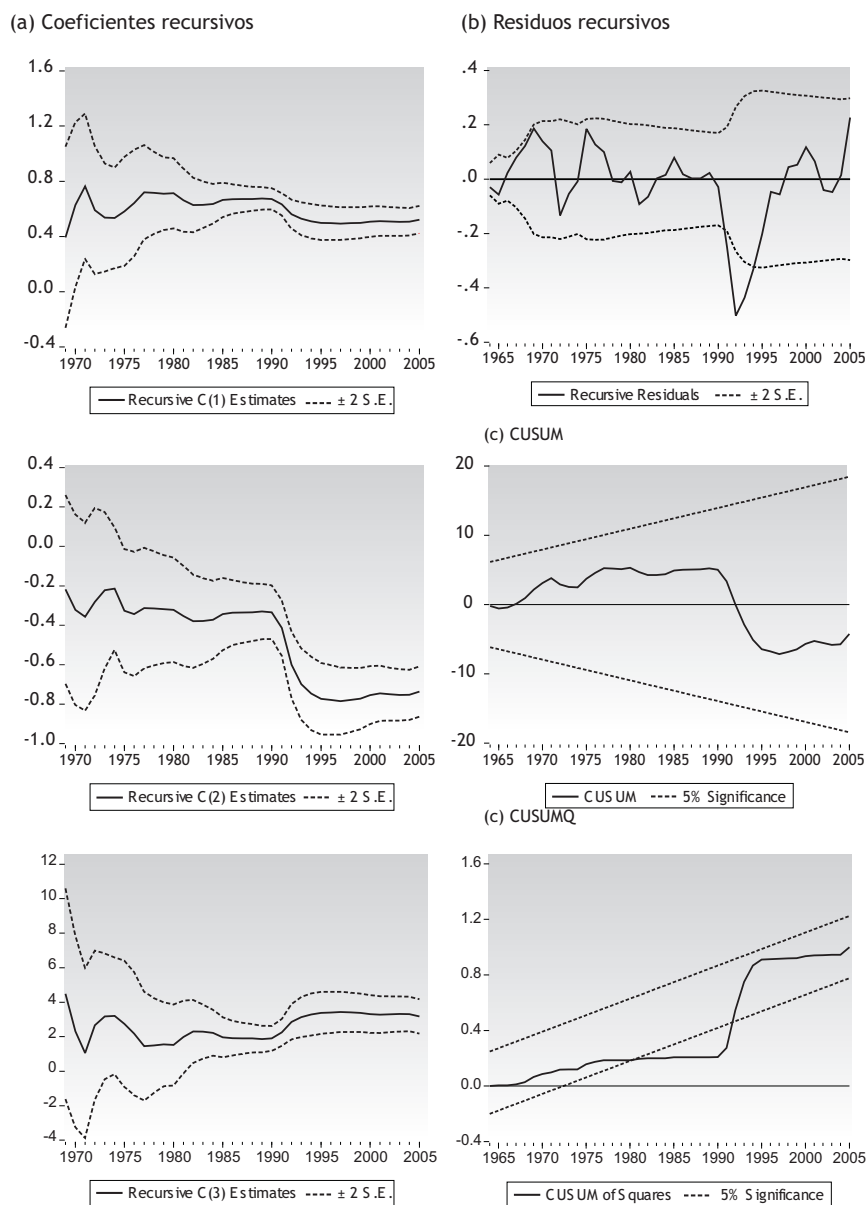
(c) CUSUMQ



Fuente: Elaboración propia



**Figura 13.3: Resultados de las pruebas de estabilidad de los parámetros de la función de demanda de importaciones en Cuba, estimada a partir del modelo  $LM_t = \alpha_1 LY_t + \alpha_2 LP_t + \alpha_3$**



Fuente: Elaboración propia

Estos análisis podemos corroborarlos, haciendo el test de Chow para los años en que tenemos sospecha que ocurren los cambios estructurales. Por ejemplo, si sometemos nuestra función de demanda de importaciones a esta prueba, en su variante loglineal, confirmamos el cambio estructural en 1975 y 1991. Mostramos los resultados de estos tests en la tabla 13.6.

**Tabla 13.6: Función de demanda de importaciones cubanas  
loglineal, tests de Chow**

|                        | Valor    | Probabilidad |
|------------------------|----------|--------------|
| <b>Año 1975</b>        |          |              |
| Estadístico-F          | 18.58754 | 0.000000     |
| Razón de verosimilitud | 39.95161 | 0.000000     |
| <b>Año 1991</b>        |          |              |
| Estadístico-F          | 24.39265 | 0.000000     |
| Razón de verosimilitud | 47.54362 | 0.000000     |

*Fuente: Elaboración propia*

En este epígrafe hemos mostrado las múltiples herramientas a nuestra disposición para investigar sobre la estabilidad de los parámetros de la regresión. Al test de Chow que ya conocíamos, hemos añadido los coeficientes y residuos recursivos, y las pruebas en base a ellos como CUSUM y CUSUMQ, que nos permiten también indagar por estos cambios.

### 13.7 Algunos Consejos sobre Cómo Conducir Trabajos Empíricos y Escribir sus Informes

¿Por qué hacemos estudios empíricos o econométricos? Uno de los propósitos es de apoyo a la investigación teórica. Los economistas teóricos proponen sistemáticamente hipótesis que deben ser testeadas y contrastadas con la realidad. Recuerden que la realidad siempre es cierta, sin embargo, las teorías no necesariamente. Una tarea fundamental de la ciencia es la de buscar evidencia empírica que contradiga las hipótesis teóricas, las que son, entonces, reveladas y reemplazadas por nuevas hipótesis que explican más acertadamente la realidad. Esta es una demanda primordial al economista, que no siempre está en capacidad de satisfacer.

Así, cuando nos enfrentamos a la escritura del informe resumen de nuestro trabajo empírico, lo primero que debemos dejar claramente establecidas son las hipótesis que vamos a someter a escrutinio. Nuestras hipótesis no son más que respuestas tentativas a la pregunta o problema de investigación que nos ocupa, que en este caso debemos expresar en términos de parámetros de un modelo que deben ser explorados.

Debemos dar un respaldo teórico al estudio que vamos a realizar, para evitar la tentación de la medida sin teoría. Será preciso que busquemos en la producción bibliográfica, las diferentes teorías desarrolladas para explicar el fenómeno que estamos estudiando. Debemos analizarlas críticamente y, si se precisa, adaptarlas al contexto para el que van a ser verificadas empíricamente, tomando en cuenta

los elementos idiosincrásicos que pueden influir en adición a los hechos estilizados que explica la teoría. No debe faltar una mención a aquellos investigadores que, antes que nosotros, han trillado ya este terreno.

Tampoco debemos dejar de presentar nuestra muestra. Para comenzar, declarar las fuentes de nuestros datos, explicar quién los proveyó; esto es imprescindible para garantizar el principio de replicabilidad de nuestros estudios por otros colegas. Una buena costumbre es el compartir las bases de datos con otros investigadores, después de que publiquemos nuestro trabajo, permitiendo la reproducción y control de nuestros resultados.

Asimismo, debemos describir nuestros datos, para lo que nos apoyaremos en sus estadísticas básicas, entre ellas media, mediana, valor máximo, mínimo, desvío estándar, coeficiente de variación, y la cantidad de observaciones. Todas ellas pueden ser obtenidas con los programas econométricos que circulan, y también pueden ser calculadas en una simple hoja de Excel.

Los gráficos de nuestras series pueden igualmente ayudar mucho a comprender algunas de sus características: posibles tendencias comunes, probables observaciones atípicas, qué podemos esperar sobre si son estacionarios o no, su estacionariedad si se trata de datos interanuales. Del mismo modo nos interesa la matriz de correlación entre las series; será saludable que tener altos coeficientes de correlación entre las variables explicativas y la dependiente, y que la correlación sea escasa para las explicativas entre sí. Y para terminar el escrutinio de nuestros datos, tenemos que someter las series de tiempo a las pruebas de estacionariedad, que aun no hemos presentado, pero que a la altura del capítulo 16 comprenderán por qué son importantes y constituyen un estudio obligado.

Después de este preámbulo, ya estamos en condiciones de presentar el (o los) modelo(s) analítico(s) en que basaremos nuestros estudios. Estos modelos se asentarán en la teoría económica que hemos explorado, que brinda el marco conceptual necesario. Pero recuerden que las teorías analizan la realidad despojándola de los detalles, para dar una explicación estilizada a las relaciones entre variables económicas que se observan en la práctica. A esa relación argumentada por la teoría debemos buscarle una forma funcional y si la teoría no aporta elementos sobre esto, deben ser nuestros datos quienes lo hagan. Asimismo, debemos ser conscientes de las diferencias entre las variables teóricas y nuestras variables observadas: si no disponemos de datos para las variables teóricas, ni de representantes adecuados para ellos, debemos buscar otra teoría para explicar el fenómeno bajo estudio para la que dispongamos de series. También las especificidades de la dimensión temporal-espacial concreta que analizamos brindan elementos para finalmente plantear un modelo que pueda ser estimado.

A partir del modelo estimable, podremos estimar las regresiones, hacer las pruebas de diagnóstico, comentar sus resultados y extraer las conclusiones necesarias para proceder al perfeccionamiento de nuestro modelo. Las múltiples pruebas de diagnóstico nos facilitan indagar sobre si la evidencia empírica permite no rechazar que los supuestos sobre los que se construye un MRLN sean válidos en

nuestro caso de estudio. No tiene ningún sentido someter a evaluación un modelo sobre el cual no estamos seguros que sea estadísticamente correcto, pues de no serlo, la inferencia estadística que podamos hacer a partir del mismo estará totalmente errada: las pruebas de hipótesis sobre los estimadores tendrán baja potencia, no podremos rechazar lo que es falso, en fin, podemos equivocarnos mucho.

Después de realizadas pruebas de diagnóstico, es que debemos decidir cuáles problemas corregir o remediar. Para eso hace falta mucha experiencia, pues a veces una única falla —como por ejemplo, una variable relevante omitida en la especificación de la media condicionada o parte sistemática del modelo—, puede ser la responsable del resto de los errores detectados. Igual que la falta de datos observables nos lleva a buscar otro modelo teórico, estas pruebas nos pueden hacer retroceder a varios momentos previos a la estimación: reespecificar el modelo porque están faltando variables relevantes o porque la forma funcional no es correcta, reestimarlos con otros métodos porque el que hemos aplicado no es el adecuado, modificar el planteamiento del MRLN a un modelo dinámico, etc., etc.

Y en todo este proceso, como siempre hemos aconsejado, no deben perder de vista que en Econometría no hay almuerzo gratis: debemos proceder con prudencia y evitando perfeccionismos inútiles que más pueden perjudicar que beneficiar. El sentido del *trade-off* es muy importante. Debemos mantener nuestro modelo tan simple como sea posible. ¡Recuerden siempre las características de un buen modelo que tratamos al inicio de este capítulo!

Los resultados de las pruebas de diagnóstico pueden ser muy tediosos para los no especialistas, por eso les aconsejamos que sólo incluyan un resumen de resultados en el informe del trabajo empírico y las estrategias aplicadas para el perfeccionamiento del modelo en cuestión. Todos los detalles de estas pruebas debemos ubicarlos en anexos al informe, para que los especialistas puedan saciar su curiosidad acerca de la validez de los resultados a que arribamos.

Finalmente, todo trabajo debe ser coronado con conclusiones y recomendaciones, que son el pináculo de nuestros esfuerzos, concentrado en mensajes útiles para los investigadores, pero también para aquellos que deben ser asesorados en la toma de decisiones a partir de la aplicación de estos instrumentos.

Para escribir este tipo de trabajos necesitamos apoyarnos en un conjunto de herramientas informáticas a nuestra disposición. Una de ellas son los programas de edición de textos. Muchos de ellos tienen incorporados módulos para la escritura de ecuaciones, que resultan imprescindibles en estos empeños. Si se dispone de un programa especializado para la escritura de ecuaciones, mucho mejor, pues en Econometría es muy frecuente el uso de letras griegas para designar los parámetros, de las sumatorias, de los vectores y matrices, de los subíndices y superíndices, y todo esto es realmente un trabajo pesado si no se dispone de los programas adecuados. Las ecuaciones y símbolos se pueden escribir en línea con el texto, pero también de forma desplegada. En el presente texto podrán encontrar los dos tipos de ecuaciones. También es importante numerar las ecuaciones, pero sólo aquellas a las que vayamos a hacer referencia.

Otros instrumentos muy útiles son los gráficos, como también habrán podido corroborar a lo largo de este texto. Los programas econométricos traen procesadores gráficos, que nos permiten obtener una variedad de ellos, vinculados al análisis básico de las series, a la estimación de los modelos y a algunos tests, como pueden ser los de estabilidad de los parámetros de la regresión que recién vimos en el epígrafe anterior. Sin embargo, estas rutinas gráficas son bastante rígidas y sus resultados no siempre son atractivos. En caso de que nos guste experimentar con efectos gráficos, Excel brinda buenas posibilidades. Para el diseño gráfico hay que ser especialmente creativos, pero en todo caso se pueden tomar ideas a partir del examen de gráficos en otros estudios. Tanto los gráficos que produce Excel como los que obtenemos con programas econométricos, pueden ser copiados y pegados a ficheros de texto.

Asimismo, para la presentación de resultados es muy útil la composición de tablas. En este sentido debemos ser bien selectivos acerca de los datos a informar. Las tablas de salida de los programas econométricos suelen estar bien cargadas de información, por lo que es preciso escoger lo realmente relevante. Resulta de provecho también fijarnos en los diseños de tablas en trabajos empíricos de otros autores, para extraer las lecciones pertinentes.

Lo mismo los gráficos que las tablas, debemos numerarlos, para poder hacer referencia cómodamente a ellos. Tampoco debemos olvidar que todo gráfico y tabla debe tener su fuente, de modo que quede una senda que puedan recorrer igualmente aquellos que quieran verificar nuestros hallazgos como los que pretendan seguir trabajando en nuestro campo de estudio. En el trabajo científico es muy importante mantenerse abiertos a los consejos y la crítica, tanto de colaboradores como de oponentes.

Tal como ya hemos comentado con antes, la Econometría nos ayuda a analizar las relaciones entre variables, a describir mejor los procesos, a corroborar teorías y también a predecir. Esta última es una aplicación cardinal de estas herramientas, pues los economistas estamos constantemente emplazados sobre las repercusiones futuras que podrán tener los procesos que ocurren en la actualidad. Por eso es muy importante tener la posibilidad de valorar la capacidad de predicción de un modelo econométrico fuera de la muestra que sirvió para su estimación.

En este punto de nuestro texto queremos enfatizar que, antes de lanzarnos a la tarea de seleccionar modelos para la elaboración de pronósticos, debemos estar seguros de que los modelos a comparar han pasado satisfactoriamente los tests e indicadores evaluativos que hemos desarrollado en los capítulos del siete al trece. Es decir, que no hemos logrado conseguir evidencia contraria a que estos modelos cumplan los supuestos subyacentes a un MRLN.

Igualmente los diferentes tests desarrollados en el capítulo seis con el objetivo también de evaluar la regresión estimada sólo serán válidos si realmente estamos lidiando con una realidad que se deja representar a través de un MRLN.

A los fines de evaluar la capacidad predictiva de un modelo fuera de su muestra, se han desarrollado diversos indicadores. Dos de ellos ya los hemos abordado

en el capítulo seis de este libro, específicamente en su epígrafe dedicado a evaluar la bondad de ajuste de un modelo. Se trata de los criterios de Akaike (fórmula 6.18) y de Schwartz (expresión 6.20). Recuerden que ambos criterios pretenden minimizar la suma de residuos al cuadrado, pero que también imponen un castigo o penalización por incluir un número creciente de variables explicativas en la regresión. Cuando comparemos modelos aplicando estos criterios, el mejor modelo será el que tenga el menor CIA o CIS.



# *preguntas*

Intente dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo.

1. ¿Qué entendemos por error de especificación?
2. ¿Cuáles características debe poseer un buen modelo de regresión?
3. ¿Qué consecuencias tiene la omisión de una variable relevante en la modelación econométrica de un fenómeno de interés?
4. ¿Cuáles son los efectos de la inclusión de variables no relevantes en un modelo de regresión?
5. ¿Qué consecuencias tiene adoptar una forma funcional incorrecta en un modelo de regresión?
6. ¿Cuáles procedimientos o tests tenemos a nuestra disposición para evaluar si tenemos problemas con la especificación de un modelo de regresión?
7. ¿Qué son los residuos recursivos y para qué se calculan?
8. ¿Qué son los tests CUSUM y CUSUMQ? ¿Cuál es el propósito de estas pruebas?
9. ¿Qué son los criterios de información de Akaike y de Schwartz?



# *ejercicios*



- 13.1 A partir de los datos de la tabla B.1 del CD acompañante y de la información que ha ido acumulando a lo largo de este texto, estime una función de demanda de importaciones que sirva de apoyo a las tareas de planificación del sector externo de la economía cubana. Le sugerimos realizar este ejercicio en dos momentos.

Justo al terminar el estudio hasta la Parte Tres del libro.

Después de haber rebasado la Parte Cuatro.

- a. Finalmente, elabore un pronóstico para las importaciones correspondientes a  $\mathbf{X}_{2006}^A = [1 \ 10 \ 0]$  y  $\mathbf{X}_{2006}^B = [1 \ 10 \ 7,3]$ .
  - b. Escriba un pequeño reporte del trabajo realizado anteriormente.
- 13.2 Utilizando los datos de la tabla B.4 del CD acompañante y también tomando en consideración el trabajo pretérito sobre este conjunto de datos, estime una función de crecimiento de la economía cubana, que sirva a los planificadores macroeconómicos cubanos para evaluar la robustez de las proyecciones obtenidas por otros métodos cuantitativos. Le sugerimos igualmente realizar este ejercicio en dos momentos.

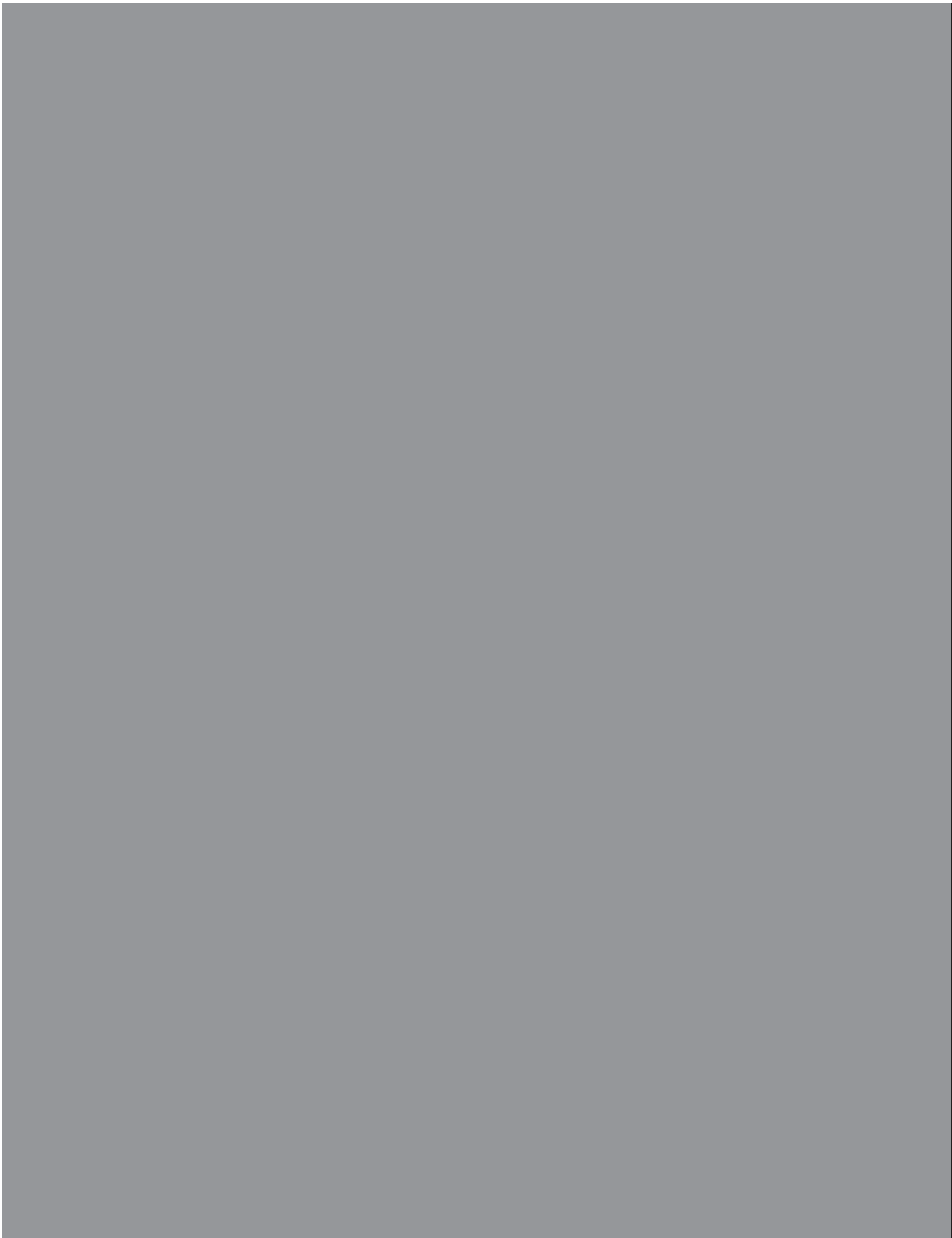
Una vez concluida la lectura de la Parte Tres de este texto.

Después de haber estudiado la Parte Cinco.

- a. Contraste los resultados de crecimiento del plan económico correspondiente a 2008 con los resultados que obtendría a partir del modelo finalmente estimado.
  - b. Escriba un sucinto informe del trabajo realizado.
- 13.3 Con la información de la encuesta de hogares sueca, recogida en la tabla B.6 del CD acompañante y teniendo en cuenta los análisis realizados en el ejercicio 7.6, así como los conocimientos adquiridos hasta este punto, estime una función de formación de salarios para ese contexto correctamente especificada.
- c. Elabore un reporte sobre los resultados del trabajo econométrico realizado.
  - d. Evalúe si existe discriminación de la mujer en cuanto al ingreso salarial.
  - e. Examine asimismo si el estado civil tiene una influencia en el salario.







• PARTE CUATRO

• MODELOS NO-LINEALES  
• Y DATOS DE PANEL



En esta parte hemos reunido un grupo de modelos que guardan poca relación entre sí, salvo el que exceden claramente las capacidades del método de mínimos cuadrados. En primer lugar, tenemos los modelos en que la variable dependiente es discreta o puede ser observada solamente en un rango limitado de valores. En segundo lugar, consideramos modelos que son estimados a partir de paneles de datos y que tratan de capturar la información adicional que fluye del hecho de que la muestra es un panel. Finalmente, tenemos modelos de frontera o con distribuciones asimétricas.

CAPÍTULO 14



## VARIABLE DEPENDIENTE CUALITATIVA O LIMITADA

En este capítulo discutimos una familia de modelos que se caracterizan porque su variable dependiente es discreta o definida con un soporte o dominio limitado. Estos modelos han probado tener una gran utilidad, especialmente en el ámbito de la microeconomía y de la toma de decisiones, pero nos llevan al terreno de las estimaciones no-lineales.

### 14.1 Variable Dependiente Binaria

Con frecuencia los agentes económicos, individuos o firmas, enfrentan decisiones mutuamente excluyentes. Es necesario elegir entre dos alternativas que se excluyen mutuamente. Por ejemplo, una persona puede participar en la Fuerza de Trabajo o estar excluida de ella. Un deudor puede pagar un crédito o puede no hacerlo. Necesitamos, entonces, construir un modelo estadístico que nos permita examinar la determinación de **variables binarias**, o de elección del tipo sí o no. Estas variables binarias son a menudo representadas por variables que pueden asumir sólo dos valores; el valor 1 cuando una determinada opción es elegida y el valor 0 en el caso contrario.

Hemos visto ya modelos en que se utilizan variables binarias; éste es el caso de los modelos que utilizan variables mudas como variables explicativas. Pero en este caso tenemos una diferencia decisiva: la variable binaria es la variable dependiente o variable explicada. Veremos, más adelante, que el método de mínimos cuadrados no es el método de estimación adecuado y que preferiremos las estimaciones máximo verosímiles.

### 14.1.1 El modelo lineal de probabilidad

Tratemos de formular un modelo lineal que explica una variable binaria. Consideremos, por ejemplo, el caso clásico del problema de transporte público. En este ejemplo se supone que hemos hecho una encuesta entre un cierto número de residentes de un determinado barrio. La pregunta central es si utilizan el transporte público o no para dirigirse a su trabajo. También recogimos información sobre las características de los individuos encuestados, y sobre el tiempo que les toma llegar a su trabajo. Este es un ejemplo importante, pues entrega una información relevante para una mejor planificación de las redes de transporte público.

Tratemos de formular un modelo lineal que explique la elección de medio de transporte. Para ello suponemos que tenemos una variable binaria,  $y$ , que representa esta elección, la que admite sólo dos valores posibles: 1 cuando el individuo usa el transporte público para ir al trabajo, y 0 cuando no lo hace. Ésta es una variable estocástica discreta con un rango muy reducido, sólo dos valores. La función de probabilidad para esta variable es

$$f(y) = p^y (1-p)^{1-y}, \quad y \in \{0, 1\}$$

donde  $p$  es la probabilidad de que la variable  $y$  asuma el valor 1. Esto es  $f(y) = P[y=1]$ . Decimos que la variable  $y$  sigue una distribución binomial. La media y la varianza de esta distribución están dadas por

$$E(y) = p, \quad \text{var}(y) = p(1-p)$$

Supongamos que hay una relación lineal entre esta probabilidad y un cierto número de variables explicativas. Es decir, supongamos que

$$E(y) = p = \beta_1 + \beta_2 x$$

Esta relación nos permite especificar un **modelo lineal de probabilidad** para explicar la variable de elección  $y$  como,

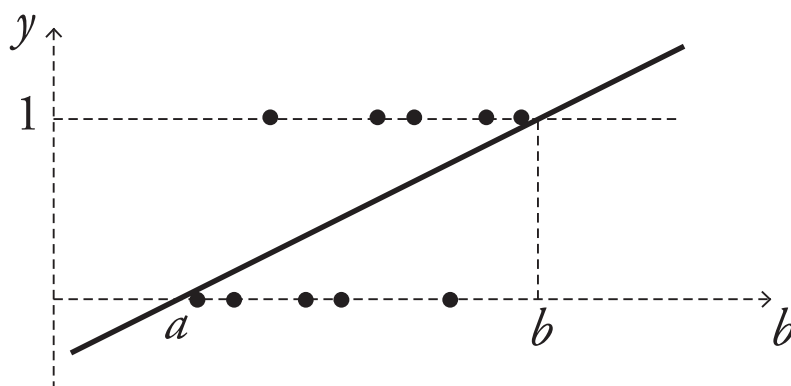
$$y_i = E(y_i) + \varepsilon_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Este modelo, y su estimación mínimo cuadrática, debe ser descartado en la práctica por varias razones. En primer lugar, este modelo es esencialmente heterocedástico. Es fácil ver que para los diferentes individuos en la población, y en la muestra considerada, la variable  $y$  no tendrá la misma probabilidad de ser igual a uno. De hecho es esta variabilidad lo que da sentido a un modelo estadístico de esta elección. Pero si cada individuo tiene un valor diferente de  $p$ , también tendrá una varianza diferentes. En consecuencia el modelo es esencialmente heterocedástico. No es necesario probar esta propiedad. Ya está implícita en la especificación del modelo.

En segundo lugar, siempre hay un riesgo de que este modelo entregue predicciones fuera de rango. Es decir, el modelo puede predecir probabilidades menores que cero o probabilidades mayores que uno.



Figura 14.1: Modelo lineal de probabilidad



Este problema se comprende fácilmente observando la Figura 14.1. En este caso la variable  $x$  toma valores en el rango  $[a, b]$ , mientras que para la variable  $y$  observamos solamente los valores cero y uno. El gráfico presenta las observaciones de la muestra como puntos oscuros, y se ha incluido la línea de regresión. Resulta obvio que, para valores de  $x$  menores de  $a$  o mayores que  $b$ , las predicciones serán fuera de rango. En efecto, este modelo está sugiriendo probabilidades mayores que uno para valores de  $x$  mayores que  $b$  y menores que cero, para valores de  $x$  menores que  $a$ .

Finalmente, tenemos que recalcar el hecho de que el modelo lineal de probabilidad ignora el hecho de que la variable dependiente tiene un rango limitado, puede variar solamente entre cero y uno. En este caso la estimación mínimo-cuadráticas nos darán estimadores sesgados. Estos problemas son muy serios y nos obligan a buscar un enfoque alternativo para explicar el comportamiento de variables binarias.

#### 14.1.2 El modelo Probit

Consideremos la función de probabilidad normal estándar, cuya fórmula es la siguiente:

$$F(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-0.5u^2} du$$

Esta función nos entrega la probabilidad de que el valor de una variable estocástica  $Z$  sea menor o igual que un valor dado igual a  $z$ . Esta función es, con frecuencia, llamada la función **Probit**, con media igual a cero y varianza igual a uno. Utilizaremos esta función para especificar un modelo alternativo que llamaremos modelo Probit.

Supongamos que tenemos una variable,  $y^*$ , que representa la potencialidad o propensión de un determinado individuo a elegir el transporte público. Esta es una variable estocástica continua que puede variar sobre todo el rango de números reales. Valores mayores de  $y^*$  implican la elección de transporte público. Valores pequeños de  $y^*$  implican que el individuo en cuestión no usará el transporte público. Más aún, suponemos que esta variable está linealmente asociada a una o varias variables explicativas. Dada una muestra de  $n$  individuos podemos, entonces, especificar el siguiente modelo estadístico

$$y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i \in \{1, 2, \dots, n\} \quad (14.1)$$

donde  $\varepsilon_i \sim N(0, \sigma^2)$ . Si bien el modelo resulta atractivo, no puede ser estimado porque la variable  $y^*$  no puede ser observada. Variables como ésta se denominan **variables latentes**. Otro ejemplo de una variable como ésta es considerar que  $y^*$  representa un índice de utilidad indirecta cuando consideramos la compra de un bien durable.

Hay una salida relativamente simple para estimar los parámetros de este modelo. Si bien no podemos observar  $y^*$ , sí podemos observar la variable  $y$ , que discutimos anteriormente, y que es igual a uno cuando un individuo elige el transporte público e igual a cero cuando no lo hace. Entonces podemos suponer que ambas variables están unidas por la siguiente relación:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases}$$

La ecuación anterior puede reescribirse como

$$\varepsilon_i = y_i^* - \beta_1 - \beta_2 x_i$$

Si suponemos que  $\sigma^2 = 1$  la variable  $\varepsilon_i$  seguirá una distribución normal estándar para todos los valores de  $i$ . Entonces, aprovechando la simetría de esta distribución podemos escribir

$$P[y_i = 1] = P[y_i^* > 0] = P[y_i^* \leq 0] = F(\beta_1 + \beta_2 x_i) \quad (14.2)$$

donde  $F$  es la función de probabilidad normal estándar presentado anteriormente. Este modelo, llamado **modelo Probit**, tiene la capacidad de producir predicciones dentro del rango que es aceptable para una probabilidad:  $[0, 1]$ . Sin embargo, este modelo es altamente no lineal y, en consecuencia, no puede ser estimado por mínimos cuadrados.

La alternativa de estimación que nos queda abierta es la estimación **máximo-verosímil**, la que ya fue discutida, en general, en el capítulo 3. Recordemos que observamos una variable binaria,  $y_i$ , relacionada a la elección examinada por el

modelo. Podemos suponer que si observamos  $y_i = 1$ , esto es evidencia de que  $y_i^* > 0$  para este individuo indexado por  $i$ . Por el contrario, si observamos que  $y_i = 0$ , consideraremos esto como evidencia de que  $y_i^* \leq 0$ . Entonces la función de verosimilitud puede ser escrita como

$$L = \prod_{i=1}^n [F(\beta_1 + \beta_2 x)]^{y_i} [1 - F(\beta_1 + \beta_2 x)]^{1-y_i}$$

Tomando logaritmos en ambos lados, obtenemos la función de log-verosimilitud:

$$\log L = \sum_{i=1}^n y_i F(\beta_1 + \beta_2 x) + \sum_{i=1}^n (1 - y_i) [1 - F(\beta_1 + \beta_2 x)]$$

Desgraciadamente no conocemos una solución analítica exacta para el problema de maximización de esta función. Sin embargo es fácil encontrar algoritmos iterativos rápidos y robustos para encontrar una solución numérica. Estos algoritmos han sido incorporados en casi todos los paquetes computacionales orientados a estadísticas y econometría.

### 14.1.3 Interpretando un modelo Probit

La interpretación de los resultados de una estimación probit no son tan directos como en el caso de una modelo lineal y conviene discutirlos con ayuda de un ejemplo.

La tabla 14.1 muestra la salida típica de un programa que estima un modelo Probit. En las primeras líneas de esta tabla, el programa simplemente nos informa del número de iteraciones que fueron necesarias para alcanzar el máximo de la función de log-verosimilitud, así como los valores que alcanza esta función en cada iteración. Después tenemos una línea titulada LR chi2(1). Esta línea nos informa del estadígrafo de prueba de razón de verosimilitud, indicador que sigue una función de distribución chi cuadrado con un grado de libertad. Esta prueba compara el modelo con otro que sólo contiene la constante. En la línea siguiente vemos que el modelo tiene un alto grado de significación global. Esta prueba estadística puede ser interpretada en forma similar a la prueba  $F$  en un modelo lineal estimado con mínimos cuadrados.

En seguida este cuadro nos informa del valor alcanzado finalmente por la función de log-verosimilitud. También se presenta un indicador de bondad de ajuste llamado pseudo- $R^2$ , que tiene un sentido similar al  $R^2$  de las estimaciones mínimo-cuadráticas. El resto de la tabla es similar al de las estimaciones mínimo cuadráticas. Dando valores a las variables explicativas y utilizando los estimados de los parámetros podemos conseguir una estimación de la probabilidad de que un individuo, con características dadas por las variables independientes, elija el transporte público. Podemos elegir un umbral entre cero y uno, por ejemplo 0,5, tal que predecimos que los individuos con una probabilidad estimada mayor que este

umbral toman el transporte público, y no lo hacen aquellos con una probabilidad estimada menor que este umbral.

**Tabla 14.1: Estimación Probit**

|                                          |              |                       |          |                 |
|------------------------------------------|--------------|-----------------------|----------|-----------------|
| Iteration 0: log likelihood = -19.095425 |              |                       |          |                 |
| Iteration 1: log likelihood = -6.5660421 |              |                       |          |                 |
| Iteration 2: log likelihood = -4.5231848 |              |                       |          |                 |
| Iteration 3: log likelihood = -4.3375906 |              |                       |          |                 |
| Iteration 4: log likelihood = -4.3371923 |              |                       |          |                 |
| Iteration 5: log likelihood = -4.3371923 |              |                       |          |                 |
|                                          |              |                       |          |                 |
| Number of obs.                           | 30           |                       |          |                 |
| LR chi2(1)                               | 29.52        |                       |          |                 |
| Prob > chi2                              | 0.0000       |                       |          |                 |
| Log likelihood                           | -4.337192    | Pseudo R <sup>2</sup> | 0.7729   |                 |
|                                          |              |                       |          |                 |
| <b>yobs</b>                              | <b>Coef.</b> | <b>Std. Err.</b>      | <b>z</b> | <b>P&gt; z </b> |
| tiempo                                   | 0.389501     | 0.16866               | 2.31     | 0.021           |
| constante                                | -9.614266    | 4.26700               | -2.25    | 0.024           |

Sin embargo, los parámetros de una estimación Probit no pueden ser interpretados de la misma forma que los parámetros de un modelo lineal. Supongamos, por ejemplo, que hemos estimado un modelo lineal como el siguiente:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$$

Entonces podremos atribuir una interpretación precisa a los parámetros estimados. Por ejemplo, podemos decir que si la variable  $x$  aumenta en una unidad, entonces la variable dependiente  $\hat{y}$  aumentará en  $\hat{\beta}_2$  unidades, no importa el valor que tenga la variable  $x$ . Es decir, la relación entre la variación de ambas variables es constante. Esto se debe a que la derivada de la variable dependiente con respecto a  $x$  es constante y precisamente igual a  $\hat{\beta}_2$ . El origen de este resultado es que la relación entre ambas variables es lineal. Pero en el caso del modelo Probit la relación no es lineal y tenemos que

$$\frac{\partial y}{\partial x} = \hat{\beta}_2 f(\hat{\beta}_1 + \hat{\beta}_2 x)$$

donde  $f$  es la función de densidad normal estándar. Este término recibe el nombre de contribución marginal y puede ser interpretado de la misma forma que los coeficientes en el caso lineal. Naturalmente, esta contribución marginal no es constante, pues depende del valor de  $x$ . Muchos programas informan automáticamente

del valor de estas contribuciones marginales, evaluadas en el valor de las medias muestrales de las variables independientes.

En el ejemplo anterior, suponemos que tiempo se refiere al tiempo que ahorran las personas cuando toman el transporte público. La media de esta variable para nuestra muestra es de 30,5 minutos. Entonces, la componente lineal de nuestro modelo Probit en (14.2) es igual a 2,27. La función de densidad estándar en este punto es igual a 0,0306. La contribución marginal de la variable tiempo es, entonces, igual a 0,012 (1,2 por ciento), obtenido multiplicando 0,0306 por 0,389501. Es decir, por cada minuto de tiempo ahorrado la probabilidad de que un individuo tome el transporte público crece en 1,2 puntos porcentuales.

## 14.2 La función Logística



Una distribución probabilística especialmente interesante es la distribución logística. La función de densidad par una variable  $Y$  que sigue esta distribución es

$$f(y) = \frac{e^{-y}}{(1 + e^{-y})^2}, \quad -\infty < y < \infty$$

La función de probabilidad correspondiente es

$$F(y) = P[Y \leq y] = \frac{1}{1 + e^{-y}}$$

Si en un modelo Probit sustituimos la función normal estándar por esta función logística, tenemos el siguiente **modelo Logit**.

$$P[L \leq \beta_1 + \beta_2 x] = F(\beta_1 + \beta_2 x) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x)}}$$

La función logística de probabilidad tiene un gráfico muy cercano a la función normal. De esta manera hay sólo pequeñas diferencias si se usa un Probit o un Logit para modelar una situación de selección binaria. El modelo Logit fue muy popular en el pasado pues resulta computacionalmente más sencillo. Sin embargo, con los recursos computacionales disponibles hoy esta diferencia se ha hecho irrelevante y, casi siempre, los problemas de elección binaria se modelan utilizando un Probit.

Sin embargo, el modelo Logit retiene su importancia por otras razones. Por ejemplo, la curva logit, el gráfico de la función logística de probabilidad, ocurre con cierta frecuencia en la naturaleza y, entonces, es la especificación natural para modelar ciertos procesos. También el modelo logit tiene la ventaja de que el rango de la función logística está entre cero y uno. Como un ejemplo relevante consideremos el caso de una variable que representa un porcentaje y, por lo tanto, tiene un recorrido entre cero y uno. Este caso podría ser el del porcentaje de servicios en el total de importaciones de bienes y servicios. Estos servicios, tales

como fletes, seguros, etc., tienden a ser una proporción del total de importaciones. Sin embargo, hay otras variables que afectan esta proporción, tales como cambios en la composición de las importaciones de bienes, desarrollo tecnológico, etc. Nuestra variable dependiente, la proporción de servicios en el total de importaciones de bienes y servicios,  $y$ , tiene un rango entre cero y uno. Esto lo podemos modelar usando un modelo logit

$$y = F(\beta_1 + \beta_2 x) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x)}}$$

Este es un modelo no lineal en los parámetros y debe ser estimado usando un enfoque máximo-verosímil. También en este caso conviene recurrir a la idea de contribuciones marginales para tener una mejor interpretación de los parámetros.

### 14.3 Otros Modelos de Variables Discretas

#### 14.3.1 Modelos multinomiales

Una clara limitación del modelo Probit es que considera únicamente dos alternativas mutuamente excluyentes, cuando en la realidad hay muchas situaciones en que es necesario considerar varias alternativas. En nuestro ejemplo de transporte anterior podemos pensar en que cada individuo puede elegir entre caminar, tomar su bicicleta, tomar un automóvil y tomar un medio de transporte público. En otras palabras, situaciones con elecciones entre alternativas múltiples son comunes y relevantes en la realidad. Una situación semejante se llama una **elección multinomial**, en oposición al caso anterior que era binomial.

Modelos de elección multinomial, también llamados Probit Multinomial, pueden ser especificados pero con un creciente grado de complejidad computacional. Dentro de estos modelos son especialmente interesantes, por resultar más tratables, es la clase de modelos llamados de **selección ordenada**. En este caso las alternativas consideradas pueden ser ordenadas de acuerdo a algún criterio preestablecido. La existencia de este orden es un elemento de información adicional que debe ser considerado cuando se considera el modelo y su estimación. En otras palabras, existe un **rango** de alternativas, dentro del cual la alternativa 1 es la más baja, la alternativa 2 es la siguiente, y así sucesivamente hasta la alternativa  $J$  que es la más alta de acuerdo a algún criterio. Diremos que hay  $J$  alternativas ordenadas. Este rango u ordenamiento es ordinal y no necesariamente cardinal.

#### 14.3.2 Modelos de cuenta

Existen situaciones en que la variable a explicar se refiere a un número de ocurrencias de un determinado suceso. Por ejemplo, supongamos que queremos estudiar la formación de colas en las cajas de un supermercado. En este caso, la variable dependiente es el número de personas ante cada caja en un determinado punto en el tiempo. El número de camiones que tratan de acceder y descargar en

un ingenio depende de factores asociados al momento de la zafra, distancias, etc. Cuantas embarcaciones de la flota pesquera salen a faenar cada día es otra variable interesante. Los modelos que tratan este tipo de variables se llaman **Modelos de Cuenta o Regresiones de Poisson**.

Las regresiones de Poisson se basan en la llamada distribución de Poisson. Una variable aleatoria que asume los valores 0, 1, 2, hasta  $J$ , sigue una distribución de Poisson cuya función de probabilidad está dada por:

$$P[Y = y] = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots, J$$

Nótese que esta función depende de un parámetro  $\lambda$ . Este parámetro es igual a la media y a la varianza. Es decir tenemos que

$$E[Y] = \text{var}[Y] = \lambda$$

Entonces, si tenemos variables explicativas  $x$  podemos escribir

$$E[Y] = \lambda = e^{\beta_1 + \beta_2 x}$$

y realizar una estimación máximo-verosímil de la función de probabilidad. Esto nos da estimados de los parámetros  $\beta$  y del parámetro  $\lambda$ .

Este enfoque ha sido criticado por la relación fija que existe entre media y varianza para la distribución de Poisson. Por esto se ha hecho muy frecuente un enfoque basado en una distribución semejante que evita este problema: la distribución **binomial negativa**.

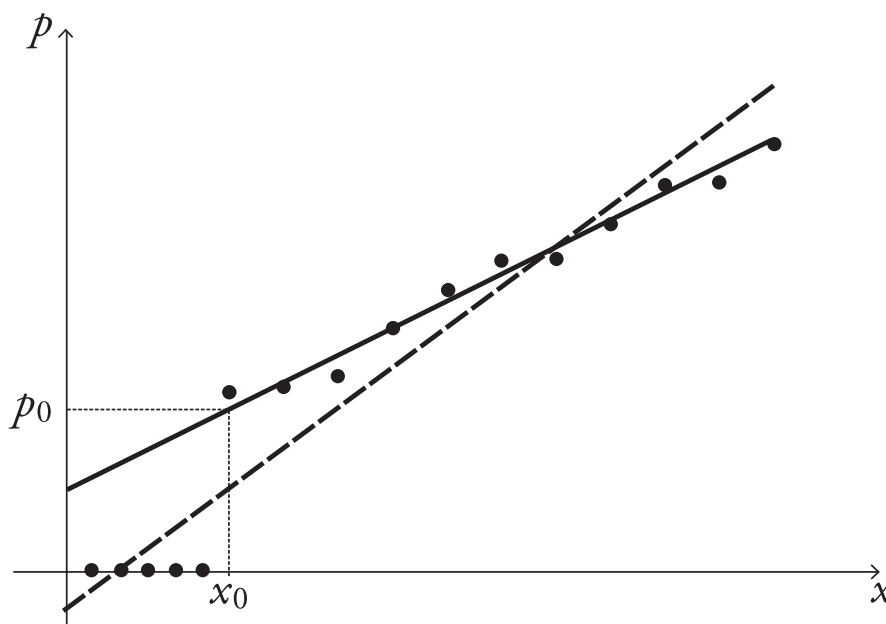
## 14.4 El Modelo Tobit

El modelo Tobit es un enfoque especialmente útil en presencia de **variables censuradas**, es decir variables que no pueden ser observadas para un cierto rango de su soporte. Posiblemente, el caso más claro de este tipo de variables lo encontramos cuando estudiamos el mercado de bienes durables, como heladeras automóviles, viviendas, etc. Estos bienes tienen diferentes precios que están relacionados a diferentes tamaños, capacidades, funcionalidad, etc. Más aún hay un precio mínimo por debajo del cual no observamos transacciones. Consideremos, por ejemplo, el precio que las familias están dispuestas a pagar por un automóvil. Este precio depende naturalmente de las características de las familias. El tamaño de la familia, su nivel de ingresos, distancia al trabajo, etc. Sin embargo, el automóvil más barato tiene un precio tal que habrán familias que no pueden comprar.

Supongamos que deseamos estudiar cuanto las familias están dispuestas a invertir en un automóvil,  $p$ , en función de características de las familias. Estas características están resumidas en una variable continua  $x$ . Tenemos una muestra de familias, representadas por puntos en el gráfico a continuación. Nótese que para las familias a la izquierda de  $x_0$  no hay demanda de automóviles pues no pueden

pagar el precio mínimo de  $p_0$ . Decimos que la variable ha sido censurada a la derecha en  $p_0$ . Debido a que para estas familias no observamos un consumo de automóviles le atribuimos un valor  $p=0$ .

Figura 14.2: Variable censurada



Si estimamos un modelo con todas las observaciones, obtenemos una línea de regresión como la representada por la línea discontinua de nuestro gráfico. Es fácil ver que la pendiente ha sido sobreestimada y subestima los valores de  $p$  a la izquierda de la muestra, y sobreestima los valores a la derecha de la muestra. Es decir, la inclusión de los ceros observados para la variable independiente introduce un sesgo en nuestra estimación mínimo-cuadrática. Si estimamos un modelo de regresión sólo con las observaciones a la derecha de  $x_0$ , excluyendo los ceros, obtenemos una línea de regresión representada por la línea continua. En esta estimación estamos ignorando la información contenida en las observaciones excluidas, lo que también introduce un sesgo. Esto es así aunque limitemos nuestras conclusiones solamente a las familias que efectivamente poseen un automóvil. Esto se debe a que las razones por las que escogemos las observaciones que van a ser excluidas son parecidas a aquellas que explican la variable  $p$ . Entonces, la muestra ya no es más aleatoria e introducimos el llamado **sesgo de selección**.

El problema puede ser resuelto por un modelo Tobit, el que puede ser racionalizado de la siguiente manera. Supongamos nuevamente una variable latente  $y^*$



relacionada linealmente con un vector  $x$  de variables explicativas. Esta variable  $y^*$  no puede ser observada, pero en cambio observamos otra variable  $y$  con las siguientes características:

$$y^* = \beta_1 + \beta_2 x$$

$$y = \begin{cases} y^* & \text{si } y^* > y_0 \\ 0 & \text{si } y^* \leq y_0 \end{cases}$$

Este modelo puede ser estimado usando métodos máximo-verosímiles. Es importante hacer notar que hay especificaciones ligeramente diferentes para este modelo que, en determinados casos, pueden ser relevantes. En general, este enfoque consiste en estimar un modelo lineal condicionado a un probit previo.

Una aplicación importante de este modelo se encuentra en estudios del mercado de trabajo. Un instrumento central en estos estudios son las ecuaciones mincerianas. En este caso explicamos el logaritmo del salario con ayuda de las características de las personas y, posiblemente, algunos controles adicionales. El problema es que, en general, cuando tomamos una muestra, un cierto número de personas no estará trabajando, por lo que observamos un salario igual a cero. En este caso el punto al cual la variable salario está censurada es el llamado salario de reserva. Este salario de reserva es aquel a partir del cual una persona acepta trabajar. Esta regresión puede ser exitosamente modelada usando un modelo Tobit.



# *preguntas*

Intente dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo.

1. ¿Qué tipo de variable dependiente incluye el modelo Probit?
2. ¿Qué distribuciones son frecuentes cuando consideramos modelos de elección binaria?
3. ¿Qué enfoque usamos para estimar modelos Probit?
4. ¿Qué tipo de variable dependiente consideramos en los modelos Tobit?



# ejercicios

- 14.1 Un estudio reciente entrevistó las familias de 770 niños en la educación primaria en escuelas pública de Montevideo. La tabla siguiente presenta los principales estadígrafos de un grupo de variables incluidas en la encuesta.

Estadígrafos Básicos

| Variables                 | Obs. | Media  | Desv. Estd. | Mín.  | Máx.   |
|---------------------------|------|--------|-------------|-------|--------|
| Computadora en casa       | 770  | 0.3571 | 0.4795      | 0     | 1      |
| Promedio educación padres | 770  | 8.0136 | 3.4933      | 0     | 18     |
| Vive con ambos padres     | 770  | 0.7013 | 0.4580      | 0     | 1      |
| Más de 50 libros en casa  | 770  | 0.2286 | 0.4202      | 0     | 1      |
| Personas por habitación   | 770  | 2.1211 | 1.4676      | 0.5   | 10     |
| Ingreso per cápita        | 770  | 1485.2 | 1299.7      | 105.5 | 8693.4 |

Los investigadores estaban interesados en los factores que determinan la disponibilidad de una computadora en casa, disponible para el trabajo escolar de los niños. Para ello estimaron un modelo Probit con los resultados informados en la tabla siguiente.

Probabilidad de tener una computadora en casa.

| Variables                 | Computadora en casa |       |
|---------------------------|---------------------|-------|
|                           | Coeficiente         | t     |
| Ingreso per cápita        | 0,0003              | 5,41  |
| Más de 50 libros en casa  | 0,538               | 4,08  |
| Vive con ambos padres     | 0,279               | 2,21  |
| Promedio educación padres | 0,091               | 4,43  |
| Personas por habitación   | -0,242              | -4,46 |
| Constante                 | -1,323              | -5,81 |
| N                         | 770                 |       |
| R <sup>2</sup>            | 0,3066              |       |
| Log verosimilitud         | -347,98             |       |
| Wald $\chi^2$ (5)         | 187,19              |       |
| Predicciones correctas    | 0,8946              |       |

Considerando estos resultados:

- Discuta la significación estadística de los parámetros.
- Calcule la contribución marginal de cada variable.
- Ordene las variables de acuerdo a la importancia creciente de su importancia.
- Estime el posible aumento en la probabilidad de disponer de una computadora en casa si el ingreso per cápita aumenta en 100 pesos.

- e. El indicador “Predicciones correctas” es un indicador del número relativo de observaciones que son estimadas correctamente usando un umbral de 0,05 en la predicción. Dado el resultado de la tabla comente la bondad de esta estimación.

CAPÍTULO 15



## DATOS DE PANEL

Las técnicas para el tratamiento de datos de panel han avanzado rápidamente en sus aplicaciones econométricas. Esto, probablemente, se debe a la mayor disponibilidad de paneles. Pero también es importante señalar que hay un más amplio reconocimiento de que la consideración de la estructura de panel de nuestras muestras nos permite un mejor tratamiento de problemas que son importantes para la Teoría Económica. Por esta razón hemos incorporado este capítulo que puede ser considerado una introducción a este tema. Es importante recalcar que esta es simplemente una introducción. El tema de la econometría de panel ha sido un tema de investigación notablemente activo durante los últimos años, con muchos resultados recientes que resultan interesantes. Por esto debemos dejar de lado una larga lista de temas relacionados. Por otro lado, la discusión detallada de los métodos de estimación y los problemas que surgen al considerar paneles, requiere de largas y trabajosas derivaciones matemáticas y estadísticas.

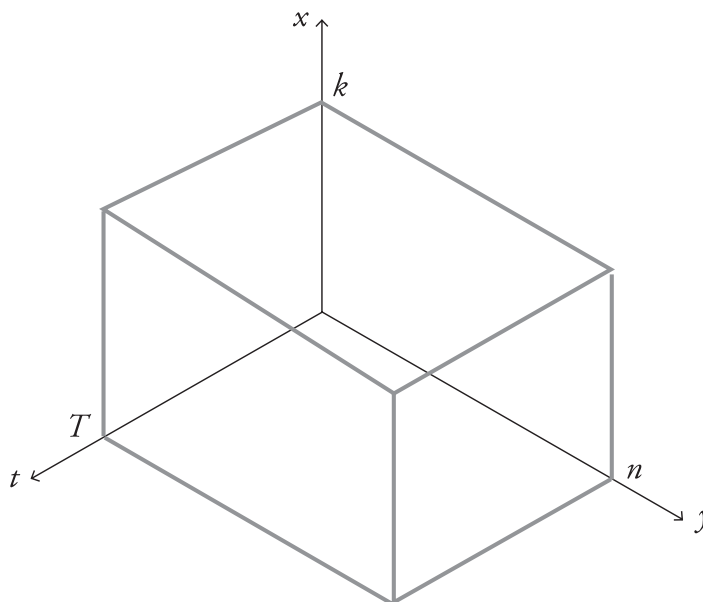
### 15.1 Introducción a los Paneles

Por ***datos de panel*** nos referimos a una muestra de corte de sección que observamos durante varios períodos de tiempo. Es decir, tenemos un cierto número de individuos, o unidades muestrales, que observamos durante varios períodos de tiempo.

El gráfico de la Figura 15.1 ilustra la estructura una muestra de panel. Tenemos  $k$  variables representadas sobre el eje vertical de un gráfico tridimensional, para  $T$  períodos de tiempo y  $n$  individuos, representados sobre los dos ejes del plano de la base. Usualmente representamos nuestras muestras como figuras rectangulares sobre un plano, en que una dimensión corresponde a las variables y la otra

a las observaciones. En este caso la representación es un paralelepípedo tridimensional. Para el caso anterior, ya que cada uno de los  $n$  individuos es observado durante  $T$  períodos, tenemos un total de  $n \times T$  observaciones. Una observación particular de una variable  $x$  será denotada por  $x_{it}$ , donde  $i$  denota al individuo o elemento muestral y  $t$  denota el período en que es observado.

Figura 15.1:  $k$  variables, para  $n$  individuos, durante  $T$  períodos de tiempo



El panel que hemos presentado en la Figura 15.1 es un **panel balanceado**; todos los elementos muestrales o individuos son observados en todos los períodos. Esto no es necesariamente así para todos los paneles. En el caso de paneles no balanceados faltan las observaciones correspondientes a algunos individuos en algunos períodos. Nuestra discusión se centra en el caso de paneles balanceados, pero puede ser fácilmente extendida al caso de paneles no balanceados.

Desde el punto de vista del contenido, hay razones técnicas para distinguir entre dos clases de paneles. Por una parte tenemos **micropaneles**. Es decir, paneles de datos microeconómicos. Generalmente tenemos muchos individuos y pocos períodos. Por esto las cuestiones asintóticas son, en general, relevantes sólo en el sentido de los individuos. Por otra parte tenemos **macropaneles**. Es decir, paneles de datos macroeconómicos. En algún sentido estos paneles pueden ser considerados colecciones de series de tiempo paralelas. En consecuencia, las cuestiones de estacionalidad pueden ser relevantes. También puede haber problemas de dependencia entre las unidades muestrales. Las propiedades asintóticas son, usualmente, relevantes tanto en el sentido de los individuos como del tiempo.

No es difícil encontrar ejemplos de paneles. A continuación discutimos algunos de estos.



### ***Micropaneles***

- Prácticamente en todos los países se realizan encuestas anuales o se llevan registros detallados de actividad industrial. Estos paneles en que cada planta industrial es un individuo en el panel, han sido importantísimos en estudios de teoría de la producción y organización industrial.
- Prácticamente en todos los países se realizan encuestas de hogares periódicas, especialmente para medir el empleo, la distribución del ingreso y los estándares de vida. Es una práctica cada vez más generalizada el dar a estas encuestas un carácter de **paneles rotatorios**, en que un cierto número de individuos o familias son entrevistados durante un cierto número de períodos consecutivos, siendo luego reemplazados por otro grupo.
- Usando los datos de las distribuidoras es posible formar paneles de consumo domiciliario de energía eléctrica.

### ***Macropaneles***

- El Fondo Monetario Internacional (<http://www.imf.org>) publica varios macropaneles en su sitio web. Especialmente importantes son el IFS (International Financial Statistics) y el DOT (Directions of Trade).
- El Banco Mundial (<http://www.wb.org>) es también una fuente importante de macropaneles.
- Otra fuente importante es CEPAL (<http://www.cepal.org>). Esta fuente es importante pues se pueden encontrar paneles que son relevantes para el caso latinoamericano.
- Otras organizaciones que publican macropaneles relevantes son FAO (AGROSTAT), la Unión Europea (EUROSTAT) y la OECD.

## **15.2 Modelos de Una Dirección y de Dos Direcciones**

El modelo lineal general de datos de panel es

$$y_{it} = \alpha + X'_{it}\beta + u_{it}, \quad i \in \{1, \dots, N\}, \quad t \in \{1, \dots, T\} \quad (15.1)$$

En esta ecuación  $\beta$  es un vector  $k \times 1$  de parámetros y  $X'_{it}$  son vectores  $k \times 1$  de variables explicativas. A partir de esta especificación muy general podemos seleccionar diferentes tipologías que admiten tratamientos diferentes para el problema de estimación, utilizando la información contenida en el carácter de panel de la muestra.

Para el modelo de componentes de error de una dirección suponemos que:

$$u_{it} = \mu_i + v_{it}, \quad i \in \{1, \dots, N\}, \quad t \in \{1, \dots, T\} \quad (15.2)$$

donde  $\mu_i$  es un **efecto individual** específico no-observable para la observación  $i$  y  $v_{it}$  son errores estocásticos con las propiedades requeridas para la estimación mínimo-cuadrática.

En forma matricial la ecuación (15.1) es

$$y = \alpha i_{NT} + X\beta + u = Z\delta + u$$

En esta ecuación,  $i_{NT}$  es un vector  $NT \times 1$  de unos,  $y$  es un vector  $NT \times 1$ ,  $X$  es  $NT \times K$ ,  $Z = [i_{NT} \ X]$  y  $\delta' = [\alpha \ \beta']$ . La ecuación (15.2) es, en forma matricial,

$$u = Z_\mu \mu + v$$

Donde  $u' = (u_{11}, \dots, u_{1T}, \dots, u_{N1}, \dots, u_{NT})$ ,  $\mu' = (\mu_1, \dots, \mu_N)$ , y  $v' = (v_{11}, \dots, v_{1T}, \dots, v_{N1}, \dots, v_{NT})$ , y  $Z_\mu = I_N \otimes i_T$ .

Hay dos enfoques para estimar este modelo: el **modelo de efectos fijos** y el **modelo de efectos aleatorios**. La elección entre estos enfoques se ha convertido en una cuestión central de la econometría de datos de panel. A continuación discutiremos estos enfoques.

Naturalmente es posible especificar un modelo con **efectos temporales** en lugar de efectos individuales. Este modelo admite un tratamiento semejante al que discutiremos aquí para los efectos individuales. Finalmente, es posible especificar un **modelo de dos direcciones** con efectos individuales y efectos temporales al mismo tiempo.

### 15.3 El Modelo de Efectos Fijos

Cuando consideramos el Modelo de Efectos Fijos suponemos que los  $\mu_i$  en la ecuación (15.2) son parámetros fijos a ser estimados. Nótese que esto equivale a suponer que tenemos un intercepto diferente por cada individuo. A este supuesto básico podemos agregar que suponemos que los  $v_{it}$  en la ecuación (15.2) son  $\text{IID}(0, \sigma^2)$ . También los  $X_{it}$  se suponen independientes de los  $v_{it}$  para todo  $i$  y todo  $t$ .

Este modelo puede reescribirse así:

$$y = \alpha i_{NT} + X\beta + Z_\mu \mu + v = Z\delta + Z_\mu \mu + v \quad (15.3)$$

donde  $Z = [i_{NT}, X]$ . Con estos supuestos el modelo de la ecuación (15.3) puede ser estimado utilizando Mínimos Cuadrados Ordinarios.

Nótese que lo que hemos hecho es esencialmente equivalente a introducir variables mudas que capturan el efecto de los  $\mu_i$ . Es decir estos efectos individuales son capturados por los coeficientes de estas variables mudas. Sin embargo, hay ciertas dificultades y problemas que limitan el uso de este modelo, o su estimación directa utilizando variables mudas.

En primer lugar, hay que tener cuidado con la llamada trampa de las variables mudas. Es decir, sólo podemos introducir  $N-1$  variables mudas. De otra manera creamos una situación de multicolinealidad perfecta que hace el modelo inestimable. Por otra parte, ciertamente hay un riesgo de sobreparametrización cuando  $N$  es grande y  $T$  es pequeño. En la práctica este enfoque, estimar efectos fijos usando variables mudas, es recomendable cuando tenemos muchos períodos y comparativamente pocos individuos. Este estimador muchas veces es llamado **MCVM (Mínimos Cuadrados con Variables Mudadas)** y **LSDV** en inglés.

Considerando que los parámetros de interés son  $\alpha$  y  $\beta$ , podemos usar un estimador alternativo a **MCMV** (mínimos cuadrados con variables mudas). Multipliquemos la ecuación (22.3) por la matriz  $Q$ , cuya derivación se encuentra en la sección (2.5) del Apéndice A, y obtenemos

$$Qy = QX\beta + Qv \quad (15.4)$$

Esto resulta así porque  $QZ_\mu = QI_{NT} = O$ . Recordemos que la matriz  $Q$  es una transformación que toma diferencias de una variable con respecto a las medias temporales. En consecuencia, como los efectos individuales son constantes en el tiempo, la matriz  $Q$  es, simplemente, una transformación que permite descontar la influencia de los efectos individuales no-observados en el modelo.

Consideremos el vector  $\tilde{y} = Qy$  con elemento típico  $(y_{it} - \bar{y}_i)$  y  $\tilde{X} = QX$  con elemento típico  $(X_{it} - \bar{X}_i)$ , donde  $\bar{y}_i$  y  $\bar{X}_i$  son medias individuales tomadas a lo largo del tiempo. Es decir, en lugar de las variables originales consideraremos las desviaciones en torno a la media de cada individuo entorno a su media en el tiempo. Ahora podemos escribir

$$\tilde{y} = \tilde{X}\beta + \tilde{v} \quad (15.5)$$

Esta transformación nos permite definir el estimador **within** (interior) de la siguiente manera

$$\begin{aligned} \tilde{\beta} &= (X'QX)^{-1} X'Qy \\ \text{var}(\tilde{\beta}) &= \sigma_v^2 (X'QX)^{-1} \end{aligned}$$

Los estimadores para  $\alpha$  y los  $\mu_i$  pueden ser recuperados tomando la media de la regresión (22.3) e introduciendo la condición de que  $\sum_{i=1}^N \mu_i = 0$ .

El enfoque anterior puede ser ilustrado con un modelo de una sola variable. En este caso tenemos el siguiente modelo:

$$y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it} \quad (15.6)$$

En primer lugar tomamos medias sobre el tiempo y tendremos,

$$\bar{y}_{i.} = \alpha + \beta \bar{x}_{i.} + \mu_i + \bar{v}_{i.} \quad (15.7)$$

Substrayendo la ecuación (15.7) de la ecuación (15.6) tenemos

$$y_{it} - \bar{y}_{i.} = \beta (x_{it} - \bar{x}_{i.}) + (v_{it} - \bar{v}_{i.}) \quad (15.8)$$

Introducimos la condición  $\sum_{i=1}^N \mu_i = 0$  y tomamos medias de la ecuación (15.6) sobre todas las observaciones. Entonces tendremos,

$$\bar{y}_{..} = \alpha + \beta \bar{x}_{..} + \bar{v} \quad (15.9)$$

Dados estos resultados, el estimador  $\tilde{\beta}$  puede ser obtenido de la ecuación (15.8) por mínimos cuadrados ordinarios. De la ecuación (15.9) podemos recuperar el estimador de  $\alpha$  como  $\tilde{\alpha} = \bar{y}_{..} - \tilde{\beta} \bar{x}_{..}$ . Finalmente, los estimadores de los efectos fijos  $\mu_i$  serán recuperados por  $\tilde{\mu}_i = \bar{y}_{i.} - \tilde{\alpha} - \tilde{\beta} \bar{x}_{i.}$

Las propiedades de estos estimadores son adecuadas, cuando podemos aceptar los correspondientes supuestos. En primer lugar hay que hacer notar que el estimador MCMV es MELI (mejor estimador lineal insesgado). En segundo lugar, estos estimadores son consistentes para  $T$  grande ( $T \rightarrow \infty$ ). Por otra parte, para un valor de  $T$  fijo y  $N$  grande ( $N \rightarrow \infty$ ), el estimador de  $\beta$  es consistente, pero no los estimadores de los efectos fijos. Esto se debe a que el número de parámetros aumenta cuando  $N$  crece. Finalmente, hay que observar que, bajo el supuesto de que la especificación de efectos fijos es correcta, la estimación de la ecuación (15.1) directamente por mínimos cuadrados dará como resultado estimadores sesgados e inconsistentes, debido al sesgo de variables omitidas.

Ambos estimadores son atractivos. Sin embargo ambos presentan algunos problemas que deben ser señalados. En primer lugar, es fácil encontrar situaciones en que estemos considerando demasiados regresores. Recordemos que la consideración de efectos fijos implica introducir  $(N-1)$  parámetros extra. La consecuencia de estos parámetros adicionales es la disminución de los grados de libertad que, en algunos casos, pueden afectar seriamente la eficiencia de los estimadores. Por otra parte, con frecuencia la introducción de un número excesivo de variables mudas implica un aumento inaceptable del grado de multicolinealidad. Finalmente, debemos señalar que el estimador *within* puede afrontar una seria dificultad que anula su utilidad. En efecto, el estimador *within* no puede considerar variables que son invariantes en el tiempo, en las cuales toda la variación tiene lugar

entre individuos. Un ejemplo es un modelo donde quisiéramos usar la variable género de los individuos como una de las variables explicativas.

Podemos probar la hipótesis de efectos fijos usando una prueba estadística centrada en la significación simultánea de las variables mudas individuales. La hipótesis a probar es:

$$H_0 : \mu_1 = \dots = \mu_{N-1} = 0 \text{ contra } H_1 : \mu_i \neq 0 \text{ para algún } i \in \{1, 2, \dots, N-1\}$$

Esta es una prueba  $F$ . La suma restringida de cuadrados de residuales (RRSS) viene de la estimación mínimo-cuadrática del modelo presentado en (15.1), y la suma irrestricta (URSS) viene de la regresión MSVM. Si  $N$  es grande podemos usar la transformación “within” para la suma irrestricta de cuadrados. El estadígrafo de prueba es

$$F = \frac{(RRSS - URSS) / (N-1)}{URSS / (NT - N - K)} \sim F_{N-1, N(T-1)-K}$$

Dado que en un panel hacemos una referencia explícita a la estructura de la muestra, en términos de individuos y de períodos, es inevitable considerar la existencia de algún grado de heterocedasticidad. Esto nos lleva a considerar la posibilidad de una estimación robusta de los errores estándar. Arellano sugirió un estimador robusto de la matriz de varianzas y covarianzas análogo al estimador robusto de White.

Finalmente queremos hacer notar que la transformación  $\tilde{y}_{it} = y_{it} - \bar{y}_i$  no es la única posible para obviar los efectos fijos. También podemos usar  $\Delta y_{it} = y_{it} - y_{i,t-1}$ . El estimador derivado de esta transformación se ha llamado estimador de **diferencia en diferencias**.

## 15.4 El Modelo de Efectos Aleatorios

En el caso del modelo de efectos aleatorios suponemos que los efectos individuales  $\mu_i$  son variables aleatorias, las que estarían idéntica e independientemente normalmente distribuidas. Es decir, suponemos que  $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ . Tal como hicimos en el caso de modelo de efectos fijos, suponemos que los términos de error estocásticos  $v_{it}$  también están idéntica e independientemente normalmente distribuidos. Es decir, suponemos que  $v_{it} \sim \text{IIN}(0, \sigma_v^2)$ . También suponemos que los  $\mu_i$  son independientes de los  $v_{it}$ , y que los  $X_{it}$  son independientes (ortogonales) de los  $v_{it}$  y los  $\mu_i$  para todo  $i$  y para todo  $t$ . Esta es la especificación apropiada cuando la muestra es de  $N$  individuos elegidos aleatoriamente de una población mayor. En este caso la inferencia que deseamos es sobre toda la población.

Dada esta especificación del modelo de efectos aleatorios o estocásticos, la matriz de varianzas y covarianzas es

$$\begin{aligned}\Omega &= E(uu') = Z_\mu E(\mu\mu') Z_\mu' + E(vv') \\ &= \sigma_\mu^2 (I_N \otimes J_T) + \sigma_v^2 (I_N \otimes I_T)\end{aligned}$$

Esta matriz  $\Omega$  de varianzas y covarianzas contiene una varianza homocedástica y correlación serial para el mismo individuo:

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_v^2, \text{ para } i = j, t = s \\ &= \sigma_\mu^2, \text{ para } i = j, t \neq s\end{aligned}$$

Entonces, el coeficiente de correlación entre  $u_{it}$  y  $u_{js}$  es

$$\begin{aligned}\rho &= \text{corr}(u_{it}, u_{js}) = 1, \text{ para } i = j, s = t \\ &= \sigma_\mu^2 / (\sigma_\mu^2 + \sigma_v^2), \text{ para } i = j, s \neq t\end{aligned}$$

La conclusión es que el considerar efectos aleatorios implica una matriz de varianzas y covarianzas con una estructura estocástica más compleja que la que esperamos en el caso de mínimos cuadrados. Básicamente esta estructura está dada por la presencia de elementos diferentes de cero fuera de la diagonal principal. Dada esta estructura estocástica el enfoque adecuado es de Mínimos Cuadrados Generalizados (GLS). Sin embargo, la matriz  $\Omega$  es muy grande y  $\Omega^{-1}$  es difícil de obtener.

Este problema se puede resolver utilizando algunas de las matrices obtenidas en el Apéndice A. (ver Apéndice A, Sección 2.4). Se puede demostrar que

$$\Omega^r = (\sigma_1^2)^r P + (\sigma_v^2)^r Q$$

donde  $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$ .

Entonces

$$\begin{aligned}\Omega^{-1} &= \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_v^2} Q \\ \Omega^{-1/2} &= \frac{1}{\sigma_1} P + \frac{1}{\sigma_v} Q\end{aligned}$$

Para utilizar estas fórmulas necesitamos estimadores consistentes de  $\sigma_1^2$  y de  $\sigma_v^2$ . Con este propósito podemos usar las siguientes fórmulas para estos estimadores:

$$\hat{\sigma}_1^2 = \frac{u'Pu}{\text{tr}(P)} = T \sum_{i=1}^N \bar{u}_i^2 / N$$

$$\hat{\sigma}_v^2 = \frac{u'Qu}{\text{tr}(Q)} = \frac{\sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2}{N(T-1)}$$

Pero la utilidad de estas fórmulas todavía depende de que encontremos un estimador consistente de  $u$ .

Amemiya propone usar el estimador within, debido a que este estimador es consistente. Es decir, propone que en la fórmula anterior se use,

$$\tilde{u} = y - \tilde{\alpha}t_{NT} - X\tilde{\beta}$$

Swamy y Arora proponen dos regresiones. Primero el estimador within anterior nos da

$$\hat{\sigma}_v^2 = \left[ y'Qy - y'QX(X'QX)^{-1}X'Qy \right] / [N(T-1) - K]$$

En segundo lugar el estimador **between** que es una regresión sobre los promedios a lo largo del tiempo. Es decir

$$\bar{y}_i = \alpha + \bar{X}_i'\beta + \bar{u}_i, \quad i \in \{1, \dots, N\}$$

Pero aquí hay sólo  $N$  observaciones, usar entonces,  $\sqrt{T}\bar{y}_i = \alpha\sqrt{T} + \sqrt{T}\bar{X}_i'\beta + \sqrt{T}\bar{u}_i$ . Como  $\text{var}(\sqrt{T}\bar{u}_i) = \sigma_1^2$  podemos estimar

$$\hat{\sigma}_1^2 = \left[ y'Py - y'PZ(Z'PZ)^{-1}Z'Py \right] / (N - K - 1)$$

Con estos estimadores de los componentes de la varianza podemos construir el estimados GLS.

Es posible demostrar que existe la siguiente relación entre estos estimadores,  $\hat{\beta}_{GLS} = W_1\tilde{\beta}_{\text{within}} + W_2\hat{\beta}_{\text{between}}$ .

Nótese que el estimador within ignora la variación entre los grupos y el estimador between ignora la variación dentro de los grupos. Si  $\sigma_\mu^2 = 0$  entonces  $\hat{\beta}_{GLS}$  se reduce a  $\hat{\beta}_{OLS}$ . Si  $T \rightarrow \infty$  entonces  $\hat{\beta}_{GLS} \rightarrow \tilde{\beta}_{\text{within}}$ .

Bajo el supuesto de efectos aleatorios, la estimación GLS basada en los verdaderos componentes de la varianza es BLUE y los estimadores son asintóticamente eficientes tanto cuando  $N$  como  $T$  tienden a infinito.

Podemos anotar algunas propiedades del estimador GLS de Swamy-Arora. Los estimadores GLS factibles son más eficientes que los del modelo MCVM, salvo para grados de libertad muy reducidos. En segundo lugar, estudios empíricos muestran

que la varianza de los estimadores GLS factibles no es nunca más de un 17 por ciento que el límite de Cramer-Rao (límite inferior asintótico de la varianza). Finalmente, estimadores más eficientes de los componentes de la varianza no necesariamente resultan en estimadores GLS factibles más eficientes.

También podemos pensar en la estimación máximo-verosímil del modelo de efectos aleatorios. La función logarítmica de verosimilitud es

$$L(\alpha, \beta, \phi^2, \sigma_v^2) = C - \frac{NT}{2} \log \sigma_v^2 + \frac{N}{2} \log \phi^2 - \frac{1}{2\sigma_v^2} u' \Sigma^{-1} u$$

donde  $C$  es una constante,  $\Omega = \sigma_v^2 \Sigma$ ,  $\phi^2 = \sigma_v^2 / \sigma_1^2$  y  $\Sigma = Q + \phi^{-2} P$ .

Sin embargo, maximizar esta función de verosimilitud es una tarea que no está exenta de problemas. En primer lugar esta función presenta problemas de concavidad. De hecho Maddala demostró que puede haber hasta dos máximos para esta función cuando  $0 < \phi^2 \leq 1$ . Breusch sugiere un procedimiento para afrontar este problema, comenzar a iterar con  $\tilde{\beta}_{\text{within}}$  como solución inicial, después comenzar a iterar con  $\hat{\beta}_{\text{between}}$  como valor inicial. Si se llega al mismo máximo este sería el máximo global.

## 15.5 ¿Efectos Fijos o Efectos Aleatorios?

El elegir entre efectos fijos y efectos aleatorios es básicamente una cuestión de especificación y contenido del modelo. Sin embargo, existe una prueba estadística que aporta evidencia empírica apoyando una u otra especificación. Esta es la llamada **Prueba de Hausman**.

Recordemos que un supuesto básico es que  $E(u_{it} | X_{it}) = 0$ . Pero los términos estocásticos del modelo contienen efectos individuales  $\mu_i$  que no son observados y que pueden estar correlacionados con los  $X_{it}$ . Entonces podría ser que  $E(u_{it} | X_{it}) \neq 0$ . En este caso los estimadores  $\beta_{\text{GLS}}$  son inconsistentes, pero los estimadores  $\beta_{\text{within}}$  siguen siendo insesgados y consistentes. El enfoque de Hausman para construir esta prueba sugiere comparar el estimador  $\beta_{\text{GLS}}$  y el estimador  $\beta_{\text{within}}$ , ambos consistentes bajo la hipótesis nula  $H_0 : E(u_{it} | X_{it}) = 0$ . Nótese que el estimador  $\beta_{\text{within}}$  es consistente tanto si  $H_0$  es cierta como si no lo es. Por su parte, el estimador  $\beta_{\text{GLS}}$  es BLUE, consistente y asintóticamente eficiente bajo  $H_0$ , pero es inconsistente si  $H_0$  no es cierta.

Para construir la prueba de Hausman consideraremos la diferencia



$$\hat{q}_1 = \beta_{\text{GLS}} - \beta_{\text{witin}}$$

Nótese que, bajo la condición de que la hipótesis nula es cierta,  $\text{plim } \hat{q}_1 = 0$  y podemos probar que

$$\begin{aligned} E(\hat{q}_1) &= 0 \\ \text{var}(\hat{q}_1) &= \text{var}(\beta_{\text{witin}}) - \text{var}(\beta_{\text{GLS}}) \end{aligned}$$

El estadígrafo de Hausman es, entonces,

$$m_1 = \hat{q}_1' [\text{var}(\hat{q}_1)]^{-1} \hat{q}_1$$

Este indicador se distribuye  $\chi_K^2$  bajo  $H_0$ , donde  $K$  es igual al número de regresores. Para hacer operacional este indicador la matriz de varianzas y covarianzas  $\Omega$  se remplaza por un estimador consistente de la misma, y el estimador GLS se convierte en un estimador GLS factible.

Una prueba alternativa asintóticamente equivalente a la anterior se obtiene de la regresión aumentada

$$y^* = X^* \beta + \bar{X} \gamma + w$$

donde  $y^* = \sigma_v \Omega^{-1/2} y$ ,  $X^* = \sigma_v \Omega^{-1/2} X$  y  $\bar{X} = QX$ . Esta prueba es equivalente a una prueba de la hipótesis nula  $\gamma = 0$ . Es decir una prueba estándar de Wald para la omisión de las variables  $\bar{X}$ .

Es posible ver que, bajo la hipótesis nula,

$$\begin{aligned} E(\hat{\gamma}) &= 0 \\ \text{var}(\hat{\gamma}) &= \text{var}(\beta_{\text{within}}) + \text{var}(\beta_{\text{between}}) \end{aligned}$$

El estadígrafo de prueba correspondiente es, entonces,

$$m_2 = \hat{\gamma}' [\text{var}(\hat{\gamma})]^{-1} \hat{\gamma}$$

y se distribuye  $\chi_K^2$ .

En realidad es posible desarrollar tres pruebas equivalentes basados en tres diferentes diferencias

$$\begin{aligned} \hat{q}_1 &= \beta_{\text{GLS}} - \beta_{\text{witin}} \\ \hat{q}_2 &= \beta_{\text{GLS}} - \beta_{\text{between}} \\ \hat{q}_3 &= \beta_{\text{within}} - \beta_{\text{between}} \end{aligned}$$

## 15.6 Consideraciones Finales

Naturalmente, las anteriores son estimaciones básicas que pueden ser sometidas a numerosas consideraciones de acuerdo a la verosimilitud de los supuestos en que estos enfoques se basan. En especial, la estructura de la muestra con direcciones diferentes de variación, a lo largo del tiempo y a través de los individuos, hace necesario plantearse el problema de los errores *no-esféricos*, *heterocedasticidad* y *correlación serial*. Diversos enfoques han sido desarrollados y publicados para tratar estos problemas. Muchos de ellos ya aparecen rutinariamente incluidos en paquetes estándar de software especializados en Econometría. También se han desarrollado técnicas que permiten acomodar en estas estimaciones el problema de paneles no-balanceados.

El hecho de la existencia y consideración de una dimensión temporal en los paneles hace inevitable el incluir otras consideraciones en que esta dimensión temporal es crítica. En primer lugar, es necesario considerar de que, en un sentido, un panel es una colección de series de tiempo paralelas. Esto hace inevitable el considerar la cuestión de la estacionaridad. En segundo lugar, es necesario considerar la posibilidad de especificar modelos dinámicos. Desgraciadamente, carecemos de espacio en este texto para desarrollar el tratamiento de estos problemas.

Finalmente, debemos mencionar que un área de notable actividad de investigación en recientes años, son los modelos no-lineales en un marco de datos de panel. Esto incluye tanto modelos de variable dependiente cualitativa como modelos de variable dependiente limitada. También hemos excluido la discusión del tratamiento de estos modelos. Poco a poco los paquetes de software econométrico han ido incluyendo rutinas para estimar estos modelos.



# *preguntas*

Trate de responder las preguntas a continuación. Con ello comprobará si ha captado e incorporado los contenidos de este capítulo.

1. Describa la estructura de una muestra de panel. De ejemplos.
2. ¿Qué es el modelo de efectos fijos?
3. ¿Qué es el modelo de efectos aleatorios?
4. ¿Qué es el estimador within?
5. ¿Por qué el modelo de de efectos aleatorios no puede ser estimado usando mínimos cuadrados?
6. ¿Qué criterios usaría para elegir entre efectos fijos y efectos aleatorios?



# ejercicios

- 15.1 Con el objeto de estudiar el comportamiento de la participación femenina en la fuerza de trabajo en Chile, se construyó un panel sobre las 13 regiones en que se divide el país. Para ello se recogió información sobre la participación femenina y un grupo de variables que estaban disponibles. En primer lugar se estimó un modelo lineal con variables mudas. Los resultados de la estimación se presentan en la siguiente tabla.

| Efectos Fijos con Variables Mudas<br>(Referencia: Región Metropolitana) |            |          |        |
|-------------------------------------------------------------------------|------------|----------|--------|
| Obs                                                                     | 221        |          |        |
| F(16,204)                                                               | 53,96      | Prob>F   | 0,0000 |
| R <sup>2</sup>                                                          | 0,8097     | Raíz-ECM | 1,6728 |
|                                                                         |            |          |        |
| Variable                                                                | Coef.      | t        | P> t   |
| Región 1                                                                | -0,2241941 | -0,36    | 0,721  |
| Región 2                                                                | -3,151917  | -7,51    | 0,000  |
| Región 3                                                                | -0,2814283 | -1,05    | 0,296  |
| Región 4                                                                | -0,9912798 | -4,23    | 0,000  |
| Región 5                                                                | -1,113838  | -8,10    | 0,000  |
| Región 6                                                                | -0,7938289 | -6,00    | 0,000  |
| Región 7                                                                | -0,1652922 | -1,26    | 0,210  |
| Región 8                                                                | -0,6547888 | -7,14    | 0,000  |
| Región 9                                                                | -0,7800642 | -7,35    | 0,000  |
| Región 10                                                               | -0,4472083 | -5,83    | 0,000  |
| Región 11                                                               | 0,5681679  | 5,09     | 0,000  |
| Región 12                                                               | -0,1177877 | -1,91    | 0,058  |
| PIB pc                                                                  | 0,257941   | 0,91     | 0,364  |
| Export. pc                                                              | 0,1508299  | 0,32     | 0,751  |
| Inv. Pub. pc                                                            | -0,0389972 | -2,21    | 0,028  |
| Desempleo                                                               | -0,0652742 | -0,87    | 0,387  |
| Constante                                                               | 56,48419   | 60,95    | 0,000  |

Test de los efectos fijos,  $H_0$ : Región 1=Región 2=....=Región 12=0.

F(12,204) = 36,82

Prob > F = 0,0000

- a. Comente estos resultados en términos de significación estadística del modelo y de los parámetros, interpretación e importancia de los mismos y validez de la hipótesis de efectos fijos.

- b. ¿Cree usted que este resultado entrega evidencia empírica en favor de una conducta idiosincrática de las regiones? Es decir, ¿hay efectos específicos por región que afectan la participación femenina?
- c. Interprete los resultados para la Región 2.

Una estimación, utilizando el estimador within dio los siguientes resultados

| Estimación within                                                                                        |            |           |        |
|----------------------------------------------------------------------------------------------------------|------------|-----------|--------|
| R <sup>2</sup> :                                                                                         |            |           |        |
| within                                                                                                   | 0,0330     | Obs/grupo | 17     |
| between                                                                                                  | 0,0454     |           |        |
| total                                                                                                    | 0,0174     |           |        |
| F(4,204) = 1,63                                                                                          |            | Prob>F    | 0,1684 |
| corr(u <sub>i</sub> , Xb)                                                                                | -0,3248    |           |        |
|                                                                                                          |            |           |        |
| Variable                                                                                                 | Coef.      | t         | P>  t  |
| PIB pc                                                                                                   | 0,257941   | 0,91      | 0,364  |
| Export. pc                                                                                               | 0,1508299  | 0,32      | 0,751  |
| Inv. Pub. pc                                                                                             | -0,0389972 | -2,21     | 0,028  |
| Desempleo                                                                                                | -0,0652742 | -0,87     | 0,387  |
| Constante                                                                                                | 53,81333   | 84,92     | 0,000  |
|                                                                                                          |            |           |        |
| σ <sub>u</sub>                                                                                           | 3,6091882  |           |        |
| σ <sub>e</sub>                                                                                           | 1,6727923  |           |        |
| ρ                                                                                                        | 0,82317048 |           |        |
| Nota: El indicador ρ representa la fracción de varianza debida a los efectos aleatorios u <sub>i</sub> . |            |           |        |

- a. ¿Hay alguna ventaja en usar el estimador within?
- b. ¿Hay algún resultado que sugiera efectos aleatorios?

CAPÍTULO 16



## MODELOS DE FRONTERA O ASIMÉTRICOS

Este capítulo se refiere a la estimación de un tipo de modelos que aparecen sobre todo en el contexto de Teoría de la Producción y de Funciones de Producción. Una discusión detallada y matemáticamente rigurosa de la Teoría de la Producción nos lleva naturalmente a la consideración de fronteras de producción. La instrumentalización empírica de este instrumento analítico implica la necesidad de la estimación estocástica de fronteras, o estimación de funciones estocásticas asimétricas. Este capítulo presenta entonces una introducción sucinta a este tipo de estimación.

### 16.1 Estimaciones Simétricas y Estimaciones Asimétricas

La necesidad de estimar modelos que incluyen una componente estocástica asimétrica surgió en la discusión de problemas de Teoría de la Producción. Un punto central en el enfoque tradicional de Teoría de la Producción y Organización Industrial es el de estimar funciones *medias* de producción o de costos. En este caso se trata de estimar modelos que pueden ser especificados como  $y_i = f(\mathbf{x}_i) + \vartheta_i$ , donde  $y_i$  es un indicador del producto y donde  $\mathbf{x}_i$  es un vector de insumos, para todo  $i \in \{1, \dots, n\}$ , una muestra de plantas industriales, por ejemplo.

El supuesto básico para la componente estocástica es que los errores se distribuyen independiente e idénticamente siguiendo una distribución normal con media cero y varianza constante sobre las unidades muestrales (firmas). Es decir, tenemos que:

$$\vartheta_i \sim IIDN(0, \sigma^2) \text{ para todo } i \in \{1, \dots, n\}$$

Hay un tratamiento econométrico estándar para este problema, usualmente con *Mínimos Cuadrados*, cuando tenemos especificaciones lineales en los parámetros.

Hay varias especificaciones para la función de producción,  $f$ , que han sido detalladamente estudiadas; por ejemplo, la función Cobb-Douglas, las funciones CES, especificaciones trans-log, etc. Algunas de estas especificaciones conllevan posibles complicaciones por no-linealidades en los parámetros de  $f$ . También podemos tener problemas derivados de las características de la distribución de  $\hat{\eta}_i$ ; errores no-esféricos, por ejemplo. Hay soluciones bien conocidas para todos estos problemas.

Sin embargo, nunca se cuestiona el carácter simétrico de la distribución de  $\hat{\eta}_i$  y se supone que estos se distribuyen simétricamente en torno al cero. Avances en la fundamentación matemáticamente rigurosa de la Teoría de la Producción, por una parte, y la introducción de medidas de eficiencia por Farrell, por otra parte, llevaron a la necesidad de estimaciones en que los errores estocásticos siguen una distribución que no es simétrica.

## 16.2 Fronteras Estocásticas

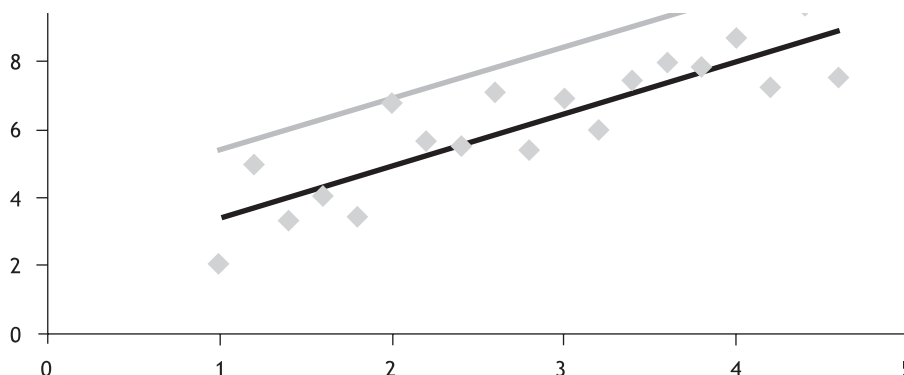
Posiblemente el problema de estimación de modelos con distribución asimétrica resulta más fácil de comprender apoyándose en una discusión de las fronteras de producción. Podemos partir de la observación de que distintas combinaciones de insumos sirven para producir distintos niveles de producto. Sin embargo, esta relación no es uno a uno. De hecho, un mismo nivel de producto puede ser el resultado de diferentes combinaciones de insumos. De esta manera tenemos un conjunto de combinaciones de insumos y niveles de producto, el que está limitado por una función  $y = f(\mathbf{x})$  siguiendo la notación anterior. Esta función, que limita el conjunto de las posibles combinaciones de insumos y niveles de producto, la interpretamos como indicando el máximo producto que es posible, dada la tecnología disponible, representada por la función  $f$ . Otra interpretación posible es considerar que la función  $f$  representa la *mejor práctica* presente para unidades productivas que tienen disponible la misma o muy similares tecnologías. Esta función se llama frontera de producción, porque las unidades productivas son observadas con resultados por *debajo* de esta función. Hay otros elementos, usualmente no observables o cuantificables, que determinan el resultado de una unidad productiva y su cercanía a la frontera de producción. Por ejemplo, la calidad de la gerencia de las unidades productivas. La especificación correcta es, entonces,  $y \leq f(\mathbf{x})$ . Estas ideas están ilustradas en el gráfico 16.1, donde suponemos la existencia de una función lineal como frontera de producción.

En dicho gráfico suponemos que en el eje vertical ( $y$ ) tenemos un índice que mide el producto de las distintas unidades productivas, mientras que en el eje horizontal tenemos un índice compuesto de los insumos ( $\mathbf{x}$ ) realmente usados por



cada una de las unidades productivas incluidas en la muestra. La línea de regresión inferior representa una función de producción tradicional o media, mientras que la línea superior representa una frontera de producción. Nótese que todas las observaciones quedan por debajo de esta frontera. La distancia de estas observaciones a la frontera es un indicador de la *eficiencia técnica* de las distintas unidades productivas. Para una unidad que se encuentra más cercana a la frontera, diremos que es más eficiente que otras unidades que están más lejos. Diversas medidas de esta distancia han sido desarrolladas, pero todas ellas requieren la estimación de la frontera de producción.

Figura 16.1: Líneas de regresión de frontera y media



Las fronteras de producción sugieren que las observaciones están a un sólo lado de la frontera. Los elementos que determinan la eficiencia de las distintas unidades productivas son, en general, no observables o difícilmente cuantificables, los que pueden suponerse aleatoriamente distribuidos entre las unidades productivas. Esto implica que necesitamos una distribución asimétrica para  $\varepsilon_i$  si queremos especificar el modelo econométrico

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \quad (16.1)$$

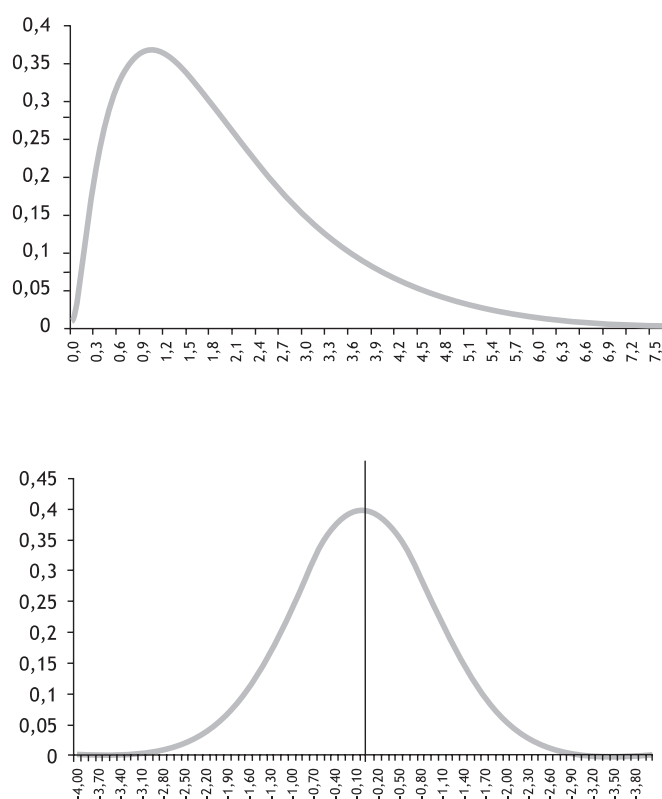
### 16.3 Estimación de Fronteras Estocásticas

El problema es, entonces, como estimar un modelo que tiene una distribución asimétrica. Una opción, cuando el modelo es lineal o transformable en una forma lineal, es estimar con mínimos cuadrados y ajustar la constante para dejar todas las observaciones a un sólo lado de la línea de regresión. El estimador resulta insesgado pero muy poco eficiente. En consecuencia, debemos abandonar las

estimaciones mínimo-cuadráticas. Una posibilidad es especificar una distribución asimétrica para la componente estocástica y utilizar estimaciones máximo-verosímiles, las que tienen la ventaja adicional de recuperar posibles no-linealidades en la frontera de producción.

Las primeras estimaciones de una frontera de producción fueron hechas utilizando una distribución gama. El gráfico 16.2 nos muestra una distribución normal y una distribución gama.

**Figura 16.2: Distribución Normal y Distribución Gama**



El primer panel en el gráfico de la figura 16.2 nos muestra la función de densidad de una distribución normal, donde es fácil observar la simetría de esta distribución en torno al cero. El segundo panel nos muestra la función de densidad de una distribución gama. Es fácil observar su asimetría. De hecho la función tiene un soporte o dominio de definición que sólo incluye los números reales positivos. Más aún, podemos ver que la probabilidad de encontrar una unidad productiva con ineficiencia cero tiende a cero. La mayoría de las unidades productivas se concentran en una región muy cercana al cero, y hay una larga cola de unidades menos productivas que se hace más y más tenue a medida que nos movemos hacia la derecha.

Las estimaciones basadas en distribuciones gama han sido criticadas porque atribuyen toda la componente estocástica al efecto de eficiencia, sin dejar espacio para otro tipo de errores estocásticos; errores de medición, por ejemplo. La consideración de errores estocásticos, además de la consideración de la eficiencia técnica, llevó a estimaciones que utilizan la suma de una distribución normal, capturando los efectos de los errores puramente estocásticos del modelo, y una distribución asimétrica, capturando los efectos de eficiencia técnica. Para esta segunda distribución, que captura los efectos de la eficiencia, se han utilizado funciones gama y funciones normales truncadas.

En este enfoque, de dos distribuciones, se trata de estimar:

$$y_i = f(\mathbf{x}_i) + \mu_i - \nu_i$$

donde  $y_i$  es un índice de producción del establecimiento  $i$  y  $\mathbf{x}_i$  es un vector de insumos del establecimiento  $i$ . Por su parte suponemos que  $\mu_i$  es un ruido blanco, con  $\mu_i \sim IIDN(0, \sigma^2)$  para todo  $i \in \{1, \dots, n\}$ . Finalmente, suponemos que  $\nu_i \geq 0$  para todo  $i \in \{1, \dots, n\}$  y que los  $\nu_i$  siguen una distribución asimétrica. Por ejemplo, una gama o una normal truncada. Suponemos que  $\nu_i$  captura las ineficiencias. Este modelo es estimable usando el método de máxima verosimilitud.

La función de verosimilitud de este problema no es demasiado compleja y se comporta bien frente a los algoritmos de maximización habituales. Hay rutinas pre-programadas en **Limdep** y en **Stata**. También hay un programa especializado, **FRONTIER**, el que está disponible en la Internet.

Nótese que el término estocástico es  $\mu_i - \nu_i$ . El problema es que, en principio, con esta estimación no es posible identificar separadamente  $\mu_i$  y  $\nu_i$ . Identificamos sólo el residual global y, en realidad quisiéramos identificar las eficiencias, asociadas a una estimación de  $\nu_i$ . Sin embargo, podemos identificar los parámetros de la distribución de  $\nu_i$ , con lo que podemos identificar, entonces, la eficiencia media de la muestra, por ejemplo. También podemos introducir restricciones, más o menos arbitrarias, para poder identificar las ineficiencias individuales. Estas estimaciones son utilizadas a menudo para establecer una jerarquía que ordena los establecimientos o unidades productivas según su índice de eficiencia.

Cuando consideramos una versión de datos de panel, el problema es claramente uno de efectos aleatorios, en que los efectos individuales siguen una distribución asimétrica. Este es el enfoque preferido en la mayoría de los estudios del sector industrial, donde es frecuente el contar con paneles.

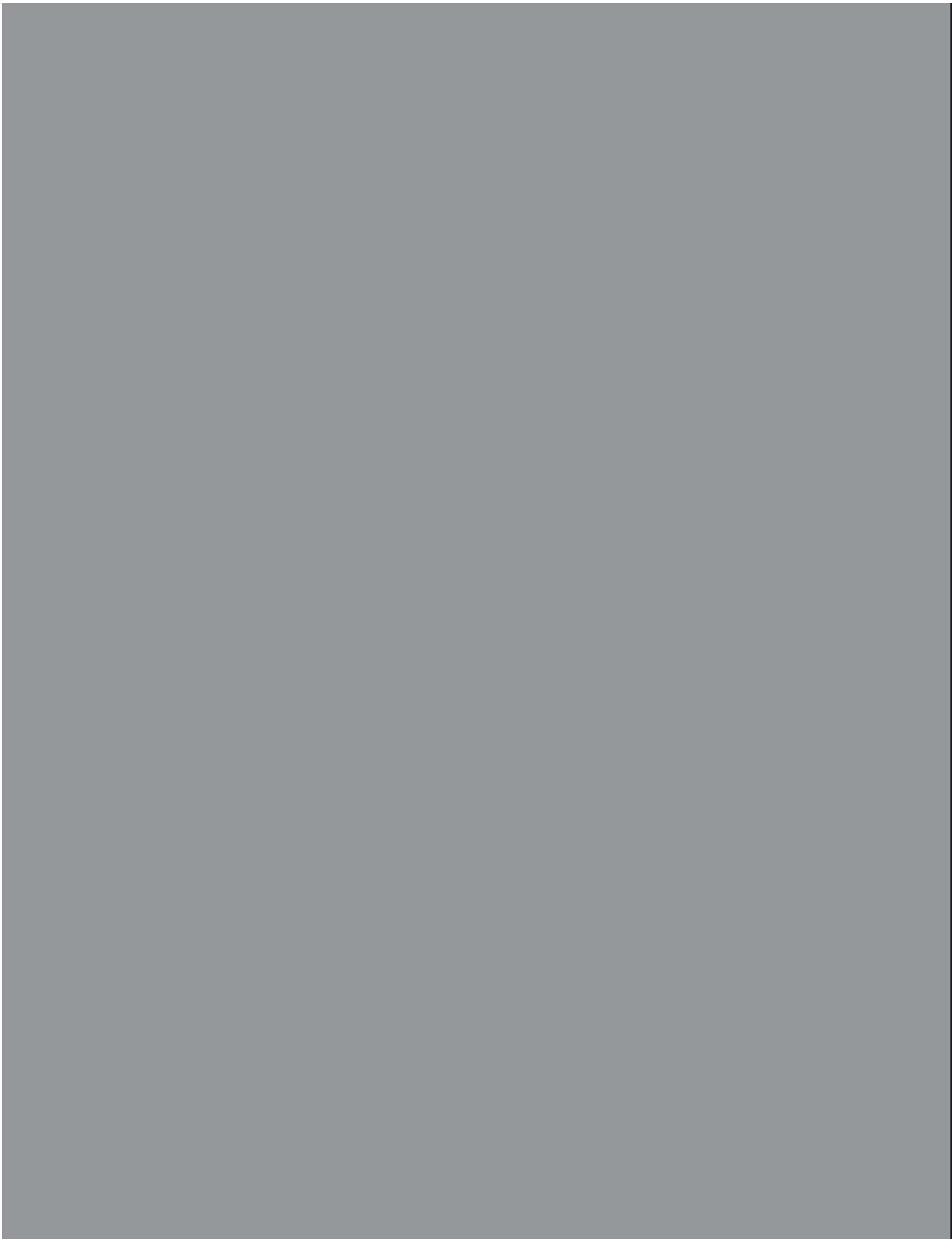


## *preguntas*

Intente dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo.

1. ¿Qué es un modelo de frontera?
2. ¿Cuáles son las distribuciones asimétricas más frecuentemente usadas en estimaciones de frontera?
3. Indique una aplicación de los modelos de frontera.





• PARTE CINCO

• INTRODUCCIÓN  
• A SERIES DE TIEMPO





En esta parte del libro profundizamos en algunos temas que es necesario conocer para trabajar con modelos que involucran datos organizados en el tiempo. Se siguen aplicando los conceptos básicos que se han venido desarrollando sobre el análisis de regresión, la inferencia estadística, los estimadores, sus propiedades y los supuestos. La econometría de series de tiempo, aunque tiene sus particularidades, no se distingue radicalmente de los contenidos estudiados dentro del modelo lineal normal de regresión. Inicialmente se examinan algunas propiedades de las series de tiempo, la prueba Dickey-Fuller para contrastar la estacionariedad, los componentes de las series de tiempo y los modelos ARIMA. Hasta aquí el análisis es univariante, es decir, se trabaja con una sola serie de tiempo. En las siguientes secciones se pasa al análisis multivariante, donde se exploran modelos que relacionan diferentes series de tiempo. Se estudian los modelos ARDL, la cointegración, el mecanismo de corrección de errores y los vectores autorregresivos (VAR). Como siempre, el objetivo último de la estimación de los modelos será realizar pruebas de inferencia estadística y confeccionar predicciones; ambas nos proporcionan información empírica que nos ayudan a entender mejor el fenómeno en estudio y nos facilitan la toma de decisiones. También en esta parte se adicionan a las explicaciones ejemplos reales de la economía cubana.

CAPÍTULO 17



## SERIES DE TIEMPO Y SUS PROPIEDADES

Llamamos serie de tiempo a un conjunto de mediciones de cierto fenómeno registradas secuencialmente en el tiempo. En general, la información de series de tiempo puede tener una periodicidad anual, trimestral, mensual, semanal, diaria, entre otras; de hecho, en los mercados financieros es ahora posible recolectar información con una frecuencia de minutos.

En términos estadísticos se representa que las distintas observaciones de una serie de tiempo son una realización de un proceso estocástico subyacente; también conocido como proceso generador de los datos (PGD). El proceso estocástico se define como una sucesión de variables aleatorias:

$$\{y_t\}, t = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$$

donde  $t$  se refiere al período. Los datos que observamos de las distintas series de tiempo son una única realización del PGD. De la misma manera como se utiliza la información de una muestra de datos de corte transversal para inferir sobre una población, en las series de tiempo se utiliza la realización para inferir sobre el proceso estocástico subyacente.

A continuación vamos a definir diferentes procesos estocásticos teóricos y sus propiedades. En el análisis empírico uno de los primeros pasos que tiene que cumplir el investigador es identificar que proceso estocástico puede estar generando la serie de tiempo con que está trabajando. Una vez identificado el PGD, el investigador conoce diferentes propiedades de sus series de tiempo, a partir de las cuales, tiene una mejor comprensión del tema en estudio, puede elegir con mayor facilidad el tipo de modelo que va a utilizar para la estimación y está más capacitado para confeccionar predicciones.

## 17.1 Procesos Estocásticos Estacionarios y no Estacionarios

Un tipo de proceso estocástico que ha recibido gran atención entre los analistas de series de tiempo es el llamado **proceso estocástico estacionario**. Las propiedades de un proceso  $y_t$  estacionario<sup>1</sup> son las siguientes:

$$\text{Media: } E(y_t) = \mu$$

$$\text{Varianza: } E(y_t - \mu) = \sigma^2$$

$$\text{Covarianza: } E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$$

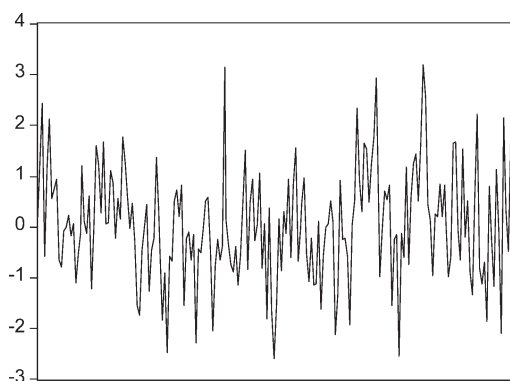
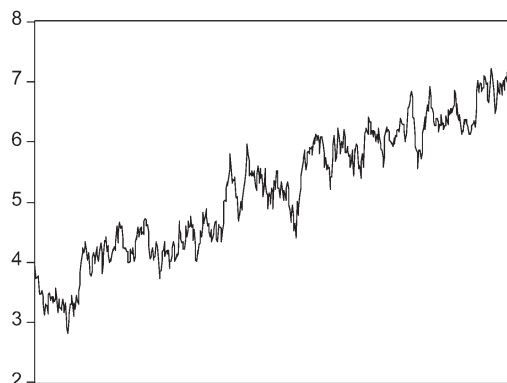
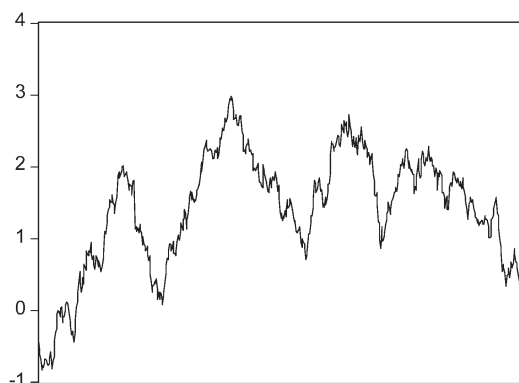
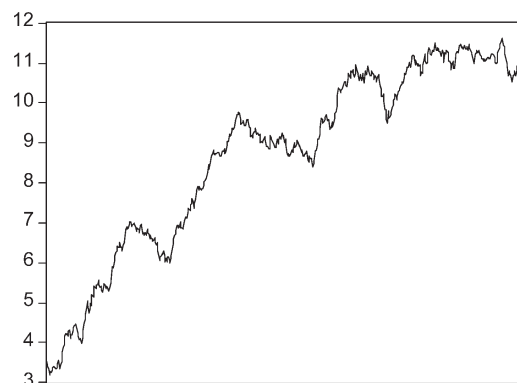
La media y varianza son independientes del tiempo, y el valor de la covarianza entre dos observaciones sólo depende de la distancia o rezago temporal pero no del período en que se mida; es decir, la covarianza depende del rezago  $k$  pero es invariante en el tiempo.

En la figura 17.1 se muestra el gráfico de línea de una serie generada por un proceso estocástico estacionario. La variable oscila alrededor de su valor medio, que en este caso es cero, con una amplitud promedio que no cambia en el tiempo. Este proceso es estacionario tanto en media como en varianza. En la figura 17.2 se grafica una serie generada por un proceso estacionario alrededor de una **tendencia determinista**. Es un proceso no estacionario en media, dado que la media siempre es mayor a medida que pasa el tiempo; sin embargo, es estacionario en varianza. Ambos procesos tienden a volver —convergen— a su media cuando se han desviado de ella, en el primer caso tiende a cero y en el segundo tiende a la tendencia determinista, que en este ejemplo es positiva.

En las figuras 17.3 y 17.4 se muestran las series de dos procesos no estacionarios, tanto en media como en varianza. Las variables tienen períodos de menor varianza y otros en que aumenta la volatilidad. Ambas variables se dice que presentan una **tendencia estocástica**; aunque en el paseo aleatorio con deriva hay un elemento determinista, dado por la deriva positiva que en este ejemplo se le añadió al proceso. Si bien el paseo aleatorio con deriva sigue una trayectoria errática, en el largo plazo el proceso siempre crece.

Numerosas variables de la economía siguen un **proceso evolutivo**: la media depende positivamente del tiempo; por tanto, en principio serían procesos no estacionarios en media. En este caso casi siempre se encuentra el Producto Interno Bruto, las series de stock de capital y trabajo, entre otras.

<sup>1</sup> Se refiere a la estacionariedad débil o estacionariedad covarianza.

**Figura 14.1: Serie estacionaria****Figura 14.2: Serie no estacionaria en media****Figura 17.3: Paseo aleatorio****Figura 17.4: Paseo aleatorio con deriva**

*Fuente: Elaboración propia, series simuladas*

Desde el punto de vista de la predicción, los procesos estacionarios, los no estacionarios en media y el paseo aleatorio con deriva, son relativamente más fáciles de predecir; se tiene la certidumbre de que en el largo plazo la serie tenderá a su media, a su tendencia determinista o a la deriva, respectivamente. Por el contrario, el paseo aleatorio es el proceso que tiene mayor incertidumbre. Tal proceso se suele observar en las series de tiempo asociadas a los mercados financieros: el precio de un bono, el precio de una acción, el valor de una moneda (en un régimen cambiario flexible), etc. Si se encontraran series de tiempo de activos financieros estacionarias sería muy fácil hacer dinero mediante la especulación: comprando el activo cuando el precio se encuentra por debajo del valor medio y vendiendo cuando lo supera. Dada la teoría económica de las expectativas racionales y los mercados financieros eficientes, se plantea que no es posible predecir el precio medio de los activos financieros, por tanto, deben seguir un paseo aleatorio.

Un proceso estocástico más general es el autorregresivo, que incluye como casos particulares los distintos procesos estacionarios y no estacionarios vistos anteriormente. El más simple es el **proceso autorregresivo de orden uno** o **AR(1)**:

$$y_t = a_0 + a_1 y_{t-1} + u_t \quad (17.1)$$

donde  $u_t$  es el término de error o perturbación aleatoria, que en la literatura de series de tiempo se suele denominar como **shock** o **innovación**; pero al igual que en el modelo lineal de regresión, recoge el efecto promedio de las variables relevantes no incluidas en el modelo. Diferentes supuestos que se habían formulado sobre el término de error en las partes anteriores del libro los resumiremos ahora en dos conceptos. Diremos que el término de error  $u_t$  es un **ruido blanco** cuando al mismo tiempo se cumple que tiene media cero, varianza constante y no presenta autocorrelación de ningún orden:

$$E(u_t) = 0$$

$$Var(u_t) = E[u_t^2] = \sigma^2$$

$$Covar(u_t, u_{t-k}) : E[u_t \cdot u_{t-k}] = 0 \quad \forall k \geq 1$$

Si además se cumple que el término de error sigue una distribución normal, entonces diremos que es un **ruido blanco gaussiano**.

Los procesos autorregresivos pueden verse desde un punto de vista matemático como una **ecuación en diferencia**.<sup>2</sup> En las ecuaciones en diferencia la variable es expresada en función de sus propios rezagos, el tiempo y otras variables explicativas. Como en la mayoría de las series de tiempo se observa que el comportamiento presente de la variable depende de sus valores pasados, en los modelos de series de tiempo casi siempre va a aparecer algún término autorregresivo para recoger esta dependencia. Por tanto, la econometría de series de tiempo estará casi siempre estimando ecuaciones en diferencia, o sistemas de ecuaciones en diferencia, como los modelos VAR que se verán posteriormente. En contraste con las formulaciones que usualmente presenta la matemática, en econometría las ecuaciones en diferencia contienen componentes estocásticos.<sup>3</sup>

Este punto de vista es importante debido a que muchos de los análisis en los modelos en series de tiempo se deben hacer en la **solución de la ecuación en diferencia** estimada.<sup>4</sup> Al no contener rezagos de la variable dependiente, en la solución queda explícito el impacto de las variables explicativas en la variable dependiente, así como, el efecto de los shocks. En la solución se puede analizar más fá-

<sup>2</sup> Este enfoque por ejemplo se presenta en el libro de Enders (1995).

<sup>3</sup> Para tener la analogía con las formulaciones matemáticas, el componente estocástico se puede ver como otra variable explicativa.

<sup>4</sup> La solución de una ecuación en diferencia expresa la variable dependiente en función del tiempo, otras variables y alguna condición inicial en algunos casos, es decir, no contiene rezagos de la variable dependiente.

cilmente la trayectoria temporal de la variable y sus propiedades, si es estable, explosiva y a partir de aquí identificar si el proceso es o no estacionario.

Partiendo de una condición inicial  $y_0$ , se puede encontrar la siguiente solución para el proceso autorregresivo de orden uno que presentamos en (17.1):<sup>5</sup>

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i u_{t-i} \quad (17.2)$$

Como habíamos apuntado, en la solución queda más explícita la trayectoria en el tiempo de la variable y el efecto de los shocks. Se puede apreciar que si  $|a_1| = 1$  o  $|a_1| > 1$  la solución no es estable, la sucesión de valores en el tiempo de  $y_t$  no converge. Si  $|a_1| = 1$  la solución será:

$$y_t = a_0 t + y_0 + \sum_{i=0}^{t-1} u_{t-i} \quad (17.3)$$

La variable  $y_t$  cada vez alcanza mayores valores, en cada período acumula  $(a_0 + u_t)$  unidades. Si  $|a_1| > 1$  entonces  $y_t$  seguiría una trayectoria explosiva; esta propiedad no se observa en la práctica en las series económicas, al menos, no en el largo plazo.

La **condición de estabilidad** de un **proceso AR(1)** requiere que  $|a_1| < 1$ . Si asumimos que  $t$  es suficientemente grande, en el límite la solución tendría la expresión:

$$y_t = a_0 / (1 - a_1) + \sum_{i=0}^{\infty} a_1^i u_{t-i} \quad (17.4)$$

debido a que  $\sum_{i=1}^{t-1} a_1^i$  converge a  $1/(1 - a_1)$  y  $a_1^t y_0$  converge a cero.

El valor esperado de esta expresión es  $a_0 / (1 - a_1)$ , el cual es independiente del tiempo. También se puede demostrar que son independientes del tiempo la va-

<sup>5</sup> Iterando a partir de una condición inicial  $y_0$  se llega a la solución, con tres iteraciones podemos sacar la expresión general:

$$y_1 = a_0 + a_1 y_0 + u_1$$

$$y_2 = a_0 + a_1 y_1 + u_2 = a_0 + a_1(a_0 + a_1 y_0 + u_1) + u_2 = a_0(1 + a_1) + a_1^2 y_0 + a_1 u_1 + u_2$$

$$y_3 = a_0 + a_1 y_2 + u_3 = a_0 + a_1[a_0(1 + a_1) + a_1^2 y_0 + a_1 u_1 + u_2] + u_3 = a_0(1 + a_1 + a_1^2) + a_1^3 y_0 + a_1^2 u_1 + a_1 u_2 + u_3$$

Se utilizan diferentes métodos para hallar la solución de una ecuación en diferencia. A partir de una condición inicial —o sin condición inicial— se puede iterar hacia delante o hacia atrás la expresión; también se utiliza el método de los coeficientes indeterminados, o encontrar las raíces de la ecuación característica en la parte reducida de la ecuación. Por estos métodos se llega a una solución general que se divide en la solución homogénea y la solución particular. El objetivo es llegar a una expresión que no contenga la variable dependiente en la parte derecha de la ecuación. La solución debe satisfacer la ecuación en diferencia en todos sus puntos, la sustitución de la solución en la ecuación debe resultar en una identidad.

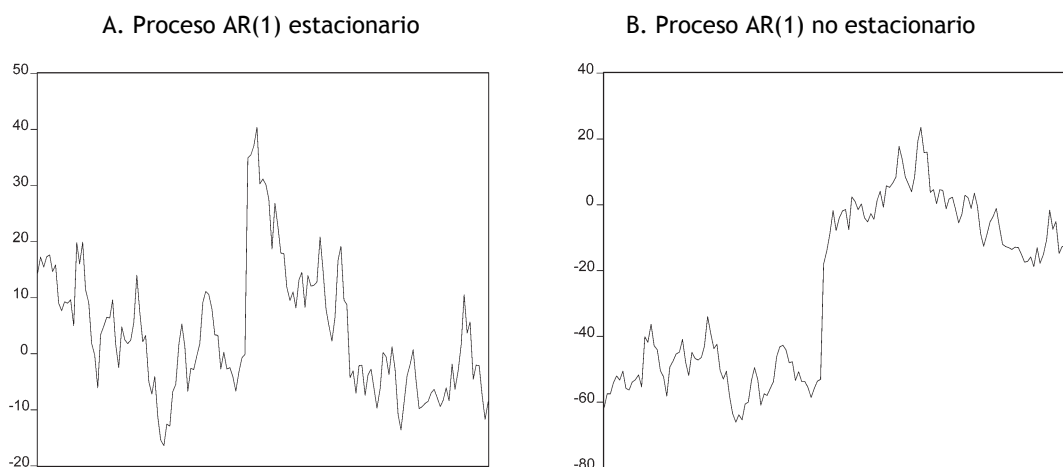
rianza y la covarianza; completándose las tres propiedades que posee una serie estacionaria según definimos al inicio de esta sección.

En resumen, para que un proceso estocástico autorregresivo de orden uno sea estacionario, este debe satisfacer la condición de estabilidad y además  $t$  debe ser suficientemente grande. Debido a esto último, los investigadores deben tener cuidado cuando trabajan con datos generados por un proceso “nuevo”, los cuales podrían ser no estacionarios debido a la presencia de condiciones iniciales. Si se toman datos de un proceso que ha comenzado recientemente la realización pudiera ser no estacionaria. En la práctica, usualmente se asume que el proceso estocástico subyacente ha estado ocurriendo por un infinito periodo de tiempo.

En la solución del proceso AR(1) se encuentran dos formas de **respuesta del proceso a los shocks**. Si  $|a_1| < 1$  (proceso estacionario) se ve en la expresión (17.4) que el efecto de los pasados  $u_t$  se vuelve sucesivamente menor en el tiempo; están multiplicados por  $a_1^i$  que cada vez se hace más pequeño. Si se reduce la magnitud de  $|a_1|$  aumenta la velocidad de convergencia.

Pero si  $|a_1| = 1$  (proceso no estacionario) entonces se observa en la expresión (17.3) que  $y_t$  contiene la suma de los  $u_t$ ; los shocks tienen un efecto permanente sobre el valor de  $y_t$ . Por ello, al estudiar la estabilidad y estacionariedad de una serie, estamos obteniendo información sobre el mecanismo de propagación o respuesta de la variable ante los shocks; cuestión que puede ser útil para el análisis económico y la formación de predicciones. En la figura 17.5 se simula el efecto que tiene un shock positivo sobre un proceso AR(1) estacionario y sobre un proceso AR(1) no estacionario.

**Figura 17.5: Respuesta a los shocks**



Fuente: Elaboración propia, series simuladas

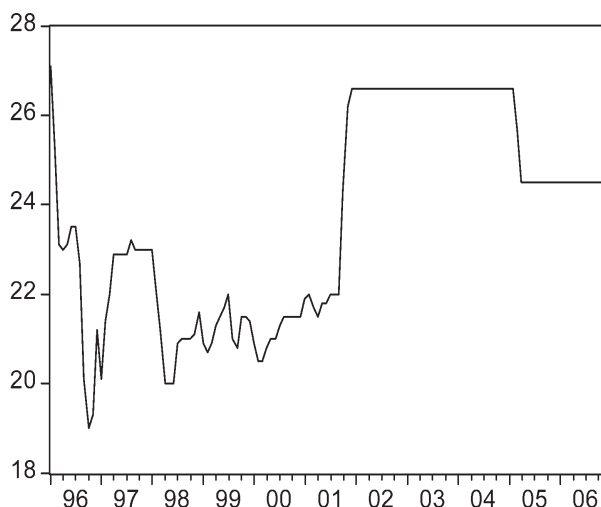


Sobre los procesos estacionarios los shocks tendrán un efecto transitorio, sobre los no estacionarios el efecto será permanente. Los procesos estacionarios se dice que tienen la propiedad de tener memoria limitada, el pasado reciente es más importante que el pasado lejano y los shocks tienden a perder efecto. En cambio, las series con una tendencia estocástica se dice que tienen memoria ilimitada, los valores observados hoy dependen mucho de las perturbaciones pasadas.

Los gráficos que hemos mostrado hasta el momento son de series de tiempo teóricas simuladas. En las figuras 17.6 y 17.7 se grafican dos series de tiempo de la economía cubana. En la figura 17.6 se tiene la evolución mensual del tipo de cambio en la red de casas de cambio (Cadeca). Hasta finales del año 2001 el tipo de cambio parecía ser estacionario, oscilaba alrededor de una media de 21 pesos cubanos por un peso convertible (CUC). Sin embargo, como resultado de los efectos en la economía cubana de los acontecimientos del 11 de septiembre en EE.UU. el tipo de cambio recibió un shock, el cual tuvo un efecto permanente sobre la variable, es decir, se propagó como si fuera un proceso no estacionario. El tipo de cambio promedio se devaluó a 26 pesos cubanos y no regresó a la media de 21 pesos cubanos. Lo cual es consistente con el hecho de que, aún cuando el régimen cambiario no ha sido completamente flexible, el Banco Central de Cuba no se ha comprometido con un tipo de cambio fijo o banda de flotación. Con este ejemplo queremos mostrar que las valoraciones económicas también deben formar parte del análisis que hacemos sobre la propiedad de estacionariedad de una serie de tiempo.<sup>6</sup>

**Figura 17. 6: Tipo de cambio de Cadeca**

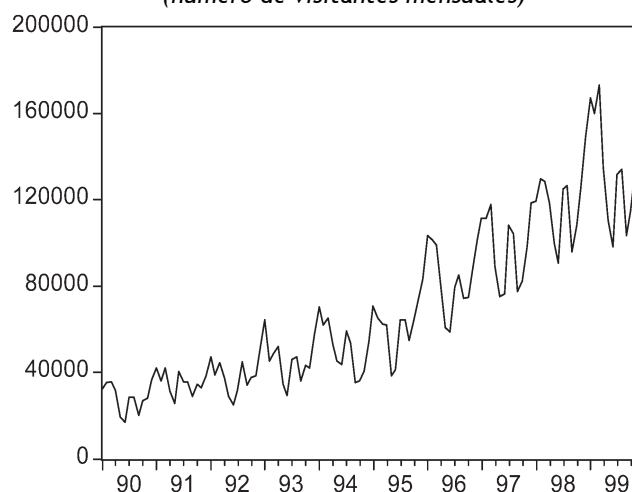
*(Pesos por un CUC. Promedio mensual entre el tipo de cambio de compra y de venta)*



*Fuente: Elaboración propia a partir de datos del Banco Central de Cuba.*

<sup>6</sup> Otro elemento a tomar en cuenta es que la valoración que hacemos sobre la no estacionariedad del tipo de cambio del peso cubano puede estar afectada por la presencia de condiciones iniciales, dado que es un proceso nuevo que se inicia con la creación de la institución financiera no bancaria Cadeca en 1994.

**Figura 17.7: Arribo de turistas a Cuba**  
(número de visitantes mensuales)



Fuente: Elaboración propia a partir de datos del Anuario Estadístico de la ONE

En la práctica, el examen gráfico de la realización de la variable es sólo un primer paso para evaluar la estacionariedad del proceso estocástico subyacente que la genera. Para tener un resultado más riguroso debemos recurrir a determinadas pruebas estadísticas, tal y como veremos en la sección 17.2.<sup>7</sup> El gráfico de la serie no siempre —o casi nunca— nos brinda una visión clara y definitiva sobre la estacionariedad. Por ejemplo, en la figura 17.7 se ilustra la evolución desde 1990 hasta 1999 de los datos mensuales del arribo de turistas a Cuba. Se destacan determinados “picos” y “valles” que se repiten todos los años, los cuales obedecen a la estacionalidad, componente que suele observarse en este tipo de variable y que marca las llamadas temporadas altas y bajas de la demanda de turismo. La estacionalidad de alguna manera no dejar observar con claridad si la variable tiene una trayectoria estacionaria. Lo que se observa es que la serie en el largo plazo siempre tiende a crecer, ha seguido un proceso evolutivo. Pero podría ser que la variable esté generada por un proceso estacionario alrededor de una tendencia o, en cambio, por un paseo aleatorio con una deriva positiva. Más adelante, aplicamos la prueba Dickey-Fuller a la serie del tipo de cambio y a la serie del arribo de turistas.

Llegado a este punto, podemos identificar las expresiones analíticas de los procesos que generaron las series estacionarias y no estacionarias ilustradas al inicio de la sección en las figuras 17.1, 17.2, 17.3 y 17.4:

<sup>7</sup> Algunos textos sugieren examinar también la estacionariedad con el correlograma muestral de la serie. En este libro no incorporamos dicho procedimiento, nos concentramos en la prueba de Dickey-Fuller.

Estacionario:  $y_t = a_0 + a_1 y_{t-1} + u_t$  con  $|a_1| < 1$  (17.5)

No estacionario en media:  $y_t = a_1 y_{t-1} + a_2 t + u_t$  con  $|a_1| < 1$  (17.6)

Paseo aleatorio:  $y_t = y_{t-1} + u_t$  (17.7)

Paseo aleatorio con deriva:  $y_t = a_0 + y_{t-1} + u_t$  (17.8)

En (17.6) tendríamos que  $a_2 > 0$  debido a que se graficó el proceso con una tendencia determinista positiva. En (17.8) el intercepto  $a_0$  hace la función de deriva, tal y como se aprecia en la solución de este proceso (ecuación 17.3). El proceso sigue una tendencia temporal que tiene asociada el coeficiente  $a_0$ , a la cual se le añade una sumatoria de shocks con un efecto permanente. El proceso que se ilustró en la figura 17.4 vendría generado por la expresión (17.8) con  $a_0 > 0$  dado que tiene una deriva positiva.

A continuación definimos las **condiciones de estabilidad y estacionariedad de procesos autorregresivos de mayor orden**. Un proceso estocástico  $AR(p)$  incluye la sumatoria de  $p$  términos autorregresivos:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + u_t \quad (17.9)$$

La condición de estabilidad de un proceso  $AR(p)$  se define en las raíces del polinomio autorregresivo. Este método también nos lleva a la condición de estabilidad de un  $AR(1)$  equivalente a la que obtuvimos en la solución encontrada por iteración. Para definir la condición de estabilidad solamente se necesita trabajar con la parte reducida (variable dependiente y sus rezagos) de la expresión (17.9):

$$y_t - \sum_{i=1}^p a_i y_{t-i} = 0 \quad (17.10)$$

Nos valemos del **operador de rezagos**  $L$ , del que haremos uso en múltiples ocasiones en el resto de las secciones de esta parte del libro. Para escribir términos autorregresivos con el operador de rezagos se define que  $y_{t-i} \equiv L^i y_t$ . La expresión (17.10) utilizando el operador de rezagos quedaría como sigue:

$$\begin{aligned} y_t - a_1 L y_t - a_2 L^2 y_t - \dots - a_p L^p y_t &= 0 \\ (1 - a_1 L - a_2 L^2 - \dots - a_p L^p) y_t &= 0 \end{aligned} \quad (17.11)$$

El operador de rezagos forma un polinomio. La expresión  $1 - a_1 Z - a_2 Z^2 - \dots - a_p Z^p$  se conoce como polinomio autorregresivo o **ecuación característica inversa**. Para saber si el proceso es estable se necesita calcular las **raíces** del polinomio autorregresivo. En este polinomio se conocen los valores de  $a_i$  y  $Z$  se trata como una incógnita. Los valores de  $Z$  que solucionan este polino-

mio se conocen como las raíces características inversas. El número de raíces depende del orden del polinomio. Las raíces pueden ser reales o complejas.

La condición de estabilidad de un proceso  $AR(p)$  requiere que estas raíces, si son reales, su módulo debe ser mayor que 1, y si son complejas deben estar fuera del círculo unitario. La estabilidad sigue siendo una condición necesaria para la estacionariedad; en definitiva, para que sea estacionario el proceso, deberá además  $t$  ser suficientemente grande. Si al menos una de las raíces de la ecuación característica inversa es igual a 1, el proceso no será estacionario. Por ello, en la literatura la no estacionariedad se identifica con la presencia de **raíces unitarias**.

El proceso autorregresivo de orden uno:

$$y_t = a_0 + a_1 y_{t-1} + u_t$$

expresado con el operador de rezagos quedaría como:

$$(1 - a_1 L)y_t = a_0 + u_t$$

La ecuación característica inversa sería  $(1 - a_1 Z)$ . Como es de primer orden habría una sola raíz, que se obtiene igualando a cero y despejando  $Z$ :

$$(1 - a_1 Z) = 0$$

$$a_1 Z = 1$$

$$Z = \frac{1}{a_1}$$

Se observa la equivalencia entre las dos condiciones de estabilidad formuladas. Para que un proceso  $AR(1)$  sea estacionario, habíamos inicialmente dicho que  $|a_1| < 1$ , si esto se cumple entonces  $|Z| > 1$ , es decir, el módulo de la raíz característica inversa será mayor a la unidad. Si hay una raíz unitaria  $|Z| = 1$  entonces coincide con que  $|a_1| = 1$  y el proceso es no estacionario.

Todo el espacio que le hemos venido dedicando a la estacionariedad no sólo es por el valor que tiene para el análisis económico, sino también por las implicaciones negativas que puede tener la inclusión de variables no estacionarias en un modelo de series de tiempo. La no estacionariedad afecta las propiedades de los estimadores MCO y los procedimientos convencionales de inferencia estadística.

El problema está en que la combinación lineal de series no estacionarias es generalmente no estacionaria; la excepción es lo que se denomina cointegración, tal y como veremos más adelante. Por lo que los errores de una regresión de variables no estacionarias serán casi siempre no estacionarios. El término de error no será un ruido blanco gaussiano. Si la media y la varianza de los errores dependen del tiempo, no se cumplirán los supuestos de media cero y varianza constante (homoscedasticidad) vistos en la sección 5.2 de la parte uno del libro. Con lo cual no se garantizan las propiedades de insesgamiento y mínima varianza de los esti-

madores MCO. También se estaría violando el supuesto de que los errores se distribuyen normalmente, por lo que los procedimientos de inferencia estadística como las pruebas  $t$  y  $F$  dejan de tener validez y pueden generar resultados engañosos.

Uno de los peligros que se corre al modelar series no estacionarias es la obtención de **regresiones espurias** (Granger y Newbold 1974). En la regresión se obtienen valores elevados para el  $R^2$  y para las razones  $t$ , que llevan a inferir una relación significativa entre las variables. Los resultados engañosos de los distintos estadísticos conducen a aceptar una relación entre las variables del modelo que puede ser que no exista en el proceso estocástico subyacente.

Por tanto, antes de comenzar la estimación de un modelo de series de tiempo, una de las primeras cosas que debe hacer el investigador es contrastar la presencia de raíces unitarias en las variables involucradas, tanto la dependiente como las variables independientes. En caso de que se encuentre que las variables son no estacionarias se deberán aplicar determinados procedimientos adecuados para estos casos. Sobre estos tópicos trata la siguiente sección.

## 17.2 Prueba de Raíz Unitaria

La prueba más usada para contrastar la presencia de raíces unitarias en procesos AR(1) es la propuesta por Dickey y Fuller (1979).<sup>8</sup> Posteriormente ha sido extendida para procesos AR de mayor orden y se conoce como la **prueba de Dickey Fuller aumentado (ADF)**.

Partiendo del modelo autorregresivo de orden uno:

$$y_t = a_0 + a_1 y_{t-1} + u_t$$

sabemos que el proceso será no estacionario si  $|a_1| = 1$  (presenta una raíz unitaria).

Si en cambio  $|a_1| < 1$  entonces el proceso AR(1) será estacionario. Para contrastar la estacionariedad con la prueba ADF se toman los datos de la realización de la variable y se estima una transformación del modelo (17.1). Se restan a ambos lados de la ecuación  $y_{t-1}$ , quedando el siguiente modelo:

$$\Delta y_t = a_0 + \delta y_{t-1} + u_t \quad (17.12)$$

donde  $\Delta y_t = y_t - y_{t-1}$  y  $\delta = a_1 - 1$ .

La hipótesis nula de la prueba plantea que  $\delta = 0$ , lo que implica que  $a_1 = 1$ : el proceso tiene una raíz unitaria y por tanto es no estacionario. La hipótesis alternativa plantea que  $\delta < 0$ , lo que implica que  $a_1 < 1$ : el proceso es estacionario. Una

<sup>8</sup> Existe otra variedad de pruebas de raíz unitaria que no explicaremos en este texto, por ejemplo, la prueba de Phillips-Perron (1998) y la prueba KPSS de Kwiatkowski et al (1992).

vez estimado el modelo (17.12) se calcula la razón  $t$  —se le llama **estadístico  $\tau$  ( $\tau$ )**— asociada al coeficiente  $\delta$ . Ello se hace de la manera habitual para un contraste de significación individual, dividiendo el coeficiente estimado entre su error estándar:  $\hat{\delta}/se(\delta)$ . La particularidad radica en que no se compara con los valores de la distribución  $t$ -Student, debido a que bajo la hipótesis nula de que  $\delta = 0$  la distribución del estadístico no sigue una  $t$ -Student. Los valores críticos con que se debe comparar el valor calculado del estadístico  $\tau$  fueron inicialmente computados por Dickey-Fuller y posteriormente extendidos para diferentes especificaciones de la prueba por McKinnon (1996).

Si los valores calculados del estadístico  $\tau$  se encuentran a la izquierda de los valores críticos en la distribución, entonces se rechaza la hipótesis nula y se infiere que el proceso es estacionario. Si este es el resultado, se dice que el proceso tiene **orden de integración** 0, es  $I(0)$ . Si no rechaza la hipótesis nula, entonces hay que repetir la prueba pero con la primera diferencia de la variable. Si se rechaza la hipótesis nula utilizando en la prueba la primera diferencia, entonces se dice que la variable tiene orden de integración 1, es  $I(1)$ . Si no se rechaza la hipótesis nula se deben continuar aplicando mayores diferencias hasta rechazar la hipótesis de no estacionariedad. En general, cualquier serie de tiempo no estacionaria puede transformarse en estacionaria diferenciándola  $d$  veces, siendo  $d$  el orden de integración de la variable. La mayoría de las series económicas son  $I(1)$  y rara vez se encuentran proceso con un orden de integración mayor a 2.

Hay autores que sugieren efectuar la prueba de Dickey-Fuller partiendo de comparar si la serie es  $I(1)$  contra la alternativa de que es  $I(2)$ . En caso de que no se rechace la hipótesis nula, se pasa a comparar  $I(0)$  contra la alternativa de que es  $I(1)$ .

Estas explicaciones nos llevan al procedimiento más utilizado para lidiar con la no estacionariedad de las series de tiempo: la **diferenciación**. En dependencia del orden de integración que tengan nuestras variables, así será el número de diferencias que tenemos aplicar para incluirlas en el modelo como series estacionarias. Utilizando el **operador de diferencias**  $\Delta$ , tenemos que una variable  $y_t$  con orden  $I(1)$  deberá incorporarse al modelo con la transformación  $\Delta(y_t) = y_t - y_{t-1}$ , que no es más que calcularle la variación. Una variable  $y_t$  con orden  $I(2)$  deberá incorporarse al modelo con la transformación  $\Delta^2(y_t) = \Delta\Delta(y_t) = \Delta(y_t - y_{t-1})$ , que no es más que calcularle la aceleración. En general, una variable  $y_t$  con orden  $I(d)$  deberá incorporarse al modelo con la transformación  $\Delta^d(y_t)$ . Los programas econométricos usualmente tienen comandos para aplicar el operador de diferencias. Las diferencias también pueden expresarse con el operador de rezagos  $L$ :  $\Delta^d(y_t) = (1 - L)^d y_t$ .

Al aplicar la primera diferencia a un paseo aleatorio ( $y_t = y_{t-1} + u_t$ ) se obtiene como resultado una serie ruido blanco:

$$y_t - y_{t-1} = u_t$$

Por construcción es estacionaria, dado que definimos invariantes en el tiempo los momentos de primero y segundo orden de  $u_t$ :  $E(u_t) = 0$ ,  $Var(u_t) = \sigma^2$  y  $Covar(u_t, u_{t-k}) = 0 \quad \forall k \geq 1$ .

De igual forma, diferenciando la expresión de un paseo aleatorio con deriva ( $y_t = a_0 + y_{t-1} + u_t$ ) también construimos una serie estacionaria:

$$y_t - y_{t-1} = a_0 + u_t$$

Obtenemos una serie que su media, varianza y covarianza no dependen del tiempo. Tiene todas las propiedades de un ruido blanco excepto que la media no es cero, sino  $a_0$ . En conclusión, los procesos paseo aleatorio y paseo aleatorio con deriva tienen orden de integración 1.

Como muchas series de tiempo no estacionarias son  $I(1)$ , en la mayoría de los modelos multivariantes se incluirían las variables con una primera diferencia y se estaría estimando cómo el crecimiento de las variables explicativas afecta el crecimiento de la variable dependiente. Cuando intervienen en el modelo variables con distintos órdenes de integración, y por tanto con diferencias de distintos órdenes, el modelo puede utilizarse para efectuar predicciones, sin embargo es algo confusa la interpretación económica de los parámetros estimados.

Para efectuar la prueba ADF se pueden incluir en el modelo (17.12) más rezagos de la variable dependiente para así contrastar la estacionariedad de procesos autorregresivos de mayor orden. Además, la prueba se puede realizar en el modelo sin constante o adicionando una tendencia determinista. Así, se estarían contrastando distintas formas de no estacionariedad, entre ellas el paseo aleatorio y el paseo aleatorio con deriva.

Las diferentes especificaciones del modelo (17.12) pueden hacer que cambien los resultados del contraste. Es conveniente probar varias especificaciones para estar más seguros de los resultados. Algunos paquetes estadísticos tienen incorporado la selección automática de rezagos a partir del criterio de información de Akaike u otro. En la selección de la constante y la tendencia pueden servir las pruebas  $t$  asociadas a cada uno de los términos; el modelo (17.12) no sería espurio debido a que tiene la variable dependiente diferenciada.<sup>9</sup>

Se le debe prestar atención a la especificación del modelo utilizada en la prueba ADF con el fin de reconocer bien el resultado que nos está ofreciendo la prueba y las medidas remediales que, en caso de ser necesario, debemos aplicar. Por ejemplo, si se le adiciona una tendencia al modelo (17.12) se estarían contrastando en la prueba ADF las siguientes hipótesis:

<sup>9</sup> Si la variable es  $I(2)$  o de un orden mayor entonces sí pudiera ser espurio el modelo con una sola diferencia en la variable dependiente. Varios autores recomiendan examinar los resultados de la prueba con las tres especificaciones: 1) con constante y tendencia, 2) sólo con constante y 3) sin constante y sin tendencia.

$H_0: \delta = 0$  en el modelo  $\Delta y_t = a_0 + \delta y_{t-1} + a_2 t + u_t \Rightarrow y_t$  no estacionario ni en media ni en varianza

$H_1: \delta < 0$  en el modelo  $\Delta y_t = a_0 + \delta y_{t-1} + a_2 t + u_t \Rightarrow y_t$  estacionario en varianza, no estacionario en media

En ambas hipótesis se está asumiendo que el proceso es no estacionario en media y lo que se está contrastando si es o no estacionario en varianza. Si se rechaza la hipótesis nula, la prueba no nos dice que el proceso es estacionario, sino que es estacionario en varianza, pero, al tener una tendencia  $t$ , es no estacionario en media. Como medida remedial para este tipo de no estacionariedad se propone remover el componente de tendencia de la variable: con restar del proceso  $a_2 t$  nos queda un proceso estacionario. Cuando la variable dependiente en el modelo es la que sigue un proceso de este tipo, la no estacionariedad queda resuelta con la inclusión de una tendencia temporal en la regresión como variable explicativa.

En la prueba ADF hay otros temas que de alguna manera se deben tener presentes, aunque en la práctica no siempre hay una solución a la mano para los mismos. Perron (1989) propone una prueba de raíz unitaria para cuando la variable ha sufrido cambios estructurales en el período muestral, ya sea un cambio en el nivel y/o en la pendiente. La prueba puede tener también dificultades para diferenciar entre procesos estacionarios y no estacionarios cuando las raíces del polinomio asociado al operador de rezago son cercanas pero menores a la unidad.

Cuando se trabaja con datos con una frecuencia mayor a la de un año la estacionalidad puede complicar el contraste sobre la estacionariedad. En el polinomio del operador de rezagos pueden aparecer raíces de módulo unidad debido a la estacionalidad, estas se conocen como raíces unitarias estacionales. La prueba ADF no contrasta estas raíces, la prueba sólo contrasta la presencia de raíces unitarias en el componente regular (frecuencia cero). Una alternativa seguida por algunos para contrastar la estacionariedad en la frecuencia cero, es desestacionalizar la serie y aplicar la prueba de Dickey- Fuller a la variable desestacionalizada. Sin embargo Ghysels (1990) demuestra que el ajuste estacional reduce la potencia de las pruebas de raíz unitaria debido al efecto de alisamiento que producen los filtros sobre las series. Existen pruebas que permiten contrastar la estacionariedad en la frecuencia cero y en las frecuencias estacionales al mismo tiempo. La más conocida es la desarrollada por Hylleberg, Engle, Granger y Yoo (1990) para datos trimestrales y extendida por Beaulieu y Miron (1992) para datos mensuales.<sup>10</sup>

A pesar de todos estos inconvenientes y pruebas alternativas, la prueba ADF se sigue usando en la mayoría de los trabajos aplicados de series de tiempo. En los ejemplos que mostramos a continuación también haremos uso de ella.

<sup>10</sup> Una aplicación de esta prueba a series de la economía cubana puede encontrarse en Vidal, Hernández, Pérez et al (2001).



### 17.2.1 Aplicación de la prueba ADF a series de tiempo de la economía cubana

Continuamos trabajando con los datos mensuales del tipo de cambio del peso cubano en Cadeca y del arribo de turistas. Empezamos por el tipo de cambio. Los datos utilizados están en la tabla B.9 en el CD acompañante, con el objetivo de que el lector pueda replicar la prueba en su computador.

Se hace uso del programa econométrico Eviews. Se decide la inclusión o no de constante y/o tendencia a partir de la prueba t-Student sobre los coeficientes asociados a cada término, empezando por el modelo más amplio que incluye ambos términos. Se seleccionan los rezagos por el criterio de Akaike de un máximo de 13 rezagos, dado que estamos trabajando con datos mensuales.

Primero aplicamos la prueba al subperíodo enero 1996 - junio 2001 en el cual parecía observarse una trayectoria estacionaria del tipo de cambio, en función de continuar con los comentarios que expresamos alrededor de la figura 17.6. La tendencia no resultó significativa por lo que efectuamos la prueba sólo con la constante. El menor valor del criterio de Akaike se obtiene con un solo rezago. Así, estaríamos contrastando la estacionariedad del tipo de cambio con la estimación del modelo:

$$\Delta y_t = a_0 + \delta y_{t-1} + \Delta y_{t-1} + u_t$$

Los resultados del contraste se resumen en la tabla 17.1. El valor calculado del estadístico  $\tau$  (valor de la razón  $t$  asociada al coeficiente  $\delta$ ) es menor que los valores críticos incluso al 1% de nivel de significación. Por tanto se rechaza la hipótesis nula de no estacionariedad. La variable tiene orden de integración  $I(0)$ .

**Tabla 17.1: Prueba ADF a la variable tipo de cambio de Cadeca, periodo enero 1996 - junio 2001**

|                                                          |                                      |
|----------------------------------------------------------|--------------------------------------|
| <i>Transformación de la variable</i>                     | Sin diferencia                       |
| <i>Inclusión de constante y/o tendencia</i>              | Constante                            |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -4.095                               |
| <i>Valores críticos</i>                                  | -3.536 (1%), -2.907(5%), -2.591(10%) |

*Fuente: Elaboración propia*

Posteriormente efectuamos la prueba con todo el período de datos disponibles, añadiendo los datos restantes hasta diciembre de 2006 (este ejemplo tiene la dificultad de que la serie tiene un solo mes con variación en el período restante). Se mantiene un rezago en la prueba, pero ahora la tendencia sí es significativa. Se estaría contrastando la estacionariedad con el modelo:

$$\Delta y_t = a_0 + \delta y_{t-1} + a_2 t + \Delta y_{t-1} + u_t$$

Los resultados del contraste se resumen en la tabla 17.2. El valor calculado del estadístico  $\tau$  no es menor que los valores críticos a ningún nivel de significación. Por tanto no se rechaza la hipótesis nula de no estacionariedad y debemos repetir la prueba con la primera diferencia de la variable. Con la primera diferencia se

mantiene un rezago según el criterio de Akaike, pero en esta ocasión no resultaron significativas ni la constante ni la tendencia. Las hipótesis nula y alternativa del contraste estarían ahora referidas al modelo:

$$\Delta^2 y_t = \delta y_{t-1} + \Delta^2 y_{t-1} + u_t$$

Recordar que aunque el modelo tiene dos diferencias, se está efectuando el contraste de la primera diferencia; el modelo para probar la estacionariedad del nivel de la variable ya tenía una diferencia (ver explicación que hicimos sobre la prueba ADF al inicio de la sección). En el segundo cuadro de la tabla 17.2 se resume el contraste con la primera diferencia del tipo de cambio. Ahora sí se rechaza la hipótesis nula, la variable tipo de cambio es estacionaria en primera diferencia, tiene orden de integración  $I(1)$ . Como no hay constante ni tendencia en la prueba podemos inferir que el tipo de cambio es generado por un proceso estocástico de paseo aleatorio. En definitiva, la prueba ADF confirma las apreciaciones que habíamos expresado sobre el gráfico del tipo de cambio del peso cubano y el régimen cambiario que ha adoptado el Banco Central de Cuba para esta moneda.

**Tabla 17.2: Prueba ADF a la variable tipo de cambio de Cadeca, período enero 1996 - diciembre 2006**

|                                                          |                                      |
|----------------------------------------------------------|--------------------------------------|
| <i>Transformación de la variable</i>                     | Sin diferencia                       |
| <i>Inclusión de constante y/o tendencia</i>              | Constante y tendencia                |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -2.914                               |
| <i>Valores críticos</i>                                  | -4.030 (1%), -3.445(5%), -3.147(10%) |
| <i>Transformación de la variable</i>                     | Primera diferencia                   |
| <i>Inclusión de constante y/o tendencia</i>              | Ninguna                              |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -8.261                               |
| <i>Valores críticos</i>                                  | -2.583 (1%), -1.943(5%), -1.615(10%) |

*Fuente: Elaboración propia*

Pasamos ahora al segundo ejemplo, la serie arribo de turistas. Los datos utilizados se encuentran en la tabla B.10 del CD acompañante.

En el ejemplo anterior del tipo de cambio se efectuaron pruebas con distintos períodos para corroborar las apreciaciones que habíamos hecho sobre su evolución en el gráfico, pero normalmente hacemos uso de toda la información disponible. En la tabla 17.3 se resume el contraste de raíz unitaria sobre la variable arribo de turistas a Cuba con toda la serie disponible. Se realiza primeramente la prueba a la variable sin diferenciar. Se seleccionan 13 rezagos por el criterio de Akaike de un máximo de 13 rezagos, dado que seguimos trabajando con datos mensuales. Se dejan en la prueba la constante y la tendencia, aunque la tendencia no es estrictamente significativa al 5%; la probabilidad de la prueba  $t$  es de 0.051. En cualquier caso, la exclusión de la tendencia y/o la constante no hacen que varíe el no rechazo de la hipótesis nula. Con las tres especificaciones el estadístico  $\tau$  indica que la variable no es estacionaria.

Se repite la prueba con la primera diferencia, incluyendo 12 rezagos por el criterio de Akaike y sólo la constante, dado que ahora sí la tendencia es altamente no significativa. El estadístico ADF nos conduce a rechazar la hipótesis nula al 10% de nivel de significación, pero no así al 5% o 1%. Hacemos el contraste entonces con la segunda diferencia, en esta ocasión el estadístico ADF por mucho rechaza la hipótesis nula de no estacionariedad (ver último cuadro de la tabla 17.3). La serie de tiempo arribo de turistas tiene orden de integración 2.

**Tabla 17.3: Prueba ADF a la variable arribo de turistas a Cuba, período enero 1990 - diciembre 1999**

|                                                          |                                      |
|----------------------------------------------------------|--------------------------------------|
| <i>Transformación de la variable</i>                     | Sin diferencia                       |
| <i>Inclusión de constante y/o tendencia</i>              | Constante y tendencia                |
| <i>Rezagos</i>                                           | 13                                   |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -1.754                               |
| <i>Valores críticos</i>                                  | -4.047 (1%), -3.453(5%), -3.152(10%) |
| <i>Transformación de la variable</i>                     | Primera diferencia                   |
| <i>Inclusión de constante y/o tendencia</i>              | constante                            |
| <i>Rezagos</i>                                           | 12                                   |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -2.722                               |
| <i>Valores críticos</i>                                  | -3.493 (1%), -2.889(5%), -2.581(10%) |
| <i>Transformación de la variable</i>                     | Segunda diferencia                   |
| <i>Inclusión de constante y/o tendencia</i>              | ninguna                              |
| <i>Rezagos</i>                                           | 11                                   |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -11.516                              |
| <i>Valores críticos</i>                                  | -2.587 (1%), -1.943(5%), -1.615(10%) |

Fuente: Elaboración propia





# *preguntas*

Trate de dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo

1. Defina la condición de estacionariedad de un proceso  $AR(p)$
2. ¿Un paseo aleatorio es un proceso estacionario?
3. ¿Qué es una prueba de raíz unitaria?
4. ¿Cómo se determina el orden de integración de una serie de tiempo?



# ejercicios

17.1 En la siguiente tabla mostramos el resultado de la prueba de Dickey-Fuller aplicado a la producción mercantil (PM) de la provincia de Villa Clara (se usó el programa Eviews).

Null Hypothesis: PM has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

|                                        | t-Statistic | Prob.* |
|----------------------------------------|-------------|--------|
| Augmented Dickey-Fuller test statistic | -0.360898   | 0.9078 |
| Test critical values: 1% level         | -3.565430   |        |
| 5% level                               | -2.919952   |        |
| 10% level                              | -2.597905   |        |

\*MacKinnon (1996) one-sided p-values.

- ¿La prueba indica que la serie de PM es estacionaria?
- ¿Cuál es el orden de integración de la serie?
- Escriba el modelo en que se apoya la prueba (nótese en la parte superior de la tabla que la prueba se aplicó sin tendencia y con cero rezagos adicionales)

17.2 En la siguiente tabla mostramos el resultado de la prueba de Dickey-Fuller aplicado a la serie trimestral de depósitos de ahorro (AHO) de la provincia de Santiago de Cuba (se usó el programa Eviews).

Null Hypothesis: D(AHO) has a unit root

Exogenous: None

Lag Length: 4 (Automatic based on SIC, MAXLAG=10)

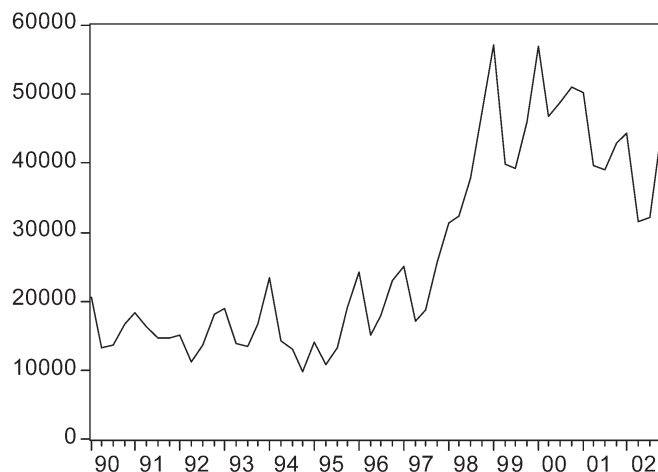
|                                        | t-Statistic | Prob.* |
|----------------------------------------|-------------|--------|
| Augmented Dickey-Fuller test statistic | -2.854497   | 0.0054 |
| Test critical values: 1% level         | -2.621185   |        |
| 5% level                               | -1.948886   |        |
| 10% level                              | -1.611932   |        |

\*MacKinnon (1996) one-sided p-values.

- ¿La prueba indica que la serie de ahorro es estacionaria?

- b. ¿Cuál es el orden de integración de la serie? (nótese en la parte superior de la tabla que la prueba se ha efectuado sobre  $D(AHO)$ )
- c. Escriba el modelo en que se apoya la prueba

17.3 En el gráfico se ilustra el comportamiento trimestral de la llegada de turistas alemanes a Cuba en el período 1990-2002.



- a. De acuerdo con el gráfico, ¿usted cree que la serie sea estacionaria?
- b. ¿Es posible identificar la ocurrencia de shocks en la serie? ¿En qué fechas?
- c. ¿Cómo fue la respuesta de la serie a los shocks identificados?
- d. Compruebe su interpretación gráfica sobre la estacionariedad, aplicando la prueba Dickey-Fuller a los datos trimestrales de la serie de turistas alemanes (ver datos en la tabla B.11 en el CD acompañante)

CAPÍTULO 18





## ANÁLISIS UNIVARIANTE DE SERIES DE TIEMPO

El análisis univariante de una serie de tiempo puede hacerse con dos objetivos fundamentales: la extracción de señales y la predicción. Inicialmente en esta sección mostramos la metodología de extracción de señales, la cual estudia los diferentes componentes que pueden describir a una serie de tiempo: tendencia, ciclo, estacionalidad e irregular. Se resumen los métodos principales que utiliza la investigación empírica aplicada para estimar estos componentes. A continuación, se pasa al análisis univariante apoyado en modelos ARIMA y que tiene como objetivo fundamental la predicción. En dicha sección se describe la metodología Box-Jenkins.

### 18.1 Componentes de las Series de Tiempo

El análisis univariante de una serie de tiempo puede hacerse desde la perspectiva del análisis clásico de series temporales o como también se conoce, la extracción de señales. Estos enfoques consideran a toda serie de tiempo integrada por cuatro componentes: tendencia, ciclo, estacionalidad e irregular. Dichos componentes caracterizan los distintos movimientos que puede presentar una serie de tiempo:

- **Tendencia:** Representa el movimiento general a largo plazo de una serie de tiempo. Constituye el elemento de crecimiento de la variable.
- **Variaciones cíclicas:** Son oscilaciones que se producen con un período superior al año, y que se deben principalmente a la alternancia de etapas de prosperidad y de depresión en la actividad económica.

- **Variaciones estacionales:** Son oscilaciones intraanuales que se repiten de manera muy similar en el mismo mes, trimestre, semestre, semana o incluso el día de cada año. Con frecuencia vienen causados por fenómenos no económicos, tales como variaciones climáticas y la regularidad de los feriados nacionales.
- **Variaciones irregulares:** También llamados residuos, son movimientos que no muestran un carácter periódico reconocible, y como tales se les considera originados por fenómenos singulares que afectan a la variable en estudio de manera más o menos casual.

No todos los componentes están presentes en todas las series de tiempo. De hecho, con una frecuencia anual en los datos el componente estacional no se observaría.

Existen dos esquemas que muestran la relación que guardan los componentes. El esquema aditivo, que se utiliza cuando los componentes son independientes entre sí o se presentan valores negativos o ceros en los datos:

$$X_t = T_t + C_t + E_t + I_t$$

donde  $X_t$  la serie de tiempo observada,  $T_t$  es la tendencia,  $C_t$  es el ciclo,  $E_t$  es el componente estacional e  $I_t$  es el irregular.

El esquema multiplicativo en cambio se utiliza cuando los componentes son dependientes entre sí o cuando el nivel de la serie es muy variable:<sup>1</sup>

$$X_t = T_t * C_t * E_t * I_t$$

En este esquema, por ejemplo, el tamaño de la estacionalidad depende del valor de los otros componentes. Por lo general, las series económicas son mejor descritas por el modelo multiplicativo.

Evidentemente, los componentes no son directamente observables, de ellos sólo podemos tener estimaciones. Originalmente el objetivo principal de la extracción de señales era eliminar las variaciones debidas a movimientos estacionales. Al eliminar la estacionalidad nos queda la **serie desestacionalizada** o ajustada de estacionalidad que denotamos con el superíndice *sa* por sus siglas en inglés:

$$X_t^{sa} = T_t * C_t * I_t$$

Si bien el desarrollo de la estadística y de la informática en las últimas décadas ha permitido aislar el resto de los componentes, muchas veces no se separa el componente cíclico de la tendencia. Entre otras razones porque en una serie se suelen superponer ciclos de diferente duración, porque se requiere de métodos estadísticos específicos para poder separarlos y porque se necesita de observacio-

<sup>1</sup> En la literatura muchas veces se habla de “componentes” cuando se usa el modelo aditivo y de “factores” cuando se usa el modelo multiplicativo. En este texto usamos indistintamente ambos términos.

nes en periodos extensos para poder identificar los ciclos más largos. Al combinar ambos componentes en uno solo queda lo que se denomina como **serie tendencia-ciclo** que denotamos con el superíndice  $tc$ :

$$X_t^{tc} = T_t * C_t$$

Podemos ver que la serie tendencia-ciclo se obtiene una vez eliminado el componente irregular de la serie ajustada de estacionalidad.

Para efectos del análisis económico, la estimación de los componentes no observables de una serie de tiempo cobra significativa relevancia. La **extracción de señales** es un instrumento que utilizan habitualmente las instituciones en los ejercicios de análisis de coyuntura.<sup>2</sup> El punto de partida de los procedimientos de extracción de señales es que el componente estacional y el componente irregular contienen las oscilaciones transitorias de la variable, mientras que el componente de tendencia, y también el ciclo, marcan un movimiento de más largo plazo que también se le conoce como evolución subyacente.

En este sentido, se considera que la información que proporciona la observación y medición de un fenómeno económico se encuentra “contaminada”. Por ello, es necesario “filtrar” los datos de modo de recuperar la “señal” que ellos contienen. Otro punto de vista del problema es considerar una serie de tiempo como señal más “ruido”, en donde el proceso de extracción de señales consiste en descomponer la serie en señal y ruido. Las series que ofrecen las señales que busca el investigador, o el hacedor de políticas, serían la serie ajustada de estacionalidad o la serie tendencia-ciclo. El ruido vendría dado entonces por la estacionalidad y las variaciones irregulares.

La extracción de señales contribuye a evaluar la situación presente de las variables y ayuda a comprender las predicciones. El reconocimiento de los movimientos estacionales e irregulares permite explicar si los cambios que se están observando o prediciendo en una variable obedecen efectivamente a aumentos o disminuciones permanentes en su nivel medio o bien a fenómenos transitorios. El diseño de políticas y la toma de decisiones requieren de estos análisis. Por ejemplo, un banco central sólo debería aumentar las tasas de interés cuando ocurren incrementos permanentes en la inflación, por ello, tales instituciones filtran los datos sobre la variación de los precios cada vez que se recibe una nueva observación. Debido al uso extendido de estas prácticas, muchas oficinas de estadísticas proveen ya las series desestacionalizadas y en algunos casos la tendencia-ciclo de los diferentes indicadores de la economía nacional.

<sup>2</sup> Sobre la extracción de señales en el análisis de coyuntura se puede consultar el libro de Espasa y Cancelo (1993).

### 18.1.1 Métodos de estimación de componentes inobservables

Se han propuesto distintos métodos para estimar los componentes inobservables de las series de tiempo. Los denominados métodos empiricistas y los métodos basados en modelos.

Los métodos empiricistas se caracterizan por implementar una descomposición basada en filtros lineales cuya estructura y parámetros no dependen de la naturaleza de los datos, sino que tienen valores prefijados. Los métodos empiricistas más representativos en economía son: las medias móviles, el alisado exponencial, el X-11 ARIMA (o su versión más reciente X-12 ARIMA), y el filtro de Hodrick-Prescott.

Respecto a los métodos de medias móviles y de alisado exponencial, la idea central es definir, a partir de la serie de tiempo observada, una nueva serie que suaviza los efectos ocasionados por la estacionalidad y el componente irregular, de manera que se pueda determinar la dirección de la tendencia.

El método de **media móvil** aplica un filtro lineal  $F$  a los valores de la serie para producir una nueva con características determinadas; no es más que una expresión lineal que transforma la serie observada  $X_t$  en una serie suavizada  $Z_t$ :  $Z_t = F(X_t)$ . Calcula para cada fecha un promedio utilizando valores a cada lado de esa fecha. El orden de la media móvil indica la cantidad de observaciones que toma para el cálculo del promedio en cada período. Una media móvil centrada de orden  $2n+1$  se puede representar como:

$$Z_t = m_{-n}X_{t-n} + \dots + m_{-1}X_{t-1} + m_0X_t + m_1X_{t+1} + \dots + m_nX_{t+n}$$

donde  $m_i$  son los parámetros de la media móvil.

Las medias móviles centradas tienen un número igual de observaciones después y antes del período especificado. La media móvil es simétrica si tiene las mismas ponderaciones. Por ejemplo, una media móvil centrada simétrica de orden tres con  $m=1/3$  se calcula como:

$$Z_t = (X_{t-1} + X_t + X_{t+1})/3 = \left( \frac{1}{3}X_{t-1} + \frac{1}{3}X_t + \frac{1}{3}X_{t+1} \right)$$

Cuando el orden de la media móvil es par, se toma una observación más para que sea centrada y se calcula la mitad de los valores extremos. Por ejemplo, una media móvil de orden cuatro se calcularía como:

$$\begin{aligned} Z_t &= \left( \frac{X_{t-2}}{2} + X_{t-1} + X_t + X_{t+1} + \frac{X_{t+2}}{2} \right) / 4 \\ &= \frac{1}{8}X_{t-2} + \frac{1}{4}X_{t-1} + \frac{1}{4}X_t + \frac{1}{4}X_{t+1} + \frac{1}{8}X_{t+2} \end{aligned}$$

Se plantea que el método de medias móviles para la estimación de la tendencia introduce en la serie un efecto conocido como de Slutsky-Yule. El efecto

Slutsky-Yule se refiere a la creación de nuevas oscilaciones en la serie suavizada después de aplicado el filtro. Para contrarrestar tal efecto, muchas veces se aplican sobre una misma serie de tiempo diferentes tipos de medias móviles que corrigen los ciclos espurios generados. También se definen medias móviles con parámetros que reducen el efecto Slutsky-Yule. Un ejemplo son las llamadas medias móviles de Spencer las cuales trabajan incluso con parámetros negativos. Más abajo se pueden ver los valores que toman los parámetros del filtro de una media móvil de Spencer de orden 15 para una serie de datos trimestrales:

$$F = 1/320 \text{ (-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3)}$$

Por su parte, el método de **alisado exponencial** construye la nueva serie a partir de la siguiente ecuación:

$$S_t = \beta S_{t-1} + (1 - \beta) X_t$$

donde  $S_t$  es el valor de la serie alisada en el momento  $t$ ,  $X_t$  es una observación de la variable en el momento  $t$  y  $0 < \beta < 1$  es la constante de alisado.

Observando la ecuación anterior, se puede derivar que cuando el valor de la constante de alisado  $\beta$  se acerca a 1, la serie alisada va a estar explicada en mayor medida por los valores precedentes de esta y en menor medida por los valores de la serie original. O sea que, será una serie más alisada de las oscilaciones que oscurecen la comprensión de los datos en la serie original. Si por el contrario, el valor de  $\beta$  se acerca a 0 los valores que tome la serie alisada dependerán más de los valores de la serie original y no habrá gran diferencia entre el comportamiento de ambas.

Para calcular  $S_t$  es necesario conocer el valor de  $\beta$  y  $S_0$ . Este último es el primer valor de la serie alisada, y puede estimarse como el promedio de un número reducido de observaciones iniciales. Por ejemplo:  $S_0 = (X_1 + X_2)/2$ . La elección de  $\beta$  se basa en criterios del investigador. La formulación presentada del método de alisado corresponde al alisado exponencial simple, el cual supone que la media de la serie de tiempo es constante, o lo que es lo mismo, que la serie de tiempo no tiene tendencia. Para variables que crecen o decrecen se utiliza el alisado exponencial doble que aplica dos veces a los datos la formulación del alisado simple:<sup>3</sup>

$$S_t = \beta S_{t-1} + (1 - \beta) X_t$$

$$D_t = \beta D_{t-1} + (1 - \beta) S_t$$

<sup>3</sup> Existen formulaciones más complejas de alisado exponencial conocidas con el nombre de Holt-Winters que también son usadas para efectuar predicciones.

donde  $D_t$  es el valor de la serie doblemente alisada en el momento  $t$ .

Si el objetivo que persigue el investigador es simplemente eliminar oscilaciones que oscurecen la lectura de los datos y observar a *grosso modo* la tendencia de la serie, tanto el método de medias móviles como el alisado exponencial son prácticos debido a su sencillez de aplicación. No obstante, existen paquetes de cómputo estadístico para realizar los cálculos con métodos más sofisticados como el X-11 ARIMA.

El **X-11 ARIMA** es el método más importante que se ha desarrollado en el campo del ajuste estacional. Realiza automáticamente la desestacionalización de la serie con diferentes filtros de medias móviles. Está ampliamente probado, ya que es utilizado por oficinas de estadísticas de todo el mundo. Se remonta a la década de los cincuenta, en el *Bureau of the Census* de los Estados Unidos, aunque la versión denominada X-11 de Shiskin, Young y Musgrave es de 1967. Posteriormente, son incorporados en el método —por Estela Bee Dagum del *Statistics Canada*— los modelos ARIMA para la proyección de la serie de tiempo a desestacionalizar; denominándolo finalmente X-11 ARIMA.

El método X-11 ARIMA consiste básicamente en:

- Estimar un modelo ARIMA de la serie de tiempo a desestacionalizar. Con este modelo se realiza la extrapolación un año hacia delante y hacia atrás de la serie. Ello ayuda a resolver el problema que tiene la media móvil para aplicar el filtro a los valores finales e iniciales del período muestral. El método además corrige aquellos puntos que pueden ser clasificados como observaciones atípicas.
- Aplica un proceso iterativo de filtrado mediante distintos tipos de medias móviles con el objetivo de obtener una estimación de los factores estacionales, para finalmente calcular la serie ajustada de estacionalidad.<sup>4</sup> Por ejemplo, si la serie que se analiza es mensual, el método crea una serie para cada uno de los doce meses y aplica sobre estas medias móviles centradas.

Los **factores estacionales** no son más que una estimación estocástica de la estacionalidad. El método ofrece un factor estacional diferente para cada mes dentro de cada año contenido en el período muestral. Ello es diferente a la estimación determinista de la estacionalidad mediante variables dummy en un modelo lineal de regresión (las variables dummy se expusieron en la sección 7.2). Con las variables dummy no se permite que la estacionalidad cambie en el tiempo, los coeficientes asociados a cada dummy ofrecen un único valor para el efecto estacional de cada mes.

Otro método empírico de extracción de señales es el **filtro de Hodrick-Prescott**. Se recurre a este procedimiento usualmente para separar la tendencia

<sup>4</sup> La versión más actual X-12 ARIMA ofrece, además de los factores estacionales y la serie ajustada de estacionalidad, una estimación del componente irregular y de la tendencia ciclo.

del ciclo. Permite estimar los componentes de tendencia y del ciclo a partir de la serie ajustada de estacionalidad. El filtro contiene un parámetro de alisado  $\lambda$ . Hodrick y Prescott calcularon para series trimestrales de los Estados Unidos un valor de  $\lambda=1.600$ . Se considera este un valor óptimo del parámetro de alisado para las series de ese país en específico ya que anula ciclos de una amplitud promedio de entre cuatro y seis años. Para el caso de una serie de tiempo anual, Hodrick y Prescott sugieren un valor de  $\lambda=100$ ; para una serie mensual sugieren  $\lambda=14.400$ . La arbitrariedad en la elección del parámetro  $\lambda$  es uno de los defectos de este método de estimación de componentes tendenciales y cíclicos. Cuando se aplica a países con un patrón cíclico diferente al de los Estados Unidos este método puede generar ciclos espurios.

En cuanto a los métodos basados en modelos, se dice que su planteamiento es más fiable que el de los empiricistas, dado que tienen en cuenta las características particulares de cada serie de tiempo estimando un modelo para cada variable. Entre los métodos disponibles se encuentran el procedimiento **TRAMO-SEATS**, que en sus siglas en inglés significa: Time Series Regression with ARIMA Noise, Missing Observations and Outliers—Signal Extraction in ARIMA Time Series.

El procedimiento TRAMO-SEATS desarrollado por Maravall en 1994 surge como un paso de avance en relación con el método X-11 ARIMA de desestacionalización<sup>5</sup>. El método se propone buscar una señal más firme que la serie desestacionalizada, una señal que fuese menos volátil y que sirviera mejor para la toma de decisiones. Propone entonces que sea utilizada la serie de tendencia-ciclo como una señal más estable, más suavizada que la serie ajustada de estacionalidad.

TRAMO-SEATS considera que los componentes se relacionan de forma multiplicativa, de manera que si se multiplica la tendencia-ciclo por los factores estacionales y por los factores irregulares, se obtiene la serie original:

$$X_t = TC_t * E_t * I_t$$

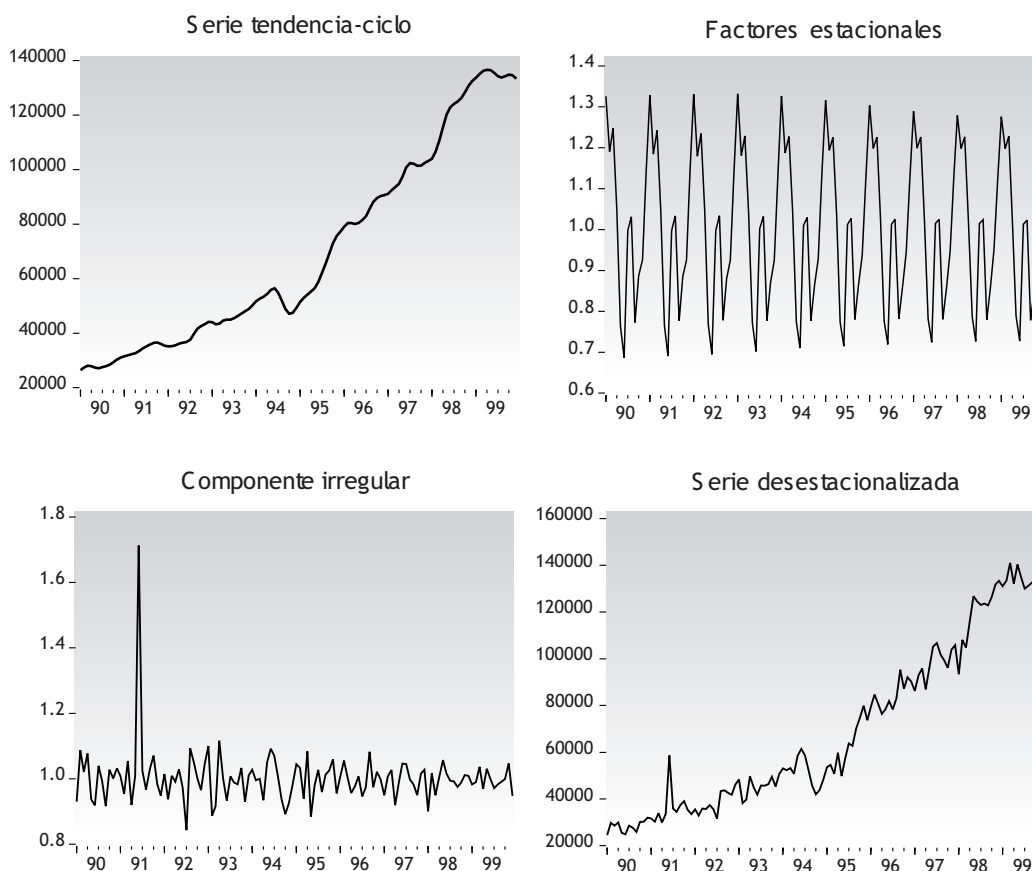
Para su aplicación existe el programa DEMETRA desarrollado en Europa por la agencia Eurostat. DEMETRA es un software bastante amigable que domina dos módulos.<sup>6</sup> Uno ofrece detalladamente el modelo ARIMA estimado y los efectos especiales que están presentes en la serie (efecto calendario y semana santa), además completa las observaciones faltantes y realiza análisis de intervención de las observaciones atípicas. El otro módulo realiza automáticamente la descomposición de la variable en tendencia-ciclo, factores estacionales y componente irregular. Los métodos empiricistas aplican un mismo filtro a todas las series, tengan o no estacionalidad, en cambio el TRAMO-SEATS aplica un filtro a la medida de los datos, teniendo como un resultado posible indicarle al investigador que su serie de tiempo no presenta estacionalidad.

<sup>5</sup> Ver Gómez y Maravall (1995)

<sup>6</sup> El programa DEMETRA también tiene el método X-12 ARIMA. El método TRAMO-SEATS, al igual que el X-11 ARIMA, X-12 ARIMA, el alisado exponencial y las medias móviles, igualmente puede encontrarse en otros paquetes estadísticos, por ejemplo, en las versiones 4 y 5 del programa Eviews.

A continuación ilustramos un ejemplo de la estimación de los componentes inobservables de la serie arribo de turistas a Cuba con el método TRAMO-SEATS, se emplean los datos de la tabla B.10 del CD acompañante.

**Figura 18.1: Componentes de la serie arribo de turistas estimados con TRAMO-SEATS**



*Fuente: Elaboración propia*

El lector debe recordar que la serie tendencia-ciclo multiplicada por los factores estacionales y por el componente irregular forman la serie original del arribo de turistas. La serie original también se forma multiplicando la serie ajustada de estacionalidad por los factores estacionales. Las series tendencia-ciclo y desestacionalizada se mantienen en la misma unidad de medida de la variable (número de visitantes). Los factores estacionales y el componente irregular se presentan como índices, revelando en cada período en qué porcentaje la variable se ubicó por encima o por debajo de la tendencia-ciclo. Por ejemplo, a mitad del año 1991 el componente irregular llegó a alcanzar el valor de 1.71, indicando que la variable se ubicó 71% por encima de la tendencia-ciclo debido a un factor irregular que afectó la variable en esa fecha, que dicho sea de paso, fue la celebración de los Juegos Panamericanos en La Habana. Como no fue un fenómeno que tuvo un im-



pacto permanente en la variable y como tampoco se debió a un evento estacional que posteriormente se repitió en la misma fecha todos los años, el método lo ubica en el componente irregular.

Los factores estacionales muestran una estacionalidad con picos a finales y principios de año y valle a mitad de años. Aunque no han ocurrido grandes cambios en la estacionalidad, sí se observa una reducción de los máximos y mínimos, se aprecia una ligera disminución de la diferencia entre la temporada alta y baja del turismo. Para más detalle, en la tabla 18.1 se exponen los factores estacionales estimados por el TRAMO-SEATS para el año 1990 y para el año 1999. Se ven los diferentes picos y valles de la demanda de turismo en el año. Los meses de mayor demanda de turismo son enero, marzo, febrero y diciembre, en este orden. El mes de menor demanda de turismo es junio, en 1990 se estima que se ubicaba 31% por debajo de la tendencia-ciclo y en 1999 se estima en 27%.

**Tabla 18.1: Factores estacionales de la serie arribo de turistas a Cuba estimados con el método TRAMO-SEATS**

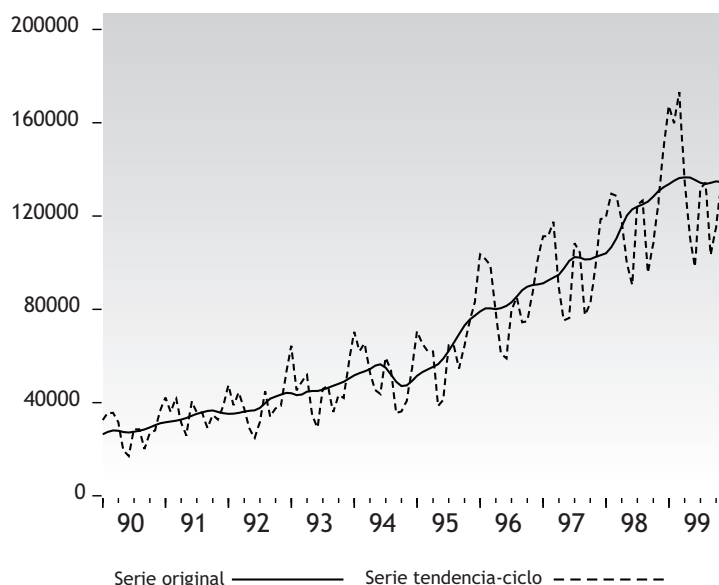
|            | 1990 | 1999 |
|------------|------|------|
| Enero      | 1.33 | 1.28 |
| Febrero    | 1.19 | 1.20 |
| Marzo      | 1.25 | 1.23 |
| Abril      | 1.05 | 1.02 |
| Mayo       | 0.76 | 0.79 |
| Junio      | 0.69 | 0.73 |
| Julio      | 1.00 | 1.01 |
| Agosto     | 1.03 | 1.02 |
| Septiembre | 0.77 | 0.78 |
| Octubre    | 0.89 | 0.86 |
| Noviembre  | 0.93 | 0.95 |
| Diciembre  | 1.14 | 1.12 |

*Fuente: Elaboración propia*

En la figura 18.1 se aprecia lo que habíamos comentado sobre la señal más estable en el movimiento subyacente de la variable que brinda la tendencia-ciclo, en comparación con la serie desestacionalizada. En la figura 18.2 se superponen la serie original con la tendencia-ciclo. Para los análisis de coyuntura que se puedan hacer sobre el turismo cubano, la serie tendencia-ciclo ofrece señales importantes sobre el comportamiento de la variable que no están visibles en la serie original. En la serie original, los componentes estacional e irregular no dejan ver bien la trayectoria de más largo plazo, cuesta trabajo identificar cuándo el turismo está realmente cayendo o cuándo empieza a recuperarse. En cambio, en la tendencia-ciclo queda más claro por ejemplo el efecto que realmente tuvo sobre el turismo

la “crisis de los balseros” en 1994, o el estancamiento que presentó el arribo de turistas en 1999.<sup>7</sup>

**Figura 18.2: Serie arribo de turistas a Cuba y su tendencia-ciclo, estimada con el método TRAMO-SEATS**



*Fuente: Elaboración propia*

Pasamos entonces al otro método de extracción de señales basado en modelos que expondremos en este texto y que tendría un carácter opcional en un curso de pregrado. Este es el método de los modelos estructurales univariantes de series temporales.

### **18.1.2 Modelos estructurales univariantes (opcional)**

Los **modelos estructurales univariantes de series temporales**, desarrollados fundamentalmente por Andrew Harvey desde la década de los ochenta, admiten la posibilidad de que cada uno de los componentes inobservables de una serie de tiempo posea una naturaleza estocástica. No modela los componentes de manera determinista, sino que el supuesto de partida es que evolucionan aleatoriamente en el tiempo. La mayor utilidad del método es que separa la tendencia del componente cíclico, permitiendo estimar más de un ciclo de diferentes períodos en la serie.<sup>8</sup>

<sup>7</sup> Estos análisis tienen un mayor valor cuando se realizan en el momento en que están ocurriendo los acontecimientos.

<sup>8</sup> Ver Harvey (1989) y Harvey y Koopman (1992)

El modelo estructural univariante no necesariamente debe ser definido en términos de todos los componentes inobservables. El ejemplo más sencillo de un modelo estructural, lo constituye una serie cuyas observaciones oscilan alrededor de un nivel medio que permanece constante en el tiempo. Si dichas oscilaciones son estacionarias y no están correlacionadas entre sí, una formulación capaz de recoger dicho comportamiento es:

$$y_t = \mu_0 + \varepsilon_t$$

donde  $\varepsilon_t$  es un proceso ruido blanco con varianza constante  $\sigma_\varepsilon^2$ . Se trata de un modelo con componente irregular  $\varepsilon_t$  y componente de nivel  $\mu_0$  fijo o determinista. La formulación anterior se puede flexibilizar para permitir que el nivel de la serie cambie en el tiempo, dando lugar a un modelo en el que el nivel en cada momento del tiempo es el nivel del período anterior más un elemento aleatorio:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t \end{aligned}$$

donde  $\eta_t$  es un ruido blanco con varianza  $\sigma_\eta^2$ . Este modelo sirve para representar el comportamiento de una serie de tiempo sin estacionalidad ni ciclos, cuyo nivel medio cambia en el tiempo (nivel estocástico) pero que no tiene una tendencia sistemática a crecer o a decrecer. Se conoce por el nombre de paseo aleatorio con ruido. Cuando  $\sigma_\eta^2=0$ , el nivel de la serie es determinista y se obtiene el modelo más sencillo antes especificado.

Ahora bien, al componente de nivel se le puede añadir una tendencia:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta + \eta_t \end{aligned}$$

donde  $\beta$  es una constante que mide la tasa de crecimiento promedio de la serie, es decir, la pendiente de la tendencia, es lo que se conoce como tasa de crecimiento subyacente. En este modelo el nivel cambia aleatoriamente en el tiempo pero la tasa de crecimiento promedio de la serie,  $\beta$  es constante. Si queremos flexibilizar la dinámica de la serie, permitiendo que dicha tasa pueda cambiar a lo largo del tiempo, podemos utilizar la siguiente especificación:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta + \eta_t \\ \beta_t &= \beta_{t-1} + \xi_t \end{aligned}$$

donde  $\xi_t$  es un ruido blanco con varianza  $\sigma_\xi^2$ . El término de perturbación  $\xi_t$  otorga a la pendiente un carácter estocástico. Este es conocido como modelo de tendencia lineal local y representa el comportamiento de una serie de tiempo sin estacionalidad ni ciclos, cuya tasa de crecimiento cambia a lo largo del tiempo. Si  $\sigma_\xi^2=0$ , se retorna al modelo anterior, es decir, la serie tiene pendiente pero con una tasa de crecimiento medio constante a lo largo del tiempo. Es posible también

mantener el carácter estocástico de la pendiente ( $\sigma_{\xi}^2 \neq 0$ ) y al mismo tiempo suponer que  $\sigma_{\eta}^2 = 0$ , o sea, el nivel no es estocástico. Abajo se muestra la formulación de este caso particular. Tal modelo de nivel fijo y pendiente estocástica se denomina de tendencia alisada:

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_t \\ \beta_t &= \beta_{t-1} + \xi_t\end{aligned}$$

Si las observaciones de la serie en cuestión se registran con periodicidad inferior a la anual, suelen aparecer efectos estacionales y, por ello, conviene introducir este componente en el modelo. Uno apropiado sería el siguiente:

$$\begin{aligned}y_t &= \mu_t + \gamma_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_t + \eta_t \\ \beta_t &= \beta_{t-1} + \xi_t\end{aligned}$$

donde  $\gamma_t$  es el componente estacional. Si las fluctuaciones estacionales son estables, este componente se puede considerar determinista o fijo. Se recomienda modelar la estacionalidad de manera determinista cuando el número de años para los que se dispone de información es pequeño y no haya suficientes observaciones que permitan la consideración de un patrón estacional cambiante. Pero en caso contrario, se puede modelar que el componente estacional es estocástico y evoluciona a lo largo del tiempo. La estacionalidad basada en dummy estacionales es una forma de considerar que el patrón estacional evoluciona aleatoriamente en el tiempo si se admite que la suma de los efectos estacionales a lo largo del año no es estrictamente nula. Se define que la estacionalidad no es exactamente igual de año en año, sino igual a un término de perturbación aleatoria ( $\omega_t$ ). Es decir, los valores del componente estacional en los instantes del tiempo  $t$ ,  $t-1, \dots$ ,  $t-(s-1)$  están ligados por la relación:

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$$

donde  $\omega_t$  es ruido blanco con varianza  $\sigma_{\omega}^2$ ,  $s$  es la frecuencia de la estacionalidad ( $s=12$  para series mensuales). Cuando  $\sigma_{\omega}^2 = 0$  se obtiene una estacionalidad determinista.

Como habíamos dicho, el modelo estructural univariante permite separar la tendencia del ciclo, quedando la representación más general de un modelo estructural con todos los componentes:

$$\begin{aligned}y_t &= \mu_t + \psi_t + \gamma_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_t + \eta_t \\ \beta_t &= \beta_{t-1} + \xi_t\end{aligned}$$

$$\sum_{j=0}^{s-1} \gamma = \omega_t$$

donde  $\psi_t$  representa el componente cíclico.

Para representar los ciclos se manejan las funciones seno y coseno, dado que son funciones periódicas. Para realizar las estimaciones econométricas de los modelos estructurales existe el programa STAMP (*Structural Time Series Analyser, Modeller and Predictor*). El programa STAMP utiliza la siguiente formulación para lograr una representación estocástica del ciclo:

$$\begin{aligned}\psi_t &= \cos \lambda_c \psi_{t-1} + \text{sen} \lambda_c \psi_{t-1}^* + \kappa_t \\ \psi_t^* &= -\text{sen} \lambda_c \psi_{t-1} + \cos \lambda_c \psi_{t-1}^* + \kappa_t^*\end{aligned}$$

donde  $\kappa_t$  y  $\kappa_t^*$  son ruidos blancos no correlacionados entre sí ni con ninguna de las otras perturbaciones del modelo y con varianza común  $\sigma_\kappa^2$ ; el parámetro  $\lambda_c$  es la frecuencia medida en radianes, es decir, representa el número de veces que el ciclo se repite en un período de tiempo de longitud  $2\pi$ . El período del ciclo es  $2\pi/\lambda_c$ . Por otra parte, para recoger los comportamientos de ciclos falsos que caracterizan a muchas series económicas, se incluye el factor de amortiguación  $\rho$  que aporta mayor flexibilidad al ciclo:

$$\begin{aligned}\psi_t &= \rho \cos \lambda_c \psi_{t-1} + \rho \text{sen} \lambda_c \psi_{t-1}^* + \kappa_t \\ \psi_t^* &= -\rho \text{sen} \lambda_c \psi_{t-1} + \rho \cos \lambda_c \psi_{t-1}^* + \kappa_t^*\end{aligned}$$

En la formulación anterior si  $|\rho| < 1$  el ciclo es estacionario.

En la terminología de los modelos estructurales univariantes las varianzas (o desviación estándar) de las perturbaciones aleatorias que afectan a los distintos componentes se denominan **hiperparámetros**. Cuando estimamos un modelo estructural univariante los hiperparámetros constituyen un resultado esencial que debemos examinar para entender la especificación que se nos propone para los datos disponibles. Hemos ido apuntando en las diferentes especificaciones, que cuando el hiperparámetro es diferente de cero el componente asociado es estocástico, en caso contrario, el componente es determinista.

En la tabla 18.2 exponemos las estimaciones que hicieron Vidal y Fundora (2004) sobre los componentes del PIB cubano con el programa STAMP. Se empleó la transformación logarítmica de la serie anual del PIB a los precios del año base 1981, comprendiendo el período 1903-2003.

Como esta variable es de frecuencia anual no se realizó la estimación de la estacionalidad. Fijándonos en las estimaciones de los hiperparámetros (desviaciones estándar de las perturbaciones de los componentes), podemos darnos cuenta de que la tendencia del PIB está formada por una pendiente determinista y un nivel estocástico, debido a que en el primer caso es cero y no así en el segundo. Por otro lado, el componente irregular no está presente en la serie; la estimación del hiperparámetro correspondiente a este es cero. Ello significa que la serie del PIB

se puede explicar por el componente de tendencia y por el componente cíclico. Se estimaron dos ciclos, uno corto de 7.6 años y otro largo de 38.4 años, teniendo ambos un comportamiento estocástico y estacionario, según lo demuestran los valores que toman los hiperparámetros de los dos ciclos (son distintas de cero) y el coeficiente  $\rho$  (es menor que 1). Los resultados de estos hiperparámetros muestran que el ciclo largo presenta mayor variabilidad que el ciclo corto.

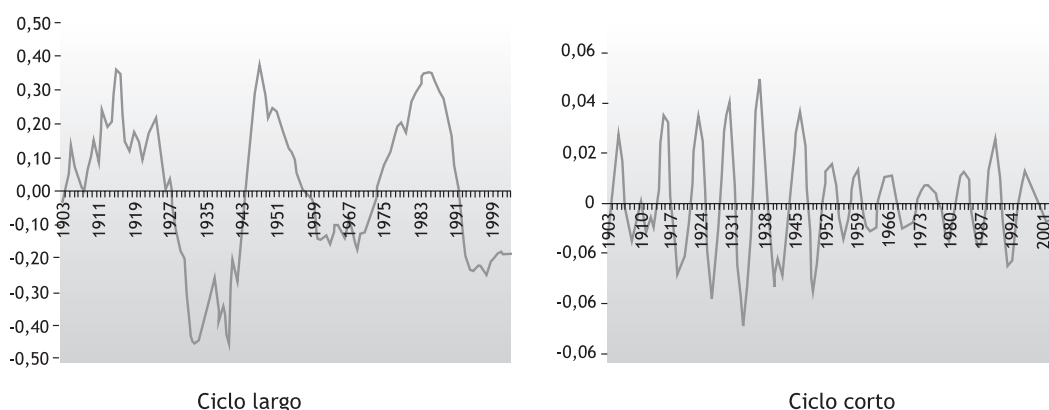
**Tabla 18.2: Estimación de un modelo estructural univariante para el PIB cubano, período 1903-2003**

|                                  | Estimaciones |
|----------------------------------|--------------|
| $\sigma_\varepsilon$ (irregular) | 0.000        |
| $\sigma_\eta$ (nivel)            | 0.044        |
| $\sigma_\xi$ (pendiente)         | 0.000        |
| Ciclo corto (años)               | 7.6          |
| Ciclo largo (años)               | 38.4         |
| $\sigma_\kappa$ (ciclo corto)    | 0.018        |
| $\sigma_\kappa$ (ciclo largo)    | 0.077        |
| $\rho$ (ciclo corto)             | 0.87         |
| $\rho$ (ciclo largo)             | 0.93         |

Fuente: Vidal y Fundora (2004)

En la figura 18.3 se grafican los ciclos estimados. Se puede observar que el patrón cíclico es irregular, la evolución y duración de un ciclo no es exactamente igual a otro; esto es porque la estimación de estos es estocástica. Las condiciones que generaron una onda cíclica en un período determinado no son las mismas para todos los ciclos. Ello es más realista que esperar un patrón cíclico determinista, sin diferencias entre los ciclos.

**Figura 18.3: Ciclos del PIB de Cuba estimados con el programa STAMP, período 1903-2003**



Fuente: Vidal y Fundora (2004)

## 18.2 Metodología Box-Jenkins

Los métodos de estimación univariante que hemos visto hasta el momento pueden utilizarse para conformar predicciones de la serie estudiada y de sus componentes. Sin embargo, el mayor uso en la investigación empírica ha estado dirigido hacia la extracción de señales. En esta sección estaremos más enfocados en la predicción al analizar la metodología Box-Jenkins de estimación de modelos ARIMA.

Los **modelos ARIMA** comenzaron a usarse a partir de que Box y Jenkins en 1976 publicaran el libro titulado *Time Series Analysis, Forecasting and Control* con un análisis integral de esta metodología. Los modelos ARIMA intentan describir lo mejor posible la naturaleza del proceso estocástico que genera la serie temporal. Para ello, no incluyen variables económicas independientes, utilizan únicamente la información contenida en la serie de tiempo que se quiere predecir. Son modelos que no tienen un valor teórico muy revelante. Su estimación no requiere del conocimiento del investigador de un modelo teórico subyacente ni de la imposición de supuestos económicos. Explican las variaciones de la variable dependiente a partir de sus propios rezagos y de una combinación lineal del término de error. Parten de la filosofía de “...dejar que los datos hablen por sí mismos”.

En este sentido, se plantea que son modelos “baratos”: requieren de poca información, la estimación de los mismos consume relativamente poco tiempo al analista y la metodología es fácil de implementar en los diferentes paquetes estadísticos actualmente disponibles. Por ello, se utilizan como “*benchmark*” en la elaboración de predicciones; se justifica construir modelos más complicados si estos producen menores errores de predicción que el modelo univariante ARIMA.

Con los modelos ARIMA se ha probado que se obtienen buenas predicciones de corto plazo. El modelo recoge un grupo de regularidades que se han observado en el pasado en la trayectoria de la variable y las extrapola hacia el futuro. Unos pocos períodos hacia adelante tales características suelen repetirse y no sufrir grandes cambios, por ejemplo, uno o dos trimestres. Los modelos ARIMA pueden construirse de manera que sean flexibles y otorguen un mayor peso a las experiencias recientes, de manera que las predicciones se adapten con rapidez a los últimos acontecimientos. Las desventajas de los modelos ARIMA evidentemente estarían en que brindan mucha menos información que los modelos multivariantes, que para períodos más extensos las predicciones podrían acumular grandes errores y que no permiten simular el efecto que tendría sobre la variable dependiente los cambios en otras variables de la economía.

### 18.2.1 Formulación de los modelos ARIMA

Las siglas ARIMA tienen el siguiente significado:

**AR:** Representa los términos autorregresivos que se incluyen en el modelo. Como ya habíamos explicado en el Capítulo 14, una serie de tiempo puede ser generada por procesos autorregresivos de diferentes órdenes:

$$\text{AR}(1): y_t = \alpha_1 y_{t-1} + \mu_t$$

$$\text{AR}(2): y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \mu_t$$

$$\text{AR}(p): y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \mu_t$$

donde  $\mu_t$  representa una serie de errores estocásticos ruido blanco.

**MA:** Representa el promedio móvil de los términos de error contemporáneos y rezagados que se incluyen en el modelo, los cuales se nombran como términos de medias móviles.<sup>9</sup> Se incluyen términos de medias móviles de diferentes órdenes:

$$\text{MA}(1): y_t = \mu_t - \beta_1 \mu_{t-1}$$

$$\text{MA}(2): y_t = \mu_t - \beta_1 \mu_{t-1} - \beta_2 \mu_{t-2}$$

$$\text{MA}(q): y_t = \mu_t - \beta_1 \mu_{t-1} - \beta_2 \mu_{t-2} - \cdots - \beta_q \mu_{t-q}$$

donde  $\mu_t$  representa una serie de errores estocásticos ruido blanco.<sup>10</sup> El proceso  $y_t$  es un ruido blanco; si bien  $\mu_t$  es un ruido blanco, su combinación lineal no lo es.

**I:** Se refiere al orden de integración de la variable según el resultado que se obtuvo en el contraste de la estacionariedad. Indica que en el modelo ARIMA la variable debe incluirse con el número de diferencias requerido para que sea estacionaria acorde con su orden de integración. Si vamos a modelar un proceso estacionario, entonces le llamaríamos simplemente modelo ARMA. Así, el modelo ARIMA o ARMA debe tener incluida la información referida a la estacionariedad de la variable, cuestión esta que utilizará el modelo a la hora de computar las predicciones. El contraste de estacionariedad de la variable a predecir se encuentra dentro del primer paso que debe seguir el investigador en el proceso de estimación de un modelo ARIMA.

Se puede modelar una serie de tiempo simplemente mediante un proceso puramente AR o mediante un proceso puramente MA. Sin embargo, pueden combinarse términos autorregresivos con términos de medias móviles para formar un modelo ARMA( $p, q$ ):

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \mu_t - \beta_1 \mu_{t-1} - \beta_2 \mu_{t-2} - \cdots - \beta_q \mu_{t-q}$$

<sup>9</sup> Aunque tienen el mismo nombre, no se refiere al filtro de medias móviles que estudiamos como método de extracción de señales.

<sup>10</sup> Por convención y para utilizar la notación más habitual en los textos de series de tiempo, se plantean los términos autorregresivos con signos positivos y los de medias móviles con signo negativo. No obstante, en la práctica, los coeficientes estimados tienen cualquier signo.



donde  $p$  indica el orden de los términos autorregresivos y  $q$  el orden de los términos de medias móviles.  $\delta$  es el intercepto del modelo, el cual también puede estar presente en modelos puros AR( $p$ ) o MA( $q$ ).

Otra notación equivalente para representar un proceso ARMA( $p, q$ ) se hace con el operador de rezagos  $L$  (recordar que  $L^i y_t \equiv y_{t-i}$ ):

$$A(L)y_t = \delta + B(L)u_t \quad (18.1)$$

donde  $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$  y  $B(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$  y sus respectivas ecuaciones características inversas son:  $A(Z) = 1 - \alpha_1 Z - \alpha_2 Z^2 - \dots - \alpha_p Z^p$  y  $B(Z) = 1 - \beta_1 Z - \beta_2 Z^2 - \dots - \beta_q Z^q$  y  $Z$  las raíces correspondientes (ver Capítulo 17).

Si el proceso es no estacionario con orden de integración  $d$ , se estaría estimando un modelo ARIMA( $p, d, q$ ):

$$A(L)\Delta^d y_t = \delta + B(L)u_t$$

o lo que es lo mismo,

$$A(L)(1-L)^d y_t = \delta + B(L)u_t$$

Partiendo de la formulación (18.1) es fácil ver que, en principio, existen formas equivalentes de representar un proceso estocástico con los términos AR y MA. Un modelo AR puede transformarse en un modelo puro de medias móviles:

$$\text{AR:} \quad A(L)y_t = \delta + u_t$$

$$\text{MA:} \quad y_t = (\delta + u_t) / A(L)$$

A su vez, un proceso MA puede transformarse en un proceso puro autorregresivo:

$$\text{MA:} \quad y_t = \delta + B(L)u_t$$

$$\text{AR:} \quad (y_t - \delta) / B(L) = u_t$$

De la misma manera, todo proceso ARMA tiene una representación equivalente como un proceso puro AR o MA. Tales equivalencias las habíamos practicado en el Capítulo 17 en las ecuaciones en diferencia que forman los procesos autorregresivos y en sus soluciones formadas por la suma de errores estocásticos. Todo esto nos da paso a definir dos de las condiciones más importantes que deben cumplir los modelos ARIMA para que pueda efectuarse su estimación y predicción: los modelos ARIMA deben ser **estacionarios** e **invertibles**. Dichas condiciones también se requieren para computar las equivalencias que acabamos de presentar entre modelos AR, MA y ARMA.

La estacionariedad en los modelos ARIMA garantiza la constancia en el tiempo de los momentos de primero y segundo (media, varianza y covarianza), permitiendo así la estimación y predicción de la variable en estudio. Posibilita definir una

distribución de probabilidades que mediante sus momentos de primero y segundo orden caractericen el proceso estocástico subyacente.

Los modelos ARIMA pueden verse como una ecuación en diferencias donde la solución será un proceso MA. La condición de estacionariedad se encontrará en la parte reducida de la ecuación en diferencias, es decir, en la parte autorregresiva del modelo. Aplicando lo que aprendimos en la sección 17.1, decimos que un modelo ARIMA (modelo 18.1) cumple la condición de estabilidad y estacionariedad (suponiendo  $t$  suficientemente grande) si el módulo de las raíces del polinomio  $A(Z)$  es mayor a uno. Si en el polinomio  $A(Z)$  hay raíces unitarias el modelo es no estacionario; se deberá entonces diferenciar la variable hasta que se transforme en estacionaria.

Todo proceso  $AR(p)$  o  $ARMA(p,q)$  puede transformarse en un proceso  $MA(\infty)$  estable, siempre y cuando las raíces del polinomio autorregresivo tengan un módulo mayor a la unidad. Para ilustrarlo nos valemos de un proceso  $AR(1)$  y de las siguientes propiedades del operador de rezagos: 1)  $Lc = c$ , siendo  $c$  una constante y 2) cuando  $|\alpha| < 1$ , la suma infinita  $(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 \dots)y_t = y_t / (1 - \alpha L)$ .

En un proceso  $AR(1)$   $y_t = \delta + \alpha_1 y_{t-1} + u_t$ , expresado con el operador de rezagos  $(1 - \alpha_1 L)y_t = \delta + u_t$ , hallamos su solución despejando  $(1 - \alpha_1 L)$  de manera tal que no aparezcan en la expresión rezagos de  $y_t$ :

$$y_t = \delta / (1 - \alpha_1 L) + u_t / (1 - \alpha_1 L)$$

Con las propiedades mencionadas del operador de rezagos, y recordando que  $\delta \sum_{i=1}^t \alpha_1^i$  converge a  $\delta / (1 - \alpha_1)$ , el proceso  $AR(1)$  puede ser expresado como un proceso  $MA(\infty)$  estable:

$$y_t = (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 \dots)\delta + (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 \dots)u_t$$

$$y_t = (\delta + \alpha_1 \delta + \alpha_1^2 \delta + \alpha_1^3 \delta \dots) + (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 \dots)u_t$$

$$y_t = \delta / (1 - \alpha_1) + (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 \dots)u_t$$

$$y_t = \delta / (1 - \alpha_1) + \sum_{i=0}^{\infty} \alpha_1^i u_{t-i}$$

Nótese que conseguimos transformar el proceso  $AR(1)$  en un  $MA(\infty)$  utilizando las propiedades del operador de rezagos si  $|\alpha_1| < 1$ , y para que ello se cumpla sabemos que debe cumplirse que  $|Z| > 1$ , es decir, el módulo de la raíz característica inversa tiene que ser mayor a la unidad. Si hay una raíz unitaria el proceso  $AR(1)$  no puede transformarse en un proceso  $MA(\infty)$  estable. La expresión a la que llegamos valiéndonos de las propiedades del operador de rezagos es la misma que la solución que en el Capítulo 17 habíamos formulado en la ecuación (17.4).

En cuanto a la condición de invertibilidad de los modelos ARMA se dice que un proceso es invertible cuando puede ser representado por un proceso autorregresivo estable. La invertibilidad garantiza que la serie puede expresarse en función de su pasado. La necesidad de invertibilidad pone de manifiesto el interés de realizar predicciones futuras a partir de los valores pasados. La metodología Box-Jenkins para estimar modelos ARIMA supone que el proceso es invertible.

Se define que un modelo ARIMA es invertible cuando el módulo de las raíces del polinomio que forma el operador de rezagos en la parte de medias móviles son mayores que la unidad. En (18.1) se dirá que el modelo es invertible si el módulo de las raíces (valores de  $Z$ ) que solucionan  $B(Z) = 0$  son mayores a la unidad.

Todo proceso  $MA(q)$  o  $ARMA(p,q)$  puede transformarse en un proceso  $AR(\infty)$  estable, siempre y cuando el módulo de las raíces del polinomio de medias móviles sean mayores a la unidad. Para ilustrarlo nos valemos de un proceso  $MA(1)$  y de las propiedades del operador de rezagos  $L$  anteriormente mencionadas. Un proceso  $MA(1)$   $y_t = \delta + u_t - \beta_1 u_{t-1}$ , expresado con el operador de rezagos  $y_t = \delta + (1 - \beta_1 L)u_t$ , podemos transformarlo en un proceso AR despejando  $B(L)$ :

$$y_t / (1 - \beta_1 L) = \delta / (1 - \beta_1 L) + u_t$$

Con las propiedades del operador de rezagos, y recordando que  $\delta \sum_{i=1}^t \beta_1^i$  converge a  $\delta / (1 - \beta_1)$ , el proceso  $MA(1)$  puede ser expresado como un proceso  $AR(\infty)$  estable:

$$(1 + \beta_1 L + \beta_1^2 L^2 + \beta_1^3 L^3 \dots) y_t = \delta / (1 - \beta_1) + u_t$$

$$y_t = \delta / (1 - \beta_1) - \sum_{i=1}^{\infty} \beta_1^i y_{t-i} + u_t \quad (18.2)$$

Nótese que podemos transformar el proceso  $MA(1)$  en un  $AR(\infty)$  utilizando las propiedades del operador de rezagos si  $|\beta_1| < 1$ , y para que ello se cumpla debe cumplirse que  $|Z| > 1$  en  $B(Z)$ , es decir, el módulo de la raíz en el polinomio de medias móviles tiene que ser mayor a la unidad. Si  $|Z| \leq 1$  en  $B(Z)$ , entonces  $|\beta_1| \geq 1$ , como consecuencia el modelo no será invertible y por tanto no puede ser estimado con la metodología Box-Jenkins.

Como hemos visto la condición de estacionariedad está referida a la parte autorregresiva del modelo ARIMA y la invertibilidad a la parte de medias móviles. Por definición todo modelo  $AR(p)$  puro es invertible y todo modelo  $MA(q)$  puro es estacionario. Lo que habría que verificar en un proceso  $AR(p)$  puro es si es estacionario, así como tendríamos que verificar que un proceso  $MA(q)$  puro es invertible.

Pasando a otro análisis de la formulación de los modelos ARIMA, en la sección 17.1 veíamos que en la solución de la ecuación en diferencia se encontraba de

manera explícita la respuesta del proceso a los shocks. Habíamos visto que si el proceso era estacionario el efecto de los shocks se volvía sucesivamente menor en el tiempo, pero si era no estacionario el efecto era permanente. Esta información la incorporamos al modelo ARIMA con el orden de las diferencias con que se incluye a la variable. Ahora bien, al estimar el modelo ARIMA obtendremos una mayor cantidad de información sobre la forma de la respuesta del proceso a los shocks. Los términos AR y MA y los coeficientes asociados que se estiman definen una **función impulso respuesta** de la variable a los shocks.<sup>11</sup>

La función impulso respuesta no es más que la respuesta de  $y_t$  en el tiempo a un incremento unitario y por una sola vez de  $u_t$  (shock o innovación positiva). Si estimamos un modelo ARIMA o AR puro tendríamos que computar la solución de la ecuación en diferencias, es decir, transformar el modelo en un MA puro. Si estimamos un MA, la función impulso respuesta la obtenemos directamente. Una vez que tenemos la representación de medias móviles:

$$y_t = \delta + u_t + \sum_{i=1}^q \beta_i u_{t-i}$$

donde puede ser que  $q = \infty$ , la función impulso respuesta se define como la derivada parcial de  $y_{t+i}$  con respecto a  $u_t$  en cada período posterior al shock:

$$\partial y_{t+i} / \partial u_t \quad \forall i \geq 0$$

lo cual es equivalente a:

$$\partial y_t / \partial u_{t-i} \quad \forall i \geq 0$$

La respuesta contemporánea, es decir, en el mismo período en que ocurre el shock ( $i = 0$ ) es  $\partial y_t / \partial u_t$  y se conoce como multiplicador de impacto o de corto plazo que en los modelos ARIMA normalmente es igual a 1. Los multiplicadores intermedios vendrían dados por:

$$\partial y_t / \partial u_{t-i} = \beta_i \quad \forall i \geq 1$$

Los siguientes serían algunos ejemplos de modelos ARIMA y sus correspondientes funciones impulso respuesta (FIR):

$$\text{MA}(1): \quad y_t = u_t + 0.8u_{t-1} \quad \text{FIR: } \{1, 0.8\}$$

$$\text{MA}(2): \quad y_t = u_t + 0.4u_{t-1} + 0.5u_{t-2} \quad \text{FIR: } \{1, 0.4, 0.5\}$$

<sup>11</sup> En la mayoría de los modelos de series de tiempo las relaciones no son sólo contemporáneas, sino que toman varios períodos, formando así una función impulso respuesta. Siempre que hayan rezagos de las variables involucradas, el efecto de los shocks y de las variables independientes van a ser una función del tiempo. En los trabajos aplicados de series de tiempo es mucho más usual ver el uso de las funciones impulso respuesta en los modelos de vectores autorregresivos (VAR), los cuales vamos a estudiar en el Capítulo 21.

$$\text{AR}(1): \quad y_t = 0.7y_{t-1} + u_t \quad \text{FIR: } \{1, 0.7, (0.7)^2, (0.7)^3, (0.7)^4 \dots\}$$

$$\text{AR}(1): \quad y_t = -0.7y_{t-1} + u_t \quad \text{FIR: } \{1, -0.7, (-0.7)^2, (-0.7)^3, (-0.7)^4 \dots\}$$

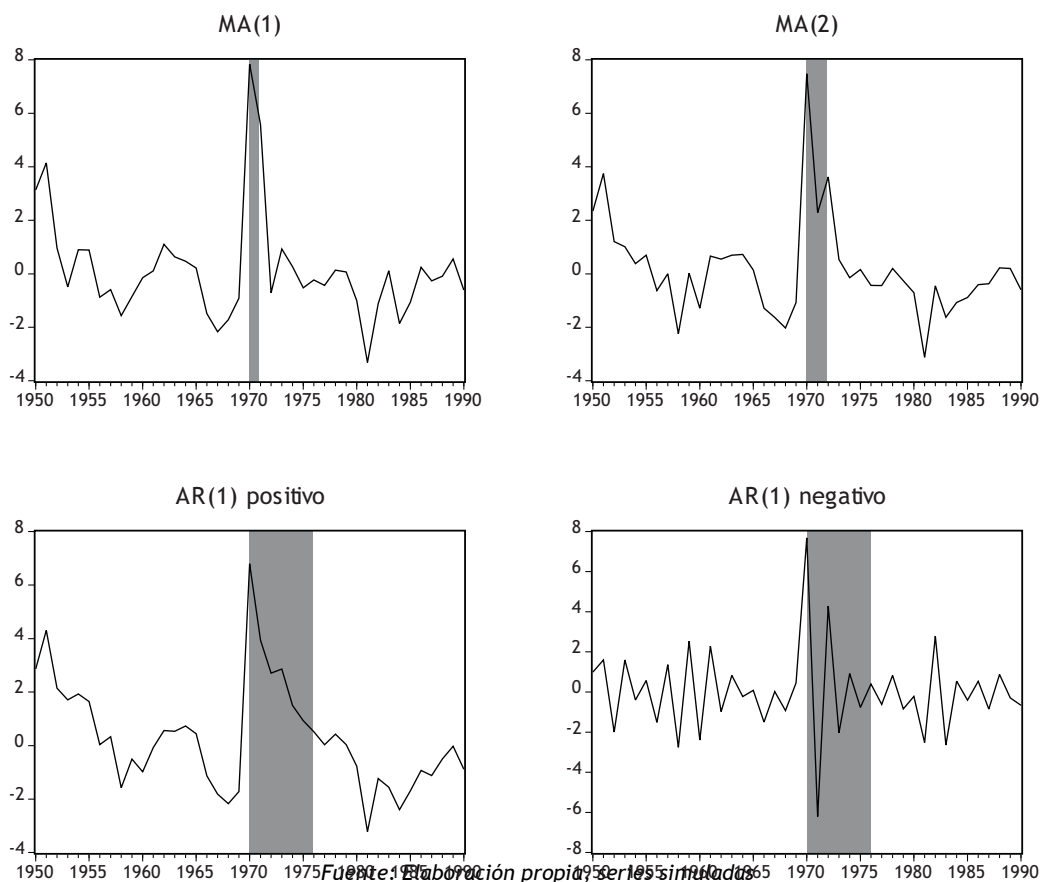
En los últimos dos ejemplos definimos la FIR valiéndonos de la solución (18.2) de un modelo AR(1). Las FIR de los MA siempre están explícitas y vemos que el número de períodos que se propaga un shock depende del orden de los términos MA. La cantidad de respuestas con valores no nulos será  $q+1$ . En los modelos AR puros o ARIMA la respuesta hay que calcularla y cuando los términos autorregresivos tienen un orden mayor a 1 el álgebra puede ser un poco complicada. No obstante, en todos los casos la respuesta se propagará por infinitos períodos, pero convergiendo a cero a medida que transcurre el tiempo (dada la condición de estacionariedad del modelo).

Para tener una visión gráfica, en la figura 18.4 se simula la respuesta a un shock positivo que tienen los cuatro modelos que acabamos de usar de ejemplos. No son exactamente las funciones impulso respuesta, puesto que la variable después que recibe el shock más grande, que adicionamos en el año 1970, sigue recibiendo otros pequeños shocks en los períodos sucesivos. Esto es más cercano a como en la práctica se observan los fenómenos económicos. El proceso  $y_t$  depende en todo el período de la serie de shocks  $u_t$ .<sup>12</sup> La respuesta es más visible a partir del shock más grande en 1970, pero en realidad la variable se mantiene todo el tiempo respondiendo a los shocks que va recibiendo, superponiéndose o anulándose en el tiempo las sucesivas respuestas.

En la figura se aprecia que el shock positivo en 1970 tiene un efecto contemporáneo en la variable, propagándose por un solo período en el modelo MA(1) y por dos períodos en el modelo MA(2). En el modelo MA(2) la respuesta al shock llega hasta el año 1972, siendo mayor el efecto en este último año que en 1971, coincidiendo así con los coeficientes que habíamos visto en la FIR de este modelo: la respuesta un período después de que ocurre el shock es 0.4 y dos períodos después es 0.5. Se aprecia que la respuesta en los modelos AR se propaga por más períodos, disminuyendo gradualmente el efecto. Después de seis años ya no se distingue en el gráfico algún efecto del shock. La respuesta del modelo AR(1) con coeficiente negativo es sinusoidal, se alternan valores positivos y negativos. En la FIR de este modelo se ve que  $y_t$  va a tener valores positivos cuando el exponente es par, y negativos cuando el exponente es impar. En la economía rara vez se observa esta forma de respuesta.

<sup>12</sup> Como sabemos, hay en cada observación un elemento aleatorio no recogido en la parte sistemática o determinista del modelo.

Figura 18.4: Respuesta a los shocks en modelos MA y AR



Con las funciones impulso respuesta terminamos con los elementos que queríamos mostrar sobre la formulación de los modelos ARIMA. Pasamos ahora a explicar el procedimiento de estimación de modelos ARIMA.

### 18.2.2 Las fases de la metodología

La metodología Box-Jenkins de estimación de modelos ARIMA está compuesta de cuatro pasos o fases:

**Fase 1. Identificación:** Consiste en determinar el orden de integración de la variable y el orden de los términos autorregresivos y de medias móviles que se van a incluir en el modelo  $ARIMA(p,d,q)$ . Entre las identificaciones posibles algunos de los órdenes  $p$ ,  $d$  y  $q$  podrían ser cero. Para identificar  $d$  se recurre al análisis gráfico, al correlograma y a alguna prueba de raíz unitaria como la prueba ADF. Para identificar  $p$  y  $q$  se recurre al correlograma y a las distintas formas que pueden adoptar dentro de este la función de autocorrelación y la función de autocorrelación parcial; esto lo mostraremos más adelante. Solamente anotar que es en esta fase donde se requiere determinada experiencia del econometrista. El resto de las fases están prácticamente au-

tomatizadas en los programas estadísticos disponibles, pero la identificación requiere de cierto “arte” en el investigador. La mejor identificación de  $p$  y  $q$  no se logra en el primer intento, el procedimiento exige de varias identificaciones provisionales.

**Fase 2. Estimación:** Una vez identificados  $p$ ,  $d$  y  $q$  se estima el modelo ARIMA con esta especificación. Se estiman, con los datos disponibles de la variable, los parámetros asociados a los términos autorregresivos y de medias móviles: los  $\hat{\alpha}_i$  y los  $\hat{\beta}_i$ . Los programas estadísticos utilizan como método de estimación los MCO, pero también se recurre a los estimadores Máximo-Verosímiles y a procedimientos no lineales.

**Fase 3. Validación del modelo:** Se verifica la significatividad de los parámetros con los t-estadísticos. Se chequea la estacionariedad e invertibilidad del modelo en las raíces de los polinomios  $A(Z)$  y  $B(Z)$ . Con las diferentes pruebas que se estudiaron en la parte tres del libro, y de la misma forma que vimos se procedía en el modelo lineal normal de regresión, se validan cada uno de los supuestos sobre los residuos; de manera tal que se verifique que las innovaciones son un ruido blanco gaussiano.

En la validación de los supuestos se debe hacer énfasis en el supuesto de no autocorrelación. El modelo ARIMA debe recoger toda la dependencia de la variable con su pasado. En la medida que esto se logre, mejores serán las predicciones que podremos hacer de la trayectoria futura de la variable en base a la información presente. Si el correlograma mantiene coeficientes de autocorrelación significativos, o no se validan el resto de los supuestos, se deben repetir los pasos 1 y 2. La estimación de modelos ARIMA es un proceso iterativo.

Puede encontrarse más de una especificación del modelo ARIMA que cumpla con los requisitos anteriores. En la sección 18.2.1 advertimos que existen representaciones equivalentes de modelos ARIMA. Diferentes modelos ARIMA pueden acercarse a la naturaleza del proceso estocástico que genera la serie temporal. De hecho, es aconsejable tener más de una variante de modelo ARIMA. Para decidir entonces entre las distintas variantes se evalúa cuál modelo ajusta mejor los datos. Ello puede hacerse observando el error estándar de la regresión o el  $R^2$ , no obstante, la metodología Box-Jenkins propone usar los criterios de información de Akaike y Schwartz (ver sección 6.4) para seguir el principio de *parquedad*, o parsimonia.

En este sentido, se debe elegir el modelo que logra el mejor ajuste de los datos con la menor cantidad de parámetros. En la sección 18.2.1 vimos que un modelo MA(1) invertible podía representarse como un AR( $\infty$ ) y que un modelo AR(1) estacionario podía transformarse en un MA( $\infty$ ). De estas especificaciones equivalentes el MA(1) y el AR(1) serían los modelos con una mayor parquedad. En los polinomios  $A(L)$  y  $B(L)$  puede haber también redundancia de parámetros. Cuando se incorporan al modelo más términos AR y MA aumenta el ajuste de los datos, pero a costa de reducir los grados de li-

bertad. Un argumento muy importante a favor del modelo que cumple mejor el principio de parquedad es que produce mejores predicciones. Algunos autores plantean como recomendación para la parquedad que se debería trabajar con un límite de  $p \leq 4$  y  $q \leq 4$ .

**Fase 4. Predicción:** Una vez que se seleccionan los mejores candidatos de modelos ARIMA se prueba la capacidad predictiva de los mismos. Generalmente se deja reservado para esta fase algunas observaciones del final del período de datos disponibles. De esta manera, se puede comparar la predicción que realiza el modelo ARIMA con los datos verdaderos y calcular diferentes medidas de errores de predicción. Esta es la prueba definitiva para elegir finalmente el modelo que usaremos para predecir la variable en estudio.

A esta altura del libro, el lector está en condiciones de implementar las fases 2 y 3 debido a que no hay muchas diferencias con la estimación y validación de un modelo de regresión con datos de corte transversal. A continuación explicamos la fase 1 que es la que más elementos nuevos contiene.

### 18.2.3 Fase de identificación

#### Identificación de $d$

Para identificar  $d$  conocemos del Capítulo 17 el análisis gráfico y la prueba ADF. Sólo vamos a añadir otros tres puntos importantes. El primero es llamar la atención sobre la **sobrediferenciación**. La varianza de una serie suele disminuir al diferenciarla, pero si diferenciamos más veces de las adecuadas entonces la varianza aumenta. Adicionalmente, la diferenciación de una variable estacionaria produce un proceso no invertible. Por ejemplo, un proceso no estacionario en media pero estacionario en varianza no debería diferenciarse (para recordar las medidas remediales para este tipo de no estacionariedad ver sección 17.2). Si se diferencia se crea un proceso ARMA(1,1) no invertible:

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 t + u_t \\ y_t - y_{t-1} &= a_1 y_{t-1} + a_2 t + u_t - (a_1 y_{t-2} + a_2 t + u_{t-1}) \\ (1-L)y_t &= a_1 y_{t-1} - a_1 L y_{t-1} + u_t - u_{t-1} \\ (1-L)y_t &= (1-a_1 L)y_{t-1} + (1-L)u_t \end{aligned}$$

En el polinomio asociado al término de medias móviles la raíz no es mayor a la unidad, por lo que no se cumple la condición de invertibilidad.

Otro punto importante es que usualmente en los modelos ARIMA además de la diferencia se le aplica a la variable la transformación logarítmica, esto es, se le aplica la **diferenciación logarítmica**. Si la variable aumenta a lo largo del tiempo, puede ser que después que diferenciamos, la varianza no sea constante en el tiempo. La diferenciación con el logaritmo reduce más la varianza de la serie y ayuda a resolver la heterocedasticidad, favoreciendo tener un proceso estacionario. Si se



desea efectuar ambas transformaciones debemos aplicar primero los logaritmos. Al tomar la diferenciación logarítmica la serie resultante aproximadamente es la tasa de variación de la variable:

$$(1-L)\ln y_t = \ln y_t - \ln y_{t-1} \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

El tercer y último punto es la **diferenciación estacional**. Cuando la frecuencia de los datos es mayor al año lo más usual es que esté presente en la variable el componente estacional. Box- Jenkins (1976) señalan que *“la característica principal de las series de tiempo con períodos  $s$ , es que las observaciones que están separadas por períodos  $s$  son similares”*.

Como el objetivo que perseguimos con el modelo ARIMA es predecir el valor “completo” de la variable, no tendría mucho sentido aplicar un método de desestacionalización. Si estimamos el modelo ARIMA con los datos desestacionalizados evidentemente tendríamos que pronosticar la variable sin estacionalidad. Para predecir la variable incluyendo la estacionalidad debemos estimar este componente dentro del modelo ARIMA.

Para estimar la estacionalidad en la metodología Box-Jenkins existen dos opciones:

1) Usar variables dicótomicas (ver sección 7.2). Al modelo  $ARIMA(p,d,q)$  se le añadiría un juego de dummies:

$$A(L)(1-L)^d y_t = \delta + B(L)u_t + \sum_{i=1}^{s-1} m_i D_i$$

donde  $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$ ,  $B(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$  y  $D_i$  es una variable dummy estacional. El índice  $s$  es la frecuencia estacional ( $s=12$  en datos mensuales y  $s=4$  en datos trimestrales). Se incluye una dummy menos para no caer en la trampa de la variable cualitativa.

2) Aplicar a la variable una diferencia estacional. Para series mensuales empleamos la transformación:

$$(1-L^{12})y_t = y_t - y_{t-12}$$

Para series trimestrales sería:

$$(1-L^4)y_t = y_t - y_{t-4}$$

Si se aplica conjuntamente la diferencia estacional con el logaritmo, la serie resultante aproximadamente es la tasa de variación interanual:

$$(1-L^s)\ln y_t \approx \frac{y_t - y_{t-s}}{y_{t-s}}$$

Se puede aplicar más de una diferencia estacional. A las diferencias estacionales se les denota como  $\Delta_s^D$ , con el superíndice  $D$  para distinguirla del orden de las diferencias regulares  $d$ . Sin embargo, lo más usual es trabajar con  $D=1$ .

$$\Delta_s^D y_t = (1 - L^s)^D y_t$$

$$\Delta_{12}^1 y_t = (1 - L^{12})^1 y_t = y_t - y_{t-12}$$

$$\Delta_4^2 y_t = (1 - L^4)^2 y_t = (1 - L^4)(y_t - y_{t-4})$$

La estimación de la estacionalidad mediante las dummies supone el componente estacional como un proceso determinista. Los coeficientes estimados para cada dummy  $\hat{m}_i$  son constantes en todo el período. En cambio, con la diferenciación estacional se representa el componente estacional como un proceso estocástico.

La diferencia estacional no exime de usar la diferencia regular. La diferencia estacional no garantiza la estacionariedad de la serie. Si estamos trabajando con una serie no estacionaria que presenta estacionalidad debemos aplicar las dos diferencias. Si además creemos que es necesaria la transformación logarítmica, aplicamos la transformación:

$$\Delta^d \Delta_s^D \ln y_t = (1 - L)^d (1 - L^s)^D \ln y_t$$

Por ejemplo, en la investigación empírica se ha comprobado que una gran cantidad de series económicas que tienen una frecuencia mensual requieren de la transformación:

$$\Delta \Delta_{12} \ln y_t = (1 - L)(1 - L^{12}) \ln y_t$$

La diferenciación de la variable le proporciona al modelo ARIMA una mayor flexibilidad para adaptar las predicciones a los acontecimientos más recientes. Al aplicar la diferencia regular y estacional a la variable estamos modelando la tendencia y la estacionalidad con una especificación estocástica, y estamos permitiendo que ambos componentes puedan variar en el tiempo. No obstante, como ya apuntamos, hay que tomar cuidados con la posibilidad de estar sobrediferenciando la serie.

Transformar la variable para la fase de identificación y estimación no significa que tengamos que predecir al final la transformación de la variable. Los programas estadísticos normalmente permiten pronosticar la serie en la unidad de medida original aún cuando en el modelo la hayamos incluido con un grupo de transformaciones.

### Identificación de $p$ y $q$

Para determinar el orden de los términos autorregresivos y de medias móviles que se van a incluir en el modelo ARIMA( $p, d, q$ ) la metodología Box-Jenkins propone observar el **correlograma** de la variable. Como vimos en el capítulo 10, el correlograma recoge la estructura de dependencia temporal de la serie y está compuesto por la función de autocorrelación (**FAC**) y la función de autocorrelación

parcial (**FACP**). Nos apoyamos en estas funciones para medir la relación que una variable tiene con ella misma en períodos anteriores. La FAC que denotamos como  $\rho_k$  mide la correlación entre  $y_t$  y  $y_{t-k}$ . La FACP igualmente mide la correlación entre  $y_t$  y  $y_{t-k}$  pero después de eliminar el efecto de las correlaciones intermedias. Tanto la FAC y la FACP indicarán una correlación alta positiva (negativa) entre  $y_t$  y  $y_{t-k}$  en la medida que el coeficiente de autocorrelación se acerque a 1 positivo (negativo).

Como conocemos del capítulo 10, el correlograma que nos presenta el paquete estadístico es una estimación del correlograma verdadero, computado con las observaciones disponibles  $T$ , pudiéndose estimar hasta un máximo  $T/4$  valores de la FAC y la FACP. Si la variable ha sido diferenciada, tenemos que tomar en cuenta que se reducen las observaciones inicialmente disponibles  $T^1$ . Después de diferenciar la serie nos quedaría  $T = T^1 - d - D$ .

La estimación de las FAC y la FACP es posible si el proceso estocástico subyacente es estacionario y **ergódico**. Un proceso estacionario se dice que es ergódico si los valores de  $y_t$  suficientemente alejados en el tiempo están prácticamente incorrelacionados:  $\lim_{k \rightarrow \infty} \rho_k = 0$ . Si un proceso no es ergódico, las observaciones estarán muy relacionadas, por lo que al aumentar el tamaño del período muestral se aporta poca información adicional para la estimación. La condición de ergodicidad permite llevar a cabo estimaciones consistentes de los parámetros del proceso.

Hacemos aquí también uso del estadístico  $Q$  que estudiamos en el capítulo 10:

$$Q = T(T+2) \sum_{j=1}^k \rho_j^2 / (T-j)$$

Si  $y_t$  viene generado por un proceso estacionario y ergódico,  $Q$  se distribuirá asintóticamente  $\chi^2$  con  $k$  grados de libertad. La hipótesis nula del contraste plantea que todos los coeficientes de autocorrelación son ceros hasta el rezago  $k$  ( $\rho_j = 0$ ,  $j = 1, \dots, k$ ). Si se rechaza la nula, entonces al menos un coeficiente de autocorrelación es distinto de cero hasta el rezago  $k$  ( $\exists \rho_j \neq 0$ ,  $j = 1, \dots, k$ ). La prueba se repite para cada rezago verificando si hay o no autocorrelación hasta el orden  $k$ .

Inicialmente en esta fase de identificación, cuando le pedimos al computador que nos proporcione el correlograma de la variable, debemos obtener un correlograma con autocorrelación debido a que las observaciones de la mayoría de las series temporales están correlacionadas en el tiempo; de hecho, en esta información nos apoyamos para hacer las predicciones. Asimismo, se espera que después de añadir los términos AR y MA que representan el proceso estocástico que genera la serie, obtengamos un correlograma “limpio” de autocorrelación.

Para llegar a ello, debemos aprender a leer la información que nos proporciona un correlograma con autocorrelación. La FAC y la FACP toman diferentes formas

en función del tipo de proceso, sea un AR puro, un MA puro, un ARMA, y en dependencia también del orden de cada término. Veamos algunos ejemplos.

En la figura 18.5 se grafica el correlograma de un proceso AR(1). El correlograma de un proceso AR(1) va a tener un “pico” o autocorrelación significativa en el primer rezago de la FACP. Si el coeficiente del término AR(1) en el proceso es positivo, la FAC va a decrecer exponencialmente, como en la figura 18.5.A. Si el coeficiente es negativo, la FAC va a decrecer de manera sinusoidal, como en la figura 18.5.B. El signo del AR(1) sólo es para ilustrar los patrones que puede tomar la FAC, porque en realidad el signo y el valor del coeficiente asociado al término AR(1) se estimaría en la siguiente fase de la metodología Box-Jenkins.

**Figura 18.5: Correlograma de un proceso AR(1),  $y_t = \alpha_1 y_{t-1} + u_t$**

**A. AR(1) positivo ( $\alpha_1 > 0$ )**

| Autocorrelation                                                                    | Partial Correlation                                                                | AC       | PAC    | Q-Stat | Prob  |
|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|----------|--------|--------|-------|
|  |  | 1 0.830  | 0.830  | 746.87 | 0.000 |
|                                                                                    |                                                                                    | 2 0.692  | 0.009  | 1266.3 | 0.000 |
|                                                                                    |                                                                                    | 3 0.593  | 0.052  | 1647.9 | 0.000 |
|                                                                                    |                                                                                    | 4 0.494  | -0.044 | 1912.7 | 0.000 |
|                                                                                    |                                                                                    | 5 0.408  | -0.007 | 2094.2 | 0.000 |
|                                                                                    |                                                                                    | 6 0.348  | 0.029  | 2226.4 | 0.000 |
|                                                                                    |                                                                                    | 7 0.288  | -0.028 | 2316.6 | 0.000 |
|                                                                                    |                                                                                    | 8 0.247  | 0.033  | 2383.0 | 0.000 |
|                                                                                    |                                                                                    | 9 0.216  | 0.008  | 2433.8 | 0.000 |
|                                                                                    |                                                                                    | 10 0.173 | -0.045 | 2466.4 | 0.000 |
|                                                                                    |                                                                                    | 11 0.150 | 0.039  | 2491.1 | 0.000 |
|                                                                                    |                                                                                    | 12 0.141 | 0.027  | 2512.9 | 0.000 |

**B. AR(1) negativo ( $\alpha_1 < 0$ )**

| Autocorrelation                                                                     | Partial Correlation                                                                 | AC        | PAC    | Q-Stat | Prob  |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------|--------|--------|-------|
|  |  | 1 -0.773  | -0.773 | 648.36 | 0.000 |
|                                                                                     |                                                                                     | 2 0.598   | -0.001 | 1035.9 | 0.000 |
|                                                                                     |                                                                                     | 3 -0.458  | 0.009  | 1263.9 | 0.000 |
|                                                                                     |                                                                                     | 4 0.381   | 0.075  | 1421.8 | 0.000 |
|                                                                                     |                                                                                     | 5 -0.311  | 0.011  | 1527.1 | 0.000 |
|                                                                                     |                                                                                     | 6 0.251   | -0.008 | 1595.5 | 0.000 |
|                                                                                     |                                                                                     | 7 -0.179  | 0.052  | 1630.3 | 0.000 |
|                                                                                     |                                                                                     | 8 0.107   | -0.052 | 1642.8 | 0.000 |
|                                                                                     |                                                                                     | 9 -0.039  | 0.051  | 1644.5 | 0.000 |
|                                                                                     |                                                                                     | 10 0.003  | 0.009  | 1644.5 | 0.000 |
|                                                                                     |                                                                                     | 11 0.019  | 0.001  | 1644.9 | 0.000 |
|                                                                                     |                                                                                     | 12 -0.049 | -0.046 | 1647.6 | 0.000 |

Fuente: Elaboración propia, series simuladas

En la figura 18.6 se grafica el patrón que seguirían la FAC y la FACP de un proceso MA(1). En esta ocasión los picos que indican que es un MA se observan en la FAC, a diferencia de lo que sucede con los procesos AR donde los picos aparecen en la FACP. Como el orden del MA es 1 entonces el pico de la FAC aparece en el primer rezago. Si el coeficiente asociado al término MA(1) en el proceso es negativo, la FACP va a decrecer sinusoidal, como en la figura 18.6.B. Si el coeficiente es positivo, la FAC va a decrecer de manera exponencialmente como en la figura 18.6.A.

**Figura 18.6: Correlograma de un proceso MA(1),  $y_t = u_t + \beta_1 u_{t-1}$**

**A. MA(1) positivo ( $\beta_1 > 0$ )**

| Autocorrelation                                                                    | Partial Correlation                                                                | AC                            | PAC | Q-Stat | Prob |
|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------|-----|--------|------|
|  |  | 1 0.527 0.527 144.95 0.000    |     |        |      |
|  |  | 2 0.063 -0.297 147.03 0.000   |     |        |      |
|  |  | 3 0.034 0.224 147.63 0.000    |     |        |      |
|  |  | 4 0.033 -0.132 148.20 0.000   |     |        |      |
|  |  | 5 0.023 0.103 148.47 0.000    |     |        |      |
|  |  | 6 0.001 -0.087 148.47 0.000   |     |        |      |
|  |  | 7 0.036 0.130 149.17 0.000    |     |        |      |
|  |  | 8 0.094 0.002 153.85 0.000    |     |        |      |
|  |  | 9 0.085 0.043 157.68 0.000    |     |        |      |
|  |  | 10 0.049 -0.009 158.97 0.000  |     |        |      |
|  |  | 11 -0.006 -0.042 158.98 0.000 |     |        |      |
|  |  | 12 -0.014 0.033 159.09 0.000  |     |        |      |

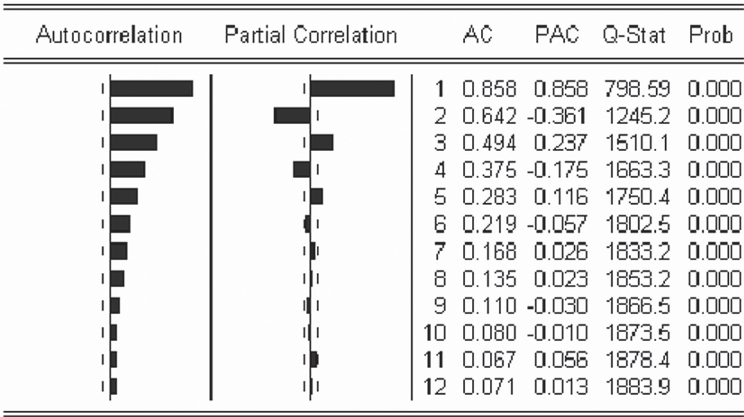
**B. MA(1) negativo ( $\beta_1 < 0$ )**

| Autocorrelation                                                                     | Partial Correlation                                                                 | AC        | PAC    | Q-Stat | Prob  |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------|--------|--------|-------|
|  |  | 1 -0.469  | -0.469 | 238.89 | 0.000 |
|                                                                                     |                                                                                     | 2 -0.062  | -0.362 | 243.07 | 0.000 |
|                                                                                     |                                                                                     | 3 0.073   | -0.203 | 248.79 | 0.000 |
|                                                                                     |                                                                                     | 4 -0.007  | -0.125 | 248.85 | 0.000 |
|                                                                                     |                                                                                     | 5 -0.049  | -0.142 | 251.46 | 0.000 |
|                                                                                     |                                                                                     | 6 0.063   | -0.051 | 255.79 | 0.000 |
|                                                                                     |                                                                                     | 7 -0.043  | -0.070 | 257.77 | 0.000 |
|                                                                                     |                                                                                     | 8 -0.014  | -0.085 | 257.99 | 0.000 |
|                                                                                     |                                                                                     | 9 0.068   | 0.005  | 263.01 | 0.000 |
|                                                                                     |                                                                                     | 10 -0.048 | -0.016 | 265.53 | 0.000 |
|                                                                                     |                                                                                     | 11 -0.007 | -0.024 | 265.58 | 0.000 |
|                                                                                     |                                                                                     | 12 -0.003 | -0.055 | 265.59 | 0.000 |

Fuente: Elaboración propia, series simuladas

Cuando el orden de los términos AR y MA es mayor o cuando el proceso estocástico subyacente es un ARMA, entonces la identificación del proceso a partir del correlograma se vuelve un poco más difícil. En la figura 18.7 se grafica el patrón que seguirían la FAC y la FACP de un proceso ARMA(1,1). Ambas funciones decrecerían a partir del primer rezago, la FAC de manera exponencialmente y la FACP de manera sinusoidal.

Figura 18.7: Correlograma de un proceso ARMA(1,1)



Fuente: Elaboración propia, series simuladas

En la tabla 18.1 se resume el patrón que siguen los diferentes procesos AR, MA y ARMA con distintos órdenes. En la práctica, la FAC y FACP no coinciden exactamente con los patrones teóricos. En primer lugar, porque solamente se pueden observar la FAC y FACP estimadas con los datos que tenemos a mano de la serie de tiempo, no son la FAC y FACP del proceso estocástico subyacente. En segundo lugar, la serie de tiempo probablemente tendrá estacionalidad, cambios estructurales y/o observaciones atípicas, factores que dificultan la lectura del correlograma. Así, la tarea del investigador es buscar, en un proceso iterativo, qué modelo teórico se aproxima más al correlograma que se observa. De la fase de estimación se regresa a la fase de identificación en varias ocasiones hasta encontrar la mejor especificación de  $p$  y  $q$ .

Tabla 18.3: Correlogramas teóricos para procesos AR, MA y ARMA

|               |                                                                                  |                                                                                  |
|---------------|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Modelo        | FAC                                                                              | FACP                                                                             |
| AR( $p$ )     | Decrece exponencialmente o con un patrón sinusoidal                              | Picos grandes hasta el rezago $p$                                                |
| MA( $q$ )     | Picos grandes hasta el rezago $q$                                                | Decrece exponencialmente o con un patrón sinusoidal                              |
| ARMA( $p,q$ ) | Decrece exponencialmente o con un patrón sinusoidal, comenzando en el rezago $q$ | Decrece exponencialmente o con un patrón sinusoidal, comenzando en el rezago $p$ |

Fuente: Elaboración propia

Volviendo al tema de la estacionalidad. Cuando en la serie está presente el componente estacional, en el correlograma pueden aparecer correlaciones significativas en los rezagos de orden  $s$  o múltiplos de  $s$ . La diferenciación estacional modela una parte de este componente. Sin embargo, una vez diferenciada la serie e identificado el modelo regular ( $p$  y  $q$ ), podemos observar a través del correlograma si se producen rebotes en estos rezagos múltiplos de  $s$ . En ese caso, el modelo ARIMA se puede extender para incluir también términos AR y/o MA en los rezagos estacionales. Si trabajamos por ejemplo con datos mensuales, puede ser conveniente probar incluir un término AR(12) y/o MA(12). En datos trimestrales, puede ser conveniente probar incluir un término AR(4) y/o MA(4). Ello permite estimar la **correlación serial estacional** que puede existir en los datos y que no queda recogida con la diferencia estacional. Para distinguirlo del orden de los términos AR y MA no estacionales  $p$  y  $q$ , al orden de los términos **AR estacionales** se le denota como  $P$  y al orden de los términos **MA estacionales** se le denota como  $Q$ . El modelo lo denotaríamos como **ARIMA( $p, d, q$ )( $P, D, Q$ )**:

$$A(L)A(L)^s \Delta^d \Delta_s^D \ln y_t = \delta + B(L)B(L)^s u_t$$

donde:

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p,$$

$$A(L)^s = (1 - \phi_1 L^s - \phi_2 L^{2s} - \dots - \phi_p L^{ps})$$

$$B(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$$

$$B(L)^s = (1 - \theta_1 L^s - \theta_2 L^{2s} - \dots - \theta_Q L^{Qs})$$

$$\Delta^d = (1 - L)^d$$

$$\Delta_s^D = (1 - L^s)^D$$

Una vez finalizada la etapa de identificación del modelo tenemos que estimar con los datos disponibles los parámetros  $\delta$ ,  $\alpha_i$ ,  $\phi_i$ ,  $\beta_i$  y  $\theta_i$ . Se ha probado que un alto porcentaje de las series de tiempo en economía pueden ser modeladas con un ARIMA(0,1,1)(0,1,1). A esta especificación se le conoce como *modelo de líneas aéreas*, dado que fue empleada inicialmente por Box y Jenkins para estimar y predecir la demanda de los pasajeros por este servicio.

Hasta aquí llegamos con la parte de la modelación univariante. Transitamos ahora, en los siguientes capítulos, hacia la modelación multivariante de series temporales.





# *preguntas*

Intente dar respuesta a las preguntas a continuación. Este esfuerzo le servirá para comprobar si ha entendido los contenidos de este capítulo

1. Mencione los cuatro componentes que caracterizan una serie de tiempo
2. ¿El componente irregular está presente en la serie desestacionalizada o en la serie tendencia-ciclo?
3. Enumere los diferentes métodos que son usados para estimar los componentes de una serie de tiempo.
4. ¿Qué es una función impulso-respuesta?
5. ¿Cuándo un proceso ARMA es invertible?



# ejercicios

18.1 Aplique el método TRAMO-SEATS con los datos de los arribos trimestrales de turistas alemanes a Cuba (tabla B.11 del CD acompañante) para:

- Estimar la serie ajustada de estacionalidad
- Estimar la serie tendencia-ciclo
- Analizar la diferencia entre ambas

NOTA: Para aplicar el método TRAMO-SEATS puede usar el programa Demetra o el programa Eviews.

18.2 Escriba la formulación de un proceso MA(3) y de un ARMA(2,1). Presente también la formulación utilizando el operador de rezagos.

18.3 Identifique el orden de  $p$  y  $q$  que sugieren los siguientes correlogramas.

| Autocorrelation | Partial Correlation | AC        | PAC    | Q-Stat | Prob  |
|-----------------|---------------------|-----------|--------|--------|-------|
|                 |                     | 1 0.524   | 0.524  | 142.85 | 0.000 |
|                 |                     | 2 0.278   | 0.005  | 183.11 | 0.000 |
|                 |                     | 3 0.045   | -0.141 | 184.18 | 0.000 |
|                 |                     | 4 0.040   | 0.096  | 185.02 | 0.000 |
|                 |                     | 5 0.023   | 0.001  | 185.29 | 0.000 |
|                 |                     | 6 0.026   | -0.012 | 185.63 | 0.000 |
|                 |                     | 7 0.048   | 0.057  | 186.86 | 0.000 |
|                 |                     | 8 0.099   | 0.075  | 192.00 | 0.000 |
|                 |                     | 9 0.087   | -0.013 | 195.97 | 0.000 |
|                 |                     | 10 0.056  | -0.009 | 197.61 | 0.000 |
|                 |                     | 11 0.017  | -0.001 | 197.77 | 0.000 |
|                 |                     | 12 -0.011 | -0.025 | 197.83 | 0.000 |

| Autocorrelation | Partial Correlation | AC       | PAC    | Q-Stat | Prob  |
|-----------------|---------------------|----------|--------|--------|-------|
|                 |                     | 1 0.583  | 0.583  | 177.02 | 0.000 |
|                 |                     | 2 0.399  | 0.089  | 259.96 | 0.000 |
|                 |                     | 3 0.180  | -0.128 | 276.87 | 0.000 |
|                 |                     | 4 0.050  | -0.055 | 278.17 | 0.000 |
|                 |                     | 5 0.040  | 0.086  | 279.02 | 0.000 |
|                 |                     | 6 0.044  | 0.038  | 280.04 | 0.000 |
|                 |                     | 7 0.066  | 0.019  | 282.32 | 0.000 |
|                 |                     | 8 0.106  | 0.061  | 288.20 | 0.000 |
|                 |                     | 9 0.086  | -0.016 | 292.09 | 0.000 |
|                 |                     | 10 0.072 | -0.002 | 294.80 | 0.000 |
|                 |                     | 11 0.035 | -0.014 | 295.45 | 0.000 |
|                 |                     | 12 0.005 | -0.011 | 295.46 | 0.000 |

18.4 Estime un modelo ARIMA para la serie de los arribos trimestrales de turistas alemanes a Cuba (tabla B.11 del CD). Siga las diferentes fases de la metodología Box-Jenkins. Trabaje con la serie en logaritmo.



CAPÍTULO 19



## FUNCIÓN DE TRANSFERENCIA

A partir del presente capítulo transitamos a la modelación multivariante de series temporales. Comenzamos con los llamados modelos de función de transferencia, los cuales son regresiones lineales que involucran datos organizados en el tiempo y añaden rezagos o retardos de las variables en su especificación. La inclusión de rezagos de la variable dependiente y/o de las variables explicativas permite considerar relaciones dinámicas e introduce algunos elementos particulares en la interpretación de los resultados. Asimismo, se deben estudiar algunos aspectos nuevos asociados al proceso de estimación, aunque en la inmensa mayoría de los casos se siguen utilizando los estimadores Mínimos Cuadrados Ordinarios y la comprobación de los supuestos que condicionan sus propiedades.

Además de la interpretación y estimación de funciones de transferencia, en el presente capítulo desarrollamos el tema del análisis de intervención. El mismo es una continuación del tópico sobre las variables mudas estudiado en el capítulo 7.

### 19.1 Modelos Multivariantes Dinámicos

En los capítulos iniciales del libro especificábamos el modelo de regresión lineal simple sin rezagos:

$$y_t = \delta + c_0 x_t + u_t$$

De esta manera, estábamos restringiendo el efecto de  $x_t$  sobre  $y_t$  a un solo período, considerando solamente el efecto contemporáneo (momento  $t$ ). Sin embargo, podríamos pensar que el efecto demora varios períodos en manifestarse completamente, es decir, que existe una relación dinámica.

En teoría económica se pueden encontrar varias razones por las cuales debemos especificar rezagos en las relaciones que estamos estimando con nuestro modelo:

- Razones sicológicas: Inercia e importancia de los hábitos en el comportamiento de los individuos y empresas.
- Razones tecnológicas: Los efectos de las nuevas tecnologías no son inmediatos.
- Razones institucionales: Las instituciones suelen trabajar bajo determinadas reglas y contratos laborales y con los proveedores que añaden rigidez a la toma de decisiones.

En un modelo de regresión lineal, hay dos formas fundamentales de especificar relaciones dinámicas entre las variables. Una de ellas es incluyendo rezagos de la variable explicativa:

$$y_t = \delta + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + \dots + c_k x_{t-k} + u_t \quad (19.1)$$

Este se conoce como **modelo de rezagos distribuidos**. Podemos incluir  $k$  rezagos de  $x_t$ , lo cual limita hasta cuántos períodos hacia adelante tienen lugar los efectos de  $x_t$  sobre  $y_t$ . Nótese que es lo mismo decir que  $y_t$  depende del valor contemporáneo y los valores rezagados de  $x_t$ , que decir que  $x_t$  tiene efecto sobre  $y_t$  en el período contemporáneo y en los períodos subsiguientes. Por ejemplo, con  $k = 2$  tendríamos el siguiente modelo:

$$y_t = \delta + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + u_t \quad (19.2)$$

Si  $t$  estuviera referido a años, diríamos que los cambios unitarios en  $x_t$  impactan ese mismo año a  $y_t$  en una magnitud  $c_0$ . Para conocer el efecto de  $x_t$  sobre  $y_t$  al próximo año, es decir, sobre  $y_{t+1}$ , debemos actualizar el modelo en un período:

$$y_{t+1} = \delta + c_0 x_{t+1} + c_1 x_t + c_2 x_{t-1} + u_{t+1}$$

De esta forma, queda explícito que el impacto de  $x_t$  el próximo año va a ser igual a  $c_1$ . Al cabo de dos años el impacto sería igual a  $c_2$ ; actualizando el modelo por dos períodos ello queda más evidente:

$$y_{t+2} = \delta + c_0 x_{t+2} + c_1 x_{t+1} + c_2 x_t + u_{t+2}$$

El coeficiente  $c_0$  se conoce como **multiplicador de impacto o de corto plazo** y los restantes  $c_j$  son los **multiplicadores intermedios**. Todos ellos serían resultados de las derivadas parciales:

$$\partial y_t / \partial x_{t-j} = c_j \quad \forall j \geq 0$$

Se puede además computar el efecto de  $x_t$  sobre la variable dependiente hasta determinado período, esto es, el efecto acumulado hasta el período  $n$ . Para ello, sumamos los multiplicadores intermedios:

$$\sum_{j=0}^n c_j$$

El efecto completo o **multiplicador de largo plazo** se refiere entonces a la suma de todos los multiplicadores intermedios hasta el período  $k$ :

$$\sum_{j=0}^k c_j$$

En el ejemplo con dos rezagos (19.2), el multiplicador de largo plazo vendría dado por:

$$\sum_{j=0}^2 c_j = c_0 + c_1 + c_2$$

Normalmente, esperamos que los consecutivos  $c_j$  tengan el mismo signo y sean decrecientes. Ello es más coherente con las explicaciones teóricas psicológicas, tecnológicas e institucionales que vimos justifican los rezagos en un modelo de regresión.

El modelo (19.1) puede ser fácilmente extendido a más de una variable explicativa para tener un modelo múltiple de rezagos distribuidos. Con dos variables explicativas,  $x_1$  y  $x_2$ , quedaría:

$$y_t = \delta + c_{10}x_{1t} + \dots + c_{1k}x_{1t-k} + c_{20}x_{2t} + \dots + c_{2m}x_{2t-m} + u_t$$

Al igual que estudiamos en el capítulo 18 en los modelos univariantes, en los modelos multivariantes también tenemos **funciones impulso respuesta**. En este caso, estaría definida por la respuesta en el tiempo de  $y_t$  a un shock en  $x_t$ . El shock lo seguimos entendiendo como un incremento unitario y por una sola vez de  $x_t$ . La función impulso respuesta la construimos con los consecutivos multiplicadores intermedios  $\{c_0, c_1, c_2, \dots, c_k\}$ .

Si tuviésemos más de una variable independiente en el modelo, habría tantas funciones impulso respuesta como variables independientes. Las funciones impulso respuesta nos resumen las diferentes formas que tienen las relaciones dinámicas entre la variable dependiente y las variables a la derecha de la ecuación. Adicionalmente, podemos evaluar la respuesta de  $y_t$  a un shock en  $u_t$ , tal y como lo hacíamos en los modelos ARIMA; ello serviría para examinar la respuesta en el tiempo de  $y_t$  ante un shock propio.

Como alternativa a los modelos de rezagos distribuidos, tenemos los **modelos autorregresivos multivariantes**. Tales modelos, toman en cuenta que otra forma

de especificar relaciones dinámicas en un modelo lineal de regresión es rezagando la variable dependiente. Ello sería similar a lo que hicimos en la metodología Box-Jenkins donde teníamos modelos MA y modelos AR, pero ahora en un contexto multivariante.

Un modelo autorregresivo multivariante con una sola variable explicativa y con un solo rezago de la variable dependiente (modelo autorregresivo multivariante de orden uno) tendría la siguiente forma:

$$y_t = \delta + \alpha_1 y_{t-1} + c_0 x_t + u_t \quad (19.3)$$

Nótese que la relación entre  $x_t$  y  $y_t$  no se reduce a un período. Cualquier impacto que  $x_t$  tenga sobre  $y_t$  se propagará hacia el futuro debido al término autorregresivo  $y_{t-1}$ . La dependencia de  $y_t$  con su pasado propaga los efectos en los períodos subsiguientes.

Tal y como ya conocemos, el modelo (19.3) es una ecuación en diferencias donde no queda explícito el impacto de la variable explicativa. Debemos obtener la solución de la ecuación en diferencias. Para ello nos valemos nuevamente del operador de rezagos:

$$\begin{aligned} y_t &= \delta + \alpha_1 L y_t + c_0 x_t + u_t \\ y_t - \alpha_1 L y_t &= \delta + c_0 x_t + u_t \\ (1 - \alpha_1 L) y_t &= \delta + c_0 x_t + u_t \\ y_t &= \delta / (1 - \alpha_1 L) + c_0 / (1 - \alpha_1 L) x_t + u_t / (1 - \alpha_1 L) \end{aligned} \quad (19.4)$$

Así, comprobamos que el efecto de  $x_t$  sobre  $y_t$  no se reduce a  $c_0$ , sino que es igual a  $c_0 / (1 - \alpha_1 L)$ ; este último término veremos a continuación que es el multiplicador de largo plazo.

Para despejar los diferentes multiplicadores debemos desarrollar la solución de la ecuación en diferencias (19.4). Utilizamos las propiedades del operador de rezagos que estudiamos en la sección 18.2.1, las cuales además requieren que  $|\alpha_1| < 1$ :

$$y_t = \delta / (1 - \alpha_1) + (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 \dots) c_0 x_t + u_t / (1 - \alpha_1 L)$$

Llegamos por esta vía a un modelo de rezagos distribuidos infinito:

$$y_t = \mu + c_0 x_t + c_0 \alpha_1 x_{t-1} + c_0 \alpha_1^2 x_{t-2} + c_0 \alpha_1^3 x_{t-3} + \dots + e_t \quad (19.5)$$

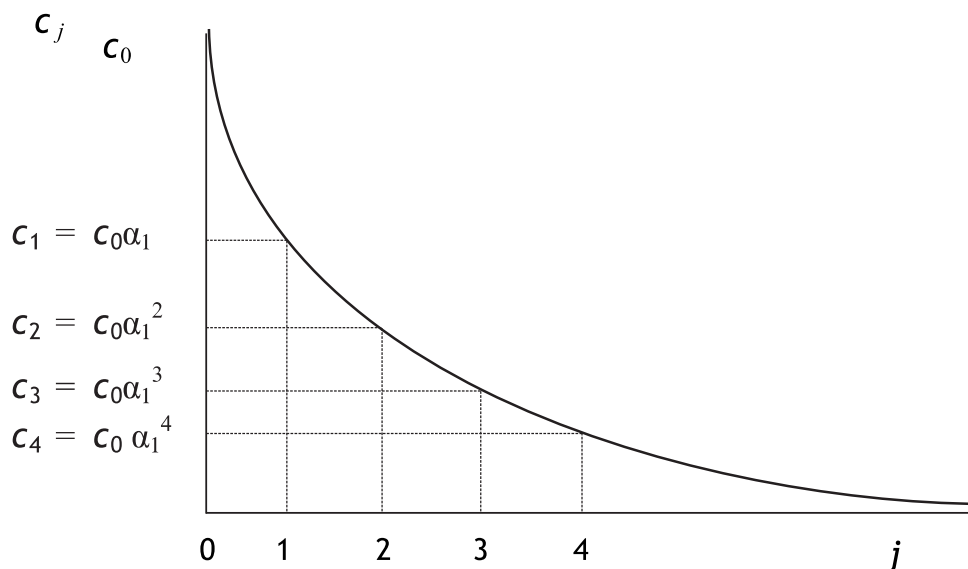
donde  $\mu = \delta / (1 - \alpha_1)$  y  $e_t = u_t / (1 - \alpha_1 L)$ .

El multiplicador de corto plazo sería  $c_0$ , el multiplicador intermedio para el primer período sería  $c_0 \alpha_1$ , el del segundo período sería  $c_0 \alpha_1^2$  y el del período  $k$  vendría dado por  $c_0 \alpha_1^k$ .



Todos los multiplicadores intermedios dependen de los dos parámetros del modelo autorregresivo (19.3): el coeficiente  $c_0$  asociado a la variable explicativa  $x_t$  y el coeficiente  $\alpha_1$  asociado al término autorregresivo  $y_{t-1}$ . Recuerde que para poder aplicar las propiedades del operador de rezago y llegar a la solución de la ecuación en diferencias supusimos que  $|\alpha_1| < 1$ . Ello implica que los multiplicadores intermedios van a ir decreciendo geométricamente (ver figura 19.1).

Figura 19.1: Estructura de rezagos geométricos



Fuente: Elaboración propia

La figura 19.1 dibuja la forma que tiene la función impulso respuesta de un modelo autorregresivo multivariante de orden uno (ecuación 19.3) con  $|\alpha_1| < 1$ .

El efecto acumulado de  $x_t$  sobre  $y_t$ , por ejemplo, hasta dos períodos, vendría dado por  $c_0 + c_0\alpha_1 + c_0\alpha_1^2$ ; mientras que el multiplicador de largo plazo sería:

$$c_0 + c_0\alpha_1 + c_0\alpha_1^2 + c_0\alpha_1^3 \dots = c_0(1 + \alpha_1 + \alpha_1^2 + \alpha_1^3 \dots) = c_0 / (1 - \alpha_1)$$

Vemos que tanto los modelos autorregresivos multivariantes como los modelos de rezagos distribuidos permiten considerar relaciones dinámicas entre las variables. Los segundos, siempre que no tengan infinitos rezagos, tienen bien delimitado cuantos períodos demora el efecto de la variable explicativa; después de  $k$  períodos el efecto es nulo. En cambio los modelos autorregresivos propagan el efecto de la variable explicativa infinitamente, aunque a medida que nos alejamos del momento  $t$  los multiplicadores intermedios tienden a cero.

La equivalencia que conocíamos entre una ecuación en diferencias y su solución nos sirvió en el capítulo 18 para entender la equivalencia entre los procesos autorregresivos y los procesos de medias móviles. Ahora la utilizamos para

examinar la equivalencia entre los modelos autorregresivos multivariantes y los modelos de rezagos distribuidos.

En la literatura sobre econometría también se pueden encontrar diferentes transformaciones de modelos dinámicos, basadas en supuestos sobre las expectativas y la manera en que se ajustan las variables en el tiempo. Las más conocidas son las transformaciones de Koyck, los modelos de Expectativas Adaptativas y los modelos de Ajuste Parcial. Ellos igualmente permiten transformar modelos de rezagos distribuidos en modelos autorregresivos.

Los modelos de rezagos distribuidos demandan estimar una mayor cantidad de parámetros, son más dados a la multicolinealidad y reducen los grados de libertad. Por eso, para la estimación preferimos los modelos autorregresivos multivariantes dado que tenemos que estimar menos parámetros y ganamos en grados de libertad. Sin embargo, para el análisis de las relaciones dinámicas entre las variables preferimos los modelos de rezagos distribuidos, esto es, preferimos la solución de la ecuación en diferencias pues es donde podemos examinar las funciones impulso respuesta.

En la práctica, muchas veces debemos combinar rezagos de las variables independientes con varios rezagos de la variable dependiente. Así, llegamos a una expresión más general de los modelos dinámicos que se conoce como **modelo autorregresivo de rezagos distribuidos o modelos ARDL** (por sus siglas en inglés). Con una sola variable independiente quedaría:

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + \dots + c_k x_{t-k} + u_t \quad (19.6)$$

A no ser que contemos con un modelo teórico que nos indique la especificación dinámica que tenemos que considerar, en la práctica el econométrico debe elegir  $p$  y  $q$  en función de los datos disponibles de las variables. Una vez seleccionada la cantidad de rezagos se estima el modelo, con el cual se pueden efectuar predicciones y analizar las relaciones dinámicas en las funciones impulso respuesta.

Si finalmente se estima un modelo de rezagos distribuidos, las funciones impulso respuesta se obtienen directamente de la regresión. Pero si estimamos un modelo autorregresivo o un modelo **ARDL**, entonces tenemos que construir las funciones impulso respuesta realizando las transformaciones necesarias para llegar a la solución de la ecuación en diferencias.

Hasta aquí hemos visto la formulación de diferentes tipos de modelos multivariantes dinámicos. Nótese que nos hemos mantenido con una sola ecuación. Así, la función de transferencia no es más que el procedimiento de estimación de los diferentes casos de modelos dinámicos uniecuacionales vistos anteriormente.

Más adelante en esta parte cinco del libro veremos que la función de transferencia no es la única manera de estimar y analizar relaciones dinámicas entre las series temporales, hay otros modelos más complejos que se pueden usar en determinadas circunstancias. Adelantándonos un poco a los contenidos que restan en esta parte del libro, podemos apuntar que la estimación de las funciones de transferencia no considera las relaciones de cointegración que veremos en el capítulo

20, además, se desarrolla en un ambiente uniecuacional a diferencia de los modelo de Vectores Autorregresivos (VAR) que estudiaremos en el capítulo 21.

## 19.2 Estimación de una Función de Transferencia

Hay tres elementos que distinguen a las funciones de transferencia con relación a otros modelos multivariantes de series temporales:

- Son siempre uniecuacionales.
- Se supone que las variables a la derecha de la ecuación son exógenas. De hecho, este supuesto es el que nos posibilita mantenernos con una sola ecuación.
- Las series se incluyen en su transformación estacionaria.

La formulación de una función de transferencia con una sola variable independiente podemos escribirla como:

$$y_t = \delta + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + c_0 x_{t-b} + c_1 x_{t-b-1} + \dots + c_d x_{t-b-s} + u_t + \beta_1 u_{t-1} + \dots + \beta_q u_{t-q} \quad (19.7)$$

Mantenemos en la formulación los rezagos de la variable dependiente (términos autorregresivos de orden  $p$ ) y los rezagos de la variable independiente, pero añadimos una estructura de medias móviles de orden  $q$  que pudiera estar presente o no en los residuos.

El número de rezagos de la variable independiente que anteriormente habíamos denotado como  $k$  ahora los descomponemos en  $b+s$ , donde  $b$  representa el número de períodos que transcurre antes de que  $x_t$  comience a tener efecto sobre la variable  $y_t$ . Si se encuentra que  $b \neq 0$  entonces se dice que  $x_t$  es un **indicador adelantado**. En este caso, los valores de  $x_t, x_{t-1}, x_{t-2}, \dots$  pueden usarse para predecir la trayectoria futura de  $y_t$ .

Una formulación más compacta basándonos en el operador de rezagos queda como sigue:

$$A(L)y_t = \delta + C(L)x_{t-b} + B(L)u_t \quad (19.8)$$

donde

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

$$C(L) = 1 - c_1 L - c_2 L^2 - \dots - c_s L^s$$

$$B(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$$

La función impulso respuesta de  $y_t$  a los shocks en  $x_t$  vendría dada por  $C(L)/A(L)$ . Un shock en  $x_t$  empezaría a impactar a  $y_t$  después de  $b$  períodos. Si

$b=0$  entonces el impacto sería inmediato, es decir, existiría una relación contemporánea. Si  $s \neq 0$  o  $p \neq 0$  entonces el shock se propagaría hacia el futuro. Si  $p=0$  entonces la respuesta completa de  $y_t$  ante un shock en  $x_t$  sería finita y tomaría exactamente  $b+s$  períodos.

Considerando ahora que tanto la variable dependiente como la independiente puedan ser no estacionarias reescribimos (19.8) como:

$$A(L)\Delta^{d_1}y_t = \delta + C(L)\Delta^{d_2}x_{t-b} + B(L)u_t \quad (19.9)$$

donde  $\Delta^d = (1-L)^d$ , siendo  $d_1$  es el orden de integración de la variable dependiente y  $d_2$  el orden de integración de la variable independiente, los cuales pueden o no coincidir.

Al igual que en el procedimiento ARIMA, el primer paso para estimar una función de transferencia es determinar el orden de integración de las variables. Una vez contrastada la presencia de raíces unitarias se diferencian las series tantas veces como sea necesario para transformarlas en estacionarias.

El segundo paso también tiene una similitud con la metodología Box-Jenkins. Debemos identificar  $b$ ,  $s$  y  $p$  antes de pasar a estimar los parámetros. En el procedimiento ARIMA nos apoyábamos en las funciones de autocorrelación y autocorrelación parcial para tener una primera impresión de la dependencia de la serie con su pasado. Ahora, en correspondencia con el ambiente multivariante en el que estamos trabajando, debemos identificar la dependencia temporal entre la variable dependiente y la/s variable/s independientes. Para ello nos apoyamos en la **función de correlación cruzada (FCC)**.

En los capítulos 10 y 15, para computar la función de autocorrelación dividíamos la autocovarianza entre la varianza del proceso para llegar al coeficiente de autocorrelación en distintos rezagos u órdenes. Ahora, para computar la función de correlación cruzada, lo que tenemos en el numerador de la expresión es la covarianza cruzada, esto es, la covarianza entre la variable dependiente  $y_t$  y sucesivos rezagos de la variable independiente  $x_t$ :

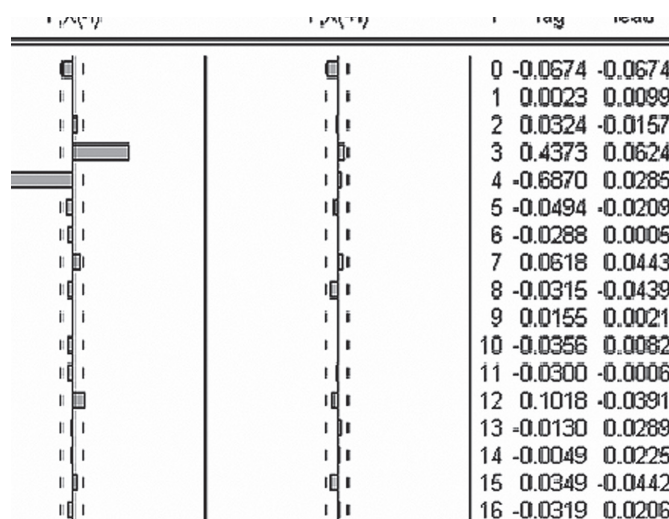
$$\rho_{yx}^{(-i)} = \frac{\text{covarianza cruzada entre } y_t \text{ y } x_{t-i}}{\sigma_y \sigma_x} = \frac{\sum (y_t - \bar{y})(x_{t-i} - \bar{x}) / k}{\sigma_y \sigma_x}$$

Al dividir la covarianza cruzada por las desviaciones estándar obtenemos el coeficiente de correlación cruzado que denotamos como  $\rho_{yx}^{(-i)}$  y que mide la correlación entre  $y_t$  y  $x_{t-i}$ . Los consecutivos coeficientes de correlación cruzados forman la función de autocorrelación cruzada.

Se puede además computar la función de correlación cruzada para  $i > 0$ , para así observar la correlación entre  $y_t$  y sucesivos adelantos de  $x_t$  ( $x_{t+i}$ ). Los paquetes estadísticos usualmente presentan el correlograma cruzado con la correlación

cruzada de la variable dependiente  $y_t$  y los rezagos y adelantos de  $x_t$ . Un ejemplo de ello se muestra en la figura 19.2.

Figura 19.2: Correlograma cruzado



Fuente: Elaboración propia, series simuladas

En la figura 19.2 mostramos el correlograma hasta  $i=16$ . Para  $i=0$  las correlaciones siempre coincidirán en las dos columnas, indicando la correlación contemporánea entre las series; este sería el coeficiente de correlación simple. Para  $i>0$  las correlaciones no tienen que coincidir, la función de correlación cruzada no es simétrica. Correlaciones significativas en la columna izquierda del correlograma indican que los cambios en  $x$  se adelantan a los cambios en  $y$ . Correlaciones significativas en la columna derecha indican que los cambios en  $x$  suceden después que los cambios en  $y$ . En principio, esperamos que estas últimas sean no significativas. En la especificación de la función de transferencia (19.7) ubicamos a  $y_t$  en función de los rezagos de  $x_t$  y no de sus adelantos.

En el ejemplo de la figura 19.2 se observan correlaciones significativas entre  $y_t$  y  $x$  en los rezagos 3 y 4. La correlación entre  $y_t$  y  $x_{t-3}$  es de 0.437. La correlación entre  $y_t$  y  $x_{t-4}$  es de -0.687. Un shock positivo en  $x_t$  impacta a la variable dependiente tres meses después de manera positiva y cuatro meses después de manera negativa.

Ahora bien, ¿cómo podemos identificar  $b$ ,  $s$  y  $p$  a partir del correlograma cruzado? Como sabemos, en la práctica nunca dispondremos del correlograma del proceso estocástico subyacente sino sólo de estimaciones del mismo a partir de los datos disponibles de  $y_t$  y  $x_t$ . Vamos entonces a mostrar algunos patrones teóricos que puede seguir la FCC en dependencia de la relación dinámica que presenten  $y_t$  y  $x_t$ . Vamos a aprender a leer la información que nos proporciona un correlograma

cruzado, conociendo de antemano que en la práctica la *FCC* no coincide exactamente con los patrones teóricos. La tarea del investigador es buscar que modelo teórico se aproxima más al correlograma estimado por el paquete estadístico.

### 19.2.1 Identificación de $b$ , $s$ y $p$

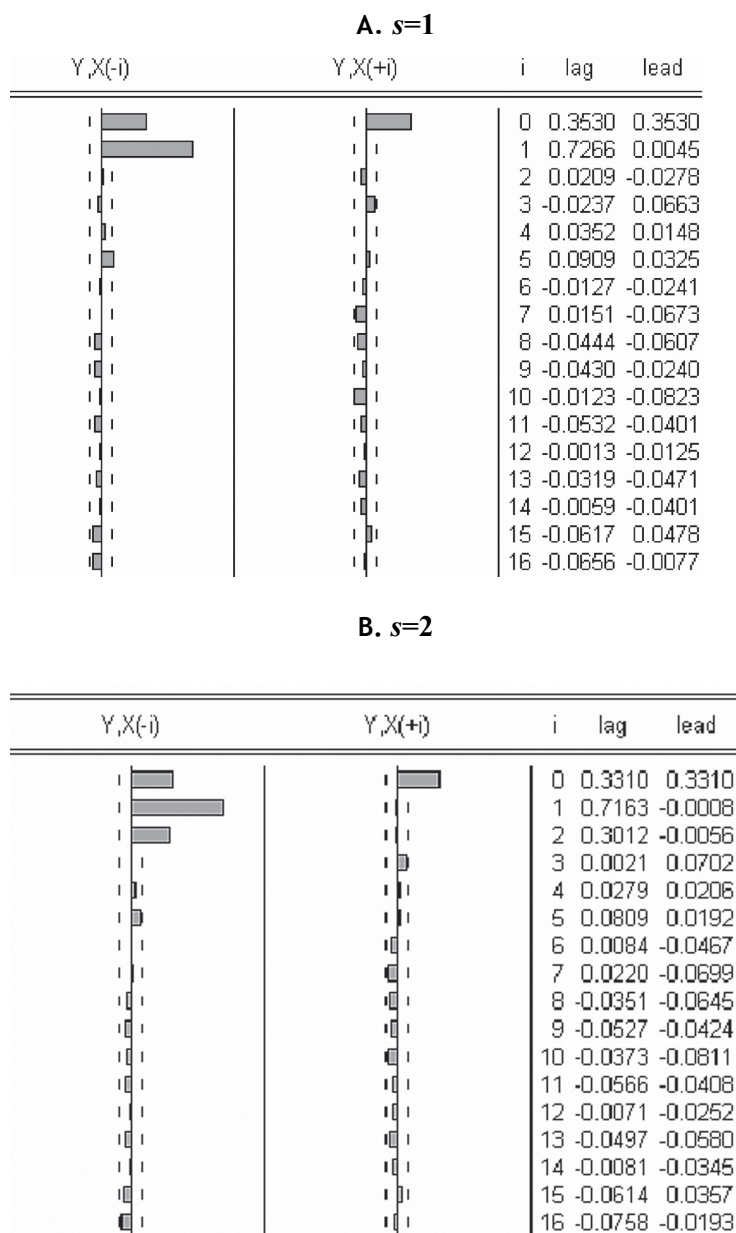
Dada la especificación de la función de transferencia, en la cual  $y_t$  está en función de los rezagos de  $x_t$  y no de sus adelantos, esperamos que las correlaciones significativas estén en la columna izquierda del correlograma cruzado donde se ubican los  $\rho_{yx}(-i)$ . Si por el contrario encontrásemos  $\rho_{yx}(+i)$  significativos en diferentes adelantos, estos podrían ser una indicación de que nuestro modelo está mal especificado. Puede ser un indicio de que los rezagos de  $y_t$  explican a  $x_t$ , es decir,  $x_t$  no es exógena, tal y como habíamos supuesto. Los modelos VAR que estudiaremos en el capítulo 21 podrían ser entonces una mejor especificación para estimar las relaciones dinámicas, o tal vez haya que considerar a  $x_t$  como variable dependiente y a los rezagos de  $y_t$  como variables independientes.

En el caso que se cumpla la relación de causalidad inicialmente especificada en (19.7), esto es, que sean únicamente significativos los  $\rho_{yx}(-i)$ , entonces en la columna izquierda del correlograma cruzado tenemos toda la información que necesitamos para identificar  $b$ ,  $s$  y  $p$ .

De forma general podemos guiarnos por los siguientes patrones en la *FCC*:

- El valor de  $b$  se determina por el número de los primeros  $\rho_{yx}(-i)$  iguales a cero.
- El valor de  $s$  se determina por el número de  $\rho_{yx}(-i)$  distintos de cero menos uno, antes de que se anule o comience a decrecer la *FCC*.
- El valor de  $p$  se determina por el patrón de comportamiento de la *FCC* a partir de los rezagos mayores a  $b+s$  (similar a como se identifica el orden de los AR en un modelo ARIMA). Si la *FCC* se anula para los  $i > b+s$ , entonces  $p$  es igual a cero.

En la figura 19.3 se muestran dos correlogramas cruzados con  $b=0$  dado que la correlación contemporánea es significativa y con  $p=0$  dado que la *FCC* se anula a partir del último rezago significativo. En el caso A se tienen dos  $\rho_{yx}(-i)$  distintos de cero por lo tanto  $s=1$ . En el caso B se tienen tres  $\rho_{yx}(-i)$  distintos de cero por lo tanto  $s=2$ .

**Figura 19.3: Correlograma cruzado con  $b=0$  y  $p=0$** 

Fuente: Elaboración propia, series simuladas

Si nuestros datos de las variables  $y_t$  y  $x_t$  presentasen un correlograma como el de la figura 19.3.A entonces deberíamos pasar a estimar el siguiente modelo:

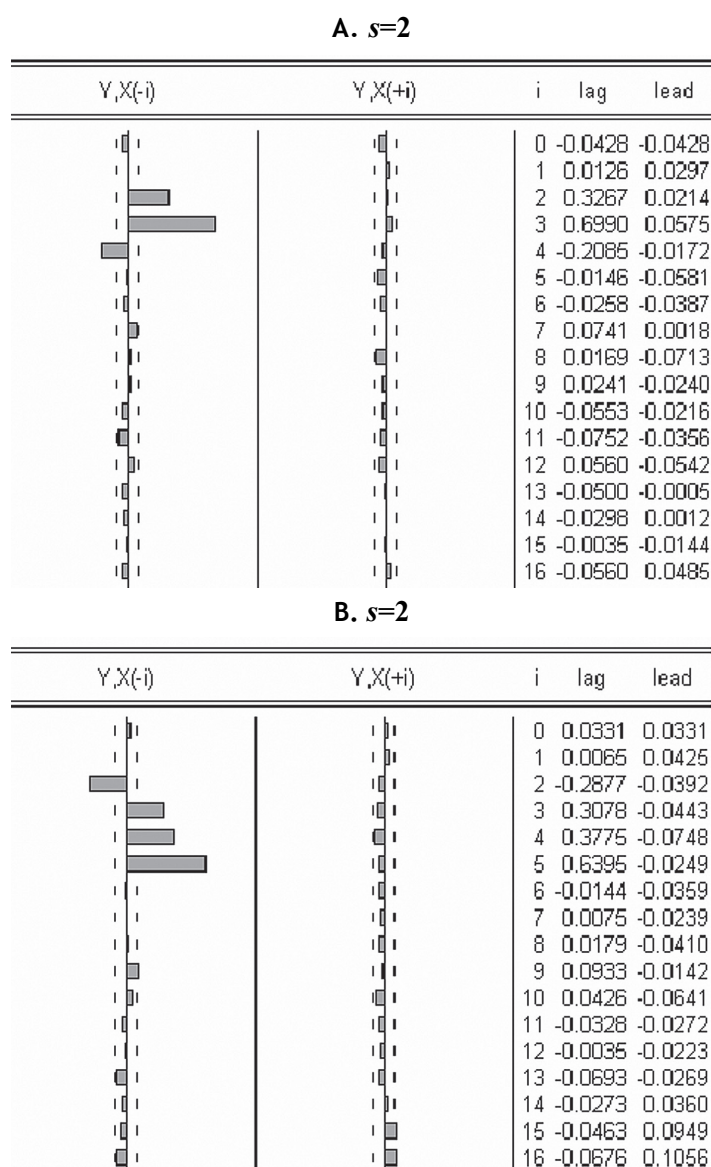
$$y_t = \delta + c_0 x_t + c_1 x_{t-1} + u_t$$

Si nuestro correlograma fuese como el de la figura 19.3.B, entonces deberíamos estimar la función de transferencia con la siguiente especificación:

$$y_t = \delta + c_0 x_t + c_1 x_{t-1} + c_2 x_{t-2} + u_t$$

En la figura 19.4 se muestran dos correlogramas cruzados con  $b=2$  dado que la correlación es significativa a partir del tercer rezago y con  $p=0$  dado que la FCC se anula a partir del último rezago significativo. En el caso A se tienen tres  $\rho_{yx}(-i)$  distintos de cero por lo tanto  $s=2$ . En el caso B se tienen cuatro  $\rho_{yx}(-i)$  distintos de cero por lo tanto  $s=3$ . Ambos correlogramas cruzados sugieren que la variable  $x_t$  es un indicador adelantando de  $y_t$ .

Figura 19.4: Correlograma cruzado con  $b=2$  y  $p=0$



Fuente: Elaboración propia, series simuladas



El modelo sugerido por el correlograma cruzado de la figura 19.4.A sería:

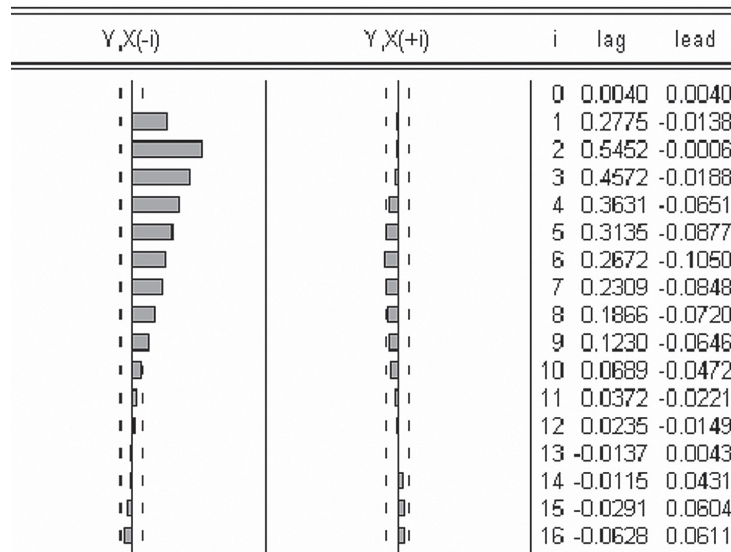
$$y_t = \delta + c_2x_{t-2} + c_3x_{t-3} + c_4x_{t-4} + u_t$$

Mientras que el sugerido por el correlograma cruzado de la figura 19.4.B sería:

$$y_t = \delta + c_2x_{t-2} + c_3x_{t-3} + c_4x_{t-4} + c_5x_{t-5} + u_t$$

En la figura 19.5 se muestra un correlograma cruzado generado con  $b=1$ ,  $s=1$  dado que se tienen dos  $\rho_{yx}(-i)$  distintos de cero antes de que comienza a decrecer la FCC. La FCC no se anula a partir del tercer rezago sino que decrece de forma exponencial, por tanto, podemos interpretar que  $p=1$ .

Figura 19.5: Correlograma cruzado con  $b=1$ ,  $s=1$  y  $p=1$



Fuente: Elaboración propia, series simuladas

En la figura 19.6 se tiene el mismo caso  $b=1$  y  $s=1$  pero con  $p=2$ , lo cual hace que la FCC decrezca con un patrón sinusoidal.

La función de transferencia que nos propone estimar el correlograma de la figura 19.5 es:

$$y_t = \delta + \alpha_1 y_{t-1} + c_1 x_{t-1} + c_2 x_{t-2} + u_t$$

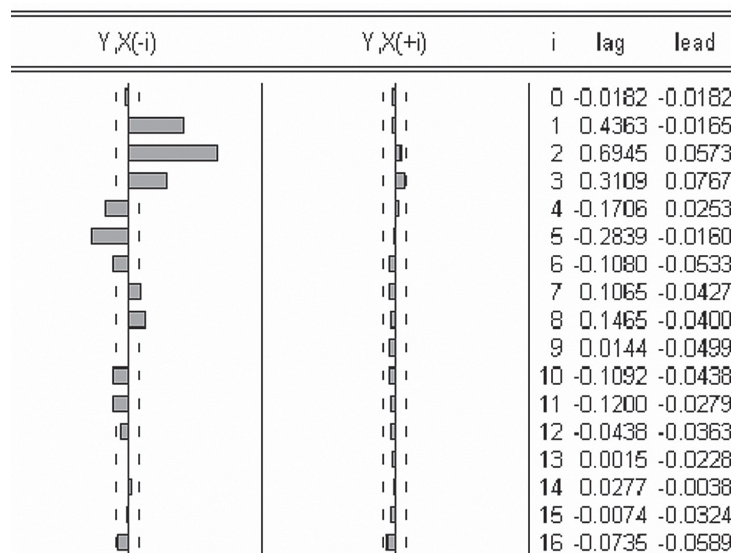
Por su parte, el correlograma de la figura 19.6 nos propone estimar:

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + c_1 x_{t-1} + c_2 x_{t-2} + u_t$$

Para la identificación con el correlograma cruzado del orden de  $b$ ,  $s$  y  $p$  hay dos aspectos adicionales que debemos tener presente. Primero, el correlograma tenemos que construirlo con las transformaciones estacionarias de la variable dependiente y la variable independiente. Segundo, el correlograma cruzado pudiera estar “contaminado” por la autocorrelación de la variable independiente. Es de-

cir, la autocorrelación de  $x_t$  distorsiona la FCC entre ella y  $y_t$ . Por tanto debemos **preblanquear** las series antes de computar el correlograma.

**Figura 19.6: Correlograma cruzado con  $b=1$ ,  $s=1$  y  $p=2$**



Fuente: Elaboración propia, series simuladas

Debemos asegurarnos de cumplimentar el siguiente procedimiento antes de pedirle al paquete estadístico que nos muestre el correlograma cruzado:

1. Transformar las series en estacionarias aplicando el número de diferencias que corresponda.
2. Estimar un modelo AR para la variable independiente de manera que recoja toda la autocorrelación. Por ejemplo, podría ser un AR(1):

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + u_t$$

El modelo para preblanquear las series pudiera incluir las variables dummy de intervención que veremos en el próximo acápite.

3. Tomar los residuos del modelo anterior  $u_t$ , que sería la serie  $x_t$  preblanqueada:

$$x_t^p = x_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{t-1}$$

4. Aplicar el mismo filtro a la variable dependiente usando los parámetros estimados en el paso 3. En este ejemplo sería transformar  $y_t$  con el filtro:

$$y_t^p = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 y_{t-1}$$

5. Computar el correlograma cruzado con las series preblanqueadas  $y_t^p$  y  $x_t^p$  donde se identifican  $b$ ,  $s$  y  $p$  de la manera que ya explicamos.

El procedimiento visto hasta el momento puede generalizarse. Recordemos que la función de transferencia puede contener más de una variable independiente, por lo que habría que repetir el procedimiento de identificación de  $b$ ,  $s$  y  $p$  para cada una de ellas. Posteriormente se combinan en un modelo común.

Un factor favorable para la identificación de  $b$ ,  $s$  y  $p$  lo representa el hecho de que la  $FCC$  no depende de  $B(L)$  en la ecuación (19.8). Los términos de medias móviles se identifican en una segunda etapa para la estimación de la función de transferencia que le sigue a la primera etapa en la que identificamos  $b$ ,  $s$  y  $p$ .

En la segunda etapa usamos las mismas herramientas que conocemos para estimar y validar un modelo de regresión lineal. La identificación de  $b$ ,  $s$  y  $p$  con el correlograma cruzado es sólo un punto de partida para empezar a aplicar en la función de transferencia el procedimiento estándar de estimación por MCO de los parámetros, la inferencia estadística, la validación de los supuestos y la predicción.

Además de los rezagos de la variable dependiente, las variables independientes y sus rezagos, y los términos de medias móviles, una función de transferencia puede requerir incluir variables de intervención. Sobre este asunto trata la última sección del capítulo.

### 19.3 Análisis de Intervención

Este acápite es una continuidad de la sección 7.2 donde se expusieron las variables dicotómicas. Ahora vamos a extender el uso de variables cualitativas o dummy para corregir observaciones atípicas en modelos de series de tiempo. Ello puede aplicarse no sólo a las funciones de transferencia que venimos estudiando sino también a los ARIMA y a otros modelos de series de tiempo que nos faltan por examinar en lo que resta de esta parte del libro.

Normalmente, las series de tiempo se encuentran influenciadas por sucesos extraordinarios como cambios de política económica, guerras, crisis económicas, huelgas, modificaciones en la definición de los datos, etc., que provocan movimientos bruscos sobre su comportamiento. Como consecuencia de estos hechos extraordinarios aparecen datos anómalos, inconsistentes con el resto de la serie. A estos datos anómalos los denominaron Box y Tiao (1975) **observaciones atípicas** o **outliers**.

Si las observaciones atípicas no son detectadas y corregidas por el investigador, estas pueden tener un impacto negativo sobre el proceso de modelización de las series de tiempo y sobre la predicción. La presencia de atípicos en la serie puede ocasionar varios problemas:

- Alteran la estructura de los estadísticos que se usan en la especificación de un modelo. La función de autocorrelación ( $ACF$ ), la función de autocorrelación parcial ( $PACF$ ) y la  $FCC$  de una serie temporal contaminada por la presencia de observaciones atípicas puede conducir a errores de

especificación importantes en el modelo o sesgos en las estimaciones de los parámetros.

- La precisión de las predicciones se deteriora debido al sesgo en la estimación de los parámetros del modelo. Los errores de predicción son mayores cuando las observaciones atípicas ocurren en las últimas observaciones del período muestral.
- Puede distorsionar los contrastes de raíz unitaria.
- No se cumple el supuesto clásico de normalidad en los errores. En consecuencia, se afectan las pruebas estadísticas  $t$  y  $F$  pudiendo estas sugerir erróneamente que el modelo es significativo, o lo contrario.

Debido a los problemas que causan las observaciones atípicas, todo investigador empírico debe siempre considerar, como uno de los pasos en el proceso de estimación, la detección y corrección de las posibles observaciones atípicas.

Cuando se logra detectar observaciones atípicas en una serie, se debe comenzar siempre por examinar si la causa es un error en la transcripción de los datos o algún error en la construcción de los datos publicados que puede ser subsanado. Si la causa de la anomalía no son errores subsanables en los datos, habrá que indagar si es debida a algún acontecimiento extraordinario en particular.

En la literatura existen determinados medios para lidiar con las observaciones atípicas, estos se resumen en lo que se conoce como *análisis de intervención*. El mismo fue inicialmente desarrollado por Box y Tiao (1975). En esencia, el análisis de intervención constituye un procedimiento para identificar, analizar y estimar acontecimientos extraordinarios en la modelación de series temporales.

La identificación y estimación de outliers en las series de tiempo constituye una herramienta útil para comprender por qué una serie se comporta de determinada manera. La presencia de outliers en una serie debe ser entendida como una oportunidad para aprender acerca de aspectos esenciales del fenómeno que se está analizando.

El análisis de intervención parte de que existen diferentes tipos de acontecimiento que provocan outliers. En correspondencia, se clasifican cuatro tipos fundamentales de observaciones atípicas:

- **Atípico aditivo (AO, additive outlier):** Es un evento que afecta a la serie en una única observación, que toma un valor más grande o más pequeño que lo esperado en una fecha puntual. Después de esta perturbación la serie regresa a su comportamiento anterior como si nada hubiera pasado.
- **Cambio de nivel (LS, level shift):** Es una perturbación que cambia el nivel (o la media) de la serie a partir de un punto determinado de la misma, o sea, esta perturbación modifica el valor de todas las observaciones de la serie desde el momento en que ocurre el atípico.
- **Cambio transitorio (TC, transitory change):** Es un suceso cuyo efecto sobre la serie observada es temporal. Se dice que es una generalización de AO y LS en el sentido de que causa un impacto inicial como un AO pero su

efecto también se extiende a las observaciones siguientes. El impacto de un TC no es permanente, se va diluyendo gradualmente.

- **Cambio de pendiente (SC, slope change):** Es un acontecimiento que cambia el ritmo de crecimiento de la serie a partir de un determinado momento. Modifica la pendiente del componente de tendencia de la serie desde el momento en que ocurre el atípico.

Así, para enfrentar las observaciones atípicas no basta con detectar que nuestra serie esté afectada por las mismas, sino que se debe también identificar qué tipo de acontecimiento está determinando el comportamiento anómalo. Hay que seleccionar entre observaciones atípicas AO, LS, TC y SC.

En el proceso de estimación, las observaciones atípicas que más nos interesan son las que están asociadas a la variable dependiente. Si nuestra variable dependiente presenta valores atípicos, entonces la podemos denotar con un supraíndice  $a$ :  $y_t^a$ . Así, en un modelo ARIMA o una función de transferencia por ejemplo veremos que los residuos también van a presentar valores atípicos:

$$y_t^a = \delta + \alpha_1 y_{t-1} - \beta_1 u_{t-1} + u_t^a$$

$$y_t^a = \delta + \alpha_1 y_{t-1} + c_1 x_{t-1} + u_t^a$$

Si existen observaciones atípicas en  $y_t^a$  entonces estas no serán recogidas en la parte sistemática del modelo y aparecerán en la parte aleatoria  $u_t^a$ . Si la variable dependiente presenta observaciones atípicas observaremos valores “grandes” en la serie de los residuos del modelo. Se plantea que aquellos residuos que alcancen un valor de más de 3 veces el valor de su desviación típica pueden ser considerados outliers. Estos podrán ser negativos o positivos en dependencia del evento que afectó a la serie. En un período muestral determinado es posible encontrar más de un residuo que sobrepasa 3 veces la desviación estándar.

Uno de los supuestos que se han formulado sobre los residuos es que se distribuyen normalmente. En una distribución Normal es muy poco probable encontrar valores que sobrepasan 3 veces la desviación estándar. Otro elemento es que cuando hay residuos positivos “grandes” la media de los residuos tiende a ser mayor a la mediana. La distribución de probabilidad de los residuos tiene una asimetría derecha (right skewed). En cambio, cuando hay residuos negativos “grandes” la media de los residuos tiende a ser menor a la mediana. La distribución de probabilidad de los residuos tiene una asimetría izquierda (left skewed). En ambos casos la distribución de probabilidad de los residuos, al no ser simétrica, no es Normal.

Por lo anterior es que se utiliza el histograma de los residuos para detectar la presencia de observaciones atípicas. Como sabemos, el histograma no es más que un gráfico de distribución de frecuencia, donde se representa por el eje horizontal los intervalos de clases de la serie con bandas verticales y por el eje vertical la frecuencia relativa. Cuando existen datos que están muy alejados del resto de la serie, resulta muy difícil incluirlos en algún intervalo de clase. En estos casos, los

outliers deben tratarse como casos especiales y estarán distanciados del resto de las clases que se representen en el histograma. O sea que, si se observa en el histograma valores de los residuos que se alejan mucho del resto de las clases se puede deducir que la serie presenta observaciones atípicas.

Una forma de contrastar que existen observaciones atípicas, es precisamente realizar la prueba de normalidad en los residuos a partir del estadístico Jarque-Bera que estudiamos en el capítulo 11. Así, si la probabilidad asociada al estadístico es menor que el alfa prefijado, los errores no se distribuyen normalmente y la serie puede estar afectada por observaciones atípicas. No obstante, en el capítulo 11 vimos que la no normalidad en los residuos puede tener otros orígenes; esto se debe tener presente.

Ahora bien, una vez que se conoce que la serie está afectada por observaciones atípicas, se debe identificar el período de comienzo de cada una de ellas en el gráfico de los residuos del modelo estimado. Un simple gráfico de línea de los residuos  $u_t^a$  sirve para localizar la fecha de las observaciones atípicas dado que apreciaremos valores excesivamente “grandes” con signo positivo o negativo.

Luego de localizadas las observaciones atípicas, se debe pasar a recoger información sobre el tipo de efecto que tienen sobre la serie en cuestión, en otras palabras, ya sea AO, LS, TC o SC. Ello se realiza como parte del procedimiento de estimación de las observaciones atípicas.

Para la estimación de las observaciones atípicas se utilizan las llamadas **variables de intervención**, las cuales son variables cualitativas que se incluyen en los modelos para recoger los efectos de los eventos extraordinarios en la variable dependiente. En el modelo se deben incluir variables de intervención hasta tanto los residuos se distribuyan normalmente y no presenten valores más allá de 3 veces la desviación estándar. En la literatura se recomienda manejar como límite máximo de variables de intervención el 5% del total de observaciones.

En los ejemplos que mostramos de modelo ARIMA y función de transferencia, debemos ver que los residuos no presenten observaciones atípicas una vez que incluimos las variables de intervención ( $I_t^1 + I_t^2 + \dots I_t^j$ ):

$$y_t^a = \delta + \alpha_1 y_{t-1} - \beta_1 u_{t-1} + \omega_1 I_{t1}^1 + \omega_2 I_{t2}^2 + \dots \omega_m I_{tm}^m + u_t$$

$$y_t^a = \delta + \alpha_1 y_{t-1} + c_1 x_{t-1} + \omega_1 I_{t1}^1 + \omega_2 I_{t2}^2 + \dots \omega_m I_{tm}^m + u_t$$

El investigador debe determinar las distintas fechas  $t1, t2, \dots, tm$  en las que parece haber observaciones atípicas. Se construyen seguidamente variables dummy de intervención para cada fecha y se introducen en el modelo junto con el resto de las variables independientes. Finalmente se estiman los coeficientes  $\omega_1, \omega_2, \dots, \omega_m$  asociados a cada variable de intervención. Para comprobar que se han identificado correctamente las observaciones atípicas se deben efectuar contrastes de significación individual de cada uno de estos coeficientes.

Cada uno de los  $\omega_j$  arroja un valor estimado del efecto que tuvo sobre la variable dependiente el evento extraordinario de cada fecha específica. Los  $\omega_j$  proveen información realmente relevante para el análisis del fenómeno en estudio.

Para construir las variables de intervención hay que tomar en cuenta, además de la fecha en que comienza el evento extraordinario, el tipo de observación atípica que corresponde (AO, LS, TC o SC). A continuación, se muestran los diferentes tipos de variables de intervención que se usan; en la figura 19.7 se grafica cada caso:

- La variable de intervención tipo AO se traduce en una variable que contiene un 1 en ese período y 0 en el resto.
- La variable de intervención tipo LS contiene 0 hasta el momento en que ocurre la perturbación y 1 en adelante.
- La tercera variable de intervención tipo TC se construye como una variable que toma valor 0 hasta el momento en que ocurre la perturbación, en ese momento la variable toma valor 1 y luego toma valores de  $\delta^*x_{t-1}$ , donde  $x_{t-1}$  es la observación anterior de la propia variable de intervención y  $\delta$  es un parámetro, con valores dentro del intervalo  $0 < \delta < 1$ , que determina la velocidad con que el efecto del atípico desaparece en el nivel de la serie.

Para el atípico aditivo  $\delta = 0$ , lo que significa que desaparece rápido el efecto de la perturbación. El cambio de nivel corresponde a  $\delta = 1$ , o sea, la perturbación tiene un efecto permanente en el nivel de la serie. Se recomienda un valor de  $\delta = 0.7$  para la variable cambio transitorio para que esta tenga una caída suave y para distinguir claramente un cambio transitorio de un atípico aditivo o de un cambio de nivel.

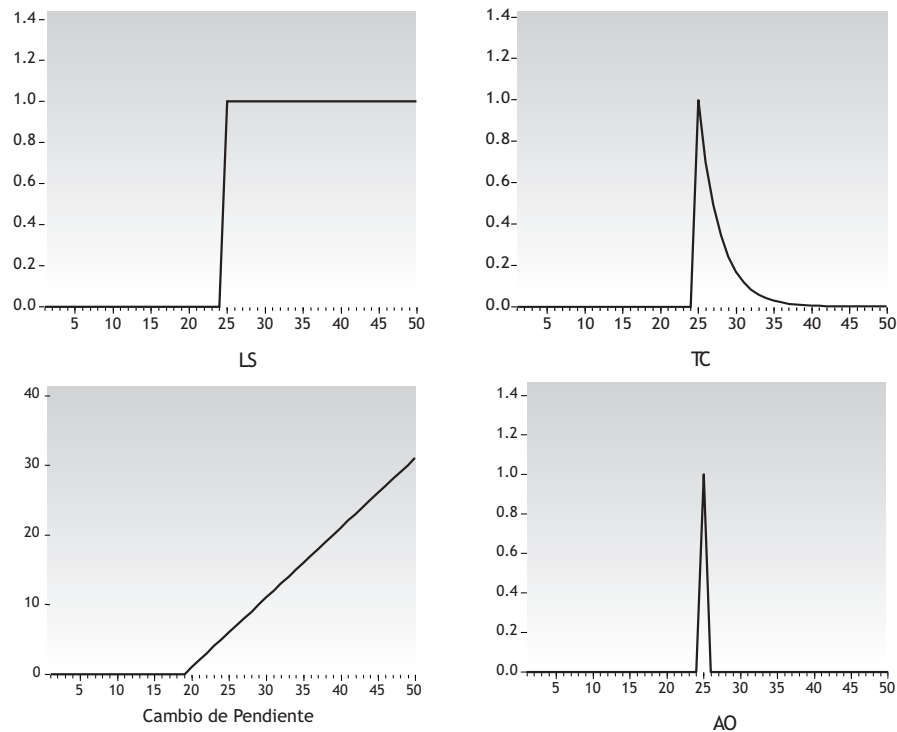
- La variable de intervención tipo SC contiene 0 hasta el momento en que ocurre el acontecimiento extraordinario y a partir del mismo toma sucesivamente los valores 1, 2, 3..., creciendo en forma ascendente.

Apoyándonos en todo lo que hemos examinado sobre el análisis de intervención, resumimos a continuación los pasos que debe seguir el investigador para llevarlo a cabo:

1. Estimación del modelo sin las variables de intervención. Ver el Histograma y la prueba de Jarque- Bera para conocer si existen observaciones atípicas.
2. Hacer un gráfico con los residuos. Detectar la fecha del mayor atípico.

Se construye la variable de intervención AO para esa fecha. Se incluye en el modelo la variable AO con suficientes rezagos. Se excluyen los rezagos no significativos. Si existen rezagos significativos esto es un indicio de que el evento no es puntual sino que tuvo un efecto más allá de la fecha inicial. Si este es el caso se debe valorar la posibilidad de sustituir la variable AO por un LS o por un TC, para ello se utilizan los conocidos criterios de información AIC y SIC.

**Figura 19.7 Variables de intervención**



*Fuente: Elaboración propia*

3. Se continúa el procedimiento con otros atípicos. Es decir, el análisis de intervención no se realiza al mismo tiempo para todos los outliers, sino que se hace de uno en uno empezando por el mayor.
4. Se revisa la especificación del modelo, pues una vez incluidas las variables de intervención este puede cambiar significativamente.

Aquí concluimos este capítulo sobre las funciones de transferencia enriquecido con este último acápite sobre el análisis de intervención, el cual constituye una herramienta muy necesaria para diferentes tipos de modelación de series temporales.



# *preguntas*

Intente dar respuesta a las siguientes preguntas. Así sabrá si ha captado los principales mensajes de este capítulo.

1. ¿Qué es un modelo de rezagos distribuidos?
2. ¿Cuál es la formulación de un modelo autorregresivo multivariante?
3. ¿En una función de transferencia, las series no se deben diferenciar?
4. ¿Cuáles son los tipos fundamentales de observaciones atípicas que podemos encontrar en una serie de tiempo?
5. ¿Para qué se usan las variables de intervención?



# ejercicios

- 19.1 A continuación se presenta el resultado de estimar una función de transferencia que relaciona la demanda trimestral de divisas en las casas de cambio (CADECA) con los montos de salarios pagados (SAL) trimestralmente por las empresas:

$$\Delta \ln CADECA_t = -0.001 + 0.7\Delta \ln CADECA_{t-1} + 0.15\Delta \ln SAL_t + e_t$$

- ¿Qué tipo de modelo multivariante dinámico es este?
  - ¿Podría ser esta una regresión espuria?
  - ¿Los coeficientes estimados están acordes con lo que debemos esperar teóricamente?
  - ¿Cuál es el valor del multiplicador de corto plazo estimado?
  - ¿Cuáles son los valores estimados para los multiplicadores intermedios para el primer y segundo trimestre? Interprete el significado de los mismos.
  - ¿Cuál es el multiplicador de largo plazo? Interprete el significado del mismo.
  - Construya un gráfico con la función impulso-respuesta en un período de dos años.
  - ¿Es el salario un indicador adelantado de la demanda de divisas?
- 19.2 A continuación se presenta el resultado de estimar una función de transferencia que relaciona los datos anuales de la inflación (INFLA) con la brecha entre el salario y la productividad (SALPROD) y las variaciones del tipo de cambio (VTC):

$$INFLA_t = 0.002 + 0.95INFLA_{t-1} + 0.2SALPROD_{t-1} + 0.09VTC_t + 0.18VTC_{t-1} + e_t$$

- ¿Esto podría ser esta una regresión espuria?
  - ¿Cuál variable tiene un impacto más inmediato en la inflación?
  - ¿Cuál variable tiene un mayor impacto en el largo plazo en la inflación?
  - ¿Alguna de las dos variables independientes es un indicador adelantado?
  - ¿Los coeficientes estimados están acordes con lo que debemos esperar teóricamente?
  - Construya un gráfico con las funciones impulso-respuesta en un período de 6 años.
- 19.3 Identifique  $b$ ,  $s$  y  $p$  en los siguientes correlogramas cruzados. ¿Qué especificaciones deberíamos usar para estimar las funciones de transferencia?



CAPÍTULO 20



## COINTEGRACIÓN Y MECANISMO DE CORRECCIÓN DE ERRORES

En el presente capítulo mostramos el desarrollo de una metodología alternativa a los ARIMA y a las funciones de transferencia para la modelación y predicción de series de tiempo. Se puede afirmar que en los últimos años esta ha sido la metodología preferida por los investigadores en los análisis empíricos de disímiles temas económicos. Debido al carácter introductorio del libro, nos referimos básicamente a la metodología de Engle y Granger (1987). La misma permite entender más fácilmente la lógica y conceptos asociados a la cointegración.

Inicialmente se define qué es cointegración estadísticamente y su interpretación en términos económicos. Posteriormente se explica una de las pruebas que pueden ser usadas para contrastar si un grupo de variables están cointegradas, esto es, la prueba ADF sobre los residuos. En un tercer acápite se muestra el modelo conocido como Mecanismo de Corrección de Errores, el cual corresponde utilizar en la estimación y predicción de variables cointegradas; ello, como parte de la estimación bietápica de Engle y Granger.

### 20.1 Definición de Cointegración

En los capítulos previos habíamos visto que la solución que procedía, una vez que encontrábamos que las series de tiempo de nuestro modelo eran no estacionarias, era la diferenciación. De esta forma evitábamos que la no estacionariedad afectara los supuestos sobre el término de error y evadíamos las regresiones espurias. Así lo vimos en la metodología ARIMA y en las funciones de transferencia.

El concepto de cointegración ahora nos enseña que no todos los modelos que mantienen sin diferenciar series no estacionarias son espurios, que la

diferenciación no es la mejor opción si nuestras series de tiempo siguen una relación de cointegración.

La no estacionariedad es una propiedad de tipo dominante, es decir, la combinación lineal de procesos con distinto orden de integración es del mismo orden que el proceso de orden mayor. Por tanto, los residuos de una regresión con variables no estacionarias —que son precisamente una combinación lineal de las variables involucradas— también serían no estacionarios, dejándose de cumplir con ello el supuesto de que son un ruido blanco.

Ahora bien, existe una excepción en esta regla: la combinación lineal de procesos no estacionarios puede ser estacionaria. Precisamente, la excepción de esta regla es lo que conocemos como cointegración. Este concepto fue formulado por Engle y Granger (1987).

De manera más formal, si tenemos un vector  $Y_t (m \times 1)$ , se dice que sus componentes están cointegrados de órdenes  $d$  y  $b$ , y se denota por  $Y_t \rightarrow CI(d, b)$  si:

1. Todos los componentes de  $Y_t$  son integrados del mismo orden  $d$ ,  $I(d)$ .
2. Existe un vector  $\beta$ , no nulo, tal que, la combinación lineal  $\beta' Y_t = u_t$  es integrada de orden  $I(d-b)$  con  $b > 0$ .

Al vector  $\beta$  se le denomina **vector de cointegración**. El caso más común es aquel donde  $d = b = 1$ . Esto es, todos los elementos de  $Y_t$  serán  $I(1)$  y para que estén cointegrados  $u_t$  debe ser  $I(0)$ , es decir, estacionario.

En un modelo uniecuacional de dos variables  $y_t = \beta_1 x_t + u_t$ , que también puede expresarse como:

$$\begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix} (y_t \ x_t) = u_t$$

podemos decir que  $y_t$  y  $x_t$  están cointegradas cuando las dos variables son  $I(1)$  y los residuos  $u_t$  son  $I(0)$ . Es decir, la combinación lineal de las dos variables no estacionarias  $(y_t - \beta_1 x_t)$  produce un proceso estacionario. El vector de cointegración es precisamente  $\begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix}$ .

Anteriormente habíamos estudiado que para estimar la relación entre dos variables  $I(1)$  debíamos especificar el modelo en diferencias para que no fuera espurio,

$$\Delta(y_t) = a_1 \Delta(x_t) + u_t$$

Lo que estamos diciendo ahora es que podemos estimar el modelo con las variables en niveles,

$$y_t = \beta_1 x_t + u_t$$

siempre y cuando estas estén cointegradas. La parte sistemática de este modelo se conoce como **relación de cointegración**.

Económicamente, la relación de cointegración se interpreta como una relación de equilibrio de largo plazo. En el vector de cointegración están los parámetros de esta relación, si el modelo se plantea en logaritmo, el vector  $\beta$  tiene las **elasticidades de largo plazo**. El término de error  $u_t$  recoge entonces las posibles desviaciones que puedan darse con respecto a este equilibrio. Como  $u_t$  es estacionario las desviaciones son transitorias. Así, los shocks van a tener un efecto permanente sobre las variables no estacionarias  $y_t$  y  $x_t$ , pero si estas están cointegradas, los shocks van a tener un impacto transitorio en la relación de largo plazo.

El concepto de cointegración reviste una particular importancia por el hecho de que la teoría económica en la mayoría de los casos se refiere a relaciones de equilibrio entre determinadas variables, por ejemplo:

Función de consumo:  $Consumo_t = \beta_0 + \beta_1 Ingreso_t$

Demanda de dinero:  $AgregadoMonetarioReal_t = \beta_0 + \beta_1 Ingreso_t + \beta_2 TasaInterés_t$

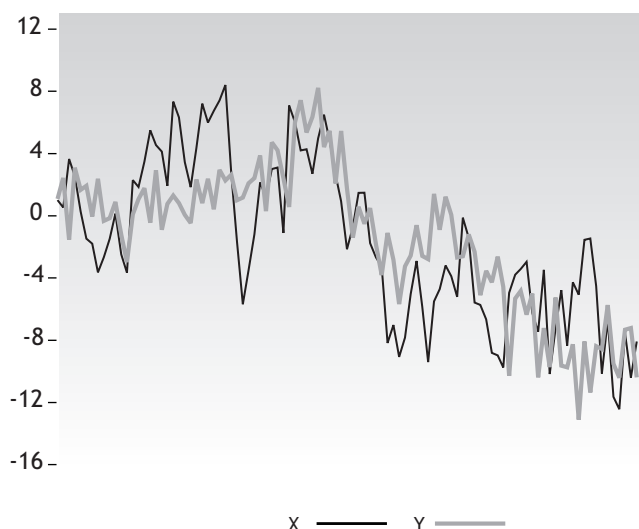
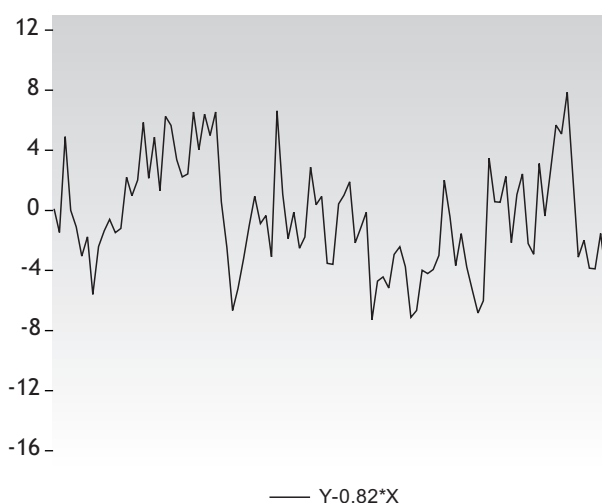
Paridad de intereses:  $InterésInterno_t = \beta_0 + \beta_1 InterésExterno_t + \beta_2 DevaluacionEsperada_t$

Todas las ecuaciones tendrían asociado un término de perturbación  $u_t$  para recoger las variaciones del Consumo, el Agregado Monetario y el Tipo de Interés Interno que no se explican por la función teórica. Pero, de ser cierta la relación propuesta por la teoría económica, las desviaciones de la función teórica sólo deberían ser transitorias; en el largo plazo, el Consumo, el Agregado Monetario y el Tipo de Interés Interno deben tender a la relación de equilibrio propuesta. Para probar entonces que se cumple la teoría habría que contrastar que las variables están cointegradas, es decir, que  $u_t \rightarrow I(0)$ .

En la figura 20.1.A se simulan dos paseos aleatorios cointegrados. Se observa como los shocks afectan permanentemente la media de las variables, pero existe una relación positiva de equilibrio que lleva a que ambas terminen moviéndose en el mismo sentido. En determinados períodos se alejan una de otra, pero esto sólo sucede en el corto plazo, en el largo plazo tienden a moverse juntas. Las desviaciones de este equilibrio de largo plazo se grafican en la figura 20.1.B El vector de cointegración para estas series es:

$$\begin{pmatrix} 1 \\ -0.82 \end{pmatrix}$$

es decir, la relación de cointegración es  $y_t = 0.82x_t$ . Por tanto las desviaciones del equilibrio de largo plazo se computan como  $(y_t - 0.82x_t)$ . Se observa en la figura que queda un proceso estacionario donde los shocks tienen un efecto transitorio.

**Figura 20.1: Variables cointegradas****A. Relación de equilibrio de largo plazo****B. Desviaciones del equilibrio del largo plazo**

Fuente: Elaboración propia, series simuladas

El anterior, por supuesto, es un ejemplo simulado. En la práctica, sólo contamos con los datos disponibles de las variables, con los cuales debemos estimar el vector de cointegración en una regresión de las series en niveles y probar si los residuos son estacionarios. En los siguientes acápites explicamos el proceso práctico de contraste y estimación de la cointegración. Pero antes, debemos añadir algunos otros conceptos relacionados con la definición de cointegración.

Las variables que están cointegradas presentan **tendencias estocásticas comunes**. Volvamos al mismo ejemplo de dos variables  $I(1)$ ,  $y_t$  y  $x_t$ , las cuales están cointegradas con un vector  $(1 - \beta)'$ . Ambas variables se pueden descomponer



en una parte estacionaria y un término que representa la tendencia estocástica común  $w_t$ :

$$\begin{aligned}x_t &= w_t + \tilde{x}_t \\ y_t &= \beta_1 w_t + \tilde{y}_t\end{aligned}$$

donde  $w_t \rightarrow I(1)$ ,  $\tilde{x}_t \rightarrow I(0)$  y  $\tilde{y}_t \rightarrow I(0)$ .

La combinación lineal de estas variables sería:

$$u_t = y_t - \beta_1 x_t = \beta_1 w_t + \tilde{y}_t - \beta_1 (w_t + \tilde{x}_t) = \tilde{y}_t - \beta_1 \tilde{x}_t$$

En este caso  $u_t \rightarrow I(0)$  dado que el parámetro  $\beta_1$  anuló la tendencia estocástica común  $w_t$ , quedando  $u_t$  en función de los procesos estacionarios  $\tilde{y}_t$  y  $\tilde{x}_t$ .

Otro aspecto es que el vector de cointegración no es único, basta con multiplicarlo por cualquier valor distinto de cero para que obtengamos un nuevo vector que también anula la tendencia estocástica común. Por ejemplo, si en vez de  $(1-\beta_1)'$  tomamos  $k(1-\beta_1)'$ , también llegamos a que  $u_t \rightarrow I(0)$ :

$$u_t = ky_t - k\beta_1 x_t = k\beta_1 w_t + k\tilde{y}_t - k\beta_1 (w_t + \tilde{x}_t) = k\tilde{y}_t - k\beta_1 \tilde{x}_t$$

Ahora bien, si normalizamos en  $y_t$ , es decir, fijamos que el primer elemento del vector de cointegración tiene que ser igual a 1, entonces el vector de cointegración sí es único para el caso de dos variables, cualquier otro nos daría como resultado un proceso  $u_t$  no estacionario. Por ejemplo, si en vez de considerar a  $(1-\beta_1)'$  tomamos  $(1-\alpha_1)'$ , entonces:

$$u_t = y_t - \alpha_1 x_t = \beta_1 w_t + \tilde{y}_t - \alpha_1 w_t - \beta_1 \tilde{x}_t = (\beta_1 - \alpha_1) w_t + \tilde{y}_t - \alpha_1 \tilde{x}_t$$

La combinación lineal sería estacionaria,  $u_t \rightarrow I(0)$ , si y solo si  $\alpha_1 = \beta_1$ .

El máximo de vectores de cointegración linealmente independientes que podemos obtener de 2 variables cointegradas es 1. En general, el máximo de vectores de cointegración linealmente independientes que podemos obtener de  $m$  variables integradas del mismo orden es  $m-1$ , pues si normalizamos el vector, nos quedarán, como máximo,  $m-1$  vectores de cointegración distintos. Al número de vectores de cointegración linealmente independientes se le denomina **rango de cointegración**.

## 20.2 Contraste de Cointegración sobre los Residuos

Existen varias formas para probar la existencia de una relación de cointegración entre un conjunto variables. En este texto utilizaremos una muy sencilla que consiste en probar si los residuos de la regresión de cointegración presentan un

orden de integración menor que el de las variables implicadas. En el caso más común, todas las variables son  $I(1)$  y probaríamos que la combinación lineal resultante de ellas es  $I(0)$ , es decir, que los residuos son estacionarios.

Para contrastar la estacionariedad de los residuos aplicamos la prueba de Dickey Fuller ampliado (ADF) que estudiamos previamente en la sección 14.2, sólo que en este caso, en vez de aplicarla sobre variables, lo hacemos sobre una serie de residuos. Esto es, aplicamos la prueba ADF sobre el término de error resultante de la regresión en niveles de las variables que estamos probando si siguen o no una relación de cointegración.

En resumen, el contraste de cointegración de dos variables  $y_t$  y  $x_t$  contendría los siguientes pasos:

1. Inicialmente debemos probar que ambas variables presentan el mismo orden de integración. Aplicamos la prueba ADF tanto a  $y_t$  como a  $x_t$ . Si presentan distintos órdenes, por ejemplo  $y_t \rightarrow I(1)$  y  $x_t \rightarrow I(2)$ , entonces se puede concluir aquí el procedimiento, puesto que por la primera parte de la definición de cointegración veíamos que se requiere que todas las variables sean integradas del mismo orden. Otro caso es que ambas variables fuesen  $I(0)$  en el cual tampoco se seguiría el procedimiento. No habría una tendencia estocástica común que estimar. En ambos casos procederíamos a utilizar la metodología de función de transferencias que analizamos en el capítulo anterior.

Si en cambio,  $y_t$  y  $x_t$  son no estacionarias y tienen el mismo orden de integración, entonces se procede con los siguientes pasos que constituyen propiamente el contraste de cointegración sobre los residuos.

2. Estimar por Mínimos Cuadrados Ordinarios los parámetros de la regresión en niveles:

$$y_t = \beta_1 x_t + u_t$$

Este modelo puede estimarse con el logaritmo de las series y añadiendo constante y/o tendencia:

$$\ln y_t = \beta_0 + \beta_1 \ln x_t + u_t$$

No se añaden rezagos al modelo, ni es necesario probar los supuestos sobre los residuos o la significatividad estadística de las variables. El signo de los coeficientes sí se debe chequear que concuerde con la teoría económica.

3. Tomar la serie de los residuos  $u_t$  y probar que presenten un orden de integración menor que las variables, en el caso más común, las variables serían  $I(1)$  y contrastaríamos si  $u_t \rightarrow I(0)$ . Aplicamos la prueba Dickey Fuller sobre  $u_t$ , de la misma forma que lo hicimos en la sección 17.2 con diferentes series de tiempo.

Se estimaría el modelo  $\Delta u_t = \delta u_{t-1} + e_t$  y se prueba con el estadístico  $\tau$  (*tau*) si  $\delta$  es significativamente distinto de cero. Las hipótesis nula y alternativa son las siguientes:

$$H_0: \delta = 0 \Rightarrow u_t \text{ no estacionario} \Rightarrow \text{No hay cointegración}$$

$$H_1: \delta < 0 \Rightarrow u_t \text{ estacionario} \Rightarrow \text{Hay cointegración}$$

Usualmente se añaden rezagos al contraste para recoger la autocorrelación, con lo cual estaríamos aplicando la prueba de Dickey Fuller aumentado (ADF). Por otra parte, si la regresión en niveles del paso 2 se estimó con constante, entonces en el contraste ADF no hace falta especificar constante.

Otro punto importante es que los valores calculados del estadístico  $\tau$  (*tau*) en esta ocasión no se comparan con los valores críticos de Dickey-Fuller o de McKinnon (1996). Tales valores críticos no son apropiados para contrastar cointegración. Nótese que  $u_t$  no es una realización directa de un proceso estocástico subyacente, sino que es una serie construida mediante la expresión  $u_t = y_t - \hat{\beta}_1 x_t$  después de estimar  $\hat{\beta}_1$ . Engle y Yoo (1987) tabulan valores críticos para el contraste de cointegración sobre los residuos que toman en cuenta tal aspecto.

Si en el paso 3 no rechazamos la hipótesis nula, al tener valores calculados del estadístico  $\tau$  (*tau*) mayores que los valores críticos de Engle y Yoo (1987), entonces el modelo estimado en el paso 2 es espurio. La prueba indica que  $y_t$  y  $x_t$  no presentan una relación de largo plazo, habrá que estimar el modelo en diferencias, no en niveles.

Si por el contrario en el paso 3 rechazamos la hipótesis nula al tener valores calculados del estadístico  $\tau$  (*tau*) menores que los valores críticos de Engle y Yoo (1987), entonces el modelo estimado en el paso 2 no es espurio. De hecho, el modelo estimado representaría la relación de equilibrio de largo plazo, siendo  $\hat{\beta}_1$  la elasticidad de largo plazo (para el modelo en logaritmos).

Los pasos anteriores pueden fácilmente generalizarse para el caso de más de dos variables.

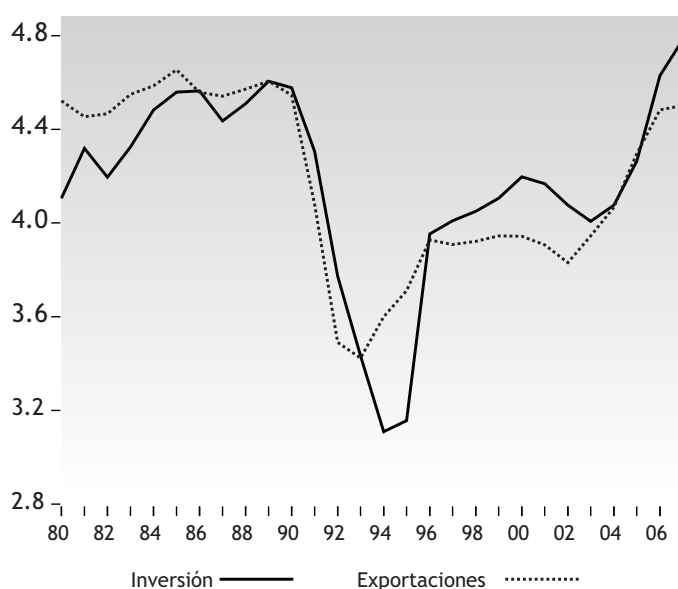
### 20.2.1 Ejemplo del contraste de cointegración

Vamos a desarrollar un ejemplo sencillo, que sirve para ilustrar el procedimiento, aun cuando la cantidad de observaciones disponibles de las series es sólo de 28. Tenemos datos anuales de las inversiones (INV) e importaciones (M) cubanas desde 1980 hasta el año 2007 tomadas como índice (1989=100). El lector puede replicar las estimaciones tomando los datos de la tabla B.12 del CD acompañante. Trabajaremos con las series en logaritmo.

Partimos de la hipótesis teórica de que ambas variables tienen una relación de largo plazo, pensando que, dada la insuficiente oferta doméstica de bienes de capital, el nivel de inversiones en el país depende del nivel de importaciones.

En la figura 20.2 se observa que la caída de las inversiones en los inicios de los años noventa coincidió efectivamente con la contracción de las importaciones. Una vez que el país pudo expandir las importaciones, la inversión empezó a recuperarse. Las serie de inversión y la serie de importaciones no se mueven todos los años en el mismo sentido, sin embargo, en el largo plazo, ambas terminan coincidiendo en sus caídas, estancamientos y aceleraciones.

**Figura 20.2: Inversión e importaciones cubanas**  
(logaritmo de las series)



Fuente: Elaboración propia a partir de datos del Anuario Estadístico de la ONE

Para corroborar la impresión que nos ofrece el gráfico desarrollamos los tres pasos que explicamos anteriormente para contrastar la cointegración:

1. Con la prueba ADF determinamos el orden de integración de las series. En las tablas 20.1 y 20.2 se resumen los resultados.

**Tabla 20.1: Prueba ADF a la variable inversión, período 1980- 2007**  
(logaritmo de la serie)

|                                                          |                                      |
|----------------------------------------------------------|--------------------------------------|
| <i>Transformación de la variable</i>                     | Sin diferencia                       |
| <i>Inclusión de constante y/o tendencia</i>              | Constante                            |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -2.338                               |
| <i>Valores críticos</i>                                  | -3.711 (1%), -2.981(5%), -2.630(10%) |
| <i>Transformación de la variable</i>                     | Primera diferencia                   |
| <i>Inclusión de constante y/o tendencia</i>              | Ninguna                              |
| <i>Rezagos</i>                                           | 0                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -3.246                               |
| <i>Valores críticos</i>                                  | -2.657 (1%), -1.954(5%), -1.609(10%) |

Fuente: Elaboración propia

**Tabla 20.2: Prueba ADF a la variable importaciones, período 1980- 2007**  
(logaritmo de la serie)

|                                                          |                                      |
|----------------------------------------------------------|--------------------------------------|
| <i>Transformación de la variable</i>                     | Sin diferencia                       |
| <i>Inclusión de constante y/o tendencia</i>              | Constante                            |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -2.303                               |
| <i>Valores críticos</i>                                  | -3.711 (1%), -2.981(5%), -2.630(10%) |
| <i>Transformación de la variable</i>                     | Primera diferencia                   |
| <i>Inclusión de constante y/o tendencia</i>              | Ninguna                              |
| <i>Rezagos</i>                                           | 1                                    |
| <i>Valor calculado del estadístico <math>\tau</math></i> | -3.645                               |
| <i>Valores críticos</i>                                  | -2.660 (1%), -1.955(5%), -1.609(10%) |

Fuente: Elaboración propia

Fue necesario diferenciar ambas series una vez para que fuesen estacionarias. Por tanto, se concluye que tienen el mismo orden de integración, en este caso, son  $I(1)$ . Así, verificamos la primera condición requerida para que dos procesos estén cointegrados.

2. Estimamos por MCO la regresión de las variables en los niveles de los logaritmos. Se obtuvo el siguiente resultado, en el cual el signo del coeficiente asociado a las importaciones está acorde con lo esperado teóricamente:

$$\ln INV_t = 0.374 + 0.908 * \ln M_t \quad (20.1)$$

se      (0.481) (0.114)

3. Tomamos del modelo 20.1 la serie de los residuos y le aplicamos la prueba ADF. El resultado se muestra en la tabla 20.3.

**Tabla 20.3: Prueba ADF sobre los residuos**

|                                                                                                  |                                   |
|--------------------------------------------------------------------------------------------------|-----------------------------------|
| <i>Inclusión de constante y/o tendencia</i>                                                      | Ninguna                           |
| <i>Rezagos</i>                                                                                   | 1                                 |
| <i>Valor calculado del estadístico <math>\tau</math></i>                                         | -3.056                            |
| <i>Valores críticos</i><br><i>Engle y Yoo (1987) Tabla 3</i><br><i>Tamaño de muestra 50 obs.</i> | -4.12 (1%), -3.29(5%), -2.90(10%) |

*Fuente: Elaboración propia*

El valor calculado del estadístico es -3.056. En Engle y Yoo (1987) no aparecen valores críticos para el caso de una muestra de 28 observaciones, así que tomamos los valores críticos del menor tamaño de muestra que es 50. Con estos valores críticos se rechaza la hipótesis nula de raíz unitaria en los residuos al 10% de nivel de significación. La prueba indica que los residuos son estacionarios, por tanto, se concluye que las variables están cointegradas al 10%.

El modelo (20.1) entonces no es espurio, sino que recoge la relación de cointegración entre las variables estimando una elasticidad de largo plazo de 0.908. Se interpreta que, por cada 10% que aumentan las importaciones, en el largo plazo las inversiones crecen 9.08%.

## 20.3 Mecanismo de Corrección de Errores

Si se prueba que un conjunto de variables están cointegradas, el modelo más apropiado para estimar la relación entre ellas es el Mecanismo de Corrección de Errores (MCE). El modelo de MCE combina variables en niveles y variables en diferencia, las primeras recogen las relaciones de largo plazo, mientras que las segundas nos dan las elasticidades de corto plazo. Como su nombre lo indica, estos modelos funcionan de tal manera que corrigen los desajustes que puedan ocurrir en las relaciones de largo plazo; toman en cuenta que las variables cointegradas van a tender a sus niveles de equilibrio.

El enunciado formal entre cointegración y MCE se conoce en la literatura como **Teorema de Representación de Granger** y plantea lo siguiente:

- Si un vector de variables es  $CI(1,1)$ , existe un Mecanismo de Corrección de Errores válido para representar el Proceso Generador de Datos (PGD).

- Si el PGD de un conjunto de variables admite una representación MCE, estas están cointegradas.

La formulación de un Mecanismo de Corrección de Errores para dos variables  $y_t$  y  $x_t$  integradas del mismo orden  $I(1)$  y cointegradas, es decir  $CI(1,1)$ , se representa de la siguiente manera:

$$\Delta y_t = \mu_1 + A_1(L) \Delta y_t + C_1(L) \Delta x_t - \gamma_1 [y_{t-1} - \beta_1 x_{t-1}] + \varepsilon_{1t} \quad (20.2)$$

donde  $A_1(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$  y  $C_1(L) = c_0 + c_1 L + c_2 L^2 + \dots + c_q L^q$ ;  $\varepsilon_{1t}$  se supone ruido blanco.

Si no fuese por el término entre corchetes, la ecuación (20.2) fuera una simple función de transferencias que incluye rezagos de la variable explicativa en  $C_1(L)$  y rezagos de la variable dependiente en  $A_1(L)$ .

Nótese que el término entre corchetes recoge las desviaciones del equilibrio de largo plazo rezagadas un período. El modelo (20.2) se apoya en una relación de cointegración entre las variables  $y_t = \beta_1 x_t$ . Las desviaciones de este equilibrio conocemos que son  $u_t = y_t - \beta_1 x_t$ . Al rezagarlos un período queda  $u_{t-1} = y_{t-1} - \beta_1 x_{t-1}$ .

Como las series son  $I(1)$  se incluyen en primera diferencia. En principio, todas las variables que aparecen en estas ecuaciones son estacionarias excepto el término entre corchetes. Para que este término sea estacionario, es necesario que las variables que lo conforman estén cointegradas, pues en caso contrario la ecuación no estaría equilibrada.

El MCE admite la inclusión de dummies e incluso otras variables que no formen parte de la relación de cointegración. Estas últimas deben siempre incorporarse en su transformación estacionaria. Usualmente también se trabaja con la transformación logarítmica de las series.

Como vemos, el MCE, conjuga la modelización de relaciones de corto y largo plazo. Pues, por una parte, tenemos a las variables diferenciadas y sus retardos (en caso de que sean necesarios) que nos dan las relaciones de corto plazo, es decir, recogen como los cambios en la variable explicativa  $\Delta(x_t)$ , afectan a los cambios en la variable dependiente  $\Delta(y_t)$ . Los parámetros estimados en  $C_1(L)$ ,  $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_q$ , son las elasticidades de corto plazo. Por otra parte tenemos en el término entre corchetes las variables en nivel que nos da la elasticidad de largo plazo  $\beta_1$ .

En el modelo de MCE (20.2) la expresión  $\gamma_1 [y_{t-1} - \beta_1 x_{t-1}]$  se conoce como **término de corrección de errores**. En cada período  $t$  este podrá ser distinto de cero sólo cuando se produzcan desviaciones del equilibrio. Dicho término se incluye en la ecuación para recoger los impactos que tienen sobre  $\Delta(y_t)$  las desviaciones del equilibrio de largo plazo. Como las variables  $y_t$  y  $x_t$  están cointegradas, cualquier

desequilibrio en el momento  $t$  tenderá a corregirse en los períodos subsiguientes en una proporción igual a  $\gamma_1$ .

Justamente,  $\gamma_1$  se nombra **parámetro de velocidad de ajuste** ya que cuanto mayor sea su valor más rápidamente se corregirán los desequilibrios. Por ejemplo, si  $\gamma_1 = 1/3$  entonces los ajustes hacia el equilibrio demoran tres períodos; si  $\gamma_1 = 0.5$ , entonces, en cada período que sigue al desequilibrio se corrige la mitad, bastando dos períodos (que pueden ser años, trimestres, meses, etc., en dependencia del período a que se refiere  $t$ ) para que se recupere el equilibrio de largo plazo.

El signo de  $\gamma_1$  debe ser negativo cuando la relación de largo plazo entre  $y_t$  y  $x_t$  es positiva ( $\beta_1 > 0$  en la relación de cointegración  $y_t = \beta_1 x_t$ ). De esta manera:

- cuando  $y_t$  se va por encima de su valor de largo plazo,  $[y_{t-1} - \beta_1 x_{t-1}]$  será positivo. Como  $\gamma_1 < 0$ , entonces  $\gamma_1 [y_{t-1} - \beta_1 x_{t-1}]$  será negativo en la ecuación (20.2). Por tanto, el término de corrección de errores afectará negativamente a  $\Delta(y_t)$ , provocando así que  $y_t$  disminuya en los próximos períodos para retornar al equilibrio.
- Cuando  $y_t$  se queda por debajo de su valor de largo plazo,  $[y_{t-1} - \beta_1 x_{t-1}]$  será negativo. Como  $\gamma_1 < 0$ , entonces  $\gamma_1 [y_{t-1} - \beta_1 x_{t-1}]$  será positivo;  $y_t$  aumentará en los próximos períodos para retornar al equilibrio.

Antes de pasar a ver la manera en que se estiman los MCE, vamos a mencionar algunas complicaciones adicionales que pueden llevar este tipo de modelos.

En el ejemplo mostrado en la ecuación (20.2) estamos considerando un solo término de corrección de errores debido a que estamos trabajando con dos variables. Pero sabemos que en el caso de  $m$  variables cointegradas pueden existir hasta  $m-1$  relaciones de cointegración, por tanto habría que incorporar en el modelo de MCE  $m-1$  términos de corrección de errores con sus correspondientes elasticidades de largo plazo y parámetros de velocidad de ajuste.

El modelo (20.2) es válido para los MCE que puedan darse entre variables integradas de primer orden, es decir,  $I(1)$ . Para relacionar variables cointegradas con órdenes de integración superiores será necesario aplicar MCE más complejos que sobrepasan los objetivos del presente texto.

El modelo (20.2) supone que la variable  $x_t$  es exógena a los cambios en  $y_t$ . Si existiera causalidad en los dos sentidos habría que incorporar al MCE una segunda ecuación que recoja los efectos que sobre  $\Delta(x_t)$  tienen tanto  $\Delta(y_t)$  como las desviaciones del equilibrio de largo plazo. Habría que tomar en cuenta que después una desviación del equilibrio de largo plazo, se ajustarán en los próximos períodos tanto la variable  $y_t$  como  $x_t$ :



$$\begin{aligned}\Delta y_t &= \mu_1 + A_1(L)\Delta y_t + C_1(L)\Delta x_t - \gamma_1[y_{t-1} - \beta_1 x_{t-1}] + \varepsilon_{1t} \\ \Delta x_t &= \mu_2 + A_2(L)\Delta x_t + C_2(L)\Delta y_t - \gamma_2[y_{t-1} - \beta_1 x_{t-1}] + \varepsilon_{2t}\end{aligned}\quad (20.3)$$

donde  $A_1(L)$ ,  $A_2(L)$ ,  $C_1(L)$ ,  $C_2(L)$  son polinomios de rezagos;  $\varepsilon_{1t}$  y  $\varepsilon_{2t}$  se suponen ruido blanco.

El modelo (20.3) se conoce en la literatura como **Vector de Corrección de Errores (VEC)** y combina la cointegración con los modelos VAR; estos últimos los analizaremos en el próximo capítulo.

Llegado a este punto del análisis, podemos ver que si se prueba que existe una relación de cointegración entre un conjunto de variables, la estimación de la misma mediante un modelo de Función de Transferencia no es aconsejable, debido a que se estará eliminando de la relación dinámica entre las variables el término de corrección de errores. El modelo de Función de Transferencia presentará errores de especificación al no incluir los efectos que tienen sobre la variable las desviaciones del equilibrio de largo plazo.

A continuación explicamos la metodología bietápica de Engle y Granger para estimar los MCE.

### 20.3.1 Estimación bietápica del MCE. Ejemplo con datos cubanos

En la metodología bietápica de Engle y Granger no se estiman al mismo tiempo las relaciones de largo plazo y corto plazo, sino que en una primera etapa se estima la relación de cointegración, y en una segunda etapa se estima el MCE empleando los residuos de la relación de largo plazo estimados en la primera etapa.

Las dos etapas las detallamos a continuación. Continuamos con el ejemplo de dos series de tiempo, con  $y_t$  como variable dependiente y  $x_t$  como variable explicativa.

- **Etapas 1:** En esta etapa justamente lo que habría que desarrollar serían los tres pasos que referimos anteriormente en la sección 20.2: habría que determinar el orden de integración de las variables, estimar por MCO la relación de largo plazo  $y_t = \beta_1 x_t + u_t$  y tomar los residuos  $\hat{u}_t$  para efectuarles la prueba ADF.
- **Etapas 2:** Una vez que estimamos en la etapa previa la relación de largo plazo de las variables y probamos que están cointegradas, pasaríamos a estimar en esta segunda etapa el Mecanismo de Corrección de Errores.

Para ello, tomamos los residuos estimados en la relación de largo plazo y los retardamos un período. De la primera etapa tenemos  $\hat{\beta}_1$  que permite construir  $\hat{u}_{t-1} = y_{t-1} - \hat{\beta}_1 x_{t-1}$ . Así, para estimar el MCE sustituimos  $[y_{t-1} - \beta_1 x_{t-1}]$  en el modelo (20.2) por  $\hat{u}_{t-1}$ :

$$\Delta y_t = \mu_1 + A_1(L)\Delta y_t + C_1(L)\Delta x_t - \gamma_1 \hat{u}_{t-1} + \varepsilon_{1t} \quad (20.4)$$

El modelo (20.4) es el que se estima en la segunda etapa utilizando nuevamente los estimadores MCO.

Queda al investigador la decisión de determinar el número de rezagos a incluir de las variables diferenciadas, para ello se pueden usar los criterios de información AIC y SIC, así como, los estadísticos t-Student y F-Fisher.

Como en cualquier regresión convencional se debe probar que se cumplen los diferentes supuestos, así como, observar el  $R^2$ , contrastar la significación global del modelo o la significación individual de los parámetros u otra hipótesis de interés para el análisis económico que se está desarrollando.

Dado que el Teorema de Representación de Granger plantea que el MCE es válido para representar la dinámica de dos variables cointegradas, se supone entonces que en (20.4)  $\gamma_1$  es no nulo. Así, una manera de comprobar que nuestras variables están cointegradas, y que hemos seguido correctamente todos los pasos de la metodología de Engle y Granger, es contrastar que  $\hat{\gamma}_1$  es significativamente distinto de cero.

Una vez estimado el MCE se pueden efectuar predicciones un período hacia delante si se disponen datos de las variables explicativas que presenten una relación contemporánea en el corto plazo con la variable dependiente. Nótese que si en el MCE no quedan variables explicativas sin rezagar, entonces no se requiere de información adicional para predecir el próximo período, puesto que el término de corrección de errores se encuentra rezagado.

## Ejemplo

Continuamos con el ejemplo de la dependencia de las inversiones a las importaciones (los datos se encuentran en la tabla B.12 del CD acompañante). De hecho, ya cumplimos la primera etapa de la metodología de Engle y Granger en la sección 20.2.1. Transitamos entonces directamente a la segunda etapa con la estimación del Mecanismo de Corrección de Errores.

Inicialmente se incluyeron en el modelo dos rezagos de las inversiones y de las importaciones. Después de eliminar los no significativos se obtuvo la siguiente estimación del MCE:

$$\begin{aligned} \Delta \ln INV_t &= 0.010 + 0.369 * \Delta \ln INV_{t-1} + 0.607 * \Delta \ln M_t - 0.477 * errorLP_{t-1} \\ \text{se} \quad &(0.029) \quad (0.121) \quad (0.185) \quad (0.157) \end{aligned} \quad (20.5)$$

$$R^2_{ajustado} = 0.637$$

Se comprobó que el modelo cumple con los supuestos. Los residuos no están autocorrelacionados, son homoscedásticos y tienen una distribución normal.

Se observa que el término de corrección de errores es significativo. El parámetro de velocidad de ajuste estimado es de 0.477, indicando que las desviaciones del equilibrio de largo plazo se ajustan casi completamente en dos años.

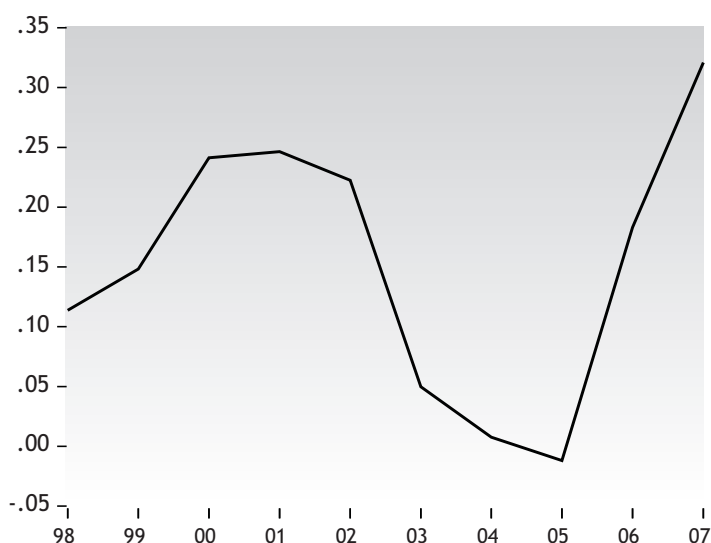
La elasticidad de corto plazo de las importaciones (0.369) es menor que la elasticidad de largo plazo anteriormente estimada (0.908). Cualquier aumento o disminución de las importaciones impactará en mayor magnitud a las inversiones en el largo plazo que en el corto plazo; ocurriendo el efecto casi completo de largo plazo en dos años.

El  $R^2$  ajustado de 0.637 es relativamente alto para un modelo de MCE que intenta explicar no sólo la dinámica de largo plazo, sino también el corto plazo, es decir, lo que sucede cada año con las inversiones.

El Ministerio de Economía y Planificación de Cuba planificó para el año 2008 un incremento de 7.2% en las importaciones. En base a este escenario, el MCE estimado proyecta para el año 2008 una contracción de las inversiones del orden de 4.2%, de acuerdo con el valor central de predicción.

A pesar de que las importaciones van a crecer en el 2008 y en el corto plazo esto tiene un impacto positivo en la inversión, el MCE está computando una anulación de tal efecto con el ajuste que deben tener las inversiones hacia su equilibrio de largo plazo. Si graficamos la serie de desviaciones de la relación de largo plazo (residuos estimados del modelo (20.1)) vemos que presenta valores positivos en los últimos dos años 2006 y 2007 (ver figura 20.3). Significa que la inversión se ha ido por encima de su equilibrio de largo plazo con relación al nivel de las importaciones. En consecuencia, debe ocurrir un ajuste de las mismas a la baja.

**Figura 20.3: Desviación de la relación de largo plazo entre inversión e importaciones**



Fuente: Elaboración propia

Se concluye así el presente capítulo referido a la cointegración y al Mecanismo de Corrección de Errores. En el próximo capítulo estudiamos los Vectores Autorregresivos, que es otra de las herramientas estadísticas que tienen los investigadores para estimar modelos multivariantes de series temporales.



# *preguntas*

Intente dar respuesta a las siguientes preguntas. Así sabrá si ha captado los principales mensajes de este capítulo.

1. ¿Cuándo se puede decir que dos variables están cointegradas?
2. Explique los pasos del contraste de cointegración sobre los residuos.
3. ¿Cuál es la formulación de un mecanismo de corrección de errores con dos variables? Identifique las elasticidades de largo y de corto plazo
4. ¿Cómo se interpreta el parámetro de velocidad de ajuste?



# ejercicios

- 20.1 En el siguiente modelo se relaciona la producción anual agrícola cubana (PA) con los factores capital (K) y trabajo (L) en los últimos 40 años. Según la prueba de raíz unitaria, las variables involucradas resultaron ser I(1). Siguiendo la metodología bietápica de Engle y Granger, se estimaron los siguientes modelos:

$$\begin{aligned} \ln PA_t &= 0.274 + 0.702 \ln K_t + 0.113 \ln L_t \\ \text{se} \quad &(0.481) \quad (0.114) \quad (0.045) \\ R^2 &= 0.98 \end{aligned} \quad (1)$$

$$\begin{aligned} \Delta \ln PA_t &= 0.012 + 0.451 \Delta \ln PA_{t-1} + 0.025 \Delta \ln K_t + 0.214 \Delta \ln L_t - 0.253 \text{errorLP}_{t-1} \\ \text{se} \quad &(0.661) \quad (0.106) \quad (0.015) \quad (0.026) \quad (0.008) \\ R^2 &= 0.66 \end{aligned} \quad (2)$$

- ¿Cuál de ellos es el Mecanismo de Corrección de Errores?
  - ¿Qué condición debe cumplirse para que el modelo (1) no sea espurio?
  - Identifique las elasticidades de corto y largo plazo
  - ¿Los factores de producción tienen un impacto significativo sobre la producción agrícola cubana en el largo y en el corto plazo?
  - ¿Qué interpretación usted haría del término  $0.253 \text{errorLP}_{t-1}$ ?
- 20.2 En la tabla B.4 del CD que acompaña a este texto aparecen datos anuales del Producto Interno Bruto de Cuba (PIB), las exportaciones de bienes y servicios a precios corrientes, y el valor constantes de la formación bruta de capital fijo. Estamos interesados en construir un modelo para predecir el comportamiento futuro del PIB. Utilizando la transformación logarítmica de las variables:
- Determine el orden de integración de las variables.
  - Pruebe si están cointegradas las variables utilizando la prueba ADF sobre los residuos.
  - Estime por MCO un Mecanismo de Corrección de Errores.
  - Verifique y corrija los supuestos de los estimadores MCO. Si corresponde, aplique análisis de intervención.
  - Examine si son significativos los coeficientes estimados.
  - Analice el signo de las elasticidades.



CAPÍTULO 21





## VECTORES AUTORREGRESIVOS

El presente es el último capítulo de esta parte del libro dedicada a la econometría de series temporales. Pretendemos inicialmente introducir la formulación de los modelos de Vectores Autorregresivos (VAR) y las restricciones que se usan para identificar el sistema. Posteriormente mostraremos las herramientas que han hecho a los VAR tan populares para la estimación y análisis de las relaciones dinámicas entre series temporales, nos referimos a las funciones impulso respuesta, la descomposición de la varianza y la causalidad de Granger. Finalmente, veremos estas aplicaciones en un ejemplo con datos reales de la economía cubana vinculado a la política monetaria y los precios.

### 21.1 Formulación del VAR

Los modelos de Vectores Autorregresivos (VAR) fueron propuestos por Sims (1980) como una alternativa a los modelos macroeconómicos que imponían “restricciones increíbles” para identificar los parámetros de interés. Sims critica los supuestos que se utilizaban para excluir o preestablecer la exogeneidad de determinadas variables en los modelos.<sup>1</sup>

<sup>1</sup> ‘[...] most of the restrictions on existing models are false, and the models are nominally over-identified.’

‘[...] existing large models contain too many incredible restrictions, [...]’

‘It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous. Of course, some restrictions, if only on lags length, are essential, so by “unrestricted” here I mean “without restrictions based on supposed a priori knowledge”.’ (Sims 1980 p.14-15).

En general, en los VAR todas las variables relevantes se tratan como endógenas y todas las ecuaciones tienen los mismos regresores. Es un sistema de ecuaciones dinámicas donde cada serie de tiempo se explica en función de su propio pasado, así como del valor contemporáneo y de los rezagos del resto de las series del sistema. De esta forma, se estiman las relaciones en el tiempo entre las variables usando el menor número de supuestos económicos. Constituye una metodología flexible donde el investigador solamente tiene que elegir las variables y seleccionar la cantidad de rezagos, imponiendo pocas restricciones teóricas a priori.

Un modelo VAR con tres series de tiempo y un solo rezago tiene la siguiente formulación:

$$\begin{aligned} y_t &= a_{12}w_t + a_{13}z_t + c_{11}y_{t-1} + c_{12}w_{t-1} + c_{13}z_{t-1} + u_{1t} \\ w_t &= a_{21}y_t + a_{23}z_t + c_{21}y_{t-1} + c_{22}w_{t-1} + c_{23}z_{t-1} + u_{2t} \\ z_t &= a_{31}y_t + a_{32}w_t + c_{31}y_{t-1} + c_{32}w_{t-1} + c_{33}z_{t-1} + u_{3t} \end{aligned} \quad (21.1)$$

Con el objetivo de imponer la menor cantidad de restricciones, los VAR son sistemas simétricos, cada serie de tiempo se incluye con la misma cantidad de rezagos en todas las ecuaciones. El modelo VAR viene caracterizado por dos dimensiones,  $k$  y  $p$ , referidas respectivamente al número de variables endógenas y al orden de rezagos considerados. En nuestro ejemplo,  $k=3$  y  $p=1$ .

Los  $u_t$  representan los shocks asociados a cada una de las variables. Los llamaremos shocks estructurales. Se supone que son errores ruido blanco no correlacionados. Esto es, además del supuesto de no autocorrelación se añade que no están correlacionadas las series de los errores de las diferentes ecuaciones. Se dice que son **errores ortogonales**.

El sistema (21.1) se conoce como el **VAR en su forma estructural**. Esta forma del VAR que incluye las relaciones contemporáneas no puede estimarse directamente. Como consecuencia de la simultaneidad entre las variables, los errores van a estar correlacionados con los regresores, violándose un supuesto que se había visto era necesario para que los estimadores Mínimos Cuadrados Ordinarios fueran insesgados.

El supuesto se viola debido a que los regresores que recogen las relaciones contemporáneas no son exógenos. En la primera ecuación los regresores  $w_t$  y  $z_t$  no son exógenos a la variable dependiente: en la segunda ecuación del sistema VAR se ve que  $w_t$  depende de  $y_t$ , así como en la tercera  $z_t$  depende de  $y_t$ . Por tanto,  $w_t$  y  $z_t$  van a estar correlacionados con  $u_{1t}$ . Lo mismo sucede en la segunda y tercera ecuación del sistema VAR.

Por esta razón, las estimaciones se realizan en un VAR que no incluye las relaciones contemporáneas, el cual se conoce como **VAR reducido** o estándar.

Primeramente, escribimos el modelo VAR estructural en forma matricial. Las relaciones contemporáneas de (21.1) se pasan hacia la izquierda de las ecuaciones y se definen las siguientes matrices:

$$\begin{aligned}
y_t - a_{12}w_t - a_{13}z_t &= c_{11}y_{t-1} + c_{12}w_{t-1} + c_{13}z_{t-1} + u_{1t} \\
a_{21}y_t - w_t - a_{23}z_t &= c_{21}y_{t-1} + c_{22}w_{t-1} + c_{23}z_{t-1} + u_{2t} \\
a_{31}y_t - a_{32}w_t - z_t &= c_{31}y_{t-1} + c_{32}w_{t-1} + c_{33}z_{t-1} + u_{3t}
\end{aligned}$$

$$\begin{pmatrix} 1 & -a_{12} & -a_{13} \\ -a_{21} & 1 & -a_{23} \\ -a_{31} & -a_{32} & 1 \end{pmatrix} \begin{pmatrix} y_t \\ w_t \\ z_t \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ w_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}$$

$$AX_t = C(L)X_{t-1} + u_t \quad (21.2)$$

donde  $X$  es un vector con todas las variables del esquema. La matriz  $A$ , que está formada por los coeficientes  $a_{ij}$  con signo cambiado, los cuales recogen las relaciones contemporáneas entre las variables; es una matriz simétrica de dimensión  $(k \times k)$ .  $C(L)$  contiene los coeficientes que relacionan a cada variable con los rezagos del resto de las variables y con sus propios rezagos. Puede ser un polinomio en el operador de rezagos en el caso que se incluyan más rezagos.

El sistema VAR estructural se transforma en un sistema en forma reducida premultiplicando el modelo (21.2) por  $A^{-1}$ :

$$A^{-1}AX_t = A^{-1}C(L)X_{t-1} + A^{-1}u_t$$

simplificando y definiendo  $B(L) = A^{-1}C(L)$  y  $e_t = A^{-1}u_t$  se llega a la forma reducida del sistema VAR:

$$X_t = B(L)X_{t-1} + e_t \quad (21.3)$$

Los errores del modelo VAR estándar  $e_t$  no coinciden con las series de los shocks estructurales  $u_t$ , puesto que existen relaciones contemporáneas entre las variables que no se recogen en la estimación del sistema reducido. Como consecuencia, los  $e_t$ , aunque siguen siendo ruido blanco, presentan correlaciones contemporáneas entre ellos. Se dice que son **errores no ortogonales**.

En  $B(L)$  están los coeficientes que relacionan a cada variable con sus propios rezagos (parte autorregresiva de cada ecuación del VAR) y los rezagos del resto de las variables. Para verificar la estacionariedad del VAR debemos computar las raíces de la parte autorregresiva; en total hay  $kp$  raíces. Para que el sistema VAR sea estacionario el módulo de las raíces inversas tiene que ser mayor a la unidad.

Para las predicciones y para el análisis de las relaciones entre las variables el modelo VAR se trata como un sistema. Sin embargo, la estimación de los parámetros se realiza para cada ecuación independientemente por Mínimos Cuadrados

Ordinarios.<sup>2</sup> En nuestro ejemplo, sería como estimar tres funciones de transferencia. Una vez estimados los parámetros del modelo reducido, se recuperan en una segunda etapa los parámetros estructurales. Al modelo VAR se le pueden añadir constante, tendencia o variables dummy para recoger la estacionalidad, pero siempre manteniendo la simetría del VAR, esto es, deben estar en todas las ecuaciones.

En el espíritu de no imponer restricciones y mantener la simetría del sistema, los modelos VAR tienden a estar sobreparametrizados. Se incluyen suficientes rezagos para poder captar las relaciones dinámicas entre las variables, pero por cada rezago adicional, los grados de libertad se reducen en  $k$  en cada ecuación y en  $k^2$  en el sistema. Además puede haber multicolinealidad entre los regresores. Esto hace que aumente el error estándar de los estimadores, con lo que se afecta la precisión de los mismos y las pruebas  $t$  y  $F$  reducen su potencia. No obstante, los estimadores MCO siguen siendo insesgados, así como las predicciones que se derivan del VAR.

La sobreparametrización tiene estos costos, pero el caso contrario puede producir errores de especificación con consecuencias aún más negativas. Si se imponen inapropiadamente restricciones cero a los coeficientes, los estimadores y las predicciones serán sesgados. Como sabemos, el costo de las variables omitidas es mayor al de las redundantes.

La cantidad de rezagos del VAR se puede seleccionar usando los criterios de información de Akaike (AIC) o Schwartz (SIC). También se puede usar el estadístico de razón de verosimilitud (RV), tal y como veremos más adelante.

No hay un consenso entre los econometristas sobre si diferenciar o no las variables en el VAR. Cuando el objetivo es tener predicciones de corto plazo es más común diferenciar las series de tiempo involucradas. En cambio, en las investigaciones empíricas es más usual no diferenciar las variables, aun cuando sigan una trayectoria no estacionaria. Con el objetivo de no perder la información sobre la relación de largo plazo, muchos de los estudios realizados con sistemas VAR utilizan las series en niveles.<sup>3</sup> De esta manera, se capturan las relaciones de cointegración que puedan existir. Un VAR en niveles puede ser consistente con un sistema de variables cointegradas, no así un VAR en diferencias.<sup>4</sup>

<sup>2</sup> Cuando en un modelo de varias ecuaciones los errores están correlacionados se suele usar la técnica SUR: modelo de regresión aparentemente no relacionado. No obstante, como en los VAR las variables situadas en la derecha de cada ecuación son las mismas, no se produce ninguna ganancia con esta técnica, los estimadores MCO arrojan los mismos resultados.

<sup>3</sup> Ver por ejemplo los trabajos de Sims (1980), Gordon y Leeper (1994), Bernanke (1986) y Rudebusch (1998)

<sup>4</sup> Ver Hamilton (1994) pp. 579-581.

## 21.2 Identificación del Sistema

No es posible recuperar toda la información del VAR estructural a partir de la estimación del VAR reducido. Existe un problema de identificación, hay más parámetros estructurales que reducidos.

Una vez estimado el VAR reducido (21.3) se obtienen los coeficientes de  $B(L)$  y la matriz de varianzas y covarianzas de los  $e_t$ :

$$\Sigma_e = E(e_t e_t') = E \left( \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} (e_1 \ e_2 \ e_3) \right) = \begin{pmatrix} E(e_1^2) & E(e_1 e_2) & E(e_1 e_3) \\ E(e_2 e_1) & E(e_2^2) & E(e_2 e_3) \\ E(e_3 e_1) & E(e_3 e_2) & E(e_3^2) \end{pmatrix}$$

Esta es una matriz simétrica de dimensión  $(k \times k)$  que en la diagonal principal tiene la información sobre la varianza de cada serie de errores y fuera de la diagonal principal se encuentran las covarianzas entre los errores de las distintas ecuaciones. Sabemos que los errores del VAR reducido son no ortogonales, por tanto, los valores de la matriz fuera de la diagonal principal no son ceros. La correlación entre los errores, que parece ser un problema, realmente lo que nos permite es obtener información relevante para recuperar los parámetros estructurales. Al ser simétrica ( $E(e_i e_j) = E(e_j e_i)$ ), la matriz de varianzas y covarianzas aporta  $(k^2+k)/2$  elementos distintos. En nuestro ejemplo de tres variables, aportaría 6 elementos distintos.

Del VAR estructural (21.2) hace falta recuperar los coeficientes de  $C(L)$  y la matriz de varianzas y covarianzas de los shocks estructurales. Conocemos que la serie de los shocks no están correlacionadas, es decir, el VAR estructural tiene errores ortogonales. La matriz de varianzas y covarianzas es una matriz diagonal:

$$\Sigma_u = E(u_t u_t') = \begin{pmatrix} E(u_1^2) & E(u_1 u_2) & E(u_1 u_3) \\ E(u_2 u_1) & E(u_2^2) & E(u_2 u_3) \\ E(u_3 u_1) & E(u_3 u_2) & E(u_3^2) \end{pmatrix} = \begin{pmatrix} E(u_1^2) & 0 & 0 \\ 0 & E(u_2^2) & 0 \\ 0 & 0 & E(u_3^2) \end{pmatrix}$$

con sólo  $k$  elementos desconocidos; en nuestro ejemplo, 3 elementos desconocidos.

Del VAR en su forma estructural (21.2) se precisa además obtener los  $k^2$  componentes de la matriz  $A$ , que se reducen a  $k^2 - k$  al ser iguales a uno los elementos de la diagonal principal. En un VAR de tres variables serían 6 elementos.

En balance, los parámetros de  $B(L)$  se agotan en el cálculo de  $C(L)$ ; así, se deben recuperar  $k^2$  parámetros estructurales con  $(k^2+k)/2$  parámetros reducidos. Esto es, hay  $(k^2-k)/2$  que no se pueden recuperar. Para identificar el sistema se deben imponer entonces al menos  $(k^2-k)/2$  restricciones. En un VAR de tres variables se necesitaría imponer 3 restricciones.

Para imponer restricciones, lo más usual es hacer cero algunos elementos de la matriz  $A$  del VAR estructural. La forma más utilizada para imponer restricciones en la matriz  $A$  es la conocida como **descomposición de Cholesky**.<sup>5</sup> Ella permite identificar el sistema a partir de darle un orden a las variables, es decir, definiendo el orden en que las variables se afectan unas a otras. En el ejemplo con que venimos trabajando, si se asume el orden  $y_t$ ,  $w_t$  y  $z_t$ , queda la siguiente matriz  $A$ :

$$\begin{pmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & -a_{32} & 1 \end{pmatrix}$$

Significa que  $y_t$  no está afectada contemporáneamente por  $w_t$  y  $z_t$ . A su vez,  $w_t$  no está afectada contemporáneamente por  $z_t$ . En cambio,  $z_t$  sí se ve afectada por las otras dos variables en el propio mes (en caso de que los datos sean mensuales).

Si se impone como restricción el orden  $z_t$ ,  $w_t$  y  $y_t$ , la matriz  $A$  es:

$$\begin{pmatrix} 1 & -a_{12} & -a_{13} \\ 0 & 1 & -a_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

Ahora es  $z_t$  la que no está afectada contemporáneamente por el resto de las variables y  $y_t$  la que se ve afectada en el mismo mes por las otras dos variables del sistema. Ahora  $w_t$  no está afectada contemporáneamente por  $y_t$ .

Nótese que la descomposición de Cholesky siempre lleva al sistema a estar exactamente identificado. En un VAR con tres variables necesitamos imponer 3 restricciones, que son exactamente la cantidad de ceros que tiene la matriz  $A$  en ambos casos mostrados. Con la descomposición de Cholesky la matriz  $A$  siempre puede verse como una matriz triangular. El sistema VAR siempre se asume como un modelo recursivo; la descomposición de Cholesky no permite simultaneidad en el modelo estructural. Cuando estamos trabajando con pocas variables la descomposición de Cholesky puede ser válida, pero en un VAR con más de cuatro o cinco variables resulta bastante arbitrario tener que imponer un modelo recursivo.

Para permitir simultaneidad, la sobreidentificación del sistema, e imponer restricciones consistentes con la teoría económica, se propone la **metodología de los**

<sup>5</sup> Las restricciones en la matriz  $A$  son conocidas como restricciones de corto plazo. También existe la opción de imponer restricciones de largo plazo en la función impulso-respuesta como las desarrolladas por Blanchard y Quah (1989). Recientemente se ha desarrollado una literatura sobre restricciones de signo en las funciones impulso respuesta (ver por ejemplo Uhlig 2005).

**VAR estructurales (SVAR).**<sup>6</sup> Las identificaciones estructurales fueron presentadas inicialmente por Sims (1986) y Bernanke (1986). Estos autores proponen restringir la matriz  $A$  en base a un modelo teórico. El modelo teórico depende de las variables y mercados que estemos modelando. Nótese que el modelo teórico no está predeterminando la exogeneidad de ninguna variable, sólo está restringiendo las relaciones contemporáneas.

En particular, las identificaciones estructurales suelen hacer cero la fila de las variables instrumentos de política económica. Por ejemplo, cuando se incluye en un VAR monetario el tipo de interés de corto plazo que controla el Banco Central, es común restringir que no reacciona contemporáneamente al resto de las variables del sistema.

Los instrumentos de política en los VAR se especifican como variables endógenas, debido a que las instituciones toman las decisiones en función del entorno económico. Sin embargo, se supone que sobre ellas no hay un efecto contemporáneo del resto de las variables, es decir, se supone que ellas sólo pueden reaccionar con rezagos. Este supuesto se apoya en el hecho de que, primero, existe una demora en la recolección de los datos por los departamentos de estadística, lo cual dificulta que la autoridad conozca de manera inmediata de la ocurrencia de shocks en la economía. Segundo, se considera que las instituciones toman las decisiones con cierta prudencia y prefieren confirmar con nuevos datos la ocurrencia de shocks antes de modificar las condiciones de la política económica. Con una frecuencia mensual en los datos es realmente defendible este supuesto.

En nuestro ejemplo de tres variables, podemos suponer que las variables  $y_t$  y  $z_t$  son instrumentos de política económica y que por tanto se puede esperar que ninguna de ellas reacciona contemporáneamente al resto de las variables. Tenemos entonces un sistema sobreidentificado con una restricción más que la necesaria. La matriz  $A$  queda de la siguiente manera:

$$\begin{pmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & -a_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

Puede en cambio suponerse que solamente la variable  $z_t$  es un instrumentos de política económica y que existe una teoría económica que propone que el coeficiente  $a_{23}$  es igual a 1. Entonces la matriz  $A$  con la identificación estructural es:

<sup>6</sup> En la literatura sobre los VAR se usa el término *estructural* para referirse a la identificación *estructural* y también para el VAR *estructural* que mantiene en su especificación las relaciones contemporáneas.

$$\begin{pmatrix} 1 & -a_{12} & -a_{13} \\ -a_{21} & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Este es un sistema VAR exactamente identificado y que permite la simultaneidad entre  $y_t$  y  $w_t$  ( $-a_{12} \neq 0$  y  $-a_{21} \neq 0$ ).

Una vez identificado el sistema, podemos calcular los parámetros del VAR estructural a partir de los parámetros estimados en el VAR reducido. Necesitamos recuperar los parámetros de  $A$  y  $\Sigma_u$  con la información contenida en  $\Sigma_e$ . Ellos están relacionados por la expresión:

$$\Sigma_u = A\Sigma_e A' \quad (21.4)$$

Los parámetros estructurales se calculan resolviendo el sistema de ecuaciones que forma la expresión (21.4); algo que normalmente hace el programa estadístico. La expresión (21.4) proviene de la transformación que efectuamos para llegar de la forma estructural del VAR (21.2) a su forma reducida (21.3). Recordar que quedaba que  $e_t = A^{-1}u_t$ , o lo que es lo mismo,  $u_t = Ae_t$ . Sustituyendo esta ecuación en  $\Sigma_u$ ,

$$\Sigma_u = E(u_t u_t') = E(Ae_t (Ae_t)')$$

y dado que  $(Ae_t)'$  es equivalente a  $(A'e_t')$ , podemos llegar a (21.4):

$$\Sigma_u = E(Ae_t A'e_t') = AE(e_t e_t')A' = A\Sigma_e A'$$

Los resultados que se obtienen para el VAR estructural son sensibles a los supuestos de identificación. La robustez de las conclusiones suelen ser bajas. Sólo cuando las relaciones contemporáneas son muy cercanas a cero, los resultados son independientes de los supuestos de identificación.

### 21.3 Utilidades del VAR

Los modelos VAR se estiman con dos objetivos fundamentales: para hacer predicciones a corto plazo y para analizar las relaciones dinámicas entre las variables. Las predicciones se hacen directamente con el VAR en su forma reducida. Teniendo los datos de las series de tiempo hasta el período  $t$  se puede pronosticar la trayectoria que seguirán en los períodos sucesivos  $t+h$ . El sistema VAR pronostica al unísono la trayectoria futura de todas las series considerando las diferentes interacciones entre ellas y la estructura autorregresiva de las ecuaciones. Nótese que para las predicciones no se requieren los parámetros que recogen las relaciones contemporáneas (momento  $t$ ) entre las variables. Las predicciones se computan a



partir de  $t+1$  con los datos que ya disponemos de todas las variables en el momento  $t$ .

Las relaciones dinámicas se examinan con dos herramientas principales: las funciones impulso respuesta (FIR) y la descomposición de la varianza. Dichas herramientas construyen determinadas simulaciones que sí requieren de las relaciones contemporáneas, la cuales sabemos están presentes en el VAR estructural. Ambas simulan los efectos de los shocks asociados a cada una de las variables. Para que tengan un sentido económico se trabaja con los shocks estructurales, es decir, después de haber completado el proceso de ortogonalización de los errores. Así, una vez que se estimó el VAR reducido, se definieron las restricciones que permiten identificar el sistema y se recuperaron los parámetros estructurales, se construyen las FIR y la descomposición de la varianza en el VAR estructural.

A diferencia de lo que sucede en las funciones de transferencia que vimos en el capítulo 16, en el VAR no puede simularse la respuesta de la variable dependiente ante un incremento de la variable independiente; y no puede hacerse porque no hay variables independientes, todas las series en el VAR son endógenas. Las relaciones en el VAR se analizan observando que pasa con las variables una vez que ocurre un shock. Como el VAR es un sistema, cada variable está sujeta, además de a sus propios shocks, a los shocks en el resto de las variables.

Así, las **funciones impulso respuesta** en el VAR simulan la respuesta que acontece sobre el resto de las variables, cuando una variable se sale del equilibrio que describe el sistema, es decir, cuando la serie recibe un shock o impulso. En el VAR estructural de nuestro ejemplo (modelo 21.1), un shock en  $w_t (u_{2t})$  tendría una respuesta en  $y_t$  a través del valor contemporáneo de  $w_t$  en la primera ecuación del sistema. Pero también a través del efecto que  $w_t$  tiene sobre  $z_t$  en la tercera ecuación del sistema, dado que  $z_t$  también está presente en la ecuación de  $y_t$ . Los shocks además se propagan por varios períodos debido a los rezagos y términos autorregresivos en las ecuaciones. Toda esta diversidad de interacciones en el tiempo se resume en las funciones impulso respuesta.

Aquí volvemos a aplicar la teoría matemática de las ecuaciones en diferencia. Con cierta analogía, tenemos en el capítulo 18 los procesos AR y su solución MA donde vemos la FIR. El modelo VAR constituye un sistema de ecuaciones en diferencia, al cual podemos encontrarle una solución que llamamos **vector de medias móviles (VMA)**:

$$\begin{bmatrix} y_t \\ w_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{w} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) & \phi_{13}(i) \\ \phi_{21}(i) & \phi_{22}(i) & \phi_{23}(i) \\ \phi_{31}(i) & \phi_{32}(i) & \phi_{33}(i) \end{bmatrix} \begin{bmatrix} u_{1t-i} \\ u_{2t-i} \\ u_{3t-i} \end{bmatrix} \quad (21.5)$$

El sistema (21.5) es la solución del VAR estructural (21.1).<sup>7</sup> Cada variable se expresa, además de en función de sus propios shocks estructurales, en función de los shocks estructurales presentes y pasados del resto de las variables. Como es requerido, en la solución no aparece ninguna variable endógena rezagada. Queda así explícito el efecto en el tiempo de los shocks  $u_1$ ,  $u_2$  y  $u_3$ . Se tienen  $k^2$  funciones impulso respuesta en un VAR de  $k$  variables; 9 en el sistema VAR de 3 variables, 3 por cada variable. Cada FIR describe la trayectoria en el tiempo de la variable ante un cambio en  $u_{jt}$ . Los  $\phi$  contienen los valores de las respuestas, siendo los  $\phi(0)$  multiplicadores de impacto y los  $\phi(i)$  ( $\forall i \geq 1$ ) los multiplicadores intermedios.

Por ejemplo, las tres funciones impulso respuesta de la variable  $y_t$  se definen como:

$$\partial y_t / \partial u_{jt-i} = \phi_{1j} \quad \forall i \geq 0 \quad j = 1, 2, 3$$

La respuesta de  $y_t$  ante un shock en  $w_t(u_{2t})$  viene dada por:

$$\partial y_t / \partial u_{2t-i} = \phi_{12} \quad \forall i \geq 0$$

La respuesta de  $y_t$  ante un shock propio ( $u_{1t}$ ) viene dada por:

$$\partial y_t / \partial u_{1t-i} = \phi_{11} \quad \forall i \geq 0$$

Es común analizar el efecto de un shock positivo igual a una desviación estándar, que representaría un shock “típico” en la variable. Es usual representar la respuesta en un gráfico que tiene en el eje de las abscisas el tiempo y en las ordenadas el valor de la respuesta ( $\phi$ ) en cada período  $i$ . Si se cumple la condición de estacionariedad en el VAR entonces la representación VMA será estable; las FIR a medida que transcurre el tiempo deben ir convergiendo a cero. Por ello, se grafica por lo general la función impulso respuesta en un período de dos o tres años (con datos mensuales o trimestrales). Al final del capítulo ilustramos un ejemplo con datos de la economía cubana.

Se estila también construir la **FIR acumulada** en  $n$  períodos. La respuesta acumulada de  $y_t$  a un shock en  $w_t(u_{2t})$  sería:

$$\sum_{i=0}^n \phi_{12}(i)$$

Si se toma la respuesta acumulada en  $\infty$  períodos se está computando el multiplicador de largo plazo; si el sistema es estable, dicha respuesta será finita:

<sup>7</sup> No se describe en este texto el algebra de esta solución porque es bastante engorrosa.

$$\sum_{i=0}^{\infty} \phi_{12}(i)$$

Por su parte, la **descomposición de la varianza** provee de información acerca de la importancia relativa de cada variable a la hora de explicar el resto de las variables en el VAR. Esta información se obtiene de descomponer los errores de predicción entre los shocks estructurales propios y los shocks estructurales del resto de los indicadores para distintos horizontes.

También se emplea para ello la representación VMA en (21.5). Resumida en su forma matricial, se tiene que:

$$X_t = \mu + \sum_{i=0}^{\infty} \phi_i u_{t-i}$$

$$X_{t+h} = \mu + \sum_{i=0}^{\infty} \phi_i u_{t+h-i}$$

De donde el error de predicción  $h$  períodos hacia delante tiene la siguiente expresión:

$$X_{t+h} - E_t X_{t+h} = \sum_{i=0}^{\infty} \phi_i u_{t+h-i}$$

Para examinar la importancia relativa de cada shock, se toma entonces la varianza de los errores de predicción para distintos horizontes y se descompone porcentualmente entre los shocks de cada una de las variables. Este porcentaje tiende a converger a partir de determinado horizonte temporal. La explicación de cada variable por sus propios shocks da una idea del grado de exogeneidad de la serie al resto de las variables incluidas en el VAR.

La descomposición de la varianza normalmente se presenta en forma de tabla, ilustrando el porcentaje en que los errores de predicción de cada variable son explicados por los shocks en el resto de las variables del VAR en un horizonte de tres a cuatro años (con datos mensuales o trimestrales). También hay un ejemplo de esta herramienta al final del capítulo.

Se pueden computar bandas de confianza para las FIR y la descomposición de la varianza y de esta forma inferir si son o no significativas las relaciones entre las variables.

En los modelos VAR no se efectúan pruebas de significación individual de los regresores con el estadístico  $t$ . La significación o no de un único coeficiente no dice mucho, pues en la relación entre las variables están involucrados varios rezagos, además de que no debemos romper la simetría del VAR porque un coeficiente sea no significativo. También habíamos comentado que la sobreparametrización y la multicolinealidad reducen la potencia de la prueba  $t$ .

La prueba  $F$  de significación global si se emplea más dentro de lo que se ha dado en conocer como causalidad de Granger.

### 21.3.1 Causalidad de Granger

La prueba de causalidad de Granger verifica el orden de precedencia entre las series. En un VAR de dos variables y tres rezagos ( $k = 2$  y  $p = 3$ ):

$$\begin{aligned} y_t &= b_{11}y_{t-1} + b_{12}y_{t-2} + b_{13}y_{t-3} + b_{14}w_{t-1} + b_{15}w_{t-2} + b_{16}w_{t-3} + e_{1t} \\ w_t &= b_{21}y_{t-1} + b_{22}y_{t-2} + b_{23}y_{t-3} + b_{24}w_{t-1} + b_{25}w_{t-2} + b_{26}w_{t-3} + e_{2t} \end{aligned}$$

donde  $e_{1t}$  y  $e_{2t}$  son errores ruidos blancos gaussianos, se dice que  $w_t$  causa en el sentido de Granger a la variable  $y_t$ , si es significativamente distinto de cero al menos uno de los coeficientes asociados a los rezagos de  $w_t$  en la ecuación donde  $y_t$  es la variable endógena. Si por el contrario  $b_{14} = b_{15} = b_{16} = 0$  entonces  $w_t$  no causa en el sentido de Granger a  $y_t$ .

Asimismo, se dice que  $y_t$  causa en el sentido de Granger a la variable  $w_t$  si es significativamente distinto de cero al menos uno de los coeficientes asociados a los rezagos de  $y_t$  en la ecuación donde  $w_t$  es la variable endógena. Si por el contrario  $b_{24} = b_{25} = b_{26} = 0$  entonces  $y_t$  no causa en el sentido de Granger a  $w_t$ .

Hacemos entonces dos contrastes con las correspondientes hipótesis nula y alternativa:

#### Contraste 1

$$H_0: b_{14} = b_{15} = b_{16} = 0 \quad \Rightarrow w_t \text{ no causa en el sentido de Granger a } y_t$$

$$H_1: \text{algún } b_{1j} \neq 0, j = 4, 5, 6 \quad \Rightarrow w_t \text{ causa en el sentido de Granger a } y_t$$

#### Contraste 2

$$H_0: b_{24} = b_{25} = b_{26} = 0 \quad \Rightarrow y_t \text{ no causa en el sentido de Granger a } w_t$$

$$H_1: \text{algún } b_{2j} \neq 0, j = 4, 5, 6 \quad \Rightarrow y_t \text{ causa en el sentido de Granger a } w_t$$

Se pueden dar cuatro casos según el resultado de los contrastes:

1. Si se rechaza  $H_0$  en el contraste 1 y no se rechaza en el contraste 2, se dice que hay causalidad unidireccional de  $w_t$  hacia  $y_t$ .
2. Si se rechaza  $H_0$  en el contraste 2 y no se rechaza en el contraste 1, se dice que hay causalidad unidireccional de  $y_t$  hacia  $w_t$ .
3. Si se rechaza  $H_0$  en ambos contrastes se dice que hay causalidad en los dos sentidos entre las variables.
4. Si no se rechaza  $H_0$  en ambos contrastes se dice que las variables son independientes.

Para efectuar cada contraste se calcula el estadístico  $F$  con la suma de cuadrados de la ecuación restringida y sin restringir, de la manera que se estudió en la

sección 6.5.2. Para el contraste 1 se calcula el estadístico  $F$  en la primera ecuación del VAR. La ecuación sin restringir conserva los rezagos de  $w_t$ , mientras que la ecuación restringida excluye los rezagos de  $w_t$ . Para el contraste 2 estimamos la segunda ecuación del VAR con y sin los rezagos de  $y_t$ . Tomando la suma de residuos sin restringir al cuadrado ( $SRC$ ) y la suma de residuos restringidos al cuadrado ( $SRRC$ ) calculamos la  $F$ :

$$F = \frac{T - c}{m} \frac{(SRRC - SRC)}{SRC}$$

donde  $T$  = total de observaciones,  $c$  = cantidad de parámetros estimados en la ecuación no restringida (en nuestro ejemplo  $c = 6$ ) y  $m$  = cantidad de parámetros restringidos, que coincide con el orden  $p$  de rezagos del VAR (en nuestro ejemplo  $m = 3$ ). Los valores calculados se comparan con los valores críticos de la distribución  $F$  con  $m$  y  $(T - c)$  grados de libertad.

¿Qué otras conclusiones pueden extraerse de los resultados de la causalidad de Granger? Cuando una variable ( $w_t$ ) causa en sentido de Granger a otra ( $y_t$ ) se dice que tiene capacidad predictiva sobre esa variable. Sus valores rezagados mejoran las predicciones que pueden efectuarse de  $y_t$ , por tanto se justifica mantener a  $w_t$  en la ecuación de  $y_t$ . Cuando hay causalidad en los dos sentidos se está encontrando evidencia que corrobora la necesidad de usar un VAR para predecir conjuntamente ambas variables.

Cuando una variable ( $w_t$ ) no causa en sentido de Granger a otra ( $y_t$ ) no puede decirse que  $y_t$  es exógena a  $w_t$ . Nótese que las pruebas de causalidad de Granger se hacen en un VAR reducido. Para verificar la exogeneidad necesitaríamos incluir en la prueba las relaciones contemporáneas entre las variables. Dado su carácter introductorio, en este texto no abordamos los contrastes para verificar los distintos tipos de exogeneidad que se estudian en la literatura (débil, fuerte y superfuerte).

Pasando a otro aspecto de la causalidad de Granger. Cuando en el VAR intervienen más de dos variables la prueba  $F$  pierde utilidad. Por ejemplo, tomemos un modelo VAR con tres variables y dos rezagos ( $k = 3$  y  $p = 2$ ):

$$\begin{aligned} y_t &= b_{11}y_{t-1} + b_{12}y_{t-2} + b_{13}w_{t-1} + b_{14}w_{t-2} + b_{15}z_{t-1} + b_{16}z_{t-2} + e_{1t} \\ w_t &= b_{21}y_{t-1} + b_{22}y_{t-2} + b_{23}w_{t-1} + b_{24}w_{t-2} + b_{25}z_{t-1} + b_{26}z_{t-2} + e_{2t} \\ z_t &= b_{31}y_{t-1} + b_{32}y_{t-2} + b_{33}w_{t-1} + b_{34}w_{t-2} + b_{35}z_{t-1} + b_{36}z_{t-2} + e_{3t} \end{aligned} \quad (21.6)$$

No tiene mucho valor que para probar la causalidad de  $w_t$  hacia  $y_t$  efectúemos una prueba  $F$  contrastando en la primera ecuación que  $b_{13} = b_{14} = 0$ . El efecto de  $w_t$  sobre  $y_t$  puede ocurrir a través de la tercera ecuación del sistema. Si en la tercera ecuación  $b_{33} \neq 0$  o  $b_{34} \neq 0$  los rezagos de  $w_t$  afectarán a  $y_t$  por

intermedio de  $z_t$ . En general, la inferencia estadística en un VAR de más de dos variables requiere contrastar hipótesis que involucren todas las ecuaciones del sistema (*cross-equation restrictions*).

Para ello se propone utilizar el **estadístico de razón de verosimilitud (RV)** calculado en base a la estimación del modelo VAR sin restringir y la estimación del modelo VAR restringido:

donde:

$|\Sigma_e|$ : determinante de la matriz de varianzas y covarianzas de los errores del VAR sin restringir

$|\Sigma_e^r|$ : determinante de la matriz de varianzas y covarianzas de los errores del VAR restringido

$T$ : total de observaciones

$c$ : número de parámetros estimados en cada ecuación del modelo no restringido. Si el modelo no tiene constante, ni tendencia o variables dummy, entonces  $c = kp$  (por ejemplo, si el VAR no restringido es (21.6) entonces  $c = 6$ )

Este estadístico tiene una distribución asintótica  $\chi^2$  con grados de libertad igual al número de parámetros restringidos en el VAR. La prueba tiene en la hipótesis nula el modelo restringido. Si el valor calculado del estadístico es mayor que los valores críticos se rechaza la hipótesis nula.

Para probar la causalidad de Granger contrastamos la capacidad predictiva de la variable sobre el resto de las variables del VAR. En nuestro ejemplo (modelo 21.6) para probar si  $w_t$  causan en el sentido de Granger al resto de las variables tenemos como modelo no restringido:

$$\begin{aligned} y_t &= b_{11}y_{t-1} + b_{12}y_{t-2} + b_{13}w_{t-1} + b_{14}w_{t-2} + b_{15}z_{t-1} + b_{16}z_{t-2} + e_{1t} \\ z_t &= b_{31}y_{t-1} + b_{32}y_{t-2} + b_{33}w_{t-1} + b_{34}w_{t-2} + b_{35}z_{t-1} + b_{36}z_{t-2} + e_{3t} \end{aligned}$$

mientras que el modelo restringido excluye los rezagos de  $w_t$ :

$$\begin{aligned} y_t &= b_{11}y_{t-1} + b_{12}y_{t-2} + b_{15}z_{t-1} + b_{16}z_{t-2} + e_{1t} \\ z_t &= b_{31}y_{t-1} + b_{32}y_{t-2} + b_{35}z_{t-1} + b_{36}z_{t-2} + e_{3t} \end{aligned}$$

La hipótesis nula del contraste es  $b_{13} = b_{14} = b_{33} = b_{34} = 0$  y la alternativa que al menos uno de estos coeficientes es distinto de cero. El mismo análisis se efectuará para el resto de las series. Tal hipótesis nos brinda información que nos ayuda a detectar si se debe incorporar la variable en el VAR.

Pueden construirse otros contrastes con restricciones en todas las ecuaciones. Por ejemplo, ver si de manera conjunta dos o más variables no son causadas en el sentido de Granger por el resto de las variables del VAR (*block exogeneity test*).

El estadístico *RV* también es útil para seleccionar la cantidad de rezagos a incorporar al VAR. Se contrasta la hipótesis nula de que los coeficientes asociados a un cierto retardo  $p$  son todos ceros. El modelo VAR no restringido tiene  $p$  rezagos y el restringido  $p - 1$  rezagos. Si los datos son mensuales se puede comenzar con el rezago 13. Se va reduciendo sucesivamente el orden hasta obtener el primer rechazo. En nuestro ejemplo, si queremos verificar si el orden de rezagos es dos o uno, el modelo no restringido es el VAR (21.6) mientras que el restringido viene dado por:

$$\begin{aligned}y_t &= b_{11}y_{t-1} + b_{13}w_{t-1} + b_{15}z_{t-1} + e_{1t} \\w_t &= b_{21}y_{t-1} + b_{23}w_{t-1} + b_{25}z_{t-1} + e_{2t} \\z_t &= b_{31}y_{t-1} + b_{33}w_{t-1} + b_{35}z_{t-1} + e_{3t}\end{aligned}$$

Debe siempre comprobarse con alguna prueba estadística que el orden de rezagos que propone el estadístico *RV* o los criterios de información (AIC y SIC) elimina la autocorrelación de los errores.

## 21.4 Un VAR con los Precios en Cuba

Probablemente, uno de los mayores usos que han tenido los VAR ha sido en el terreno de la política monetaria.<sup>8</sup> Las variables que comúnmente están presentes en los VAR monetarios son: el tipo de interés, algún agregado monetario, el tipo de cambio, el producto y la inflación. Los VAR permiten estimar los efectos de los shocks monetarios sin tener que contar con un modelo macroeconómico estructural para la economía. El resultado más buscado cuando se estima un VAR monetario es la respuesta en el tiempo de las variables objetivos finales de la política monetaria (inflación y producto) a los shocks en la variable objetivo operativo del Banco Central (casi siempre el tipo de interés interbancario).

Para el caso cubano vamos a presentar un ejemplo sencillo estimando las relaciones dinámicas entre las series de tiempo tipo de los precios regulados (PR) y los precios no regulados (P). Los datos utilizados se muestran en la tabla B.13 del CD acompañante, con el objetivo de que el lector pueda replicar el ejemplo en su computador.

En correspondencia con la alta participación que tiene el Estado en la economía cubana, los precios regulados ocupan el 40% dentro del total del Índice de Precios al Consumidor (IPC). Los precios regulados son manejados por las autoridades monetarias y fiscales en función de los objetivos de la política económica. Es interesante analizar que efecto tienen los precios regulados en la inflación en

<sup>8</sup> Se pueden ver por ejemplo los trabajos de Gordon y Leeper (1994); Christiano, Eichenbaun y Evans (1996); Bernanke y Mihov (1998); y Evans y Kuttner (1998), entre muchos otros. En Rudebusch (1998) se realiza una crítica a los VAR como herramienta para medir la política monetaria, en el mismo número de la revista *International Economic Review* se puede ver la respuesta de Sims (1998).

los mercados con precios no regulados (respuesta de P a un shock en PR) y a su vez ver como reaccionan las autoridades ante un aumento de la inflación en los mercados no regulados (respuesta de PR a un shock en P).

Se usan datos mensuales comprendiendo el período enero de 1996 hasta diciembre de 2004. Se les elimina la estacionalidad con el programa TRAMO-SEATS y se trabaja con la transformación logarítmica. Las variables no son diferenciadas, con el objetivo de obtener información sobre la relación de largo plazo. Las estimaciones se efectúan con el programa econométrico Eviews 5. En Vidal (2008) se puede encontrar la estimación de un VAR estructural para la política monetaria cubana incluyendo otras variables además de los precios regulados y no regulados.

Los criterios de información (AIC y SIC) proponen estimar el VAR con dos rezagos y el estadístico *RV* propone siete rezagos. Al chequear la autocorrelación con el estadístico *LM* vemos que esta no se elimina con dos rezagos y sí con siete rezagos (tabla 21.1). Con dos rezagos la prueba indica autocorrelación de orden 3, 6 y 9, al 5% de nivel de significación. Finalmente se estima el VAR con  $p=7$ .

**Tabla 21.1: Prueba de autocorrelación**

( $H_0$ : No hay autocorrelación de orden  $p$  en los residuos)

| Rezago | Estadístico LM (probabilidad) |       |
|--------|-------------------------------|-------|
|        | $P=2$                         | $P=7$ |
| 1      | 0.798                         | 0.406 |
| 2      | 0.497                         | 0.846 |
| 3      | 0.038                         | 0.303 |
| 4      | 0.461                         | 0.186 |
| 5      | 0.725                         | 0.953 |
| 6      | 0.010                         | 0.464 |
| 7      | 0.825                         | 0.458 |
| 8      | 0.520                         | 0.711 |
| 9      | 0.040                         | 0.212 |
| 10     | 0.770                         | 0.967 |
| 11     | 0.965                         | 0.846 |
| 12     | 0.073                         | 0.723 |

Fuente: Resultados del programa Eviews 5.

Para verificar la estacionariedad del VAR el programa Eviews computa las raíces y su módulo (ver tabla 21.2). Como estamos trabajando con un VAR con dos variables y siete rezagos, la parte autorregresiva tiene 14 raíces. De ellas, el programa calcula 12 raíces complejas y 2 reales. Como la condición de estacionariedad la hemos venido definiendo sobre el módulo de las raíces inversas, calculamos en la tercera columna la inversa de cada valor de la segunda columna. Vemos entonces que el módulo de todas las raíces inversas es mayor a la unidad, cumpliéndose así la condición de estacionariedad.



Tabla 21.2: Raíces de la parte autorregresiva del VAR

| Raíces                | Módulo raíces | Módulo raíces inversas |
|-----------------------|---------------|------------------------|
| 0.931652 - 0.053674i  | 0.933196      | 1.071586               |
| 0.931652 + 0.053674i  | 0.933196      | 1.071586               |
| 0.043305 - 0.870996i  | 0.872072      | 1.146694               |
| 0.043305 + 0.870996i  | 0.872072      | 1.146694               |
| -0.733569 - 0.444925i | 0.857952      | 1.165566               |
| -0.733569 + 0.444925i | 0.857952      | 1.165566               |
| 0.672399 - 0.466257i  | 0.818240      | 1.222135               |
| 0.672399 + 0.466257i  | 0.818240      | 1.222135               |
| -0.748635             | 0.748635      | 1.335764               |
| 0.486825 - 0.284742i  | 0.563983      | 1.773103               |
| 0.486825 + 0.284742i  | 0.563983      | 1.773103               |
| 0.072651 - 0.520461i  | 0.525507      | 1.902924               |
| 0.072651 + 0.520461i  | 0.525507      | 1.902924               |
| 0.315104              | 0.315104      | 3.173555               |

Fuente: Resultados del programa Eviews 5.

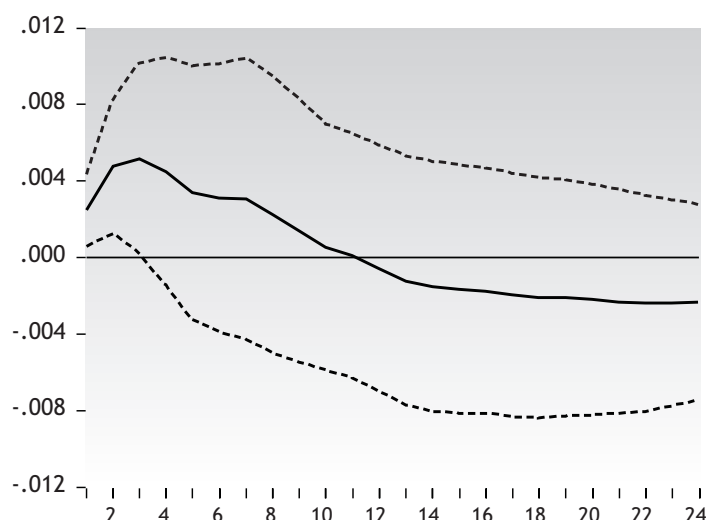
Para identificar el sistema VAR de dos variables se requiere imponer una sola restricción. Utilizamos la descomposición de Cholesky. Como la variable PR es fundamentalmente determinada por las autoridades económicas, establecimos el orden PR y después P. Con lo cual suponemos que los precios regulados no reaccionan contemporáneamente a los precios no regulados, sólo podrían reaccionar con rezagos. Con esta identificación el paquete estadístico computa las FIR y la descomposición de la varianza.

En total hay cuatro FIR. En el texto mostramos la respuesta de P a un shock en PR (figura 21.1) y la respuesta de PR a un shock en P (figura 21.2). Se simula la respuesta mensual durante dos años ante un shock positivo en cada variable igual a una desviación estándar, que representa la medida de un shock “típico”. Las funciones impulso respuesta contienen las bandas de confianza de 90% de probabilidad (2 desviaciones estándar). Para que una respuesta positiva sea significativa la banda mínima tiene que ser mayor que cero. Para que una respuesta negativa sea significativa la banda máxima tiene que ser menor que cero.

En la figura 21.1 se observa que un shock positivo igual a una desviación estándar en los precios regulados provoca un incremento de los precios no regulados bastante inmediato. Se obtiene una respuesta estadísticamente significativa contemporánea y con uno y dos meses de rezago (en la figura serían los meses 1, 2 y 3). Después de transcurridos tres meses la respuesta empieza a disminuir y se hace

no significativa. Con dos rezagos (mes 3) se alcanza la respuesta máxima con 0.52% (0.052 en la figura). En los tres meses de respuesta significativa los precios no regulados acumulan un incremento de 1.24%.<sup>9</sup>

**Figura 21.1: Respuesta de los precios no regulados (P) a un shock en los precios regulados (PR)**  
(eje horizontal: 24 meses a partir del mes en que ocurre el shock (mes 1))  
(eje vertical: respuesta en cada mes (1=100%))  
(línea punteada:  $\pm 2$  desviaciones estándar)



Fuente: Resultados del programa Eviews 5

En la figura 21.2 se ve que la respuesta contemporánea de PR a un shock positivo en P es cero porque la impusimos de esta manera con la descomposición de Cholesky. Con uno y dos rezagos se obtiene una respuesta positiva significativa. La respuesta posteriormente se reduce, pero rebrota casi después de transcurrido un año, con valores que están en el límite de la banda mínima significativa.

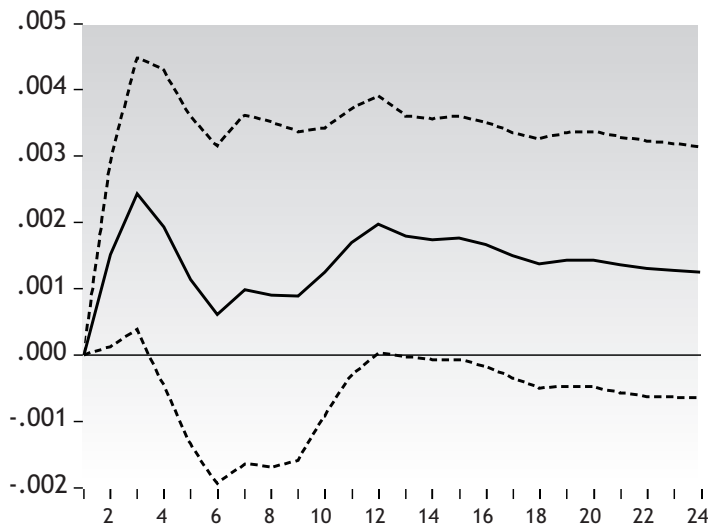
La FIR de PR indica que las autoridades económicas reaccionan aumentando los precios regulados cuando ocurre un shock inflacionario en los mercados no regulados. El incremento de los precios regulados reduce el déficit fiscal y aumenta la recaudación de liquidez monetaria por parte del Estado. Ello puede explicar el signo de la respuesta.

Sin embargo, con la FIR anterior hemos descubierto que esta no es la mejor política, dado que un aumento en PR genera todavía más inflación en los precios no regulados. Si bien el aumento de PR puede contraer el déficit fiscal y la liquidez monetaria, también tiene un efecto costo sobre las producciones privadas y cooperativas por la vía de las tarifas eléctricas, el transporte y los precios de los alimentos.

<sup>9</sup> El paquete estadístico Eviews permite también ver la FIR como tabla, de donde se obtiene el valor exacto de cada respuesta.

**Figura 21.2: Respuesta de los precios regulados (PR) a un shock en los precios no regulados (P)**

(eje horizontal: 24 meses a partir del mes en que ocurre el shock (mes 1)  
eje vertical: respuesta en cada mes (1=100%)  
línea punteada:  $\pm 2$  desviaciones estándar)



Fuente: Resultados del programa Eviews 5

En la tabla 21.3 se expone la descomposición de la varianza de las dos series de tiempo. Ella dejar ver que los shocks en PR explican el 9.6% de P y que los shocks en P explican un porcentaje bastante mayor de PR: 22%. No obstante, en ambos casos las variables están explicadas en un porcentaje alto por shocks propios. Si bien existe una relación significativa entre estos índices de precio, el poder explicativo de cada uno sobre el otro es bajo, existe un grado relativamente alto de exogeneidad. Otras variables relevantes deben incluirse para aumentar el poder explicativo y en consecuencia el poder predictivo de este VAR.

**Tabla 21.3: Descomposición de la varianza del error de pronóstico de 36 meses**  
(cada fila suma 100%)

| Variable explicada: | por shocks en: |      |
|---------------------|----------------|------|
|                     | P              | PR   |
| P                   | 90.4           | 9.6  |
| PR                  | 22.0           | 78.0 |

Fuente: Resultados del programa Eviews 5



# *preguntas*

Intente dar respuesta a las siguientes preguntas. Así sabrá si ha captado los principales mensajes de este capítulo.

1. ¿Cómo se determina el número de ecuaciones y el número de rezagos del VAR?
2. ¿Cuál es la diferencia entre el VAR estructural y el VAR reducido?
3. ¿Qué es la descomposición de Cholesky? ¿En qué se diferencia de la identificación estructural?
4. ¿Una vez estimado el VAR, qué herramientas utilizamos para analizar las relaciones dinámicas entre las variables involucradas?
5. ¿Cómo probamos la causalidad de Granger?



# ejercicios



- 21.1 Se estimó un modelo VAR con cuatro variables. La forma estructural de este modelo es:

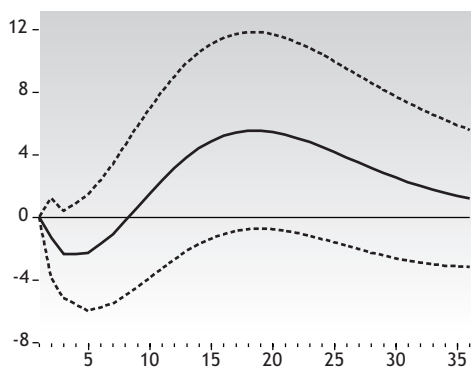
$$AX_t = C(L)X_{t-1} + u_t$$

donde  $X$  es un vector con todas las variables, la matriz  $A$  está formada por los coeficientes  $a_{ij}$  con signo cambiado, los cuales recogen las relaciones contemporáneas entre las variables y  $C(L)$  contiene los coeficientes que relacionan a cada variable con los rezagos del resto de las variables y con sus propios rezagos.

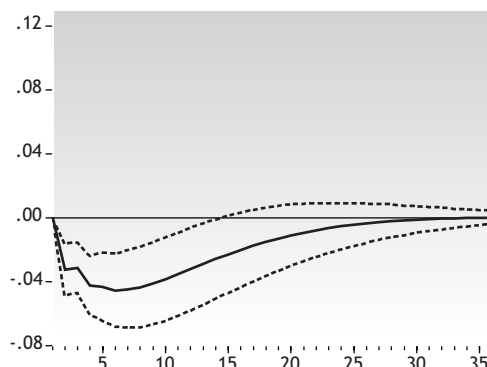
Para identificar el sistema y poder recuperar los parámetros estructurales se definieron las siguientes restricciones en la matriz  $A$ :

$$\begin{pmatrix} 1 & -a_{12} & -a_{13} & -a_{14} \\ 0 & 1 & -a_{23} & -a_{24} \\ 0 & 0 & 1 & -a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

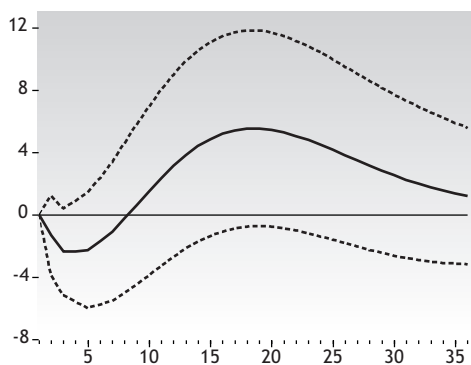
- ¿Estas restricciones son suficiente para identificar el sistema?
  - ¿El sistema está sobreidentificado?
  - ¿Qué tipo de identificación se está utilizando?
  - Interprete las restricciones cero sobre los parámetros  $a_{ij}$ .
- 21.2 De las siguientes funciones impulso-respuesta diga,
- ¿Qué representan las funciones impulso respuesta?
  - ¿Es significativamente distinta de cero la respuesta en el tiempo?
  - ¿En qué período es significativa la respuesta?
  - Analice la forma dinámica de la respuesta.



(a)



(b)



(c)

21.3 Se estimó un modelo VAR monetario para Cuba con datos mensuales de tres variables: el Índice de Precios al Consumidor (P), un agregado monetario (M1) y el tipo de cambio de las casas de cambio (TC). Una vez estimado el VAR se obtuvo los siguientes resultados para la descomposición de la varianza:

|                      | por innovaciones en |      |      |
|----------------------|---------------------|------|------|
|                      | M1                  | P    | TC   |
| Variables explicadas |                     |      |      |
| M1                   | 89.7                | 4.0  | 6.2  |
| P                    | 50.9                | 40.8 | 8.3  |
| TC                   | 39.6                | 3.2  | 57.2 |

- Indique la variable que más depende de los shocks en el resto de las variables en conjunto.
- Indique la variable con un mayor grado de exogeneidad a los shocks en el resto de las variables.
- ¿Cómo se interpretaría el valor de 50.9 en la primera columna?

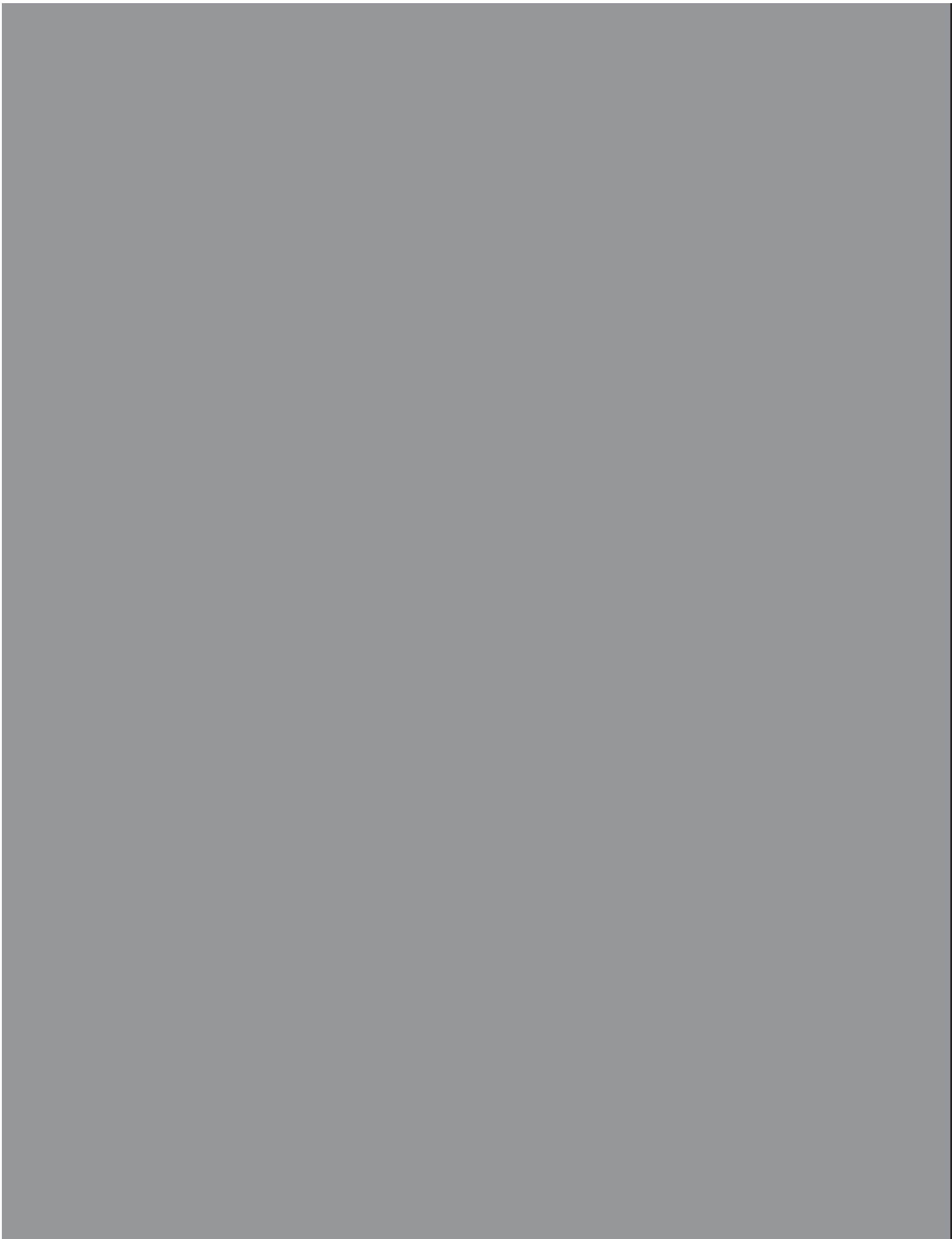
d. ¿Cómo se interpretaría el valor de 57.2 en la tercera columna?

21.4 Reestime el VAR usado como ejemplo en el acápite 21.3 añadiendo ahora la variable tipo de cambio (ver tablas B.9 y B.13 del CD). Es decir, sería un VAR con los precios no regulados (P), los precios regulados (PR) y el tipo de cambio (TC).

- a. Desarrolle la prueba de causalidad de Granger para estas tres variables.
- b. Obtenga las funciones impulso-respuesta y la descomposición de la varianza con el siguiente orden para la descomposición de Cholesky: TC, PR y P. Analice los resultados.
- c. Obtenga las funciones impulso-respuesta y la descomposición de la varianza con el siguiente orden para la descomposición de Cholesky: PR, TC y P. Examine si se modifican los resultados.







• PARTE SEIS

• TRES MODELOS  
• ECONOMÉTRICOS  
• PARA CUBA



La última parte de este texto de Econometría la dedicamos a presentar tres estudios empíricos realizados en el contexto cubano. Uno de ellos de carácter macroeconómico, se refiere a la predicción del PIB de Cuba. El otro modelo se aplica en el ámbito microeconómico y se trata de la estimación de una función de producción para la agricultura cañera en Cuba con el propósito de conducir un análisis de productividad total de factores o efectividad conjunta de los factores de producción para esta actividad. El tercero se centra en el sector turístico y una aproximación a su competitividad dentro del Caribe insular mediante una estimación de datos panel.

CAPÍTULO 22



## MODELO PARA LA PREDICCIÓN DEL PIB CUBANO

El objetivo de este capítulo es presentar de manera resumida un modelo econométrico sencillo que pueda ser utilizado para la predicción del crecimiento económico cubano. Tal modelo surge en el Centro de Estudios de la Economía Cubana con el objetivo de disponer de una proyección propia del Producto Interno Bruto cada año.

El modelo econométrico hace depender el PIB real de las exportaciones corrientes y del gasto del presupuesto para inversión. Con datos anuales desde 1981 se estimó un mecanismo de corrección de errores, partiendo de una relación de cointegración entre las variables. Con el modelo se pueden hacer cada año predicciones del crecimiento real del PIB cubano, dado un escenario para las exportaciones corrientes y el gasto del presupuesto en inversión.

Un elemento que enmarcó la especificación del modelo fue la disponibilidad de información. Se necesitaban series históricas de las variables involucradas para un período relativamente extenso, y más importante aún, tenían que ser datos que pudieran seguirse actualizando regularmente a partir de su publicación en una fecha temprana del año.

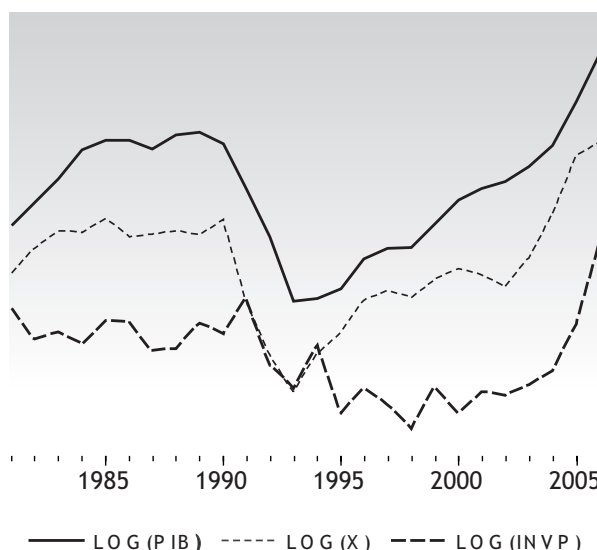
Dado el carácter centralizado de la economía cubana, el Estado tiene determinado control sobre muchas de las variables y a partir de ello se planifica su comportamiento para el año en curso. En este sentido, era conveniente que el crecimiento del PIB estuviera en función de alguna de estas variables, de manera tal, que se pudiera efectuar la predicción tomando en cuenta las decisiones y planes del Estado para el futuro cercano.

A la hora de elegir las distintas opciones de modelo, se priorizó de acuerdo a su capacidad predictiva por encima de su utilidad analítica. La sencillez también fue un rasgo deseado, en función de poder entender mejor las predicciones que ofrece el modelo en base a distintos escenarios.

## 22.1 Datos

Se toman datos anuales desde 1981 del anuario de la Oficina Nacional de Estadísticas (ONE) del PIB a precios constantes de 1997, las exportaciones corrientes y el gasto del presupuesto en inversión. En la serie del PIB, a partir del 2000 se toma el PIB revaluado con la nueva metodología para la medición de los servicios sociales, antes de esta fecha se hace el cambio de base manteniendo la tasa de crecimiento real de la metodología anterior. Los datos están disponibles en el CD acompañante, en la tabla B.14.

Figura 22.1: Evolución de las series en logaritmo



En el gráfico se observa una trayectoria similar entre las variables. La correlación del PIB con las exportaciones (0.92) es mayor que la correlación con los gastos del presupuesto en inversión (0.63).

¿Qué fundamentos teóricos sugieren colocar al PIB en función de las exportaciones y la inversión?

Ambas variables, según la contabilidad nacional, forman parte del cálculo del PIB por el lado del gasto. Pero ellas también contienen información sobre variables no incluidas en el modelo y que también afectan el PIB. Por el papel que juega el Estado en la economía cubana, muchos sectores se dinamizan una vez que crece el gasto del presupuesto en inversión. Las inversiones tienen un efecto multiplicador al expandir la demanda doméstica de diferentes producciones y servicios. Pero sobre todo, determina la acumulación del factor capital, indispensable para sostener el crecimiento de la economía en el largo plazo.

Por su parte, la inclusión de las exportaciones sigue la idea del modelo teórico de Crecimiento Restringido por la Balanza de Pagos (CRBP) de Thirlwall (1979). Este modelo destaca el rol de las exportaciones debido a que ellas generan



automáticamente ingresos externos para pagar las importaciones; una diferencia clave con los otros componentes de la demanda. De esta forma se puede reflejar el hecho de que el crecimiento del PIB depende de la importación de insumos intermedios para el funcionamiento corriente de la economía y de bienes de capital para lograr una mayor expansión de las fuerzas productivas. Así, un aumento de las exportaciones incrementa la disponibilidad de moneda extranjera para pagar importaciones, beneficiándose por esta vía el crecimiento económico. Al tomar las exportaciones en términos corrientes se recoge implícitamente parte de las variaciones en los términos de intercambio, los cuales pueden también ayudar a relajar las restricciones externas y tener un efecto sobre el PIB.

Otras estimaciones del modelo CRBP para Cuba pueden encontrarse en Moreno-Brid (2000), Mendoza y Robert (2002), Mañalich y Quiñones (2004), Alonso y Sánchez-Egozcue (2005), Cribeiro y Triana (2005) y Vidal y Fundora (2008).

En la Tabla 22.1 se muestran los resultados de la prueba de Dickey-Fuller aumentado (ADF). Las tres series son no estacionarias en niveles e integradas de orden 1,  $I(1)$ .

**Tabla 22.1: Prueba de estacionariedad de las series  
(Logaritmo de las series, período 1981-2006)**

| Serie   | ADF     | Constante y Tendencia | Rezagos |
|---------|---------|-----------------------|---------|
| PIB     | 0.44    | No                    | 1       |
| X       | -1.78   | constante             | 1       |
| INVP    | 0.31    | No                    | 0       |
| D(PIB)  | -2.49** | No                    | 0       |
| D(X)    | -3.01*  | No                    | 0       |
| D(INVP) | -5.64*  | No                    | 0       |

(1) \* Estacionaria al 5%, \*\* Estacionaria al 10%

## 22.2 Estimación del Modelo

Como las tres series tienen el mismo orden de integración, primero se prueba que las variables siguen una relación de largo plazo, es decir, están cointegradas. En la tabla 22.2 se presenta la relación de cointegración estimada tal y como aparece en la salida del programa Eviews 5.0.

**Tabla 22.2: Relación de largo plazo  
Crecimiento del PIB (%)**

| Dependent Variable: LOG(PIB) |                 |                       |             |           |
|------------------------------|-----------------|-----------------------|-------------|-----------|
| Method: Least Squares        |                 |                       |             |           |
| Date: 01/15/07 Time: 14:44   |                 |                       |             |           |
| Sample: 1981 2006            |                 |                       |             |           |
| Included observations: 26    |                 |                       |             |           |
| Variable                     | Coefficient     | Std. Error            | t-Statistic | Prob.     |
| C                            | 6.624913        | 0.355725              | 18.62370    | 0.0000    |
| LOG(X)                       | <b>0.363011</b> | 0.039611              | 9.164325    | 0.0000    |
| LOG(INVP)                    | <b>0.086379</b> | 0.053303              | 1.620526    | 0.1188    |
| R-squared                    | 0.869662        | Mean dependent var    |             | 10.35302  |
| Adjusted R-squared           | 0.858328        | S.D. dependent var    |             | 0.156516  |
| S.E. of regression           | 0.058911        | Akaike info criterion |             | -2.717397 |
| Sum squared resid            | 0.079823        | Schwarz criterion     |             | -2.572232 |
| Log likelihood               | 38.32617        | F-statistic           |             | 76.73207  |
| Durbin-Watson stat           | 0.617678        | Prob(F-statistic)     |             | 0.000000  |

Se destaca en las elasticidades de largo plazo que las exportaciones tienen un mayor efecto relativo sobre el PIB que el gasto en inversión del presupuesto. En la tabla 22.3 se observa que los residuos de esta relación son estacionarios al 1% y ello permite decir que las variables están cointegradas.<sup>1</sup>

<sup>1</sup> No obstante, el valor de ADF no es menor que los valores críticos para una muestra de 50 observaciones reportados por Engle y Yoo (1987).

Tabla 22.3: Prueba ADF sobre residuos de la relación de largo plazo

|                                                  |           |  |             |        |
|--------------------------------------------------|-----------|--|-------------|--------|
| Null Hypothesis: RESID01 has a unit root         |           |  |             |        |
| Exogenous: None                                  |           |  |             |        |
| Lag Length: 1 (Automatic based on SIC, MAXLAG=8) |           |  |             |        |
|                                                  |           |  | t-Statistic | Prob.* |
| Augmented Dickey-Fuller test statistic           |           |  | -2.897333   | 0.0057 |
| Test critical values:                            | 1% level  |  | -2.664853   |        |
|                                                  | 5% level  |  | -1.955681   |        |
|                                                  | 10% level |  | -1.608793   |        |
| *MacKinnon (1996) one-sided p-values.            |           |  |             |        |

Se estima entonces un mecanismo de corrección de errores donde el crecimiento del PIB real se hace depender del crecimiento de las exportaciones corrientes y del crecimiento del gasto del presupuesto para inversión, así como de los residuos de la relación de largo plazo rezagados un período. En la tabla 22.4 se presenta la estimación del Mecanismo de Corrección de Errores tal y como aparece en la salida del programa Eviews 5.0.

Se incluyen en el modelo rezagos del PIB así como de las exportaciones y del gasto en inversión del presupuesto, para tomar en cuenta que la relación entre las variables no sólo ocurre en el propio año sino que puede demorar varios años en manifestarse completamente. A pesar de que los rezagos no son significativos se mantienen en la ecuación debido a que mejora el criterio de Akaike. Según las elasticidades de corto plazo, las exportaciones siguen teniendo un mayor efecto sobre el crecimiento económico que el gasto del presupuesto para inversiones.

**Tabla 22.4: Mecanismo de corrección de errores**

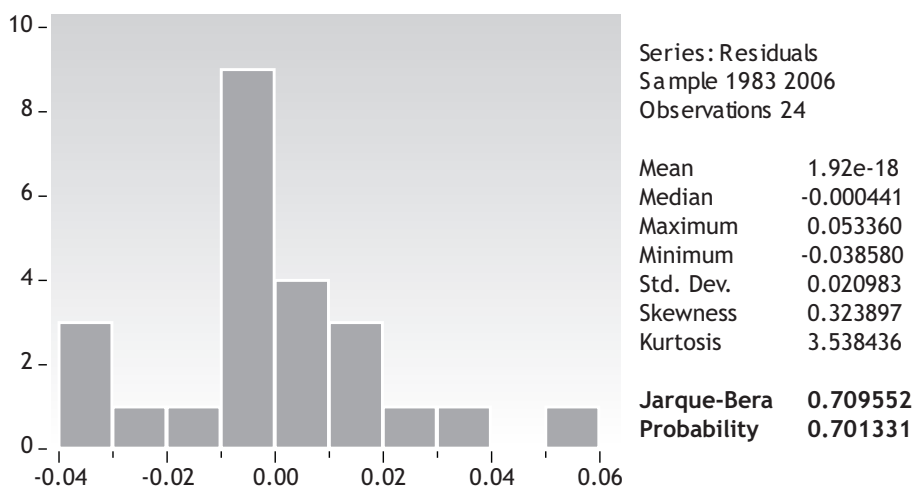
| Dependent Variable: DLOG(PIB)                       |                 |                       |             |           |
|-----------------------------------------------------|-----------------|-----------------------|-------------|-----------|
| Method: Least Squares                               |                 |                       |             |           |
| Date: 01/15/07 Time: 14:45                          |                 |                       |             |           |
| Sample(adjusted): 1983 2006                         |                 |                       |             |           |
| Included observations: 24 after adjusting endpoints |                 |                       |             |           |
| Variable                                            | Coefficient     | Std. Error            | t-Statistic | Prob.     |
| C                                                   | 0.005588        | 0.005237              | 1.066948    | 0.3009    |
| RESID01(-1)                                         | -0.351298       | 0.129222              | -2.718555   | 0.0146    |
| DLOG(PIB(-1))                                       | 0.398087        | 0.134832              | 2.952458    | 0.0089    |
| DLOG(X)                                             | <b>0.203146</b> | 0.029727              | 6.833788    | 0.0000    |
| DLOG(INVP)                                          | <b>0.080215</b> | 0.029581              | 2.711686    | 0.0148    |
| DLOG(X(-1))                                         | -0.044238       | 0.058656              | -0.754200   | 0.4610    |
| DLOG(INVP(-1))                                      | 0.037016        | 0.029944              | 1.236158    | 0.2332    |
| R-squared                                           | 0.904966        | Mean dependent var    |             | 0.015563  |
| Adjusted R-squared                                  | <b>0.871425</b> | S.D. dependent var    |             | 0.068067  |
| S.E. of regression                                  | 0.024407        | Akaike info criterion |             | -4.349409 |
| Sum squared resid                                   | 0.010127        | Schwarz criterion     |             | -4.005809 |
| Log likelihood                                      | 59.19290        | F-statistic           |             | 26.98055  |
| Durbin-Watson stat                                  | 2.175651        | Prob(F-statistic)     |             | 0.000000  |

### 22.3 Validación del Modelo

El modelo cumple los supuestos de los mínimos cuadrados ordinarios de no autocorrelación y normalidad de los residuos. En la tabla 22.5 se muestran los resultados de las pruebas realizadas con el programa Eviews 5.0.

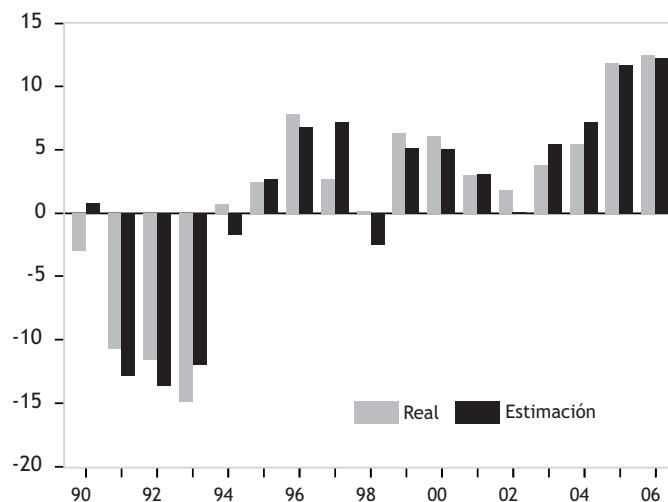
Tabla 22.5: Supuestos de no autocorrelación y normalidad

| Breusch-Godfrey Serial Correlation LM Test: |              |             |              |
|---------------------------------------------|--------------|-------------|--------------|
| F-statistic                                 | 0.981<br>274 | Probability | 0.46<br>7848 |
| Obs*R-squared                               | 6.965<br>001 | Probability | 0.22<br>3256 |



En la figura 22.2 se muestra el crecimiento real del PIB y la estimación dentro de la muestra que hace el modelo desde 1990. Se observa que la estimación logra recoger una parte importante de la trayectoria del crecimiento del PIB, permite explicar los años de mayor caída del PIB (1991-1993), la recuperación posterior, así como la aceleración del crecimiento desde el 2004. El  $R^2$  ajustado del modelo es de 87%.

Figura 22.2: Capacidad de predicción dentro de la muestra: 1990-2006



Para evaluar la capacidad de predicción fuera de la muestra, se estima el modelo de corrección de errores hasta 1996 y se hace una predicción un período hacia delante (estática) desde 1997 hasta el 2006. A partir de los errores de predicción se toman diferentes medidas que se muestran en la Tabla 22.6. Se observa que el error absoluto medio es de 1.78% y que el sesgo de las predicciones es bajo.

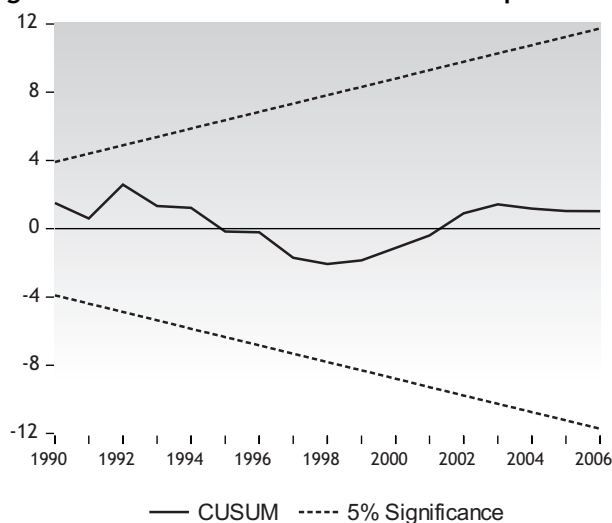
**Tabla 22.6: Capacidad de predicción fuera de la muestra.**  
**Evaluación de los errores de predicción**  
**un período hacia delante: 1997-2006.**

|                                |                 |
|--------------------------------|-----------------|
| Forecast: PIBF                 |                 |
| Actual: DLOG(PIB)              |                 |
| Forecast sample: 1997 2006     |                 |
| Included observations: 10      |                 |
| Root Mean Squared Error        | 0.022846        |
| Mean Absolute Error            | <b>0.017836</b> |
| Mean Absolute Percentage Error | 166.3977        |
| Theil Inequality Coefficient   | 0.183010        |
| Bias Proportion                | <b>0.063321</b> |
| Variance Proportion            | 0.053298        |
| Covariance Proportion          | 0.883381        |

La prueba de CUSUM (la cual estudiamos en el capítulo 13) indica que en el período de estimación se obtiene una relación estable entre las variables.<sup>2</sup>

<sup>2</sup> En el trabajo de Vidal y Fundora (2008) sí se encuentra inestabilidad en los parámetros en un período más extenso: 1950-2006.

Figura 22.3: Prueba de estabilidad de los parámetros



## 22.4 Predicción para 2007

Dado un escenario para el 2007 en que las exportaciones corrientes crezcan un 7.0% y el gasto del presupuesto para inversión sea de 4817.3 MMP (según el presupuesto aprobado para el 2007), se pronosticó para ese año un crecimiento del PIB real para la economía cubana **alrededor de 8%**. Tomando unas bandas de 1.5% (aproximadamente 50% de probabilidad) se obtendría una banda mínima de crecimiento de 6.5% y un máximo de 9.5%.

Tabla 22.7: Pronóstico del PIB cubano para 2007

|                                       | Mínimo | Centro*    | Máximo |
|---------------------------------------|--------|------------|--------|
| Crecimiento real del PIB (porcentaje) | 6.5    | <b>8.0</b> | 9.5    |

Un resultado menor en las exportaciones durante el 2007 presionaría a un crecimiento del PIB más cerca del mínimo, y viceversa. Así, el pronóstico estuvo sujeto a la evolución que efectivamente tuvieron la exportación de servicios profesionales, el turismo, el azúcar y el níquel, fundamentalmente.

Finalizado el año 2007 el crecimiento observado del PIB cubano fue de 7.3%, lo cual corrobora la capacidad predictiva del modelo fuera de la muestra. Este mismo modelo pronosticó para 2008 un crecimiento del PIB cubano de 5.1% y el dato efectivamente observado al terminar el año fue 4.1%. Para 2009 la predicción fue de 1.0% y el dato finalmente observado fue de 1.4%. El Centro de Estudios de la Economía Cubana pudo anticipar la desaceleración de la economía desde 2007 utilizando técnicas econométricas de series de tiempo (ver el Boletín Cuatrimestral del Centro en [www.ceec.uh.cu](http://www.ceec.uh.cu))

CAPÍTULO 23





# EFFECTIVIDAD CONJUNTA DE LOS FACTORES DE PRODUCCIÓN EN LA AGRICULTURA CAÑERA CUBANA

## 23.1 Introducción

Una de las cuestiones que se han puesto de manifiesto en numerosos estudios sobre la agricultura cubana,<sup>1</sup> es la relativamente débil respuesta de la producción de este origen a los incrementos en la dotación de insumos del sector. Esta incoherencia en la respuesta fue especialmente notable con anterioridad a 1989, cuando todavía se disfrutaba de restricciones débiles en el abastecimiento de insumos para estas producciones.

A partir de 1993, la política económica se ha dirigido a conseguir condiciones de oferta agropecuaria más favorables: a través del fomento de explotaciones más pequeñas, más manejables y mejor organizadas, donde se pueden aplicar tecnologías de mayor sustentabilidad; abriendo nuevos espacios de mercado, que ayudan a mitigar la distorsión en los precios de acopio que regían para este sector; modificando algunos de estos precios de acopio e introduciendo nuevos incentivos a los productores. Todas estas medidas deben lograr una mejor asignación de los todavía hoy escasos recursos a disposición del sector, y por consiguiente un incremento en su eficiencia y competitividad. Es importante entonces también conseguir alguna evidencia empírica acerca de los efectos de estos cambios de política en la evolución de la producción agropecuaria.

Esta aplicación se concentra particularmente en el estudio de la efectividad conjunta de los factores de producción en la agricultura cañera cubana, fuente de

<sup>1</sup> Ver Bu y col. (1996), Fernández (1997 y 2002), García y Perera (1997), García y Fernández (1998), García y col. (1999), Nova (1997), Pérez (1999), Rodríguez (2000), García (2002), García y Aguilar (2002).

la materia prima fundamental para la producción azucarera, que hasta hace pocos años fuera el principal renglón de exportación de este país y todavía se mantiene entre los principales rubros. Su propósito es brindar evidencia acerca de cuales son los factores que explican la variabilidad de esta producción, que están en la base también de la pérdida de competitividad del azúcar cubano en el mercado internacional y en mercados específicos de destino de este producto.<sup>2</sup>

## 23.2 El Marco y el Modelo Teórico

Para analizar la efectividad en la utilización de los principales recursos productivos en una actividad agrícola, así como en cualquier otro sector de la economía, usualmente se calculan índices que relacionan la producción con los distintos factores, tales como la productividad del trabajo (producción contra trabajadores), el rendimiento de la tierra (producción contra superficie en explotación) o el rendimiento de la maquinaria agrícola (producción contra maquinaria). Tal análisis sólo toma en cuenta la efectividad de cada recurso de forma individual; es decir, al calcular la productividad del trabajo o el rendimiento de la maquinaria se están adjudicando los resultados productivos a la acción exclusiva del recurso en cuestión, y se hace abstracción de la influencia del resto de los factores que participan en el proceso de producción. Sin embargo, en realidad todos los factores son importantes y no es posible obtener resultados si alguno de ellos falla; de modo que la magnitud y la dinámica de cada uno de los índices de efectividad individuales lleva implícita la influencia de la utilización del resto de los factores que se excluyen del análisis en cada caso.

El análisis de la efectividad conjunta de los factores de producción –también denominada productividad total de los factores– supone justamente segregar el efecto de los diferentes factores y su proporcionalidad, con el propósito de evaluar la influencia de **factores de tipo cualitativo** sobre el crecimiento de la producción. Entre estos factores cualitativos podemos mencionar, por ejemplo, el nivel tecnológico de la maquinaria utilizada, el nivel de calificación de la fuerza de trabajo y la organización de la producción.

La productividad total de los factores puede interpretarse también como **las ganancias en eficiencia que logra la economía por una mejor asignación de los recursos** y no por incrementos cuantitativos en las cantidades de los mismos.

Estas son las influencias que trata de capturar la productividad total de los factores: el efecto de los factores cualitativos y el efecto de una mejor asignación de los recursos físicos y técnicos disponibles.

Existe una estrecha relación entre competitividad y eficiencia, por lo que en los análisis de la situación competitiva de cualquier rubro se recomienda tomar en consideración, además de índices que caractericen los avances o retrocesos en los

---

<sup>2</sup> Ver García y Guerra (2005, pp. 121-123).

mercados de destino,<sup>3</sup> también índices que den cuenta de los procesos de cambio técnico o del estado de la eficiencia de la producción. Es decir, se trata de realizar estudios de efectividad conjunta de los factores.

Esta consideración es de la mayor importancia, ya que entre los más importantes campos de actuación de la política económica están la transferencia de tecnología y los programas de promoción. Los esfuerzos desplegados en dotar de una mejor tecnología a la agricultura y la asignación de factores productivos a ese sector están en la base de su desempeño competitivo.

Entre los enfoques para medir la efectividad conjunta de los factores (ECF) se encuentran: la construcción de *índices* a partir de la *contabilidad del crecimiento*, la *estimación econométrica* de funciones o fronteras de producción y la aplicación de métodos *no paramétricos* para la estimación de fronteras de producción.<sup>4</sup>

La construcción de índices a partir de la contabilidad del crecimiento se basa en comparar la tasa de variación de la producción con la tasa de variación media ponderada de los insumos productivos. El factor de ponderación habitualmente utilizado es la participación de cada insumo en los costos totales. La relación entre ambas tasas de variación es una medida de la tasa de cambio de la efectividad multifactorial agregada. Si, por ejemplo, la producción crece más rápidamente que el conjunto de factores, esto quiere decir que existe otra fuente de crecimiento de la producción que no reside en la variación de los factores en sí, sino en una mejor combinación entre ellos o en un incremento de su calidad.

Sin embargo, para calcular las participaciones de los factores en los costos es preciso contar con información sobre los *precios de mercado* de los distintos factores. A falta de mercados para determinados factores se pueden aplicar los otros enfoques. Por ejemplo, en Zhang (1997) se presenta un estudio sobre la ECF en la producción de cereales de la antigua URSS que se basa en la estimación de una función de producción promedio. Esta metodología atribuye la discrepancia entre el crecimiento observado en la producción y su crecimiento potencial —originado por la variación de los distintos factores ponderados por las elasticidades obtenidas para cada uno de ellos según una función de producción Cobb-Douglas— a los cambios en la efectividad conjunta de los factores.<sup>5</sup>

Una función de producción de tipo Cobb-Douglas para el sector agropecuario presenta el siguiente modelo:

$$Q_t = \prod_{i=1}^n F_{it}^{\alpha_i} e^{u_t} \quad (23.1)$$

Donde  $Q_t$  es la producción;  $F_{it}$  el factor de producción  $i$ ; y  $t$  es un subíndice que indica el período temporal. Entre los principales factores de producción en el

<sup>3</sup> Se trata de los indicadores de competitividad ex-post.

<sup>4</sup> Ver Zepeda (2001, pp. 4-6) y Fulginiti y Perrin (1997, p. 374).

<sup>5</sup> Ver Zhang (1997, pp. 204-206).

sector agropecuario están la fuerza de trabajo ( $L_t$ ), la tierra en explotación ( $T_t$ ), los fertilizantes ( $F_t$ ), los herbicidas y pesticidas ( $H_t$ ), y la maquinaria agrícola y los animales de trabajo ( $E_t$ ). Por lo general en esta modelación se incluye alguna variable para reflejar la influencia de las condiciones climáticas, como pueden ser las precipitaciones ( $P_t$ ).

El modelo (23.1) se puede linealizar, aplicando logaritmos, lo que se denota anteponiendo una “ $L$ ” a cada variable.

$$LQ_t = \sum_{i=1}^n \alpha_i L F_{it} + u_t \quad (23.2)$$

Los coeficientes  $\alpha_i$  del modelo (23.2) son las elasticidades de la producción respecto al factor correspondiente. Las elasticidades estimadas son las que se utilizan para calcular el crecimiento potencial de la producción según la variación de los factores.

La derivada con respecto al tiempo del logaritmo de una variable puede ser interpretada como la tasa de crecimiento de dicha variable. Si se utiliza la minúscula de cada variable para denotar su tasa de crecimiento, la derivada con respecto al tiempo de la ecuación (23.2) conduce a la ecuación (23.3) a continuación.

$$q_t = \sum_{i=1}^n \alpha_i f_{it} + \partial u_t / \partial t \quad (23.3)$$

De acuerdo a la metodología discutida anteriormente, el último término de (23.3) —la derivada de los residuos—, mediría los cambios en la ECF.

Lógicamente, este método *residual* para determinar la contribución de la ECF tiene ciertas debilidades. Por ejemplo, los residuos de un modelo econométrico incorporan otros elementos estocásticos y, posiblemente, la influencia de las variables excluidas de dicho modelo. Al tratarse de una medición indirecta, se pueden estar atribuyendo aportes a la ECF que sencillamente no le corresponden. Además, la estimación de la función de producción con el instrumental econométrico es muy exigente en cuanto a cantidad de datos.

Otro problema vinculado a este método es el relacionado con la forma funcional de la función de producción.<sup>6</sup> La investigación en este campo se enfocó fundamentalmente al desarrollo de formas funcionales *adecuadas* para la representación de la tecnología. Entre las especificaciones funcionales más importantes se encuentran: la Cobb-Douglas, así denominada en honor a sus creadores e introducida en 1928; las funciones de producción de elasticidad de sustitución constante, conocidas por sus siglas en inglés CES,<sup>7</sup> desarrolladas por Arrow, Chenery, Minhas y

<sup>6</sup> Entre los objetos de estudio de la microeconomía ocupan un lugar bien importante la búsqueda de una adecuada descripción formal de la tecnología y las posibles medidas de la eficiencia, ambos íntimamente relacionados. De hecho, obtener medidas sobre la eficiencia (o ineficiencia) del proceso productivo ha sido la principal motivación para formalizar la relación producto-insumos.

<sup>7</sup> Siglas en inglés por *constant elasticity of substitution*.

Solow en 1961; y la especificación transcendental-logarítmica,<sup>8</sup> introducida por Christensen, Jorgenson y Lau en 1973.

En la presente aplicación se ha empleado justamente el método residual anteriormente descrito para la determinación de la ECF. Una de las principales razones que motivó la experimentación con esta técnica fue encontrar medidas de eficiencia que no descansaran demasiado en los precios de los productos y los factores, toda vez que en Cuba no existen mercados para algunos de los medios de producción —como es el caso de la tierra— y los precios son fundamentalmente precios administrados. Además, resulta importante tener una medida que tome en cuenta el efecto conjunto, que se consigue a partir de la consideración de una relación de producción, más que de cada factor de producción de forma aislada.

### 23.3 Los Datos, el Modelo Estimado y los Resultados

La estimación de la función de producción cañera que se presenta se realizó a partir de una serie de tiempo para el período 1975-2001 (ver tabla B.8 en el CD acompañante). Como medida de los resultados del proceso productivo se utilizó la cantidad de caña producida ( $Q_t$ ), en millones de toneladas métricas. Como medida de los factores de producción se utilizó: para el caso de la tierra ( $T_t$ ), el área a zafra en miles de hectáreas; para los insumos de tipo químico, se contó con información sobre los fertilizantes aplicados ( $F_t$ ), en miles de toneladas de sustancia activa; no se pudo obtener información ni para el potencial energético de la maquinaria agrícola, ni para el consumo de combustible para todo el período que abarca el estudio, que son las variables que habitualmente se utilizan para modelar los medios de trabajo,<sup>9</sup> por lo que se optó por utilizar los datos disponibles para la inversión en la agricultura cañera ( $INV_t$ ), a falta de una medida más precisa; tampoco se ha conseguido información sobre la fuerza de trabajo activa en las labores de la caña para todo el período de la muestra.<sup>10</sup> Así, los resultados del estudio pueden presentar algún sesgo al faltar algunos de los factores en la especificación y/o por aplicar medidas no muy adecuadas. Para modelar la influencia de las condiciones meteorológicas, tan importantes para el desempeño de cualquier actividad agrícola, se incorporaron las precipitaciones ( $P_t$ ) captadas por la red de pluviómetros de las zonas cañeras.

<sup>8</sup> Abreviadamente, se le denomina *translog*.

<sup>9</sup> Se trata de variables *proxies*.

<sup>10</sup> Si bien estudios anteriores sobre eficiencia, realizados para la totalidad del sector, arrojaron que la propia disponibilidad de fuerza de trabajo —medida a partir de los trabajadores promedio anuales—, no influye significativamente en la producción agropecuaria (ver García (2002) y García y Aguilar (2002)). Esto puede obedecer a que la fuerza de trabajo ocupada en las labores del sector estuvo atada durante la mayor parte del período estudiado a una nómina fija en el caso de las empresas estatales.

La fuente de la información en todos los casos ha sido el *MINAZ*<sup>11</sup> para la mayor parte del período. En el caso de las inversiones y las precipitaciones, variables para las que faltaron algunos datos, éstos se completaron a partir de estimaciones propias que se sustentaron en el volumen de inversiones para todo el sector agropecuario y la lluvia media anual, reportados por la ONE (2003).

En el anexo 23.1 se muestra la dinámica de las diferentes variables para el período que se analiza. Con el objetivo de facilitar su comparación y para evitar problemas de escala, estas variables están representadas como índices con base en 1975, año inicial del período.

La forma funcional utilizada para la función de producción fue lineal en los logaritmos de las variables, es decir, una función tipo Cobb-Douglas. Así, los regresores pueden interpretarse como elasticidades y ser después utilizados para ponderar los crecimientos de los respectivos factores en el cálculo de la ECF. Si bien es cierto que esa forma funcional impone restricciones a las elasticidades de sustitución de los factores entre sí, la experiencia indica que resulta una buena aproximación para otras formas más flexibles. Además, esas formas flexibles tienen como desventaja la multicolinealidad entre variables explicativas por “diseño”, lo que hace las estimaciones poco eficientes.

La especificación estimada de inicio fue la siguiente:

$$LQ_t = \alpha_0 + \alpha_1 LT_t + \alpha_2 LF_t + \alpha_3 LINV_t + \alpha_4 LP_t + u_t \quad (23.4)$$

Se calculó la matriz de correlación lineal entre las variables involucradas en el modelo (23.4), como parte de las estadísticas básicas. También se investigó la estacionariedad de las series de tiempo para las variables, aplicando las pruebas de DF y ADF. En el anexo 23.1 se resumen los resultados. Se debe enfatizar que el tamaño de la muestra objeto de estudio es pequeño (apenas 27 observaciones), sin que sea posible ampliarlo por el momento.<sup>12</sup> En estas condiciones es más difícil garantizar que los resultados de las pruebas estadísticas sean “correctos”, sobre todo en el caso de pruebas como la *ADF*, que tienen validez asintótica. Para la realización de las pruebas *ADF* se aplicó el método sugerido por Doornik y Hendry (1994).<sup>13</sup>

No fue posible rechazar que se tratara de procesos integrados en el caso de las variables en niveles; sólo la variable precipitaciones podría ser estacionaria. El

<sup>11</sup> Queremos agradecer expresamente al Dr. Nelson Labrada y los compañeros del área de economía del Ministerio del Azúcar (*MINAZ*), así como al Ms. Federico Sulroca, por su contribución con información para llevar adelante este estudio.

<sup>12</sup> La serie se extendió sólo hasta el año 2001, pues a partir del 2002 es que se despliega ya totalmente el programa de reestructuración de la agroindustria azucarera. En el momento en que se desarrolló esta aplicación, apenas habían transcurrido cuatro años después de esta medida, tiempo demasiado breve como para poder demostrar desde el punto de vista estadístico el cambio estructural que debe haberse verificado en la función de producción.

<sup>13</sup> Ver Doornik y Hendry (1994, pp. 94-95). Este método consiste en determinar el tamaño del rezago aplicando la siguiente estrategia: comenzar el análisis con un número elevado de rezagos y seleccionar el mayor que resulte significativo en la especificación dada.

hecho de que al menos dos variables del conjunto ostenten el mismo orden superior de integración  $-I(1)$  en este caso—, abre la posibilidad de poder encontrar una relación de cointegración entre ellas, que permita estimar la función de producción en niveles. Entonces, se exploró la posibilidad de cointegración: se estimó el modelo de regresión lineal simple (MRLS) originalmente propuesto y se investigó la estacionariedad de sus residuos, no encontrándose evidencia contraria.

Se realizó una variedad de pruebas para verificar la validez de los supuestos implícitos en este modelo. Los problemas de especificación vinculados al MRLS se presentan en el anexo 23.2. Al atender a esos problemas, el modelo quedó especificado como un modelo de regresión lineal dinámico (MRLD), según se muestra en la ecuación (23.5):

$$LQ_t = \alpha_0 + \alpha_1 LT_t + \alpha_2 LF_t + \alpha_3 LT_{t-1} + \alpha_4 LF_{t-1} + \alpha_5 LINV_{t-1} + \alpha_6 LP_{t-1} + u_t \quad (23.5)$$

Los resultados de la estimación del modelo (23.5) se recogen de forma resumida en la tabla 23.1.<sup>14</sup>

**Tabla 23.1: Estimación del modelo dinámico para la función de producción cañera**

| Variabes          | Coefficientes | t         | Probabilidad |
|-------------------|---------------|-----------|--------------|
| Constante         | -5.093504     | -3.271566 | 0.0040       |
| $LT_t$            | 1.082355      | 4.348194  | 0.0003       |
| $LF_t$            | 0.091816      | 3.556150  | 0.0021       |
| $LT_{t-1}$        | -0.479896     | -2.676877 | 0.0149       |
| $LF_{t-1}$        | 0.109637      | 3.409469  | 0.0029       |
| $LP_{t-1}$        | 0.444228      | 4.775201  | 0.0001       |
| $LINV_{t-1}$      | 0.129440      | 14.24227  | 0.0000       |
| No. observaciones | 26            |           |              |
| $R^2$ ajustado    | 0.964802      |           |              |
| F                 |               | 115.2097  | 0.0000       |
| Durbin Watson     |               | 1.462217  |              |

Fuente: Elaboración propia.

Entre las variables contemporáneas sólo resultan significativas la tierra y los fertilizantes. Las inversiones y las precipitaciones tienen un efecto retardado en un año, lo que es coherente con las características de la producción cañera: las inversiones ciertamente deben influir después de cierto *lag*, que no debe ser excesivo en el caso de una actividad agrícola donde los componentes fundamentales

<sup>14</sup> El anexo 23.2 recoge un resumen de las pruebas de incorrecta especificación para el MRLD.

son maquinaria agrícola (de rápida asimilación) y fomento de plantaciones; en el caso de las precipitaciones, ellas son particularmente importantes en el período de crecimiento de la plantación, es decir, antes de la zafra y no durante la misma. Los fertilizantes muestran además un efecto residual, pues queda significativo su primer rezago. Resulta también interesante que el área a zafra anterior influye negativamente en la producción de la siguiente. Ello apunta a la importancia de atender a la necesaria continuidad del ciclo productivo: el área que se lleva a zafra cada año es el factor que más influye en la producción de caña (ver la tabla 23.1), sin embargo, si se exceden los límites que garantizan la zafra futura ello pone riesgo los resultados por venir.

A partir de las elasticidades obtenidas para los factores significativos, se ponderó su crecimiento por período y se sustrajo del crecimiento de la producción para determinar la contribución de la ECF. Los resultados obtenidos se muestran en la tabla 23.2.

**Tabla 23.2: Efectividad conjunta de los factores de producción en la actividad cañera, 1976-2001**

| Tasa media de crecimiento*                                                                                                                                    | 1976-80      | 1980-85      | 1985-89      | 1989-94      | 1994-97       | 1997-2001    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|--------------|--------------|--------------|---------------|--------------|
| <b>Producción de caña</b>                                                                                                                                     | <b>5.1%</b>  | <b>2.3%</b>  | <b>3.0%</b>  | <b>-9.4%</b> | <b>-1.7%</b>  | <b>-3.9%</b> |
| <b>Factores</b>                                                                                                                                               | <b>2.0%</b>  | <b>3.4%</b>  | <b>0.3%</b>  | <b>-7.1%</b> | <b>9.8%</b>   | <b>6.5%</b>  |
| Tierra                                                                                                                                                        | 3.3%         | 0.1%         | 0.3%         | -1.2%        | 0.0%          | -4.9%        |
| Tierra rezagada                                                                                                                                               | 4.3%         | 1.4%         | -1.4%        | -1.0%        | 84.9%         | -3.3%        |
| Fertilizante                                                                                                                                                  | 5.6%         | 5.1%         | -1.0%        | -26.9%       | 1.0%          | -4.1%        |
| Fertilizante rezagado                                                                                                                                         | 12.7%        | 5.5%         | 0.3%         | -16.9%       | 81.5%         | -19.4%       |
| Inversión rezagada                                                                                                                                            | -11.4%       | 17.5%        | 0.3%         | -11.8%       | -53.7%        | 109.7%       |
| <b>Clima</b>                                                                                                                                                  | <b>4.3%</b>  | <b>-1.3%</b> | <b>3.4%</b>  | <b>-1.3%</b> | <b>3.3%</b>   | <b>-1.9%</b> |
| Precipitaciones rezagadas                                                                                                                                     | 9.6%         | -2.9%        | 7.5%         | -2.9%        | 7.4%          | -4.4%        |
| <b>ECF</b>                                                                                                                                                    | <b>-1.1%</b> | <b>0.1%</b>  | <b>-0.6%</b> | <b>-1.0%</b> | <b>-14.8%</b> | <b>-8.4%</b> |
| *Las tasas de crecimiento para la producción y los factores por período se han obtenido como el promedio de las de un año a otro al interior de cada período. |              |              |              |              |               |              |

Fuente: Elaboración propia.

Con anterioridad a 1980 la ECF fue negativa, en consecuencia, los crecimientos en la disponibilidad de factores y las favorables condiciones climáticas no se trasladaron en su totalidad a la producción. El único tramo con contribución positiva corresponde a los años 1980-1985. Este fue el período con mayor crecimiento de los factores, pero también con mayor contracción en el régimen de lluvias. En la segunda mitad de los 80 el aporte de la ECF vuelve a ser negativo. En este lapso los factores de producción crecen poco, pero los climáticos resultaron favorables.

Entre 1989 y 1994 la caída de la producción se explica mayormente por la abrupta disminución de los suministros de origen importado (observar en la tabla



23.2 las contracciones en la disponibilidad de fertilizantes e inversión). El régimen de lluvias tampoco fue favorable. Pero la producción decreció más que lo que ayudarían a explicar la adversidad del clima y la contracción de los factores. Las difíciles condiciones que se crearon y la desorganización que esto generó en las entidades productoras, que no estaban acostumbradas —y por lo tanto no estaban preparadas— a operar con semejante escasez de insumos, puede haber sido la responsable de este comportamiento. Por ejemplo, entre los problemas que aquejan a la agroindustria están algunos de tipo organizativo como la prolongación excesiva de las contiendas, los incumplimientos de los cronogramas y volúmenes de siembra, así como de las atenciones para el control de malezas.<sup>15</sup>

La contribución de la ECF después de 1994 se vuelve más desfavorable, si bien entre 1997-2001 tiende a mejorar. Entre 1994 y 1997, ante la dramática reducción precedente, se decide restituir parte de los apoyos en recursos a esta actividad en forma de prefinanciamientos conseguidos a través de créditos externos, que debían posibilitar la atención integral a la agroindustria en determinadas provincias. En diferentes análisis sobre la actividad cañero-azucarera se señala que aun cuando el propósito fue garantizar una cobertura integral a la producción, esto no se logró.<sup>16</sup>

La información contenida en la tabla 23.2 pone de manifiesto la asincronía en la provisión de recursos: entre 1994-1997 mejora la asignación de fertilizantes, pero la inversión disminuye; sin embargo, en el período posterior (1997-2001) se incrementa la inversión y disminuyen los fertilizantes. Como los recursos no alcanzan a cubrir todas las necesidades, unas veces se apoya ciertos insumos y en otras ocasiones la inversión, y éste es justamente el problema. Si no se tienen recursos para atender todos los requerimientos lo más conveniente es disminuir la escala y garantizar las proporciones correctas, pues hacer otra cosa significa una pérdida inútil de recursos: se gasta y los resultados son malos en términos de eficiencia, lo que perjudica la competitividad de la fase agrícola y de toda la cadena agroexportadora. En el período posterior a 1994 se insistió en continuar plantando grandes extensiones de tierra, lo que indica que los enfoques de crecimiento extensivo se mantuvieron en este sector.<sup>17</sup> Revertir esta situación es precisamente lo que intenta el proceso de reestructuración que acomete la agroindustria en la actualidad.

## 23.4 Conclusiones y Recomendaciones

- La función de producción promedio estimada para la caña permite corroborar la importancia de atender a la necesaria continuidad del ciclo productivo: el área que se lleva a zafra cada año es el factor que más influye en la

<sup>15</sup> Ver Marquetti (2003, p. 6).

<sup>16</sup> Ibid, p. 31-32 y Nova (2001, p. 65).

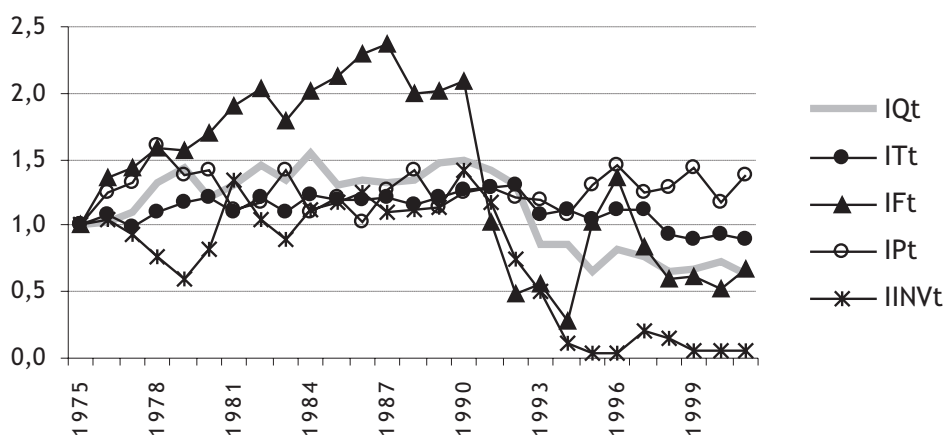
<sup>17</sup> Ver Marquetti (2003, p. 31).

- producción de caña, sin embargo, si se exceden los límites que garantizan la zafra futura ello pone riesgo los resultados por venir.
- Los resultados obtenidos sobre la contribución de la efectividad conjunta de factores a los resultados de la producción cañera arrojan que el único tramo con contribución positiva corresponde a los años 1980-1985. Entre 1989 y 1994 la producción cañera decreció más que lo que ayudarían a explicar la adversidad del clima y la contracción de los factores. Las difíciles condiciones que se crearon y la desorganización que esto generó en las entidades productoras, responden por este comportamiento. La contribución de la ECF después de 1994 vuelve más desfavorable, si bien entre 1997-2001 tiende a mejorar.
- Se pone de manifiesto la asincronía en la provisión de recursos. Si no se dispone de recursos para atender todos los requerimientos, lo más conveniente es disminuir la escala y garantizar las proporciones correctas. Hacer otra cosa significa una pérdida inútil de recursos: se gasta y los resultados son malos en términos de eficiencia, lo que perjudica la competitividad de la fase agrícola y de toda la cadena agroexportadora.
- Resulta muy importante conseguir una más elevada capacidad de ajuste fino de la asignación de recursos a nivel microeconómico. No basta con otorgar determinada prioridad a una actividad a nivel de la planificación y asignar cierta cantidad de divisas. Las proporciones en que se verifica la asignación definitivamente resultan críticas para lograr eficiencia. Así, se impone la revisión del mecanismo de asignación actual.
- Resultaría importante poder desarrollar estudios de efectividad para otras ramas de la producción agrícola que hoy son básicas para la agroexportación, como la producción tabacalera y citrícola, y también para renglones con potencial como las frutas y los vegetales.

## Anexo Capítulo 23

### A.23.1 Estadísticas básicas para las series relativas a la función de producción para la caña de azúcar

Figura A23.1: Producción de caña, factores de producción, inversiones y precipitaciones (1975=1.0)



Fuente: Elaboración propia a partir de MINAZ y ONE (2002).

### Análisis de correlación

Tabla A1: Matriz de correlación entre variables de la función de producción de caña de azúcar

| Variables | $LQ_t$  | $LT_t$  | $LF_t$  | $LP_t$  | $LINV_t$ |
|-----------|---------|---------|---------|---------|----------|
| $LQ_t$    | 1.000   |         |         |         |          |
| $LT_t$    | 0.8025  | 1.000   |         |         |          |
| $LF_t$    | 0.7192  | 0.4900  | 1.000   |         |          |
| $LP_t$    | -0.1213 | -0.1624 | 0.08589 | 1.000   |          |
| $LINV_t$  | 0.8814  | 0.6457  | 0.63570 | -0.2469 | 1.000    |

Fuente: Elaboración propia

Se observa una correlación lineal de la producción con los fertilizantes aplicados superior a 70%, con la tierra por encima del 80% y con las inversiones rebasa el 88%. La relación lineal entre la producción de caña y las precipitaciones contemporáneas es baja. Asimismo, se detectan coeficientes de correlación relativamen-

te altos entre algunas variables explicativas del modelo, lo que puede conducir a multicolinealidad aproximada. Esta situación es lógica, si se tiene en cuenta que estos factores de producción pueden ser complementarios. Ocurre en el caso de la tierra y las inversiones, y no es de extrañar ya que las inversiones contienen, además de la expansión del parque de maquinaria, el fomento de plantaciones. También se da entre las inversiones y los fertilizantes, factores que tienen un componente importado de consideración y entonces pueden moverse al unísono, en dependencia de las coyunturas externas.

### Análisis de estacionariedad

No fue posible rechazar que se tratara de procesos integrados en el caso de las variables en niveles; sólo la variable precipitaciones podría ser estacionaria. Fue necesario diferenciar una vez para obtener rechazo de la existencia de raíz unitaria para el resto de las variables. De modo que, en principio, se puede asumir que  $LP_t$  es  $I(0)$ , y que las demás variables son  $I(1)$ .

Tabla A2: Resultados de las pruebas de raíz unitaria, estadísticos DF o ADF\*

| Variables                      | Modelo (a) | Modelo (b) | Modelo (c) |
|--------------------------------|------------|------------|------------|
| <b>En niveles</b>              |            |            |            |
| $LQ_t$                         | -0.66145   | -0.50981   | -2.0515    |
| $LT_t$                         | -0.28080   | -1.5689    | -2.0756    |
| $LF_t$                         | -0.37381   | -1.5900    | -2.6553    |
| $LP_t$                         | 0.19028    | -5.0591**  | -4.9568**  |
| $LINV_t$                       | -1.1264    | -0.55972   | -1.7777    |
| <b>En primeras diferencias</b> |            |            |            |
| $DLQ_t$                        | -5.3779**  | -5.3631**  | -5.8585**  |
| $DLT_t$                        | -7.0749**  | -6.9853**  | -4.7478**  |
| $DLF_t$                        | -5.1972**  | -5.1168**  | -5.0276**  |
| $DLP_t$                        | -7.7697**  | -7.6170**  | -7.3996**  |
| $DLINV_t$                      | -3.8332**  | -4.4922**  | -4.6325**  |

\* La hipótesis nula en esta prueba es que  $\rho = 1$ . Un asterisco significa que se rechaza  $H_0$  al 5%; dos asteriscos, rechazo al 1%. Los resultados sombreados corresponden a las pruebas ADF.

Fuente: Elaboración propia.

### A.23.2 Problemas de especificación y resumen de las pruebas de diagnóstico para la función de producción de caña de azúcar

Se realizó una variedad de pruebas para verificar la validez de los supuestos implícitos en el modelo MRLS. Estos tests sugieren varios problemas. Por una parte, una posible omisión de rezagos de las variables  $LT_t$ ,  $LF_t$  y  $LINV_t$ . Se rechazan los supuestos relativos a la linealidad y la homoscedasticidad. Es probable que se presenten problemas de multicolinealidad aproximada entre variables explicativas. Asimismo se sugiere que habría un cambio estructural en los regresores en la primera mitad de los 90.

Para tomar en consideración estas cuestiones se trabajó en los problemas de dinámica, incorporando los rezagos que estaban faltando en la especificación de la media condicionada de la producción de caña. También se intentó modelar el cambio estructural a la altura del año 1993, cuando comienzan a aplicarse medidas para la transformación de la economía cubana con mayor intensidad. Sin embargo, una vez atendidos los problemas de dinámica, el cambio estructural no se sostiene. Los resultados de la estimación y el resumen de los test de diagnóstico para el MRLD se presentan seguidamente.

EQ( 4) Modelling LQ by OLS

The present sample is: 1976 to 2001

| Variable | Coefficient | Std.Error | t-value | t-prob | PartR $\hat{y}$ |
|----------|-------------|-----------|---------|--------|-----------------|
| Constant | -5.0935     | 1.4197    | -3.588  | 0.0020 | 0.4039          |
| LT       | 1.0824      | 0.18437   | 5.871   | 0.0000 | 0.6446          |
| LF       | 0.091816    | 0.032449  | 2.830   | 0.0107 | 0.2965          |
| LT_1     | -0.47990    | 0.17378   | -2.762  | 0.0124 | 0.2864          |
| LP_1     | 0.44423     | 0.11290   | 3.935   | 0.0009 | 0.4490          |
| LINV_1   | 0.12944     | 0.014157  | 9.143   | 0.0000 | 0.8148          |
| LF_1     | 0.10964     | 0.040538  | 2.705   | 0.0141 | 0.2780          |

$R^2 = 0.973249$   $F(6, 19) = 115.21$  [0.0000]  $\sigma = 0.0580852$   $DW = 1.46$

RSS = 0.06410396152 for 7 variables and 26 observations

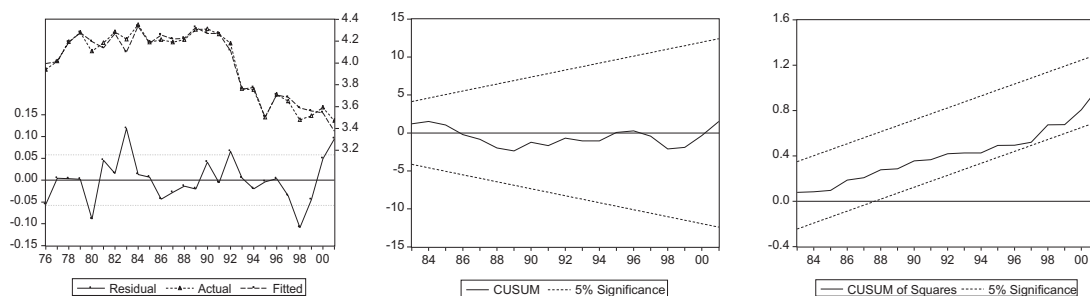
AR 1-  $2F(2, 17) = 0.43126$  [0.6566]

ARCH 1  $F(1, 17) = 0.25597$  [0.6194]

Normality  $\chi^2(2) = 2.4746$  [0.2902]

$\chi^2 F(12, 6) = 0.36985$  [0.9327]

RESET  $F(1, 18) = 0.6199$  [0.4413]



| Dependent Variable: LQ                                           |             |                       |             |           |
|------------------------------------------------------------------|-------------|-----------------------|-------------|-----------|
| Method: Least Squares                                            |             |                       |             |           |
| Date: 03/12/04 Time: 15:59                                       |             |                       |             |           |
| Sample(adjusted): 1976 2001                                      |             |                       |             |           |
| Included observations: 26 after adjusting endpoints              |             |                       |             |           |
| White Heteroskedasticity-Consistent Standard Errors & Covariance |             |                       |             |           |
| Variable                                                         | Coefficient | Std. Error            | t-Statistic | Prob.     |
| LT                                                               | 1.082355    | 0.248921              | 4.348194    | 0.0003    |
| LF                                                               | 0.091816    | 0.025819              | 3.556150    | 0.0021    |
| LT(-1)                                                           | -0.479896   | 0.179275              | -2.676877   | 0.0149    |
| LF(-1)                                                           | 0.109637    | 0.032157              | 3.409469    | 0.0029    |
| LP(-1)                                                           | 0.444228    | 0.093028              | 4.775201    | 0.0001    |
| LINV(-1)                                                         | 0.129440    | 0.009088              | 14.24227    | 0.0000    |
| C                                                                | -5.093504   | 1.556901              | -3.271566   | 0.0040    |
| R-squared                                                        | 0.973249    | Mean dependent var    |             | 3.995033  |
| Adjusted R-squared                                               | 0.964802    | S.D. dependent var    |             | 0.309602  |
| S.E. of regression                                               | 0.058085    | Akaike info criterion |             | -2.629007 |
| Sum squared resid                                                | 0.064104    | Schwarz criterion     |             | -2.290289 |
| Log likelihood                                                   | 41.17709    | F-statistic           |             | 115.2097  |
| Durbin-Watson stat                                               | 1.462217    | Prob(F-statistic)     |             | 0.000000  |

Este modelo pasó satisfactoriamente la mayoría de las pruebas de diagnóstico correspondientes a un MRLD. La posible colinealidad aproximada entre variables es un problema que persiste en esta especificación, lo que dificulta deslindar los

aportes de cada una de ellas. Ello, sin embargo no tiene por que ser relevante a los efectos de determinar la ECF, que equivale a la primera derivada del residuo de la función de producción. También se encuentra heteroscedasticidad simple y multiplicativa, por lo que se aplicó para la estimación el método de White para tener en cuenta esta circunstancia.

CAPÍTULO 24





## MODELO DE DATOS PANEL PARA TURISMO

A continuación se presentan algunos resultados de un modelo econométrico de datos panel que intenta explicar los flujos de turistas en el Caribe insular a partir de sus determinantes por el lado de la oferta. Por esta vía Vidal y Pérez (2008) obtienen una aproximación a la competitividad del turismo cubano en el área.

### 24.1 La teoría: El Enfoque de Oferta

Los modelos econométricos usados internacionalmente para explicar los flujos o ingresos turísticos se clasifican en dos tipos: los que utilizan un enfoque de demanda y los que se apoyan en los determinantes por el lado de la oferta.

Los enfoques de demanda son básicamente una función de demanda y son los más recurridos a la hora de efectuar predicciones. En ellos se especifica el arribo de turistas en función del PIB de los países emisores, algún índice de precios relativos, los costos de transportación y la población en el país de origen.

Los enfoques de oferta, por su parte, son en principio una función de producción extendida. Se especifican los ingresos turísticos en función de la cantidad de habitaciones disponibles (proxy del capital), la población en el país de destino (proxy del trabajo), el desarrollo socio económico del país de destino, el nivel de seguridad, el nivel de educación y salud, la infraestructura para las comunicaciones, la apertura de la economía, la inversión extranjera, entre otras características del país receptor del turismo.

Se critica de los enfoques de demanda que ignoran las particularidades del producto turístico en cada país. Los destinos turísticos no son homogéneos, no se deben ignorar las ventajas comparativas de cada país y el rol activo que estos

desempeñan en atraer mayor cantidad de turistas. En la práctica, se observa que algunos destinos emergen mientras otros se desaceleran. El desarrollo socio económico del país es determinante para su competitividad como destino turístico: infraestructura, nuevas tecnologías, clusters, seguridad, sanidad, nivel de educación, cultura, conservación del medio ambiente, etc. De ahí la mayor validez de los enfoques de oferta para entender la dinámica de los flujos turísticos internacionales.

## 24.2 Los Datos y el Modelo Empírico

Para las estimaciones se tomaron datos anuales en el período 1993-2006 de indicadores de 31 países del Caribe insular. En la tabla 24.1 se relacionan las variables. Los países que integran el panel son: Anguila, Antigua y Barbuda, Aruba, Bahamas, Barbados, Belice, Bermuda, Bonaire, Islas Vírgenes Británicas, México (Cancún y Cozumel), Islas Caimán, Cuba, Curazao, Dominica, República Dominicana, Granada, Guadalupe, Guyana, Haití, Jamaica, Martinica, Montserrat, Puerto Rico, St. Kitts and Nevis, Lucía, St. Martin, St. Vincent and Granadines, Suriname, Trinidad y Tobago, Turks and Caicos Islands, e Islas Vírgenes Estadounidenses. El panel de datos se encuentra bastante desbalanceado. No se tienen todas las observaciones de las variables para todos los países en los distintos años.

Tabla 24.1: Datos para el modelo

| Variables                         | Unidad de Medida                        |
|-----------------------------------|-----------------------------------------|
| Ingresos (ING)                    | Millones de USD                         |
| Habitaciones (HAB)                |                                         |
| Población (POB)                   |                                         |
| Apertura (APERT)                  | Porcentaje $(X+M)/PIB$                  |
| PIB per cápita (PIBP)             | Miles de USD                            |
| Inversión extranjera directa      | Millones de USD                         |
| Usuarios de Internet              | Por cada mil habitantes                 |
| Homicidios                        | Por cada cien mil habitantes            |
| Esperanza de vida                 | Años                                    |
| Prevalencia del VIH               | Porcentaje de la población (15-49 años) |
| Nivel de educación secundario     | Porcentaje de la población              |
| Nivel de educación universitario  | Porcentaje de la población              |
| Índice de Desarrollo Humano (ONU) |                                         |

Como modelo empírico se aplicaron las tres variantes más recurridas en las estimaciones de datos panel:

Efecto fijo:  $\log ING_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it}$

Efecto aleatorio:  $\log ING_{it} = X'_{it}\beta + \alpha + \mu_i + \varepsilon_{it}$

Panel dinámico:  $\log ING_{it} = \gamma \log ING_{it-1} + X'_{it}\beta + \alpha_i + \varepsilon_{it}$

donde  $X_{it}$  contiene las distintas variables explicativas del modelo.

Como no se tiene toda la información de las variables en  $X_{it}$ , gran parte de los factores que explican las diferencias entre los ingresos turísticos en el Caribe estará sintetizada en los efectos individuales. Los efectos individuales ( $\alpha_i, \mu_i$ ) recogerán la heterogeneidad inobservable entre los distintos destinos. En otras palabras, recogerán factores que explican las diferencias de ingresos entre los países una vez descontados los factores de escala y la disponibilidad de capital y trabajo. De esta forma, los efectos individuales son utilizados por Vidal y Pérez (2008) para aproximar las diferencias de competitividad entre los países que componen el Caribe insular.

### 24.3 Resultados: Aproximación a la Competitividad

Los resultados de la estimación con las variables que finalmente resultaron significativas se presentan en la tabla 24.2.

**Tabla 24.2: Estimación del modelo**  
Variable dependiente  $\log(ING)$   
(*t* valores en paréntesis, corregidos por heteroscedasticidad)

|                         | Efectos fijos     | Efectos aleatorios | Dinámico          |
|-------------------------|-------------------|--------------------|-------------------|
| $\log(ING(-1))$         |                   |                    | 0.073<br>(3.00)*  |
| $\log(HAB)$             | 0.912<br>(6.26)*  | 0.921<br>(8.00)*   | 0.564<br>(3.44)*  |
| $\log(POB)$             | 0.167<br>(0.343)  | 0.153<br>(2.35)*   | 1.599<br>(7.52)*  |
| $\log(PIBP)$            | 0.255<br>(2.227)* | 0.229<br>(2.21)*   | 0.039<br>(9.15)*  |
| APERT                   | 0.826<br>(7.924)* | 0.786<br>(4.66)*   | 0.952<br>(15.02)* |
| R <sup>2</sup> ajustado | 0.987             | 0.607              | --                |
| Error Estándar          | 0.190             | 0.191              | 0.231             |

Las variables cantidad de habitaciones, población, PIB per cápita, y la apertura de la economía resultaron significativas y con el signo acorde con lo esperado. Los modelos presentan un alto poder explicativo dado el valor del  $R^2$  y el reducido error estándar.

Una vez estimados los modelos se tomaron los efectos individuales para tener una aproximación a la competitividad. En la tabla 24.3 se presenta el resultado para los seis principales destinos turísticos del Caribe insular. Se ordenan según la competitividad que arroja el modelo y se construye un índice en porcentaje.

**Tabla 24.3: Competitividad en el Caribe (efectos individuales del modelo)**

| Estimación con efectos fijos |       |        | Estimación con efectos aleatorios |       |        |
|------------------------------|-------|--------|-----------------------------------|-------|--------|
|                              | Orden | Índice |                                   | Orden | Índice |
| República Dominicana         | 4     | 100.0  | República Dominicana              | 4     | 100.0  |
| Cuba                         | 5     | 97.1   | Puerto Rico                       | 5     | 97.5   |
| Bahamas                      | 6     | 93.8   | Cuba                              | 6     | 96.9   |
| Puerto Rico                  | 7     | 93.5   | Bahamas                           | 7     | 96.0   |
| Jamaica                      | 14    | 73.1   | Jamaica                           | 14    | 74.0   |
| México                       | 18    | 40.1   | México                            | 18    | 40.5   |

En ambas estimaciones República Dominicana aparece como el país con mayor competitividad entre los seis destinos del Caribe insular y es el cuarto en toda la región. Con el modelo de efectos fijos, Cuba es el segundo en competitividad entre los seis destinos fundamentales y el quinto dentro del total. Con el modelo de efectos aleatorios Cuba es el tercero después de Puerto Rico y el sexto dentro del total. El índice indica que Cuba tiene aproximadamente un 3% menos de competitividad que República Dominicana y está bien por encima de Jamaica y México.

En conclusión, la estimación con el modelo de datos panel apoya la validez del enfoque de oferta para explicar el turismo en el Caribe y conduce a rechazar la hipótesis de que el turismo cubano tiene una baja competitividad en el Caribe. A pesar de las debilidades que todavía presenta el sector turístico cubano, este tiene potencialidades destacadas que se descubren con un enfoque de oferta y quedan implícitas en los efectos individuales, entre ellas una de las más importantes podría ser la seguridad para el turista que visita la isla, así como las condiciones sanitarias y el nivel de educación de la población.



APÉNDICES



## Apéndice Metodológico

### 1. Propiedades de las Sumatorias

La **sumatoria** es un operador de uso frecuente en Estadística y Econometría. Comencemos por las definiciones básicas relacionadas con este operador. Consideremos  $n$  valores de una variable, posiblemente no todos iguales, que denotaremos por  $x_i$ , donde  $i$  es un índice que varía entre 1 y  $n$ . En realidad el rango de variación del índice  $i$  puede ser cualquier intervalo de números enteros que contenga  $n$  números. La siguiente fórmula define la **sumatoria** de esta variable:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n. \quad (\text{A.1})$$

Los límites de la sumatoria indican que el índice  $i$  varía entre 1 y  $n$ . El índice  $i=1$  es el límite inferior de la sumatoria y  $n$  es el límite superior. Utilizando directamente la definición presentada en (A.1) es fácil demostrar las siguientes propiedades de la sumatoria:

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad (\text{A.2})$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i, \quad (\text{A.3})$$

donde  $a$  es una constante. Las propiedades (A.2) y (A.3) pueden ser combinadas en la fórmula

$$\sum_{i=1}^n (ax_i + by_i) = a \sum_{i=1}^n x_i + b \sum_{i=1}^n y_i. \quad (\text{A.4})$$

La sumatoria, definida para una variable  $x_i$  puede ser generalizada para una función  $f$  de esta misma variable. Este resultado se presenta en la siguiente fórmula:

$$\sum_{i=1}^n f(x_i) = f(x_1) + f(x_2) + \cdots + f(x_n). \quad (\text{A.5})$$

La notación puede ser simplificada eliminando total o parcialmente los límites de la sumatoria, cuando éstos son fácilmente implicados por el contexto. La siguiente fórmula muestra algunas de las simplificaciones posibles:

$$\sum_x f(x_i) = \sum_i f(x_i) = \sum_{i=1}^n f(x_i).$$

También podemos tener sumatorias dobles, operando sobre dos diferentes variables. Esta posibilidad es presentada en la siguiente fórmula:

$$\sum_{i=1}^n \sum_{j=1}^m f(x_i, y_j) = \sum_{i=1}^n [f(x_i, y_1) + f(x_i, y_2) + \cdots + f(x_i, y_m)]. \quad (\text{A.6})$$

Esta definición puede ser generalizada a más variables. Una importante regla, la que puede ser fácilmente demostrada utilizando las definiciones (A.1) y (A.6), indica que en el caso de sumatorias múltiples, el orden en que estas operan puede ser alterado sin afectar el resultado. Es decir, tenemos que:

$$\sum_{i=1}^n \sum_{j=1}^m f(x_i, y_j) = \sum_{j=1}^m \sum_{i=1}^n f(x_i, y_j). \quad (\text{A.7})$$

Muchas veces necesitamos modificar los índices de la sumatoria. La siguiente fórmula, fácilmente demostrable usando la definición (A.1), nos indica como hacerlo. Supongamos que  $k$  es un número entero positivo. Entonces:

$$\sum_{i=1}^n f(x_i) = \sum_{i=1+k}^{n+k} f(x_{i-k}) = \sum_{i=1-k}^{n-k} f(x_{i+k}). \quad (\text{A.8})$$

Observemos que la constante  $k$  aparece con distinto signo en los límites de la sumatoria que en el interior de la misma.

## 2. Álgebra de Matrices

Las matrices son un instrumento y un vehículo apropiado para expresar muchas de nuestras ideas en Estadística y en Econometría. Este apéndice pretende desarrollar un mínimo de ideas sobre Álgebra de Matrices. Los lectores que sientan que necesitan conocimientos más profundos y detallados sobre el tema deberían recurrir a algún texto estándar de matemáticas.

### 2.1. Definiciones Básicas

Llamamos **matriz** a un conjunto rectangular y ordenado de números reales. Naturalmente, es posible utilizar otras definiciones, matemáticamente más rigurosas, pero que están fuera del alcance de este texto. Usualmente representamos las matrices por letras latinas mayúsculas o utilizando los símbolos siguientes:



$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

La matriz  $A$  en la fórmula anterior ha sido representada de dos maneras. En el primer caso la representamos por el **elemento genérico**  $a_{ij}$  que corresponde al elemento que se encuentra en la fila  $i$  y la columna  $j$  de la matriz. En el segundo caso, simplemente escribimos ordenadamente todos sus elementos. Esta segunda representación muestra, claramente, que se trata de un arreglo rectangular de números, organizados en filas y columnas. En el ejemplo anterior tenemos  $m$  filas y  $n$  columnas. Por eso muchas veces escribimos  $A_{m \times n}$ , haciendo explícito el número de filas y columnas. El número de filas y columnas es la dimensión de la matriz. Es convencional escribir primero el número de filas y luego el de columnas. En general, las matrices son rectangulares, pero a veces el número de filas es igual al número de columnas. En este caso diremos que las matrices son cuadradas y escribiremos simplemente  $A_n$ , cuando la matriz tiene  $n$  filas y  $n$  columnas.

Un elemento sumamente importante en una matriz es su **diagonal principal**. Ésta comienza en la esquina superior izquierda y avanza diagonalmente hasta el fin de la matriz. El número de elementos de la diagonal principal es igual al número de filas o columnas, dependiendo de cual sea menor.

$$\begin{bmatrix} 2 & -1 & 3 & 1 \\ -4 & 0 & 2 & 1 \\ 1 & 7 & 4 & 9 \end{bmatrix}, \begin{bmatrix} 3 & 1 & -5 \\ 0 & 1 & 5 \\ 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 3 & 0 & 0 \\ 1 & 1 & 0 \\ 5 & -5 & 2 \end{bmatrix}, \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

En el ejemplo anterior, la primera matriz es **rectangular** con 3 filas y 4 columnas. Hemos marcado la diagonal principal para esta matriz. La siguiente matriz es **cuadrada** y la llamaremos matriz **triangular superior**, porque sólo los elementos en la diagonal principal o sobre ella son diferentes de cero. El caso siguiente se trata de una matriz triangular inferior. Una matriz que es al mismo tiempo triangular superior y triangular inferior, se la llama **matriz diagonal** porque tiene elementos diferentes de cero sólo sobre la diagonal principal, como es el caso en la última matriz en el ejemplo anterior.

## 2.2. Operaciones con Matrices

Es posible definir varias operaciones importantes y útiles con matrices. En primer lugar, diremos que dos matrices son iguales cuando todos los elementos corres-

pondientes (en la misma posición) son iguales. Es decir, no es suficiente que ambas matrices tengan los mismos elementos, sino que además estos deben estar en el mismo orden. Por ejemplo, las matrices:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ y } \begin{bmatrix} b & d \\ c & a \end{bmatrix}.$$

son diferentes a pesar de que tienen los mismos elementos.

La primera operación que consideraremos es muy simple y se llama **transposición**. Transponemos una matriz cuando reemplazamos cada fila por la correspondiente columna. También podemos decir que transponemos una matriz cuando reemplazamos cada columna por la correspondiente fila. Consideremos las siguientes matrices:

$$A = \begin{bmatrix} 1 & -3 \\ 5 & 4 \\ 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 5 & 3 \\ -3 & 4 & 2 \end{bmatrix}.$$

Entonces podemos decir que la matriz  $B$  es la **transpuesta** de la matriz  $A$  y escribiremos  $B = A'$ . Algunos usan un  $T$  como exponente, en lugar del apóstrofe, para indicar transposición. Podemos observar que la transpuesta de la transpuesta es la matriz original. Es decir, en el ejemplo anterior  $A = B'$ . También señalaremos que cuando una matriz que es igual a su transpuesta, diremos que esta matriz es **simétrica**. Es decir la matriz  $A$  es simétrica si y solo si  $A = A'$ . Otra operación, que asocia un número real a cada matriz, es la **traza**. Para una matriz  $A$  la traza es igual a la suma de los elementos de su diagonal principal. Por ejemplo para una matriz  $A = [a_{ij}]$  de dimensión  $m \times n$  la traza es

$$\text{tr}(A) = \sum_{i=1}^m a_{ii},$$

suponiendo que  $m \leq n$ . Para las matrices  $A$  y  $B$ , presentada anteriormente, tenemos que

$$\text{tr}(A) = \text{tr}(B) = 1 + 4 = 5.$$

También podemos definir una **suma de matrices**, gracias a la siguiente fórmula. Si tenemos dos matrices  $A$  y  $B$  con elementos genéricos  $a_{ij}$  y  $b_{ij}$  respectivamente, entonces podemos escribir:

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}].$$

Consideremos el siguiente ejemplo:

$$\begin{bmatrix} 3 & -2 \\ 1 & 5 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} -1 & 3 \\ 2 & -2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 2 & 5 \end{bmatrix},$$

resultado que obtenemos sumando los elementos correspondientes en cada matriz. Nótese que sólo podemos sumar matrices que tienen el mismo número de filas y el mismo número de columnas, para obtener una matriz con el mismo número de filas y el mismo número de columnas. Tales matrices, que pueden ser sumadas, se llaman **conformables**.

Otra operación útil es el **producto escalar**. En este caso intervienen una matriz y un escalar y el resultado será una matriz con las mismas dimensiones que la matriz original. En el contexto de este apéndice y las aplicaciones utilizadas en este libro la palabra escalar significa simplemente un número real. La fórmula de producto escalar es:

$$\alpha A = \alpha [a_{ij}] = [\alpha a_{ij}],$$

donde  $\alpha$  es un escalar y  $A$  es una matriz. Es decir, en el producto escalar debemos multiplicar cada elemento de la matriz por el escalar. Estas operaciones siguen ciertas reglas que describimos a continuación.

En primer lugar, la suma de matrices es conmutativa y asociativa. Formalmente:

$$\begin{aligned} A + B &= B + A && \text{(conmutatividad),} \\ A + (B + C) &= (A + B) + C && \text{(asociatividad).} \end{aligned}$$

Es decir, podemos sumar las matrices en cualquier orden. También podemos agruparlas de todas las maneras posibles para ser sumadas sin alterar el resultado.

Para cada dimensión posible, existe una matriz  $N$ , llamada matriz nula o matriz cero, con la siguiente propiedad:

$$A + N = N + A = A,$$

para cualquier matriz  $A$ . Es fácil ver que todos los elementos de la matriz  $N$  deben ser ceros. Toda matriz  $A$  tiene asociada una matriz  $-A$  tal que  $A + (-A) = N$ . Es fácil ver que tenemos  $-A = (-1)A$  y escribimos simplemente  $A + (-A) = A - A = N$ .

Existe una doble distributividad para la multiplicación escalar. La siguiente fórmula, donde  $\alpha$  y  $\beta$  son escalares y  $A$  y  $B$  son matrices, presenta esta propiedad:

$$(\alpha + \beta)A = \alpha A + \beta A,$$

$$\alpha(A + B) = \alpha A + \alpha B.$$

Nótese que la transpuesta de una suma es la suma de las transpuestas. Es decir,

$$(A + B)' = A' + B'.$$

La **multiplicación de matrices** es una operación más compleja que se realiza de acuerdo a la siguiente fórmula:

$$A_{m \times n} B_{n \times s} = C_{m \times s} = \left[ c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \right].$$

Esta fórmula indica que debemos multiplicar una fila de la primera matriz por una columna de la segunda matriz. Esta multiplicación se realiza sumando los productos de los elementos correspondientes. Esta multiplicación da como resultado un elemento en la misma fila y columna que lo originan. El siguiente ejemplo contribuye a aclarar esta regla:

$$\begin{bmatrix} 1 & 4 & 2 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 12 \\ 5 & 9 \end{bmatrix}.$$

En este caso 12, el elemento en la primera fila y segunda columna de la matriz resultado es el producto de la primera fila de la primera matriz por la segunda columna de la segunda matriz. Es fácil ver que  $12 = 1 \times 2 + 4 \times 1 + 2 \times 3$ .

Debemos ser muy cuidadosos cuando multiplicamos matrices. En primer lugar estas deben ser conformables. Esto significa que la multiplicación sólo es posible cuando el número de columnas de la primera matriz es igual al número de filas de la segunda matriz. En segundo lugar la multiplicación de matrices no es conmutativa y el orden en que se realiza la multiplicación resulta decisivo. Esto es, en general, tenemos que  $AB \neq BA$ . En el ejemplo anterior resulta claro que no podemos cambiar el orden de la multiplicación, porque entonces las matrices ya no son conformables.

En cuanto a la relación entre multiplicación y transposición podemos anotar las dos siguientes propiedades.

$$(\alpha A)' = \alpha A',$$

$$(AB)' = B' A'.$$

Podemos observar que en la segunda de estas propiedades, transpuesta del producto de dos matrices, el orden de multiplicación se invierte. Es importante seguir este orden cuidadosamente ya que la multiplicación de matrices no es conmutativa.

### 2.3. Determinantes e Inversas

Cada matriz cuadrada tiene asociado un número que denominamos como su **determinante**. El cálculo del determinante de una matriz cuadrada es una operación que rápidamente resulta pesada y engorrosa desde el punto de vista aritmético. En efecto, el número de operaciones aritméticas básicas necesarias para calcular el determinante aumenta muy rápidamente cuando aumenta el tamaño de la matriz. Muy rápidamente resulta impráctico el cálculo de los determinantes para matrices de tamaño relativamente modesto. Afortunadamente existe una amplia disposición de recursos computacionales para realizar cálculos con matrices. De hecho, el programa Excel responde a un diseño basado en matrices. De esta manera resulta muy sencillo realizar las operaciones básicas con matrices disponiendo estas en una hoja Excel. La función Excel MMULT() permite multiplicar matrices. La función DET() permite calcular el determinante de una matriz.

Nosotros nos limitaremos sólo a dos casos de cálculo de determinantes. En primer lugar, consideremos una matriz de orden 1; es decir, con una sola columna y una sola fila. En este caso el determinante es simplemente igual al único elemento de la matriz. Por ejemplo, para la matriz  $A = [a]$ , tenemos que:

$$|A| = |a| = a.$$

Nótese la notación que utilizamos para los determinantes, encerrando la matriz entre dos líneas verticales. En segundo lugar, consideremos una matriz de orden 2; es decir, con dos filas y dos columnas. En este caso tenemos:

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Es decir, el determinante es igual al producto de los dos elementos en la diagonal principal, menos el producto de los dos elementos restantes.

Para cada dimensión tenemos una matriz cuadrada y diagonal, con todos los elementos de la diagonal principal iguales a uno, que denominamos como **matriz identidad**. Esta matriz identidad tiene la siguiente propiedad. Para una matriz con  $m$  filas y  $n$  columnas tenemos que:

$$A_{m \times n} I_n = I_m A_{m \times n} = A_{m \times n}.$$

Por ejemplo, la matriz identidad para la dimensión 3 es igual a:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ciertas matrices cuadradas tiene asociada otra matriz, que llamamos su matriz inversa, y que representamos como  $A^{-1}$ . Esta matriz inversa tiene la siguiente propiedad:

$$AA^{-1} = A^{-1}A = I.$$

Nótese la analogía con el recíproco de un número para la matriz inversa y con el número 1 para la matriz identidad. Las matrices cuadradas que admiten una inversa se llaman **matrices no-singulares**. Debemos insistir que no todas las matrices cuadradas tienen una inversa. Si una matriz cuadrada  $A$  es no-singular, entonces tenemos que  $|A| \neq 0$ . Recíprocamente, si tenemos que  $|A| \neq 0$ , entonces  $A$  no es no-singular y su inversa  $A^{-1}$  existe. Otro enfoque al problema de la existencia de la inversa de una matriz cuadrada es el **rango de la matriz**. El símbolo  $\text{rank}(A)$  denota el rango de la matriz  $A$ . Este es un número entero que indica el número de filas o columnas **linealmente independientes**<sup>1</sup>. Si el rango de una matriz cuadrada es igual al número de sus filas o columnas, diremos que esta matriz es de rango completo y que admite una inversa. Consideremos una matriz cuadrada  $A$  con  $n$  filas y  $n$  columnas. Entonces, si  $\text{rank}(A)=n$ , diremos que  $A$  es de rango completo, que  $|A| \neq 0$  y que su inversa existe.

El cálculo de la inversa es aún más demandante que el cálculo del determinante, desde el punto de vista aritmético. Consideraremos sólo dos situaciones. En primer lugar, para una matriz con sólo una fila y una columna, la inversa tiene un solo elemento igual al recíproco del único elemento de la matriz original. En segundo lugar, para matrices con dos filas y columnas tenemos que:

---

<sup>1</sup> Para entender el significado de dependencia o independencia lineal, por favor consultar un texto de Álgebra Lineal.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Nótese que el denominador de la fracción que encabeza la fórmula de la inversa es, simplemente, igual al determinante de la matriz  $A$ . La matriz que aparece en la fórmula puede ser fácilmente calculada de acuerdo a la siguiente regla: los elementos de la diagonal principal intercambian sus posiciones, los elementos restantes cambian su signo. Para matrices de orden mayor que dos podemos usar la función MINV() de Excel.

## 2.4 Producto de Kronecker

El producto de Kronecker es una operación matricial que ocurre con frecuencia en Econometría de Datos de Panel. Esta operación puede ser definida de la siguiente manera: Sea  $A$  una matriz  $m \times n$  y  $B$  una matriz  $p \times q$ . Entonces el **producto Kronecker** de  $A$  y  $B$  es la matriz  $mp \times nq$

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}, \quad A_{m \times n} = [a_{ij}] \text{ y } B_{p \times q} = [b_{ij}].$$

Esta definición puede ser ilustrada con el siguiente ejemplo: Consideremos las siguientes matrices:

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix}.$$

Entonces,

$$A \otimes B = \begin{bmatrix} 2B & 3B \\ 3B & B \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 & 3 & 6 & 9 \\ 6 & 4 & 2 & 9 & 6 & 3 \\ 2 & 8 & 4 & 3 & 12 & 6 \\ 3 & 6 & 9 & 1 & 2 & 3 \\ 9 & 6 & 3 & 3 & 2 & 1 \\ 3 & 12 & 6 & 1 & 4 & 2 \end{bmatrix}.$$

La siguiente es una lista de las principales propiedades del producto de Kronecker. Estas propiedades son presentadas sin las demostraciones correspondientes, las que pueden derivarse directamente de la definición anterior.

- a) Sea  $A$  una matriz  $m \times n$ ,  $B$  una matriz  $r \times s$ ,  $C$  una matriz  $n \times p$  y  $D$  una matriz  $s \times t$ . Entonces

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

Esta es una matriz  $mr \times pt$ .

- b) Para todo  $A$  y  $B$ , tenemos  $(A \otimes B)' = A' \otimes B'$ .
- c) Si  $A$  y  $B$  son simétricas, entonces  $A \otimes B$  es también simétrica.
- d) Si  $A$  y  $B$  son no-singulares, entonces  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ .
- e) Para todo  $A$  y  $B$ , tenemos  $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B) = \text{tr}(B \otimes A)$ .
- f) Sea  $A$  una matriz  $m \times m$  y  $B$  una matriz  $n \times n$ . Entonces

$$\det(A \otimes B) = (\det A)^m (\det B)^n = \det(B \otimes A).$$

## 2.5. Algunas matrices importantes para Econometría de Datos de Panel

A continuación presentamos algunos resultados que son críticos para la discusión de la Econometría de Datos de Panel.

Sea  $i_T$  un vector de unos de dimensión  $T \times 1$ . Es decir, un vector de  $T$  elementos, en que cada posición está ocupada por un uno. Consideremos la matriz  $Z_\mu = I_N \otimes i_T$ . Entonces,

$$Z_\mu Z_\mu' = (I_N \otimes i_T)(I_N' \otimes i_T') = I_N \otimes i_T i_T' = I_N \otimes \varnothing_T,$$

Donde  $\varnothing_T$  es una matriz cuadrada de unos de dimensión  $T \times T$ . Es decir, cada una de las posiciones de la matriz está ocupada por un uno. Ahora consideremos



$$\begin{aligned}
P &= Z_{\mu} (Z'_{\mu} Z_{\mu})^{-1} Z'_{\mu}, \\
&= Z_{\mu} \left[ (I'_N \otimes I'_T) (I_N \otimes I_T) \right]^{-1} Z'_{\mu}, \\
&= Z_{\mu} (I_N \otimes I'_T I_T)^{-1} Z'_{\mu}, \\
&= Z_{\mu} \left( \frac{1}{T} I_N \right) Z'_{\mu}, \\
&= \frac{1}{T} Z_{\mu} Z'_{\mu}, \\
P &= I_N \otimes \bar{\varnothing}_T.
\end{aligned}$$

Donde  $\bar{\varnothing}_T = \varnothing_T / T$ . Usamos este resultado para definir

$$Q = I_{NT} - P.$$

Algunas veces decimos que la matriz  $P$  es la matriz de **proyección** de  $Z_{\mu}$ . La matriz  $Q$  es una matriz de **desviaciones** porque si consideramos un vector  $y = [y_{it}]$ , entonces tendremos que

$$Qy = [y_{it} - \bar{y}_{i.}], \quad \text{donde} \quad \bar{y}_{i.} = \frac{1}{T} \sum_{t=1}^T y_{it}.$$

Es decir, la matriz  $Q$  nos permite calcular desviaciones en torno a las medias parciales, calculadas a lo largo de la variable  $t$ . El objetivo de esta discusión ha sido el obtener la matriz  $P$  y, sobre todo, la matriz  $Q$ . En consecuencia resulta importante anotar algunas propiedades de las matrices  $P$  y  $Q$ , aunque omitimos las demostraciones correspondientes.

- a) Tanto  $P$  como  $Q$  son simétricas e idempotentes. Esto es  $P = P'$ ,  $P = P^2$ ,  $Q = Q'$ , y  $Q = Q^2$ .
- b) Usando la propiedad de que una matriz idempotente tiene rango igual a su traza, tenemos que

c)

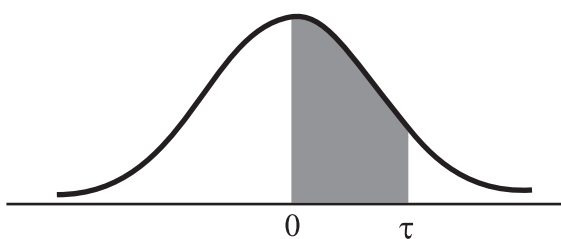
$$\text{rank}(P) = \text{tr}(P) = N,$$

$$\text{rank}(Q) = \text{tr}(Q) = N(T-1).$$

d)  $P$  y  $Q$  son ortogonales. Es decir,  $PQ = 0$ .

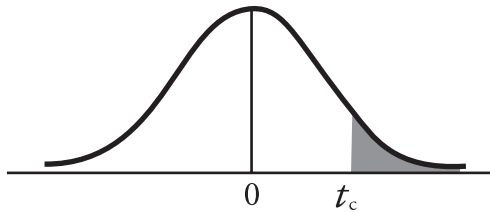
e)  $P + Q = I_{NT}$ .

## B. Tablas Estadísticas.



Área bajo la Distribución Normal Estándar.

| z   | 0,00   | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0,0 | 0,0000 | 0,0040 | 0,0080 | 0,0120 | 0,0160 | 0,0199 | 0,0239 | 0,0279 | 0,0319 | 0,0359 |
| 0,1 | 0,0398 | 0,0438 | 0,0478 | 0,0517 | 0,0557 | 0,0596 | 0,0636 | 0,0675 | 0,0714 | 0,0753 |
| 0,2 | 0,0793 | 0,0832 | 0,0871 | 0,0910 | 0,0948 | 0,0987 | 0,1026 | 0,1064 | 0,1103 | 0,1141 |
| 0,3 | 0,1179 | 0,1217 | 0,1255 | 0,1293 | 0,1331 | 0,1368 | 0,1406 | 0,1443 | 0,1480 | 0,1517 |
| 0,4 | 0,1554 | 0,1591 | 0,1628 | 0,1664 | 0,1700 | 0,1736 | 0,1772 | 0,1808 | 0,1844 | 0,1879 |
| 0,5 | 0,1915 | 0,1950 | 0,1985 | 0,2019 | 0,2054 | 0,2088 | 0,2123 | 0,2157 | 0,2190 | 0,2224 |
| 0,6 | 0,2257 | 0,2291 | 0,2324 | 0,2357 | 0,2389 | 0,2422 | 0,2454 | 0,2486 | 0,2517 | 0,2549 |
| 0,7 | 0,2580 | 0,2611 | 0,2642 | 0,2673 | 0,2704 | 0,2734 | 0,2764 | 0,2794 | 0,2823 | 0,2852 |
| 0,8 | 0,2881 | 0,2910 | 0,2939 | 0,2967 | 0,2995 | 0,3023 | 0,3051 | 0,3078 | 0,3106 | 0,3133 |
| 0,9 | 0,3159 | 0,3186 | 0,3212 | 0,3238 | 0,3264 | 0,3289 | 0,3315 | 0,3340 | 0,3365 | 0,3389 |
| 1,0 | 0,3413 | 0,3438 | 0,3461 | 0,3485 | 0,3508 | 0,3531 | 0,3554 | 0,3577 | 0,3599 | 0,3621 |
| 1,1 | 0,3643 | 0,3665 | 0,3686 | 0,3708 | 0,3729 | 0,3749 | 0,3770 | 0,3790 | 0,3810 | 0,3830 |
| 1,2 | 0,3849 | 0,3869 | 0,3888 | 0,3907 | 0,3925 | 0,3944 | 0,3962 | 0,3980 | 0,3997 | 0,4015 |
| 1,3 | 0,4032 | 0,4049 | 0,4066 | 0,4082 | 0,4099 | 0,4115 | 0,4131 | 0,4147 | 0,4162 | 0,4177 |
| 1,4 | 0,4192 | 0,4207 | 0,4222 | 0,4236 | 0,4251 | 0,4265 | 0,4279 | 0,4292 | 0,4306 | 0,4319 |
| 1,5 | 0,4332 | 0,4345 | 0,4357 | 0,4370 | 0,4382 | 0,4394 | 0,4406 | 0,4418 | 0,4429 | 0,4441 |
| 1,6 | 0,4452 | 0,4463 | 0,4474 | 0,4484 | 0,4495 | 0,4505 | 0,4515 | 0,4525 | 0,4535 | 0,4545 |
| 1,7 | 0,4554 | 0,4564 | 0,4573 | 0,4582 | 0,4591 | 0,4599 | 0,4608 | 0,4616 | 0,4625 | 0,4633 |
| 1,8 | 0,4641 | 0,4649 | 0,4656 | 0,4664 | 0,4671 | 0,4678 | 0,4686 | 0,4693 | 0,4699 | 0,4706 |
| 1,9 | 0,4713 | 0,4719 | 0,4726 | 0,4732 | 0,4738 | 0,4744 | 0,4750 | 0,4756 | 0,4761 | 0,4767 |
| 2,0 | 0,4772 | 0,4778 | 0,4783 | 0,4788 | 0,4793 | 0,4798 | 0,4803 | 0,4808 | 0,4812 | 0,4817 |
| 2,1 | 0,4821 | 0,4826 | 0,4830 | 0,4834 | 0,4838 | 0,4842 | 0,4846 | 0,4850 | 0,4854 | 0,4857 |
| 2,2 | 0,4861 | 0,4864 | 0,4868 | 0,4871 | 0,4875 | 0,4878 | 0,4881 | 0,4884 | 0,4887 | 0,4890 |
| 2,3 | 0,4893 | 0,4896 | 0,4898 | 0,4901 | 0,4904 | 0,4906 | 0,4909 | 0,4911 | 0,4913 | 0,4916 |
| 2,4 | 0,4918 | 0,4920 | 0,4922 | 0,4925 | 0,4927 | 0,4929 | 0,4931 | 0,4932 | 0,4934 | 0,4936 |
| 2,5 | 0,4938 | 0,4940 | 0,4941 | 0,4943 | 0,4945 | 0,4946 | 0,4948 | 0,4949 | 0,4951 | 0,4952 |
| 2,6 | 0,4953 | 0,4955 | 0,4956 | 0,4957 | 0,4959 | 0,4960 | 0,4961 | 0,4962 | 0,4963 | 0,4964 |
| 2,7 | 0,4965 | 0,4966 | 0,4967 | 0,4968 | 0,4969 | 0,4970 | 0,4971 | 0,4972 | 0,4973 | 0,4974 |
| 2,8 | 0,4974 | 0,4975 | 0,4976 | 0,4977 | 0,4977 | 0,4978 | 0,4979 | 0,4979 | 0,4980 | 0,4981 |
| 2,9 | 0,4981 | 0,4982 | 0,4982 | 0,4983 | 0,4984 | 0,4984 | 0,4985 | 0,4985 | 0,4986 | 0,4986 |
| 3,0 | 0,4987 | 0,4987 | 0,4987 | 0,4988 | 0,4988 | 0,4989 | 0,4989 | 0,4989 | 0,4990 | 0,4990 |

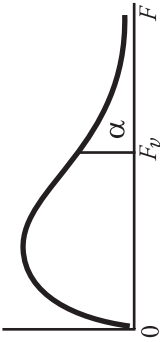


Valores Críticos para la distribución *t* de Student.

| DF | $\alpha = 0,10$ | $\alpha = 0,05$ | $\alpha = 0,025$ | $\alpha = 0,005$ |
|----|-----------------|-----------------|------------------|------------------|
| 1  | 3,078           | 6,314           | 12,706           | 63,656           |
| 2  | 1,886           | 2,920           | 4,303            | 9,925            |
| 3  | 1,638           | 2,353           | 3,182            | 5,841            |
| 4  | 1,533           | 2,132           | 2,776            | 4,604            |
| 5  | 1,476           | 2,015           | 2,571            | 4,032            |
| 6  | 1,440           | 1,943           | 2,447            | 3,707            |
| 7  | 1,415           | 1,895           | 2,365            | 3,499            |
| 8  | 1,397           | 1,860           | 2,306            | 3,355            |
| 9  | 1,383           | 1,833           | 2,262            | 3,250            |
| 10 | 1,372           | 1,812           | 2,228            | 3,169            |
| 11 | 1,363           | 1,796           | 2,201            | 3,106            |
| 12 | 1,356           | 1,782           | 2,179            | 3,055            |
| 13 | 1,350           | 1,771           | 2,160            | 3,012            |
| 14 | 1,345           | 1,761           | 2,145            | 2,977            |
| 15 | 1,341           | 1,753           | 2,131            | 2,947            |
| 16 | 1,337           | 1,746           | 2,120            | 2,921            |
| 17 | 1,333           | 1,740           | 2,110            | 2,898            |
| 18 | 1,330           | 1,734           | 2,101            | 2,878            |
| 19 | 1,328           | 1,729           | 2,093            | 2,861            |
| 20 | 1,325           | 1,725           | 2,086            | 2,845            |
| 21 | 1,323           | 1,721           | 2,080            | 2,831            |
| 22 | 1,321           | 1,717           | 2,074            | 2,819            |
| 23 | 1,319           | 1,714           | 2,069            | 2,807            |
| 24 | 1,318           | 1,711           | 2,064            | 2,797            |
| 25 | 1,316           | 1,708           | 2,060            | 2,787            |
| 26 | 1,315           | 1,706           | 2,056            | 2,779            |

Continúa en la página siguiente.

| DF       | $\alpha = 0,10$ | $\alpha = 0,05$ | $\alpha = 0,025$ | $\alpha = 0,005$ |
|----------|-----------------|-----------------|------------------|------------------|
| 27       | 1,314           | 1,703           | 2,052            | 2,771            |
| 28       | 1,313           | 1,701           | 2,048            | 2,763            |
| 29       | 1,311           | 1,699           | 2,045            | 2,756            |
| 30       | 1,310           | 1,697           | 2,042            | 2,750            |
| 40       | 1,303           | 1,684           | 2,021            | 2,704            |
| 50       | 1,299           | 1,676           | 2,009            | 2,678            |
| 60       | 1,296           | 1,671           | 2,000            | 2,660            |
| 70       | 1,294           | 1,667           | 1,994            | 2,648            |
| 80       | 1,292           | 1,664           | 1,990            | 2,639            |
| 90       | 1,291           | 1,662           | 1,987            | 2,632            |
| 100      | 1,290           | 1,660           | 1,984            | 2,626            |
| 110      | 1,289           | 1,659           | 1,982            | 2,621            |
| 120      | 1,289           | 1,658           | 1,980            | 2,617            |
| $\infty$ | 1,282           | 1,645           | 1,960            | 2,576            |

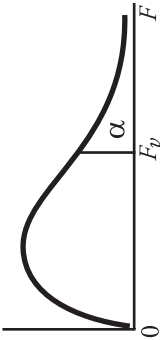


Valores Críticos de Cola Derecha para la Distribución F de Fisher, 5% superior.

| $v_2 \backslash v_1$ | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     | 12     | 15     | 20     | 24     | 30     | 40     | 60     | 120    | ∞      |
|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1                    | 161,45 | 199,50 | 215,71 | 224,58 | 230,16 | 233,99 | 236,77 | 238,88 | 240,54 | 241,88 | 243,90 | 245,95 | 248,02 | 249,05 | 250,10 | 251,14 | 252,20 | 253,25 | 254,32 |
| 2                    | 18,51  | 19,00  | 19,16  | 19,25  | 19,30  | 19,33  | 19,35  | 19,37  | 19,38  | 19,40  | 19,41  | 19,43  | 19,45  | 19,45  | 19,46  | 19,47  | 19,48  | 19,49  | 19,50  |
| 3                    | 10,13  | 9,55   | 9,28   | 9,12   | 9,01   | 8,94   | 8,89   | 8,85   | 8,81   | 8,79   | 8,74   | 8,70   | 8,66   | 8,64   | 8,62   | 8,59   | 8,57   | 8,55   | 8,53   |
| 4                    | 7,71   | 6,94   | 6,59   | 6,39   | 6,26   | 6,16   | 6,09   | 6,04   | 6,00   | 5,96   | 5,91   | 5,86   | 5,80   | 5,77   | 5,75   | 5,72   | 5,69   | 5,66   | 5,63   |
| 5                    | 6,61   | 5,79   | 5,41   | 5,19   | 5,05   | 4,95   | 4,88   | 4,82   | 4,77   | 4,74   | 4,68   | 4,62   | 4,56   | 4,53   | 4,50   | 4,46   | 4,43   | 4,40   | 4,37   |
| 6                    | 5,99   | 5,14   | 4,76   | 4,53   | 4,39   | 4,28   | 4,21   | 4,15   | 4,10   | 4,06   | 4,00   | 3,94   | 3,87   | 3,84   | 3,81   | 3,77   | 3,74   | 3,70   | 3,67   |
| 7                    | 5,59   | 4,74   | 4,35   | 4,12   | 3,97   | 3,87   | 3,79   | 3,73   | 3,68   | 3,64   | 3,57   | 3,51   | 3,44   | 3,41   | 3,38   | 3,34   | 3,30   | 3,27   | 3,23   |
| 8                    | 5,32   | 4,46   | 4,07   | 3,84   | 3,69   | 3,58   | 3,50   | 3,44   | 3,39   | 3,35   | 3,28   | 3,22   | 3,15   | 3,12   | 3,08   | 3,04   | 3,01   | 2,97   | 2,93   |
| 9                    | 5,12   | 4,26   | 3,86   | 3,63   | 3,48   | 3,37   | 3,29   | 3,23   | 3,18   | 3,14   | 3,07   | 3,01   | 2,94   | 2,90   | 2,86   | 2,83   | 2,79   | 2,75   | 2,71   |
| 10                   | 4,96   | 4,10   | 3,71   | 3,48   | 3,33   | 3,22   | 3,14   | 3,07   | 3,02   | 2,98   | 2,91   | 2,85   | 2,77   | 2,74   | 2,70   | 2,66   | 2,62   | 2,58   | 2,54   |
| 11                   | 4,84   | 3,98   | 3,59   | 3,36   | 3,20   | 3,09   | 3,01   | 2,95   | 2,90   | 2,85   | 2,79   | 2,72   | 2,65   | 2,61   | 2,57   | 2,53   | 2,49   | 2,45   | 2,40   |
| 12                   | 4,75   | 3,89   | 3,49   | 3,26   | 3,11   | 3,00   | 2,91   | 2,85   | 2,80   | 2,75   | 2,69   | 2,62   | 2,54   | 2,51   | 2,47   | 2,43   | 2,38   | 2,34   | 2,30   |
| 13                   | 4,67   | 3,81   | 3,41   | 3,18   | 3,03   | 2,92   | 2,83   | 2,77   | 2,71   | 2,67   | 2,60   | 2,53   | 2,46   | 2,42   | 2,38   | 2,34   | 2,30   | 2,25   | 2,21   |
| 14                   | 4,60   | 3,74   | 3,34   | 3,11   | 2,96   | 2,85   | 2,76   | 2,70   | 2,65   | 2,60   | 2,53   | 2,46   | 2,39   | 2,35   | 2,31   | 2,27   | 2,22   | 2,18   | 2,13   |

Continúa en la página siguiente.

| $v_2 \backslash v_1$ | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 12   | 15   | 20   | 24   | 30   | 40   | 60   | 120  | `    |
|----------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 15                   | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 | 2,48 | 2,40 | 2,33 | 2,29 | 2,25 | 2,20 | 2,16 | 2,11 | 2,07 |
| 16                   | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,42 | 2,35 | 2,28 | 2,24 | 2,19 | 2,15 | 2,11 | 2,06 | 2,01 |
| 17                   | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 | 2,38 | 2,31 | 2,23 | 2,19 | 2,15 | 2,10 | 2,06 | 2,01 | 1,96 |
| 18                   | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,34 | 2,27 | 2,19 | 2,15 | 2,11 | 2,06 | 2,02 | 1,97 | 1,92 |
| 19                   | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 | 2,31 | 2,23 | 2,16 | 2,11 | 2,07 | 2,03 | 1,98 | 1,93 | 1,88 |
| 20                   | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 | 2,28 | 2,20 | 2,12 | 2,08 | 2,04 | 1,99 | 1,95 | 1,90 | 1,84 |
| 21                   | 4,32 | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,49 | 2,42 | 2,37 | 2,32 | 2,25 | 2,18 | 2,10 | 2,05 | 2,01 | 1,96 | 1,92 | 1,87 | 1,81 |
| 22                   | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 | 2,23 | 2,15 | 2,07 | 2,03 | 1,98 | 1,94 | 1,89 | 1,84 | 1,78 |
| 23                   | 4,28 | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,44 | 2,37 | 2,32 | 2,27 | 2,20 | 2,13 | 2,05 | 2,01 | 1,96 | 1,91 | 1,86 | 1,81 | 1,76 |
| 24                   | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 | 2,18 | 2,11 | 2,03 | 1,98 | 1,94 | 1,89 | 1,84 | 1,79 | 1,73 |
| 25                   | 4,24 | 3,39 | 2,99 | 2,76 | 2,60 | 2,49 | 2,40 | 2,34 | 2,28 | 2,24 | 2,16 | 2,09 | 2,01 | 1,96 | 1,92 | 1,87 | 1,82 | 1,77 | 1,71 |
| 26                   | 4,23 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 | 2,15 | 2,07 | 1,99 | 1,95 | 1,90 | 1,85 | 1,80 | 1,75 | 1,69 |
| 27                   | 4,21 | 3,35 | 2,96 | 2,73 | 2,57 | 2,46 | 2,37 | 2,31 | 2,25 | 2,20 | 2,13 | 2,06 | 1,97 | 1,93 | 1,88 | 1,84 | 1,79 | 1,73 | 1,67 |
| 28                   | 4,20 | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 | 2,12 | 2,04 | 1,96 | 1,91 | 1,87 | 1,82 | 1,77 | 1,71 | 1,65 |
| 29                   | 4,18 | 3,33 | 2,93 | 2,70 | 2,55 | 2,43 | 2,35 | 2,28 | 2,22 | 2,18 | 2,10 | 2,03 | 1,94 | 1,90 | 1,85 | 1,81 | 1,75 | 1,70 | 1,64 |
| 30                   | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 | 2,09 | 2,01 | 1,93 | 1,89 | 1,84 | 1,79 | 1,74 | 1,68 | 1,62 |
| 40                   | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 | 2,00 | 1,92 | 1,84 | 1,79 | 1,74 | 1,69 | 1,64 | 1,58 | 1,51 |
| 60                   | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 | 1,92 | 1,84 | 1,75 | 1,70 | 1,65 | 1,59 | 1,53 | 1,47 | 1,39 |
| 120                  | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,18 | 2,09 | 2,02 | 1,96 | 1,91 | 1,83 | 1,75 | 1,66 | 1,61 | 1,55 | 1,50 | 1,43 | 1,35 | 1,25 |
| `                    | 3,84 | 3,00 | 2,60 | 2,37 | 2,21 | 2,10 | 2,01 | 1,94 | 1,88 | 1,83 | 1,75 | 1,67 | 1,57 | 1,52 | 1,46 | 1,39 | 1,32 | 1,22 | 1,00 |



Valores Críticos de Cola Derecha para la Distribución F de Fisher, 1% superior.

| $v_2 \backslash v_1$ | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      | 12      | 15      | 20      | 24      | 30      | 40      | 60      | 120     | ∞       |
|----------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1                    | 4052,18 | 4999,34 | 5403,53 | 5624,26 | 5763,96 | 5858,95 | 5928,33 | 5980,95 | 6022,40 | 6055,93 | 6106,68 | 6156,97 | 6208,66 | 6234,27 | 6260,35 | 6286,43 | 6312,97 | 6339,51 | 6365,59 |
| 2                    | 98,50   | 99,00   | 99,16   | 99,25   | 99,30   | 99,33   | 99,36   | 99,38   | 99,39   | 99,40   | 99,42   | 99,43   | 99,45   | 99,46   | 99,47   | 99,48   | 99,48   | 99,49   | 99,50   |
| 3                    | 34,12   | 30,82   | 29,46   | 28,71   | 28,24   | 27,91   | 27,67   | 27,49   | 27,34   | 27,23   | 27,05   | 26,87   | 26,69   | 26,60   | 26,50   | 26,41   | 26,32   | 26,22   | 26,13   |
| 4                    | 21,20   | 18,00   | 16,69   | 15,98   | 15,52   | 15,21   | 14,98   | 14,80   | 14,66   | 14,55   | 14,37   | 14,20   | 14,02   | 13,93   | 13,84   | 13,75   | 13,65   | 13,56   | 13,46   |
| 5                    | 16,26   | 13,27   | 12,06   | 11,39   | 10,97   | 10,67   | 10,46   | 10,29   | 10,16   | 10,05   | 9,89    | 9,72    | 9,55    | 9,47    | 9,38    | 9,29    | 9,20    | 9,11    | 9,02    |
| 6                    | 13,75   | 10,92   | 9,78    | 9,15    | 8,75    | 8,47    | 8,26    | 8,10    | 7,98    | 7,87    | 7,72    | 7,56    | 7,40    | 7,31    | 7,23    | 7,14    | 7,06    | 6,97    | 6,88    |
| 7                    | 12,25   | 9,55    | 8,45    | 7,85    | 7,46    | 7,19    | 6,99    | 6,84    | 6,72    | 6,62    | 6,47    | 6,31    | 6,16    | 6,07    | 5,99    | 5,91    | 5,82    | 5,74    | 5,65    |
| 8                    | 11,26   | 8,65    | 7,59    | 7,01    | 6,63    | 6,37    | 6,18    | 6,03    | 5,91    | 5,81    | 5,67    | 5,52    | 5,36    | 5,28    | 5,20    | 5,12    | 5,03    | 4,95    | 4,86    |
| 9                    | 10,56   | 8,02    | 6,99    | 6,42    | 6,06    | 5,80    | 5,61    | 5,47    | 5,35    | 5,26    | 5,11    | 4,96    | 4,81    | 4,73    | 4,65    | 4,57    | 4,48    | 4,40    | 4,31    |
| 10                   | 10,04   | 7,56    | 6,55    | 5,99    | 5,64    | 5,39    | 5,20    | 5,06    | 4,94    | 4,85    | 4,71    | 4,56    | 4,41    | 4,33    | 4,25    | 4,17    | 4,08    | 4,00    | 3,91    |
| 11                   | 9,65    | 7,21    | 6,22    | 5,67    | 5,32    | 5,07    | 4,89    | 4,74    | 4,63    | 4,54    | 4,40    | 4,25    | 4,10    | 4,02    | 3,94    | 3,86    | 3,78    | 3,69    | 3,60    |
| 12                   | 9,33    | 6,93    | 5,95    | 5,41    | 5,06    | 4,82    | 4,64    | 4,50    | 4,39    | 4,30    | 4,16    | 4,01    | 3,86    | 3,78    | 3,70    | 3,62    | 3,54    | 3,45    | 3,36    |
| 13                   | 9,07    | 6,70    | 5,74    | 5,21    | 4,86    | 4,62    | 4,44    | 4,30    | 4,19    | 4,10    | 3,96    | 3,82    | 3,66    | 3,59    | 3,51    | 3,43    | 3,34    | 3,25    | 3,17    |
| 14                   | 8,86    | 6,51    | 5,56    | 5,04    | 4,69    | 4,46    | 4,28    | 4,14    | 4,03    | 3,94    | 3,80    | 3,66    | 3,51    | 3,43    | 3,35    | 3,27    | 3,18    | 3,09    | 3,00    |

Continúa en la página siguiente.



| $v_2 \backslash v_1$ | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 12   | 15   | 20   | 24   | 30   | 40   | 60   | 120  | `    |
|----------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 15                   | 8,68 | 6,36 | 5,42 | 4,89 | 4,56 | 4,32 | 4,14 | 4,00 | 3,89 | 3,80 | 3,67 | 3,52 | 3,37 | 3,29 | 3,21 | 3,13 | 3,05 | 2,96 | 2,87 |
| 16                   | 8,53 | 6,23 | 5,29 | 4,77 | 4,44 | 4,20 | 4,03 | 3,89 | 3,78 | 3,69 | 3,55 | 3,41 | 3,26 | 3,18 | 3,10 | 3,02 | 2,93 | 2,84 | 2,75 |
| 17                   | 8,40 | 6,11 | 5,19 | 4,67 | 4,34 | 4,10 | 3,93 | 3,79 | 3,68 | 3,59 | 3,46 | 3,31 | 3,16 | 3,08 | 3,00 | 2,92 | 2,83 | 2,75 | 2,65 |
| 18                   | 8,29 | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,84 | 3,71 | 3,60 | 3,51 | 3,37 | 3,23 | 3,08 | 3,00 | 2,92 | 2,84 | 2,75 | 2,66 | 2,57 |
| 19                   | 8,18 | 5,93 | 5,01 | 4,50 | 4,17 | 3,94 | 3,77 | 3,63 | 3,52 | 3,43 | 3,30 | 3,15 | 3,00 | 2,92 | 2,84 | 2,76 | 2,67 | 2,58 | 2,49 |
| 20                   | 8,10 | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,70 | 3,56 | 3,46 | 3,37 | 3,23 | 3,09 | 2,94 | 2,86 | 2,78 | 2,69 | 2,61 | 2,52 | 2,42 |
| 21                   | 8,02 | 5,78 | 4,87 | 4,37 | 4,04 | 3,81 | 3,64 | 3,51 | 3,40 | 3,31 | 3,17 | 3,03 | 2,88 | 2,80 | 2,72 | 2,64 | 2,55 | 2,46 | 2,36 |
| 22                   | 7,95 | 5,72 | 4,82 | 4,31 | 3,99 | 3,76 | 3,59 | 3,45 | 3,35 | 3,26 | 3,12 | 2,98 | 2,83 | 2,75 | 2,67 | 2,58 | 2,50 | 2,40 | 2,31 |
| 23                   | 7,88 | 5,66 | 4,76 | 4,26 | 3,94 | 3,71 | 3,54 | 3,41 | 3,30 | 3,21 | 3,07 | 2,93 | 2,78 | 2,70 | 2,62 | 2,54 | 2,45 | 2,35 | 2,26 |
| 24                   | 7,82 | 5,61 | 4,72 | 4,22 | 3,90 | 3,67 | 3,50 | 3,36 | 3,26 | 3,17 | 3,03 | 2,89 | 2,74 | 2,66 | 2,58 | 2,49 | 2,40 | 2,31 | 2,21 |
| 25                   | 7,77 | 5,57 | 4,68 | 4,18 | 3,85 | 3,63 | 3,46 | 3,32 | 3,22 | 3,13 | 2,99 | 2,85 | 2,70 | 2,62 | 2,54 | 2,45 | 2,36 | 2,27 | 2,17 |
| 26                   | 7,72 | 5,53 | 4,64 | 4,14 | 3,82 | 3,59 | 3,42 | 3,29 | 3,18 | 3,09 | 2,96 | 2,81 | 2,66 | 2,58 | 2,50 | 2,42 | 2,33 | 2,23 | 2,13 |
| 27                   | 7,68 | 5,49 | 4,60 | 4,11 | 3,78 | 3,56 | 3,39 | 3,26 | 3,15 | 3,06 | 2,93 | 2,78 | 2,63 | 2,55 | 2,47 | 2,38 | 2,29 | 2,20 | 2,10 |
| 28                   | 7,64 | 5,45 | 4,57 | 4,07 | 3,75 | 3,53 | 3,36 | 3,23 | 3,12 | 3,03 | 2,90 | 2,75 | 2,60 | 2,52 | 2,44 | 2,35 | 2,26 | 2,17 | 2,06 |
| 29                   | 7,60 | 5,42 | 4,54 | 4,04 | 3,73 | 3,50 | 3,33 | 3,20 | 3,09 | 3,00 | 2,87 | 2,73 | 2,57 | 2,49 | 2,41 | 2,33 | 2,23 | 2,14 | 2,03 |
| 30                   | 7,56 | 5,39 | 4,51 | 4,02 | 3,70 | 3,47 | 3,30 | 3,17 | 3,07 | 2,98 | 2,84 | 2,70 | 2,55 | 2,47 | 2,39 | 2,30 | 2,21 | 2,11 | 2,01 |
| 40                   | 7,31 | 5,18 | 4,31 | 3,83 | 3,51 | 3,29 | 3,12 | 2,99 | 2,89 | 2,80 | 2,66 | 2,52 | 2,37 | 2,29 | 2,20 | 2,11 | 2,02 | 1,92 | 1,80 |
| 60                   | 7,08 | 4,98 | 4,13 | 3,65 | 3,34 | 3,12 | 2,95 | 2,82 | 2,72 | 2,63 | 2,50 | 2,35 | 2,20 | 2,12 | 2,03 | 1,94 | 1,84 | 1,73 | 1,60 |
| 120                  | 6,85 | 4,79 | 3,95 | 3,48 | 3,17 | 2,96 | 2,79 | 2,66 | 2,56 | 2,47 | 2,34 | 2,19 | 2,03 | 1,95 | 1,86 | 1,76 | 1,66 | 1,53 | 1,38 |
| `                    | 6,63 | 4,61 | 3,78 | 3,32 | 3,02 | 2,80 | 2,64 | 2,51 | 2,41 | 2,32 | 2,18 | 2,04 | 1,88 | 1,79 | 1,70 | 1,59 | 1,47 | 1,32 | 1,00 |

**Estadígrafo  $d$  de Durbin-Watson: puntos  $d_L$  y  $d_U$  para un 5% de significación.**

| $n$ | $k=1$ |       | $k=2$ |       | $k=3$ |       | $k=4$ |       | $k=5$ |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15  | 1,077 | 1,361 | 0,945 | 1,543 | 0,814 | 1,750 | 0,685 | 1,977 | 0,562 | 2,220 |
| 16  | 1,106 | 1,371 | 0,982 | 1,539 | 0,857 | 1,728 | 0,734 | 1,935 | 0,615 | 2,157 |
| 17  | 1,133 | 1,381 | 1,015 | 1,536 | 0,897 | 1,710 | 0,779 | 1,900 | 0,664 | 2,104 |
| 18  | 1,158 | 1,392 | 1,046 | 1,535 | 0,933 | 1,696 | 0,820 | 1,872 | 0,710 | 2,060 |
| 19  | 1,180 | 1,401 | 1,075 | 1,535 | 0,967 | 1,685 | 0,859 | 1,848 | 0,752 | 2,022 |
| 20  | 1,201 | 1,411 | 1,100 | 1,537 | 0,998 | 1,676 | 0,894 | 1,828 | 0,792 | 1,991 |
| 21  | 1,221 | 1,420 | 1,125 | 1,538 | 1,026 | 1,669 | 0,927 | 1,812 | 0,828 | 1,964 |
| 22  | 1,240 | 1,429 | 1,147 | 1,541 | 1,053 | 1,664 | 0,958 | 1,797 | 0,863 | 1,940 |
| 23  | 1,257 | 1,437 | 1,168 | 1,543 | 1,078 | 1,660 | 0,986 | 1,786 | 0,895 | 1,919 |
| 24  | 1,273 | 1,446 | 1,188 | 1,546 | 1,101 | 1,657 | 1,013 | 1,775 | 0,925 | 1,902 |
| 25  | 1,288 | 1,454 | 1,206 | 1,550 | 1,123 | 1,654 | 1,038 | 1,767 | 0,953 | 1,886 |
| 26  | 1,302 | 1,461 | 1,224 | 1,553 | 1,143 | 1,652 | 1,062 | 1,759 | 0,979 | 1,873 |
| 27  | 1,316 | 1,468 | 1,240 | 1,556 | 1,162 | 1,651 | 1,083 | 1,753 | 1,004 | 1,861 |
| 28  | 1,328 | 1,476 | 1,255 | 1,560 | 1,181 | 1,650 | 1,104 | 1,747 | 1,028 | 1,850 |
| 29  | 1,341 | 1,483 | 1,270 | 1,563 | 1,198 | 1,650 | 1,124 | 1,743 | 1,050 | 1,841 |
| 30  | 1,352 | 1,489 | 1,284 | 1,567 | 1,214 | 1,650 | 1,143 | 1,739 | 1,070 | 1,833 |
| 31  | 1,363 | 1,496 | 1,297 | 1,570 | 1,229 | 1,650 | 1,160 | 1,735 | 1,090 | 1,825 |
| 32  | 1,373 | 1,502 | 1,309 | 1,573 | 1,244 | 1,650 | 1,177 | 1,732 | 1,109 | 1,819 |
| 33  | 1,383 | 1,508 | 1,321 | 1,577 | 1,258 | 1,651 | 1,193 | 1,730 | 1,127 | 1,813 |
| 34  | 1,393 | 1,514 | 1,332 | 1,580 | 1,271 | 1,652 | 1,208 | 1,728 | 1,144 | 1,807 |
| 35  | 1,402 | 1,519 | 1,343 | 1,584 | 1,283 | 1,653 | 1,222 | 1,726 | 1,160 | 1,803 |
| 36  | 1,411 | 1,524 | 1,354 | 1,587 | 1,295 | 1,654 | 1,236 | 1,725 | 1,175 | 1,799 |
| 37  | 1,419 | 1,530 | 1,364 | 1,590 | 1,307 | 1,655 | 1,249 | 1,723 | 1,190 | 1,795 |
| 38  | 1,427 | 1,535 | 1,373 | 1,594 | 1,317 | 1,656 | 1,261 | 1,723 | 1,204 | 1,792 |
| 39  | 1,435 | 1,540 | 1,382 | 1,597 | 1,328 | 1,658 | 1,273 | 1,722 | 1,218 | 1,789 |
| 40  | 1,442 | 1,544 | 1,391 | 1,600 | 1,338 | 1,659 | 1,285 | 1,721 | 1,231 | 1,786 |
| 45  | 1,475 | 1,566 | 1,430 | 1,615 | 1,383 | 1,666 | 1,336 | 1,720 | 1,287 | 1,776 |
| 50  | 1,503 | 1,585 | 1,462 | 1,628 | 1,421 | 1,674 | 1,378 | 1,721 | 1,334 | 1,771 |
| 55  | 1,527 | 1,601 | 1,490 | 1,640 | 1,452 | 1,681 | 1,414 | 1,724 | 1,374 | 1,768 |
| 60  | 1,549 | 1,616 | 1,514 | 1,652 | 1,480 | 1,689 | 1,444 | 1,727 | 1,408 | 1,767 |
| 65  | 1,567 | 1,629 | 1,536 | 1,662 | 1,503 | 1,696 | 1,471 | 1,731 | 1,438 | 1,767 |
| 70  | 1,583 | 1,641 | 1,554 | 1,671 | 1,524 | 1,703 | 1,494 | 1,735 | 1,464 | 1,768 |
| 75  | 1,598 | 1,652 | 1,571 | 1,680 | 1,543 | 1,709 | 1,515 | 1,739 | 1,486 | 1,770 |
| 80  | 1,611 | 1,662 | 1,586 | 1,688 | 1,560 | 1,715 | 1,534 | 1,743 | 1,507 | 1,772 |
| 85  | 1,624 | 1,671 | 1,600 | 1,696 | 1,575 | 1,721 | 1,551 | 1,747 | 1,525 | 1,774 |
| 90  | 1,635 | 1,679 | 1,612 | 1,703 | 1,589 | 1,726 | 1,566 | 1,751 | 1,542 | 1,776 |
| 95  | 1,645 | 1,687 | 1,623 | 1,709 | 1,602 | 1,732 | 1,579 | 1,755 | 1,557 | 1,778 |
| 100 | 1,654 | 1,694 | 1,634 | 1,715 | 1,613 | 1,736 | 1,592 | 1,758 | 1,571 | 1,780 |

$n$  = número de observaciones.

$k$  = número de variables explicativas.

**Estadígrafo  $d$  de Durbin-Watson: puntos  $d_L$  y  $d_U$  para un 2,5% de significación.**

| $n$ | $k=1$ |       | $k=2$ |       | $k=3$ |       | $k=4$ |       | $k=5$ |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15  | 0,949 | 1,222 | 0,827 | 1,405 | 0,706 | 1,615 | 0,588 | 1,848 | 0,478 | 2,099 |
| 16  | 0,980 | 1,235 | 0,864 | 1,403 | 0,748 | 1,594 | 0,636 | 1,806 | 0,527 | 2,035 |
| 17  | 1,009 | 1,248 | 0,899 | 1,403 | 0,788 | 1,578 | 0,680 | 1,773 | 0,574 | 1,983 |
| 18  | 1,035 | 1,261 | 0,930 | 1,405 | 0,825 | 1,567 | 0,720 | 1,746 | 0,619 | 1,939 |
| 19  | 1,060 | 1,274 | 0,959 | 1,407 | 0,859 | 1,558 | 0,758 | 1,724 | 0,660 | 1,902 |
| 20  | 1,082 | 1,286 | 0,988 | 1,410 | 0,890 | 1,551 | 0,794 | 1,705 | 0,699 | 1,871 |
| 21  | 1,104 | 1,297 | 1,012 | 1,415 | 0,920 | 1,546 | 0,826 | 1,691 | 0,734 | 1,845 |
| 22  | 1,124 | 1,308 | 1,036 | 1,419 | 0,947 | 1,543 | 0,858 | 1,678 | 0,769 | 1,823 |
| 23  | 1,144 | 1,319 | 1,059 | 1,424 | 0,973 | 1,541 | 0,887 | 1,668 | 0,801 | 1,804 |
| 24  | 1,161 | 1,329 | 1,080 | 1,429 | 0,997 | 1,539 | 0,914 | 1,659 | 0,830 | 1,787 |
| 25  | 1,178 | 1,339 | 1,099 | 1,435 | 1,019 | 1,539 | 0,939 | 1,652 | 0,859 | 1,773 |
| 26  | 1,194 | 1,348 | 1,118 | 1,439 | 1,041 | 1,538 | 0,964 | 1,646 | 0,886 | 1,761 |
| 27  | 1,208 | 1,358 | 1,135 | 1,445 | 1,061 | 1,539 | 0,986 | 1,641 | 0,911 | 1,751 |
| 28  | 1,222 | 1,367 | 1,153 | 1,450 | 1,080 | 1,540 | 1,007 | 1,637 | 0,934 | 1,742 |
| 29  | 1,236 | 1,375 | 1,168 | 1,455 | 1,098 | 1,541 | 1,028 | 1,634 | 0,958 | 1,734 |
| 30  | 1,249 | 1,383 | 1,183 | 1,460 | 1,115 | 1,542 | 1,047 | 1,632 | 0,978 | 1,727 |
| 31  | 1,261 | 1,391 | 1,197 | 1,465 | 1,132 | 1,544 | 1,066 | 1,630 | 0,999 | 1,721 |
| 32  | 1,273 | 1,399 | 1,211 | 1,469 | 1,147 | 1,546 | 1,083 | 1,628 | 1,018 | 1,715 |
| 33  | 1,284 | 1,406 | 1,224 | 1,474 | 1,163 | 1,548 | 1,099 | 1,627 | 1,037 | 1,711 |
| 34  | 1,294 | 1,413 | 1,236 | 1,479 | 1,176 | 1,550 | 1,115 | 1,626 | 1,054 | 1,707 |
| 35  | 1,305 | 1,420 | 1,248 | 1,484 | 1,190 | 1,553 | 1,131 | 1,626 | 1,071 | 1,704 |
| 36  | 1,315 | 1,426 | 1,259 | 1,488 | 1,203 | 1,555 | 1,145 | 1,625 | 1,087 | 1,701 |
| 37  | 1,324 | 1,433 | 1,270 | 1,493 | 1,215 | 1,557 | 1,159 | 1,625 | 1,102 | 1,698 |
| 38  | 1,333 | 1,439 | 1,281 | 1,497 | 1,227 | 1,560 | 1,173 | 1,625 | 1,117 | 1,695 |
| 39  | 1,342 | 1,445 | 1,291 | 1,501 | 1,238 | 1,562 | 1,185 | 1,626 | 1,131 | 1,693 |
| 40  | 1,350 | 1,451 | 1,300 | 1,506 | 1,249 | 1,564 | 1,197 | 1,626 | 1,144 | 1,692 |
| 45  | 1,388 | 1,477 | 1,343 | 1,525 | 1,298 | 1,576 | 1,252 | 1,630 | 1,204 | 1,687 |
| 50  | 1,420 | 1,500 | 1,380 | 1,543 | 1,338 | 1,588 | 1,297 | 1,636 | 1,255 | 1,685 |
| 55  | 1,447 | 1,520 | 1,411 | 1,559 | 1,373 | 1,600 | 1,335 | 1,642 | 1,297 | 1,686 |
| 60  | 1,471 | 1,538 | 1,438 | 1,573 | 1,404 | 1,610 | 1,369 | 1,649 | 1,333 | 1,688 |
| 65  | 1,492 | 1,554 | 1,461 | 1,587 | 1,430 | 1,620 | 1,398 | 1,655 | 1,365 | 1,691 |
| 70  | 1,511 | 1,568 | 1,482 | 1,599 | 1,453 | 1,630 | 1,424 | 1,662 | 1,393 | 1,695 |
| 75  | 1,528 | 1,582 | 1,501 | 1,610 | 1,474 | 1,638 | 1,446 | 1,668 | 1,418 | 1,699 |
| 80  | 1,543 | 1,594 | 1,518 | 1,619 | 1,493 | 1,647 | 1,467 | 1,674 | 1,441 | 1,703 |
| 85  | 1,557 | 1,605 | 1,534 | 1,629 | 1,510 | 1,654 | 1,485 | 1,680 | 1,461 | 1,707 |
| 90  | 1,570 | 1,614 | 1,548 | 1,638 | 1,525 | 1,662 | 1,502 | 1,686 | 1,479 | 1,711 |
| 95  | 1,582 | 1,624 | 1,560 | 1,646 | 1,539 | 1,668 | 1,517 | 1,691 | 1,495 | 1,715 |
| 100 | 1,593 | 1,633 | 1,573 | 1,654 | 1,552 | 1,675 | 1,532 | 1,696 | 1,511 | 1,718 |

$n$  = número de observaciones.

$k$  = número de variables explicativas.

**Estadígrafo  $d$  de Durbin-Watson: puntos  $d_L$  y  $d_U$  para un 1% de significación.**

| $n$ | $k=1$ |       | $k=2$ |       | $k=3$ |       | $k=4$ |       | $k=5$ |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15  | 0,811 | 1,069 | 0,700 | 1,252 | 0,591 | 1,465 | 0,487 | 1,705 | 0,390 | 1,967 |
| 16  | 0,844 | 1,087 | 0,738 | 1,253 | 0,633 | 1,447 | 0,532 | 1,664 | 0,437 | 1,901 |
| 17  | 0,873 | 1,102 | 0,773 | 1,255 | 0,672 | 1,432 | 0,574 | 1,631 | 0,481 | 1,847 |
| 18  | 0,902 | 1,118 | 0,805 | 1,259 | 0,708 | 1,422 | 0,614 | 1,604 | 0,522 | 1,803 |
| 19  | 0,928 | 1,133 | 0,835 | 1,264 | 0,742 | 1,416 | 0,650 | 1,583 | 0,561 | 1,767 |
| 20  | 0,952 | 1,147 | 0,862 | 1,270 | 0,774 | 1,410 | 0,684 | 1,567 | 0,598 | 1,736 |
| 21  | 0,975 | 1,161 | 0,889 | 1,276 | 0,803 | 1,408 | 0,718 | 1,554 | 0,634 | 1,712 |
| 22  | 0,997 | 1,174 | 0,915 | 1,284 | 0,832 | 1,407 | 0,748 | 1,543 | 0,666 | 1,691 |
| 23  | 1,017 | 1,186 | 0,938 | 1,290 | 0,858 | 1,407 | 0,777 | 1,535 | 0,699 | 1,674 |
| 24  | 1,037 | 1,199 | 0,959 | 1,298 | 0,881 | 1,407 | 0,805 | 1,527 | 0,728 | 1,659 |
| 25  | 1,055 | 1,210 | 0,981 | 1,305 | 0,906 | 1,408 | 0,832 | 1,521 | 0,756 | 1,645 |
| 26  | 1,072 | 1,222 | 1,000 | 1,311 | 0,928 | 1,410 | 0,855 | 1,517 | 0,782 | 1,635 |
| 27  | 1,088 | 1,232 | 1,019 | 1,318 | 0,948 | 1,413 | 0,878 | 1,514 | 0,808 | 1,625 |
| 28  | 1,104 | 1,244 | 1,036 | 1,325 | 0,969 | 1,414 | 0,901 | 1,512 | 0,832 | 1,618 |
| 29  | 1,119 | 1,254 | 1,053 | 1,332 | 0,988 | 1,418 | 0,921 | 1,511 | 0,855 | 1,611 |
| 30  | 1,134 | 1,264 | 1,070 | 1,339 | 1,006 | 1,421 | 0,941 | 1,510 | 0,877 | 1,606 |
| 31  | 1,147 | 1,274 | 1,085 | 1,345 | 1,022 | 1,425 | 0,960 | 1,509 | 0,897 | 1,601 |
| 32  | 1,160 | 1,283 | 1,100 | 1,351 | 1,039 | 1,428 | 0,978 | 1,509 | 0,917 | 1,597 |
| 33  | 1,171 | 1,291 | 1,114 | 1,358 | 1,055 | 1,432 | 0,995 | 1,510 | 0,935 | 1,594 |
| 34  | 1,184 | 1,298 | 1,128 | 1,364 | 1,070 | 1,436 | 1,012 | 1,511 | 0,954 | 1,591 |
| 35  | 1,195 | 1,307 | 1,141 | 1,370 | 1,085 | 1,439 | 1,028 | 1,512 | 0,971 | 1,589 |
| 36  | 1,205 | 1,315 | 1,153 | 1,376 | 1,098 | 1,442 | 1,043 | 1,513 | 0,987 | 1,587 |
| 37  | 1,217 | 1,322 | 1,164 | 1,383 | 1,112 | 1,446 | 1,058 | 1,514 | 1,004 | 1,585 |
| 38  | 1,227 | 1,330 | 1,176 | 1,388 | 1,124 | 1,449 | 1,072 | 1,515 | 1,019 | 1,584 |
| 39  | 1,237 | 1,337 | 1,187 | 1,392 | 1,137 | 1,452 | 1,085 | 1,517 | 1,033 | 1,583 |
| 40  | 1,246 | 1,344 | 1,197 | 1,398 | 1,149 | 1,456 | 1,098 | 1,518 | 1,047 | 1,583 |
| 45  | 1,288 | 1,376 | 1,245 | 1,424 | 1,201 | 1,474 | 1,156 | 1,528 | 1,111 | 1,583 |
| 50  | 1,324 | 1,403 | 1,285 | 1,445 | 1,245 | 1,491 | 1,206 | 1,537 | 1,164 | 1,587 |
| 55  | 1,356 | 1,428 | 1,320 | 1,466 | 1,284 | 1,505 | 1,246 | 1,548 | 1,209 | 1,592 |
| 60  | 1,382 | 1,449 | 1,351 | 1,484 | 1,317 | 1,520 | 1,283 | 1,559 | 1,248 | 1,598 |
| 65  | 1,407 | 1,467 | 1,377 | 1,500 | 1,346 | 1,534 | 1,314 | 1,568 | 1,283 | 1,604 |
| 70  | 1,429 | 1,485 | 1,400 | 1,514 | 1,372 | 1,546 | 1,343 | 1,577 | 1,313 | 1,611 |
| 75  | 1,448 | 1,501 | 1,422 | 1,529 | 1,395 | 1,557 | 1,368 | 1,586 | 1,340 | 1,617 |
| 80  | 1,465 | 1,514 | 1,440 | 1,541 | 1,416 | 1,568 | 1,390 | 1,595 | 1,364 | 1,624 |
| 85  | 1,481 | 1,529 | 1,458 | 1,553 | 1,434 | 1,577 | 1,411 | 1,603 | 1,386 | 1,630 |
| 90  | 1,496 | 1,541 | 1,474 | 1,563 | 1,452 | 1,587 | 1,429 | 1,611 | 1,406 | 1,636 |
| 95  | 1,510 | 1,552 | 1,489 | 1,573 | 1,468 | 1,596 | 1,446 | 1,618 | 1,425 | 1,641 |
| 100 | 1,522 | 1,562 | 1,502 | 1,582 | 1,482 | 1,604 | 1,461 | 1,625 | 1,441 | 1,647 |

$n$  = número de observaciones.

$k$  = número de variables explicativas.

Valores Críticos para la distribución  $\chi^2$ .

| DF  | $\alpha = 0,10$ | $\alpha = 0,05$ | $\alpha = 0,01$ | $\alpha = 0,005$ |
|-----|-----------------|-----------------|-----------------|------------------|
| 1   | 2,71            | 3,84            | 6,63            | 7,88             |
| 2   | 4,61            | 5,99            | 9,21            | 10,60            |
| 3   | 6,25            | 7,81            | 11,34           | 12,84            |
| 4   | 7,78            | 9,49            | 13,28           | 14,86            |
| 5   | 9,24            | 11,07           | 15,09           | 16,75            |
| 6   | 10,64           | 12,59           | 16,81           | 18,55            |
| 7   | 12,02           | 14,07           | 18,48           | 20,28            |
| 8   | 13,36           | 15,51           | 20,09           | 21,95            |
| 9   | 14,68           | 16,92           | 21,67           | 23,59            |
| 10  | 15,99           | 18,31           | 23,21           | 25,19            |
| 11  | 17,28           | 19,68           | 24,72           | 26,76            |
| 12  | 18,55           | 21,03           | 26,22           | 28,30            |
| 13  | 19,81           | 22,36           | 27,69           | 29,82            |
| 14  | 21,06           | 23,68           | 29,14           | 31,32            |
| 15  | 22,31           | 25,00           | 30,58           | 32,80            |
| 16  | 23,54           | 26,30           | 32,00           | 34,27            |
| 17  | 24,77           | 27,59           | 33,41           | 35,72            |
| 18  | 25,99           | 28,87           | 34,81           | 37,16            |
| 19  | 27,20           | 30,14           | 36,19           | 38,58            |
| 20  | 28,41           | 31,41           | 37,57           | 40,00            |
| 21  | 29,62           | 32,67           | 38,93           | 41,40            |
| 22  | 30,81           | 33,92           | 40,29           | 42,80            |
| 23  | 32,01           | 35,17           | 41,64           | 44,18            |
| 24  | 33,20           | 36,42           | 42,98           | 45,56            |
| 25  | 34,38           | 37,65           | 44,31           | 46,93            |
| 26  | 35,56           | 38,89           | 45,64           | 48,29            |
| 27  | 36,74           | 40,11           | 46,96           | 49,64            |
| 28  | 37,92           | 41,34           | 48,28           | 50,99            |
| 29  | 39,09           | 42,56           | 49,59           | 52,34            |
| 30  | 40,26           | 43,77           | 50,89           | 53,67            |
| 40  | 51,81           | 55,76           | 63,69           | 66,77            |
| 50  | 63,17           | 67,50           | 76,15           | 79,49            |
| 60  | 74,40           | 79,08           | 88,38           | 91,95            |
| 70  | 85,53           | 90,53           | 100,43          | 104,21           |
| 80  | 96,58           | 101,88          | 112,33          | 116,32           |
| 90  | 107,57          | 113,15          | 124,12          | 128,30           |
| 100 | 118,50          | 124,34          | 135,81          | 140,17           |
| 110 | 129,39          | 135,48          | 147,41          | 151,95           |
| 120 | 140,23          | 146,57          | 158,95          | 163,65           |
| 200 | 226,02          | 233,99          | 249,45          | 255,26           |

**Valores críticos de  $\tau$  de la distribución Dickey-Fuller.**  
(Cola izquierda de la distribución).

|                                               | T        | 0,01  | 0,025 | 0,05  | 0,1   |
|-----------------------------------------------|----------|-------|-------|-------|-------|
| <b>Sin constante y sin tendencia temporal</b> | 25       | -2,66 | -2,26 | -1,95 | -1,60 |
|                                               | 50       | -2,62 | -2,25 | -1,95 | -1,61 |
|                                               | 100      | -2,60 | -2,24 | -1,95 | -1,61 |
|                                               | 250      | -2,58 | -2,23 | -1,95 | -1,62 |
|                                               | 300      | -2,58 | -2,23 | -1,95 | -1,62 |
|                                               | $\infty$ | -2,58 | -2,23 | -1,95 | -1,62 |
| <b>Con constante y sin tendencia temporal</b> | 25       | -3,75 | -3,33 | -3,00 | -2,62 |
|                                               | 50       | -3,58 | -3,22 | -2,93 | -2,60 |
|                                               | 100      | -3,51 | -3,17 | -2,89 | -2,58 |
|                                               | 250      | -3,46 | -3,14 | -2,88 | -2,57 |
|                                               | 300      | -3,44 | -3,13 | -2,87 | -2,57 |
|                                               | $\infty$ | -3,43 | -3,12 | -2,86 | -2,57 |
| <b>Con constante y con tendencia temporal</b> | 25       | -4,38 | -3,95 | -3,60 | -3,24 |
|                                               | 50       | -4,15 | -3,80 | -3,50 | -3,18 |
|                                               | 100      | -4,04 | -3,73 | -3,45 | -3,15 |
|                                               | 250      | -3,99 | -3,69 | -3,43 | -3,13 |
|                                               | 300      | -3,98 | -3,68 | -3,42 | -3,13 |
|                                               | $\infty$ | -3,96 | -3,66 | -3,41 | -3,12 |

T es el tamaño de la muestra.

L primera línea indica lo niveles de significación estadística.

Se tabulan tres modelos distintos.



BIBLIOGRAFÍA





---

## REFERENCIAS BIBLIOGRÁFICAS

- Alonso, José A. y Sánchez-Egozcue, Jorge M. (2005): "La Competitividad desde una Perspectiva Macro: La Restricción Externa al Crecimiento". *Tecnología, Competitividad y Capacidad Exportadora de la Economía Cubana: El Desafío de los Mercados Globales*. La Habana
- Beaulieu, J.J. y Miron, J.A. (1993): "Seasonal Unit Roots in Aggregate U.S. Data". *Journal of Econometrics*, 55, pp. 305-328.
- Bernanke, B. (1986): "Alternative Explanations of Money-Income Correlation". *Carnegie-Rochester Conference Series on Public Policy* 25.
- Bernanke, B. y Mihov (1998): "Measuring Monetary Policy". *Quarterly Journal of Economics* 113(3): 869- 902.
- Blanchard, O. y D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review* 79: 655- 673.
- Box G.E.P. y G.C. Tiao (1975): "Intervention Analysis with Applications to Cartier bressoncartier bressonc Economic and Environmental Problems". *Journal of the American Statistical*, Vol. 70 No. 349, 70-79
- Box, George E. P. y Gwilym M. Jenkins (1976): *Time Series Analysis: Forecasting and Control*. Revised Edition, Oakland, CA: Holden-Day.
- Brundenius, Claes (2000): "The Role of Human Capital in Cuban Economic Development, 1959-1999". *Working Papers*, 00.8, Centre for Development Research, Copenhagen.
- Bu, Angel y col. (1996): "Las UBPC y su Necesario Perfeccionamiento". *Cuba: Investigación Económica*, Año 2, No. 2, pp. 15-42, abril-junio.
- Cassoni, Adriana (1991): "Pruebas de Diagnóstico en el Modelo Económico I y II". Documentos de Trabajo, Centro de Investigación y Docencia Económicas, Facultad de Ciencias Sociales, México.
- Christiano, L., M. Eichenbaun y Ch. Evans (1996): "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds". *The Review of Economics and Statistics* 78: 16-34.

- 
- Comité Estatal de Estadísticas (1990): *Anuario Estadístico de Cuba 1988*. La Habana.
- Cribeiro, Yordanka y Triana, Liliam (2005): “Las Elasticidades en el Comercio Exterior Cubano: Dinámica de Corto y Largo Plazo”. Tesis de Diploma, Facultad de Economía, Universidad de La Habana.
- Dickey, D.A. y W.A. Fuller (1979): “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. *Journal of the American Statistical Association*, 74, 427-431.
- Dresdner Cid, Jorge y Felipe Vázquez Lavín (2004): *Nociones de Econometría Intermedia*. Universidad de Concepción, Facultad de Ciencias Económicas y Administrativas, Departamento de Economía, Proyecto de Docencia 98-049, Vicerrectoría Dirección de Docencia, Concepción.
- Doornik, Jurgen A. y David F. Hendry (1994): *PcGive Student 8.0, an Interactive Econometric Modelling System*. International Thomson Publishing/Duxbury Press.
- Doornik, J. A. y H. Hansen (1994): “A Practical Test of Multivariate Normality”. Unpublished paper, Nuffield College (citados por Doornik y Hendry, 1994).
- Dornbusch, Rudiger, Stanley Fischer y Richard Startz (2002): *Macroeconomía*. Octava Edición, McGraw-Hill, Madrid.
- Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution Function*. S. I. A. M., Philadelphia (citado por Spanos, 1986).
- Enders, W. (1995): *Applied Econometric Time Series*. Wiley Series in Probability and Mathematical Statistics.
- Engle, R. y Granger, C. (1987): “Co-integration and Error-correction. Representation, Estimation, and Testing”. *Econometrica* 35: 251-276.
- Engle, R. y Yoo, Byung Sam (1987): “Forecasting and Testing in Co-integrated Systems”. *Journal of Econometrics* 35: 143-159
- Espasa, A. y Cancelo, J. R. (1993): *Métodos Cuantitativos para el Análisis de la Coyuntura Económica*. Ed. Alianza, Madrid.
- Evans, Ch. y K. Kuttner (1998): “Can VAR Describe Monetary Policy?”. *Conference Papers 6 Topics in Monetary Policy Modeling*. BIS.
- Fernández de Bulnes, Carlos (2001): “Estimación de la Productividad Total de Factores en Cuba 1980-2000”. Estudio inédito, La Habana.
- Fernández Domínguez, Pablo (1997): “El Nuevo Modelo Agrario: un Desafío en el Umbral del Tercer Milenio”. Mención en el Concurso Nacional de Economía Raúl León Torras, La Habana.
- Fernández Domínguez, Pablo (2002): “El Sector Agropecuario en Cuba: Evolución y Perspectivas”. *Cuba: el Sector Agropecuario y las Políticas Agrícolas ante los Nuevos Retos*, MEP-ASDI-DECON-FCS, Montevideo.
- Fulginiti, Lilyan E. y Richard K. Perrin (1997): “LCD Agriculture: Nonparametric Malmquist Productivity Indexes”. *Journal of Development Economics*, Vol. 53, pp. 373-390.

- García Álvarez, Anicia y Eislen Guerra Boza (2005): "Las Agroexportaciones Cubanas: Demanda y Competitividad". *Cuba, Crecer desde el Conocimiento*, Editorial de Ciencias Sociales, La Habana.
- García Álvarez, Anicia y Renato Aguilar (2002): "Eficiencia Técnica y Producto Potencial en el Agro Cubano". *La Economía Cubana en el 2001*, publicación anual del CEEC, La Habana.
- García Álvarez, Anicia (2002): "Productividad y Factores de Producción en la Agricultura Cubana". *Cuba, Reflexiones sobre su Economía*, Editorial Félix Varela, La Habana.
- García, Anicia y Elio Perera (1997): "Diagnóstico Preliminar de la Ganadería". *Escenarios Económicos hasta el 2000*. INIE-CITMA, La Habana.
- García, Anicia y Pablo Fernández (1998): "Los Productos Lácteos en Cuba: Posibilidades de Sustituir su Importación". Proyecto de investigación del PNCT: "La Economía Cubana Actual, Retos y Perspectivas", INIE-MEP, La Habana.
- García, Anicia, Pablo Fernández, Clara Duarte, Arturo Costales, y otros colaboradores, (1999): *Ganadería vacuna, problemas y perspectivas*. MEP-INIE-MINAG, La Habana.
- Ghysels, E. (1990): "Unit-Root Test and the Statistical Pitfalls of Seasonal Adjustment: the Case of U.S Postwar Real Gross National Product". *Journal of Business and Economic Statistics*, 8, 2, pp. 145-152.
- Gómez, V. y Maravall, A. (1995): *Programs TRAMO-SEATS update*. Florence: European University Institute, Series: EUI working papers ECO 0046.
- Gordon, D. B. y E. Leeper (1994): "The Dynamic Impacts of Monetary Policy: an Exercise in Tentative Identification". *Journal of Political Economy* 102: 1228- 1247.
- Granger, C. W. J. y Newbold, P. (1974): "Spurious Regressions in Econometrics". *Journal of Econometrics* 2, pp. 111–120
- Gujarati, Damodar (2004): *Econometría*. Cuarta edición, McGraw-Hill, México.
- Hamilton, J. D. (1994): *Time Series Analysis*. Princenton University Press, Princeton.
- Harvey, Andrew C. y Koopman, Siem J. (1992): "Diagnostic Checking of Unobserved Time Series Models". *Journal of Business and Economic Statistics*, Vol. 10. No. 4.
- Harvey, Andrew C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Hodrick, R. J. y E. C. Prescott (1997): "Postwar U.S. Business Cycles: An Empirical Investigation". *Journal of Money, Credit, and Banking*, 29, 1-16.
- Hylleberg, S., Engle, R.F., Granger, C. W.J. y Yoo, B.S. (1990): "Seasonal Integration and Co-Integration". *Journal of Econometrics*, 44, pp. 215-228.
- Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt y Yongcheol Shin (1992): "Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root". *Journal of Econometrics*, 54, 159-178.
- Ljung, G. y G. Box (1979): "On a Measure of Lack of Fit in Time Series Models". *Biometrika*, 66, 265-270.

- 
- Lorenzo, Fernando (2003): Notas del curso: “Análisis de Coyuntura y Técnicas de Extracción de Señales”, Montevideo.
- MacKinnon, James G. (1996): “Numerical Distribution Functions for Unit Root and Cointegration Tests”. *Journal of Applied Econometrics*, 11, 601-618.
- Marquetti Nodarse, Hiram (2003): “Los Dilemas de la Reestructuración de la Industria Azucarera Cubana”. *Boletín Electrónico Avances de Investigación*, sitio <http://www.uh.cu/centros/ceec>, CEEC, abril, La Habana.
- Mañalich, Isis y Quiñones, Nancy (2004): “Sustitución de Importaciones. Un Desafío Impostergable”. Publicación Electrónica del Evento 42 Aniversario de los Estudios Económicos en la Facultad de Economía de la Universidad de La Habana.
- Mendoza, Yenniel y Robert, Leonel (2002): “El Crecimiento Económico y las Restricciones en el Sector Externo. Una Aplicación al Caso Cubano”. INIE, La Habana.
- Moreno-Brid, Juan Carlos (2000): “Crecimiento Económico y Escasez de Divisas”. *La Economía Cubana. Reformas Estructurales y Desempeño en los Noventa*. CEPAL/Fondo de Cultura Económica, México.
- Nova González, Armando (1997): “Cuba: ¿Evolución o Transformación Agrícola?”. Material docente, Centro de Estudios de Economía y Planificación, Ministerio de Economía y Planificación, La Habana.
- Nova González, Armando (2001): “Las Unidades Básicas de Producción Cooperativa y las Granjas Cañeras entre 1993 y el 2000”. *La Economía Cubana en el 2000: Desempeño Macroeconómico y Transformaciones Empresariales*, publicación anual del CEEC, La Habana, pp. 59-72.
- ONE (Oficina Nacional de Estadísticas de Cuba): *Anuario Estadístico de Cuba, varios años*, La Habana, en [www.one.cu](http://www.one.cu)
- Pérez Villanueva, Omar E. (1999): “La Reestructuración de la Economía Cubana. El Proceso en la Agricultura”. *La Última Reforma Agraria del Siglo*, Editorial Nueva Sociedad, Caracas, pp. 71-105,.
- Phillips, P.C.B. and P. Perron (1988): “Testing for a Unit Root in Time Series Regression”. *Biometrika*, 75, 335-346.
- Pindyck, Robert S. y Daniel L. Rubinfeld (2001): *Econometría: Modelos y Pronósticos*. McGraw-Hill, México.
- Rodríguez Castellón, Santiago (2000): “Transformaciones Agrarias en Cuba: Propuestas para el Desarrollo de una Agricultura Sostenible”. Tesis en opción al grado de Doctor en Ciencias Económicas”, Centro de Estudios de la Economía Cubana, Universidad de La Habana.
- Rodríguez García, José Luis (2005): “Se ha iniciado una nueva etapa de la Revolución”. Informe sobre los resultados económicos del año 2005 y perspectivas para el 2006, en *Granma*, 25 de diciembre, La Habana.
- Rudebusch, G. D. (1998): “Do Measures of Monetary Policy in a VAR Make Sense?”. *International Economic Review* 39: 907- 931.

- Sims, Ch. A. (1980): "Macroeconomics and Reality". *Econometrica* 48: 1-49.
- Sims, Ch. A. (1986): "Are Forecasting Models Usable for Policy Analysis?". *Federal Reserve Bank of Minneapolis, Quarterly Review*: 2-16.
- Sims, Ch. A. (1998): "Comment on Glenn Rudebusch: Do Measures of Monetary Policy in a VAR Make Sense?". *International Economic Review* 39: 933- 941.
- Spanos Aris (1986): *Statistical Foundations of Econometric Modelling*. Cambridge University Press, Cambridge.
- Thirlwall, A. P. (1979): "The Balance of Payments Constraint as an Explanation of International Growth Rate Differences". *Banca Nazionale del Lavoro, Quarterly Review*.
- Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". *Journal of Monetary Economics*, 52(2): 381-419
- Vidal, Daniel Luis (2000): "Informe de Consultoría en Sistemas Modernos de Manejo y Gestión de Complejos Azucareros". Proyecto TCP/CUB/8822: Manejo conservacionista y sostenible del cultivo de la caña de azúcar, FAO-INICA, abril.
- Vidal, Pavel, Hernández Eduardo, Carlos Pérez, Mercedes García, Guillermo Gil y Maria Llorente (2001): "Estacionariedad y Estacionalidad en Series de la Economía Cubana". Banco Central de Cuba, en <http://www.bc.gov.cu/antiores/Otros/InvestPolMon/Estacionariedad%20HEGY.pdf>
- Vidal, Pavel y Annia Fundora (2004): "Tendencia y Ciclos en el Producto Interno Bruto de Cuba: Estimación con un Modelo Estructural Univariante de Series Temporales". Publicación del evento 42 Aniversario de los Estudios Económicos. Universidad de la Habana, en <http://www.bc.gov.cu/Antiores/Otros/InvestPolMon/Ciclos%20PIB.pdf>
- Vidal, Pavel (2007): "Modelo para la predicción del PIB cubano". *Economía y Gerencia en Cuba: Avances de Investigación*, Boletín Cuatrimestral del CEEC enero-abril, sitio web del CEEC <http://www.uh.cu/centros/ceec/index800.htm>
- Vidal, Pavel y Annia Fundora (2008): "Relación Comercio-Crecimiento en Cuba: Estimación con el Filtro de Kalman". *Revista de la CEPAL* 94, Santiago de Chile, abril 2008.
- Vidal, Pavel, (2008): "Política Monetaria en Cuba. Estimación con un Modelo VAR Estructural". *Revista Principios* 12, España, septiembre.
- Zepeda, Lydia (2001): "Agricultural Investment, Production Capacity and Productivity". Zepeda, Lydia (ed): *Agricultural investment and productivity in developing countries*, FAO Economic and Social Development Paper, en sitio [www.fao.org/DOCREP/003/X9447E/X9447E00.HTM](http://www.fao.org/DOCREP/003/X9447E/X9447E00.HTM)
- Zhang, Bin (1997): "Total Factor Productivity of Grain Production in the Former Soviet Union". *Journal of Comparative Economics*, 24, pp. 202-209.

## ACRÓNIMOS

|               |                                                           |
|---------------|-----------------------------------------------------------|
| <b>ADF</b>    | Dickey-Fuller aumentado                                   |
| <b>AIC</b>    | criterio de información de Akaike                         |
| <b>ANCOVA</b> | análisis de covarianza                                    |
| <b>ANOVA</b>  | análisis de varianza                                      |
| <b>AO</b>     | observación atípica de tipo aditiva (additive outlier)    |
| <b>AR</b>     | autorregresivo                                            |
| <b>CADECA</b> | red de casas de cambio                                    |
| <b>CF</b>     | costo fijo                                                |
| <b>CRBP</b>   | crecimiento restringido por la balanza de pagos           |
| <b>CT</b>     | costo total                                               |
| <b>CTU</b>    | costo total unitario                                      |
| <b>CUC</b>    | peso cubano convertible                                   |
| <b>CUSUM</b>  | sumas acumuladas de residuos recursivos                   |
| <b>CUSUMQ</b> | sumas acumuladas de residuos recursivos al cuadrado       |
| <b>CV</b>     | costo variable                                            |
| <b>CVU</b>    | costo variable unitario                                   |
| <b>EE</b>     | error de estimación                                       |
| <b>EEM</b>    | error de especificación de la media                       |
| <b>EMCO</b>   | estimadores Mínimo Cuadráticos Ordinarios                 |
| <b>FAC</b>    | función de autocorrelación                                |
| <b>FACP</b>   | función de autocorrelación parcial                        |
| <b>FCC</b>    | función de correlación cruzada                            |
| <b>FIR</b>    | función impulso respuesta                                 |
| <b>FRM</b>    | función de regresión muestral                             |
| <b>FRP</b>    | función de regresión poblacional                          |
| <b>GC</b>     | gasto de consumo                                          |
| <b>IPC</b>    | Índice de Precios al Consumidor                           |
| <b>LS</b>     | observación atípica de tipo cambio de nivel (level shift) |
| <b>MA</b>     | medias móviles                                            |
| <b>MCE</b>    | modelo de corrección de error                             |
| <b>MCG</b>    | Mínimos Cuadrados Generalizados                           |
| <b>MCGF</b>   | Mínimos Cuadrados Generalizados Factibles                 |

|             |                                                                                                                                                                                            |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>MCO</b>  | Mínimos Cuadrados Ordinarios                                                                                                                                                               |
| <b>MCR</b>  | Mínimos Cuadrados Recursivos                                                                                                                                                               |
| <b>MELI</b> | Mejor estimador lineal insesgado                                                                                                                                                           |
| <b>MFP</b>  | Ministerio de Finanzas y Precios                                                                                                                                                           |
| <b>MLNR</b> | Modelo Lineal Normal de Regresión                                                                                                                                                          |
| <b>ONE</b>  | Oficina Nacional de Estadísticas de Cuba                                                                                                                                                   |
| <b>P</b>    | precios no regulados                                                                                                                                                                       |
| <b>PGD</b>  | proceso generador de los datos                                                                                                                                                             |
| <b>PIB</b>  | Producto Interno Bruto                                                                                                                                                                     |
| <b>PR</b>   | precios regulados                                                                                                                                                                          |
| <b>RV</b>   | estadístico de razón de verosimilitud                                                                                                                                                      |
| <b>SC</b>   | observación atípica de tipo cambio de pendiente (slope change)                                                                                                                             |
| <b>SCE</b>  | suma de cuadrados explicada (SSE en inglés); también se puede denotar como SCR que es la suma de cuadrados de la regresión (SSR en inglés)                                                 |
| <b>SIC</b>  | criterio de información de Schwartz                                                                                                                                                        |
| <b>SRC</b>  | suma de residuos al cuadrado (SSR en inglés); también se denota como SCE (SSE en inglés), por suma de cuadrados del error (que es justo “su contrario”, lo mismo en inglés que en español) |
| <b>SRRC</b> | suma de residuos restringidos al cuadrado                                                                                                                                                  |
| <b>STC</b>  | suma total de cuadrados                                                                                                                                                                    |
| <b>SVAR</b> | VAR estructural                                                                                                                                                                            |
| <b>TC</b>   | observación atípica de tipo cambio transitorio (transitory change)                                                                                                                         |
| <b>TCL</b>  | teorema central del límite                                                                                                                                                                 |
| <b>VAR</b>  | vectores autorregresivos                                                                                                                                                                   |
| <b>VMA</b>  | vector de medias móviles                                                                                                                                                                   |

