

**ECONOMETRÍA
DE SERIES DE
TIEMPO:**

**ENFOQUE DE
MONTE CARLO**

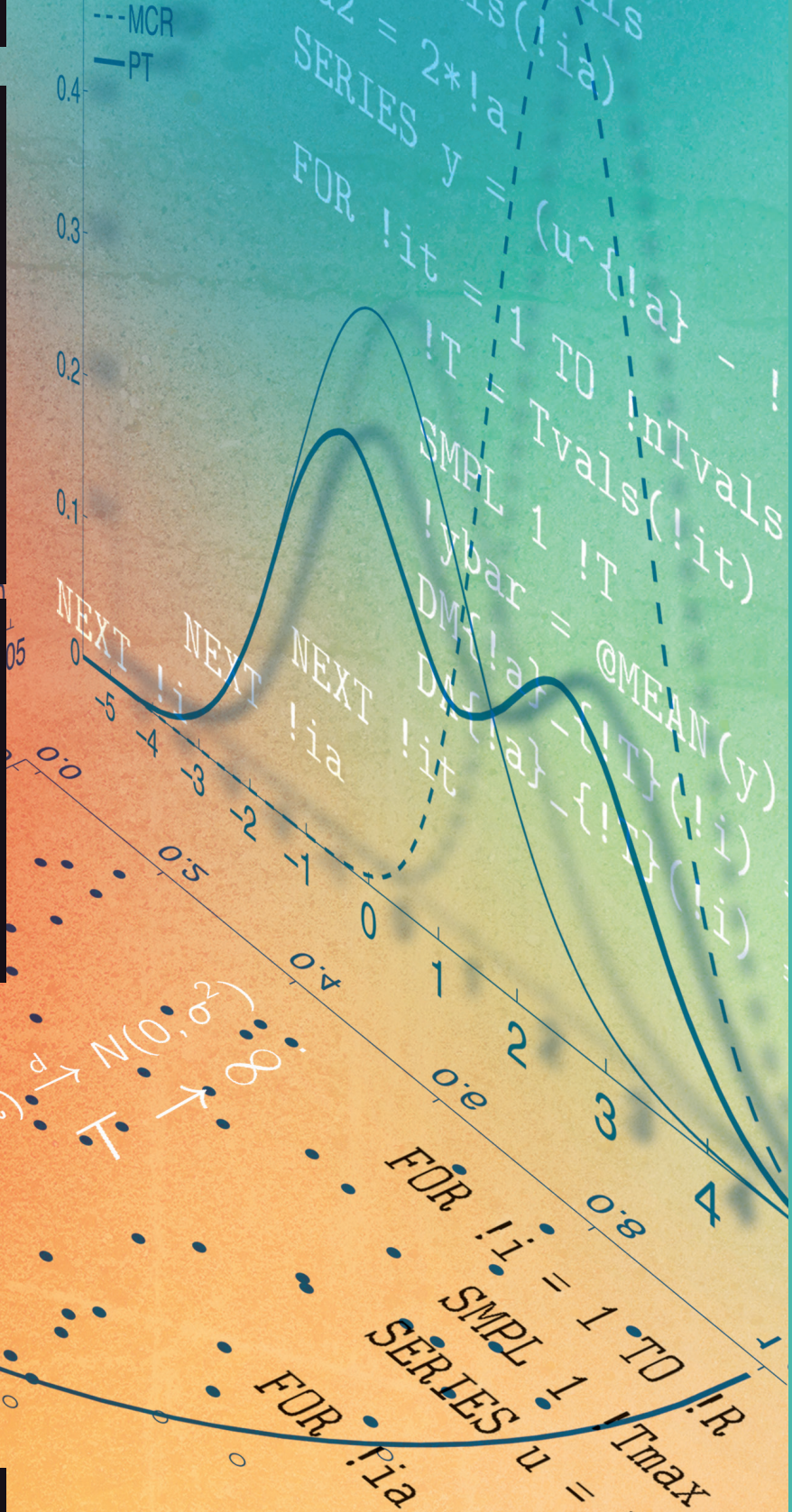
DIEGO WINKELRIED

85

Apuntes de Estudio



**UNIVERSIDAD
DEL PACÍFICO**



**ECONOMETRÍA
DE SERIES DE
TIEMPO:**

**ENFOQUE DE
MONTE CARLO**

DIEGO WINKELRIED

85

Apuntes de Estudio



**UNIVERSIDAD
DEL PACÍFICO**

© Diego Winkelried, 2017

De esta edición

© Universidad del Pacífico
Av. Salaverry 2020
Lima 11, Perú
www.up.edu.pe

ECONOMETRÍA DE SERIES DE TIEMPO: ENFOQUE DE MONTE CARLO

Diego Winkelried

1a. edición: marzo 2016, julio 2017

Diseño de la carátula: Icono Comunicadores

Tiraje: 300 ejemplares

ISBN: 978-9972-57-352-1

Hecho el Depósito Legal en la Biblioteca Nacional del Perú: 2017-08995

BUP

Winkelried, Diego.

Econometría de series de tiempo : enfoque de Monte Carlo / Diego Winkelried. -- 1a edición.

-- Lima : Universidad del Pacífico, 2017.

159 p. -- (Apuntes de estudio ; 85)

1. Econometría
 2. Análisis de series de tiempo
 3. Método de Monte Carlo
- I. Universidad del Pacífico (Lima)

330.015195 (SCDD)

Miembro de la Asociación Peruana de Editoriales Universitarias y de Escuelas Superiores (Ape-su) y miembro de la Asociación de Editoriales Universitarias de América Latina y el Caribe (Eulac).

La Universidad del Pacífico no se solidariza necesariamente con el contenido de los trabajos que publica. Prohibida la reproducción total o parcial de este texto por cualquier medio sin permiso de la Universidad del Pacífico.

Derechos reservados conforme a Ley.

Prólogo

La econometría de series de tiempo es un área de estudio fascinante a nivel teórico y su importancia práctica no puede exagerarse. Todo investigador que en el curso de sus labores deba lidiar con series de tiempo, datos que comprenden desde cotizaciones bursátiles hasta agregados macroeconómicos, requiere un conocimiento práctico de las técnicas, potencialmente sofisticadas, involucradas en la correcta inferencia estadística y proyección. Hoy por hoy, casi todo artículo de investigación financiera o macroeconómica que contenga una sección empírica, apela al uso de estos métodos estadísticos.

Desafortunadamente, los resultados más interesantes y relevantes en este campo son, en general, complejos y requieren del manejo de material matemático y estadístico que normalmente va más allá de lo cubierto a nivel de pregrado. Ante ello, es común que incluso los profesores más exigentes se vean obligados a hacer breves recuentos bibliográficos y, finalmente, a proveer de una suerte de “recetario”, casi dogmático, en la aplicación de los métodos de series de tiempo. El alumno exitoso suele aprender el recetario prolijamente, pero, inevitablemente, tanto su capacidad crítica como la interpretación que puede brindar a los fenómenos sobre los cuales centra su inferencia se ven mermadas ante la falta de un entendimiento más profundo de las propiedades de los estimadores que utiliza.

Un aliado poderoso del docente en el objetivo de comunicar eficazmente estos resultados complejos son las simulaciones de Monte Carlo. Esencialmente, un estudio de Monte Carlo descansa en dos principios de estadística ampliamente conocidos (la Ley Débil de Grandes Números y, usualmente, el Teorema del Límite Central) para llegar, de una manera muy gráfica y de fácil comunicación, a las mismas conclusiones a las que se llegaría de seguir el camino formal del análisis matemático y estadístico avanzado. Hoy en día, cuando el costo del poder computacional es bajo, implementar y ejecutar ejercicios de simulación es realmente accesible. En mi experiencia docente, he percibido cómo el ilustrar principios complejos con los resultados de simples ejercicios de Monte Carlo alimenta y fortalece el espíritu crítico de los alumnos, lo que repercute finalmente en un aprendizaje sólido de conocimientos que, como econometristas aplicados, van a necesitar para destacar en el momento de realizar un análisis empírico.

El propósito de este texto es, justamente, proveer un breve recuento teórico de algunos importantes resultados de la econometría de series de tiempo que interactúa con el diseño y, finalmente, con los resultados de simulaciones de Monte Carlo. El texto está dirigido a dos tipos de lector. Primero, al estudiante que busca una guía para consolidar y profundizar los conocimientos adquiridos sobre econometría de series de tiempo durante sus estudios,

para que luego sean utilizados rigurosamente en las aplicaciones empíricas de su interés. Segundo, al instructor o al docente que podría encontrar un valor didáctico en la forma como se escribió este texto, y en cómo se estructuran los diseños de simulación para responder preguntas específicas (muchas de las cuales aparecen en el salón de clase). Es por ello que los programas que implementan las simulaciones preparadas para este texto se encuentran disponibles para ser replicados en casa o en clase (véase el vínculo de descarga en la página 155, al final del texto). Asimismo, el lector también podrá acceder a versiones de alta resolución de las figuras que aparecerán a lo largo del texto, para analizarlas con más detalle o para utilizarlas en sus propias presentaciones. El uso de este material es libre, aunque estaré agradecido de que se reconozca el tiempo dedicado a prepararlo y difundirlo, citando adecuadamente este texto.

El contenido de este texto se basa en parte de las notas de clase que elaboré al dictar el curso Econometría II de la carrera de Economía de la Universidad del Pacífico, durante el año 2014, y el curso Tópicos de Econometría Avanzada de la Maestría de Economía de la Universidad del Pacífico, durante los años 2011 a 2015. También he utilizado, aunque en menor proporción, material preparado para los módulos de Econometría dictados como parte del Curso de Extensión Universitaria de Economía y del Curso de Actualización para Profesores de Economía, ofrecidos por el Banco Central de Reserva del Perú durante los años 2014 y 2015.

La exposición de los temas desarrollados en estas páginas está pensada para estudiantes de pregrado. Por ello, no se es particularmente riguroso en la derivación de algunos resultados asintóticos, y se evita hacer referencias explícitas a los pormenores de procesos de Weiner, a resultados (sumamente relevantes) como el Teorema del Límite Central Funcional, o al uso del operador de rezagos. Con los estudiantes de postgrado, muchas veces a su pesar, sí nos detenemos en estos tecnicismos. No obstante, considero que este grupo más avanzado podría igualmente beneficiarse al complementar su estudio con lo ofrecido en este texto.

Es recomendable que el lector esté familiarizado con varios temas, sobre todo de estadística, para que pueda seguir con fluidez la discusión acá presentada. Idealmente, al nivel de los alumnos que empiezan a llevar el curso de Econometría II de la Universidad del Pacífico. Estos temas incluyen: (1) manipulación básica de variables aleatorias, por ejemplo para el cálculo de esperanzas y covarianzas; (2) planteamiento y análisis de pruebas de hipótesis, así como las características de las distribuciones que aparecen al estudiar este tema (normal, chi-cuadrado, t de Student, F de Snedecor); (3) estimación de modelos de regresión por mínimos cuadrados; y (4) una primera noción de teoría asintótica, que incluya distinguir entre modos de convergencia (en probabilidad o en distribución), así como los postulados básicos de la Ley de Grandes Números y del Teorema del Límite Central.

En este punto, es conveniente aclarar que este texto no es, ni pretende ser, un libro de texto. Simplemente no está en la capacidad, ni es su objetivo, de brindar una discusión amplia y completa de los temas de series de tiempo que se discuten en estas páginas.

APUNTES DE ESTUDIO

La idea es que se utilice como material adicional, como el complemento a las notas y lecturas de clases, o quizá como el ejemplo o caso ilustrativo que ayude a entender mejor ciertos conceptos o relaciones. A nivel de pregrado, aunque no sea tan introductorio, no he encontrado una mejor referencia que Enders (2010), un libro muy bien escrito, con muchos ejemplos y aplicaciones. A un nivel más intermedio, Hayashi (2000) es una excelente referencia. Aunque se trate de un libro de econometría general, los capítulos donde se discuten tópicos de series de tiempo son bastante claros, con una notación muy amigable que facilita la lectura. En el postgrado, el monumental Hamilton (1994) sigue vigente. Sin embargo, mis favoritos son Maddala y Kim (1998) y Banerjee, Dolado, Galbraith y Hendry (1993), que, en mi opinión, son de lectura obligatoria para quien quiera aventurarse de lleno en el interesante mundo del análisis de series de tiempo no estacionarias.

Los programas que implementan las simulaciones son codificados íntegramente en Econometric Views (EViews), que es el *software* utilizado en la enseñanza de series de tiempo y otros temas econométricos en la Universidad del Pacífico. Estos programas fueron elaborados y probados en los laboratorios de esta universidad, bajo su licencia. La versión utilizada fue inicialmente la 8, aunque se ha hecho un esfuerzo por adecuar los códigos para que puedan ser ejecutados sin complicaciones con la versión 7 (y, muy probablemente, con versiones más antiguas). Es importante mencionar que estos programas se ponen a disposición del lector exclusivamente para su uso académico y docente. No está permitido el uso de estos códigos con fines comerciales. Asimismo, no asumo responsabilidad por (improbables) daños que la ejecución de estos programas pueda ocasionar.

La elección de este paquete informático no debería ser una limitante para entender la lógica detrás de las simulaciones y la estructura de su codificación. El lenguaje de programación de EViews es bastante intuitivo y muy comparable con los de otros paquetes estadísticos. No obstante, es altamente recomendable que el lector haya tenido algún tipo de acercamiento previo a este *software*. Una ventaja de EViews es que el uso del *software* y sus comandos de programación están cuidadosamente documentados en los archivos de ayuda del paquete. Muchas veces, consultar esta fuente es suficiente para entender un código. Adicionalmente, el libro de Castro y Rivas-Llosa (2007) provee una excelente y muy pedagógica introducción a la programación en EViews.

El lector notará a lo largo de este texto un estilo en la redacción de los programas. Este estilo facilita la lectura del código y proviene de una época en que los editores de texto eran monocromáticos y no se tabulaban ni corregían automáticamente. En particular los comandos, que son palabras “prohibidas” o “reservadas”, son escritos en MAYÚSCULAS mientras que los objetos son escritos en minúsculas, siempre con caracteres tipográficos (se recomienda, en general, que los programas sean leídos con una fuente de ancho fijo, *monospaced font*). Por ejemplo, EQUATION eq.LS y C x declara una ecuación de nombre eq (minúsculas) con el comando EQUATION (mayúsculas), en donde se ejecuta una estimación de mínimos cuadrados con la instrucción LS (mayúsculas), al regresar la serie y (minúsculas) sobre un intercepto C (mayúsculas, por ser un nombre

reservado por EViews para una serie llena de unos) y una variable x (minúsculas). Esta convención es puramente una cuestión de estilo. El lenguaje de programación de EViews no es sensible a si los caracteres son escritos en mayúsculas o minúsculas. EViews interpretará EQUATION EQ.LS Y C X y equation eq.ls y c x exactamente de la misma manera.

Los programas son autocontenidos y cada ejercicio de simulación se asocia con un único programa. En todos los casos, los programas crean en sus primeras líneas un *workfile*, completamente nuevo y sin nombre (UNTITLED, que es el nombre por defecto), sobre el cual realizar los cálculos. Se recomienda, además, ejecutar estos programas en modo Quiet (el modo alternativo y, muchas veces, por defecto es Verbose, que es considerablemente más lento). Como resultado, al término de las simulaciones se generan gráficos y cuadros muy similares a los reportados en este texto. Dado que las simulaciones tienen, por definición, un componente aleatorio, y por posibles errores de redondeo, es poco probable que se obtengan exactamente los mismos resultados que se presentan luego. No obstante, sí se puede esperar que sean lo suficientemente parecidos como para que las diferencias no tengan ninguna consecuencia práctica.

Las simulaciones que presentaremos utilizan un gran número de repeticiones, usualmente 100,000. La razón es poder presentar al lector resultados con un mínimo, casi imperceptible, error de simulación. El costo es que la ejecución de algunos de estos programas en una computadora con estándares promedio (digamos, un procesador de 2 GHz y una memoria RAM de 2 GB) puede tardar algunas horas. Sin embargo, y es una práctica que sigo en mis clases, la ejecución de estos mismos programas con un número de, digamos, 2,000 repeticiones es una cuestión de uno o dos minutos. Los resultados, ciertamente, se verán contaminados con algo de error de simulación, pero los patrones serán lo suficientemente claros como para comunicar el mensaje último del ejercicio de simulación.

El texto está organizado de la siguiente manera. El capítulo I describe los fundamentos detrás del uso de simulaciones para el análisis estadístico, enfoque conocido como Integración de Monte Carlo. El capítulo II discute los conceptos de estacionariedad y ergodicidad, y sus implicancias para la estimación e inferencia con series de tiempo. El capítulo III extiende el análisis a un contexto en donde las series pueden ser no estacionarias. El capítulo IV se centra en el análisis de cointegración. Al final del texto, se presentan un listado y una breve descripción de los programas de EViews utilizados. Son en total 25 programas (brevemente documentados en la página 147), 23 de los cuales abordan algún problema estadístico de interés (los 2 restantes ilustran el método de Integración de Monte Carlo). Si bien es cierto que los estudios de Monte Carlo responden preguntas específicas, válidas estrictamente bajo los supuestos del diseño de la simulación, considero que los programas provistos cubren muchas situaciones de interés, fácilmente adaptables. Las simulaciones se utilizan para explorar distribuciones muestrales, calcular sesgos y valores críticos de algunas pruebas de hipótesis, estudiar la tasa de cobertura de intervalos de confianza, calcular la potencia estadística de contrastes de hipótesis, entre otros. Se espera ofrecer en versiones futuras los códigos de las simulaciones en otros paquetes como Matlab.

APUNTES DE ESTUDIO

Se han dejado de lado algunos temas que podrían ser de interés, y que posiblemente sean incorporados en versiones futuras de este texto, de los cuales enfatizo tres. Primero, el estudio de las distorsiones generadas por la presencia de quiebres estructurales en el contraste de raíces unitarias, en especial la reducción de la potencia de ciertas pruebas de hipótesis. Ello ocurre básicamente porque, al ignorar un cambio estructural, el modelo que describe los datos bajo la hipótesis alternativa se encuentra mal especificado. El tema se aborda acá de una manera indirecta, ya que son varios los casos y ejemplos en donde se ilustra que el no rechazo de una hipótesis nula es el resultado de una hipótesis alternativa pobre, más que de una hipótesis nula satisfactoria. Segundo, el análisis de cointegración *à la* Johansen se trata muy marginalmente. En general, el enfoque del texto es hacia el análisis de una sola ecuación de regresión, ya que los resultados más interesantes se generalizan muy fácilmente cuando se pasa a un mundo de sistemas de ecuaciones (digamos, vectores autorregresivos). De hecho, los resultados para sistemas son repeticiones de los resultados para ecuaciones individuales. Finalmente, se excluye el análisis de modelos dinámicos no lineales, por ejemplo modelos de la familia ARCH. Ello responde a que el énfasis del texto es comprender los cambios fundamentales para la inferencia de transitar de un mundo estacionario a uno no estacionario, y en general no es particularmente informativo lo que los modelos no lineales revelen sobre este punto en particular.

Deseo expresar mi agradecimiento al Centro de Investigación de la Universidad del Pacífico (representado por su directora durante 2014, Cynthia Sanborn) por financiar e impulsar la publicación y difusión de este texto. Asimismo, agradezco al Departamento Académico de Economía (representado por su jefe durante 2014, Roberto Urrunaga) y a la Dirección de la Maestría en Economía de esta misma universidad (representada por sus directores durante el periodo 2011 a 2015, Eduardo Morón y Juan Mendoza), por compartir mi pasión por la econometría y acompañarme en la ambición de formar futuras generaciones de jóvenes economistas con una sólida orientación econométrica. César Urquiza es un miembro notable de este grupo de jóvenes, y su aporte a la redacción de muchos de los códigos que dan contenido a este texto es especialmente reconocido. Quisiera, además, expresar mi agradecimiento a dos evaluadores anónimos, cuyos acertados comentarios ayudaron a mejorar la exposición de este texto, así como a María Elena Romero del Fondo Editorial de la Universidad del Pacífico por su solícita labor editorial. Agradezco también a Javier Zúñiga por haberme provisto desinteresadamente la plantilla de LaTeX sobre la que este texto fue procesado. Cabe enfatizar que el contenido de este texto, así como todos sus errores u omisiones, son de mi entera responsabilidad. Finalmente, este texto está dedicado a mis alumnos, pasados y presentes; a la larga, fue inspirado por ustedes y escrito pensando en ustedes.

Diego Winkelried Quezada

Lima, noviembre de 2015

Índice

Prólogo	iii
Índice	viii
I Integración de Monte Carlo	1
I.1 Integración de Monte Carlo	1
I.2 Aproximando π	4
I.3 Muestreando distribuciones conocidas	8
I.4 Eficiencia del promedio y la mediana muestrales	12
I.5 Inferencia en modelos de regresión	15
I.6 Leyes asintóticas importantes	24
II Series de tiempo estacionarias	29
II.1 Generando dependencias: procesos ARMA	30
II.2 Leyes asintóticas importantes, revisadas	34
II.3 Inferencia en la práctica: estimación de la varianza de largo plazo	43
II.4 Estimación de modelos ARMA	47
II.5 Contrastes de autocorrelación residual	52
II.6 Criterios de información y selección de modelos	60
III No estacionariedad	69
III.1 Modelo de tendencia lineal	69
III.2 Procesos integrados y no estacionariedad	76
III.3 Efectos sobre la inferencia	81
III.4 Removiendo tendencias	85
III.5 Pruebas de raíz unitaria	89
III.6 Potencia y consistencia de las pruebas de Dickey y Fuller	97
III.7 Incrementando la potencia con una prueba LM	101
III.8 ¿Retorno a la normalidad?	104
III.9 Proyecciones y raíces unitarias	106
III.10 Parámetros fastidiosos y posibles soluciones	111
III.11 Estimación e inferencia con series no estacionarias	124
IV Cointegración	131
IV.1 Representaciones	131
IV.2 Ausencia de cointegración y regresiones espurias	134
IV.3 Estimación e inferencia	138
IV.4 Inferencia sobre el multiplicador de largo plazo	142
Resumen de programas	147
Bibliografía	157

Lista de gráficos

I.1	Aproximación de π	6
I.2	Colas anchas y eficiencia relativa del promedio y la mediana	14
I.3	Desempeño de los estimadores MCO, MCR y PT	22
I.4	Distribuciones muestrales de los estimadores MCO, MCR y PT	23
I.5	Ley de Grandes Números y Teorema del Límite Central	26
II.1	Leyes asintóticas para procesos dependientes	40
II.2	Consistencia de la primera autocorrelación muestral	42
II.3	Sesgo del estimador MCO de ϕ en el modelo $y_t = \phi y_{t-1} + \varepsilon_t$	49
II.4	Consistencia y normalidad asintótica de estimadores de mínimos cuadrados	53
II.5	Correlograma de residuos	55
II.6	Prueba de autocorrelación residual de Breusch-Godfrey	58
II.7	Potencia de las pruebas Q y LM de ausencia de correlación serial	59
III.1	Superconsistencia	73
III.2	Equivalencia asintótica de estimadores de ϕ	75
III.3	Comportamiento asintótico de promedios muestrales de un paseo aleatorio	83
III.4	Distribuciones empíricas de ratios t para la pendiente de la tendencia	89
III.5	Distribuciones de Dickey-Fuller	93
III.6	Probabilidad de rechazo en las pruebas de Dickey y Fuller	99
III.7	Probabilidad de rechazo en las pruebas DF y LM	103
III.8	Funciones de distribución empíricas de τ_ϕ para un proceso con tendencia	105
III.9	Superconsistencia de ρ en la regresión básica de Dickey y Fuller	113
III.10	Parámetros fastidiosos y posibles soluciones	114
III.11	Prueba de Phillips y Perron	116
III.12	Prueba de Dickey y Fuller Aumentada	117
III.13	Superconsistencia en la regresión de Dickey y Fuller Aumentada	126
III.14	Equivalencia asintótica en la regresión de Dickey y Fuller Aumentada	127
IV.1	Función de distribución empírica de estadísticos de regresión	136
IV.2	Superconsistencia del MLP en el procedimiento de Engle y Granger	140
IV.3	Equivalencias asintóticas en el procedimiento de Engle y Granger	141
IV.4	Inferencia sobre el multiplicador de largo plazo	145

Lista de cuadros

I.1	Aproximación de cantidades poblacionales para distribuciones conocidas	10
II.1	Probabilidad de cobertura con el estimador HAC de Newey-West	46
II.2	Desempeño de criterios de información	65
III.1	Convergencia de promedios de series estacionarias y no estacionarias	84
III.2	Contrastes de raíz unitaria de Dickey y Fuller	95
III.3	Desempeño predictivo de tres estrategias de proyección	110
III.4	Probabilidad de rechazo real (probabilidad nominal de 5 %)	122

I Integración de Monte Carlo

El punto de partida de todo análisis estadístico es la formulación de un **modelo poblacional**, muchas veces denominado **proceso generador de datos**, que establece las relaciones bajo estudio entre un grupo de variables aleatorias. El fin último de los ejercicios de estimación e inferencia en el análisis empírico es revelar o descubrir aspectos relevantes del modelo poblacional, a partir de **muestras** (típicamente de tamaño finito) que constituyen el conjunto de información a disposición del investigador. Así, es imperativo evaluar tanto las propiedades de los estimadores utilizados como las posibilidades de inferencia sobre las cantidades poblacionales de interés. Para ello, el investigador debe tener en cuenta tanto las características del modelo poblacional como las reglas del muestreo.

Usualmente, es prohibitivamente complicado caracterizar exacta o analíticamente el comportamiento de los estadísticos muestrales que se utilizan en la práctica. El uso de técnicas de simulación responde a la promesa de que pueden aproximarse, con niveles de precisión arbitrariamente altos, las propiedades de estos estadísticos (en concreto, sus funciones de distribución) mediante métodos de Monte Carlo. Bajo este enfoque, se le da al investigador la oportunidad de muestrear directamente, bajo reglas conocidas y controladas, de un proceso generador de datos hipotetizado para así calcular los estadísticos de interés en esta muestra ficticia. La repetición de este procedimiento (un gran número de veces) aproxima numéricamente las propiedades deseadas de los estadísticos muestrales.

Esta es la esencia del método de Integración de Monte Carlo que se discute en la sección 1.1. El funcionamiento del método es ilustrado con dos aplicaciones, en las secciones 1.2 y 1.3. Finalmente, las secciones 1.4 y 1.6 estudian las propiedades de estadísticos muestrales, en especial de promedios, mediante simulaciones.

1.1 Integración de Monte Carlo

El resultado fundamental detrás de los métodos de simulación es la **Ley Débil de Grandes Números**, a veces denominada el **Teorema de Khinchine**. Considere una variable aleatoria ξ cuya función de distribución es $f_{\xi}(\cdot)$, y tome independientemente (aleatoriamente) R realizaciones de esta función de distribución, $\xi_{(i)} \sim f_{\xi}(\xi)$, para así conseguir una muestra o colección de valores $\{\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(R)}\}$. Considere, además, una función $g(\cdot)$ para formar otra variable aleatoria $g(\xi)$. La evaluación de esta función para toda realización de ξ da como resultado una nueva colección de valores, $\{g(\xi_{(1)}), g(\xi_{(2)}), \dots, g(\xi_{(R)})\}$. La Ley Débil de

Grandes Números establece que el promedio muestral de los R valores de $g(\xi_{(i)})$ se aproxima (en estricto, converge en probabilidad) a la esperanza poblacional de $g(\xi)$ conforme R se incrementa. Explícitamente,

$$\frac{1}{R} \sum_{i=1}^R g(\xi_{(i)}) \xrightarrow{p} \mathbb{E}(g(\xi)) \quad \text{para } R \rightarrow \infty. \quad (1.1)$$

Esta ley indica que la esperanza de una variable aleatoria puede aproximarse “muestreando” repetidas veces de $f_{\xi}(\cdot)$ y tomando promedios a las observaciones de este muestreo. Puesto de otra forma, **promedios muestrales** se acercan, y a la larga convergen, a **esperanzas poblacionales**¹.

Este resultado es de suma importancia, ya que, básicamente, toda cantidad poblacional de interés que caracteriza el comportamiento aleatorio de ξ puede ser expresada como una esperanza y, por tanto, puede ser cuantificada a través de simulaciones.

Un primer grupo de resultados se relaciona con los **momentos** de ξ . El p -ésimo momento de ξ , donde p es un número entero positivo, se define como $\mathbb{E}(\xi^p)$, por lo que una aplicación directa de (1.1), con $g(\xi) = \xi^p$, conlleva

$$\frac{1}{R} \sum_{i=1}^R \xi_{(i)}^p \xrightarrow{p} \mathbb{E}(\xi^p).$$

El primer momento, $p = 1$, corresponde a la media poblacional y es aproximado por

$$\bar{\xi}_R = \frac{1}{R} \sum_{i=1}^R \xi_{(i)},$$

el promedio simple de $\{\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(R)}\}$.

El segundo momento se relaciona con el cálculo de la **varianza**, quizá la medida de dispersión más común en estadística. En concreto, se tiene que $\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2$. El primer término es aproximado por el promedio muestral de $\xi_{(i)}^2$, mientras que el segundo, por el cuadrado de $\bar{\xi}_R$. De este modo, la Ley Débil de Grandes Números garantiza que la varianza muestral converge a la poblacional,

$$V_R = \frac{1}{R} \sum_{i=1}^R (\xi_{(i)} - \bar{\xi}_R)^2 = \frac{1}{R} \sum_{i=1}^R \xi_{(i)}^2 - \bar{\xi}_R^2 \xrightarrow{p} \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2 \equiv \mathbb{V}(\xi).$$

¹ El calificativo de **débil** se refiere al hecho de que el Teorema de Khinchine establece un **límite en probabilidad**, que es una forma relativamente general (o “débil”) de convergencia de variables aleatorias. Existen leyes fuertes de grandes números (por ejemplo, la Ley de Kolmogorov) que involucran tipos de convergencia más exigentes, como convergencia casi segura (*almost surely*). No obstante, para fines del presente texto la distinción entre leyes débiles o fuertes no es importante, ya que el interés se centra en situaciones o muestreos en donde **ambas** se cumplen. Por ello y por motivos pedagógicos, únicamente se discuten las implicancias de la ley débil.

² Esta es una aplicación del **Teorema de Slutsky** que, en términos simples, establece que los límites probabilísticos pueden ser manipulados como límites ordinarios o determinísticos: si $\text{plim } a_R = a$, entonces para cualquier función $g(\cdot)$ bien comportada, $\text{plim } g(a_R) = g(a)$. Así, el límite probabilístico del cuadrado de $\bar{\xi}_R$ es igual al cuadrado del límite probabilístico de $\bar{\xi}_R$.

Note que la varianza muestral, normalmente, divide la suma de desvíos cuadráticos respecto al promedio entre $R - 1$ en lugar de R . Por supuesto, esta distinción no es relevante cuando R es lo suficientemente grande.

Otra aplicación útil del resultado en (1.1) es el **cálculo de probabilidades**, que involucra una elección particular de la función $g(\cdot)$. Suponga que el soporte de ξ (es decir, el conjunto de todos los posibles valores que esta variable aleatoria toma) se denota como Ξ . Por definición, la esperanza de $g(\xi)$ es un promedio ponderado de todos los posibles valores de esta variable, donde las ponderaciones vienen dadas por la probabilidad de ocurrencia de cada valor. Para un soporte continuo³,

$$\mathbb{E}(g(\xi)) = \int_{\Xi} g(t) f_{\xi}(t) dt. \quad (1.2)$$

Considere ahora que el soporte de Ξ se divide en dos regiones complementarias, es decir mutuamente excluyentes $\{\Xi_A \cap \Xi_B\} = \{\emptyset\}$ y exhaustivas $\{\Xi_A \cup \Xi_B\} = \{\Xi\}$. Sea, además, $\mathbf{1}\{x\}$ una **función indicador** tal que $\mathbf{1}\{x\} = 1$ si x es verdadero y $\mathbf{1}\{x\} = 0$ si x es falso. Se tiene que

$$\mathbb{E}(\mathbf{1}\{\xi \in \Xi_A\}) = \int_{\Xi_A} 1 \cdot f_{\xi}(t) dt + \int_{\Xi_B} 0 \cdot f_{\xi}(t) dt = \int_{\Xi_A} f_{\xi}(t) dt = \Pr(\xi \in \Xi_A).$$

Es decir, la esperanza de $\mathbf{1}\{x\}$ es la probabilidad de que x sea verdadero. Así, tomando la colección $\{\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(R)}\}$ pueden identificarse aquellas realizaciones que satisfagan $\xi_{(i)} \in \Xi_A$, y la Ley Débil de Grandes Números indica que la **frecuencia relativa** de este evento, que es el promedio muestral del indicador $\mathbf{1}(\xi_{(i)} \in \Xi_A)$, converge a la probabilidad deseada

$$\frac{1}{R} \sum_{i=1}^R \mathbf{1}\{\xi_{(i)} \in \Xi_A\} \xrightarrow{p} \Pr(\xi \in \Xi_A). \quad (1.3)$$

Usualmente, la **función de distribución acumulada** de ξ describe completamente el comportamiento aleatorio de esta variable. Por definición, $F_{\xi}(A) = \Pr(\xi \leq A)$. Dado que $F_{\xi}(\cdot)$ se define a partir de una probabilidad, puede ser aproximada por una frecuencia relativa. Esta frecuencia relativa se conoce como **función de distribución empírica** (FDE),

$$\text{FDE}_{\xi}(A) = \frac{\text{Número de veces en las que } \xi_{(i)} \leq A}{R} = \frac{1}{R} \sum_{i=1}^R \mathbf{1}(\xi_{(i)} \leq A) \xrightarrow{p} F_{\xi}(A).$$

Del mismo modo, en ocasiones el interés se centra en la **función de densidad** de ξ , $f_{\xi}(\cdot)$, que es igual a la primera derivada de la función de distribución acumulada. Existen varias formas de escribir la relación entre $f_{\xi}(\cdot)$ y $F_{\xi}(\cdot)$; la más conveniente para nuestros

³ Las esperanzas pueden representarse como integrales. De ahí el nombre de **Integración** de Monte Carlo que recibe el método de aproximar esperanzas mediante simulaciones.

propósitos es la variación de la probabilidad acumulada en un intervalo infinitesimalmente pequeño centrado en ξ . A saber,

$$f_{\xi}(\xi) = \lim_{h \rightarrow 0} \frac{F_{\xi}(\xi + \frac{1}{2}h) - F_{\xi}(\xi - \frac{1}{2}h)}{h}.$$

Al reemplazar $F_{\xi}(\cdot)$ por su aproximación $FDE_{\xi}(\cdot)$ y para un valor de h finito y pequeño,

$$\frac{FDE_{\xi}(\xi + \frac{1}{2}h) - FDE_{\xi}(\xi - \frac{1}{2}h)}{h} = \frac{\text{Número de veces en las que } \xi - \frac{1}{2}h \leq \xi_{(i)} \leq \xi + \frac{1}{2}h}{Rh} \xrightarrow{p} f_{\xi}(\xi),$$

conforme $R \rightarrow \infty$ y $h \rightarrow 0$. El lector familiarizado con estadísticas descriptivas habrá notado que la definición anterior corresponde a la frecuencia relativa utilizada en el cálculo de un **histograma**, dividida entre h . Ciertamente, en el contexto de simulaciones con un R lo suficientemente grande y un ancho de intervalo h lo suficientemente pequeño (una práctica común es $h \simeq R^{-1/3}$), el histograma de $\{\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(R)}\}$ aproxima razonablemente bien a h veces la función de densidad⁴.

1.2 Aproximando π

Un ejemplo clásico sobre el funcionamiento del método de Integración de Monte Carlo es la aproximación del número $\pi \simeq 3.1416$ a partir del área de un círculo. El conjunto

$$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

define a un círculo en \mathbb{R}^2 , centrado en el origen y con radio igual a $r = 1$. Se sabe que el área de este círculo es igual a $\pi r^2 = \pi$.

Por simetría, si únicamente se considera la región del círculo correspondiente al primer cuadrante, es decir los pares (x, y) que pertenecen al conjunto

$$A(x, y) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1, x \geq 0 \wedge y \geq 0\},$$

se tiene una región cuya área es igual a $\frac{1}{4}\pi$. El conjunto $A(x, y)$ puede ser representado gráficamente en el plano (x, y) , como el área por debajo la función implícita $x^2 + y^2 = 1$,

⁴ Una mejor manera de aproximar o estimar la densidad de una variable aleatoria es a través del método no paramétrico "*kernel density*". Este método guarda muchas similitudes con la lógica detrás de la construcción de un histograma, pero produce una versión suavizada (el histograma es generalmente un gráfico de barras discretas), con mejores propiedades de convergencia para variables aleatorias continuas. Escapa al alcance de este texto profundizar sobre este tema. Sin embargo, es bueno mencionar que dado que versiones automáticas del método se encuentran implementadas en paquetes como EViews, muchos de los resultados en secciones posteriores se presentan como densidades kernel, que son, dados los números elevados de repeticiones R que utilizamos, cualitativamente similares a los que se obtendrían con un histograma.

APUNTES DE ESTUDIO

a la derecha de $x = 0$, y por encima de $y = 0$. Puesto de otro modo, al expresar y como una función explícita de x (o viceversa), se consigue la siguiente igualdad:

$$\int_0^1 4\sqrt{1-t^2} dt = \pi.$$

La integral al lado izquierdo de la identidad anterior puede ser aproximada mediante el método de Monte Carlo. En particular, note que si x tiene como soporte al intervalo $[0, 1]$ (los límites de integración) y como función de distribución $f_x(t) = 1$, entonces la integral en cuestión corresponde al valor esperado de la variable aleatoria $g(x) = 4\sqrt{1-x^2}$. Estas propiedades son satisfechas si $x \sim \text{Uniforme}(0, 1)$. Así pues, tras muestrear R valores $\{x_{(1)}, x_{(2)}, \dots, x_{(R)}\}$ de una distribución uniforme y calcular $\{g(x_{(1)}), g(x_{(2)}), \dots, g(x_{(R)})\}$, la Ley Débil de Grandes Números indica que el promedio muestral de los R valores de $g(x_{(i)})$ se aproximará a π conforme R se incrementa.

En el gráfico I.1(a) se presenta el valor de este promedio como función de R , el número de observaciones promediadas, para dos muestreos distintos e independientes. Se aprecia cómo opera la Ley Débil de Grandes Números. Inicialmente, con valores reducidos de R , la calidad de la aproximación es baja. Conforme R se incrementa, el promedio se va acercando a π . Note que la convergencia hacia π no es necesariamente monotónica, sino que es más bien irregular. El comportamiento errático del promedio muestral como función de R es simplemente una manifestación de la naturaleza aleatoria de x . De hecho, la Ley Débil de Grandes Números predice que el promedio de $\{g(x_{(1)}), g(x_{(2)}), \dots, g(x_{(R)})\}$ podría dejar de pensarse como una variable aleatoria únicamente en el límite, cuando $R \rightarrow \infty$.

Esta simulación puede ser implementada fácilmente en EViews, y se encuentra disponible en el programa `Prog11_pi.prg`. Operando sobre un *workfile* con R observaciones, se genera una serie que contiene R números aleatorios independientes provenientes de una distribución uniforme, a través del comando

```
SERIES x = RND.
```

Luego, se genera una nueva serie con la transformación de estos datos $g(x)$, que será promediada con el propósito de conseguir la aproximación deseada, a través del comando

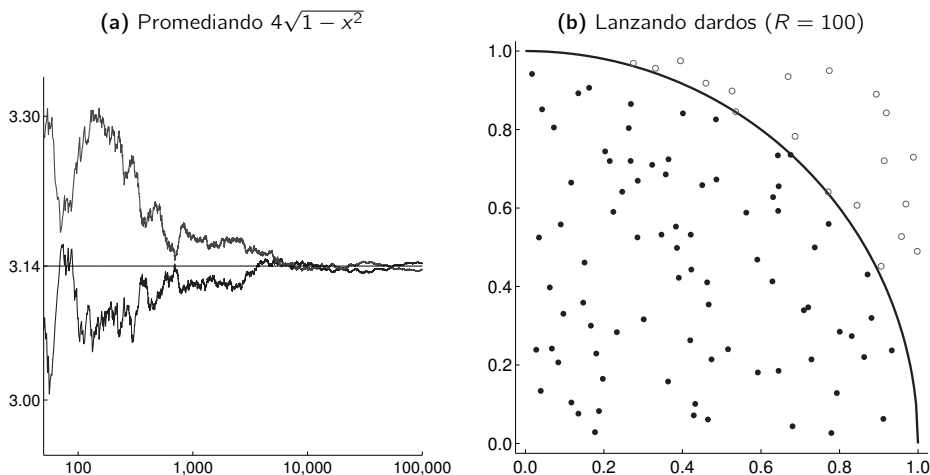
```
SERIES gx = 4*@SQRT(1 - x^2).
```

El gráfico I.1(a) presenta los promedios acumulados de gx : el valor de la i -ésima observación corresponde a la suma de las primeras i observaciones en gx , dividida entre i . El comando

```
SERIES prom = @CUMSUM(gx)/@CUMSUM(1)
```

genera una serie `prom` que contiene, precisamente, estos promedios acumulados. Al aplicar el comando `@CUMSUM(.)` a un objeto `serie`, se obtiene como resultado otro objeto `serie`

Gráfico I.1 Aproximación de π



Notas: (Prog11_pi.prg) En el panel (a), el eje horizontal muestra el valor de R (en escala logarítmica), mientras que el eje vertical muestra el promedio muestral de $g(x) = 4\sqrt{1-x^2}$, donde $x \sim \text{Uniforme}(0, 1)$. Se presentan los resultados de dos muestreos independientes. La línea horizontal corresponde al valor de π . En el panel (b), la línea continua es la curva $x^2 + y^2 = 1$ en el plano (x, y) . Los puntos sólidos (81 de 100) representan pares (x, y) que satisfacen $x^2 + y^2 \leq 1$, mientras que los puntos vacíos (19 de 100) corresponden a pares que satisfacen $x^2 + y^2 > 1$. El área cubierta por los puntos sólidos converge a $\frac{1}{4}\pi$ conforme $R \rightarrow \infty$.

con las sumas acumuladas de la serie original. Más formalmente, ante la instrucción $z = @CUMSUM(w)$, EViews procede recursivamente partiendo de $z(1) = w(1)$ y, para valores de i mayores de 1, desarrollando $z(i) = z(i - 1) + w(i)$, hasta la última observación disponible. Así, el numerador de $\text{prom}, @CUMSUM(gx)$, reditúa una serie con las sumas parciales de los elementos de gx , mientras que el denominador, $@CUMSUM(1)$, es la suma acumulada de una serie llena de unos, es decir una serie cuya i -ésima observación es igual a i . Se puede obtener un gráfico similar al gráfico I.1(a) al visualizar la serie prom , por ejemplo con el comando prom.LINE .

Una manera alternativa de aplicar el método de Integración de Monte Carlo en el cálculo de π , descansa en la aproximación de una probabilidad a partir de una frecuencia relativa (véase la ecuación I.3), y se ilustra en el gráfico I.1(b). El área total del gráfico es igual a uno, y corresponde al área del cuadrado formado por los vértices $(0, 0)$, $(0, 1)$, $(1, 1)$ y $(1, 0)$. Suponga ahora que se “lanzan dardos” a este cuadrado, y que la probabilidad de que el dardo caiga sobre un punto (x_0, y_0) sobre el cuadrado es igual a la probabilidad de que caiga sobre cualquier otro punto (x_1, y_1) . Así, si se lanzara un número muy grande de dardos (en estricto, infinitos dardos), todos los puntos del cuadrado serían cubiertos.

El cuadrado, a su vez, puede ser separado en dos regiones. La primera corresponde al conjunto $A(x, y)$ definido previamente; en el gráfico I.1(b), los dardos que caen dentro

APUNTES DE ESTUDIO

de este conjunto se marcan con puntos sólidos. La segunda región es simplemente el complemento del conjunto $A(x, y)$; los dardos que caen en este conjunto se marcan con puntos vacíos. Luego, en el escenario donde se lanzan infinitos dardos, puede deducirse que la proporción de puntos sólidos respecto al total de puntos será exactamente igual al área del conjunto $A(x, y)$, que sabemos es igual a $\frac{1}{4}\pi \simeq 0.7854$.

En el gráfico I.1(b) se lanzaron $R = 100$ dardos, de los cuales 81 cayeron dentro del conjunto $A(x, y)$. La proporción de puntos oscuros respecto al total de puntos es 0.81, que es una aproximación bastante cruda de 0.7854. Sin embargo, en línea con lo establecido por la Ley Débil de Grandes Números, las aproximaciones mejoran conforme R se incrementa: en una simulación con $R = 1,000$ se obtiene una proporción igual a 0.7950, mientras que con $R = 10,000$ la aproximación asciende a 0.7827.

El programa `Prog11_pi.prg` también contiene la implementación del “lanzamiento de dardos”. El hecho de que el dardo caiga en cualquier punto del cuadrado con la misma probabilidad, implica que tanto x como y son variables uniforme e independientemente distribuidas, cada una con soporte $[0, 1]$. El comando `SERIES y = RND` genera una serie de números uniformemente distribuidos e independientes de los contenidos en la serie x , generada previamente. Por su parte, el comando

```
SERIES ind = (x^2 + y^2 <= 1)
```

genera una variable binaria (*dummy*) que toma únicamente los valores de 0 o 1. Esta sintaxis explota las capacidades de EViews para manipular operadores lógicos, como la función indicador `1{·}`. Así, `ind` es igual a uno si el par de observaciones (x, y) pertenece al conjunto $A(x, y)$, y es igual a cero de otro modo. Las observaciones con `ind = 1` corresponden a los puntos oscuros del gráfico I.1(b), mientras que las observaciones con `ind = 0` corresponden a los puntos claros. De esta manera, el promedio de `ind` es la proporción que aproxima a $\frac{1}{4}\pi$. Una aproximación de π , entonces, puede ser recuperada a través del comando

```
!pi_aprox = 4*@MEAN(ind)
```

que genera un escalar temporal o volátil (es decir, que se elimina automáticamente tras la ejecución del programa) de nombre `!pi_aprox` con 4 veces el promedio muestral de `ind`. Si se desea que este escalar se preserve en el *workfile* luego de ejecutado el programa, debe optarse por el comando `SCALAR pi_aprox = 4*@MEAN(ind)` (note que el uso del símbolo `!` ya no es permitido). El programa `Prog11_pi.prg` varía el tamaño muestral sobre el que se toma este promedio (el comando relevante es `SMPL 1 !obs`, donde `!obs` es el número de observaciones), consiguiendo así distintas aproximaciones para distintos valores de R .

Para finalizar, es importante mencionar que no es necesario establecer resultados adicionales para generalizar la Ley Débil de Grandes Números a contextos multivariados. Nótese que bajo el segundo enfoque, donde el promedio muestral de `ind` busca aproximar a

la esperanza $\mathbb{E}(\text{ind})$, el método de Monte Carlo efectivamente aproxima una integral doble. Conviene verificar los detalles. Primero, el valor esperado del indicador de pertenencia al conjunto $A(x, y)$ viene dado por

$$\mathbb{E}(\text{ind}) = \int_0^1 \int_0^1 \mathbf{1}\{t^2 + s^2 \leq 1\} f_{xy}(t, s) dt ds = \int_0^1 \int_0^1 \mathbf{1}\{t^2 + s^2 \leq 1\} f_x(t) f_y(s) dt ds,$$

donde $f_{xy}(\cdot, \cdot)$ es la **función de distribución conjunta** de las variables aleatorias x e y . Dado que x e y son independientes, se cumple que la función de distribución conjunta es igual al producto de las funciones de distribución marginales, $f_{xy}(t, s) = f_x(t) f_y(s)$. Más aún, al ser ambas variables uniformemente distribuidas en el soporte $[0, 1]$, $f_x(t) = f_y(s) = 1$. Así, se tiene que

$$\mathbb{E}(\text{ind}) = \int_0^1 \int_0^1 \mathbf{1}\{t^2 + s^2 \leq 1\} dt ds = \int_0^1 \left(\int_0^1 \mathbf{1}\{t^2 + s^2 \leq 1\} ds \right) dt.$$

La integral interna puede descomponerse como la suma de dos integrales. En la primera, el dominio de integración son los valores de s entre 0 y $s(t) \equiv \sqrt{1-t^2}$, región en donde el par (t, s) pertenece al conjunto $A(\cdot, \cdot)$. En la segunda, el dominio de integración abarca los valores de s entre $s(t)$ y 1, y corresponde a la región fuera del conjunto $A(\cdot, \cdot)$. De esta manera, la esperanza de ind se simplifica a

$$\mathbb{E}(\text{ind}) = \int_0^1 \left(\int_0^{s(t)} 1 \cdot ds + \int_{s(t)}^1 0 \cdot ds \right) dt = \int_0^1 \left(\int_0^{s(t)} ds \right) dt.$$

Finalmente, al resolver la integral interior resultante se consigue el resultado deseado

$$\mathbb{E}(\text{ind}) = \int_0^1 s(t) dt = \int_0^1 \sqrt{1-t^2} dt = \frac{\pi}{4}.$$

1.3 Muestreando distribuciones conocidas

El método de Integración de Monte Carlo se basa en un resultado asintótico (la Ley Débil de Grandes Números) que emerge cuando $R \rightarrow \infty$. Por ello, la calidad de las aproximaciones obtenidas con este método dependerá del valor de R elegido. Desafortunadamente, no existen “fórmulas mágicas” o criterios ampliamente aceptados para elegir un valor adecuado del número de repeticiones. Lo único que puede concluirse con certeza es que cuanto más grande R , mejor será la calidad de la aproximación⁵.

⁵ Existen varias **técnicas de reducción de varianza** (variables antitéticas, variables de control, muestreos de importancia, entre otras) que permiten incrementar la precisión de las simulaciones sin necesidad de incrementar R . Estas técnicas no son desarrolladas en este texto, ya que para los problemas que se analizan el uso de “fuerza bruta”, es decir el simple incremento R , es una solución que acarrea un costo computacional razonable. Para mayor detalle sobre estas técnicas, consúltese el capítulo 13 de Davidson y MacKinnon (1993).

Pero ¿cuán grande debería ser R ? Quizá la mejor respuesta es “lo más grande posible”. Ello dependerá de la complejidad del problema bajo estudio y de la rapidez con la que el procesador pueda realizar las simulaciones. Para problemas simples o fáciles de evaluar, valores razonables de R podrían ser algunas decenas de miles, mientras que para problemas más complejos muchas veces nos tenemos que contentar con unos cientos de repeticiones. A lo largo de este texto, se utilizarán valores relativamente elevados de R , con el propósito de aminorar la influencia de los errores de simulación en nuestras aproximaciones. No obstante, las principales conclusiones cualitativas usualmente pueden obtenerse con un número considerablemente menor de repeticiones, del orden de unos pocos miles.

El programa `Prog12_MC.prg` implementa un ejercicio de Monte Carlo diseñado para ilustrar la sensibilidad de las aproximaciones al valor de R . Este programa genera números aleatorios de un grupo de distribuciones conocidas $y \sim f_y(\cdot)$ cuyos momentos de interés tienen una forma analítica exacta, y compara estos valores con las aproximaciones obtenidas para $R = 10^i$ repeticiones, donde $i = 2, 3, 4, 5$. En particular, el interés se centra en la media de la distribución μ , su varianza σ^2 y el valor del percentil 90, Q_{90} .

Recuerde que si $F_y(\cdot)$ denota la función de distribución acumulada de y , el α -ésimo percentil es el (mínimo) valor Q_α que satisface $F_y(Q_\alpha) = \frac{1}{100}\alpha$. No debería sorprender el hecho de que el percentil muestral, calculado tras ordenar ascendentemente los R valores de y en la muestra e identificar aquel valor que se ubica en la $\frac{R}{100}\alpha$ -ésima posición, converge en probabilidad a Q_α conforme $R \rightarrow \infty$ ⁶.

El cuadro I.1 presenta los resultados. Las filas etiquetadas con μ corresponden a la media; las etiquetadas con σ^2 , a la varianza; y las de nombre Q_{90} , al percentil 90. Los valores teóricos de cada una de estas cantidades se muestran en la columna $R \rightarrow \infty$. Dependiendo del estadístico de interés y de la distribución de la que provienen los datos, las aproximaciones obtenidas con $R = 100$ repeticiones pueden resultar razonables, aunque en términos generales los errores de aproximación son notorios, de entre 5% y 20%. Asimismo, en todos los casos, se aprecia una mejora sustancial en la precisión de estas aproximaciones al incrementar el número de repeticiones a $R = 1,000$. Las aproximaciones son casi indistinguibles de los valores teóricos con $R = 10,000$ y, más aún, con $R = 100,000$.

La estructura del programa `Prog12_MC.prg` es bastante simple. En un *workfile* con 10^5 observaciones, se generan números aleatorios almacenados en la serie y . Luego, se ejecuta un bucle que incrementa el tamaño muestral secuencialmente de manera exponencial: de 10^2 a 10^3 , a 10^4 y a 10^5 . Finalmente, para cada tamaño muestral se calculan y almacenan los estadísticos de interés. Explícitamente,

⁶ Se consigue este resultado al notar que el percentil muestral q_α se define como $FDE_y(q_\alpha) = \frac{1}{100}\alpha$, y por tanto $FDE_y(q_\alpha) = F_y(Q_\alpha)$, donde $FDE_y(\cdot)$ es la función de distribución empírica de y . Se sabe que para cualquier A , $\text{plim } FDE_y(A) = F_y(A)$, por lo que $\text{plim } FDE_y(q_\alpha) = F_y(q_\alpha)$. Al utilizar las igualdades resultantes, se concluye que $\text{plim } q_\alpha = Q_\alpha$.

CAPÍTULO I. INTEGRACIÓN DE MONTE CARLO

Cuadro I.1 Aproximación de cantidades poblacionales para distribuciones conocidas

			$R = 10^2$	$R = 10^3$	$R = 10^4$	$R = 10^5$	$R \rightarrow \infty$
Uniforme	$U(0, 1)$	μ	0.546	0.492	0.498	0.500	0.500
		σ^2	0.088	0.083	0.082	0.083	0.083
		Q_{90}	0.914	0.897	0.896	0.901	0.900
		$4 \cdot \text{EEN}$	0.114	0.036	0.011	0.004	
Normal	$N(0, 1)$	μ	-0.056	-0.013	0.000	0.001	0.000
		σ^2	0.858	1.052	1.006	0.998	1.000
		Q_{90}	1.012	1.283	1.278	1.282	1.282
		$4 \cdot \text{EEN}$	0.444	0.124	0.039	0.012	
t de Student	$t(4)$	μ	-0.312	-0.028	0.007	0.001	0.000
		σ^2	2.583	2.156	2.015	2.001	2.000
		Q_{90}	1.304	1.554	1.529	1.534	1.533
		$4 \cdot \text{EEN}$	0.633	0.182	0.056	0.018	
Chi cuadrado	$\chi^2(3)$	μ	3.300	3.140	3.033	3.007	3.000
		σ^2	6.202	6.155	6.090	6.043	6.000
		Q_{90}	6.791	6.405	6.239	6.268	6.251
		$4 \cdot \text{EEN}$	0.934	0.312	0.096	0.030	
Exponencial	$\exp(5)$	μ	4.749	4.910	4.958	5.001	5.000
		σ^2	21.818	24.918	24.854	25.043	25.000
		Q_{90}	9.941	11.253	11.474	11.547	11.513
		$4 \cdot \text{EEN}$	1.922	0.597	0.199	0.062	
Beta	$\mathcal{B}(3, 2)$	μ	0.613	0.602	0.604	0.601	0.600
		σ^2	0.040	0.039	0.039	0.040	0.040
		Q_{90}	0.862	0.862	0.859	0.858	0.857
		$4 \cdot \text{EEN}$	0.074	0.024	0.008	0.002	

Notas: (Prog12_MC.prg) Las filas etiquetadas con μ corresponden a la media; las etiquetadas con σ^2 , a la varianza; y las de nombre Q_{90} , al percentil 90. Los valores teóricos de cada una de estas cantidades se muestran en la columna $R \rightarrow \infty$. En el caso de las medias y varianzas se tiene lo siguiente: para $y \sim U(a, b)$, $\mu = (a + b)/2$ y $\sigma^2 = (b - a)^2/12$; para $y \sim N(0, 1)$, $\mu = 0$ y $\sigma^2 = 1$; para $y \sim t(\nu)$, $\mu = 0$ y $\sigma^2 = \nu/(\nu - 2)$; para $y \sim \chi^2(\nu)$, $\mu = \nu$ y $\sigma^2 = 2\nu$; para $y \sim \exp(a)$, $\mu = a$ y $\sigma^2 = a^2$; y, finalmente, para $y \sim \mathcal{B}(a, b)$, $\mu = a/(a + b)$ y $\sigma^2 = ab/[(a + b)^2(a + b - 1)]$. En el caso de los percentiles, se utilizan, respectivamente, los siguientes comandos de integración exacta de EViews: @QNORM(0.9), @QTDIST(0.90, !v), @QCHISQ(0.90, !v), @QEXP(0.90, !a) y @QBETA(0.90, !a, !b). Las filas de nombre $4 \cdot \text{EEN}$ muestran el valor aproximado del ancho del intervalo al 95 % de confianza de la media.

```

FOR !i = 2 TO 5
    !obs = 10^!i
    SMPL 1 !obs
    Cálculo y almacenamiento del promedio, @MEAN(y)
    Cálculo y almacenamiento de la varianza, @VAR(y)
    Cálculo y almacenamiento del percentil 90, @QUANTILE(y, 0.90)
NEXT !i

```

APUNTES DE ESTUDIO

Las distribuciones utilizadas y los comandos que generan los correspondientes números aleatorios son los siguientes (considere que los escalares !a, !b, !mu, !sigma2 y !v han sido definidos previamente en el programa):

Uniforme, $y \sim U(a, b)$,	SERIES y = !a + (!b - !a)*RND
Normal, $y \sim N(\mu, \sigma^2)$,	SERIES y = !mu + @SQRT(!sigma2)*NRND
t de Student, $y \sim t(\nu)$,	SERIES y = @RTDIST(!v)
Chi cuadrado, $y \sim \chi^2(\nu)$,	SERIES y = @RCHISQ(!v)
Exponencial, $y \sim \exp(a)$,	SERIES y = @REXP(!a)
Beta, $y \sim \mathcal{B}(a, b)$,	SERIES y = @RBETA(!a, !b)

Al revisar los resultados del cuadro I.1, se confirma que para problemas relativamente simples, algunos miles de repeticiones proveen aproximaciones bastante aceptables. No obstante, considerando que la ejecución del programa Prog12_MC.prg tarda tan solo unos segundos en una computadora moderna (que no es particularmente sofisticada), el costo computacional de utilizar algunas decenas de miles repeticiones, y obtener así aproximaciones realmente buenas, no es alto.

Otro punto por resaltar de los resultados del cuadro I.1 es una concavidad en la precisión conforme R crece: la mejora de pasar de $R = 100$ a $R = 1,000$ es cualitativamente similar a la mejora de pasar de $R = 1,000$ a $R = 10,000$ y, a su vez, de pasar de $R = 10,000$ a $R = 100,000$. Es decir, se consiguen incrementos aproximadamente lineales en precisión ante aumentos exponenciales en el número de repeticiones.

La explicación de este fenómeno se relaciona con el hecho de que el promedio muestral \bar{y} , que es una aproximación de μ , es en sí mismo una variable aleatoria. En estricto, \bar{y} es un **estimador** de μ , y al ser aleatorio presenta cierta dispersión. Dado que todas las observaciones en la muestra $\{y_1, y_2, \dots, y_R\}$ provienen de la misma distribución, comparten los mismos momentos: $\mathbb{E}(y_i) = \mu$ y $\mathbb{V}(y_i) = \sigma^2$, para todo $i = 1, 2, \dots, R$, una consecuencia de que el muestreo sea **idéntico**. Más aún, dado que el muestreo es también **independiente**, la covarianza entre cualquier par de observaciones es igual a cero, $\mathbb{C}(y_i, y_j) = 0$ para $i = 1, 2, \dots, R, j = 1, 2, \dots, R$ e $i \neq j$. Con ello,

$$\mathbb{E}(\bar{y}) = \frac{1}{R} \mathbb{E} \left(\sum_{i=1}^R y_i \right) = \frac{1}{R} \sum_{i=1}^R \mathbb{E}(y_i) = \frac{1}{R} \sum_{i=1}^R \mu = \mu,$$

y

$$\mathbb{V}(\bar{y}) = \frac{1}{R^2} \mathbb{V} \left(\sum_{i=1}^R y_i \right) = \frac{1}{R^2} \sum_{i=1}^R \mathbb{V}(y_i) + \text{covarianzas} = \frac{1}{R^2} \sum_{i=1}^R \sigma^2 = \frac{\sigma^2}{R}.$$

El promedio es una variable aleatoria centrada en μ y con una desviación estándar igual a σ/\sqrt{R} . Al valor $\text{EEN} = \hat{\sigma}/\sqrt{R}$, donde $\hat{\sigma}$ es la desviación estándar muestral, se le conoce como **error estándar numérico** y cuantifica la dispersión de \bar{y} alrededor de μ . En particular,

el **Teorema del Límite central** (que se estudia con detalle en la sección I.6) permite concluir que cuando R es lo suficientemente grande, \bar{y} se comporta como una variable normalmente distribuida. De este modo, puede construirse un intervalo de confianza para el valor de μ como $\bar{y} \pm z_\alpha \cdot \text{EEN}$, donde z_α es el α -ésimo percentil de una variable distribuida como normal estándar. Para un intervalo del 95% de confianza, $z_{2.5} \simeq 1.96$. El ancho de este intervalo, $2 \cdot z_\alpha \cdot \text{EEN}$, da una medida precisa de la calidad de \bar{y} como aproximación de μ .

En consecuencia, dado que EEN tiene como numerador \sqrt{R} , se explica la concavidad en la precisión. Cuando R se ve multiplicado por 10, el ancho de los intervalos de confianza asociados con la incertidumbre de estas aproximaciones se divide entre $\sqrt{10} \simeq 3.16$. Este es exactamente el comportamiento reportado en las filas de nombre $4 \cdot \text{EEN}$ en el cuadro I.1, que contienen el resultado de evaluar $4 \cdot \text{STDEV}(\bar{y}) / \text{SQRT}(\text{!obs})$. Desde la perspectiva del ancho del intervalo de confianza para la media, se confirma que las simulaciones con $R = 1,000$ resultan en aproximaciones bastante aceptables, y que las aproximaciones pueden ser excelentes si se dispone de suficiente poder computacional como para utilizar $R \geq 10,000$.

I.4 Eficiencia del promedio y la mediana muestrales

A continuación, se utiliza el método de Integración de Monte Carlo para estudiar un problema estadístico interesante. En econometría, es común encontrar varios estimadores alternativos, con propiedades distintas, para un mismo parámetro poblacional de interés. Surge, entonces, la necesidad de contar con criterios que permitan al investigador discriminar entre estimadores y , a la larga, preferir un estimador sobre otro.

Una aplicación ampliamente documentada, y bastante ilustrativa, sobre esta disyuntiva es la comparación entre el promedio y la mediana muestrales como estimadores de un parámetro de locación μ . En adelante, considere una variable aleatoria y y cuya función de densidad $f_y(\cdot)$ es **simétrica**, por lo que el parámetro de locación μ es igual tanto a la media como a la mediana poblacional. Considere una muestra de tamaño T con observaciones $\{y_1, y_2, \dots, y_T\}$ extraídas aleatoriamente de $f_y(\cdot)$. Además, sea \bar{y} el promedio muestral y sea \tilde{y} la mediana muestral de y , calculados con esta muestra (mantenemos la dependencia de T de estos estadísticos implícita para aliviar la notación).

La **eficiencia relativa** del promedio respecto a la mediana se define como

$$\text{ER}(\bar{y}, \tilde{y}) = \frac{\text{V}(\tilde{y})}{\text{V}(\bar{y})}.$$

En palabras, ER es simplemente la razón de varianzas de los dos estimadores alternativos⁷.

⁷ Un criterio de comparación más amplio es la razón de errores cuadráticos medios, $\mathbb{E}((\bar{y} - \mu)^2) / \mathbb{E}((\tilde{y} - \mu)^2)$. Cuando los estimadores de μ en cuestión son insesgados, lo que de hecho ocurre con el promedio y la mediana muestrales de datos provenientes de distribuciones simétricas $\mathbb{E}(\bar{y}) = \mathbb{E}(\tilde{y}) = \mu$, se tiene que $\mathbb{E}((\bar{y} - \mu)^2) = \text{V}(\bar{y})$ y $\mathbb{E}((\tilde{y} - \mu)^2) = \text{V}(\tilde{y})$, por lo que la razón de errores cuadráticos medios equivale al criterio ER.

APUNTES DE ESTUDIO

Al involucrar varianzas, el criterio ER alude a la precisión con la que los estimadores aproximan a μ en una muestra. La interpretación más directa del criterio es la siguiente: se necesita una muestra de $T \times \text{ER}(\bar{y}, \tilde{y})$ observaciones para que la varianza de la media sea igual a la varianza de la mediana calculada en una muestra de T observaciones. De esta forma, si $\text{ER}(\bar{y}, \tilde{y}) > 1$, entonces se favorece a la mediana \tilde{y} como un estimador de μ con mejores propiedades; mientras que si $\text{ER}(\bar{y}, \tilde{y}) < 1$, entonces la media \bar{y} es el estimador dominante.

Como se discutió en la sección anterior, para un muestreo aleatorio e idéntico (todas las observaciones en la muestra provienen de la misma distribución), donde $\mathbb{V}(y) = \sigma^2$ es la varianza de la observación típica en la muestra,

$$\mathbb{V}(\bar{y}) = \frac{\sigma^2}{T}.$$

Si bien es cierto que no se cuenta con un resultado exacto para la mediana, cuando T es relativamente grande se tiene que

$$\mathbb{V}(\tilde{y}) \simeq \frac{1}{4T f_y(\mu)^2},$$

donde $f_y(\mu)$ es el valor de la función de densidad de y evaluada en la mediana μ (típicamente, es el máximo valor de la función de densidad). Con ello,

$$\text{ER}(\bar{y}, \tilde{y}) \simeq 4\sigma^2 f_y(\mu)^2.$$

En el caso de una distribución normal con parámetro de locación (media) μ y varianza σ^2 , se tiene como función de densidad a

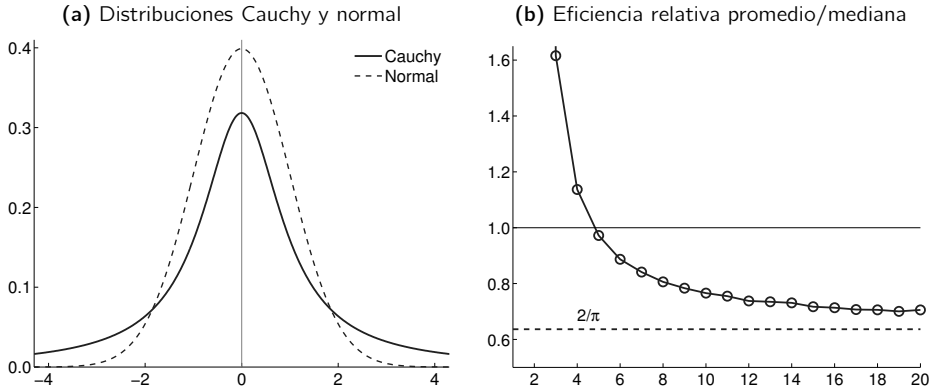
$$f_y(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu}{\sigma} \right)^2 \right\},$$

por lo que $f_y(\mu)^2 = (2\pi\sigma^2)^{-1}$ y, finalmente, $\text{ER}(\bar{y}, \tilde{y}) = 2/\pi \simeq 0.64 < 1$: la media es un estimador de μ más eficiente que la mediana. La mediana requiere de 100 observaciones para alcanzar el nivel de precisión que la media consigue con 64 observaciones.

Se sabe que la mediana es un estimador de locación **robusto**, insensible a observaciones extremas o *outliers*. Para datos normalmente distribuidos, las observaciones extremas son lo suficientemente infrecuentes como para que la pérdida en eficiencia (la mayor varianza de la mediana relativa al promedio) resulte ser un precio muy alto que pagar por una mayor robustez. No obstante, la mediana podría ser el estimador dominante en un contexto de observaciones extremas más frecuentes.

Investigamos esta posibilidad a través del estudio de Monte Carlo disponible en el programa `Prog13_Eficiencia.prg`. Acá, se muestrean datos provenientes de distribuciones t de Student con ν grados de libertad. Esta distribución es sumamente conveniente, ya que,

Gráfico I.2 Colas anchas y eficiencia relativa del promedio y la mediana



Notas: (Prog13_Eficiencia.prg) El panel (a) ilustra que la distribución Cauchy (o, en términos más generales, distribuciones t de Student con pocos grados de libertad) presenta colas mucho más anchas que la distribución normal y, por tanto, una mayor probabilidad de ocurrencia de valores extremos. El panel (b) muestra el valor simulado de $ER(\bar{y}, \tilde{y})$, en el eje vertical, como función del número de grados de libertad ν , en el eje horizontal.

además de ser simétrica, permite calibrar la ocurrencia de eventos extremos únicamente con el parámetro ν .

Cuando ν es grande, el comportamiento de la distribución t de Student es muy parecido al de una distribución normal estándar (de hecho, cuando $\nu \rightarrow \infty$ ambas distribuciones coinciden). Por el contrario, cuando ν es pequeño, la distribución t de Student presenta **colas anchas**, asignando así una mayor probabilidad a la ocurrencia de valores extremos. El caso límite ocurre con $\nu = 1$, que es cuando la distribución t de Student pasa a ser una distribución **Cauchy**. Esta distribución es famosa en estadística porque ninguno de sus momentos (media, varianza, asimetría o curtosis) está bien definido. El gráfico I.2(a) compara las densidades de una distribución Cauchy y una distribución normal estándar, y se aprecia que la primera presenta colas considerablemente más anchas que la segunda. Cuando $\nu > 1$, la media de la distribución t de Student se encuentra bien definida, y es igual a cero, mientras que cuando $\nu > 2$, también lo está su varianza, que es igual a $\nu/(\nu - 2)$.

En el programa, dado el número de grados de libertad !v, se muestrean números aleatorios de una distribución t de Student con el comando

```
SERIES y = @RTDIST(!v),
```

que se ejecuta para un tamaño muestral de !T observaciones, aplicando antes la instrucción SMPL 1 !T. En la iteración !i, tras el muestreo, se almacenan los estimadores

APUNTES DE ESTUDIO

```
promedio(!i) = @MEAN(y)
mediana(!i) = @MEDIAN(y) ,
```

donde las series (o vectores) promedio y mediana, previamente definidas, contienen !R observaciones. Al término de las !R repeticiones, se almacenan los criterios de eficiencia relativa en un vector de nombre ER,

```
ER(!v) = @VAR(promedio)/@VAR(mediana) .
```

Este procedimiento se repite para varios valores de !v.

El gráfico 1.2(b) presenta los resultados de simulaciones con $T = 100$ observaciones y $R = 100,000$ repeticiones. En particular, se grafica la aproximación de Monte Carlo de $ER(\bar{y}, \tilde{y})$ como función de ν . Dado el gran número de repeticiones empleado en las simulaciones, se espera que el error de simulación en el cálculo de estos criterios sea realmente reducido. Se confirma que la mediana es un estimador más eficiente que la media para muestreos con valores extremos recurrentes. En particular, se aprecia que $ER > 1$ para $\nu \leq 4$. Para $\nu = 1$, la eficiencia relativa es infinitamente más favorable para la mediana, lo que se refleja en valores simulados de ER realmente elevados (recuerde que en este caso la media de la distribución no existe), mientras que para $\nu = 2$ se tiene que $ER \simeq 8.34$ y para $\nu = 3$, $ER \simeq 1.62$. Cuando $\nu = 5$, la eficiencia de ambos estimadores es comparable, ya que $ER \simeq 0.97$. Asimismo, se aprecia cómo $ER \rightarrow 2/\pi < 1$ conforme ν se incrementa.

1.5 Inferencia en modelos de regresión

Considere el siguiente modelo de regresión lineal,

$$y_t = \alpha + \beta x_t + \varepsilon_t ,$$

para $t = 1, 2, \dots, T$. Suponemos que los T valores $\{x_1, x_2, \dots, x_T\}$ se mantienen fijos a lo largo de muestreos repetidos, mientras que ε_t es una variable aleatoria idéntica e independientemente distribuida, de media cero y varianza σ^2 . Estos supuestos se imponen por simplicidad, para facilitar el álgebra que sigue. Cabe mencionar que se llegaría a resultados similares si se permitiese que x_t sea aleatorio, pero independiente de ε_t . El interés se centra en estimar e inferir sobre el parámetro β .

Disponemos de tres alternativas. En primer lugar, de acuerdo con lo indicado por este proceso generador de datos, se estiman los parámetros α y β directamente por mínimos cuadrados ordinarios (MCO), al regresar y_t sobre una constante y x_t . En este caso, el estimador de β es igual a

$$b_{\text{MCO}} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2} ,$$

donde \bar{x} e \bar{y} son los promedios muestrales de x_t e y_t , respectivamente. Al reemplazar la definición de y_t en la expresión de b_{MCO} se consigue que

$$b_{MCO} = \beta + \frac{\sum_{t=1}^T (x_t - \bar{x}) \varepsilon_t}{\sum_{t=1}^T (x_t - \bar{x})^2},$$

expresando así al estimador b_{MCO} en función del parámetro desconocido β y del término de error no observable ε_t . El segundo término es una variable aleatoria (la razón de los promedios de $(x_t - \bar{x})\varepsilon_t$ y de $(x_t - \bar{x})^2$) cuyas propiedades se transmiten al estimador MCO. La esperanza de este segundo término, dado el supuesto de que x_t no es aleatorio, es cero (la esperanza de cada uno de los ε_t), mientras que su varianza es una suma de varianzas (las covarianzas son cero, dado el supuesto de independiencia en ε_t). Así,

$$\mathbb{E}(b_{MCO}) = \beta \quad \text{y} \quad \mathbb{V}(b_{MCO}) = \frac{\sigma^2}{\sum_{t=1}^T (x_t - \bar{x})^2}.$$

El estimador MCO es **insesgado**, ya que su esperanza coincide con el parámetro sobre el que el estimador busca inferir. Por su parte, la estimación MCO ostenta una interesante propiedad de optimalidad (el conocido **Teorema de Gauss-Markov**): es el estimador que presenta menor varianza, dentro de todos los posibles estimadores lineales e insesgados.

La segunda alternativa es un **estimador restringido**. Suponga que en lugar de estimar el coeficiente α , se impone la restricción $\alpha = 0$ y se procede a estimar β . Ahora, se regresa y_t directamente sobre x_t , por lo que el estimador de mínimos cuadrados restringidos (MCR) de β es igual a

$$b_{MCR} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2}.$$

Al reemplazar la definición de y_t en la expresión de b_{MCR} se consigue que

$$b_{MCR} = \beta + \alpha \left(\frac{T\bar{x}}{\sum_{t=1}^T x_t^2} \right) + \frac{\sum_{t=1}^T x_t \varepsilon_t}{\sum_{t=1}^T x_t^2},$$

expresando así al estimador b_{MCR} en función de los parámetros desconocidos α y β y del término de error no observable ε_t . El tercer término es una variable aleatoria (la razón de los promedios de $x_t \varepsilon_t$ y de x_t^2) cuyas propiedades se transmiten al estimador MCR. La esperanza de este tercer término, dado el supuesto de que x_t no es aleatorio, es cero (la esperanza de cada uno de los ε_t), mientras que su varianza es una suma de varianzas (las covarianzas son cero, dado el supuesto de independiencia en ε_t). Por su parte, el segundo término (que involucra a α) no es aleatorio y, por tanto, no contribuye a la variabilidad del estimador. Así,

$$\mathbb{E}(b_{MCR}) = \beta + \alpha \left(\frac{T\bar{x}}{\sum_{t=1}^T x_t^2} \right) \quad \text{y} \quad \mathbb{V}(b_{MCR}) = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}.$$

A menos que $\alpha = 0$, el estimador MCR es **sesgado**, ya que su esperanza difiere del parámetro sobre el que el estimador busca inferir. Recuerde que este estimador puede entenderse como MCO aplicado a un modelo que ha impuesto la restricción $\alpha = 0$. Si esta es falsa, entonces emerge un sesgo en la estimación. Asimismo, imponer una restricción, al margen de si es verdadera o falsa, reduce la variabilidad del estimador. Es simple verificar que $\mathbb{V}(b_{\text{MCR}}) < \mathbb{V}(b_{\text{MCO}})$, a menos que $\bar{x} = 0$, en cuyo caso se tiene una igualdad. Es importante mencionar que este resultado no contradice al Teorema de Gauss-Markov, que establece un ordenamiento inequívoco entre estimados insesgados, ya que el estimador MCR es sesgado.

La comparación entre los estimadores MCO y MCR en este sencillo marco analítico, refleja resultados sumamente generales sobre el efecto que omitir variables relevantes o incluir variables redundantes ocasiona sobre la estimación de parámetros. Si $\alpha \neq 0$, entonces el intercepto es un regresor relevante, ya que aparece en el proceso generador de datos. Al excluirlo, emerge un sesgo a menos que $\bar{x} = 0$, que es la “correlación” entre el regresor omitido y el incluido. La exclusión reduce la varianza, ya que se estima una menor cantidad de coeficientes. Por otro lado, si $\alpha = 0$, entonces el intercepto no forma parte del proceso generador de datos, y el estimador que impone $\alpha = 0$ es el estimador MCO sobre una regresión correctamente especificada. El Teorema de Gauss-Markov indica que este sería el mejor estimador (lineal) disponible. Y, de hecho, se tiene que el estimador que incluye un intercepto (redundante, en esta ocasión), si bien sigue siendo insesgado, presenta una mayor varianza (salvo que $\bar{x} = 0$).

En casos generales, es posible que el estimador MCR sesgado sea preferible al estimador MCO insesgado, ya que su menor varianza podría más que compensar la presencia del sesgo. Se trata, pues, de una comparación de errores cuadráticos medios (ECM). El ECM es una distancia estadística muy aceptada como criterio de comparación de estimadores con distintas propiedades. Un estimador es mejor en esta comparación, si es que produce el menor ECM. Es simple mostrar que el ECM de un estimador es la suma de su varianza más el cuadrado de su sesgo. Así,

$$\mathbb{ECM}(b_{\text{MCO}}) = \mathbb{E}((b_{\text{MCO}} - \beta)^2) = \frac{\sigma^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \quad \text{y}$$

$$\mathbb{ECM}(b_{\text{MCR}}) = \mathbb{E}((b_{\text{MCR}} - \beta)^2) = \alpha^2 \left(\frac{T\bar{x}}{\sum_{t=1}^T x_t^2} \right)^2 + \frac{\sigma^2}{\sum_{t=1}^T x_t^2}.$$

Cuál estimador es preferible depende, entre otras cosas, de cómo se compara α^2 con σ^2 . Si α^2 es pequeño en comparación con σ^2 , entonces el sesgo del estimador MCR será pequeño y la reducción en su varianza lo hará preferible a MCO. Por el contrario, si α^2 es grande, el sesgo del estimador MCR será el que más contribuya a su ECM, haciendo que MCO sea un estimador con mejores propiedades. Se preferirá el estimador MCR si

$\text{ECM}(b_{\text{MCO}}) > \text{ECM}(b_{\text{MCR}})$, lo que implica en concreto que

$$\alpha^2 < \sigma^2 \left(\frac{\sum_{t=1}^T x_t^2}{T \sum_{t=1}^T (x_t - \bar{x})^2} \right).$$

Dado que ninguno de estos estimadores es siempre dominante, posiblemente en la práctica el investigador siga el siguiente procedimiento que busca determinar la importancia de α . Primero, se estima la regresión MCO, sin restricciones. Luego se pregunta si tiene suficiente evidencia para concluir (o no desmentir) que $\alpha = 0$. Formalmente, se estudia la hipótesis nula $H_0 : \alpha = 0$. Usualmente, se calcula el estadístico t asociado con esta hipótesis, que denotamos como τ_α , y se le compara contra un valor crítico τ_{crit} proveniente de la distribución t_{T-2} (donde se pierden 2 grados de libertad por haber estimado α y β). Puesto de modo más simple, se determina si el estimador de α es estadísticamente distinto de cero o no. Si lo es, lo que ocurre si $|\tau_\alpha| \geq \tau_{\text{crit}}$, entonces se concluye que el intercepto es un regresor importante y se decide incluirlo en la regresión. Se opta, pues, por el estimador MCO de β . En cambio, si resulta que $|\tau_\alpha| < \tau_{\text{crit}}$, podría concluirse que el intercepto no es un regresor importante, lo que llevaría a optar por el estimador MCR de β .

Es bueno mencionar que el resultado de este procedimiento **no es ni el estimador MCO ni el estimador MCR**. Es un estimador híbrido definido como

$$b_{\text{PT}} = \begin{cases} b_{\text{MCO}}, & \text{si } |\tau_\alpha| > \tau_{\text{crit}} \\ b_{\text{MCR}}, & \text{si } |\tau_\alpha| \leq \tau_{\text{crit}} \end{cases}$$

o, más compactamente,

$$b_{\text{PT}} = \mathbf{1}\{|\tau_\alpha| > \tau_{\text{crit}}\} \cdot b_{\text{MCO}} + \mathbf{1}\{|\tau_\alpha| \leq \tau_{\text{crit}}\} \cdot b_{\text{MCR}}.$$

El estimador b_{PT} , que es probablemente el que sea utilizado en la práctica, es conocido como el estimador **pre-testing** (PT) de β . Sus propiedades, a diferencia de los estimadores MCO y MCR, no son fáciles de derivar porque se trata, en estricto, de un estimador no lineal. El resultado de la prueba $|\tau_\alpha| > \tau_{\text{crit}}$ depende de las propiedades de ε_t y, por tanto, se encuentra correlacionado con b_{MCO} y b_{MCR} . Describir esta correlación es algo engorroso, pero un ejercicio de simulación permite explorar estas propiedades fácilmente.

El programa Prog14_PreTesting genera un *workfile* de !R observaciones sobre el que se ejecutarán estimaciones con un tamaño muestral predefinido de !T observaciones para varios valores de α (!alpha). Las !T observaciones del regresor x_t se generan por única vez a inicios del programa (respondiendo al supuesto de que este regresor no es aleatorio). Luego, se calculan las siguientes cantidades:

```
!sumx2 = @SUMSQ(x)
!sumxx2 = @SUMSQ(x - @MEAN(x))
!sigma2 = !T*!sumxx2/!sumx2
!sigma = @SQRT(!sigma2).
```

APUNTES DE ESTUDIO

El escalar `!sumx2` contiene a $\sum_t x_t^2$ y el escalar `!sumxx2` contiene a $\sum_t (x_t - \bar{x})^2$. Note que la varianza del error σ^2 , contenida en el escalar `!sigma2`, es calibrada de tal manera que cuando $\alpha^2 = 1$, entonces los ECM de los estimadores MCO y MCR son idénticos. Cuando $\alpha^2 < 1$, luego, el ECM del estimador MCR es el menor.

El programa realiza varias estimaciones, con diversos valores de α . Este parámetro tomará `!nalpha` valores equidistantes en el intervalo [`!alphamin`, `!alphamax`]. El cuerpo del programa contiene los siguientes dos bucles:

```
FOR !ialpha = 1 TO !nalpha
    !alpha = f(!alphamin, !alphamax, !ialpha, !nalpha)
    SMPL @ALL
    SERIES bMCO = NA
    SERIES bMCR = NA
    SERIES bMPT = NA
    SERIES rechazo = 0
    FOR !i = 1 TO !R
        SMPL 1 !T
        SERIES y = !alpha + !beta*x + !sigma*NRND
        EQUATION MCO.LS y x C
        bMCO(!i) = MCO.C(1)
        !tstatC = MCO.@TSTATS(2)
        EQUATION MCR.LS y x
        bMCR(!i) = MCR.C(1)
        IF @ABS(!tstatC) > !tcrit THEN
            bPT(!i) = MCO.C(1)
            rechazo(!i) = 1
        ELSE
            bPT(!i) = MCR.C(1)
        ENDIF
    NEXT !i
    Almacenamiento de resultados
NEXT !ialpha
```

En primer lugar, se tiene un bucle asociado con el contador `!ialpha` que determina el valor de `!alpha`. Cada vez que se inicia este bucle se generan series de nombre `bMCO`, `bMCR` y `bPT`, que contendrán las `!R` realizaciones de los estimadores bajo estudio. Asimismo, se genera una serie de nombre `rechazo` (inicialmente llena de ceros) que registrará el evento $|\tau_\alpha| > \tau_{\text{crit}}$, es decir el rechazo de $H_0 : \alpha = 0$.

Luego, se tiene el bucle, de `!R` repeticiones que varían con el contador `!i`, que implementa la Integración de Monte Carlo. Primero, se generan los datos con el

comando `SERIES y = !alpha + !beta*x + !sigma*NRND` bajo una muestra activa de `!T` observaciones, `SMPL 1 !T`. El verdadero valor de β (que no es muy relevante) se almacena en el escalar `!beta`. Asimismo, se asume normalidad: $\varepsilon_t \sim N(0, \sigma^2)$. Una vez generados los datos (note que `x` no varía con `!i`), se calcula el estimador MCO en la ecuación `EQUATION MCO.LS y x C`, se almacena b_{MCO} con el comando `bMCO(!i) = MCO.C(1)`, y se recupera, además, el ratio τ_α de significación del intercepto estimado, `!tstatC = MCO.@TSTATS(2)`. Análogamente, se ejecuta la estimación restringida `EQUATION MCR.LS y x y` y se almacena b_{MCR} con el comando `bMCR(!i) = MCR.C(1)`. Finalmente, se procede a calcular el estimador PT. Si τ_α es mayor que τ_{crit} , previamente almacenado en el escalar `!tcrit = @QTDIST(0.975, !T - 2)`, entonces el estimador PT es igual al estimador MCO, `bPT(!i) = MCO.C(1)`, de otro modo es igual al estimador MCR, `bPT(!i) = MCR.C(1)`. Note que ante el rechazo, se asigna el valor de 1 a la serie rechazo.

Al término de las `!R` repeticiones, las series `bMCO`, `bMCR`, `bPT` y `rechazo` se encuentran llenas de valores, listas para ser analizadas. Llegó el momento de almacenar los resultados. Para ello, previamente se han definido las matrices `Rsesgo`, `Rvar` y `Recm`, cada una con `!nalpha` filas y 6 columnas. La primera columna contiene los valores de α ; la segunda columna, el resultado (sesgo, varianza o ECM) obtenido mediante simulación del estimador MCO; la tercera, el resultado analítico del estimador MCO; la cuarta, el resultado obtenido mediante simulación del estimador MCR; la quinta, el resultado analítico del estimador MCR; y, finalmente, la sexta columna contendrá el resultado simulado del estimador PT. Los resultados de la Integración de Monte Carlo son, luego:

```
SMPL @ALL
Rsesgo(!ialpha, 1) = !alpha
Rsesgo(!ialpha, 2) = @MEAN(bMCO - !beta)
Rsesgo(!ialpha, 4) = @MEAN(bMCR - !beta)
Rsesgo(!ialpha, 6) = @MEAN(bPT - !beta)

Rvar(!ialpha, 1) = !alpha
Rvar(!ialpha, 2) = @VAR(bMCO - !beta)
Rvar(!ialpha, 4) = @VAR(bMCR - !beta)
Rvar(!ialpha, 6) = @VAR(bPT - !beta)

Recm(!ialpha, 1) = !alpha
Recm(!ialpha, 2) = @SUMSQ(bMCO - !beta)/!R
Recm(!ialpha, 4) = @SUMSQ(bMCR - !beta)/!R
Recm(!ialpha, 6) = @SUMSQ(bPT - !beta)/!R.
```

Por brevedad, no se presenta el registro de los resultados analíticos (columnas 3 y 5), pero puede ser consultado en el programa mismo. Note que las esperanzas son reemplazadas por promedios a lo largo de repeticiones, siguiendo los principios del método de Monte Carlo. Finalmente, también se ha definido un vector de nombre `PrMCO` con `!nalpha` filas y 2 columnas. La primera columna contiene los valores de α y la segunda, la probabilidad

APUNTES DE ESTUDIO

de que la prueba t del estimador de α rechace $H_0 : \alpha = 0$, que es convenientemente aproximada mediante una frecuencia relativa:

```
PrMCO(!ialpha, 1) = !alpha  
PrMCO(!ialpha, 2) = @MEAN(rechazo) .
```

Al término del programa, los comandos `PrMCO.XY`, `Rsesgo.XY`, `Rvar.XY` y `Recm.XY` producen gráficos de líneas con los valores de la primera columna sobre el eje horizontal, y los valores correspondientes del resto de columnas sobre el eje vertical. El gráfico I.3 presenta estos resultados, utilizando `!nalpha = 25`, `!alphamin = 0`, `!alphamax = 3` (recuerde que, por construcción, `!alpha = 1` es un punto donde el ECM de los estimadores MCO y MCR es el mismo) y $R = 100,000$ repeticiones.

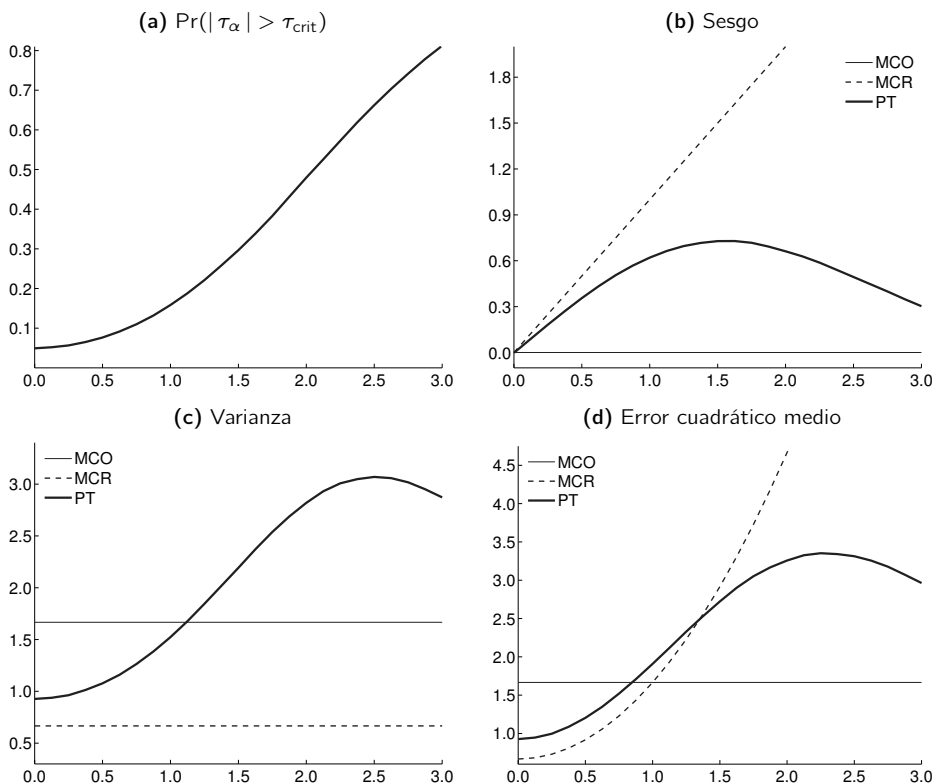
Se confirman varias conclusiones analíticas. Primero, en el panel (b) se aprecia que MCO es un estimador insesgado, mientras que el sesgo del estimador MCR crece linealmente con α . Segundo, en el panel (c) se observa que la varianza del estimador MCR es siempre menor que la del estimador MCO, y ninguna de ellas depende de α . Finalmente, en el panel (d) se aprecia que toda vez que $\alpha < 1$, el ECM del estimador MCR será menor que el de MCO; de otro modo ($\alpha > 1$), MCO será el estimador que ostente el menor ECM.

El panel (a) del gráfico I.3 muestra la probabilidad de $|\tau_\alpha| > \tau_{\text{crit}}$ como función de α . Como era de esperarse, es cada vez más probable que la estimación MCO de α arroje un coeficiente estadísticamente significativo, cuanto mayor sea el verdadero valor de α . En el límite, conforme $\alpha \rightarrow \infty$, $\Pr(|\tau_\alpha| > \tau_{\text{crit}}) \rightarrow 1$. Esta probabilidad afecta las propiedades del estimador PT. Note, en particular, que en el límite descrito, $b_{\text{PT}} \rightarrow b_{\text{MCO}}$. No obstante, para valores finitos de α el estimador PT presenta propiedades muy distintas a MCO.

En el panel (b) se aprecia que, toda vez que $\alpha \neq 0$, el estimador PT es sesgado. La razón es que siempre existe una probabilidad positiva de que $b_{\text{PT}} = b_{\text{MCR}}$. El sesgo es menor que el del estimador MCR, pero mayor que el de MCO. Llama la atención, no obstante, el comportamiento de la varianza del estimador PT, mostrada en el panel (c). En la región donde MCR domina ($\alpha < 1$), PT es más variable que MCR, ya que incorpora la fuente de incertidumbre adicional de determinar si $|\tau_\alpha| > \tau_{\text{crit}}$ o no. Del mismo modo, cuando MCO es dominante ($\alpha > 1$), PT es más variable que MCO, por el mismo motivo. Se tiene, como resumen, el siguiente resultado mostrado en el panel (d): existe una región donde MCR es el estimador dominante ($\alpha < 1$, donde el sesgo es reducido) y una región donde MCO es preferible ($\alpha > 1$, donde el sesgo es considerable), pero **en ningún caso el estimador PT domina**. Nuevamente, este resultado es el reflejo de las distorsiones que el determinar si $|\tau_\alpha| > \tau_{\text{crit}}$, una prueba sujeta a error muestral, genera sobre las propiedades del estimador PT (sobre todo por el lado de la varianza).

El gráfico I.4 muestra, para varios valores de α , las distribuciones muestrales de los tres estimadores estudiados. Estas distribuciones (que son del tipo `kernel` automáticas)

Gráfico I.3 Desempeño de los estimadores MCO, MCR y PT

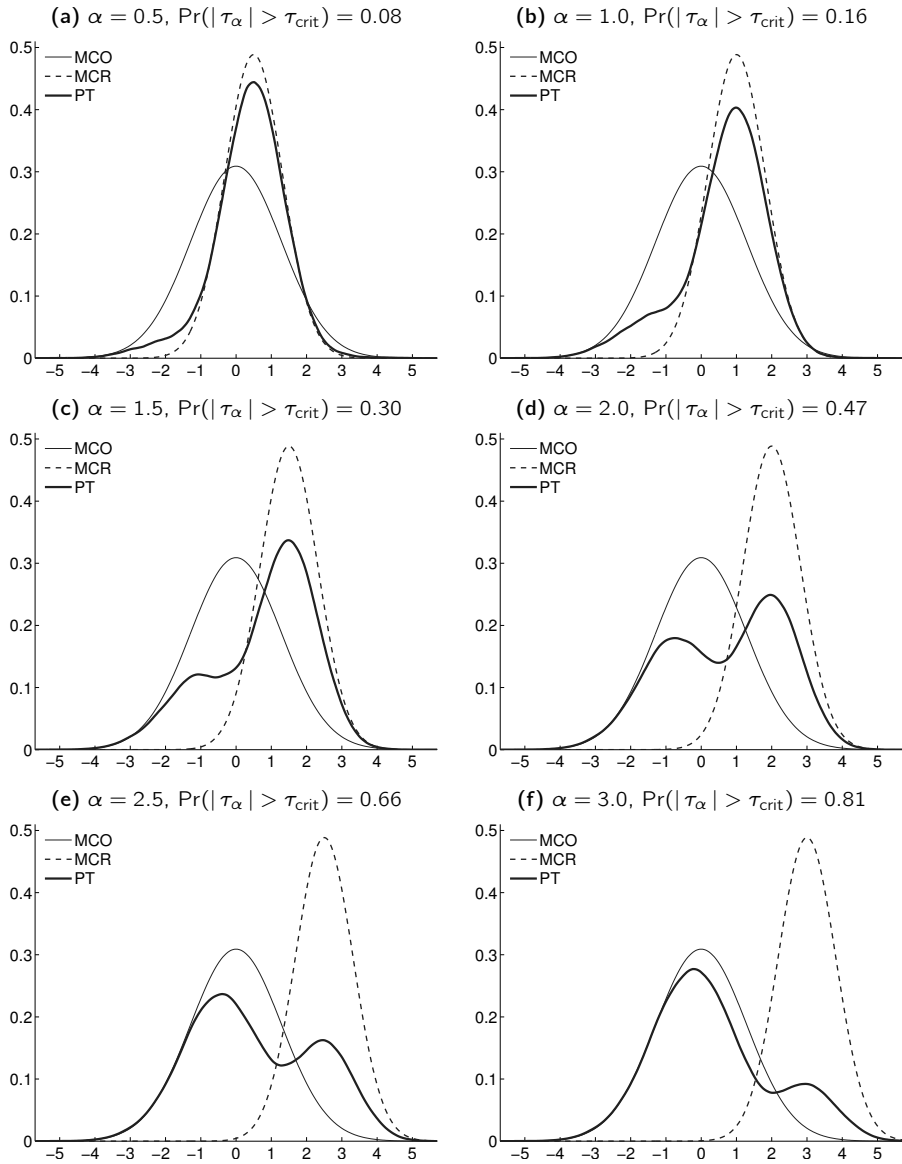


Notas: (Prog14_PreTesting.prg) El panel (a) muestra, en el eje vertical, el promedio de $1\{|\tau_\alpha| > \tau_{crit}\}$ a lo largo de $R = 100,000$ repeticiones y para 25 valores de α , en el eje horizontal. Los paneles (b), (c) y (d) muestran aproximaciones de $\mathbb{E}(b - \beta)$, $\mathbb{V}(b)$ y $\mathbb{E}((b - \beta)^2)$, respectivamente.

se generan con el comando `g.DISTPLOT(S) KERNEL`, donde `g` es un grupo que contiene las series simuladas: `GROUP g bMCO bMCR bPT`.

Dado que $\varepsilon_t \sim N(0, \sigma^2)$ en las simulaciones, es simple verificar que las distribuciones muestrales de b_{MCO} y de b_{MCR} son también normales. La primera es centrada en β , y la segunda es menos variable y centrada en $\beta + \alpha$. Luego, la distribución de b_{PT} será una **mixtura de normales**, cuyas propiedades dependerán de α , en estricto de $\Pr(|\tau_\alpha| > \tau_{crit})$. Es interesante notar cómo la distribución de b_{PT} pasa de ser muy cercana a la de b_{MCR} , a parecerse a la de b_{MCO} conforme α se incrementa. En el proceso es usual que b_{PT} presente una **distribución bimodal**. Se aprecia, además, cómo cuando $\alpha < 1$ la distribución de b_{PT} es más variable que la de b_{MCR} , mientras que es más variable que la de b_{MCO} cuando $\alpha > 1$.

Gráfico I.4 Distribuciones muestrales de los estimadores MCO, MCR y PT



Notas: (Prog14_PreTesting.prg) Densidades kernel (automáticas) para $R = 100,000$ repeticiones.

1.6 Leyes asintóticas importantes

Como se ha visto, obtener resultados analíticos en muestras finitas sobre las propiedades de estadísticos y estimadores de interés puede ser sumamente engorroso o podría requerir de supuestos muy restrictivos. La teoría asintótica provee un marco de análisis en donde se estudian estas propiedades a medida que el tamaño muestral va creciendo indefinidamente, $T \rightarrow \infty$. Este límite “elimina” la aleatoriedad observada en la muestra (digamos, la variabilidad muestral) y provee aproximaciones del comportamiento de los estadísticos en muestras grandes.

Considere una muestra de tamaño T con observaciones $\{y_1, y_2, \dots, y_T\}$ extraídas aleatoriamente de una misma función de distribución $f_y(\cdot)$. Es decir, el muestreo es idéntico e independiente. Presentamos dos resultados asintóticos sobre el comportamiento del promedio muestral \bar{y} bajo estas condiciones, y que son válidos para **cualquier función de distribución** $f_y(\cdot)$. Estos son resultados de primera importancia para el análisis econométrico porque los estimadores econométricos de interés pueden expresarse como funciones de promedios muestrales, y porque permiten al investigador prescindir de supuestos sobre las características de $f_y(\cdot)$. En la práctica, se compensa la falta de conocimiento (o de interés) sobre esta función de distribución con una mayor cantidad de datos en la muestra.

El primer resultado es el Teorema de Khinchine, presentado previamente como el principio fundamental detrás del método de Integración de Monte Carlo: el promedio muestral converge en probabilidad a la esperanza de la observación típica de la muestra,

$$\bar{y} = \frac{1}{T} \sum_{i=1}^T y_i \xrightarrow{p} \mu \quad \text{conforme} \quad T \rightarrow \infty,$$

donde $\mathbb{E}(y) = \mu$.

El segundo resultado es el **Teorema del Límite Central**, cuya versión más estilizada se conoce como el **Teorema de Lindeberg-Lévy**. Para motivar el resultado de este teorema, suponga inicialmente que $f_y(\cdot)$ es una distribución normal: $y_i \sim N(\mu, \sigma^2)$ para todo $i = 1, 2, \dots, T$. Con ello, la distribución muestral del promedio \bar{y} es igual a

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{T}\right) \quad \text{o, alternativamente,} \quad \sqrt{T}(\bar{y} - \mu) \sim N(0, \sigma^2).$$

Este es un resultado **exacto** que se cumple para cualquier tamaño muestral.

El Teorema del Límite Central provee un resultado similar asintóticamente. Nuevamente, considere que $f_y(\cdot)$ es **arbitraria**. El único requerimiento es que $\mathbb{E}(y) = \mu$ y $\mathbb{V}(y) = \sigma^2$ sean cantidades finitas. Luego,

$$\sqrt{T}(\bar{y} - \mu) \xrightarrow{d} N(0, \sigma^2) \quad \text{conforme} \quad T \rightarrow \infty.$$

Es decir que en muestras grandes los promedios muestrales tienden a comportarse como variables normalmente distribuidas **sin importar cuál es la distribución de la que provienen los datos**. La utilidad práctica de este postulado es que los ejercicios de inferencia, por ejemplo la construcción de intervalos de confianza para μ , pueden ser basados en la normalidad de \bar{y} . Es importante enfatizar que la normalidad de \bar{y} en muestras grandes es un resultado, no un supuesto.

El programa `Prog15_LeyesAsintoticas.prg` implementa una serie de experimentos de Monte Carlo con el propósito de ilustrar el funcionamiento de estas leyes asintóticas. Partiendo de $u \sim N(0, 1)$, se generan variables aleatorias de la forma

$$y = \frac{u^a - \mathbb{E}(u^a)}{\sqrt{\mathbb{E}(u^{2a}) - \mathbb{E}(u^a)^2}},$$

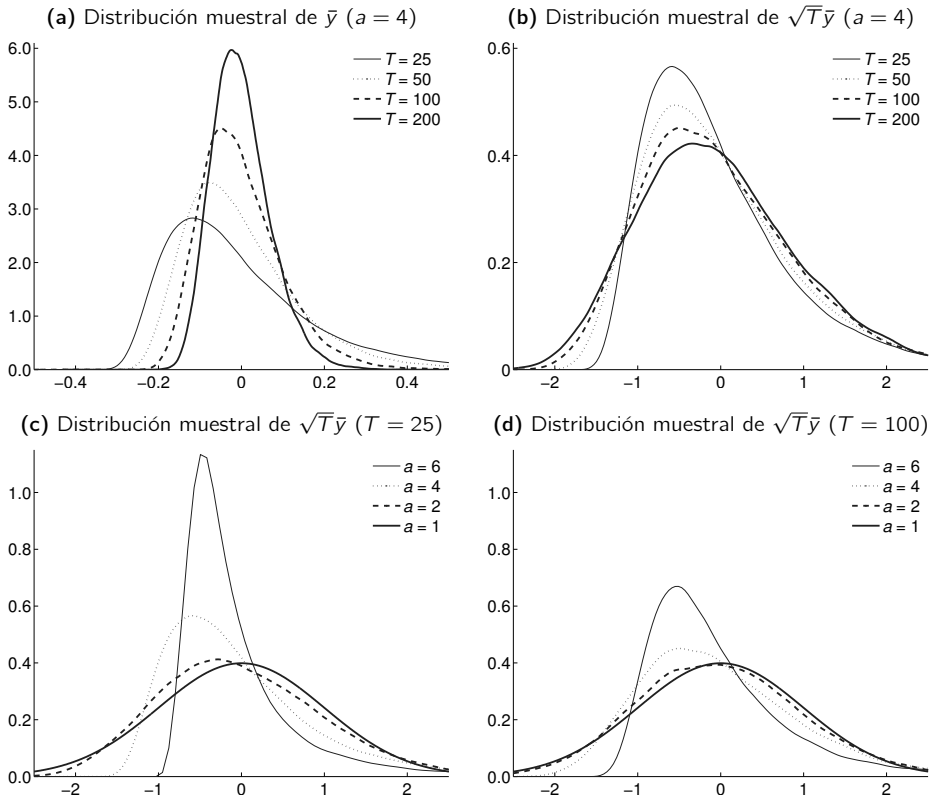
donde a es un número entero. Note que y es una variable estandarizada, de modo que para cualquier valor de a , $\mathbb{E}(y) = 0$ y $\mathbb{V}(y) = 1$. Conforme a se incrementa, la distribución de y se vuelve cada vez más asimétrica, con una cola larga hacia la derecha. El caso de $a = 1$ corresponde a $y \sim N(0, 1)$ y, por tanto, a puede interpretarse como una medida de **desvío de la normalidad**. Asimismo, se dispone de resultados analíticos para $a = 1$: $\bar{y} \sim N(0, 1/T)$ y $\sqrt{T}\bar{y} \sim N(0, 1)$. Las medias y varianzas muestrales serán las mismas para $a \neq 1$, pero las distribuciones muestrales variarán con T .

Para un valor de a se generan T números aleatorios y y se calcula su promedio \bar{y} y $\sqrt{T}\bar{y}$. Este procedimiento se repite un gran número de veces ($R = 100,000$) y se reporta la distribución muestral de estos estadísticos. Dado el gran número de repeticiones en la simulación, esta distribución simulada será prácticamente idéntica a la distribución muestral analítica (desconocida).

El panel (a) del gráfico I.5 muestra cómo opera la Ley de Grandes Números. Conforme T se incrementa, la distribución muestral de \bar{y} va concentrando cada vez más masa probabilística alrededor de $\mathbb{E}(y) = 0$. Ello refleja que muestrear cada vez más observaciones de y , provenientes de la misma distribución, provee información creciente para caracterizar tal variable aleatoria. En particular, dado que $\mathbb{V}(\bar{y}) = 1/T$, cuando T se incrementa, la dispersión de distintas realizaciones de \bar{y} alrededor de $\mathbb{E}(\bar{y}) = 0$ se amortigua. En el límite, conforme $T \rightarrow \infty$, $\mathbb{V}(\bar{y})$ irá convergiendo a cero, por lo que \bar{y} dejará de ser aleatorio. Gráficamente, la distribución muestral de \bar{y} colapsa a una masa de probabilidad igual a 1, ubicada en $\mathbb{E}(\bar{y}) = \mathbb{E}(y)$, tal y como predice la Ley Débil de Grandes Números.

El panel (b) muestra la distribución muestral de $\sqrt{T}\bar{y}$ para $a = 4$ (la distribución de y es bastante asimétrica) y para distintos valores de T . Note que a diferencia de lo ocurrido con la distribución de \bar{y} , estas distribuciones muestrales **no colapsan** conforme $T \rightarrow \infty$. La razón es simple. La multiplicación de \bar{y} por \sqrt{T} **estabiliza** la varianza del promedio y evita que esta converja a cero. En concreto, para todo T se tiene que $\mathbb{V}(\sqrt{T}\bar{y}) = T\mathbb{V}(\bar{y}) = 1$. Tras estabilizar la varianza y mantener la media, que en todo caso es cero, $\sqrt{T}\bar{y} = \sqrt{T}(\bar{y} - \mathbb{E}(\bar{y}))$, se aprecia que mayores valores de T van redituando distribuciones cada vez más cercanas a la

Gráfico I.5 Ley de Grandes Números y Teorema del Límite Central



Notas: (Prog15_LeyesAsintoticas.prg) Se muestran densidades kernel (automáticas) de \bar{y} y $\sqrt{T}\bar{y}$ para un total $R = 100,000$ repeticiones.

normal estándar. En particular, se observa cómo a medida que T se incrementa, la asimetría en las distribuciones muestrales va reduciéndose y sus modas van aproximándose a $\mathbb{E}(y) = 0$. Este es el principal postulado del Teorema del Límite Central. Es importante enfatizar que mientras que la Ley Débil de Grandes Números alude a un límite probabilístico (una esperanza), el Teorema del Límite Central describe el comportamiento de la distribución límite de una variable (que no deja de ser) aleatoria (convergencia en distribución).

Los paneles (c) y (d) permiten reflexionar sobre el alcance del Teorema del Límite Central y, en particular, sobre la idoneidad del uso de la distribución normal como aproximación de la distribución muestral. En ambos paneles, los casos donde $a = 1$ corresponden a la distribución normal estándar predicha por el Teorema del Límite Central. En el panel (c)

se aprecia que para tamaños muestrales reducidos ($T = 25$ en este caso), no hay garantía de que las aproximaciones asintóticas sean satisfactorias. Esto es particularmente cierto cuando la distribución de y es lejana a la normal ($a = 4$ y $a = 6$), características que se transmiten a las distribuciones muestrales del estadístico de interés (el promedio). Por su parte, el panel (d) muestra cómo un mayor tamaño muestral (en este caso se pasa de $T = 25$ a $T = 100$) aminora los efectos de la no-normalidad y da respaldo empírico al Teorema del Límite Central. En resumen, cuando el tamaño de la muestra es lo suficientemente grande y las distribuciones de las que provienen los datos no son muy lejanas a la normal (por ejemplo, no son muy asimétricas), las aproximaciones asintóticas proveen un marco de inferencia adecuado. Cuán grande debe ser T depende de las características poblacionales de y y es, por tanto, una pregunta abierta. Por ejemplo, para $a \leq 2$, $T = 25$ parece ser razonable, mientras que $T = 100$ provee aproximaciones aceptables para $a \leq 4$.

El programa `Prog15_LeyesAsintoticas.prg` contiene elementos de programación para la implementación de experimentos de Monte Carlo complejos que serán utilizados en varias de las simulaciones presentadas a lo largo de este texto.

En particular, se obtienen resultados para valores cambiantes de T y de a . Partiendo del supuesto de que $R > T$, se opera sobre un *workfile* de tamaño `!R`, previamente definido por el usuario. Antes de iniciar los muestreos y de calcular los estadísticos muestrales correspondientes, se definen los vectores `Tvals` y `Avals`, de dimensión `!nTvals` y `!nAvals`, respectivamente, que contienen estos valores cambiantes. En el programa, estos vectores son llenados con la sintaxis

```
Avals.FILL 1, 2, 4, 6
```

pero pueden ser llenados “a mano” (por ejemplo, `Avals(3) = 4`) sin ninguna pérdida de generalidad.

Por su parte, también previamente al muestreo, se definen varias series para almacenar los resultados de las simulaciones para cada repetición. En series de nombre `DM{!a}_{!T}` se almacenan los R promedios muestrales \bar{y} para distintas combinaciones de a y T (de ahí la dependencia de la serie de los escalares `!a` y `!T`), mientras que las series de nombre `DA{!a}_{!T}` almacenan $\sqrt{T}\bar{y}$. Las distribuciones muestrales presentadas en el gráfico I.5(a) son histogramas (en estricto, densidades kernel) de `DM{!a}_{!T}`, mientras que las distribuciones (asintóticas) en el resto de paneles del gráfico I.5 provienen de `DA{!a}_{!T}`.

El cuerpo del programa contiene tres bucles: el primero para las repeticiones (`!i` que va hasta `!R`), el segundo para los diversos valores de a (`!ia` que va hasta `!nAvals`) y el tercero para los distintos valores de T (`!it` que va hasta `!nTvals`). Para cada repetición `!i`, se genera una serie `u` de números aleatorios provenientes de una distribución normal estándar. Esta serie contiene `!Tmax` observaciones, donde este escalar es definido previamente como el máximo valor del vector de tamaños muestrales `Tvals`. Luego, utilizando estas mismas realizaciones de `u` se calcula la serie `y`, también de tamaño `!Tmax`, que depende de `!a`, que

a su vez varía con el contador !ia. Vale la pena mencionar que en la generación de y se utilizan escalares temporales de la forma !Eu{!a} y !Eu{!a2} que han sido previamente definidos a partir de las propiedades (conocidas) de la distribución normal estándar⁸. Luego, se restringen los tamaños muestrales a SMPL 1 !T, donde !T varía con el contador !it. Así, dados un valor !a y una muestra de tamaño !T, se procede a calcular el promedio muestral !ybar y a almacenar los resultados en las series DM{!a}_{!T} y DA{!a}_{!T}. El proceso se repite !R veces:

```

FOR !i = 1 TO !R
  SMPL 1 !Tmax
  SERIES u = NRND

  FOR !ia = 1 TO !nAvals
    !a = Avals(!ia)
    !a2 = 2*!a
    SERIES y = (u^{!a} - !Eu{!a})/@SQRT(!Eu{!a2} - !Eu{!a}^2)

    FOR !it = 1 TO !nTvals
      !T = Tvals(!it)
      SMPL 1 !T
      !ybar = @MEAN(y)
      DM{!a}_{!T}(!i) = !ybar
      DA{!a}_{!T}(!i) = @SQRT(!T)*!ybar
    NEXT !it
  NEXT !ia
NEXT !i

```

El gráfico I.5 muestra una selección de todos los resultados obtenidos con este programa.

⁸ En particular, se tiene que $\mathbb{E}(u^a) = 0$ cuando a es impar, mientras que $\mathbb{E}(u^a) = (a-1)(a-3)\cdots 3\cdot 1$ cuando a es par. Así, !Eu1 = 0, !Eu2 = 1, !Eu4 = 3, !Eu6 = 15, !Eu8 = 105, !Eu10 = 945 y !Eu12 = 10395.

Series de tiempo estacionarias

Una serie de tiempo es una colección de variables aleatorias

$$\mathcal{Y}_t = \{ \dots, y_{-1}, y_0, y_1, y_2, \dots, y_{t-2}, y_{t-1}, y_t \}$$

ordenadas cronológicamente y que, en general, guardan cierta dependencia. En principio, cada variable aleatoria en la colección \mathcal{Y}_t es tan solo una realización de un proceso estocástico. Por ejemplo, considere una variable aleatoria cuya función de distribución es $f_t(\cdot)$; al tomar un número al azar proveniente de $f_t(\cdot)$ se consigue y_t . El interés se centra en inferir sobre los momentos poblacionales de $f_t(\cdot)$. Sin embargo, ello no es posible con este nivel de generalidad, ya que se requiere describir las propiedades de, por ejemplo, $\mathbb{E}(y_t)$ observando únicamente y_t (el tamaño de la muestra en el período t es igual a 1), lo que lleva a cuestionamientos sobre nuestra capacidad de inferir con niveles mínimos de confiabilidad. Es preciso contar con condiciones adicionales para el análisis, y es en este punto donde el concepto de estacionariedad es sumamente útil.

Se dice que una serie y_t es un proceso **estacionario en covarianzas** cuando, para todo t , ocurre que

$$\begin{aligned} \mathbb{V}(y_t) &= \mathbb{E}((y_t - \mu_t)^2) \equiv \gamma_0, \\ \mathbb{C}(y_t, y_{t-i}) &= \mathbb{E}((y_t - \mu_t)(y_{t-i} - \mu_{t-i})) \equiv \gamma_i, \end{aligned}$$

donde $\mu_t = \mathbb{E}(y_t)$. Es decir, la varianza y las **autocovarianzas** de y_t **no dependen de t** . Como se aprecia, la autocovarianza γ_i , que es la medida de dependencia que se utilizará a lo largo de este texto, es la covarianza entre y_t e y_{t-i} y depende únicamente de i . Dado que γ_i no depende de t , presenta la siguiente simetría:

$$\gamma_i = \mathbb{C}(y_t, y_{t-i}) = \mathbb{C}(y_s, y_{s-i}) = \mathbb{C}(y_{t+i}, y_t) = \gamma_{-i}.$$

La autocovarianza, como la varianza, se expresa en las unidades de medida de y_t , al cuadrado. Una versión estandarizada (sin dimensiones, un número puro) es la **autocorrelación**, que no es más que el coeficiente de correlación entre y_t e y_{t-i} y es igual a $\rho_i = \gamma_i/\gamma_0$. Se cumple que $\rho_0 = 1$, $|\rho_i| \leq 1$ y $\rho_i = \rho_{-i}$. Cuanto más cerca está $|\rho_i|$ de 1, mayor el grado de dependencia.

Note que, hasta el momento, la media $\mathbb{E}(y_t) = \mu_t$ podría variar con t . Cuando se cumple, además, que $\mathbb{E}(y_t) = \mu$ para todo t , se tiene que y_t es un proceso **estacionario en media**.

Una implicancia de primera importancia de la estacionariedad en covarianzas es que los **momentos poblacionales pueden inferirse a partir de una muestra** del tipo

$\{y_1, y_2, \dots, y_t, \dots, y_{T-1}, y_T\}$, ya que todas estas realizaciones de la serie de tiempo comparten estos momentos. En otras palabras, la estacionariedad permite estudiar los **momentos no condicionales** de una serie de tiempo. En adelante, utilizaremos el término “estacionariedad” como sinónimo de “estacionariedad en covarianzas” (también conocida como “estacionariedad débil” o “estacionariedad de segundo orden”).

Otro concepto importante es el de **ergodicidad**, ya que la teoría asintótica de series de tiempo provee resultados útiles para procesos ergódicos. Los procesos estacionarios y los procesos ergódicos comparten una propiedad relevante: **memoria restringida** (a veces denominada *mixing*), que indica que las dependencias temporales inherentes en la serie de tiempo van desvaneciéndose para eventos cronológicamente muy distanciados o, en otras palabras, que el presente no contiene información acerca del pasado remoto o del futuro lejano. Más formalmente, el proceso $\{y_t\}$ es ergódico si para dos funciones acotadas $h : \mathbb{R}^L \rightarrow \mathbb{R}$ y $g : \mathbb{R}^K \rightarrow \mathbb{R}$,

$$\lim_{s-t \rightarrow \infty} \mathbb{E}(h(y_t, \dots, y_{t+L})g(y_s, \dots, y_{s+K})) = \mathbb{E}(h(y_t, \dots, y_{t+L}))\mathbb{E}(g(y_s, \dots, y_{s+K})).$$

En palabras, el proceso presenta memoria restringida si dos variables aleatorias que dependen de segmentos de la serie de tiempo lo suficientemente alejados son independientes.

Las condiciones para la estacionariedad de los procesos estocásticos estudiados en este texto son idénticas a las requeridas para la ergodicidad. Por ello, quizá la distinción entre un proceso estacionario y uno ergódico no será muy evidente, ni muy necesaria, en adelante. En particular, una condición suficiente para la ergodicidad es que las autocovarianzas de $\{y_t\}$ sean **absolutamente sumables**:

$$\sum_{i=0}^{\infty} |\gamma_i| = M \text{ finito.}$$

A su vez, una condición necesaria (aunque no suficiente) para garantizar la sumabilidad absoluta de las autocovarianzas es que γ_i converja a cero **lo suficientemente rápido** conforme i se incrementa. Las autocovarianzas son absolutamente sumables para los procesos estacionarios que a continuación analizamos.

II.1 Generando dependencias: procesos ARMA

Una clase muy importante, y dominante en el análisis de series de tiempo, de procesos lineales son los denominados procesos ARMA. La característica saltante de este tipo de procesos es que el patrón de dependencia, tal y como es medido por las autocovarianzas, depende de un grupo reducido de parámetros.

El punto de partida es el proceso denominado **ruido blanco**: $\{\varepsilon_t\}$ es un ruido blanco si es estacionario en covarianzas (usualmente se asume que su media es cero) y no está

APUNTES DE ESTUDIO

autocorrelacionado. Formalmente, para todo t ,

$$\begin{aligned}\mathbb{E}(\varepsilon_t) &= 0 \\ \mathbb{E}(\varepsilon_t^2) &= \sigma^2 \\ \mathbb{E}(\varepsilon_t \varepsilon_{t-i}) &= 0 \text{ para cualquier } i \neq 0.\end{aligned}$$

Es bueno mencionar que las variables analizadas en el capítulo I, provenientes de muestreos idénticos e independientes, son ruidos blancos.

La manera más simple de generar un proceso estacionario con dependencia es calculando promedios móviles de un ruido blanco. Es decir, y_t se forma a partir de combinaciones lineales de $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$, dando origen a un proceso **MA(q)** (por las siglas de *moving average*). Para ilustrar el funcionamiento de los modelos MA, considere el proceso MA(2)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

Dado que $\mathbb{E}(\varepsilon_t) = 0$ para todo t , se concluye que $\mathbb{E}(y_t) = \mu$ para todo t . Para calcular las autocovarianzas, es conveniente notar que la i -ésima autocovarianza será igual a cero a menos que la misma realización ε_t esté presente tanto en y_t como en y_{t-i} . Note que

$$\begin{aligned}y_t - \mu &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \\ y_{t-1} - \mu &= \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} \\ y_{t-2} - \mu &= \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} \\ y_{t-3} - \mu &= \varepsilon_{t-3} + \theta_1 \varepsilon_{t-4} + \theta_2 \varepsilon_{t-5},\end{aligned}$$

por lo que las autocovarianzas de orden mayor de 2 son cero. Dado $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ para todo t , se verifica que

$$\gamma_0 = \sigma^2(1 + \theta_1^2 + \theta_2^2), \quad \gamma_1 = \sigma^2\theta_1(1 + \theta_2), \quad \gamma_2 = \sigma^2\theta_2 \quad \text{y} \quad \gamma_i = 0 \quad \text{para } i \geq 3.$$

Consecuentemente, las autocorrelaciones, que no dependen de σ^2 , son

$$\rho_1 = \frac{\theta_1(1 + \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad \text{y} \quad \rho_i = 0 \quad \text{para } i \geq 3.$$

Un proceso MA es **siempre** estacionario: la media, la varianza y las autocovarianzas de y_t no dependen de t . Más aún, se aprecia que el proceso presenta memoria corta: la dependencia entre una observación de la serie de tiempo y sus valores pasados, medida a través de las autocovarianzas, desaparece después de $q = 2$ períodos.

Otra forma de generar dependencia temporal es a través de **ecuaciones en diferencias estocásticas**. Ello da origen a los procesos **autorregresivos** o **AR**. Un proceso autorregresivo de orden p , denotado como **AR(p)**, establece una relación lineal entre una variable aleatoria y sus p rezagos más una perturbación aleatoria (un ruido blanco)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

El proceso AR(p) será estacionario siempre que m_t , la solución a la ecuación en diferencias homogénea asociada

$$m_t - \phi_1 m_{t-1} - \phi_2 m_{t-2} - \phi_3 m_{t-3} - \dots - \phi_p m_{t-p} = 0$$

(que se consigue al eliminar de la ecuación en diferencias todo término que no sea y_t o sus rezagos), sea convergente. La convergencia de m_t está estrechamente vinculada al comportamiento de las **raíces del polinomio característico**

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 - \dots - \phi_p z^p,$$

es decir, de los valores z^* tales que $\Phi(z^*) = 0$. En particular, admitiendo la posibilidad de que estas raíces sean números complejos, si el **módulo** de todo z^* es mayor de uno, lo que usualmente se frasea como que z^* se encuentra **fuera del círculo unitario**, entonces y_t es un proceso estacionario.

Un caso ilustrativo es el modelo AR(1)

$$y_t = c + \phi y_{t-1} + \varepsilon_t.$$

Dado que $z^* = \phi^{-1}$, el proceso es estacionario si $|\phi| < 1$. Tras sustituciones sucesivas (partiendo de un punto arbitrario que ocurre en el pasado remoto), se consigue

$$y_t = \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i},$$

que es la **representación MA(∞)** del proceso AR(1). De ello, puede concluirse que $\mathbb{E}(y_t) = c(1 - \phi)^{-1} \equiv \mu$ para todo t . Además, para $i \geq 0$, se verifica que

$$\gamma_i = \frac{\sigma^2 \phi^i}{1 - \phi^2} \quad \text{y} \quad \rho_i = \frac{\gamma_i}{\gamma_0} = \phi^i.$$

Al ser $|\phi| < 1$, las autocovarianzas y autocorrelaciones son exponencialmente decrecientes y convergen a 0 para $i \rightarrow \infty$. Note que las autocovarianzas y las autocorrelaciones pueden determinarse recursivamente, recordándonos a la ecuación en diferencias homogénea asociada con el proceso autorregresivo: tomando a γ_0 y $\rho_0 = 1$ como condiciones iniciales, se cumple que $\gamma_i = \phi \gamma_{i-1}$ y $\rho_i = \phi \rho_{i-1}$ para $i \geq 1$.

Finalmente, los procesos ARMA son híbridos o **mixtos**. Al igual que un proceso AR, se trata de una ecuación en diferencias estocástica. No obstante, a diferencia de un modelo AR puro, el término estocástico no es un ruido blanco sino que es, en sí mismo, un proceso dependiente. El proceso **ARMA(p, q)** incluye un término autorregresivo de orden p y un término de promedios móviles de orden q ,

$$y_t = c + \phi_1 y_t + \phi_2 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}.$$

APUNTES DE ESTUDIO

Dado que un proceso MA es siempre estacionario, las condiciones de estacionariedad de un proceso ARMA(p, q) son las mismas que las de un proceso AR(p): las raíces de $\Phi(z^*) = 0$ deben caer fuera del círculo unitario.

En el caso del proceso ARMA(1,1), se tiene que

$$y_t = c + \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},$$

donde $|\phi| < 1$ asegura la estacionariedad. Tras sustituciones sucesivas, se consigue

$$y_t = \frac{c}{1-\phi} + \varepsilon_t + (\phi + \theta) \sum_{i=1}^{\infty} \phi^{i-1} \varepsilon_{t-i},$$

que es la **representación MA(∞)** del proceso ARMA(1,1). La media del proceso es, nuevamente, $\mu = c(1-\phi)^{-1}$. Asimismo, las autocovarianzas y las autocorrelaciones pueden determinarse recursivamente, siguiendo la dinámica gobernada por la ecuación en diferencias homogénea asociada con el proceso mixto. Sin embargo, la presencia de un término MA(1) altera las condiciones iniciales en estas recursiones. En este caso, tomando a γ_1 y ρ_1 como condiciones iniciales, se cumple que $\gamma_i = \phi \gamma_{i-1}$ y $\rho_i = \phi \rho_{i-1}$ para $i \geq 2$. Estas condiciones iniciales satisfacen

$$\gamma_0 = \sigma^2 \frac{(1 + 2\phi\theta + \theta^2)}{1 - \phi^2} \quad \text{y} \quad \gamma_1 = \sigma^2 \frac{(\phi + \theta)(1 + \phi\theta)}{1 - \phi^2},$$

por lo que

$$\rho_1 = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}.$$

Los procesos ARMA capturan diversos patrones de dependencia temporal, que no pueden ser generados con especificaciones exclusivamente autorregresivas o de medias móviles.

Claramente, el modelo ARMA(1,1) generaliza el modelo MA(1), al incorporar persistencia en el patrón de autocovarianzas. Estas se desvanecen en el límite ($i \rightarrow \infty$), a diferencia del caso MA(1), en donde $\gamma_i = 0$ para todo $i > 1$. Asimismo, el proceso ARMA generaliza el proceso autorregresivo. Así como pudo derivarse una representación MA(∞) bajo la condición de estacionariedad $|\phi| < 1$, es posible también encontrar una representación AR(∞) bajo la condición de **invertibilidad** $|\theta| < 1$. A saber, tras sustituciones sucesivas,

$$y_t = \frac{c}{1+\theta} + (\phi + \theta) \sum_{i=1}^{\infty} (-\theta)^{i-1} y_{t-i} + \varepsilon_t,$$

que colapsa al modelo AR(1) cuando $\theta = 0$. Este resultado lleva a una conclusión relevante para la discusión posterior: un modelo ARMA podría ser aproximado por un modelo AR(p) de orden superior, donde p podría ser grande si es que $|\theta|$ es cercano a uno. En efecto, una de las ventajas de los procesos híbridos es que pueden capturar patrones de correlación complejos de una manera muy **parsimoniosa**, es decir con pocos parámetros.

II.2 Leyes asintóticas importantes, revisadas

Considere una muestra $\{y_1, y_2, \dots, y_t, \dots, y_{T-1}, y_T\}$ de T observaciones. En la sección 1.6 se presentaron dos resultados asintóticos sumamente importantes: la Ley Débil de Grandes Números y el Teorema del Límite Central, que describían el comportamiento del promedio \bar{y} de todas las observaciones en la muestra conforme T se incrementaba. Estos resultados, que son el punto de partida de ejercicios de inferencia en muestras grandes, son válidos para muestreos independientes. En el contexto de una serie de tiempo, el supuesto de independencia es indefendible. No obstante, pueden obtenerse leyes asintóticas muy similares para procesos estacionarios de memoria restringida o procesos ergódicos. Así, la teoría "clásica" de inferencia, basada esencialmente en aproximaciones normales, resulta ser completamente válida para muestreos con este tipo de dependencia.

Es instructivo explorar la estructura de la varianza del promedio muestral. Se tiene que

$$\begin{aligned}
 \mathbb{V}(\bar{y}) &= \frac{1}{T^2} \mathbb{V} \left(\sum_{t=1}^T y_t \right) = \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \mathbb{C}(y_t, y_s) \\
 &= \frac{1}{T^2} \left[\sum_{t=1}^T \mathbb{C}(y_t, y_1) + \sum_{t=1}^T \mathbb{C}(y_t, y_2) + \dots + \sum_{t=1}^T \mathbb{C}(y_t, y_T) \right] \\
 &= \frac{1}{T^2} [(\gamma_0 + \gamma_1 + \dots + \gamma_{T-2} + \gamma_{T-1}) + (\gamma_1 + \gamma_0 + \gamma_1 + \dots + \gamma_{T-2}) + \dots \\
 &\quad \dots + (\gamma_{T-1} + \gamma_{T-2} + \dots + \gamma_1 + \gamma_0)] \\
 &= \frac{1}{T^2} [T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \dots + 2\gamma_{T-1}] \\
 &= \frac{\gamma_0}{T} + \frac{2}{T} \sum_{s=1}^{T-1} \left(1 - \frac{s}{T}\right) \gamma_s.
 \end{aligned}$$

En un muestreo independiente (por ejemplo, si y_t fuera un ruido blanco), únicamente el primer término sería distinto de cero. Dado que $\gamma_s \leq |\gamma_s|$ y que $s \leq T$, se verifica que

$$\mathbb{V}(\bar{y}) \leq \frac{\gamma_0}{T} + \frac{2}{T} \sum_{s=1}^{T-1} \left(1 - \frac{s}{T}\right) |\gamma_s| \leq \frac{\gamma_0}{T} + \frac{2}{T} \sum_{s=1}^{T-1} |\gamma_s|.$$

Dado que las autocovarianzas son absolutamente sumables, se concluye que

$$\lim_{T \rightarrow \infty} \mathbb{V}(\bar{y}) \leq \lim_{T \rightarrow \infty} \frac{\gamma_0}{T} + \left(\lim_{T \rightarrow \infty} \frac{2}{T} \right) \sum_{s=1}^{\infty} |\gamma_s| = 0.$$

Se concluye que $\mathbb{V}(\bar{y}) \rightarrow 0$ a medida que $T \rightarrow \infty$. Aplica, luego, la Ley de Grandes Números conocida como el **Teorema Ergódico**,

$$\bar{y} - \mathbb{E}(\bar{y}) \xrightarrow{p} 0.$$

APUNTES DE ESTUDIO

Este resultado es, sin duda, una muy importante generalización del Teorema de Khinchine. Acá, se permite que exista dependencia temporal, toda vez que esta sea limitada y se diluya “en el largo plazo”¹.

Asimismo, se dispone de una versión del Teorema del Límite Central (el **Teorema de Gordin**) adecuada para muestreos con dependencia. Antes de presentarla, conviene definir la **varianza de largo plazo** de la serie y_t . Esta es el límite de $\mathbb{V}(\sqrt{T}\bar{y})$. Explícitamente,

$$\Omega = \lim_{T \rightarrow \infty} T \mathbb{V}(\bar{y}) = \lim_{T \rightarrow \infty} \left[\gamma_0 + 2 \sum_{s=1}^{T-1} \left(1 - \frac{s}{T}\right) \gamma_s \right] = \gamma_0 + 2 \sum_{s=1}^{\infty} \gamma_s = \sum_{s=-\infty}^{\infty} \gamma_s,$$

donde la última igualdad usa simetría $\gamma_{-s} = \gamma_s$. Es decir, Ω es la suma de todas las autocovarianzas. Para un proceso ergódico, Ω es finita. Así²,

$$\sqrt{T}(\bar{y} - \mathbb{E}(\bar{y})) \xrightarrow{d} N(0, \Omega).$$

El programa Prog21_LeyesRevisadas.prg implementa una serie de experimentos de Monte Carlo con el propósito de ilustrar el funcionamiento de estas leyes asintóticas bajo dependencia. La idea del ejercicio es muy similar a la estudiada en la sección I.6 (programa Prog15_LeyesAsintoticas.prg), aunque con particularidades que se discuten enseguida.

Una primera distinción se encuentra en cómo se generan los datos. Estos ya no provienen de muestreos independientes, sino que poseen patrones de dependencias parametrizados como en la familia ARMA. En primer lugar, se requiere de una secuencia de ruidos blancos, generada de la manera usual, cuyas realizaciones se almacenan en la serie e . Luego, se generan los procesos dependientes con los valores en e . Para el caso de un proceso MA(1), se tiene que

```
GENR y1 = !mu + e + !theta*e(-1),
```

donde $!mu$ es un escalar predefinido que contiene el valor de μ , la media no condicional de y_1 , mientras que $!theta$ es también un escalar especificado previamente con el valor de θ . EVIEWS interpreta la serie $z(-!k)$ como el $!k$ -ésimo rezago de la serie z .

Para los procesos AR(1) y ARMA(1,1) se tiene, respectivamente,

```
GENR y2 - !phi*y2(-1) = (1 - !phi)*!mu + e
```

```
GENR y3 - !phi*y3(-1) = (1 - !phi)*!mu + e + !theta*e(-1),
```

¹ La generalización es, en estricto, más completa aún, ya que el modo de convergencia en el Teorema Ergódico es en media cuadrática (*mean squared*). La convergencia en media cuadrática implica convergencia en probabilidad, pero lo converso no ocurre necesariamente.

² Tanto el Teorema Ergódico como el Teorema de Gordin se refieren a $\bar{y} - \mathbb{E}(\bar{y})$, que es una variable aleatoria con media igual a cero. Estos planteamientos admiten casos en donde $\mathbb{E}(y_t) = \mu_t$, es decir y_t podría no ser estacionario en media. En estricto, en estas circunstancias el término $\mathbb{E}(\bar{y})$ debe reemplazarse por el límite del promedio de μ_t a lo largo de t (que, incluso, podría no estar bien definido).

donde ϕ contiene el valor de ϕ . El intercepto c en estos procesos es igual a $c = (1 - \phi)\mu$ para así asegurar que, sin importar cuál sea el valor de $|\phi| < 1$, la media no condicional de y_2 e y_3 sea también igual a μ .

Esta manera de generar procesos dependientes con componentes autorregresivos explota aspectos sumamente convenientes del lenguaje de programación de EViews. En particular, note que tanto y_2 como y_3 se generan **implícitamente** de acuerdo con las ecuaciones en diferencias que definen, respectivamente, a los procesos AR(1) y ARMA(1,1). En efecto, al encontrar EViews expresiones que involucran a una variable con sus rezagos en una misma línea de comando, el paquete resuelve interna y recursivamente la ecuación en diferencias.

No obstante, es importante ser cuidadosos con las condiciones que se utilizan para inicializar las recursiones. Considere que la muestra activa del *workfile* es `SMPL 1 !obs` al ejecutar, por ejemplo, el comando de generación del proceso AR(1). Dado que se requiere información de la observación 0, que es la inmediatamente anterior a la primera observación de la muestra activa, EViews arrojará un mensaje de error revelando una incompatibilidad, puesto que la primera observación válida en un *workfile* es la número 1. En cambio, si la muestra activa es `SMPL 2 !obs`, la recursión toma como condición inicial el valor de y en la observación 1. En general, para un proceso con p rezagos, la muestra activa debe iniciarse, como mínimo, en la observación $p + 1$.

Asimismo, si alguno de los valores iniciales que se utilizan en las recursiones es un valor omitido, NA en el lenguaje de EViews, es altamente probable que este *missing value* proliferare y que el resultado de resolver la recursión sea una serie llena de NA (EViews también arrojará un error en este caso). Para ilustrar, considere una serie gobernada por el proceso AR(1), con media igual a μ y donde la primera observación es y_1 , independiente del resto. Tras resolver la recursión $y_s = (1 - \phi)\mu + \phi y_{s-1} + \varepsilon_s$ se consigue, para $s > 1$,

$$y_s = \phi^{s-1}(y_1 - \mu) + \mu + \varepsilon_s + \phi\varepsilon_{s-1} + \phi^2\varepsilon_{s-2} + \dots + \phi^{s-3}\varepsilon_3 + \phi^{s-2}\varepsilon_2.$$

Se aprecia que si $y_1 = \text{NA}$, entonces $y_s = \text{NA}$ para todo s .

Un segunda consideración es el efecto que las condiciones iniciales podrían tener sobre los resultados. Teóricamente, se sobreentiende que los procesos ARMA han sido inicializados en el pasado remoto y, dado que poseen memoria restringida, **no se ven afectados por sus valores iniciales**. De lo contrario, considerando que observaciones recientes son más importantes al determinar el presente, el proceso presentaría una no estacionariedad: las condiciones iniciales influirían de manera diferenciada sobre las observaciones iniciales y finales del proceso, generando así dependencia temporal en los momentos de la serie.

Por ejemplo, los dos primeros momentos del proceso y_s generado previamente son

$$\mathbb{E}(y_s) = \mu + \phi^{s-1} [\mathbb{E}(y_1) - \mu] \quad \text{y} \quad \mathbb{V}(y_s) = \gamma_0 + \phi^{2(s-1)} [\mathbb{V}(y_1) - \gamma_0],$$

donde $\gamma_0 = \sigma^2/(1 - \phi^2)$. Estas propiedades difieren de las de un proceso AR(1) estacionario, cuya esperanza es μ y su varianza es γ_0 . En especial, y_s no es estacionaria pues tanto $\mathbb{E}(y_s)$

APUNTES DE ESTUDIO

como $\mathbb{V}(y_s)$ dependen de s , dependencia originada en las potenciales discrepancias del valor inicial con el resto de valores de la serie de tiempo. Pero ¿cómo lidiar con esta complicación? Una solución directa es elegir y_1 tal que $\mathbb{E}(y_1) = \mu$ y $\mathbb{V}(y_1) = \gamma_0$, lo que inmediatamente elimina la no estacionariedad. No obstante, si bien es cierto que la solución se ve simple para un proceso AR(1), podría ser engorrosa para procesos ARMA(p, q) algo más complejos.

Alternativamente, dado que $|\phi| < 1$, cuando s es grande $\phi^s \simeq 0$, por lo que $\mathbb{E}(y_s) \simeq \mu$ y $\mathbb{V}(y_s) \simeq \gamma_0$, siendo y_1 completamente arbitrario, e y_s comienza a comportarse como un proceso estacionario. Esta es la idea detrás de la construcción teórica del proceso iniciado en el pasado remoto. En términos de la simulación y generación de realizaciones de un proceso estacionario, este resultado inspira la práctica común de generar $T + T_0$ observaciones, **descartar las T_0 primeras observaciones** y retener las T últimas observaciones para el análisis. Cuando T_0 es razonablemente grande, puede considerarse que las T observaciones efectivamente utilizadas provienen del proceso estacionario de interés. Por ejemplo, si $\phi = 0.75$ y $T_0 = 50$, entonces $|\phi^s| < 10^{-6}$ para $s \geq T_0$. El argumento se ilustra con el caso AR(1), pero es general y válido para todo proceso con memoria restringida.

El programa Prog21_LeyesRevisadas.prg atiende todas estas consideraciones. El cuerpo del programa es el siguiente:

```
SMPL 1 1
SERIES y1 = !mu
SERIES y2 = !mu
SERIES y3 = !mu
FOR !i = 1 TO !R
  SMPL 1 !Tmax+!T0
  SERIES e = @RTDIST(3)/@SQRT(3)
  SMPL 2 !Tmax+!T0
  GENR y1          = !mu + e + !theta*e(-1)
  GENR y2 - !phi*y2(-1) = (1 - !phi)*!mu + e
  GENR y3 - !phi*y3(-1) = (1 - !phi)*!mu + e + !theta*e(-1)
  FOR !it = 1 TO !nTvals
    !T = Tvals(!it)
    SMPL !T0+1 !T0+!T
    FOR !a = 1 TO 3
      !ybar = @MEAN(y{!a})
      DM{!a}_{!T}(!i) = !ybar
      DA{!a}_{!T}(!i) = @SQRT(!T)*(!ybar - !mu)/!sLP{!a}
    NEXT !a
  NEXT !it
NEXT !i
```

Antes de iniciar los muestreos, las series y_1 , y_2 e y_3 son declaradas y, en cada caso, se fija la primera observación en el valor de la media no condicional. Este valor se mantiene inalterado a lo largo de repeticiones. Luego, se desarrollan tres bucles: el primero para las repeticiones (i , que va hasta R), el segundo para distintos valores del tamaño de la muestra T (i , que va hasta $nTvals$ y que toma los valores predefinidos en el vector $Tvals$), y el tercero varía el índice a para evaluar los tres procesos generados³.

Para cada repetición i , el ruido blanco almacenado en la serie e se distribuye como una variable t de Student con 3 grados de libertad, estandarizada para que su varianza sea igual a $\sigma^2 = 1$; de ahí que `@RTDIST(3)` es dividido entre su desviación estándar, `@SQRT(3)`. Como se ha discutido, este tipo de distribuciones se caracteriza por la frecuente ocurrencia de valores extremos, un comportamiento muy lejano al de variables normalmente distribuidas. El propósito es incorporar en el diseño del experimento no normalidades, para así ver operar al Teorema del Límite Central. La serie e contiene $Tmax+T0$ observaciones, donde $Tmax$ es un escalar definido previamente como el máximo valor del vector de tamaños muestrales $Tvals$, y $T0$ es el número de observaciones iniciales por descartarse. Luego, utilizando estas mismas realizaciones de e , se calculan las series y_1 , y_2 e y_3 , también de tamaño $Tmax+T0$. Note que las series son generadas bajo la muestra activa `SMPL 2 Tmax+T0`, tomando así como dada la primera observación de cada serie.

Una vez generadas las series, se restringen las muestras a `SMPL T0+1 T0+T`, donde T varía con el contador i . Es en esta indicación que se descartan las observaciones iniciales para diluir los posibles efectos de las condiciones iniciales. Así, dada una muestra de tamaño T , se procede a calcular el promedio muestral \bar{y} y a almacenar los resultados en las series `DM{a}_{T}` y `DA{a}_{T}` (previamente declaradas). El procedimiento se repite R veces. En las series de nombre `DM{a}_{T}` se almacenan los R promedios muestrales \bar{y} de los 3 procesos bajo estudio para una muestra de tamaño T (de ahí la dependencia de la serie de los escalares a y T), mientras que las series de nombre `DA{a}_{T}` almacenan el estadístico

$$\bar{z} = \sqrt{T} \left(\frac{\bar{y} - \mu}{\sqrt{\Omega}} \right),$$

donde $\sqrt{\Omega}$ es la desviación de largo plazo del proceso. Se sabe que para un proceso ARMA(p, q) de la forma

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

la varianza de largo plazo adopta una forma relativamente simple:

$$\Omega = \sigma^2 \left(\frac{1 + \theta_1 + \theta_2 + \dots + \theta_q}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \right)^2.$$

³ EViews, tradicionalmente, utilizaba el comando `SERIES` para generar objetos series definidos de manera explícita (por ejemplo, `SERIES x = log(z)`), mientras que se reservaba el comando `GENR` para series generadas de modo implícito, `GENR exp(x) = z`. En versiones recientes del paquete, desde la versión 7, los comandos `SERIES` y `GENR` son absolutamente equivalentes.

Por ello, considerando que se mantiene $\sigma^2 = 1$ en todas las simulaciones, los escalares α son previamente definidos como $\alpha = (1 + \theta)$ para el proceso MA(1), $\beta = 1/(1 - \phi)$ para el proceso AR(1), y $\alpha\beta$ para el proceso mixto ARMA(1,1).

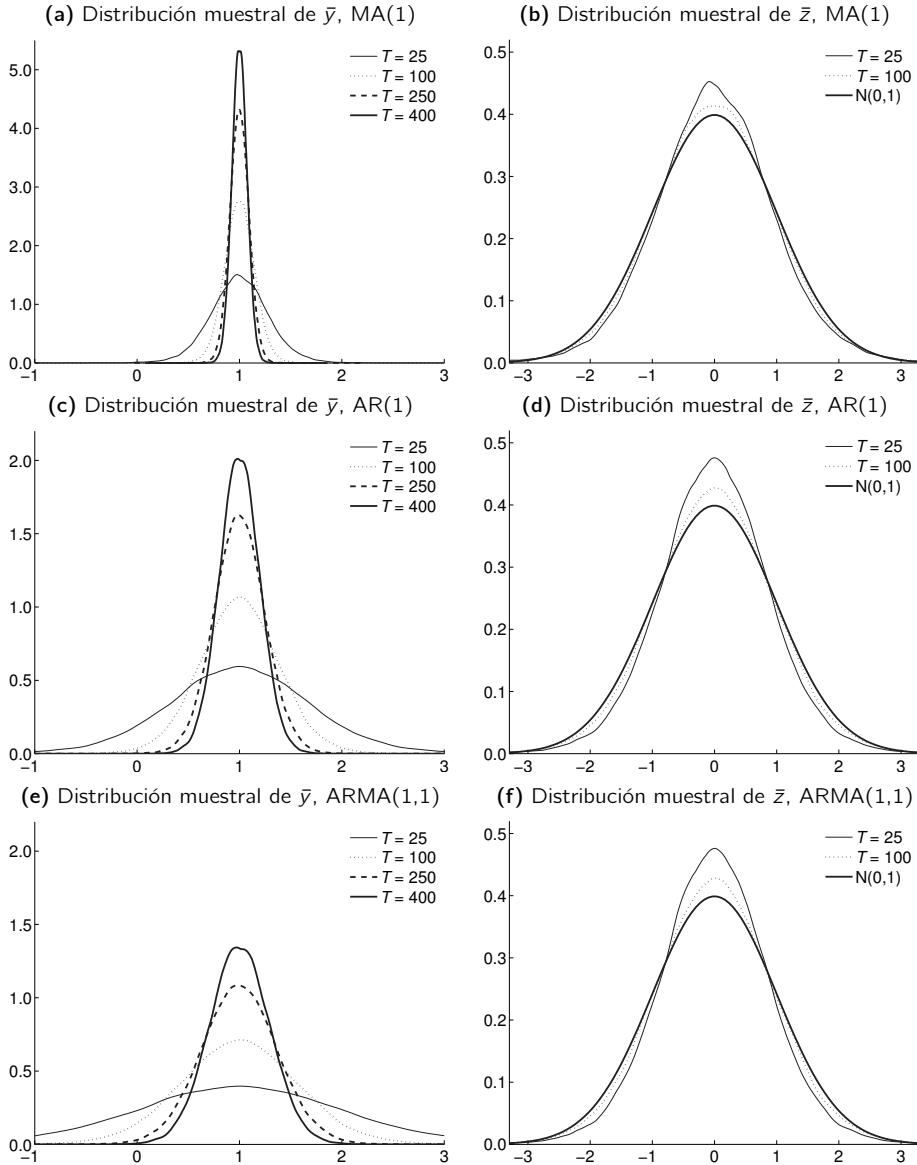
El gráfico II.1 presenta los resultados para $\sigma^2 = 1$, $\theta = 0.5$, $\phi = 0.75$ (tal que los procesos con componentes autorregresivos son estacionarios) y $\mu = 1$, donde se han eliminado las primeras $T_0 = 50$ observaciones al generar los datos. Se utilizan $R = 100,000$ repeticiones. Los paneles (a), (c) y (e) presentan la distribución muestral del promedio \bar{y} conforme T se incrementa. En estos paneles se aprecia la Ley de Grandes Números operando. En particular, cómo la distribución muestral de \bar{y} va concentrándose cada vez más en torno a μ , su límite probabilístico, hasta colapsar a este punto.

Es interesante notar que el comportamiento del proceso de convergencia descrito en el panel (a), que corresponde a un proceso MA(1), es algo distinto al descrito en los paneles (c) y (e), que corresponden a los procesos AR(1) y ARMA(1,1), respectivamente. Si bien es cierto que en los tres casos la Ley de Grandes Números aplica, la convergencia pareciera ser más lenta cuando el proceso presenta un componente autorregresivo. La explicación de este resultado tiene que ver con el valor de ϕ utilizado, específicamente con el hecho de que los procesos AR y ARMA en las simulaciones son **más persistentes** (las autocovarianzas tardan más en converger a cero) que el proceso MA. Por ejemplo, el valor de $\phi = 0.75$ indica que la autocorrelación del proceso AR(1) es $\rho_i > 0.20$ para todo $i \leq 5$, y $\rho_i > 0.10$ para todo $i \leq 8$, mientras que en el proceso MA(1) $\rho_i = 0$ para todo $i > 1$. Dado un mismo valor de σ^2 y de T , se puede concluir que la varianza del promedio muestral, que involucra a las autocovarianzas hasta el $(T - 1)$ -ésimo orden, es mayor cuanto más persistente es la serie promediada.

Intuitivamente, ocurre lo siguiente. La muestra $\{y_1, y_2, y_3\}$ contiene 3 observaciones sobre variables aleatorias cuya media es igual a μ . Así, cada observación contiene información de μ que finalmente es resumida en el promedio \bar{y} . A mayor información, menor varianza. En un proceso con poca persistencia, la dependencia entre observaciones es débil. De hecho, en el caso MA(1) y_3 ya no depende de y_1 . Ello significa que la información contenida en y_3 es completamente nueva, respecto a lo observado en y_1 . Puesto de otro modo, da lo mismo predecir y_3 con información sobre y_1 que sin ella. En cambio, en un proceso con mucha persistencia, la dependencia entre observaciones es fuerte y mucha de la información contenida en y_3 se encuentra también contenida en y_1 . Contar con y_1 es, en este caso, sumamente relevante para predecir y_3 . En otras palabras, la información **adicional** que otorga una observación extra para refinar nuestras inferencias respecto a μ , es decreciente en el nivel de persistencia de la serie. Así, la reducción en la varianza de \bar{y} , que es la fuerza detrás de la convergencia predicha por la Ley de Grandes Números, tenderá a ser menor para procesos persistentes.

Una manera más formal de pensar en el fenómeno descrito es la siguiente. La varianza de \bar{y} es aproximadamente igual a Ω/T . Para un proceso MA(1) con T_{MA} observaciones, esta

Gráfico II.1 Leyes asintóticas para procesos dependientes



Notas: (Prog21_LeyesRevisadas.prg) Densidades kernel (automáticas) de \bar{y} y $\bar{z} = \sqrt{T/\Omega}(\bar{y} - \mu)$. En las simulaciones, se utilizan $R = 100,000$ repeticiones y los valores $\sigma^2 = 1$, $\theta = 0.5$, $\phi = 0.75$ y $\mu = 1$. La escala del eje vertical de los paneles (c) y (e) difiere de la del panel (a).

expresión es $\sigma^2(1 + \theta)^2/T_{MA}$, mientras que para un proceso AR(1) con T_{AR} observaciones, la expresión correspondiente es $\sigma^2(1 - \phi)^{-2}/T_{AR}$. Al igualar ambas varianzas, se obtiene que $T_{AR} = T_{MA}(1 - \phi)^{-2}(1 + \theta)^{-2} \equiv T_{MA}A(\phi, \theta)$. Esto quiere decir que se necesitan $T_{MA}A(\phi, \theta)$ observaciones para que el promedio muestral de un proceso AR(1) tenga (aproximadamente) la misma varianza que el promedio muestral de un proceso MA(1) comparable. El factor $A(\phi, \theta)$ es creciente en ϕ , que es el parámetro que gobierna la persistencia del proceso AR(1). Para los valores utilizados en la simulación, $\phi = 0.75$ y $\theta = 0.5$, $A(\phi, \theta) \simeq 7$: cada observación adicional del proceso MA(1) reduce la varianza del mismo modo que 7(!) observaciones adicionales del proceso AR(1). Y este es un comportamiento observado en los resultados: en términos de dispersión, la distribución muestral de \bar{y} con 25 observaciones en el panel (a) es comparable con la distribución muestral de \bar{y} con 200 observaciones del panel (c) (al poner ambos gráficos en la misma escala). Recuerde que $7 \times 25 = 175$. Para el caso del proceso ARMA(1,1), que es incluso más persistente que el proceso AR(1), $A(\phi, \theta) = (1 - \phi)^{-2} \simeq 16$, lo que hace que la distribución muestral de \bar{y} con 25 observaciones en el panel (a) sea, más bien, comparable a la distribución con 400 observaciones del panel (e).

Una importante conclusión práctica extraída del análisis anterior es que en un contexto de series de tiempo con fuerte dependencia (que es usual, por ejemplo, con datos macroeconómicos), los tamaños muestrales requeridos para que las aproximaciones asintóticas sean adecuadas pueden ser considerablemente mayores que los tamaños muestrales requeridos en un contexto de poca dependencia (que es usual, por ejemplo, con datos financieros). La recomendación usual, que es mejor procurar un T lo suficientemente grande, es particularmente relevante en el análisis de series de tiempo.

Por su parte, los paneles (b), (d) y (f) del gráfico II.1 presentan la distribución del estadístico estandarizado (y estabilizado) \bar{z} . El Teorema del Límite Central indica que este converge en distribución a una variable aleatoria distribuida como una normal estándar conforme T crece. Ese es precisamente el resultado que emerge de las simulaciones. Las no normalidades observadas en muestras pequeñas (recuerde que los ruidos blancos detrás de estos procesos presentan colas anchas) se desvanecen asintóticamente.

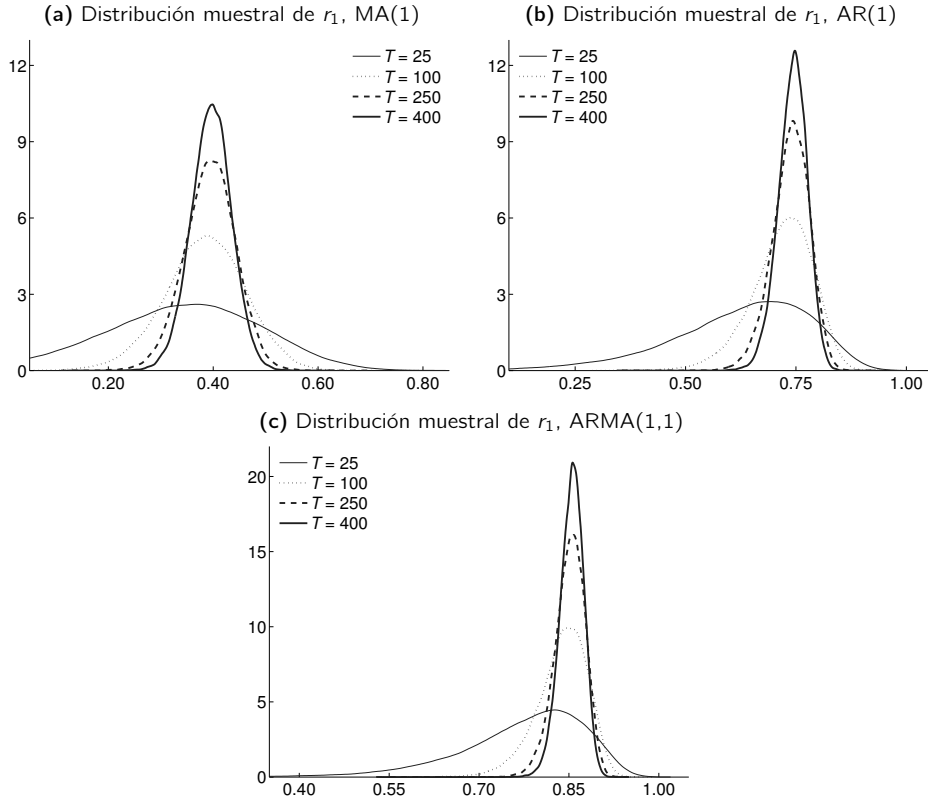
Por otro lado, el Teorema Ergódico permite encontrar el límite probabilístico de la i -ésima autocovarianza muestral

$$g_i = \frac{1}{T-i} \sum_{t=i+1}^T (y_t - \bar{y})(y_{t-i} - \bar{y}),$$

y concluir que es un estimador consistente para la i -ésima autocovarianza de y_t ⁴. En

⁴ Note que aunque g_i es el promedio sobre $T - i$ observaciones, es usual encontrar versiones de este estimador que no incorporan una corrección por las primeras i observaciones perdidas, al dividir la suma entre T en lugar de $T - i$. Esta, por supuesto, no es una distinción importante asintóticamente. Asimismo, en ocasiones el promedio \bar{y} es reemplazado por el promedio de y_t o de y_{t-i} en la muestra que va de $t = i + 1$ hasta $t = T$, o una combinación de ambos, como de hecho ocurriría en el cálculo usual de la covarianza muestral de dos series. Nuevamente, este no es un punto relevante en muestras grandes.

Gráfico II.2 Consistencia de la primera autocorrelación muestral



Notas: (Prog21_LeyesRevisadas.prg) Densidades kernel (automáticas) de r_1 con $R = 100,000$ repeticiones. Véanse las notas al gráfico II.1. La primera autocorrelación del proceso MA(1) es $\rho_1 = \theta/(1 + \theta^2) = 0.40$, la del proceso AR(1) es $\rho_1 = \phi = 0.75$ y la del proceso ARMA(1,1) es $\rho_1 = (\phi + \theta)(1 + \phi\theta)/(1 + 2\phi\theta + \theta^2) \simeq 0.86$.

concreto, es simple verificar que

$$g_i \xrightarrow{P} \mathbb{E}(y_t y_{t-i}) - \mathbb{E}(y_t) \mathbb{E}(y_{t-i}) \equiv \gamma_i,$$

para y_t ergódico y para todo i . Note que la definición de la autocovarianza γ_i permite que $\mathbb{E}(y_t) \neq \mathbb{E}(y_{t-i})$. Es decir, y_t puede no ser estacionario en media, lo que ocurre, por ejemplo, cuando $\mathbb{E}(y_t)$ contiene una tendencia determinística. Pueden establecerse resultados de consistencia similares para las autocorrelaciones muestrales: $r_i = g_i/g_0$ resulta ser un estimador cuyo límite probabilístico es ρ_i .

El programa Prog21_LeyesRevisadas.prg contiene también instrucciones para calcular y almacenar la primera autocorrelación muestral r_1 , con el propósito de estudiar sus propiedades asintóticas. El comando relevante es @COR(y,y(-1)) y los estadísticos se guardan en series de nombre DC{!a}_{!T}. El gráfico 11.2 presenta las distribuciones muestrales resultantes. Se aprecia que, en efecto, r_1 converge a ρ_1 (cuyos valores se indican en la nota al gráfico) a medida que el tamaño muestral crece. En el caso de las autocorrelaciones muestrales, no se percibe el efecto que la mayor o menor persistencia de la serie pueda tener para la inferencia, como ocurría con el caso del promedio.

11.3 Inferencia en la práctica: estimación de la varianza de largo plazo

El Teorema del Límite Central indica que, bajo circunstancias muy generales, cuando T es lo suficientemente grande (y cuando y_t es ergódico), el estadístico

$$\bar{z} = \sqrt{T} \left(\frac{\bar{y} - \mu}{\sqrt{\Omega}} \right)$$

se distribuye como una variable normal estándar. Uno de los principales usos de este resultado es el de construir **intervalos de confianza** para la media μ , lo que, a su vez, es el fundamento detrás de ejercicios de pruebas de hipótesis. Cuando $\bar{z} \sim N(0, 1)$, se sabe que la probabilidad de que \bar{z} se encuentre entre dos valores equidistantes de cero z_{crit} , $\pi = \Pr(-z_{crit} \leq \bar{z} \leq z_{crit})$, es $\pi = 90\%$ cuando $z_{crit} = 1.645$ y $\pi = 95\%$ cuando $z_{crit} = 1.960$. Con ello, es posible concluir que

$$\Pr \left(\bar{y} - z_{crit} \sqrt{\frac{\Omega}{T}} \leq \mu \leq \bar{y} + z_{crit} \sqrt{\frac{\Omega}{T}} \right) = \pi.$$

La interpretación del intervalo de confianza se desprende de la idea de muestreos repetidos. En una muestra particular, el valor del promedio que estima μ es \bar{y} ; este valor cambiaría si es que hubiera sido calculado con otra muestra, incluso manteniendo las reglas del muestreo y el proceso generador de datos (por ejemplo, el valor de μ). Las diferencias se deben al error muestral, que se manifiesta en el simple hecho de que \bar{y} es una variable aleatoria.

Así, en una muestra pueden calcularse los límites del intervalo de confianza y, si conociéramos μ (lo que ocurre al elaborar un experimento de Monte Carlo), podríamos evaluar si es que μ se encuentra dentro del intervalo calculado. Si es que el muestreo se repitiese indefinidamente, podría calcularse el intervalo de confianza en cada muestra y verificar si contiene a μ . La proporción de muestras en donde μ se encuentra contenida dentro del intervalo de confianza calculado es $\pi\%$. Dicho de otra forma, π es la probabilidad de que el intervalo de confianza contenga a μ ⁵.

⁵ Una confusión habitual es interpretar a π como la probabilidad de que μ se encuentre dentro del intervalo de confianza calculado. Esta interpretación es equivocada, entre otras cosas, porque μ es un parámetro y no una variable aleatoria. Si es correcto, no obstante, argumentar que la probabilidad de que \bar{y} se encuentre dentro del intervalo de confianza calculado es π , aunque este no pareciera ser un resultado particularmente útil para la inferencia sobre μ .

El valor de π se conoce como la **probabilidad de cobertura** y usualmente se calibra a niveles convencionales, por ejemplo al seleccionar un valor apropiado para z_{crit} . Sin embargo, es probable que en la práctica la probabilidad de cobertura no sea π . Por ejemplo, si a pesar de haber elegido cuidadosamente z_{crit} , \bar{y} no fuera normalmente distribuido. Un caso más usual tiene que ver con el hecho de que Ω , la varianza de largo plazo, **no es conocida y requiere ser estimada**.

Es importante mencionar que si contáramos con un estimador consistente de Ω , $\hat{\Omega} \xrightarrow{p} \Omega$, entonces

$$\hat{z} = \sqrt{T} \left(\frac{\bar{y} - \mu}{\sqrt{\hat{\Omega}}} \right) \xrightarrow{d} N(0, 1).$$

La diferencia entre \hat{z} y \bar{z} es que el primer estadístico utiliza $\hat{\Omega}$, mientras que el segundo asume que se conoce Ω (en ese sentido, no es un estadístico factible). Se tiene que \hat{z} y \bar{z} son **asintóticamente equivalentes**, lo que quiere decir que ambos tienen exactamente la misma distribución⁶. La equivalencia asintótica sugiere, luego, que es válido pensar en intervalos de confianza del tipo

$$\Pr \left(\bar{y} - z_{\text{crit}} \sqrt{\frac{\hat{\Omega}}{T}} \leq \mu \leq \bar{y} + z_{\text{crit}} \sqrt{\frac{\hat{\Omega}}{T}} \right) = \hat{\pi}.$$

En el límite, cuando $\hat{\Omega}$ converja a Ω , entonces $\hat{\pi}$, la probabilidad de cobertura **real**, convergerá a π , la probabilidad de cobertura **nominal**. En muestras finitas, las diferencias entre $\hat{\pi}$ y π se atribuyen, obviamente, al error muestral (o de estimación) contenido en $\hat{\Omega}$, y es interesante estudiarlas a través de simulaciones de Monte Carlo.

La varianza de largo plazo Ω es el límite de

$$\Omega_T = \gamma_0 + 2 \sum_{s=1}^{T-1} \left(1 - \frac{s}{T} \right) \gamma_s.$$

Dado que la autocovarianza muestral g_i es un estimador consistente para la autocovarianza γ_i , podría pensarse en un estimador *plug-in* de la forma

$$\tilde{\Omega}_T = g_0 + 2 \sum_{s=1}^{T-1} \left(1 - \frac{s}{T} \right) g_s.$$

No obstante, si bien este estimador es consistente, usualmente presenta un desempeño muy pobre en muestras finitas. La razón es que las autocovarianzas de orden superior son, necesariamente, calculadas con menos observaciones que las efectivamente disponibles (ya

⁶ Formalmente, si la variable aleatoria $\bar{z} \xrightarrow{d} D$, donde D es otra variable aleatoria, y $\bar{z} - \hat{z} \xrightarrow{p} 0$, luego $\hat{z} \xrightarrow{d} D$: \bar{z} y \hat{z} , al converger a la misma variable aleatoria D , son asintóticamente equivalentes.

que se pierden observaciones al considerar rezagos de y_t). En el extremo, ig_{T-1} se estima con una única observación! A la larga, el error muestral de estimar estas autocovarianzas de orden superior domina el comportamiento de $\tilde{\Omega}_T$, hasta el punto de que, por ejemplo, no se pueden garantizar aspectos tan básicos como que $\tilde{\Omega}_T \geq 0$.

En respuesta a este problema práctico, la literatura ha desarrollado una familia de estimadores **no paramétricos** (porque no dependen, por ejemplo, de la estimación de los parámetros θ o ϕ de los procesos ARMA), conocidos como estimadores **HAC** (*heteroskedasticity and autocorrelation consistent*), de la forma

$$\hat{\Omega} = g_0 + 2 \sum_{s=1}^{T-1} w(s)g_s,$$

donde $w(s)$ es un ponderador decreciente en s . La motivación de este tipo de estimadores es la de favorecer a los estimadores de γ_i que utilizan un número razonable de observaciones (cuando i es pequeño) y restarles protagonismo a los estimadores basados en pocas observaciones (cuando i es grande). De hecho, los pesos $w(s)$ decrecen más rápidamente que el término $(1 - s/T)$ usado en el estimador *plug-in*. Existen muchas elecciones para $w(s)$, siendo la más popular la propuesta inicial de Newey y West (1987),

$$\hat{\Omega} = g_0 + 2 \sum_{s=1}^q \left(1 - \frac{s}{q+1}\right) g_s,$$

donde q es un parámetro por determinar. Este estimador es consistente cuando q se incrementa con T , pero a una tasa reducida. En particular, cuando $q/T^{1/4} \rightarrow 0$.

Una ventaja importante de los estimadores de la familia HAC es que no se requiere explorar ni conocer la estructura de dependencia en los datos. Específicamente, no es necesario conocer a priori si y_t es bien caracterizado, por ejemplo, por un proceso MA o un proceso ARMA. El problema de una buena caracterización de y_t es cambiado por la determinación de un valor de q adecuado.

El programa `Prog22_Cobertura.prg` implementa un experimento de Monte Carlo que permite entender el rol de la elección de q y de las características de los datos en la determinación de la probabilidad de cobertura real $\hat{\pi}$. La generación de datos es idéntica a la descrita en la sección II.2 (programa `Prog21_LeyesRevisadas.prg`). Sin embargo, el ruido blanco que genera a los procesos dependientes es muestreado de una distribución normal estándar. Una consecuencia es que, en las simulaciones, los promedios son normalmente distribuidos para todos los valores de T y, por tanto, el valor crítico que define el ancho del intervalo de confianza, z_{crit} , proviene de una distribución normal estándar, `!zcrit = @QNORM(1 - 0.10/2)` (se utiliza una probabilidad de cobertura nominal de $\pi = 0.90$). El propósito es evaluar únicamente el efecto de la estimación de Ω sobre la probabilidad de cobertura, sin preocuparnos por el efecto que no normalidades podrían tener sobre los intervalos de confianza. A juzgar por los resultados del gráfico II.1, este supuesto no debería alterar los resultados de forma significativa para $T > 100$.

CAPÍTULO II. SERIES DE TIEMPO ESTACIONARIAS

Cuadro II.1 Probabilidad de cobertura con el estimador HAC de Newey-West

T	MA(1)		AR(1)	
	$q = T^{1/4}$	$q = T^{1/3}$	$q = T^{1/4}$	$q = T^{1/3}$
50	0.843	0.842	0.613	0.653
100	0.861	0.862	0.680	0.709
200	0.869	0.873	0.693	0.748
400	0.879	0.883	0.732	0.790
800	0.885	0.888	0.766	0.819
$\rightarrow \infty$	0.900	0.900	0.900	0.900

Notas: (Prog22_Cobertura.prg) Probabilidades de cobertura con $R = 100,000$ repeticiones y $\pi = 0.90$. En las simulaciones, se utilizan los valores $\theta = 0.5$ para el proceso MA(1) y $\phi = 0.75$ para el proceso AR(1), $\varepsilon_t \sim N(0, 1)$, y se descartan las $T_0 = 50$ observaciones iniciales. El número q es redondeado hacia abajo.

En el programa, el valor de q se determina de dos formas. Primero, $q = T^{1/4}$ (redondeado hacia el número entero inferior), lo que se implementa como `!q = @FLOOR(T^{1/4})`. Luego, se considera un valor mayor, igual a $q = T^{1/3}$, `!q = @FLOOR(T^{1/3})`. Tras determinar el valor de q , contenido en el escalar `!q`, y el tamaño muestral `!T`, se calcula el estimador HAC acumulando sumas de la siguiente manera:

```
!VLP = @VAR(y)
FOR !s = 1 TO !q
    !VLP = !VLP + 2*(1 - !s/(!q + 1))*@COV(y,y(-!s))
NEXT !s
!sLP = @SQRT(!VLP/!T).
```

Previamente, se ha definido una serie (o un vector con `!R` observaciones) de nombre `C{!a}_{!T}` (que depende del proceso `!a` y del tamaño muestral `!T`) llena de ceros. Así, se pasa a evaluar si el intervalo de confianza contiene a la media poblacional `!mu`. Si este es el caso en la repetición `!i`, la serie `C{!a}_{!T}` toma el valor de uno; de lo contrario, se mantiene su valor por defecto de cero:

```
IF (!ybar + !zcrit*!sLP > !mu) AND (!ybar - !zcrit*!sLP < !mu) THEN
    C{!a}_{!T}(!i) = 1
ENDIF
```

El promedio de `C{!a}_{!T}` a lo largo de las `!R` repeticiones aproxima a la probabilidad de cobertura real.

Los resultados con $R = 100,000$ repeticiones, diversos tamaños muestrales y para los procesos MA(1) y AR(1), se presentan en el cuadro II.1. Note que si en lugar de utilizar la desviación estándar de largo plazo estimada, hubiéramos utilizado su valor teórico

(tal y como se calcula en `Prog21_LeyesRevisadas.prg`), el muestreo normal y el gran número de repeticiones habrían implicado que todas las cifras del cuadro II.1 coincidieran con la probabilidad de cobertura nominal, 0.90. Este claramente no es el caso, y el cuadro muestra las consecuencias de reemplazar la varianza de largo plazo por su estimador HAC. En particular, se tiene que la probabilidad de cobertura real puede ser considerablemente **menor** que la nominal en muestras pequeñas, digamos $T = 50$. Ello sugiere que el intervalo de confianza no es lo suficientemente ancho como para contener al verdadero valor de la media μ , en otras palabras, que **el estimador HAC tiende a subestimar la verdadera varianza de largo plazo** (un fenómeno ampliamente documentado en la literatura). Este sesgo en el estimador HAC tiende a reducirse, la probabilidad de cobertura real se aproxima a la nominal, conforme T se incrementa, reflejando así consistencia.

Por su parte, se pueden extraer otras conclusiones importantes del cuadro II.1. Primero, el uso del estimador HAC es sumamente conveniente para procesos poco persistentes (en la simulación, el proceso MA(1) donde $\hat{\pi}$ es siempre mayor de 0.8). Note que el estimador HAC **trunca** una suma de $T - 1$ términos a una suma de q términos. Ello es completamente coincidente con lo que ocurre con la varianza de largo plazo de un proceso MA(q). En este contexto, la elección de q no es muy relevante, siempre que q sea mayor que el orden del proceso, lo que siempre ocurre en el caso MA(1) estudiado.

En cambio, cuando el proceso es persistente, como el proceso AR(1) estudiado, el desempeño del estimador HAC es menos auspicioso y la elección de q es bastante relevante. En las simulaciones, se aprecia que la elección de $q = T^{1/4}$, en general, parece incluir muy pocos términos, en un contexto en el que las autocovarianzas de orden superior pueden ser grandes. Cuando $q = T^{1/3}$, se incluye un mayor número de términos y mejora sustancialmente el desempeño del estimador HAC, aunque persisten sesgos importantes, incluso en muestras relativamente grandes.

La literatura (en especial Andrews [1991] y Newey y West [1994]) ha explorado, con varios grados de éxito, las posibilidades de automatizar la elección de q para así obtener estimadores HAC satisfactorios en circunstancias generales. Si bien es cierto que esta discusión escapa al alcance de este texto, es bueno notar que las últimas versiones de EViews implementan muchos de estos estimadores HAC (con el comando `LRRVAR`), lo que facilitaría la evaluación de su desempeño en estudios de Monte Carlo más ambiciosos.

II.4 Estimación de modelos ARMA

Las ecuaciones en diferencias que definen a los procesos ARMA son, en la práctica, tratadas como ecuaciones de regresión. A través de la estimación de los parámetros c , ϕ y θ , el investigador puede descubrir la naturaleza de dependencia del proceso y_t (que, recuerde, se encuentra completamente caracterizada por estos parámetros). Otro uso habitual de ecuaciones ARMA estimadas es la predicción de valores futuros de y_t .

La teoría de estimación de modelos de series de tiempo se basa, necesariamente, en resultados asintóticos. Para ilustrar este punto, considere el modelo AR(1) con media igual a cero,

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

donde ε_t es un ruido blanco de media cero y varianza σ^2 . Es posible estimar el parámetro ϕ si se dispone de una muestra $\{y_1, y_2, \dots, y_T\}$. El estimador de mínimos cuadrados ordinarios (MCO) presenta buenas propiedades (que luego detallamos) y es, posiblemente, el método de estimación preferido en la práctica. El estimador MCO de ϕ es

$$\hat{\phi} = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2} = \phi + \frac{\sum_{t=2}^T \varepsilon_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2}.$$

La segunda igualdad se consigue al reemplazar y_t por $\phi y_{t-1} + \varepsilon_t$ (el proceso generador de datos) y permite expresar el estimador $\hat{\phi}$ en función del parámetro desconocido ϕ y del término de error no observable ε_t . El segundo término es una variable aleatoria (la razón de los promedios de $\varepsilon_t y_{t-1}$ y de y_{t-1}^2) cuyas propiedades se transmiten al estimador MCO.

Es de interés evaluar el **sesgo** del estimador, que es la diferencia entre el valor esperado del estimador y el valor del parámetro que intenta estimar. Es decir,

$$\text{Sesgo} = \mathbb{E}(\hat{\phi}) - \phi.$$

Este sesgo es igual a la esperanza de la razón de promedios mencionada:

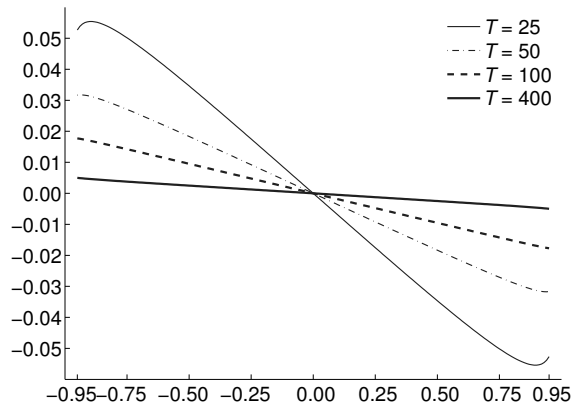
$$\text{Sesgo} = \mathbb{E} \left(\frac{\sum_{t=2}^T \varepsilon_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2} \right).$$

El numerador en la expresión anterior tiene esperanza igual a cero: dado que y_{t-1} depende de $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1\}$, todas ellas realizaciones del ruido blanco independientes de ε_t , se tiene que $\mathbb{E}(\varepsilon_t y_{t-1}) = \mathbb{E}(\varepsilon_t) \mathbb{E}(y_{t-1}) = 0$. Ello sugiere que el sesgo es, en general, reducido. Sin embargo, no es posible concluir que este sesgo sea igual a cero. La igualdad $\mathbb{E}(N/D) = \mathbb{E}(N) \mathbb{E}(1/D)$ se da únicamente cuando N y D son términos independientes, lo que no ocurre en este caso porque ambos términos dependen de la historia de ε_t . Así, aunque $\mathbb{E}(N) = 0$, se tiene que, a menos que $\phi = 0$, $\mathbb{E}(N/D) \neq 0$. La dificultad radica, precisamente, en que la dependencia de y_t induce una correlación entre N y D que no es fácil de caracterizar analíticamente para valores finitos de T^7 .

El sesgo, no obstante, puede ser determinado mediante experimentos de Monte Carlo. El programa `Prog23_SesgoAR.prg` implementa el estudio afín. Se consideran varios tamaños muestrales, almacenados en el vector `Tvals` de `!nTvals` elementos, y varios valores de ϕ , almacenados en el vector `phivals` de `!nphivals` elementos. En particular, se realizan

⁷ Para el lector interesado, Ullah (2004, sección 6.3) provee un recuento de la literatura bastante completo sobre las propiedades en muestras finitas del estimador MCO de ϕ .

Gráfico II.3 Sesgo del estimador MCO de ϕ en el modelo $y_t = \phi y_{t-1} + \varepsilon_t$



Notas: (Prog23_SesgoAR.prg) Las líneas muestran, para distintos tamaños muestrales y valores ϕ (de -0.95 a 0.95 con incrementos de 0.05), el promedio de la diferencia entre $\hat{\phi}$ y ϕ . Se utilizan $R = 100,000$ repeticiones, de modo que este promedio aproxima bastante bien al sesgo $\mathbb{E}(\hat{\phi}) - \phi$.

simulaciones para los 39 valores $\phi \in \{-0.95, -0.90, -0.85, \dots, 0.85, 0.90, 0.95\}$, es decir valores que varían de -0.95 a 0.95 con incrementos de 0.05 . Los procesos simulados son siempre estacionarios, aunque se varía su nivel de dependencia conforme ϕ se incrementa (en valor absoluto). Los casos donde $\phi \simeq 1$ son estudiados en el siguiente capítulo.

Las series son generadas con el modelo AR(1) sin intercepto, considerando $\varepsilon_t \sim N(0, 1)$, y descartando las primeras $T_0 = 50$ observaciones. Una vez generada la serie y , se calculan la series auxiliares $\text{SERIES } XX = y(-1)*y(-1)$ y $\text{SERIES } Xy = y*y(-1)$ para calcular el estimador $\hat{\phi}$ como $\hat{\phi} = \text{@SUM}(Xy)/\text{@SUM}(XX)$ ⁸. En cada repetición se almacena el estadístico $\hat{\phi} - \phi$ (donde ϕ contiene al valor de ϕ utilizado en el proceso generador de y). El promedio de esta diferencia a lo largo de un gran número de repeticiones aproxima con altos niveles de precisión al sesgo de interés. Se utilizan $R = 100,000$ repeticiones en cada experimento.

El gráfico II.3 presenta al sesgo de $\hat{\phi}$ como función de ϕ para estimaciones que utilizan distintos números de observaciones. Dado un tamaño muestral, se aprecia que el signo del sesgo depende inversamente del signo de ϕ , mientras que su magnitud depende proporcionalmente del valor absoluto de ϕ . De hecho, está ampliamente documentado que este sesgo es aproximadamente igual a $-2\phi/T$, lo que corresponde a un comportamiento

⁸ Por supuesto, puede utilizarse el objeto EQUATION de EViews para este cálculo. Se opta, sin embargo, por calcular el estimador MCO "a mano", ya que el interés se centra exclusivamente en el valor puntual del estimador y no en la gran cantidad de estadísticos involucrada en una estimación completa a través del objeto EQUATION. Ello hace más eficiente la programación de este experimento en particular, y acelera su ejecución.

que se aprecia claramente en el gráfico. Note, sin embargo, que para valores absolutos elevados de ϕ el sesgo presenta, como función de ϕ , ciertas no linealidades no capturadas por esta aproximación. Una conclusión que luego será útil recordar es que para $\phi > 0$ el sesgo es cada vez más negativo cuanto más cercano se encuentra ϕ a 1.

Por su parte, para todos los valores de ϕ considerados, se aprecia cómo el sesgo tiende a diluirse conforme el número de observaciones T se incrementa. Ello es la simple manifestación del hecho de que el estimador MCO de ϕ es **consistente**, un resultado fácil de verificar. Dado que y_t es un proceso ergódico, también lo son y_t^2 e $\varepsilon_t y_{t-1}$. Así, por el Teorema Ergódico y el Teorema de Slutsky,

$$\hat{\phi} = \phi + \frac{T^{-1} \sum_{t=2}^T \varepsilon_t y_{t-1}}{T^{-1} \sum_{t=2}^T y_{t-1}^2} \xrightarrow{p} \phi + \frac{\mathbb{E}(\varepsilon_t y_{t-1})}{\mathbb{E}(y_{t-1}^2)} = \phi + \frac{0}{\gamma_0} = \phi.$$

Por otro lado, es también sencillo verificar que el estimador MCO de ϕ es asintóticamente normal. De las expresiones anteriores, se desprende que

$$\sqrt{T}(\hat{\phi} - \phi) = \left[\frac{1}{T} \sum_{t=2}^T y_{t-1}^2 \right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=2}^T \varepsilon_t y_{t-1}.$$

El término entre corchetes converge en probabilidad a $\gamma_0 = \mathbb{E}(y_{t-1}^2)$, mientras que el segundo término es \sqrt{T} veces el promedio de la secuencia $\varepsilon_t y_{t-1}$. Nuevamente, dado que y_{t-1} depende de $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$, independientes de ε_t , se tiene no solo que $\mathbb{E}(\varepsilon_t y_{t-1}) = \mathbb{E}(\varepsilon_t) \mathbb{E}(y_{t-1}) = 0$, sino que además $\mathbb{V}(\varepsilon_t y_{t-1}) = \mathbb{V}(\varepsilon_t) \mathbb{V}(y_{t-1}) = \sigma^2 \gamma_0$ y, para todo $s \neq t$, $\mathbb{C}(\varepsilon_t y_{t-1}, \varepsilon_s y_{s-1}) = \mathbb{E}(\varepsilon_t y_{t-1} \varepsilon_s y_{s-1}) = \mathbb{E}(\varepsilon_t) \mathbb{E}(y_{t-1} \varepsilon_s y_{s-1}) = 0$. Es decir, la varianza de largo plazo de la secuencia $\varepsilon_t y_{t-1}$ es simplemente $\sigma^2 \gamma_0$. Así, por el Teorema del Límite Central, el segundo término converge en distribución a una variable normalmente distribuida, de media cero y varianza $\sigma^2 \gamma_0$. En conclusión,

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \frac{N(0, \sigma^2 \gamma_0)}{\gamma_0} \equiv N\left(0, \frac{\sigma^2}{\gamma_0}\right) = N(0, 1 - \phi^2).$$

Más allá de la normalidad asintótica, se ilustra el interesante hallazgo de que la varianza asintótica del estimador MCO de autorregresiones **no depende de σ^2** , sino de los parámetros de la parte homogénea del modelo autorregresivo.

Las consistencia de $\hat{\phi}$ y su normalidad asintótica son los ingredientes necesarios para la formulación de ejercicios de inferencia en muestras grandes, tal y como ocurre en la teoría clásica de regresión. La consistencia de $\hat{\phi}$ permite concluir que el estimador usual de la varianza del error de regresión es consistente para σ^2 , la varianza del ruido blanco⁹:

⁹ En estricto, el estimador usual divide la suma de residuos al cuadrado entre $T - 1$ y no entre T . La diferencia, por supuesto, no es relevante en muestras grandes.

$$\begin{aligned}
 s^2 &= \frac{\text{Suma de residuos al cuadrado}}{T} = \frac{1}{T} \sum_{t=2}^T (y_t - \hat{\phi} y_{t-1})^2 \\
 &= \frac{1}{T} \sum_{t=2}^T y_t^2 + \hat{\phi}^2 \frac{1}{T} \sum_{t=2}^T y_{t-1}^2 - 2\hat{\phi} \frac{1}{T} \sum_{t=2}^T y_t y_{t-1} \xrightarrow{p} \gamma_0 + \phi^2 \gamma_0 - 2\phi \gamma_1 = \sigma^2.
 \end{aligned}$$

Al tomar los límites probabilísticos, se utilizan el Teorema Ergódico, el Teorema de Slutsky y el resultado $\hat{\phi} \xrightarrow{p} \phi$. Luego, la identidad $\gamma_1 = \phi \gamma_0$, válida para el caso AR(1), permite simplificar el resultado.

Este hallazgo es de suma importancia práctica al facilitar la estimación de la varianza asintótica de $\hat{\phi}$ en el proceso de inferencia. De hecho, el estadístico t utilizado para contrastar la hipótesis $H_0 : \phi = \phi_0$ es

$$\begin{aligned}
 \tau_{\phi} &= \frac{\hat{\phi} - \phi_0}{\text{Desviación estándar estimada de } \hat{\phi}} \\
 &= \frac{\hat{\phi} - \phi_0}{\sqrt{s^2 \left[\sum_{t=2}^T y_{t-1}^2 \right]^{-1}}} = \frac{\sqrt{T}(\hat{\phi} - \phi_0)}{\sqrt{s^2 \left[\frac{1}{T} \sum_{t=2}^T y_{t-1}^2 \right]^{-1}}} \xrightarrow{d} \frac{N(0, \sigma^2 \gamma_0)}{\sqrt{\sigma^2 \gamma_0}} \equiv N(0, 1).
 \end{aligned}$$

Bajo H_0 , $\sqrt{T}(\hat{\phi} - \phi_0)$ converge a una variable normal con varianza $\sigma^2 \gamma_0$ y media cero. En el límite, el estadístico τ_{ϕ} converge a una versión estandarizada de esta variable normal.

El análisis anterior muestra cómo los resultados de la teoría clásica de regresión (básicamente, estimadores asintóticamente normales y ratios t que convergen a variables distribuidas como normal estándar) son aplicables sin modificaciones en la estimación del modelo AR(1) estacionario. Ello es también cierto para la estimación de **todo tipo de modelo ARMA estacionario**, aunque es bueno mencionar que la presencia de un componente MA en la regresión introduce una no linealidad que obliga a cambiar el método de estimación a mínimos cuadrados no lineales (MCNL).

Esta última afirmación se ilustra con el ejercicio de Monte Carlo contenido en el programa Prog24_LeyesARMA.prg. La generación de datos y el almacenamiento de resultados son idénticos a los del programa Prog21_LeyesRevisadas.prg de la sección II.2. Sin embargo, el interés se centra en los estimadores de los coeficientes ϕ y θ por mínimos cuadrados. En la repetición !i, se estiman las regresiones con los comandos

```

EQUATION eq1.LS y1 C MA(1)
EQUATION eq2.LS y2 C AR(1)
EQUATION eq3.LS y3 C AR(1) MA(1) .

```

Estos comandos definen objetos EQUATION que contienen estimaciones por mínimos cuadrados (dependiendo del caso, ordinarios o no lineales). Vale la pena mencionar lo

intuitiva y simple que resulta la estimación de estos modelos univariados en EViews¹⁰. Una vez estimados estos modelos, se ejecuta un bucle que cambia con `!a` y se tiene:

```
!coefhat = eq{!a}.C(2)
!sehat   = eq{!a}.@STDERRS(2)
DMc{!a}(!i) = !coefhat
DMt{!a}(!i) = (!coefhat - !coef{!a})/!sehat.
```

En todos los casos, se almacenan los estimados del segundo coeficiente de la regresión en la serie `DMc{!a}`: para los modelos `AR(1)` y `ARMA(1,1)`, este coeficiente corresponde a $\hat{\phi}$, mientras que para el modelo `MA(1)`, el coeficiente corresponde a $\hat{\theta}$. Asimismo, la serie `DMt{!a}` almacena el ratio t del coeficiente en cuestión. Este ratio es la diferencia entre el coeficiente estimado y el valor impuesto por la hipótesis nula (que en este caso coincide con el verdadero valor, ya que previamente se definen `!coef1 = !theta`, `!coef2 = !phi` y `!coef3 = !phi`), dividida entre la desviación estándar estimada, que se recupera con `eq.@STDERRS(.)`.

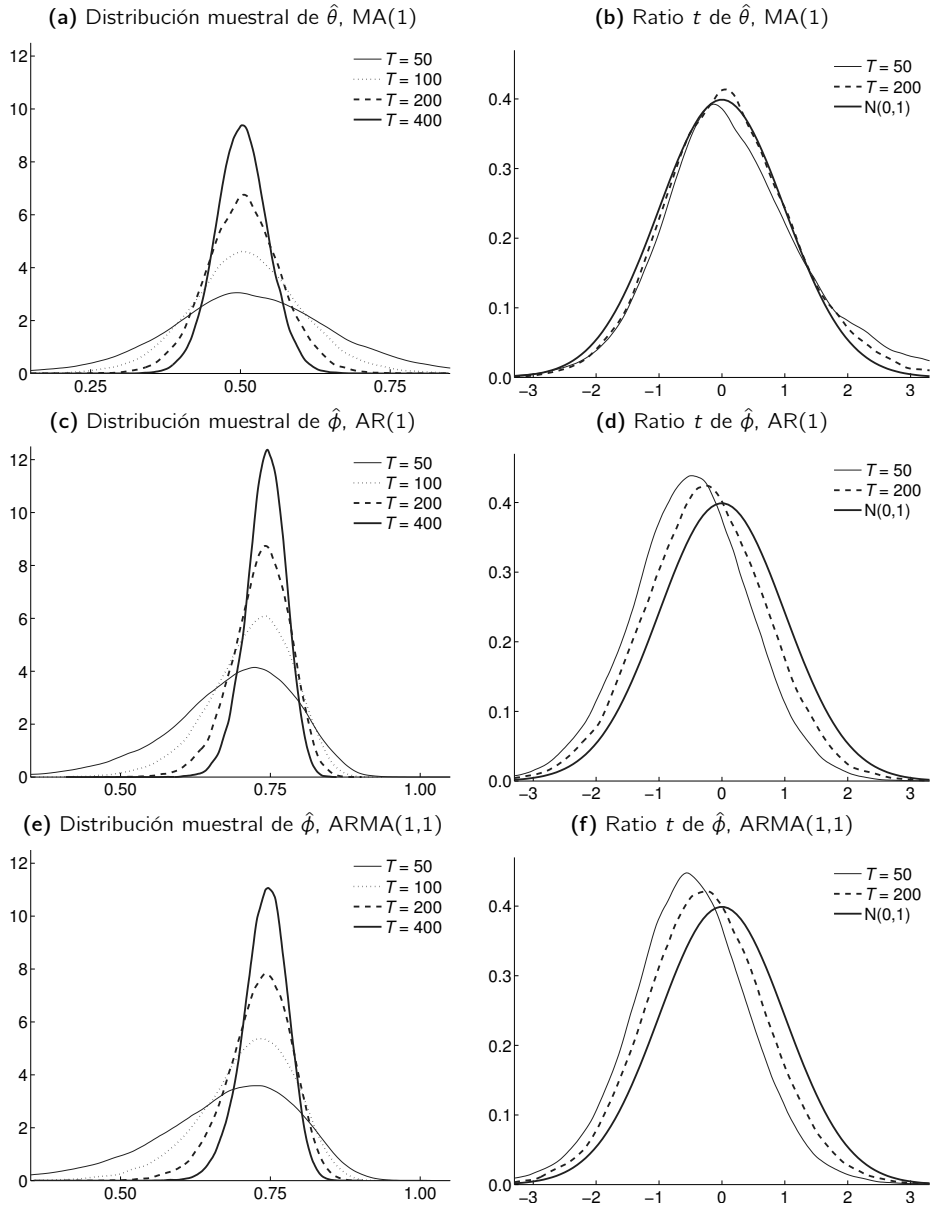
El gráfico II.4 muestra los resultados para $R = 100,000$ repeticiones. Se reportan las distribuciones muestrales de los coeficientes estimados ($\hat{\theta}$ para el modelo `MA` y $\hat{\phi}$ para los modelos `AR` y `ARMA`) y el estadístico t vinculado con la hipótesis nula de que el valor de estos coeficientes sea igual al valor efectivamente utilizado en la generación de datos. Se confirman los resultados analíticos discutidos previamente. Los estimadores de mínimos cuadrados en modelos `ARMA` estacionarios son consistentes, lo que se aprecia en la primera columna del gráfico II.4, mientras que sus respectivos ratios t son asintóticamente normales. La gran conclusión: **la teoría clásica de regresión aplica para estimadores con datos provenientes de series de tiempo estacionarias en covarianzas.**

II.5 Contrastes de autocorrelación residual

El análisis anterior ha enfatizado cómo el proceso de inferencia en un modelo de series de tiempo estacionarias puede seguir un enfoque clásico con modificaciones mínimas. Un supuesto de central importancia para la validez de esta conclusión es que, en todos los casos estudiados, el modelo estimado coincidía con el proceso generador de datos. En otras palabras, los modelos de regresión se encontraban correctamente especificados. Por supuesto, en la práctica el investigador **no conoce el proceso generador de datos** y, de hecho, parte importante del análisis empírico es descubrirlo o, en estricto, aproximarlo lo

¹⁰ En general, el comando `eq.LS y C AR(1) AR(2) ... AR(p)` estima el modelo `AR(p)` por MCO; de no incluirse un término `AR(i)`, EViews asume que el usuario impone $\phi_i = 0$. Por su parte, el comando `eq.LS y C AR(1) AR(2) ... AR(p) MA(1) MA(2) ... MA(q)` estima el modelo `ARMA(p,q)` por MCNL; de no incluirse un término `MA(i)`, EViews asume que el usuario impone $\theta_i = 0$.

Gráfico II.4 Consistencia y normalidad asintótica de estimadores de mínimos cuadrados



Notas: (Prog24_LeyesARMA.prg) Densidades kernel (automáticas) de $\hat{\theta}$, de $\hat{\phi}$ y de sus ratios t para $R = 100,000$ repeticiones. En las simulaciones, se utilizan los valores $\theta = 0.5$ y $\phi = 0.75$. Véanse las notas al gráfico II.1.

suficientemente bien. Para este propósito, se cuenta con **pruebas de hipótesis** que dan ciertas luces sobre la naturaleza de los datos.

En el contexto de los modelos de la familia ARMA, un aspecto crucial es la determinación de los órdenes p y q . Un criterio bastante aceptado sobre la adecuación de la especificación bajo estudio es la **ausencia de autocorrelación (o correlación serial) en los errores de la regresión**. Por ejemplo, suponga que estimó un modelo AR(1), de la forma $y_t = \phi y_{t-1} + u_t$, donde u_t denota el error de regresión que los residuos predicen. Si el verdadero proceso generador de datos fuera un proceso ARMA(1,1), se tendría que $u_t = \varepsilon_t + \theta \varepsilon_{t-1}$ presentaría correlación serial. Del mismo modo, si el verdadero proceso generador de datos fuera AR(2), se tendría que $u_t = \phi_2 y_{t-2} + \varepsilon_t$ también estaría autocorrelacionado. Es decir, la presencia de autocorrelación en los errores es un síntoma de mala especificación e invita a repensar la idoneidad de la regresión bajo escrutinio. En los ejemplos, si se pasara, respectivamente, de $(p, q) = (1, 0)$ a $(p, q) = (1, 1)$ o de $p = 1$ a $p = 2$, se tendría una especificación donde el error de regresión sería (o se aproximaría bastante bien a) un ruido blanco $u_t \simeq \varepsilon_t$, como los modelos vistos en secciones previas. Esta especificación es "correcta" en el sentido de que recoge plenamente los patrones dinámicos que caracterizan a y_t .

Los contrastes de hipótesis de ausencia de correlación serial más populares son las **pruebas Q**. Considere una serie de tiempo genérica u_t y sea r_i su i -ésima autocorrelación muestral, el estimador consistente de la autocorrelación ρ_i . Bajo la hipótesis nula de que u_t es un ruido blanco, es decir $H_0 : \rho_i = 0$ para todo $i > 0$, puede verificarse que $\sqrt{T}r_i$ converge a una variable aleatoria distribuida como $N(0, 1)$ y es **independiente** de $\sqrt{T}r_s$ para $i \neq s$. Con ello, se construye el **estadístico de Box y Pierce**,

$$Q = T \sum_{i=1}^K (r_i)^2 \xrightarrow{d} \chi_K^2.$$

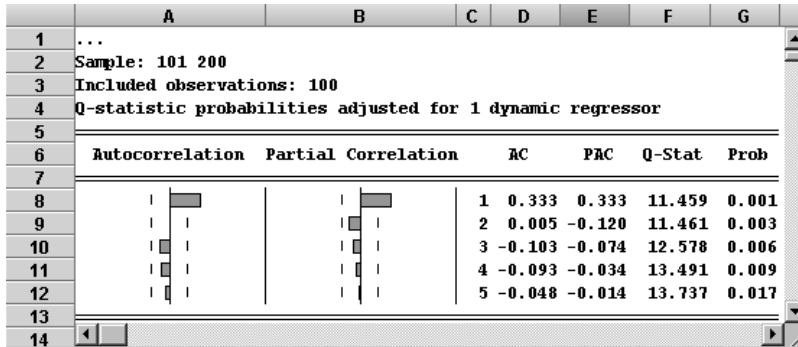
Frecuentemente, se utiliza el **estadístico de Ljung y Box** (1978) para contrastar la misma hipótesis. Este estadístico incluye correcciones por muestras finitas y es asintóticamente equivalente al estadístico Q de Box y Pierce:

$$Q = T(T+2) \sum_{i=1}^K \frac{(r_i)^2}{T-i} \xrightarrow{d} \chi_K^2.$$

Si el valor del estadístico r_i para la mayoría de valores de i es cercano a cero, entonces H_0 no podrá ser rechazada. Ello implica un valor reducido para el estadístico Q, por debajo del valor crítico dado por la distribución χ_K^2 . Con ello, se concluye que no existen correlaciones, o dependencias lineales, entre u_t y sus valores pasados. Por el contrario, si algún r_i es elevado más allá de lo que pueda atribuirse al error muestral, se rechaza H_0 y se concluye que u_t presenta dependencias con sus valores pasados.

Las pruebas Q, usualmente, son presentadas junto con el **correlograma** de la serie. Este es el gráfico (i, r_i) para $i = 1, 2, \dots, K$, donde se presentan los estimados puntuales r_i

Gráfico II.5 Correlograma de residuos



Nota: Aspecto del cuadro producido por EViews 8 tras el comando eq.CORREL(5).

junto con un intervalo de confianza para ρ_i basado en la aproximación asintótica bajo H_0 , $r_i \sim N(0, 1/T)$. En EViews, el comando u.CORREL(K) muestra el correlograma de la serie y los estadísticos Q de Ljung y Box que acumulan las primeras i autocorrelaciones, de $i = 1$ a $i = K$. En el caso de una regresión almacenada en el objeto EQUATION de nombre eq, la instrucción eq.CORREL(K) replica el análisis anterior para los **residuos de la regresión**. Un ejemplo, que luego se explica, se muestra en el gráfico II.5.

A pesar de su popularidad para el diagnóstico en modelos de regresión, las propiedades de las pruebas Q usualmente se ven afectadas cuando la serie bajo contraste se compone de los residuos de un ejercicio de estimación. Para entender por qué, y con fines ilustrativos, considere el modelo AR(1). Acá, mientras que el ruido blanco teorizado corresponde a $\varepsilon_t = y_t - \phi y_{t-1}$, los residuos que predicen este ruido blanco son iguales a $e_t = y_t - \hat{\phi} y_{t-1}$. La diferencia, por supuesto, se encuentra en el reemplazo del parámetro ϕ por el estimador $\hat{\phi}$. El estimador, recuerde, es una **variable aleatoria**, lo que introduce una fuente de aleatoriedad e incertidumbre adicional en los residuos. Esta variabilidad añadida tiende a diluirse en muestras grandes toda vez que $\hat{\phi}$ sea consistente para ϕ . Sin embargo, en muestras finitas podría tener un impacto considerable para la inferencia. Quizá la distorsión más importante se centra en que, por construcción, los residuos se encuentran correlacionados, a pesar de que los errores ε_t no lo estén. De hecho, es simple verificar que $e_t = \varepsilon_t + (\hat{\phi} - \phi)y_{t-1}$; el segundo término contiene a $(\hat{\phi} - \phi)$, que es común para todos los residuos, generando así una fuente de dependencia mutua.

Un enfoque alternativo, que también adopta como hipótesis la ausencia de correlación serial en los errores de regresión, es la denominada **prueba LM** (por el principio de Multiplicadores de Lagrange), usualmente atribuida a Godfrey (1978) y Breusch (1979). Sin ahondar en los fundamentos detrás de esta prueba, el contraste se implementa en dos pasos. Primero, se estima un modelo de regresión, por ejemplo un modelo AR(p), y se

recuperan los residuos e_t . Luego, se procede a estimar la regresión auxiliar

$$e_t = c + \sum_{j=1}^p \lambda_j y_{t-j} + \sum_{i=1}^K \beta_i e_{t-i} + \text{error}_t .$$

Esta regresión tiene como variable dependiente a los residuos de la regresión principal. Como regresores, a los regresores de la ecuación principal (los p rezagos de y_t) y K rezagos de e_t . Si valores pasados de e_t fueran útiles para explicar los valores corrientes, se tendría un claro síntoma de dependencia serial. Puesto de otro modo, los estimados de los coeficientes β , que acompañan a los rezagos del residuo, no deberían ser estadísticamente significativos cuando los errores de regresión no se encuentran autocorrelacionados. La hipótesis nula de ausencia de correlación serial en los errores es reinterpretada como $H_0 : \beta_i = 0$ para $i = 1, 2, \dots, K$.

El estadístico de contraste es TR^2 , donde R^2 es el coeficiente de determinación en la regresión auxiliar, que bajo H_0 es asintóticamente χ_K^2 . Así, bajo un escenario de dependencia serial, los rezagos de e_t ostentarían cierto poder explicativo sobre e_t , que se manifestaría en una regresión con un ajuste, medido a través de R^2 , significativo y un estadístico TR^2 elevado. Ello situaría la prueba de hipótesis en zona de rechazo. En cambio, bajo escasa correlación serial, el R^2 de la regresión auxiliar sería lo suficientemente reducido como para evitar el rechazo de H_0 .

Al igual que las pruebas Q, el enfoque LM es asintótico y su comportamiento en muestras finitas es difícil de caracterizar con precisión. Sin embargo, un estudio de Monte Carlo nos permitiría entender mejor estas propiedades y, sobre todo, comparar el desempeño de ambos contrastes de hipótesis. Este es el propósito del programa `Prog25_Autocorrelacion.prg`. En particular, el interés se centra en determinar la **potencia estadística** de ambas pruebas bajo varias hipótesis alternativas. La potencia se define como

$$\text{Potencia} = \Pr(\text{rechazar } H_0 \mid H_0 \text{ es una hipótesis falsa}) ,$$

y es igual a uno menos la probabilidad de cometer Error Tipo II. Claramente, una prueba de hipótesis es más valiosa para la inferencia correcta cuanto más potente sea. Como se ha discutido, es deseable saber si realmente los errores de la regresión en modelos tipo ARMA presentarán correlación serial, para así reformular la especificación del modelo ARMA en cuestión. Y una prueba potente es la herramienta indicada para tales fines.

El diseño de un ejercicio de Monte Carlo ilustra con claridad cuáles son los pasos para la determinación de la potencia de una prueba estadística. Considere el proceso generador de datos

$$y_t = \phi y_{t-1} + u_t , \quad \text{donde} \quad u_t = \rho u_{t-1} + \varepsilon_t ,$$

donde $|\phi| < 1$, $|\rho| < 1$ y $\varepsilon_t \sim N(0, 1 - \rho^2)$. Para todos los valores permitidos de ρ , u_t es un proceso AR(1) estacionario con media igual a cero y varianza igual a 1. Simple

manipulaciones para despejar u_t permiten concluir que

$$y_t = (\phi + \rho)y_{t-1} - \phi\rho y_{t-2} + \varepsilon_t.$$

El análisis se centra en una regresión de y_t sobre y_{t-1} (y una constante). Es decir, enfrentamos situaciones en las que equivocadamente se estima un modelo AR(1), cuando los datos corresponden a un proceso AR(2), y el interés se encuentra en verificar hasta qué punto las pruebas Q y LM mandan señales claras sobre este error de especificación.

Cuando $\rho = 0$, entonces $u_t = \varepsilon_t$ y el modelo estimado coincide con el proceso generador de datos. Esta es la situación impuesta por la hipótesis nula de las pruebas Q y LM. En muestras grandes, la probabilidad de rechazo de ambas hipótesis, basadas en los valores críticos de una distribución χ^2_K , es la probabilidad de cometer Error Tipo I (también conocida como el **tamaño estadístico**), y es una elección del investigador, por ejemplo 5%. Por otro lado, cuando $\rho \neq 0$, se sabe que la hipótesis nula de ausencia de correlación serial en los errores es **falsa**, ya que u_t tiene autocorrelaciones iguales a $\rho_i = \rho^i \neq 0$. Así, la proporción de veces que se rechaza H_0 a lo largo de muestreos repetidos converge (para un gran número de repeticiones), convenientemente, a la potencia de la prueba. Claramente, la potencia depende del valor de ρ . A priori, se espera que la potencia sea creciente en el valor absoluto de ρ , ya que cuanto mayor sea la persistencia “residual” en el modelo estimado, mayor la probabilidad de un rechazo de H_0 .

El programa `Prog25_Autocorrelacion.prg` utiliza comandos *built-in* de EViews para el cálculo de los estadísticos de interés y de sus valores- p (*p-values*). Recuerde que valores- p menores que el nivel de significación elegido por el investigador indican que el contraste cae en zona de rechazo. Como se mencionó, el comando `eq.CORREL(K)` produce un cuadro con el correlograma, los valores de los estadísticos Q (de Lung y Box) y sus valores- p . El aspecto de este cuadro, que llamaremos `_temp`, se muestra en el gráfico II.5. El interés se centra en algunas de las cantidades reportadas en este cuadro. Por ejemplo, el valor de $Q(1) = 11.837$ se ubica en la celda F8 de `_temp`, mientras que su valor- $p = 0.001$ se encuentra en la celda G8. Del mismo modo, el valor de $Q(5) = 15.285$ se ubica en la celda F12 de `_temp`, mientras que su valor- $p = 0.009$ se encuentra en la celda G12. En el contexto de un programa, es posible acceder a estos cálculos concibiendo al cuadro como una matriz, donde el listado de columnas A, B, C, ... deja de ser alfabético y pasa a ser numérico 1, 2, 3, ... Así, $Q(1) = \text{_temp}(8,6)$, $Q(5) = \text{_temp}(12,6)$ con valores- p ubicados, respectivamente, en $\text{_temp}(8,7)$ y $\text{_temp}(12,7)$. Este es un “truco” de programación recurrente para extraer estadísticos de interés automáticamente calculados por EViews pero cuyos valores son reportados en un cuadro y no como escalares (o matrices). Finalmente, el comando `eq.AUTO(K)` produce un cuadro con los resultados de la prueba LM que utiliza K rezagos del residuo. El gráfico muestra un ejemplo de este cuadro. Nuevamente, es simple ubicar las coordenadas en donde EViews reporta el estadístico TR^2 y su valor- p .

La estructura del programa `Prog25_Autocorrelacion.prg` es la siguiente. Para cada valor de ρ , la repetición `!i` (de un total de `!R`) contiene los siguientes pasos.

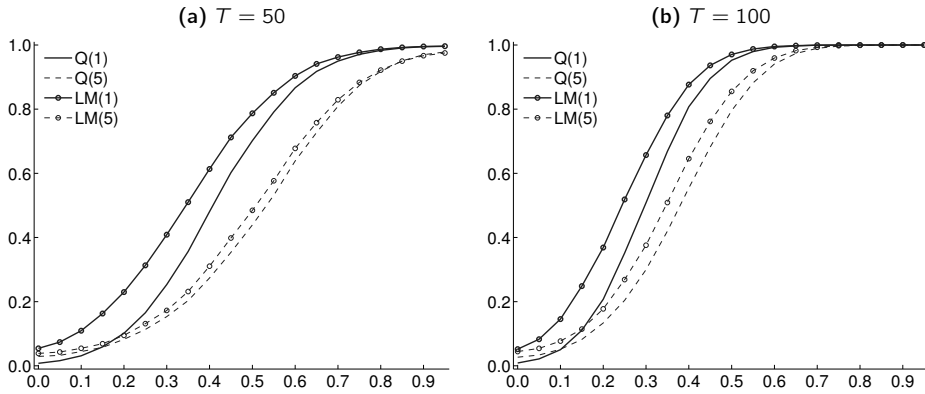
Gráfico II.6 Prueba de autocorrelación residual de Breusch-Godfrey

	A	B	C	D	E
1	Breusch-Godfrey Serial Correlation LM Test:				
2					
3	F-statistic	2.929813	Prob. F(5,93)	0.0168	
4	Obs*R-squared	13.60817	Prob. Chi-Square(5)	0.0183	
5					
6					
7	Test Equation:				
8	Dependent Variable: RESID				
9	Method: Least Squares				
10	...				
11	Sample: 101 200				
12	Included observations: 100				
13	Presample missing value lagged residuals set to zero.				
14					
15	Variable	Coefficient	Std. Error	t-Statistic	Prob.
16					
17	C	-0.006468	0.067823	-0.095367	0.9242
18	Y(-1)	-0.045224	0.056456	-0.801039	0.4252
19	RESID(-1)	0.400634	0.114732	3.491907	0.0007
20	RESID(-2)	-0.061397	0.120170	-0.510916	0.6106
21	RESID(-3)	-0.030106	0.120745	-0.249337	0.8036
22	RESID(-4)	-0.005487	0.118521	-0.046293	0.9632
23	RESID(-5)	0.017923	0.115752	0.154843	0.8773
24					
25					

Nota: Aspecto del cuadro producido por EViews 8 tras el comando eq.AUTO(5).

Primero, se genera una serie de tiempo y de acuerdo al proceso AR(2) descrito líneas arriba (ϕ se mantiene fijo). Esta serie contiene $!T0+!Tmax$ observaciones, de donde se descartarán las $!T0$ observaciones iniciales. Recuerde que $!Tmax$ es el mayor tamaño muestral considerado en nuestras simulaciones. Segundo, se restringe el tamaño muestral a $SMPL !T0+1 !T0+!T$ (donde $!T$ puede ser variante) y se estima la regresión AR(1) con el comando `EQUATION.eq y C y(-1)`. Tercero, se procede a calcular los contrastes de hipótesis. El cuadro `_temp` con las pruebas Q se genera con la instrucción `FREEZE(_temp) eq.CORREL(5)`, mientras que el cuadro con las pruebas LM se genera con `FREEZE(_temp) eq.CORREL(K)`. Cuarto, se verifica si la prueba rechazó H_0 o no. Para el caso de la prueba Q(1), previamente se ha definido un vector, de tamaño $!R$ y de nombre $Q1$, lleno de ceros. Ahora, si el valor- p del contraste Q(1) resultó ser menor que el nivel de significación (arbitrario, pero usual) de 5%, `_temp(8,7) < 0.05`, entonces la prueba Q rechazó la H_0 , un evento que se registra en $Q1(!i) = 1$. El promedio de $Q1$ a lo largo de repeticiones es nuestra estimación de la potencia de la prueba Q con $K = 1$, un nivel de significación de 5% y un valor dado de ρ . Exactamente el mismo procedimiento se repite para la prueba Q con $K = 5$ (Q5), para la prueba LM con $K = 1$ (LM1) y para la prueba LM con $K = 5$ (LM5).

Gráfico II.7 Potencia de las pruebas Q y LM de ausencia de correlación serial



Notas: (Prog25_Autocorrelacion.prg) Los paneles muestran las curvas de potencia (la frecuencia de rechazos) para las pruebas de ausencia de correlación serial en los errores de regresión, para distintos valores de ρ (eje horizontal). El proceso generador de datos es $y_t = (\phi + \rho)y_{t-1} - \phi\rho y_{t-2} + \varepsilon_t$ con $\phi = 0.75$ y $\varepsilon_t \sim N(0, 1 - \rho^2)$ y la regresión estimada corresponde al modelo AR(1) que incluye un intercepto. Se utiliza $R = 50,000$.

El gráfico II.7 muestra los resultados para los 20 valores de ρ que van de 0 a 0.95 en incrementos de 0.05, $\phi = 0.75$, los cuatro contrastes considerados y tamaños muestrales de $T = 50$ y $T = 100$. Se utilizan $R = 50,000$ repeticiones. Los resultados coinciden con lo documentado en la literatura. Considere primero el panel (a), donde $T = 50$, y fije $K = 1$. Dado el reducido tamaño muestral, ambas pruebas tienen problemas para discriminar entre un residuo verdaderamente persistente y un ruido, sobre todo cuando el valor de ρ es reducido. Ello se refleja en una baja potencia (cercana al 5%, que es la probabilidad con la que las pruebas rechazan "por casualidad"). Puesto de otro modo, dada la predominancia del error muestral en un contexto de T reducido, puede resultar difícil determinar si un proceso es serialmente correlacionado, a menos que el valor de ρ sea razonablemente elevado. De hecho, la potencia se incrementa considerablemente con ρ . Para valores de ρ cercanos a uno, la potencia es casi igual a uno. Note que la prueba LM(1) es sistemáticamente más potente que la prueba Q(1). El desempeño de las pruebas con $K = 5$ es menos auspicioso porque, desde la perspectiva del proceso generador de datos utilizado, estas pruebas se encuentran sobreparametrizadas. Se requiere de la estimación de parámetros y estadísticos adicionales para implementarlas, lo que significa una pérdida de valiosos grados de libertad con T reducido. Nuevamente, se tiene que la prueba LM(5) es sistemáticamente más potente que la prueba Q(5).

Cuando T se incrementa a $T = 100$, como se muestra en el panel (b) del gráfico II.7, el desempeño de todas las pruebas mejora considerablemente. La potencia sigue siendo baja para valores reducidos de ρ , pero es cada vez más alta para un mayor rango de valores de ρ . Con una muestra más grande, las pruebas son capaces de discriminar mejor entre casos

de procesos persistentes y ruidos. De hecho, estos patrones pueden extrapolarse y pensar en que las curvas de potencia para, digamos, $T = 1,000$ se desplazarán hacia la izquierda y llegarán a valores muy cercanos a 1 para valores cada vez menores de ρ . Este fenómeno ocurre cuando las pruebas son **consistentes**: la potencia tiende a 1 conforme $T \rightarrow \infty$, para todo valor de $\rho > 0$. Note, finalmente, que se mantiene el resultado de que la prueba LM es más potente que la prueba Q.

Una conclusión práctica de este breve estudio es que las pruebas LM deben ser favorecidas para el diagnóstico de modelos de regresión, en desmedro de las pruebas Q. Ambas pruebas contrastan, esencialmente, la misma hipótesis, tienen un comportamiento asintótico similar, pero las pruebas LM presentan mayor potencia en muestras finitas. Estudios más completos, como los de Dezhbakhsh (1990), muestran que, más aún, cuando la regresión bajo análisis incorpora un regresor distinto a los rezagos de y_t , el desempeño de las pruebas Q se ve sustancialmente deteriorado, mientras que el de las pruebas LM se mantiene. Las pruebas Q son bastante populares y siguen siendo ampliamente utilizadas. La justificación parece descansar en la costumbre de los investigadores, más que en sus propiedades estadísticas.

II.6 Criterios de información y selección de modelos

En la práctica, es posible que dispongamos de varias especificaciones que sean coherentes con la evidencia del correlograma o la evidencia de errores no autocorrelacionados brindada por las pruebas LM de Breusch y Godfrey. Por ejemplo, si el proceso generador de datos corresponde a un proceso AR(1), es altamente probable que cualquier modelo ARMA(p, q) con $p > 1$ muestre errores no autocorrelacionados. No obstante, el principio de **parsimonia** dicta elegir la especificación **más simple** (de menor p y q) entre posibles alternativas.

En este punto, los denominados **criterios de información** son particularmente útiles. Un criterio de información es un puntaje asignado a un modelo en particular, que balancea el poder explicativo o **ajuste** del modelo con su tamaño o **complejidad**. En el caso de modelos de regresión, si s^2 denota la varianza estimada de los residuos, los criterios de información toman la forma

$$CI(K) = T \ln(s^2) + C(T, K)K,$$

donde $C(T, K)$ es una función de T y posiblemente de K , el número de coeficientes estimados. En un modelo ARMA, $K = p + q$ si la regresión no incluye un término constante y $K = p + q + 1$ si la regresión incluye un intercepto. Los modelos más grandes o más complejos (mayor K) típicamente tienden a ajustar mejor, por lo que presentan un menor valor para s^{211} . El primer término de $CI(K)$ es, pues, decreciente (en estricto, no creciente)

¹¹ Es importante recordar que un común denominador de los modelos de regresión sometidos a comparación a través de los criterios de información, es que todos tienen la misma variable dependiente y_t . Así, s^2 es completamente comparable entre modelos.

APUNTES DE ESTUDIO

en K . El segundo término de $CI(K)$ es, por el contrario, una **penalidad** creciente en K . Así, bajo la óptica de los criterios de información, el mejor modelo entre varias alternativas es aquel que minimiza $CI(K)$.

Las penalidades de los dos criterios de información más utilizados en la literatura son:

$$\begin{aligned}\text{Akaike} & : C(T, K) = 2 \\ \text{Schwarz} & : C(T, K) = \ln(T).\end{aligned}$$

Note que toda vez que $\ln(T) > 2$, lo que ocurre si $T > 8$ (es decir, siempre para todo fin práctico), la penalidad impuesta por el criterio de Schwarz es mayor por cada incremento unitario en K . En otras palabras, el criterio de Schwarz, al castigar fuertemente a modelos más complejos, favorece aún más la parsimonia que el criterio de Akaike. Consecuentemente, es usual que el modelo elegido por el criterio de Schwarz tienda a ser más pequeño que el elegido por el criterio de Akaike.

Varios estudios, emblemáticamente Perron y Ng (2001, 2005), encuentran que, en la práctica, el criterio de Schwarz provee mejores resultados que el criterio de Akaike. No solo el modelo elegido por el criterio de Schwarz es más simple, sino que además tiende a explicar los datos lo suficientemente bien dentro de la muestra, y a predecirlos mejor fuera de la muestra. Al parecer, el hecho de que el criterio de Akaike típicamente favorezca a modelos sobrep parametrizados resulta ser inconveniente. Un parámetro más por estimar implica algo más de incertidumbre en el proceso de inferencia, lo que aparentemente deteriora el desempeño de las especificaciones elegidas bajo este criterio.

Tanto el criterio de Akaike como el criterio de Schwarz se basan en aproximaciones para muestras grandes. Hurvich y Tsai (1989, 1991) proveen una aproximación más refinada del criterio de Akaike, dando origen al denominado **criterio de Akaike corregido**. La penalidad en este caso pasa a ser

$$\text{Akaike corregido} : C(T, K) = 2 \left(\frac{T}{T - K - 1} \right) = 2 + 2 \left(\frac{K + 1}{T - K - 1} \right).$$

El criterio de Akaike corregido es igual al criterio de Akaike original más una penalidad **adicional**. Esta penalidad extra, que es aproximadamente igual a K veces $\kappa/(1 - \kappa)$, donde $\kappa = K/T$, es aproximadamente igual a cero en muestras grandes o, más precisamente, cuando T es muy grande en comparación con K . En muestras más cercanas a lo que usualmente se encuentra en aplicaciones, la penalidad adicional podría ser importante. De hecho, se espera que el criterio de Akaike corregido, al igual que el criterio de Schwarz, favorezca la parsimonia.

Tomando ideas de Hurvich y Tsai (1989, 1991) y de Ng y Perron (2001, 2005), el programa `Prog26_CritInfo.prg` implementa un estudio de Monte Carlo que evalúa el desempeño de los criterios de información en la selección de modelos de la familia ARMA. En términos generales, el proceso generador de datos corresponde a un modelo ARMA(3,1)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t + \theta \varepsilon_{t-1},$$

donde $\varepsilon_t \sim N(0, 1)$, mientras que los modelos “candidatos” son autorregresiones puras

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + e_t,$$

con $p = 1, 2, 3, \dots, p_{\max}$ rezagos y un intercepto. Dado un tamaño muestral T , se procede a estimar las p_{\max} autorregresiones mencionadas y a calcular los criterios de información con $K = p + 1$. El valor de p que minimiza $CI(K) \equiv CI(p + 1)$, que denotamos como p^* , provee el mejor modelo de acuerdo con el criterio $CI(K)$.

Vale la pena mencionar dos aspectos interesantes de este diseño. Primero, si el proceso generador de datos es tal que $\theta = 0$, entonces y_t es gobernado por un proceso autorregresivo puro. Dependiendo de los valores de ϕ_1 , ϕ_2 y ϕ_3 , podríamos estar hablando de un modelo AR(1), AR(2) o AR(3). Si p_{\max} es lo suficientemente grande, entonces el “modelo verdadero” se encuentra contenido dentro del conjunto de modelos candidatos. Con ello, podría pensarse en la probabilidad de que un criterio de información dado “descubra” al verdadero modelo como medida de desempeño. Por ejemplo, si $\phi_1 \neq 0$, $\phi_2 \neq 0$ y $\phi_3 = 0$, entonces el “mejor” criterio de información sería el que elija $p^* = 2$ con mayor frecuencia.

En segundo lugar, al margen de los valores de ϕ_1 , ϕ_2 y ϕ_3 , si $\theta \neq 0$, se sabe que el proceso mixto equivale a un proceso AR(∞). En estos casos, el grupo de modelos candidatos no contiene al modelo verdadero, sino que las autorregresiones candidatas (de orden finito) proveen **aproximaciones** del proceso generador de datos. Es, por tanto, deseable contar con una métrica sobre la calidad de estas aproximaciones. Una medida natural de desempeño en un modelo dinámico es el tamaño del error de predicción. Si denotamos a \tilde{y}_{T+1} como la predicción hecha con el modelo candidato (estimado con información hasta el período T) sobre el valor de y_t en el período $T + 1$, el **error de predicción cuadrático medio**

$$\text{EPCM}(y_{T+1}, \tilde{y}_{T+1}) = \mathbb{E}((y_{T+1} - \tilde{y}_{T+1})^2)$$

es la medida favorita de desempeño predictivo. Cuanto menor sea $\text{EPCM}(y_{T+1}, \tilde{y}_{T+1})$, se considera que \tilde{y}_{T+1} es un predictor con mejores propiedades. No obstante, el error cuadrático medio se ve afectado por el error muestral de estimación de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$. Así, $\text{EPCM}(y_{T+1}, \tilde{y}_{T+1})$ no solo reflejará incertidumbre sobre la naturaleza del proceso generador de datos, que es el propósito del estudio, sino que además reflejará incertidumbre sobre la estimación de los β poblacionales en la aproximación AR(p^*) del modelo ARMA.

En el caso en que el investigador conozca plenamente el proceso generador de datos, podría estimar el modelo ARMA más adecuado y obtener una predicción alternativa \hat{y}_{T+1} . El error cuadrático medio de esta predicción no reflejará incertidumbre sobre el modelo, sino exclusivamente sobre la estimación de los parámetros poblacionales. Así, podría pensarse que el error cuadrático medio de predicción **relativo**

$$\text{EPCMR} = \frac{\text{EPCM}(y_{T+1}, \tilde{y}_{T+1})}{\text{EPCM}(y_{T+1}, \hat{y}_{T+1})}$$

es una medida “neta” de error de estimación y, por consiguiente, reflejaría de una manera más clara los efectos de la incertidumbre sobre la correcta especificación ARMA. Por construcción, $EPCMR \geq 1$ y los criterios de información que reditúan valores de EPCMR más cercanos a uno han de ser favorecidos.

El cuadro II.2 muestra los resultados de ejecutar el programa Prog26_CritInfo.prg para dos tamaños muestrales, $T = 100$ y $T = 200$, y cinco procesos generadores de datos. A saber, un proceso AR(1) con $\phi_1 = 0.50$; un proceso AR(2) con $\phi_1 = 0.50$ y $\phi_2 = 0.25$; un proceso AR(2) con $\phi_1 = 0.50$, $\phi_2 = 0.25$ y $\phi_3 = 0.20$; un proceso MA(1) con $\theta = -0.75$; y, finalmente, un proceso MA(1) con $\theta = -0.99$, de modo que los coeficientes de su representación $AR(\infty)$ decaen muy lentamente. Para cada serie generada, se procede a estimar $p_{\max} = 10$ autorregresiones y a elegir, de este total, el valor de p^* que minimiza cada criterio de información. Siguiendo sugerencias de Ng y Perron (2005), las muestras son restringidas de modo que cada una de estas autorregresiones es estimada con exactamente T observaciones. Se utilizan $R = 20,000$ repeticiones¹².

En el cuadro, para cada proceso generador de datos y para cada criterio de información se muestran tres estadísticos. Primero, el valor simulado del error cuadrático medio relativo, EPCMR. Segundo, el orden promedio de las autorregresiones bajo cada criterio, $\mathbb{E}(p^*)$. Finalmente, la tasa de aciertos: el porcentaje de veces en que el orden elegido por el criterio de información coincide con el orden del proceso generador de datos, $\Pr(p^* = p)$. Este último resultado es válido solo para modelos AR puros.

Se confirma que el criterio de Schwarz tiende a elegir modelos más parsimoniosos que los criterios de Akaike. En todos los casos, el valor del rezago promedio es menor con el criterio de Schwarz. Ello es ventajoso cuando el proceso generador de datos es un proceso también parsimonioso, como un AR(1). En efecto, la tasa de aciertos en el caso del criterio de Schwarz es notablemente cercana a uno. Asimismo, bajo este criterio de información el EPCMR es también muy cercano a uno, con un desempeño algo superior a lo conseguido por los criterios de Akaike. El desempeño de todos los criterios de información mejora conforme se pasa de experimentos con $T = 100$ observaciones a experimentos con $T = 200$ observaciones. La mejora más notable la experimenta el criterio de Schwarz, que para modelos ligeramente complejos, como AR(2) o AR(3), tiende a subparametrizar el verdadero tamaño del modelo, lo que afecta tanto a la tasa de aciertos como al EPCMR cuando la muestra es pequeña.

El criterio de Akaike es claramente superior para los modelos MA considerados. Estos modelos, recuerde, pueden ser adecuadamente aproximados por autorregresiones de orden superior, en donde la penalización del criterio de Schwarz pareciera afectar sus bondades.

¹² Este número es ciertamente menor que el estándar de $R = 100,000$ repeticiones utilizado hasta el momento. La razón es que cada repetición implica la estimación de 2 (tamaños muestrales) \times 5 (procesos generadores de datos) \times 10 (modelos candidatos) = 100 regresiones, que es computacionalmente exigente. Así, $R = 20,000$ parece ser un buen balance entre el tiempo de ejecución y la precisión en la simulación.

CAPÍTULO II. SERIES DE TIEMPO ESTACIONARIAS

Llama la atención, por su parte, el excelente desempeño que tiene el criterio de Akaike corregido. Este es siempre superior al del criterio de Akaike original. En modelos simples, el criterio de Akaike corregido puede ser superado por el criterio de Schwarz, aunque con diferencias bastante marginales. En modelos algo más complejos, el criterio de Akaike corregido supera sistemáticamente a sus competidores.

Estos resultados de simulación van en línea con los presentados en Hurvich y Tsai (1989, 1991), donde se sugiere la superioridad en la práctica del criterio de Akaike corregido. Sorprende, no obstante, que a pesar de esta evidencia, las discusiones en libros de texto y en artículos académicos sigan girando en torno al binomio Akaike (original) – Schwarz. Incluso en paquetes comerciales como EViews es muy simple recuperar los criterios de información de Akaike y Schwarz tras una estimación, mientras que el criterio de Akaike corregido, simplemente, no está implementado (aunque sea sumamente simple hacerlo). Posiblemente, se requiere de un mayor número de simulaciones como las mostradas para establecer un caso documentado a favor del criterio de Akaike corregido.

El programa Prog26_CritInfo.prg contiene elementos de programación algo más sofisticados que los vistos hasta el momento. En particular, en lo relativo a la determinación del rezago elegido por cada criterio de información. Considerando que nos encontramos en la repetición !i del proceso !a y que ya se generó la serie y con un número suficiente de observaciones, se tiene lo siguiente:

```
!MinIC1 = 10000
!MinIC2 = 10000
!MinIC3 = 10000
SMPL !T0+1 !T0+!T
FOR !p = 1 TO !pmax
    EQUATION ARp.ls y C y(-1 TO -!p)
    !K = ARp.@NCOEF
    !s2 = ARp.@SSR/!T
    !IC1 = !T*LOG(!s2) + 2*!K
    !IC2 = !T*LOG(!s2) + 2*!K + 2*(!K*(!K+1))/(!T-!K-1)
    !IC3 = !T*LOG(!s2) + LOG(!T)*!K
    FOR !icrit = 1 TO 3
        IF !IC{!icrit} < !MinIC{!icrit} THEN
            !MinIC{!icrit} = !IC{!icrit}
            !p{!icrit} = !p
        ENDIF
    NEXT !icrit
NEXT !p
```

Para cada criterio de información (!icrit = 1 para el criterio de Akaike, !icrit = 2

Cuadro II.2 Desempeño de criterios de información

	Akaike			Akaike corregido			Schwarz		
	EPCMR	$\mathbb{E}(p^*)$	$\Pr(p^* = p)$	EPCMR	$\mathbb{E}(p^*)$	$\Pr(p^* = p)$	EPCMR	$\mathbb{E}(p^*)$	$\Pr(p^* = p)$
	$T = 100$								
AR(1)	1.036	1.841	0.699	1.031	1.599	0.743	1.016	1.051	0.959
AR(2)	1.037	2.557	0.566	1.032	2.312	0.599	1.036	1.583	0.521
AR(3)	1.044	3.358	0.461	1.040	3.088	0.483	1.046	2.314	0.348
MA(1), $\theta = -0.75$	1.088	4.964		1.085	4.523		1.102	2.818	
MA(1), $\theta = -0.99$	1.216	7.170		1.214	6.575		1.269	3.949	
	$T = 200$								
AR(1)	1.013	1.795	0.708	1.011	1.680	0.730	1.004	1.029	0.976
AR(2)	1.019	2.733	0.698	1.016	2.601	0.724	1.013	1.880	0.834
AR(3)	1.022	3.601	0.651	1.021	3.471	0.674	1.022	2.687	0.644
MA(1), $\theta = -0.75$	1.050	5.908		1.049	5.696		1.062	3.570	
MA(1), $\theta = -0.99$	1.144	8.572		1.142	8.396		1.189	5.432	

Notas: (Prog26.CritInfo.prg) EPCMR es la razón entre los errores de predicción cuadrático medio del modelo elegido con el criterio de información (una aproximación) relativo al modelo "verdadero": $\mathbb{E}(p^*)$ es el tamaño promedio elegido por cada criterio. Finalmente, $\Pr(p^* = p)$ es la probabilidad de que el criterio de información "descubra" al modelo verdadero, una cantidad válida únicamente cuando el proceso generador de datos es autorregresivo puro. Las simulaciones utilizan $R = 20,000$ repeticiones. Se descartan las primeras T_0 observaciones para aminorar los efectos de las condiciones iniciales. Cuando los coeficientes ϕ , no son iguales a cero, toman los valores $\phi_1 = 0.50$, $\phi_2 = 0.25$ y $\phi_3 = 0.20$.

para el criterio de Akaike corregido e $!icrit = 3$ para el criterio de Schwarz), se define el escalar $!MinIC\{!icrit\}$, que inicialmente toma un valor bastante elevado. Luego, se restringe la muestra a $SMPL !T0+1 !T0+!T$ (con efectivamente $!T$ observaciones, donde previamente se garantiza que $!T0$ sea bastante mayor que $!pmax$) y se procede a estimar, una por una, las autorregresiones. El orden de los modelos AR varía con el contador $!p$ y el comando `EQUATION ARp.ls y C y(-1 TO -!p)` genera un objeto EQUATION de nombre ARp que contiene las estimaciones. Note cómo EVIEWS entiende la sintaxis `y(-1 TO -!p)` como un listado de rezagos, desde $y(-1)$ hasta $y(-!p)$. Habiendo estimado el modelo AR, se procede a calcular los diversos criterios de información.

El condicional IF opera del siguiente modo. Si el criterio de información calculado con $!p$ rezagos $!IC\{!icrit\}$ resulta ser menor que el mínimo criterio de información registrado hasta el rezago $!p - 1$, $!MinIC\{!icrit\}$, entonces es necesario actualizar el valor de este mínimo al valor de $!IC\{!icrit\}$ y guardar en un escalar de nombre $!p\{!icrit\}$ el valor de $!p$. Este valor de $!p$ es el nuevo candidato a óptimo, ya que, hasta el momento, se asocia con el mínimo criterio de información explorado. En caso contrario, cuando el valor del criterio de información corriente es mayor que el del mínimo registrado, entonces se entiende que el rezago $!p$ no puede ser el óptimo, nada se almacena y se pasa a evaluar el siguiente modelo candidato. El haber inicializado el "mínimo" $!MinIC\{!icrit\}$ con un valor muy elevado garantiza que el programa ejecutará la actualización $!MinIC\{!icrit\} = !IC\{!icrit\}$ con el primer modelo de regresión estimado.

El resultado de ejecutar las instrucciones anteriores son tres escalares $!p1$, $!p2$ y $!p3$ con los rezagos elegidos por cada criterio de información. Luego, se tiene que:

```
FOR !icrit = 1 TO 3
    lags{!icrit}\{!a}\{!T}\{!i} = !p{!icrit}
    SMPL !T0+1 !T0+!T
    EQUATION ARIC.LS y C y(-1 to -!p{!icrit})
    SMPL !T0+!T+1 !T0+!T+1
    ARIC.FORECAST yf
    SERIES res2 = (y - yf)^2
    res2{!icrit}\{!a}\{!T}\{!i} = @MAX(res2)
NEXT !icrit
```

Los valores de $!p\{!icrit\}$ se almacenan en series de nombre $lags\{!icrit\}\{!a}\{!T}\{!i}$ que varían con el proceso generador de datos $!a$ y el tamaño muestral $!T$. Los resultados para $E(p^*)$ en el cuadro II.2 corresponden a los promedios de estas series, mientras que los resultados para $Pr(p^* = p)$ se consiguen como la proporción de veces en las que esta serie es igual a $!a$ (para $!a = 1, 2, 3$).

Luego, se estima la autorregresión elegida por el criterio de información y se almacena en el objeto EQUATION de nombre ARIC. Como antes, la muestra utilizada para tal propósito es

APUNTES DE ESTUDIO

SMPL !T0+1 !T0+!T. El resultado de esta estimación es utilizado para predecir. Para ello, el comando ARIC.FORECAST yf almacena las predicciones en la serie yf. Este comando se ejecuta para una muestra de tamaño igual a uno, SMPL !T0+!T+1 !T0+!T+1, que corresponde a la observación $T+1$. Se trata, pues, de una predicción un período en adelante. El error de predicción cuadrático se almacena en la serie de nombre `res2{!icrit}{!a}{!T}`, que también varía con el proceso generador de datos !a y el tamaño muestral !T¹³. Los resultados para EPCMR en el cuadro II.2 utilizan los promedios de estas series (en el numerador).

¹³ El uso de la instrucción @MAX(res2) se debe a que en la sintaxis `z(!i) = a`, el objeto a debe ser necesariamente un escalar. En el programa, res2 es un objeto serie de un solo valor.

III No estacionariedad

El propósito de este capítulo es estudiar los problemas de estimación e inferencia que surgen con series de tiempo que presentan algún tipo de no estacionariedad. El tema es de suma relevancia práctica desde una perspectiva econométrica, ya que es usual que las series de tiempo utilizadas en aplicaciones económicas presenten algún tipo de no estacionariedad. Muchos de los resultados estudiados en el capítulo previo, sobre todo los vinculados con las predicciones de las Leyes de Grandes Números y del Teorema del Límite Central, siguen estando vigentes, tras simples adecuaciones, en un mundo no estacionario. Muchos otros, y de gran importancia, requieren del conocimiento de otras leyes en particular, y en general de una teoría asintótica distinta. Ciertamente, este nuevo enfoque asintótico no está siempre basado en normalidad y puede resultar algo más engorroso que el de inferencia clásica. Es aquí donde el uso de simulaciones nos permite alcanzar conclusiones relevantes y muy generales sin necesidad de ahondar en detalles que escapan al alcance de este texto.

En adelante, mantendremos la siguiente notación. El símbolo ε_t se sigue reservando para denotar un ruido blanco, un proceso no autocorrelacionado con media cero y varianza σ^2 . Por su parte, denotamos como u_t a un proceso estacionario, en general autocorrelacionado, y ergódico cuya media es también igual a cero. Los problemas de estimación que involucran a u_t son, en general, de la misma naturaleza que los estudiados en el capítulo II.

Iniciamos el análisis estudiando las características de un proceso que no es estacionario, por tener una media móvil, pero cuyas covarianzas son independientes de t . Este tipo de proceso se conoce como **estacionario en tendencia** o **TS** (por las siglas del inglés *trend stationary*). Luego, consideramos los casos que motivan una mayor discusión, en donde la no estacionariedad se ve reflejada en covarianzas que dependen de t . Este tipo de proceso se conoce como **estacionario en diferencias** o **DS** (por las siglas de *difference stationary*). Buena parte del análisis se centra en el estudio de las herramientas de inferencia disponibles para distinguir entre estos dos casos.

III.1 Modelo de tendencia lineal

Considere el modelo de tendencia lineal

$$y_t = \alpha + \delta t + u_t,$$

donde u_t es un proceso estacionario de media igual a cero. En este caso, $\mathbb{E}(y_t) = \alpha + \delta t$ es una función de t , por lo que y_t **no es estacionario en media**. Sin embargo, y_t es

un proceso estacionario en covarianzas, ya que sus segundos momentos son constantes: $\mathbb{V}(y_t) = \mathbb{V}(u_t) = \gamma_0$ y $\mathbb{C}(y_t, y_{t-i}) = \mathbb{C}(u_t, u_{t-i}) = \gamma_i$ para cualquier $i \neq 0$. Este es un ejemplo de un **proceso estacionario en tendencia**: una vez removida la tendencia de y_t (es decir, la esperanza no condicional), el residuo es igual al proceso estacionario u_t . Si bien implementaremos este modelo en simulaciones, luego de parametrizar las autocovarianzas de u_t con procesos de la familia ARMA, por simplicidad analítica supondremos que u_t es igual al ruido blanco ε_t : $y_t = \alpha + \delta t + \varepsilon_t$.

El álgebra de la estimación por mínimos cuadrados de este estilizado modelo de tendencia lineal revela varios fenómenos sumamente interesantes. No es difícil verificar que los estimadores de α y de δ , para una muestra de tamaño T , dependen del promedio de dos secuencias: el ruido blanco ε_t y la variable auxiliar $w_t = (t/T)\varepsilon_t$ cuyas propiedades son: $\mathbb{E}(w_t) = (t/T)\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{V}(w_t) = \sigma^2(t/T)^2$ y $\mathbb{C}(w_t, w_s) = (ts/T^2)\mathbb{C}(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$. Si bien es cierto que la secuencia w_t no es, en estricto, estacionaria, ya que su varianza depende de t , sí se trata de una secuencia ergódica sobre la que pueden aplicarse las leyes asintóticas estudiadas en el capítulo anterior. Finalmente, note que $\mathbb{C}(w_t, \varepsilon_t) = \sigma^2(t/T)$.

Sean $\bar{\varepsilon}$ y \bar{w} los respectivos promedios muestrales, basados en T observaciones. Ambos tienen media igual a cero; además, $\mathbb{V}(\bar{\varepsilon}) \simeq \sigma^2 T^{-1}$, $\mathbb{V}(\bar{w}) \simeq \frac{1}{3}\sigma^2 T^{-1}$ y $\mathbb{C}(\bar{w}, \bar{\varepsilon}) \simeq \frac{1}{2}\sigma^2 T^{-1}$. Por el Teorema Ergódico, $\bar{\varepsilon}$ y \bar{w} convergen en probabilidad a cero, mientras que por el Teorema del Límite Central, el vector $\sqrt{T}(\bar{\varepsilon}, \bar{w})'$ converge a una distribución normal bivariada con media cero, varianzas σ^2 y $\frac{1}{3}\sigma^2$ y covarianza $\frac{1}{2}\sigma^2$. De este modo, se verifica que el estimador de α tiene un comportamiento asintótico "usual":

$$\hat{\alpha} - \alpha \simeq 4\bar{\varepsilon} - 6\bar{w} \xrightarrow{p} 0 \quad \text{y} \quad \sqrt{T}(\hat{\alpha} - \alpha) \simeq \sqrt{T}(4\bar{\varepsilon} - 6\bar{w}) \xrightarrow{d} N(0, 4\sigma^2).$$

Es decir, $\hat{\alpha}$ es consistente y asintóticamente normal.

En cambio, el estimador de δ puede escribirse como

$$\hat{\delta} - \delta \simeq \frac{12}{T}\bar{w} - \frac{6}{T}\bar{\varepsilon} \xrightarrow{p} 0.$$

Claramente, $\hat{\delta}$ es también consistente. De hecho, ocurre más aún que

$$T(\hat{\delta} - \delta) \simeq 12\bar{w} - 6\bar{\varepsilon} \xrightarrow{p} 0,$$

y, en general, que $T^n(\hat{\delta} - \delta) \xrightarrow{p} 0$ para varios valores de $n > 0$. Ello implica que $\hat{\delta}$ converge en probabilidad a δ pero a una **tasa acelerada**. Conforme T se incrementa, la brecha $\hat{\delta} - \delta$ se va acercando a cero, y lo hace lo suficientemente rápido como para que ello ocurra incluso después de multiplicarla por una función divergente como T^n ($n > 0$). En otras palabras, la velocidad con la que $\hat{\delta} - \delta$ se aproxima a cero es mayor que la velocidad con la que T^n tiende al infinito. El estimador $\hat{\delta}$ es **superconsistente**.

Sea A_T un término genérico que denota una función de T con límite finito, $A_T \rightarrow A$. La superconsistencia de $\hat{\delta}$ se centra en el hecho de que $\mathbb{V}(\hat{\delta}) \simeq A_T/T^m$ con $m > 1$, en lugar

del resultado más usual, típico en promedios de procesos ergódicos, de que $\mathbb{V}(\hat{\alpha}) \simeq A_T/T$ ($m = 1$). De hecho, una buena definición de superconsistencia alude a este fenómeno: **un estimador es superconsistente si su varianza muestral decae a una tasa mayor que T** , conforme T se incrementa. La superconsistencia, vale la pena mencionarlo, **es un fenómeno exclusivo de series de tiempo no estacionarias** y no se encuentra en otros contextos de inferencia estadística (como con datos estacionarios o con secciones transversales).

Asimismo, una implicancia de la superconsistencia es que la tasa necesaria para estabilizar las varianzas muestrales es típicamente mayor que la usual \sqrt{T} . En el caso del estimador $\hat{\delta}$,

$$T^{3/2}(\hat{\delta} - \delta) \simeq \sqrt{T}(12\bar{w} - 6\bar{\varepsilon}) \xrightarrow{d} N(0, 12\sigma^2).$$

El programa Prog31_Tendencia.prg ilustra el comportamiento, en particular el proceso de convergencia, de un estimador superconsistente a través de un simple estudio de Monte Carlo. El proceso generador de datos considerado es

$$y_t = \alpha + \delta t + u_t \quad \text{donde} \quad u_t = \phi u_{t-1} + \varepsilon_t,$$

y donde $|\phi| < 1$ garantiza que el error de la regresión en el modelo de tendencia lineal, u_t , sea un proceso estacionario en media y en covarianzas. Como ya se analizó, las autocovarianzas de y_t son iguales a las autocovarianzas del proceso AR(1) u_t .

Para cada repetición i , los datos se generan de acuerdo con

```
GENR u - !phi*u(-1) = @RTDIST(3)/@SQRT(3)
SERIES y = !alpha + !delta*time + u,
```

donde los escalares $!phi$, $!alpha$ y $!delta$ contienen valores predefinidos de los parámetros ϕ , α y δ . Los valores de estos parámetros no son particularmente relevantes, pero debe asegurarse que $|\phi| < 1$. Estas series son generadas para un gran número de observaciones y se descartan las primeras $!T0$. La serie `time` es generada previamente con el comando `SERIES time = @TREND`, que es la instrucción de EViews para generar una tendencia lineal. Note, finalmente, que se considera que ε_t es un ruido blanco de media cero, varianza igual a uno y colas anchas (una distribución t de Student con 3 grados de libertad). El motivo es, como lo ha sido en ocasiones anteriores, permitir que sea el Teorema del Límite Central el que lidie con los efectos de posibles no normalidades en los datos.

Una vez generados los datos, se procede a estimar la regresión de y_t sobre una constante y t . Para un tamaño de muestra $!T$, que es variante, las series de nombre `D{!a}delta_{!T}` almacenan el estadístico $\sqrt{T^{!a}}(\hat{\delta} - \delta)$ para valores de $!a$ entre 0 y 3. Se espera que $\sqrt{T^{!a}}(\hat{\delta} - \delta) \xrightarrow{p} 0$ para valores de $!a$ menores de 3, de acuerdo con el análisis previo. Para $!a = 3$, este estadístico debería converger a una variable aleatoria normalmente distribuida, de media 0 y varianza igual a $12\Omega_u$, donde Ω_u es la varianza de largo plazo de u_t . Esta es una pequeña generalización de nuestros resultados analíticos, donde se consideró que el error

de regresión era un ruido blanco y, por tanto, que su varianza de largo plazo equivalía a su varianza no condicional. Luego, $D\delta_{\{T\}}$ almacena una versión estandarizada que converge a una variable aleatoria distribuida como una normal estándar (VLPu contiene el valor de Ω_u). El código relevante es:

```
EQUATION detrend.LS y C time
!deltahat = detrend.C(2)
D0delta_{!T}(!i) = !deltahat - !delta
D1delta_{!T}(!i) = @SQR(!T)*(!deltahat - !delta)
D2delta_{!T}(!i) = !T*(!deltahat - !delta)
D3delta_{!T}(!i) = @SQR(!T^3)*(!deltahat - !delta)/@SQR(12*!VLPu) .
```

El gráfico III.1 presenta las distribuciones empíricas de $D\delta_{\{T\}}$ para diversos valores de T . En el panel (a) se observa la convergencia en probabilidad a cero de $\hat{\delta} - \delta$, y en los paneles (b) y (c) se verifica que esta convergencia es muy rápida, incluso más rápida que la tasa de divergencia de las funciones $T^{1/2}$ y T . Finalmente, el panel (d) muestra la convergencia en distribución a una normal estándar de este estimador, después de estabilizar su varianza con la potencia de T adecuada, $T^{3/2}$. Este último resultado permite concluir que, en el modelo de tendencia lineal, las inferencias respecto a los parámetros de la tendencia, que son los parámetros de $\mathbb{E}(y_t)$, pueden realizarse de manera convencional, utilizando aproximaciones normales.

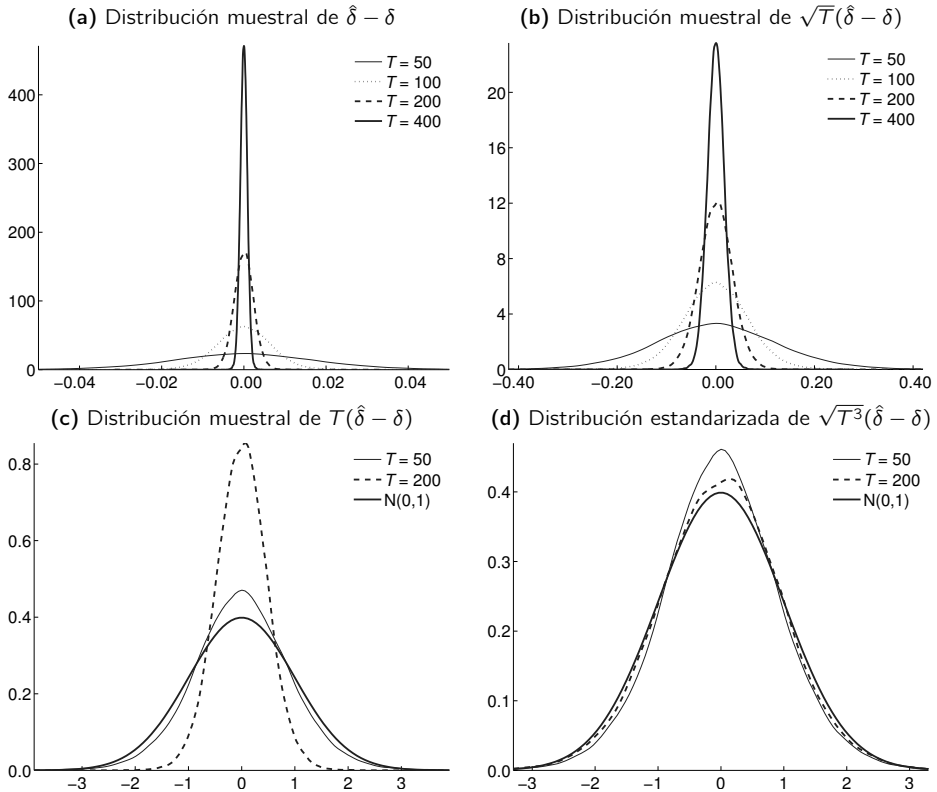
El programa Prog31_Tendencia.prg ilustra, además, otra importante implicancia de la superconsistencia. Usualmente, el interés se centra no solo en inferir sobre la media de y_t , sino especialmente sobre sus patrones de dependencia, medidos a través de sus autocovarianzas. Dado que $\mathbb{C}(y_t, y_s) = \mathbb{C}(u_t, u_s)$, ello se centra en inferencias sobre la estructura dinámica de u_t , que en el proceso generador de datos considerado se concentra en el parámetro ϕ . En pocas palabras, es de interés estimar ϕ .

Surgen dos posibilidades. La primera se enmarca en una situación que podría considerarse como "ideal". Se asume que tanto α como δ son conocidos y, por tanto, $u_t = y_t - \mathbb{E}(y_t)$ es conocido. Así, se tiene un primer estimador $\tilde{\phi}$ que se obtiene simplemente al regresar u_t sobre u_{t-1} . No obstante, en la práctica solo observamos y_t pero no u_t , lo que hace que $\tilde{\phi}$ no sea un estimador factible.

La segunda posibilidad, luego, consiste en un **procedimiento de dos etapas**, donde primero α y δ , y por consiguiente $\mathbb{E}(y_t)$, son estimados, dando origen a los residuos $e_t = y_t - \hat{y}_t$ que son, a su vez, predictores del desconocido u_t ; luego, el segundo estimador $\hat{\phi}$ se obtiene al regresar e_t sobre e_{t-1} . Note que en este procedimiento de dos etapas, $\hat{y}_t = \hat{\alpha} + \hat{\delta}t$ resulta ser un estimador de $\mathbb{E}(y_t) = \alpha + \delta t$.

El comportamiento del estimador "ideal" $\tilde{\phi}$ es idéntico al estudiado en la sección II.4 (programa Prog24_LeyesARMA.prg). En particular, sin ahondar en detalles, se puede

Gráfico III.1 Superconsistencia



Notas: (Prog31_Tendencia.prg) Se muestran las densidades kernel (automáticas) para $R = 100,000$ repeticiones. En las simulaciones, se utiliza el valor $\phi = 0.5$, y se descartan las primeras $T_0 = 50$ observaciones en la generación de datos. La escala del eje vertical difiere entre paneles. El estadístico estandarizado del panel (d) es $\sqrt{T^3}(\hat{\delta} - \delta)/\hat{\sigma}$, donde $\hat{\sigma}^2$ es 12 veces la varianza de largo plazo de u_t .

determinar que

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, 1 - \phi^2).$$

Para investigar el comportamiento del estimador factible $\hat{\phi}$, no es difícil verificar que $e_t = u_t + (\mathbb{E}(y_t) - \hat{y}_t)$. Es decir, el residuo de la regresión de y_t sobre una constante y una tendencia lineal, contiene a la variable no observable u_t , más el error de estimación de $\mathbb{E}(y_t)$. Si \mathcal{D}_T denota una variable aleatoria que converge a un límite (posiblemente aleatorio) finito, es posible (aunque algo tedioso) concluir que

$$\sqrt{T}(\hat{\phi} - \tilde{\phi}) = \sqrt{T}(\hat{\delta} - \delta)\mathcal{D}_T.$$

La diferencia $\sqrt{T}(\hat{\phi} - \tilde{\phi})$ es un término que converge a cero, en virtud de $\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} 0$, la superconsistencia de $\hat{\delta}$. Esto es cierto independientemente del comportamiento asintótico de \mathcal{D}_T . Así, tras simples manipulaciones,

$$\sqrt{T}(\hat{\phi} - \phi) = \sqrt{T}(\tilde{\phi} - \phi) + \sqrt{T}(\hat{\phi} - \tilde{\phi}) \xrightarrow{d} N(0, 1 - \phi^2) + 0 \equiv N(0, 1 - \phi^2).$$

Este resultado indica que la distribución asintótica de $\hat{\phi}$ es **exactamente la misma** que la distribución asintótica del estimador "ideal" $\tilde{\phi}$. Ambos son estimadores **asintóticamente equivalentes**. Si bien es cierto que $\hat{\phi}$ contiene una fuente de variabilidad adicional, producto de utilizar el estimador \hat{y}_t en lugar del desconocido $\mathbb{E}(y_t)$, esta incertidumbre se disipa conforme T se incrementa y, a la larga, desaparece.

La fuerza detrás de este importante resultado es, como se mencionó, la superconsistencia de $\hat{\delta}$. Intuitivamente, ocurre lo siguiente. Ambos estimadores de ϕ convergen en distribución a la tasa habitual de \sqrt{T} , mientras que el estimador de δ lo hace más aceleradamente. Así, al llegar a tamaños muestrales lo suficientemente grandes como para que los estimadores de ϕ se comporten de manera similar a lo predicho por su distribución asintótica (es decir, sean aproximadamente normales), el error muestral $\hat{\delta} - \delta$ se habrá acercado lo suficiente a cero como para no contribuir con ninguna fuente de incertidumbre adicional.

La superconsistencia justifica el uso de procedimientos de dos etapas como el descrito: en la práctica, el parámetro desconocido δ puede ser reemplazado por un estimador superconsistente $\hat{\delta}$ sin mayores implicancias (asintóticamente) para la inferencia sobre ϕ . Este es solo uno de los casos que emergen en el análisis de series de tiempo no estacionarias, algunos de los cuales serán revisados posteriormente. Note que si $\hat{\delta}$ no fuera superconsistente, el error de estimación de $\mathbb{E}(y_t)$ se desvanecería a la tasa habitual, y esta incertidumbre afectaría la distribución asintótica de $\hat{\phi}$, alejándola de la distribución asintótica del estimador "ideal" $\tilde{\phi}$.

El programa Prog31_Tendencia.prg incorpora, para cada repetición !i y dado un tamaño muestral !T, el siguiente código:

```

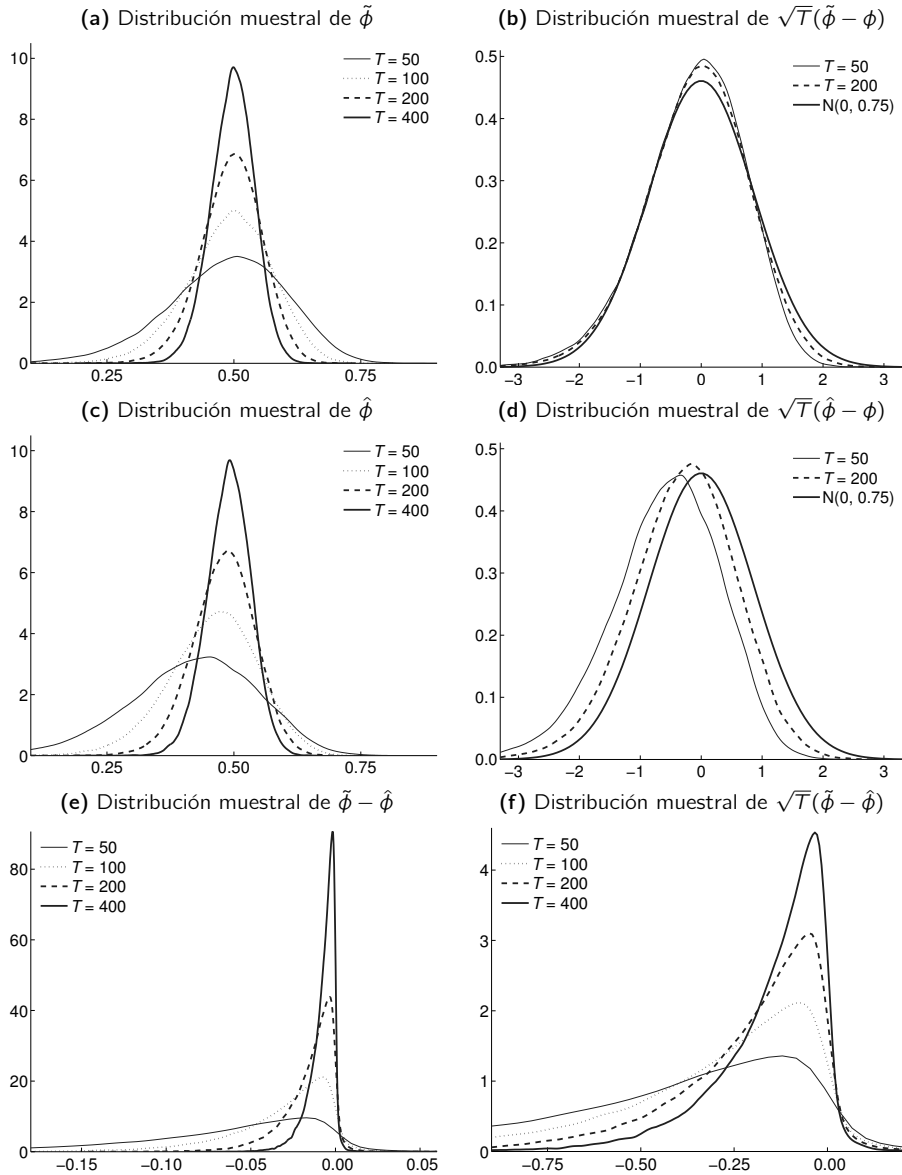
EQUATION equ.LS u u(-1)
!phitilde = equ.C(1)
DMphiu_{!T}(!i) = !phitilde
DAphiu_{!T}(!i) = @SQR(!T)*(!phitilde - !phi)

detrend.MAKERESIDS e
EQUATION eqe.LS e e(-1)
!phihat = eqe.C(1)
DMphie_{!T}(!i) = !phihat
DAphie_{!T}(!i) = @SQR(!T)*(!phihat - !phi)

DMphid_{!T}(!i) = !phihat - !phitilde
DAphid_{!T}(!i) = @SQR(!T)*(!phihat - !phitilde).

```

Gráfico III.2 Equivalencia asintótica de estimadores de ϕ



Notas: (Prog31_Tendencia.prg) Densidades kernel (automáticas) para $R = 100,000$ repeticiones, véanse las notas al gráfico III.1. El estimador $\tilde{\phi}$ proviene de una regresión de u_t sobre u_{t-1} , mientras que el estimador $\hat{\phi}$ se obtiene al regesar e_t sobre e_{t-1} , siendo e_t los residuos de una regresión de y_t sobre una constante y t .

Primero, se regresa u_t sobre su rezago (ecuación equ) para obtener el estimador "ideal" $\tilde{\phi}$ ($\tilde{\phi}$) que es almacenado en la serie $DM\phi_{!T}$. Por su parte, la variable $\sqrt{T}(\tilde{\phi} - \phi)$ se almacena en la serie $D\text{A}\phi_{!T}$. Segundo, se generan los residuos de la regresión de y_t sobre una constante y t con el comando `detrend.MAKERESIDS e`. La serie de residuos generada, e , es luego utilizada para obtener el estimador factible $\hat{\phi}$ ($\hat{\phi}$), en la ecuación eqe) que es almacenado en la serie $DM\phi_{!T}$, mientras que la variable $\sqrt{T}(\hat{\phi} - \phi)$ se almacena en la serie $D\text{A}\phi_{!T}$. Finalmente, las diferencias entre los dos estimadores se almacenan en las series $DM\phi_{!T}$ y $D\text{A}\phi_{!T}$.

El gráfico III.2 reporta los resultados. En los paneles (a) y (c) se aprecia cómo, respectivamente, $\tilde{\phi}$ y $\hat{\phi}$ son consistentes para ϕ , al concentrar cada vez más masa probabilística en torno a $\phi = 0.5$ conforme T se incrementa. En el panel (e) se confirma que ambos estimadores comparten el mismo límite probabilístico, al tener que $\tilde{\phi} - \hat{\phi} \xrightarrow{p} 0$.

Por otro lado, los paneles (b) y (d) del gráfico III.2 muestran que ambos estimadores convergen en distribución a $N(0, 1 - 0.5^2)$. Es interesante notar que aunque las diferencias en ambas distribuciones son imperceptibles para $T = 200$, sí son bastante claras para $T = 50$. El error de estimación de $\mathbb{E}(y_t)$ tiene efectos visibles sobre el sesgo y la varianza de $\hat{\phi}$ en muestras pequeñas, lo que podría causar serias distorsiones para la inferencia sobre ϕ . Finalmente, el panel (f) muestra la equivalencia asintótica entre ambos estimadores, al tener que $\sqrt{T}(\tilde{\phi} - \hat{\phi}) \xrightarrow{p} 0$. Ciertamente, la tasa de convergencia es menor que la mostrada en el panel (e), dado el factor divergente \sqrt{T} , pero la convergencia a cero es un patrón bastante obvio.

Es bueno mencionar, para concluir, que el estimador en dos etapas, por el teorema de regresiones particionadas (conocido como el Teorema de Frisch-Waugh-Lovell), es idéntico al que se obtendría directamente de la regresión aumentada

$$y_t = \alpha^* + \delta^* t + \phi y_{t-1} + \text{error}_t .$$

Esta regresión, no obstante, estima $\alpha^* = (1 - \phi)\alpha + \phi\delta$ y $\delta^* = (1 - \phi)\delta$, que no son los parámetros de $\mathbb{E}(y_t)$, sino transformaciones de estos.

III.2 Procesos integrados y no estacionariedad

A diferencia de lo visto en el modelo de tendencia lineal, la no estacionariedad de un proceso se manifiesta usualmente en autocovarianzas que dependen de t , al margen de lo que ocurra con la media. Para ver cómo, conviene definir un **proceso integrado**.

La integración es la operación inversa de las diferencias, por lo que se dice que un proceso es integrado si sus diferencias resultan ser una serie estacionaria (en estricto, ergódica). Por ello, a este tipo de procesos se le conoce como **estacionario en diferencias**. Concretamente, y_t es integrado de orden uno, lo que se denota como $y_t \sim I(1)$, si es que $\Delta y_t = u_t \sim I(0)$.

En general, $y_t \sim I(d)$, esto es y_t es un proceso integrado de orden d , si $\Delta^d y_t = u_t \sim I(0)$, es decir si se precisa tomarle a y_t diferencias d veces para obtener como resultado un proceso estacionario. Centraremos la atención en procesos $I(1)$, es decir series de tiempo cuyas primeras diferencias son estacionarias, por ser los más relevantes en aplicaciones. Note, además, que un proceso $I(0)$ no requiere de ninguna diferenciación para ser estacionario ($d = 0$), por lo que debe entenderse que $u_t \sim I(0)$ es sinónimo de estacionariedad.

Estudiaremos luego con más detalle por qué **un proceso integrado no es estacionario**. De momento, puede ilustrarse este resultado tras verificar que un proceso integrado viola las condiciones de estacionariedad discutidas en la sección II.1, por construcción. Por brevedad, suponga que $u_t \sim MA(1)$, esto es $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$, que es siempre estacionario, y ocurre que $\Delta y_t = u_t$. En términos de las representaciones ARMA discutidas en el capítulo II, la identidad $y_t = y_{t-1} + u_t$ puede expresarse como $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Es decir, y_t admite una representación ARMA(1,1) con la particularidad de que $\rho = 1$. El polinomio característico de este proceso es $\Phi(z) = 1 - \rho z = 1 - z$ y se aprecia que $\Phi(z^*) = 0$ si es que $z^* = 1$. El módulo de esta raíz es igual a uno, mostrando así que y_t **no satisface el requerimiento para ser un proceso estacionario** (es decir, $\Phi(z^*) = 0$ únicamente para $|z^*| > 1$). Se dice, en este caso, que (el polinomio característico de) y_t **contiene una raíz unitaria**, que no es más que una manera alternativa de referirse al hecho de que $y_t \sim I(1)$. Note que, en el ejemplo, $\Delta y_t \sim MA(1)$, que es estacionario.

Lo opuesto también es cierto. Considere, ahora, el proceso AR(2),

$$y_t = (\phi + \rho)y_{t-1} - \phi\rho y_{t-2} + \varepsilon_t,$$

donde $\rho = 1$ y $|\phi| < 1$. El polinomio característico es $\Phi(z) = 1 - (\phi + \rho)z + \rho\phi z^2 = (1 - \rho z)(1 - \phi z)$, y la ecuación característica $\Phi(z^*) = 0$ es resuelta para $z^* = 1/\rho = 1$ y $|z^*| = 1/|\phi| > 1$. Es decir, y_t contiene una raíz unitaria. Al considerar que $\rho = 1$, el proceso AR(2) se simplifica a

$$\Delta y_t = \phi\Delta y_{t-1} + \varepsilon_t.$$

Es decir, Δy_t se determina como un proceso AR(1) estacionario (y de media cero). Se concluye que $\Delta y_t \sim I(0)$ y, por tanto, que $y_t \sim I(1)$.

Estos ejemplos son una manifestación del siguiente principio: si $y_t \sim I(d)$, entonces la representación ARMA de y_t contiene d raíces unitarias, mientras que la representación ARMA de $\Delta^d y_t$ corresponde a la de un proceso estacionario. Conversamente, si el polinomio característico de la representación ARMA(p, q) de un proceso y_t puede factorizarse como $\Phi(z) = (1 - z)^d \Phi^*(z)$, donde $\Phi^*(z)$ es un polinomio de grado $p - d$ que satisface que $\Phi(z^*) = 0$ únicamente para $|z^*| > 1$, entonces $y_t \sim I(d)$ y la representación ARMA de $\Delta^d y_t$ tendrá como polinomio característico a $\Phi^*(z)$, por lo que $\Delta^d y_t \sim ARMA(p - d, q)$ es estacionario.

Un proceso $I(1)$ de importancia, por ser el más simple de su clase, es el **paseo aleatorio** o *random walk*. El estudio de este proceso estilizado da muchas luces de cómo la presencia de

una raíz unitaria implica no ergodicidad y no estacionariedad, y de cómo estas características alteran el comportamiento de procesos $I(1)$ con respecto a lo establecido para procesos $I(0)$. El paseo aleatorio se define como

$$z_t = z_{t-1} + \varepsilon_t,$$

donde ε_t es un ruido blanco. Partiendo de un punto inicial z_0 , tras sustituciones sucesivas, se consigue que

$$z_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots + \varepsilon_2 + \varepsilon_1 + z_0. \quad (*)$$

Se aprecia que z_t es la **acumulación de choques** (o incrementos) homocedásticos no autocorrelacionados. La palabra clave es “acumulación”, que implica una suma de variables aleatorias sin ningún tipo de descuento. La ecuación (*) nos recuerda a las representaciones MA presentadas en la sección II.1. Por ello, es útil comparar, y sobre todo contrastar, los resultados por mostrar con los de un proceso AR(1) estacionario, $u_t = \rho u_{t-1} + \varepsilon_t$ con $|\rho| < 1$. La media no condicional de este proceso es, recuerde, $\mu = 0$, y su varianza no condicional es $\gamma_0 = \sigma^2 / (1 - \rho^2)$. Note que se obtiene un paseo aleatorio cuando $\rho = 1$. La representación MA de u_t es

$$u_t = \mu + \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \cdots + \rho^{t-2} \varepsilon_2 + \rho^{t-1} \varepsilon_1 + \rho^t u_0.$$

Existen dos distinciones importantes en estas representaciones. La primera es que en un proceso estacionario es la media no condicional μ que aparece en la representación MA, mientras que z_t depende de su valor inicial z_0 . La segunda, mucho más relevante, es que la representación (*) es MA(t) y es válida toda vez que t sea **finito**. Si $t \rightarrow \infty$, la suma al lado derecho de (*) no es convergente. En cambio, en el caso estacionario, puede conseguirse una representación MA(∞) válida. Ello ocurre porque, como ya se ha sugerido, en un mundo $I(0)$ las observaciones cronológicamente distantes son ponderadas (o son descontadas) por pesos decrecientes, mientras que en el paseo aleatorio no existe tal descuento¹. Así, el efecto que una perturbación lejana ε_1 tiene sobre z_t , tiene la misma importancia que el efecto de una perturbación más cercana como ε_t . El paseo aleatorio, y los procesos $I(1)$ en general, no presentan memoria restringida y, por tanto, no son **ergódicos**.

Es simple determinar los momentos de un paseo aleatorio, una vez que condicionamos al punto de partida z_0 . Se verifica que

$$\mathbb{E}(z_t | z_0) = z_0 + \sum_{s=1}^t \mathbb{E}(\varepsilon_s) = z_0,$$

¹ Una manera inductiva de conseguir la representación MA(∞) es la siguiente: tomando a u_{t-1} como condición inicial, se tiene que $u_t = \rho u_{t-1} + \varepsilon_t$; si se toma a u_{t-2} como punto de partida (al reemplazar $u_{t-1} = \rho u_{t-2} + \varepsilon_{t-1}$), se llega a $u_t = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2}$; si se toma a u_{t-3} como punto inicial (al reemplazar $u_{t-2} = \rho u_{t-3} + \varepsilon_{t-2}$), se llega a $u_t = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \rho^3 \varepsilon_{t-3}$; este proceso se puede repetir *ad infinitum*, dando así la representación MA(∞). El principio detrás de esta inducción es el de estacionariedad: las propiedades estadísticas, tales como la representación MA, de u_t son las mismas que las de u_s para $s < t$. Alternativamente, el término $\rho^s u_{t-s}$ que emerge en el proceso inductivo tiende a cero toda vez que $|\rho| < 1$. Ello no ocurre con $\rho = 1$, negando así la posibilidad de una representación MA(∞) válida.

$$\mathbb{V}(z_t | z_0) = \sum_{s=1}^t \mathbb{V}(\varepsilon_s) = \sigma^2 t,$$

$$\mathbb{C}(z_t, z_{t-i} | z_0) = \mathbb{V}(z_{t-i} | z_0) = \sum_{s=1}^{t-i} \mathbb{V}(\varepsilon_s) = \sigma^2 (t - i),$$

y se establece la conexión entre un proceso $I(1)$ y la **no estacionariedad en covarianzas**: los momentos de z_t (en particular, las autocovarianzas) dependen de t .

Los momentos presentados, de hecho, pueden (y deben) interpretarse como el resultado de un ejercicio de predicción. Dada la condición inicial z_0 , se sabe que tras t períodos habrán ocurrido t realizaciones del ruido blanco ε_t que, acumuladas, dan como resultado z_t . Dado que estas perturbaciones tienen media cero, entonces se espera que z_t sea igual a z_0 . No obstante, se enfrenta una situación de fuerte incertidumbre en donde cada incremento en el horizonte de proyección incrementa la varianza (más precisamente, el error cuadrático medio) de esta predicción. Esta es la interpretación detrás del hecho de que $\mathbb{V}(z_t | z_0)$ no dependa de z_0 (de hecho, se ha asumido implícitamente que z_0 no correlaciona con ε_s para $s > 0$), pero sí depende del hecho de que son t las perturbaciones (sin descuento) que separan a z_0 de z_t , y en particular que $\mathbb{V}(z_t | z_0)$ crezca linealmente con t .

En el caso del proceso estacionario u_t , se tiene que

$$\mathbb{E}(u_t | u_0) = \mu + \rho^t u_0 \quad \text{y} \quad \mathbb{V}(u_t | u_0) = \sigma^2 \sum_{i=0}^{t-1} \rho^{2i}.$$

La predicción de u_t exhibe **reversión a la media**: sin importar el valor de u_0 , dado que $|\rho| < 1$, a la larga $\mathbb{E}(u_t | u_0) \rightarrow \mu$ conforme $t \rightarrow \infty$. Asimismo, **la incertidumbre está acotada**: $\mathbb{V}(u_t | u_0) \rightarrow \gamma_0$ conforme $t \rightarrow \infty$. Este último resultado es reflejo de la memoria limitada del proceso y es una forma de reversión al momento no condicional. En un paseo aleatorio, de hecho, **los momentos no condicionales** (que usualmente se obtienen como un límite $t \rightarrow \infty$) **no se encuentran bien definidos**.

Conocer las propiedades de un paseo aleatorio permite estudiar lo que ocurre con procesos $I(1)$ más generales. Un primer ejemplo es el **paseo aleatorio con tendencia** o *random walk with drift*, definido como

$$\bar{z}_t = \delta + \bar{z}_{t-1} + \varepsilon_t.$$

Tras sustituciones sucesivas, es simple verificar que

$$\bar{z}_t = \alpha + \delta t + z_t,$$

donde z_t es el paseo aleatorio **sin** tendencia que acumula ε_t y $\alpha = \bar{z}_0 - z_0$ recoge condiciones iniciales y puede ser tratado como una constante. Es decir, todo paseo aleatorio con tendencia se puede escribir como un paseo aleatorio (sin tendencia) más una tendencia

lineal. El parámetro δ , el **drift**, es la pendiente de la tendencia de \bar{z}_t . Este proceso es, en aspecto, similar al de la sección III.1, aunque sus propiedades difieren considerablemente. En particular, $\mathbb{E}(\bar{z}_t | \bar{z}_0, z_0) = \alpha + \delta t$, por lo que se trata de un proceso no estacionario en media, mientras que $\mathbb{C}(\bar{z}_t, \bar{z}_{t-j} | \bar{z}_0, z_0) = \mathbb{C}(z_t, z_{t-j} | z_0)$, por lo que se trata de un proceso que tampoco es estacionario en covarianzas.

Un segundo ejemplo, bastante más general, ocurre con

$$y_t = \delta + y_{t-1} + u_t,$$

donde u_t es un proceso estacionario arbitrario (autocorrelacionado), no necesariamente un ruido blanco. Tras sustituciones sucesivas,

$$y_t = u_t + u_{t-1} + u_{t-2} + \dots + u_2 + u_1 + y_0 + \delta t.$$

Luego, y_t es la **acumulación de choques**, en esta ocasión autocorrelacionados, más una tendencia lineal. Por ello, la estructura de los momentos condicionales de y_t dado y_0 es similar a la obtenida para \bar{z}_t . Los detalles, por supuesto, varían, ya que u_t se encuentra autocorrelacionado y este hecho se verá reflejado en una serie de términos adicionales en las expresiones para $\mathbb{V}(y_t | y_0)$ y $\mathbb{C}(y_t | y_0)$. Se mantiene, sin embargo, el importante resultado de que $\mathbb{V}(y_t | y_0) = f(t)$ creciente y $\mathbb{C}(y_t, y_{t-j} | y_0) = f(t, j)$ creciente en t .

Un ataque más formal a la relación entre y_t y z_t se obtiene a través de la denominada **descomposición de Beveridge y Nelson (1981)**. Considere que $u_t \sim \text{ARMA}(p, q)$ de la forma

$$u_t = \phi_1 u_t + \phi_2 u_{t-1} + \dots + \phi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

y defina

$$\psi = \frac{1 + \theta_1 + \theta_2 + \dots + \theta_q}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

Luego, es posible expresar u_s como

$$u_s = \psi \varepsilon_s + \eta_s - \eta_{s-1},$$

donde η_t es un proceso estacionario. Al acumular esta igualdad desde $s = 0$ hasta $s = t$, se obtiene

$$y_t = \alpha + \delta t + \psi z_t + \eta_t,$$

donde y_t es el proceso integrado general que acumula a u_t , z_t es un paseo aleatorio que acumula a ε_t y $\alpha = y_0 + \eta_0 - \psi z_0$ puede tratarse como una constante. Es decir: todo proceso integrado puede ser expresado como un paseo aleatorio (ψz_t), más una tendencia lineal ($\alpha + \delta t$), más un proceso estacionario de media cero (η_t). Cuando $\delta = 0$, dado que $z_t \sim I(1)$ y $\eta_t \sim I(0)$, ψz_t es el termino **dominante**, un concepto denominado **tendencia estocástica**, de y_t . Ello significa que, esencialmente y sobre todo asintóticamente, el comportamiento de y_t estará determinado por el comportamiento de z_t . Cuando $\delta \neq 0$, la tendencia lineal tendrá mayor preponderancia, como estudiaremos luego.

III.3 Efectos sobre la inferencia

Un buen punto de partida para estudiar las implicancias de la presencia de raíces unitarias en el comportamiento asintótico de estadísticos de interés, es el análisis del promedio muestral de un paseo aleatorio. Considere una muestra $\{z_1, z_2, \dots, z_{T-1}, z_T\}$ de T observaciones de un paseo aleatorio. El promedio muestral de estas observaciones puede ser expresado en función de los promedios de las secuencias ε_t y w_t definidas en la sección III.1. A saber, asumiendo, sin pérdida de generalidad, que $z_0 = 0$,²

$$\begin{aligned}\bar{z} &= \frac{1}{T} \sum_{t=1}^T z_t = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^t \varepsilon_s \\ &= \frac{1}{T} [\varepsilon_1 + (\varepsilon_1 + \varepsilon_2) + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) + \dots + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_{T-1} + \varepsilon_T)] \\ &= \frac{1}{T} \sum_{t=1}^T (T - t + 1) \varepsilon_t = \frac{T+1}{T} \sum_{t=1}^T \varepsilon_t - \sum_{t=1}^T \left(\frac{t}{T}\right) \varepsilon_t = T(\bar{\varepsilon} - \bar{w}) + \bar{\varepsilon}.\end{aligned}$$

A diferencia de lo que ocurre con un proceso $I(0)$, el promedio de un paseo aleatorio es igual a una variable aleatoria que **diverge**. Es decir, una variable aleatoria cuya varianza se incrementa (en lugar de reducirse) con T . Recuerde que $\mathbb{V}(\bar{\varepsilon}) \simeq \sigma^2 T^{-1}$, $\mathbb{V}(\bar{w}) \simeq \frac{1}{3} \sigma^2 T^{-1}$ y $\mathbb{C}(\bar{w}, \bar{\varepsilon}) \simeq \frac{1}{2} \sigma^2 T^{-1}$, por lo que

$$\mathbb{V}(\bar{z}) = (T+1)^2 \mathbb{V}(\bar{\varepsilon}) + T^2 \mathbb{V}(\bar{w}) - 2T(T+1) \mathbb{C}(\bar{w}, \bar{\varepsilon}) = A_T T \quad \text{donde} \quad A_T \rightarrow \frac{1}{3} \sigma^2.$$

El hecho de que $\mathbb{V}(\bar{z}) = A_T T$ sugiere que $\mathbb{V}(T^{-1/2} \bar{z}) = A_T$, por lo que se requiere dividir la suma de los z_t por una potencia mayor que T , específicamente $T^{3/2}$, para conseguir una expresión convergente. Así, se llega al resultado

$$\frac{1}{T^{3/2}} \sum_{t=1}^T z_t = \sqrt{T}(\bar{\varepsilon} - \bar{w}) + \frac{\bar{\varepsilon}}{\sqrt{T}} \xrightarrow{d} N\left(0, \frac{1}{3} \sigma^2\right) + 0 \equiv N\left(0, \frac{1}{3} \sigma^2\right),$$

donde se ha utilizado el Teorema del Límite Central. Este análisis permite concluir que los momentos de un paseo aleatorio, y en general de todo proceso integrado, presentan asintóticamente un comportamiento muy distinto al de los momentos de un proceso estacionario. Por ejemplo, el promedio, adecuadamente dividido por una potencia de T , converge **en distribución** a una variable aleatoria en lugar de converger **en probabilidad** a una expectativa, como ocurre en el caso estacionario.

Para documentar mejor esta comparación, el programa `Prog32_RandomWalk.prg` implementa un ejercicio de Monte Carlo ilustrativo. El código relevante de generación de datos es el siguiente:

² En caso $z_0 \neq 0$, el término z_0/T aparecerá en la definición de \bar{z} . Al margen de si z_0 es fijo o aleatorio, este término, claramente, se encuentra completamente dominado y es asintóticamente irrelevante.

```
SERIES e = NRND
GENR u - !rho*u(-1) = (1 - !rho)*!mu + e
GENR z - z(-1) = e
```

donde, primero, se genera una serie estacionaria u gobernada por un proceso AR(1) con media μ y coeficiente de persistencia ρ , $u_t - \rho u_{t-1} = (1 - \rho)\mu + \varepsilon_t$. La idea es que u_t se asemeje a un paseo aleatorio pero mantenga su estacionariedad. Por ello, se utilizan $\mu = 0$ y $\rho = 0.975$, un valor bastante cercano, pero estrictamente menor, a uno. Segundo, el paseo aleatorio se genera siguiendo la definición $z_t - z_{t-1} = \varepsilon_t$. La serie e recoge las realizaciones del ruido blanco ε_t . Se reportarán resultados para $\varepsilon_t \sim N(0, \sigma^2)$ con $\sigma^2 = 1$, de modo que todos los promedios involucrados sean normalmente distribuidos. Si se desea, este supuesto puede relajarse y considerar un proceso asimétrico o de colas anchas (es recomendable mantener la calibración $\sigma^2 = 1$) para ver operar al Teorema del Límite Central con mayor claridad.

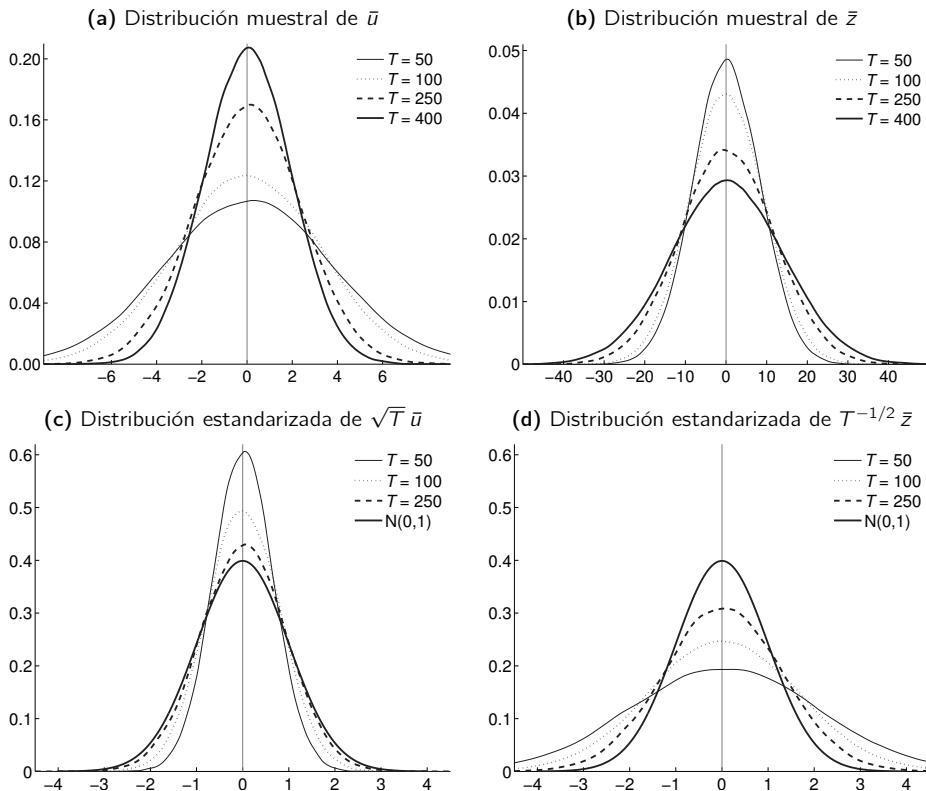
Por su parte, para un tamaño muestral T , que es variante, y en la repetición i , se calculan distintas versiones de los promedios muestrales \bar{z} e \bar{u} y se almacenan de la siguiente manera:

```
!ubar = @MEAN(u)
DMu_{!T}(!i) = !ubar
DAu_{!T}(!i) = (1 - !rho)*@SQRT(!T)*(!ubar - !mu)
!zbar = @MEAN(z)
DMz_{!T}(!i) = !zbar
DAz_{!T}(!i) = @SQRT(3/!T)*!zbar .
```

El escalar \bar{u} contiene el promedio muestral de u_t , que se almacena en la serie $DMu_{!T}(!i)$. Por el Teorema Ergódico, se espera que esta serie se concentre más en torno a cero para valores mayores de T . Además, la serie $DAu_{!T}(!i)$ recoge un estadístico que es asintóticamente normal estándar. Este es idéntico al considerado en el programa Prog21_LeyesRevisadas de la sección II.2, y es igual a $\sqrt{T/\Omega_u}(\bar{u} - \mu)$, donde Ω_u es la varianza de largo plazo de u_t que, en el caso del proceso AR(1), es igual a $\Omega_u = (1 - \rho)^{-2}$. Por su parte, el escalar \bar{z} contiene el promedio muestral del paseo aleatorio z_t , que se almacena en la serie $DMz_{!T}(!i)$. Se espera que, siguiendo el análisis de esta sección, esta serie diverja conforme T se incremente. La serie $DAz_{!T}(!i)$ recoge un estadístico que es también asintóticamente normal estándar. Este es $\sqrt{3/T}\bar{z}$. El gráfico III.3 presenta las distribuciones muestrales de los estadísticos descritos, utilizando $R = 100,000$ repeticiones.

El panel (a) del gráfico III.3 presenta la distribución muestral de \bar{u} . Al ser u_t estacionario, la Ley de Grandes Números opera. La varianza del promedio muestral va reduciéndose con el tamaño muestral, de modo que la distribución muestral de \bar{u} va concentrándose alrededor de $\mathbb{E}(\bar{u}) = 0$. En el límite, \bar{u} colapsa a 0. Por su parte, el panel (c) presenta la

Gráfico III.3 Comportamiento asintótico de promedios muestrales de un paseo aleatorio



Notas: (Prog32_RandomWalk.prg) Densidades kernel (automáticas) para $R = 100,000$ repeticiones. En las simulaciones, se generan las series $z_t = z_{t-1} + \varepsilon_t$ e $u_t = \rho u_{t-1} + \varepsilon_t$ con $z_0 = u_0 = 0$, $\phi = 0.975$ y $\varepsilon_t \sim N(0, 1)$, y se descartan las primeras $T_0 = 50$ observaciones. La escala del panel (a) es distinta de la del panel (b). El estadístico del panel (c) es multiplicado por $(1 - \rho)$, mientras que el del panel (d) es multiplicado por $\sqrt{3}$.

distribución muestral de $(1 - \rho)\sqrt{T} \bar{u}$, que se aprecia converge a $N(0, 1)$. Estos resultados son completamente comparables con los del gráfico II.1.

El panel (b) del gráfico III.3 presenta la distribución muestral de \bar{z} . Se aprecia que, en claro contraste con \bar{u} , la varianza de \bar{z} se incrementa con T y la distribución muestral se va haciendo cada vez más plana. Esta no es más que una manifestación de uno de los resultados más importantes de esta sección: \bar{z} **diverge** conforme $T \rightarrow \infty$. Por su parte, el panel (d) presenta la distribución muestral de $\sqrt{3/T} \bar{z}$. Se sabe que $\sqrt{1/T} \bar{z}$ converge en distribución a una normal con media 0 y varianza igual a $1/3$, por lo que se verifica que $\sqrt{3/T} \bar{z}$ converge a $N(0, 1)$. Más allá de las diferencias en las tasas de convergencia,

Cuadro III.1 Convergencia de promedios de series estacionarias y no estacionarias

$u_t \sim I(0)$	$z_t = z_{t-1} + \varepsilon_t \sim I(1)$
$\frac{1}{T} \sum_{t=1}^T u_t \xrightarrow{p} \mathbb{E}(u_t) = \mu$	$\frac{1}{T^{3/2}} \sum_{t=1}^T z_t \xrightarrow{d} \sigma \mathcal{D}_1$
$\frac{1}{T} \sum_{t=1}^T u_t^2 \xrightarrow{p} \mathbb{E}(u_t^2) = \gamma_0 + \mu^2$	$\frac{1}{T^2} \sum_{t=1}^T z_t^2 \xrightarrow{d} \sigma^2 \mathcal{D}_2$
$\frac{1}{T} \sum_{t=1}^T u_{t-1} \Delta u_t \xrightarrow{p} \mathbb{E}(u_{t-1} \Delta u_t) = \gamma_1 - \gamma_0$	$\frac{1}{T} \sum_{t=1}^T z_{t-1} \Delta z_t \xrightarrow{d} \sigma^2 \mathcal{D}_3$

Notas: El paseo aleatorio z_t acumula ruidos blancos de media cero y varianza σ^2 . \mathcal{D}_i ($i = 1, 2, 3$) denota una variable aleatoria. Por su parte, u_t es un proceso estacionario con media μ y autocovarianzas γ_i .

respecto al caso estacionario, el patrón de convergencia es también distinto. En el caso de \bar{u} , panel (c), las distribuciones muestrales para valores moderados de T son leptocúrticas, mientras que en el caso de \bar{z} , panel (d), son platocúrticas.

Es bueno enfatizar en las diferencias en el comportamiento asintótico de promedios de una serie estacionaria con los promedios de un paseo aleatorio. En primer lugar, **las tasas de convergencia varían**. Mientras que en el caso estacionario la suma de funciones $g(u_t)$ precisa ser dividida entre T , en el caso del paseo aleatorio la suma de funciones $g(z_t)$ ha de ser dividida entre T^n , donde típicamente $n > 1$. En segundo lugar, **el modo de convergencia es distinto**. Mientras que en el caso estacionario, el Teorema Ergódico implica la convergencia en probabilidad a escalares, en el caso no estacionario esta convergencia es en distribución a una variable aleatoria.

Si \mathcal{D}_i denota una variable aleatoria, el cuadro III.1 resume cualitativamente estas diferencias. El resultado concerniente al promedio de z_t , adecuadamente estabilizado al dividir por $T^{-1/2}$, es el estudiado en el gráfico III.3. Así, sabemos que la variable aleatoria \mathcal{D}_1 es normal con media cero y varianza $\frac{1}{3}$. Puede establecerse un resultado también en términos de funciones de distribución conocidas, en particular una variable χ_1^2 , para el promedio de la secuencia $z_{t-1} \Delta z_t$, es decir la variable aleatoria \mathcal{D}_3 , aunque ese detalle no es relevante para los fines de este texto. No obstante, no es posible caracterizar a la variable aleatoria \mathcal{D}_2 en términos de distribuciones usuales, por lo que optamos por utilizar este enfoque más cualitativo para todo promedio de funciones de un paseo aleatorio. Note que la variable \mathcal{D}_2 es el límite del promedio de z_t^2 , el cuadrado de un paseo aleatorio, dividido entre T .

Los resultados del cuadro III.1 son los ingredientes básicos para el análisis de varios problemas de interés que involucran series no estacionarias. Por ejemplo, suponga que $\Delta y_t = \delta + u_t$ es un proceso $I(1)$ con tendencia cuya descomposición de Beveridge y Nelson es $y_t = \alpha + \delta t + \psi z_t + \eta_t$. ¿Qué podemos decir, por ejemplo, respecto al promedio de esta

serie? Al promediar la identidad anterior, se consigue

$$\bar{y} = \alpha + \delta \left(\frac{T+1}{2} \right) + \psi \bar{z} + \bar{\eta}.$$

Tanto la constante α (trivialmente) como el promedio $\bar{\eta}$ convergen. El promedio del paseo aleatorio \bar{z} debe ser dividido entre $T^{1/2}$ para converger, mientras que el promedio de la tendencia lineal debe ser dividido entre T . Así, es claro que el comportamiento asintótico de \bar{y} es dominado por esta tendencia. Con ello, se concluye que

$$\frac{\bar{y}}{T} = \frac{\alpha}{T} + \frac{1}{T}(\bar{\eta}) + \frac{\psi}{\sqrt{T}} \left(\frac{\bar{z}}{\sqrt{T}} \right) + \frac{\delta}{2} \left(\frac{T+1}{T} \right) \xrightarrow{p} 0 + 0 \cdot \mathbb{E}(\eta_t) + 0 \cdot \sigma \mathcal{D}_1 + \frac{\delta}{2} \equiv \frac{\delta}{2}.$$

No obstante, si $\delta = 0$, de modo que el proceso no presenta un *drift*, entonces \bar{z} pasa a ser el término dominante. En este caso,

$$\frac{\bar{y}}{\sqrt{T}} = \frac{\alpha}{\sqrt{T}} + \frac{1}{\sqrt{T}}(\bar{\eta}) + \psi \left(\frac{\bar{z}}{\sqrt{T}} \right) \xrightarrow{d} 0 + 0 \cdot \mathbb{E}(\eta_t) + \psi \cdot \sigma \mathcal{D}_1 \equiv \sigma \psi \mathcal{D}_1,$$

de modo que el comportamiento de \bar{y} es similar al de \bar{z} . Es interesante notar que el término $\sigma\psi$ que aparece en la distribución anterior es la **desviación estándar de largo plazo** de u_t .

III.4 Removiendo tendencias

Por supuesto, los resultados de la sección anterior sobre el comportamiento anómalo de promedios de procesos integrados tendrán profundas implicancias para la inferencia con series no estacionarias. Esta sección estudia una situación bastante ilustrativa, en donde un investigador debe decidir cómo modelar una tendencia observada en los datos. La sección se basa en resultados analíticos de Durlauf y Phillips (1988), pero que son presentados de manera más accesible a través de un estudio de Monte Carlo.

Suponga que y_t es un paseo aleatorio con tendencia $y_t = \delta + y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim iid(0, \sigma^2)$. Usualmente, en macroeconometría, y_t es el logaritmo de alguna variable económica de interés (por ejemplo, algún agregado macroeconómico), por lo que δ aproxima la tasa de crecimiento porcentual de esta variable. Así, es de interés general inferir sobre esta pendiente δ . Para tal propósito, contamos con dos alternativas. Primero, utilizamos la definición (el proceso generador de datos) de y_t y escribimos la regresión

$$\Delta y_t = \delta + \text{error}_t.$$

Este es un modelo de regresión muy simple, en donde Δy_t , que es estacionario, es regresado sobre una constante. El estimador MCO en este caso es idéntico al promedio muestral de Δy_t . Luego, dado que $\Delta y_t - \delta = \varepsilon_t$, el error muestral de este estimador, $\hat{\delta} - \delta$, es simplemente el promedio del ruido blanco ε_t . Con ello, concluimos que

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \sigma^2).$$

Este es un modelo estacionario en diferencias (**DS**). Dado que $\Delta y_t = \delta + \varepsilon_t$, el modelo estimado coincide con el proceso generador de datos. En este sentido, se espera que el estimador $\hat{\delta}$ sea el que presente mejores propiedades entre alternativas. Conviene mencionar que los residuos de esta regresión son iguales a $e_t = \Delta y_t - \hat{\delta} = \varepsilon_t - (\hat{\delta} - \delta) = \varepsilon_t - \bar{\varepsilon} \sim I(0)$. La tendencia en los datos es removida tras diferenciar la serie.

El propósito último de estimar δ es conducir ejercicios de inferencia estadística sobre su valor. Considere la hipótesis nula $H_0 : \delta = \delta_0$, que puede ser contrastada mediante una prueba t . Sea $\hat{\sigma}_\delta$ la desviación estándar estimada de $\hat{\delta}$. En este caso, $\hat{\sigma}_\delta = s/\sqrt{T}$, donde s^2 es el promedio del cuadrado de los residuos, que es un estimador consistente de σ^2 , la varianza de ε_t . Luego, si a_T denota un término (posiblemente aleatorio) que converge a cero conforme T se incrementa, el ratio t que contrasta H_0 se puede escribir como

$$\tau_{\delta_0} = \frac{\hat{\delta} - \delta_0}{\hat{\sigma}_\delta} = \frac{\sqrt{T}(\hat{\delta} - \delta_0)}{s} = \frac{\sqrt{T}(\hat{\delta} - \delta)}{s} + \frac{\sqrt{T}(\delta - \delta_0)}{s} = \tau_\delta + \frac{\sqrt{T}(\delta - \delta_0)}{\sigma} + a_T.$$

En esta expresión, τ_δ es el ratio t cuando el valor impuesto por la hipótesis nula coincide con el verdadero valor de δ . Nos referimos a τ_δ como el ratio t "centrado". Es simple verificar, dada la normalidad asintótica de $\hat{\delta}$ y la consistencia de s , que τ_δ converge a una variable aleatoria normal estándar. El segundo término, por su parte, se asocia con los efectos que una hipótesis nula falsa podría tener sobre τ_{δ_0} . Así, conforme $T \rightarrow \infty$, se tiene que

$$\tau_{\delta_0} \xrightarrow{d} N(0, 1) \text{ si } H_0 \text{ es verdadera, y } |\tau_{\delta_0}| \rightarrow \infty \text{ si } H_0 \text{ es falsa.}$$

Bajo H_0 , existe una probabilidad de rechazo que da origen al Error Tipo I. Esta probabilidad es calibrada por el investigador, en este caso, a partir de los percentiles de la función de distribución normal. Es interesante notar, sin embargo, que τ_{δ_0} **diverge** si H_0 es falsa. Ello quiere decir que ante una hipótesis nula falsa, se obtendrán valores del ratio t muy grandes (y crecientes con T), lo que haría cada vez más probable que la prueba de hipótesis asociada caiga en zona de rechazo. De hecho, esta divergencia implica que conforme $T \rightarrow \infty$, $\Pr(|\tau_{\delta_0}| > \text{Valor crítico}) \rightarrow 1$, o, puesto de otro modo, $\Pr(\text{rechazar } H_0 | H_0 \text{ falsa}) \rightarrow 1$. La **potencia** de la prueba converge a su máximo valor, por lo que la prueba es **consistente**. Esta es, sin duda, una característica deseable de la prueba t , al proveer al investigador una herramienta sumamente valiosa que lo asiste en el proceso de aprender sobre δ y la naturaleza de y_t .

Una segunda manera de estimar δ es a través del modelo de tendencia lineal

$$y_t = \delta t + \text{error}_t,$$

donde por simplicidad se asume que las condiciones iniciales implican un intercepto igual a cero. Dado el supuesto de que $\Delta y_t = \delta + \varepsilon_t$, esta es una regresión mal especificada. La razón es que el proceso generador de datos es $y_t = \delta t + z_t$ y, por tanto, el error de la regresión de y_t sobre una tendencia lineal no es estacionario. Se asume, equivocadamente, que y_t es un proceso estacionario en tendencia (modelo **TS**). Compare esto con lo estudiado

en la sección III.1, donde esta regresión sí correspondía a un modelo bien especificado. Sin ahondar en detalles, es posible determinar que el estimador MCO de δ para este modelo satisface

$$\sqrt{T}(\tilde{\delta} - \delta) \xrightarrow{d} N\left(0, \frac{6}{5}\sigma^2\right).$$

Comparado con el estimador DS, se aprecia una pérdida de eficiencia, ya que ambos estimadores son consistentes y normalmente distribuidos, pero el estimador TS tiene como varianza asintótica $\frac{6}{5}\sigma^2$, mientras que la del estimador DS es σ^2 . No obstante esta pérdida de eficiencia, que en la práctica puede ser intrascendente, las mayores diferencias aparecen en el momento de hacer inferencia.

En esta ocasión, los residuos son claramente no estacionarios, $\tilde{e}_t = z_t - (\tilde{\delta} - \delta)t$, y es posible mostrar que \tilde{s}^2/T , donde \tilde{s}^2 es el promedio del cuadrado de los residuos, converge en distribución a variable aleatoria \mathcal{D}_s . Más aún, la desviación estándar estimada de $\tilde{\delta}$ es $\tilde{\sigma}_\delta \simeq (\tilde{s}^2/T)^{1/2}(1/T)$. Así, el ratio t que contrasta $H_0: \delta = \delta_0$ se puede escribir como

$$\tilde{t}_{\delta_0} = \frac{\tilde{\delta} - \delta_0}{\tilde{\sigma}_\delta} = \frac{T(\tilde{\delta} - \delta_0)}{\tilde{s}/\sqrt{T}} = \frac{T(\tilde{\delta} - \delta)}{\tilde{s}/\sqrt{T}} + \frac{T(\delta - \delta_0)}{\tilde{s}/\sqrt{T}} = \sqrt{T} \frac{\mathcal{D}_0}{\sqrt{\mathcal{D}_s}} + \frac{T(\delta - \delta_0)}{\sqrt{\mathcal{D}_s}} + a_T,$$

donde \mathcal{D}_0 denota la distribución límite de $\sqrt{T}(\tilde{\delta} - \delta)$. Se aprecia que, conforme $T \rightarrow \infty$,

$$|\tilde{t}_{\delta_0}| \rightarrow \infty \text{ en todo caso.}$$

Independientemente del valor elegido de δ_0 , el ratio t **siempre diverge**. Si H_0 es verdadera, la divergencia es a una tasa $T^{1/2}$, mientras que si H_0 es falsa, la divergencia es a una tasa T . El motivo es que el denominador $\tilde{\sigma}_\delta$ se aproxima a cero muy rápidamente, dando así la impresión de que el estimado de δ obtenido con el modelo TS es muy preciso y altamente significativo. Esta es, sin embargo, una distorsión generada por la no estacionariedad del residuo de la regresión.

En la práctica, esta significación estadística artificial puede llevar a un investigador a favorecer al modelo TS como la mejor manera de abordar la inferencia sobre δ y concluir, asimismo, que el método adecuado de remoción de la tendencia de y_t es a través del residuo de la regresión TS. Estas conclusiones son equivocadas y esconden la verdadera naturaleza de y_t , que es estacionaria en diferencias. Esto es, $\Delta y_t \sim I(0)$, mientras que $y_t - \tilde{\delta}t \sim I(1)$. Remover una tendencia lineal no reditúa un proceso estacionario, a diferencia de lo estudiado en la sección III.1, en donde el modelo TS coincidía con el proceso generador de datos³.

El programa Prog33_DurlaufPhillips.prg ilustra estas conclusiones a través de simulaciones. Ambos modelos, DS y TS, son estimados sobre datos simulados para tamaños muestrales variantes y los estadísticos t son almacenados. El proceso generador de datos es el paseo aleatorio con tendencia,

³ Note, además, la sutileza de que el estimador presentado en la sección III.1 resultaba ser superconsistente, a diferencia de $\tilde{\delta}$ que es "simplemente" consistente. Asimismo, cuando el modelo TS es el correcto, el promedio de los residuos al cuadrado resulta ser consistente para σ^2 , en lugar de converger, tras ser dividido por T , a una variable aleatoria inusual.

$$\text{GENR } y - y(-1) = \delta + e,$$

donde δ recoge el valor de δ (que no es particularmente importante) y e es una serie con observaciones de un ruido blanco normalmente distribuido (SERIES $e = \text{NRND}$). El código relevante es, para una tamaño muestral T y en la repetición i ,

```
EQUATION TS.LS y time C
Dt1TS_{!T}(!i) = (TS.C(1) - delta)/TS.@STDERRS(1)
Dt2TS_{!T}(!i) = TS.@TSTATS(1)

EQUATION DS.LS D(y) C
Dt1DS_{!T}(!i) = (DS.C(1) - delta)/DS.@STDERRS(1)
Dt2DS_{!T}(!i) = DS.@TSTATS(1) .
```

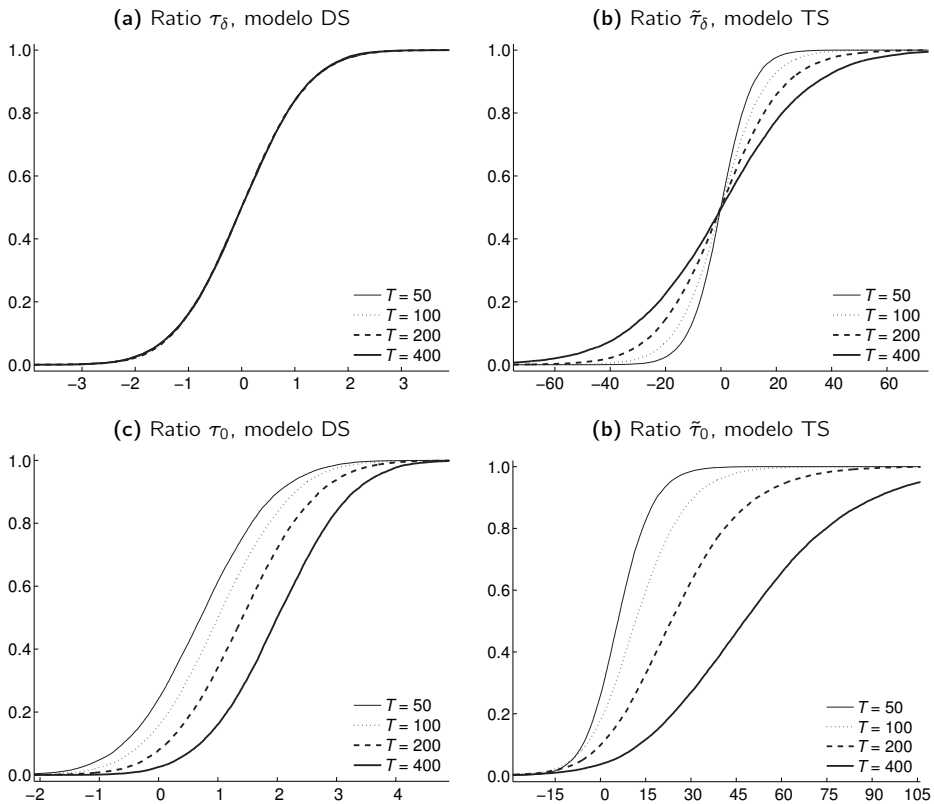
Dados los datos, se estima primero el modelo TS (ecuación TS, donde `time` es una tendencia lineal generada con el comando `SERIES time = @TREND`). Luego, se calcula τ_δ , el ratio t centrado, que se almacena en la serie `Dt1TS_{!T}`, así como el ratio τ_0 , que contrasta la hipótesis (falsa) $H_0 : \delta = 0$, que se almacena en la serie `Dt2TS_{!T}`. De manera análoga, luego se estima el modelo DS (ecuación DS) y se almacenan los ratios $\tilde{\tau}_\delta$ y $\tilde{\tau}_0$ en las series `Dt1TS_{!T}` y `Dt2TS_{!T}`, respectivamente.

El gráfico III.4 presenta las funciones de distribución empíricas de estos ratios para $R = 100,000$. En el panel (a) se presenta la función asociada con τ_δ , el ratio t centrado en el modelo DS, que es el modelo que corresponde al proceso generador de datos. Se aprecia que este ratio es básicamente distribuido como una variable normal estándar, para todos los tamaños muestrales considerados en la simulación. El panel (c) presenta el comportamiento de este ratio bajo hipótesis falsas. Se aprecia cómo las funciones de distribución empíricas del panel (a) se desplazan hacia la derecha, reflejando así divergencia del estadístico y consistencia del contraste. Según nuestros resultados analíticos, con $\sigma = 1$ y $\delta_0 = 0$, estos desplazamientos son iguales a $\sqrt{T}\delta$.

Los paneles (b) y (d) del gráfico III.4 presentan los resultados del modelo TS. En el panel (b) se aprecia cómo $\tilde{\tau}_\delta$, el ratio t centrado, diverge. El rango de variación de $\tilde{\tau}_\delta$ crece con T y, a la larga, se aprecia que podría tomar valores sumamente elevados con alta probabilidad. Se aprecia, además, un comportamiento distinto al de una variable normal, confirmando así lo descrito analíticamente. La divergencia de $\tilde{\tau}_\delta$ ante una hipótesis falsa mostrada en el panel (d) es particularmente severa, siendo esta la distorsión más importante que la no estacionariedad del residuo en la regresión TS genera.

Un mensaje es claro: en la práctica es importante que seamos capaces de distinguir entre los modelos TS y DS (otro ejemplo es el de predicción con estos modelos, que estudiamos con detalle en la sección III.9). No obstante, encontramos que el instrumental usual de inferencia podría verse distorsionado al trabajar con procesos integrados. Cómo ajustar estos procedimientos es una pregunta que las siguientes secciones intentan responder.

Gráfico III.4 *Distribuciones empíricas de ratios t para la pendiente de la tendencia*



Notas: (Prog33_DurlaufPhillips.prg) Funciones de distribución acumuladas para $R = 100,000$ repeticiones. El ratio τ_δ es centrado (la hipótesis nula impone el verdadero valor de δ), mientras que el ratio τ_0 contrasta $H_0 : \delta = 0$, que es una hipótesis falsa. El modelo TS estima δ como la pendiente de una tendencia lineal, mientras que el modelo DS estima δ como el promedio de Δy_t . El proceso generador de datos es el modelo DS.

III.5 Pruebas de raíz unitaria

Existen varios enfoques que buscan responder la pregunta de si una serie de tiempo tiene las características de un proceso estacionario o, por el contrario, es aproximada mejor como un proceso integrado, no estacionario. Las más conocidas, y que son materia de discusión en este texto, formulan el problema como una prueba de hipótesis aplicada sobre un coeficiente en una ecuación de regresión dinámica. Estas son las **pruebas de Dickey y Fuller**, estudiadas inicialmente en Dickey y Fuller (1979); Dickey, Bell y Fuller (1986); y Bhargava (1986).

A continuación, describimos los aspectos más saltantes de estos contrastes. En secciones posteriores discutimos algunas de las muchas extensiones propuestas en la literatura.

La versión más básica de estas pruebas consiste en estimar ρ en la regresión

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

y contrastar la hipótesis $H_0 : \rho = 1$. Bajo H_0 , y_t presenta una raíz unitaria y es un proceso integrado, en este caso un paseo aleatorio. La hipótesis alternativa es $H_1 : \rho < 1$, por lo que se trata de una prueba de una sola cola. Los casos explosivos $|\rho| > 1$ típicamente son ignorados por tratarse de situaciones poco representativas del comportamiento de series de tiempo económicas. Luego, bajo la hipótesis alternativa, y_t es un proceso AR(1) estacionario con media cero.

Las propiedades del estimador MCO de ρ dependen fuertemente de cuál es el supuesto mantenido sobre el verdadero valor de ρ , es decir de cuál es el proceso generador de datos. Si $\rho < 1$ (es decir, bajo la hipótesis alternativa), entonces $y_t \sim I(0)$ y $\hat{\rho}$ tendrá el comportamiento analizado en la sección II.4 (véase, en especial, el gráfico II.4):

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} N(0, 1 - \rho^2),$$

que es un resultado que proviene de aplicar el Teorema del Límite Central a promedios de procesos ergódicos. Note que la varianza asintótica de $\hat{\rho}$ es igual a $1 - \rho^2$ y es positiva si $\rho < 1$. No obstante, esta varianza es igual a cero cuando $\rho = 1$ (es decir, bajo la hipótesis nula), por lo que

$$\sqrt{T}(\hat{\rho} - 1) \xrightarrow{p} 0.$$

Esta es una típica manifestación de **superconsistencia** y es una de muchas distorsiones que $\rho = 1$ conlleva. Es acá donde es bueno enfatizar que es necesario ser cuidadosos en la inferencia, ya que en el contraste de hipótesis, **se asume que la hipótesis nula es verdad** hasta que se presente evidencia suficiente para rechazarla. Es decir, en el contexto de la prueba de Dickey y Fuller debe asumirse que y_t es un paseo aleatorio como el estudiado en la sección III.3, lo que genera comportamientos inusuales en los estadísticos involucrados en el proceso de inferencia.

El estimador MCO de ρ satisface, por construcción,

$$\hat{\rho} - 1 = \frac{\sum_t y_{t-1} \Delta y_t}{\sum_t y_{t-1}^2}.$$

Dado que $y_t \sim I(1)$, el numerador de $\hat{\rho} - 1$ dividido entre T convergerá a una variable aleatoria, mientras que para obtener convergencia en el denominador se precisa dividir entre T^2 . En este caso, **el denominador crece indefinidamente** (ya que es una suma de cuadrados) **a una tasa mayor que la del numerador**, que es la fuerza detrás de la superconsistencia de $\hat{\rho}$. Más formalmente, tras multiplicar la igualdad anterior por T ,

$$T(\hat{\rho} - 1) = \frac{\sum_t y_{t-1} \Delta y_t / T}{\sum_t y_{t-1}^2 / T^2} \xrightarrow{d} \frac{\mathcal{D}_3}{\mathcal{D}_2},$$

donde \mathcal{D}_2 y \mathcal{D}_3 son las variables aleatorias presentadas en el cuadro III.1. Así, dado que $T(\hat{\rho} - 1)$ converge a una variable aleatoria, entonces $T^n(\hat{\rho} - 1) \simeq T^{n-1}(\mathcal{D}_3/\mathcal{D}_2)$ converge a cero para cualquier $n < 1$, confirmando así la superconsistencia.

La distribución a la que converge $T(\hat{\rho} - 1)$, es decir la distribución de la variable aleatoria $\mathcal{D}_3/\mathcal{D}_2$, es bastante inusual. En particular, **no tiene una forma funcional “cerrada”** (es decir, una fórmula explícita para su función de densidad), lo que en principio complica su caracterización (por ejemplo, el cálculo de sus momentos y percentiles) analíticamente. Este hecho no es, hoy por hoy, una limitante, ya que esta distribución puede ser aproximada con altos grados de precisión a partir de simulaciones. Así, a pesar de nuestra incapacidad de obtener una “fórmula”, las técnicas de Monte Carlo nos permiten conocer todos los detalles relevantes de esta distribución.

El programa Prog34_DickeyFuller.prg, entre otras cosas, simula esta distribución. Para un tamaño muestral dado T , que en este caso no varía, se genera la serie y como un paseo aleatorio, bajo el supuesto de que $\rho = 1$. Ello se consigue, en la repetición i , con el comando GENR y $y(-1) = \text{NRND}$. Luego, se estima una regresión del tipo EQUATION eq.LS y $y(-1)$ y se procede a almacenar el coeficiente estimado $\hat{\rho}$ en la serie rhohat (previamente declarada) de tamaño R , $\text{rhohat}(i) = \text{eq.C}(1)$. La línea rotulada “A” en el panel (a) del gráfico III.5 muestra la aproximación a la distribución muestral de $\hat{\rho}$ tras $R = 100,000$ repeticiones de este experimento, mientras que la línea de nombre “A” en el panel (b) es la función de distribución empírica. Para obtener $T(\hat{\rho} - 1)$, simplemente hay que restar el valor de 1 y multiplicar por T los ejes horizontales de ambos paneles.

La distribución de $\mathcal{D}_3/\mathcal{D}_2$ amerita varios comentarios. En particular, es **muy distinta a la normal**, que es el tipo de distribución que emerge en el caso estacionario: es visiblemente asimétrica, con una cola ancha hacia la izquierda. Así, si bien la moda de la distribución es 1, su media es inequívocamente menor de 1. Ello revela un sesgo en el estimador MCO, ya que $\mathbb{E}(\hat{\rho}) < \rho = 1$. Este sesgo se relaciona con el estudiado en la sección II.4 (gráfico II.3), aunque es exacerbado por la no estacionariedad de y_t .

A pesar de la superconsistencia, que sugiere que este sesgo desaparece en muestras relativamente grandes, se tiene que $\hat{\rho}$ es un estimador **fuertemente sesgado** en muestras finitas. Ello tiene profundas implicancias sobre nuestra habilidad para discriminar una serie estacionaria ($\rho < 1$) de una no estacionaria ($\rho = 1$). En concreto, con el resultado de las simulaciones es posible determinar, mediante una frecuencia relativa, que $\text{Pr}(\hat{\rho} < 1) \simeq 0.68$. Es decir, en un contexto de muestreo repetido, **en dos tercios de las muestras generadas por un paseo aleatorio, el coeficiente estimado $\hat{\rho}$ será menor que el verdadero valor de 1**. De ignorar este fenómeno, podría concluirse equivocadamente que un estimado de, por ejemplo, $\hat{\rho} \simeq 0.95$ provee evidencia a favor de un proceso AR(1) estacionario con media cero, cuando es un evento altamente probable detrás de la estimación con un paseo aleatorio. De hecho, el procedimiento correcto es el de comparar el estimado $\hat{\rho}$ contra el valor crítico de la distribución del gráfico III.5: si se concluye que $\hat{\rho}$ es lo suficientemente menor de 1 para que $T(\hat{\rho} - 1)$ sea menor que ese valor crítico, entonces se estaría rechazando

$H_0 : \rho = 1$ a favor de H_1 .

Es poco usual explorar aproximaciones a la distribución muestral de un estimador al contrastar hipótesis en un modelo de regresión. Es mucho más común implementar una prueba t para $H_0 : \rho = 1$. Asimismo, es también muy común que en una prueba t se iguale a cero el coeficiente de interés bajo la hipótesis nula. Así, si se define $\phi = \rho - 1$, se tendría que $H_0 : \phi = 0$. Puede estimarse ϕ como $\hat{\phi} = \hat{\rho} - 1$ o directamente a través de la regresión

$$\Delta y_t = \phi y_{t-1} + \varepsilon_t \quad \text{donde} \quad \phi = \rho - 1.$$

Note que simplemente se ha restado y_{t-1} a ambos lados de $y_t = \rho y_{t-1} + \varepsilon_t$. Así, bajo $H_0 : \phi = 0$, y_t es el paseo aleatorio $\Delta y_t = \varepsilon_t$. Esta no es más que una reparametrización del modelo AR(1), que tiene como ventaja que los paquetes econométricos como EViews producen automáticamente el ratio t de $H_0 : \phi = 0$ como el resultado de sus estimaciones (y no así el ratio t de $H_0 : \rho = 1$).

Si s^2 denota el estimador usual de la varianza del error de regresión (el promedio de los residuos al cuadrado), el ratio t de la prueba de significación del estimador de ϕ es

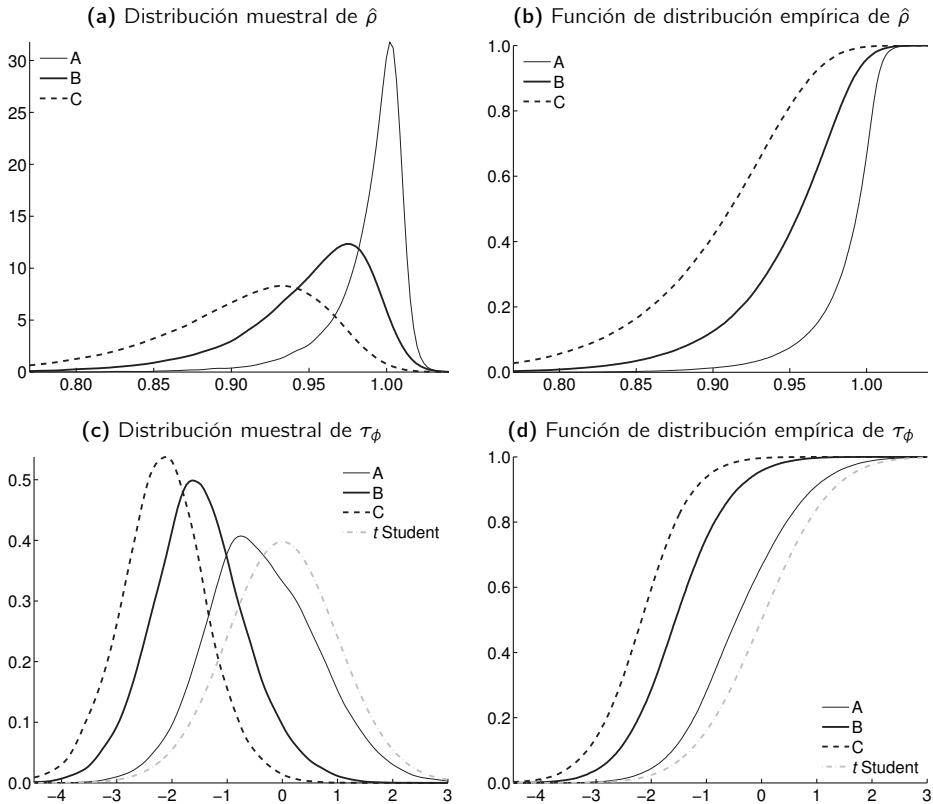
$$\tau_\phi = \frac{\sum_t y_{t-1} \Delta y_t}{\sqrt{s^2 \sum_t y_{t-1}^2}} \xrightarrow{d} \frac{\mathcal{D}_3}{\sqrt{\mathcal{D}_2}} \equiv DF_A.$$

La distribución DF_A es conocida como la **distribución de Dickey y Fuller**. Esta distribución, por supuesto, ha de determinarse mediante simulaciones, como se hace en el programa Prog34_DickeyFuller.prg. Aquí, se repite un gran número de veces la secuencia: generar un paseo aleatorio, GENR $y - y(-1) = \text{NRND}$, estimar la regresión reparametrizada, EQUATION eq.LS $D(y) \ y(-1)$, y almacenar el ratio t , `tstat(!i) = eq.@TSTATS(1)`. La línea "A" en el panel (c) del gráfico III.5 muestra la distribución DF_A y la compara con una distribución t de Student con T grados de libertad, que es válida para la inferencia con series estacionarias. El panel (d) presenta las funciones de distribución acumuladas.

El sesgo al estimar ρ (o ϕ) se traslada a la distribución DF_A , que no es centrada en cero, a diferencia de la distribución t de Student. El percentil 5 resultante de las simulaciones es `@QUANTILE(tstat, 0.05) $\simeq -1.94$` , y es menor que el percentil correspondiente a la distribución t , `@QTDIST(0.05, T) $\simeq -1.66$` . Ello quiere decir que, con un nivel de 5% de significación, se rechaza la hipótesis de raíz unitaria si el ratio t calculado, en una muestra en particular, es $\hat{\tau}_\phi < -1.94$, mientras que H_0 no es rechazada si $\hat{\tau}_\phi > -1.94$. Note que el requerimiento para rechazar H_0 es mayor que el necesario bajo estacionariedad ($\hat{\tau}_\phi < -1.66$). Ello es el resultado de haber ajustado (implícitamente) el valor crítico por el sesgo en la estimación de ρ (o ϕ), lo que se consigue al utilizar DF_A , que es la distribución que efectivamente describe el comportamiento de τ_ϕ bajo la hipótesis nula, en lugar de la distribución t del mundo estacionario.

En resumen, se especifica la **regresión de contraste** $\Delta y_t = \phi y_{t-1} + \text{error}_t$ y se contrasta, con una prueba t , si $H_0 : \phi = 0$. Los valores críticos **correctos** provienen

Gráfico III.5 Distributions de Dickey-Fuller



Notas: (Prog34_DickeyFuller.prg) Los paneles (a) y (c) muestran las densidades kernel (automáticas) de "la raíz" $\hat{\rho}$ y el ratio t que contrasta $H_0 : \rho = 1$ para un tamaño muestral de $T = 100$. Los paneles (b) y (d) presentan las funciones de distribución acumuladas correspondientes. Los casos A, B y C se refieren a las distintas especificaciones en la regresión de contraste especificadas en el cuadro III.2, para $\phi = \rho - 1$. Se utilizaron $R = 100,000$ repeticiones.

de la distribución DF_A , y se encuentran tabulados en libros de textos o implementados en paquetes econométricos como EViews. El estudio de MacKinnon (1996) es la referencia más completa al respecto. Este autor realiza un estudio de Monte Carlo de gran escala y provee fórmulas simples como

$$DF(\alpha) = \zeta_{0\alpha} + \frac{\zeta_{1\alpha}}{T} + \frac{\zeta_{2\alpha}}{T^2},$$

para aproximar, con muchísima precisión, el α -ésimo percentil de la distribución de Dickey y Fuller. Los coeficientes ζ en la expresión anterior son estimados, también con gran precisión,

por regresión: $\zeta_{0\alpha}$ es el percentil de la distribución asintótica ($T \rightarrow \infty$), mientras que los dos términos restantes incorporan un ajuste para muestras finitas. MacKinnon (1996) provee fórmulas similares para los valores críticos de otras distribuciones, que discutiremos pronto, y sus resultados serán utilizados para el cálculo de estos valores en nuestros ejercicios de simulación posteriores.

El contraste descrito no es necesariamente el más apropiado. Ello se da porque la regresión de contraste permite únicamente dos posibilidades, una de ellas potencialmente cuestionable: la primera es que y_t sea un paseo aleatorio, lo que ocurre bajo la hipótesis nula; la segunda es que sea un proceso estacionario **de media cero**, lo que ocurre bajo la hipótesis alternativa. Ante ello, ¿cómo interpretar que H_0 no sea rechazada? Puede ocurrir que, en efecto, y_t sea un paseo aleatorio; pero también puede ocurrir que y_t sea una serie estacionaria pero **en torno a una media distinta de cero**. Acá, es posible que, aunque falso, el modelo bajo H_0 sea una mejor descripción de los patrones de y_t que el modelo bajo H_1 (que también es falso), por lo que se favorece a H_0 , pero no por los méritos de H_0 sino por los deméritos de H_1 . En otras palabras, **el contraste puede confundir una mala especificación de la hipótesis alternativa con una raíz unitaria**.

Para ilustrar este punto, suponga que $u_t = \mu + \varepsilon_t$. Es decir, u_t es un ruido blanco con una media distinta de cero, por lo que sus autocovarianzas son $\gamma_0 = \mathbb{E}(u_t^2) - \mu^2 = \sigma^2$ y $\gamma_1 = \mathbb{E}(u_t u_{t-1}) - \mu^2 = 0$ y su primera autocorrelación es $\gamma_1/\gamma_0 = 0$. Ahora, considere el coeficiente estimado de una regresión de u_t sobre u_{t-1} . Este es, para una muestra grande,

$$r_1 = \frac{\sum_t u_{t-1} u_t}{\sum_t u_{t-1}^2} \simeq \frac{\mathbb{E}(u_{t-1} u_t)}{\mathbb{E}(u_{t-1}^2)} = \frac{\gamma_1 + \mu^2}{\gamma_0 + \mu^2} = \frac{\mu^2}{\sigma^2 + \mu^2} > 0.$$

Incluso se puede tener $r_1 \simeq 1$ si μ es grande relativo a σ^2 . Es decir, se podría tener la falsa impresión de que la serie u_t es persistente, por tener un coeficiente de autocorrelación muestral elevado, cuando en realidad es un ruido blanco. La fuente de confusión es, por supuesto, la mala especificación del modelo de regresión. Este es un sesgo generado por la omisión de un regresor importante, el intercepto. De hecho, es altamente probable que en una regresión bien especificada, de u_t sobre u_{t-1} y un intercepto, el coeficiente asociado con u_{t-1} (que es cero en el proceso generador de datos) no sea significativo.

Considere el siguiente proceso generador de datos:

$$y_t = \alpha + u_t \quad \text{donde} \quad u_t = \rho u_{t-1} + \varepsilon_t.$$

Bajo este planteamiento, $H_0 : \rho = 1$ implica que u_t es un paseo aleatorio y, por tanto, y_t es también un paseo aleatorio, ya que la constante α es absorbida por la condición inicial, $y_0 = \alpha + u_0$. Por el contrario, si $H_1 : \rho < 1$, entonces u_t es un proceso AR(1) estacionario con media cero, mientras que y_t es también un proceso AR(1) estacionario pero **con media α que es arbitraria**. Al reemplazar $u_t = y_t - \alpha$ en la ecuación en diferencias que gobierna a u_t , se consigue

$$y_t = (1 - \rho)\alpha + \rho y_{t-1} + \varepsilon_t \quad \text{o, compactamente,} \quad \Delta y_t = c + \phi y_{t-1} + \text{error}_t,$$

Cuadro III.2 *Contrastes de raíz unitaria de Dickey y Fuller*

Regresión de contraste	$H_0 : \phi = 0$	$H_1 : \phi < 0$	Resultado
A $\Delta y_t = \phi y_{t-1} + \text{error}_t$	$y_t = y_{t-1} + \varepsilon_t$	$y_t \sim I(0), \mathbb{E}(y_t) = 0$	$\tau_\phi \xrightarrow{d} DF_A$
B $\Delta y_t = c + \phi y_{t-1} + \text{error}_t$	$y_t = y_{t-1} + \varepsilon_t$	$y_t \sim I(0)$	$\tau_\phi \xrightarrow{d} DF_B$
C $\Delta y_t = c + d t + \phi y_{t-1} + \text{error}_t$	$y_t = \delta + y_{t-1} + \varepsilon_t$	$y_t \sim TS$	$\tau_\phi \xrightarrow{d} DF_C$

donde, como antes, $\phi = \rho - 1$. Así, al simplemente permitir que la regresión de contraste incluya un intercepto, la prueba de Dickey y Fuller permite discriminar entre un paseo aleatorio y un proceso estacionario más general.

Análogamente, puede pensarse en una formulación del problema tal que y_t presente una clara tendencia. Considere ahora el proceso generador de datos

$$y_t = \mu + \delta t + u_t \quad \text{donde} \quad u_t = \rho u_{t-1} + \varepsilon_t.$$

Ahora, $H_0 : \rho = 1$ implica que u_t es un paseo aleatorio y, por tanto, y_t es la suma de una tendencia lineal $\alpha + \delta t$, más un paseo aleatorio; es decir, y_t es un paseo aleatorio con tendencia $\Delta y_t = \delta + \varepsilon_t$. Por su parte, si $H_1 : \rho < 1$, entonces u_t es un proceso AR(1) estacionario con media cero, mientras que y_t es también un proceso AR(1) estacionario pero **en torno a la tendencia** $\alpha + \delta t$. Al reemplazar $u_t = y_t - \alpha - \delta t$ en la ecuación en diferencias que gobierna a u_t , se consigue

$$y_t = (1 - \rho)\alpha + \rho\delta + (1 - \rho)\delta t + \rho y_{t-1} + \varepsilon_t \quad \text{o} \quad \Delta y_t = c + d t + \phi y_{t-1} + \text{error}_t.$$

Luego, al permitir que la regresión de contraste incluya un intercepto y una tendencia lineal, la prueba de Dickey y Fuller permite discriminar entre un proceso DS más general y un proceso estacionario TS, también más general. El cuadro III.2 presenta un resumen de las hipótesis nulas y alternativas detrás de diversas especificaciones de la regresión de contraste.

Un fenómeno particular (y ciertamente raro) de las regresiones con series no estacionarias es que la forma de las distribuciones de los estadísticos $T(\hat{\rho} - 1)$ y τ_ϕ **depende de la especificación del modelo en el que ρ se estima**, o en otras palabras del modelo bajo la hipótesis alternativa. En particular, dependen de qué **regresores determinísticos** (intercepto o tendencia lineal, o ambos) son incluidos en la regresión de contraste. Por ello, en el cuadro III.2 aparecen, según sea el caso, las distribuciones DF_A , DF_B y DF_C . Estas distribuciones también son simuladas en el programa Prog34_DickeyFuller.prg. Se tiene, para la repetición i ,

```
EQUATION eqA.LS D(y) y(-1) ' Modelo sin términos determinísticos
rhatA(!i) = eqA.C(1) + 1
tstatA(!i) = eqA.OTSTATS(1)
```

EQUATION eqB.LS D(y) y(-1) C ' Modelo con constante

rhohatB(!i) = eqB.C(1) + 1

tstatB(!i) = eqB.@TSTATS(1)

EQUATION eqC.LS D(y) y(-1) C time ' Modelo con constante y tendencia

rhohatC(!i) = eqC.C(1) + 1

tstatC(!i) = eqC.@TSTATS(1) .

Para una misma realización del paseo aleatorio, se estiman tres versiones de la ecuación de contraste: una básica sin regresores determinísticos (eqA), una que incluye un intercepto (eqB) y una que incluye, además, la tendencia lineal time (eqC). Estas ecuaciones estiman ϕ , por lo que el estimador de ρ es $\hat{\rho} = \hat{\phi} + 1$, un estadístico que se almacena, según sea el caso, en las series rhohatA, rhohatB y rhohatC. Asimismo, las series tstatA, tstatB y tstatC, recopilan los estadísticos τ_ϕ que contrastan la hipótesis nula de raíz unitaria.

El gráfico III.5 muestra los resultados. Conforme se van incluyendo regresores determinísticos, la asimetría en la distribución muestral de $\hat{\rho}$ se vuelve más pronunciada. Ello ocurre porque el sesgo en $\hat{\rho}$ se incrementa. Con el resultado de las simulaciones es posible determinar, mediante una frecuencia relativa, que $\Pr(\hat{\rho} < 1) \simeq 1$ para los casos B y C. Es decir, bajo muestreo repetido, **en casi todas las muestras generadas por un paseo aleatorio, el coeficiente estimado $\hat{\rho}$ será menor de 1** en estos contrastes.

Respecto al estadístico τ_ϕ , el sesgo de estimar ρ se traslada a las distribuciones DF_B y DF_C , que se desplazan hacia la izquierda de la distribución DF_A . En el caso de DF_B , el percentil 5 obtenido en las simulaciones se reduce de cerca de -1.94 a $@@QUANTILE(tstatB, 0.05) \simeq -2.90$, mientras que el percentil correspondiente a la distribución DF_C es $@@QUANTILE(tstatC, 0.05) \simeq -3.45$. Dado que estos percentiles constituyen los valores críticos para el contraste de la hipótesis de raíz unitaria, se aprecia que el requerimiento para rechazar H_0 es progresivamente mayor, conforme se agregan términos determinísticos a la regresión de contraste. En otras palabras, para un mismo proceso generador de datos es menos probable el rechazo de H_0 en el caso C que en el caso B y, a su vez, que en el caso A. Luego estudiaremos las implicancias de este comportamiento.

Dado que las distribuciones DF_A , DF_B y DF_C son visiblemente distintas, surge la inquietud: en la práctica, ¿qué versión del contraste es recomendable utilizar? O sea, ¿por cuál regresión de contraste se debe optar y qué valores críticos se deben utilizar? La respuesta es la respuesta a otra pregunta: si se rechazara H_0 y favoreciéramos a H_1 , ¿qué modelo estacionario es el que describiría mejor los datos? Si la respuesta es un modelo sin tendencia que permite, además, restringir la media a cero, entonces el contraste de no estacionariedad corresponde al caso A; si la respuesta es un modelo sin tendencia pero con una media distinta de cero, el contraste adecuado es el del caso B; finalmente, si se percibe una tendencia en los datos, habría que favorecer el contraste C. Por ello, una práctica altamente recomendable antes de embarcarse en el contraste de raíz unitaria es **visualizar los datos** para identificar “al ojo” medias distintas de cero, o tendencias.

III.6 Potencia y consistencia de las pruebas de Dickey y Fuller

Regresemos a la regresión de contraste básica, $\Delta y_t = \phi y_{t-1} + \text{error}_t$. Hemos analizado el comportamiento de la estimación por MCO y la inferencia con ratios t bajo $H_0 : \phi = 0$. Vale la pena preguntarse: ¿qué ocurre bajo $H_1 : \phi < 0$? En particular, ¿cuáles son las propiedades del ratio τ_ϕ en este contexto?

Recuerde que los promedios de funciones de $y_t \sim I(0)$ convergen, por el Teorema Ergódico, en probabilidad a esperanzas. De este modo,

$$\begin{aligned} \tau_\phi &= \frac{\sum_t y_{t-1} \Delta y_t}{\sqrt{s^2 \sum_t y_{t-1}^2}} = \sqrt{T} \left(\frac{\sum_t y_{t-1} \Delta y_t / T}{\sqrt{s^2 \sum_t y_{t-1}^2 / T}} \right) \xrightarrow{p} \left(\lim_{T \rightarrow \infty} \sqrt{T} \right) \frac{\mathbb{E}(y_{t-1} \Delta y_t)}{\sqrt{\sigma^2 \mathbb{E}(y_{t-1}^2)}} \\ &= \left(\lim_{T \rightarrow \infty} \sqrt{T} \right) \frac{\gamma_1 - \gamma_0}{\sqrt{\sigma^2 \gamma_0}} = -\infty, \end{aligned}$$

conforme $T \rightarrow \infty$. El ratio τ_ϕ **diverge** hacia menos infinito cuando $y_t \sim I(0)$. Ello quiere decir que cuando H_0 es falsa, se obtendrán valores negativos del ratio τ_ϕ muy grandes (y crecientes con T), lo que haría cada vez más probable que la prueba de hipótesis asociada caiga en zona de rechazo. Esta divergencia implica que conforme $T \rightarrow \infty$, $\Pr(\tau_\phi < \text{Valor crítico}) \rightarrow 1$ o, puesto de otro modo, $\Pr(\text{rechazar } H_0 | H_0 \text{ falsa}) \rightarrow 1$. La **potencia** de la prueba de Dickey y Fuller converge a su máximo valor (el Error Tipo II converge a cero), por lo que la prueba es **consistente**. En el límite, el contraste de Dickey y Fuller rechazará H_0 con probabilidad igual a uno si el proceso generador de datos es estacionario. Si bien es cierto que el caso analizado es el más básico, se obtiene la misma conclusión en los casos B (cuando la regresión de contraste incorpora un intercepto) y C (intercepto y tendencia lineal).

La consistencia es un resultado asintótico. En muestras finitas, no obstante, la potencia de la prueba puede ser considerablemente menor de 1, dependiendo del tamaño muestral y del proceso generador de datos. Para analizar estas posibilidades, el programa Prog35_DFPotencia.prg implementa un estudio de Monte Carlo.

El proceso generador de datos es el proceso AR(1) con media cero, $y_t = \rho y_{t-1} + \varepsilon_t$, donde ρ toma distintos valores menores o iguales a uno. En particular, se define un vector `rhovec`, de dimensión `nrho`, que contiene una grilla de posibles valores de ρ . En los experimentos ejecutados, se consideran 31 valores equidistantes de ρ entre 0.5 y 1. Así, dado un valor de ρ contenido en el escalar `rho`, en la repetición `i`, los datos se generan según

$$\text{GENR } y = \text{rho} * y(-1) + \text{NRND},$$

donde se asume que $\varepsilon_t \sim N(0, 1)$. Considerando que se descartan `T0` observaciones iniciales, se incorpora el efecto de un tamaño muestral cambiante. Para una muestra de

tamaño $!T$ se restringe la muestra con $SMPL \ !T0+1 \ !T0+!T$, se estima la ecuación de Dickey y Fuller, y se pasa a tomar una decisión sobre el contraste de raíz unitaria:

```

EQUATION eqA.LS D(y) y(-1)
IF eqA.@TSTATS(1) < (-1.9393 - 0.398/!T) THEN
    rejectionA{!T}(!i) = 1
ENDIF
EQUATION eqB.LS D(y) y(-1) C
IF eqB.@TSTATS(1) < (-2.8621 - 2.738/!T - 8.36/!T^2) THEN
    rejectionB{!T}(!i) = 1
ENDIF
EQUATION eqC.LS D(y) y(-1) C time
IF eqC.@TSTATS(1) < (-3.4126 - 4.039/!T - 17.83/!T^2) THEN
    rejectionC{!T}(!i) = 1
ENDIF .

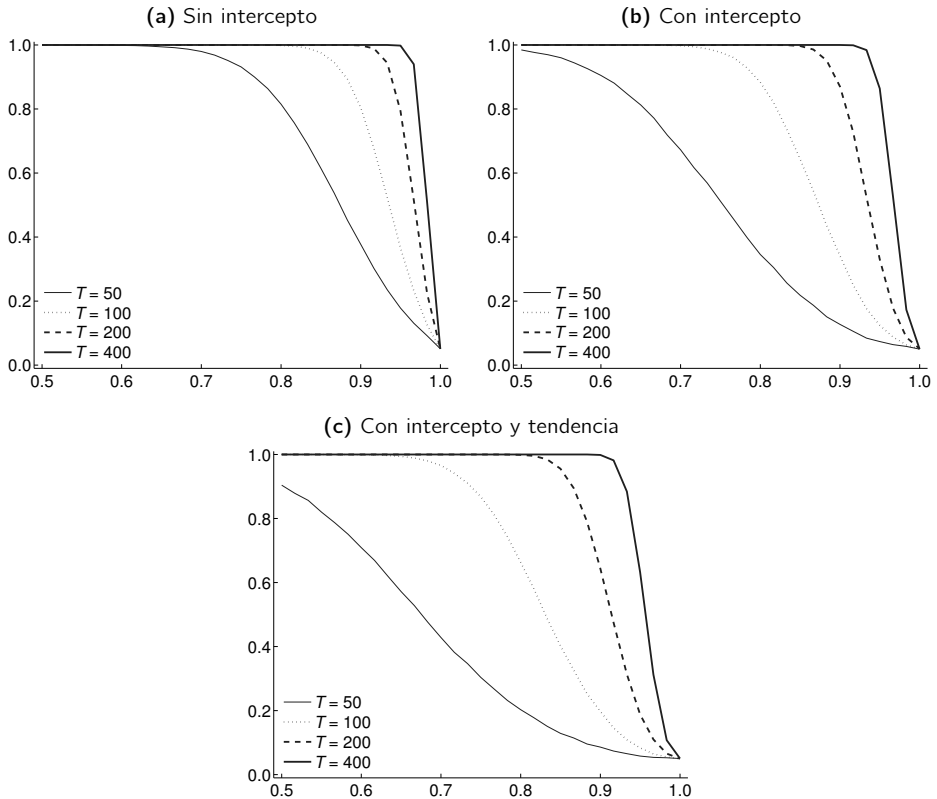
```

Las ecuaciones eqA, eqB y eqC, como ya se ha visto, contienen los resultados de las regresiones de contraste A, B y C, respectivamente. El ratio τ_ϕ se recupera, para cada caso, con los escalares eq.@TSTATS(1), que corresponden al primer regresor $y(-1)$. Así, si este es menor que el valor crítico correspondiente —calculado según las fórmulas de MacKinnon (1996) para los tres casos y un nivel de 5% de significación—, la hipótesis de raíz unitaria es rechazada. El evento del rechazo se registra al volver 1 la $!i$ -ésima observación del vector rejectionA{!T}, rejectionB{!T} o rejectionC{!T}. Estos son vectores de $!R$ observaciones, que claramente dependen de $!T$, y que al declararse contienen únicamente valores de 0. Al término de las repeticiones, los promedios de los elementos de estos vectores, @MEAN(rejectionA{!T}), @MEAN(rejectionB{!T}) y @MEAN(rejectionC{!T}), contienen aproximaciones de la probabilidad de rechazo de las respectivas pruebas, con $!T$ observaciones y un valor de ρ igual a $!rho$. Estas proporciones son, finalmente, almacenadas en la matrices ResultadosA, ResultadosB y ResultadosC, que tienen tantas filas como distintos valores de ρ y tantas columnas como distintos valores de T . Luego, se repite el análisis para el próximo valor de ρ .

El gráfico III.6 muestra los resultados con $R = 100,000$ repeticiones. El panel (a) muestra “curvas de potencia” para el primer contraste. Sobre el eje horizontal se ordenan los distintos valores de ρ considerados, mientras que en el eje vertical se tiene la proporción de rechazos. Las distintas líneas se asocian con distintos tamaños muestrales. El panel (b) es análogo pero para la prueba que incluye un intercepto en la regresión de contraste, y el panel (c) considera el contraste con intercepto y tendencia.

Es bueno mencionar que para valores de $\rho < 1$, la proporción de rechazos aproxima $\Pr(\text{rechazar } H_0 | H_0 \text{ falsa})$, que es la potencia de la prueba; por otro lado, para $\rho = 1$ la proporción de rechazos aproxima $\Pr(\text{rechazar } H_0 | H_0 \text{ verdadera})$, que es la probabilidad de

Gráfico III.6 Probabilidad de rechazo en las pruebas de Dickey y Fuller



Notas: (Prog35_DFPotencia.prg) Curvas de potencia con $R = 100,000$ repeticiones. El proceso generador de datos es el modelo AR(1), $y_t = \rho y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$. Las figuras muestran la proporción de rechazos de $H_0 : \rho = 1$ (eje vertical) ante distintos valores de ρ (eje horizontal).

cometer Error Tipo I. Dado que se han elegido los valores críticos correspondientes al 5% de significación, no debe sorprendernos que todas las curvas colapsen al punto 0.05 cuando $\rho = 1$. Además, se aprecia la consistencia del contraste: incrementos en T incrementan, primero, la probabilidad de rechazo de las pruebas para un valor dado de ρ y, segundo, el rango de valores de $\rho < 1$ para los cuales la potencia es 1.

Más allá de verificar la consistencia de las pruebas de Dickey y Fuller, el gráfico III.6 revela dos resultados interesantes. Primero, la potencia decrece rápidamente conforme ρ se va acercando a uno. De hecho, en muestras finitas, **las pruebas de raíz unitaria pueden tener dificultades para diferenciar un paseo aleatorio de un proceso estacionario persistente**. Dependiendo del tamaño muestral y del contraste utilizado, la prueba podría tener baja

potencia incluso para valores reducidos de ρ , del orden de $\rho \simeq 0.70$. Es decir, las pruebas de raíz unitaria bien podrían confundir un proceso AR(1) con parámetro de persistencia de $\rho \simeq 0.70$ con un paseo aleatorio.

Segundo, las curvas del panel (a) se ubican sistemáticamente por encima de las curvas comparables del panel (b). Se concluye que, en muestras finitas, **la presencia de regresores determinísticos en la regresión de contraste reduce la potencia de la prueba**. Esta es una conclusión, en general, cierta. La reducción de la potencia es aún más notoria en el panel (c), donde la regresión de contraste incluye una tendencia lineal.

La literatura estadística documenta ampliamente que el **sobreajuste** (*overfitting*, es decir estimar modelos sobreparametrizados, con variables redundantes) típicamente se asocia con una reducción de la potencia de pruebas de hipótesis. En el caso de los contrastes de Dickey y Fuller, la pérdida de potencia ante la presencia de regresores determinísticos es una consecuencia directa del desplazamiento de la distribución DF descrito en la sección anterior, que conlleva que el requerimiento para rechazar H_0 sea más exigente conforme se incluyen más términos determinísticos en la regresión de contraste. Ello ocurre, a su vez, por la necesidad de estimar ciertos parámetros en la ecuación de contraste, el intercepto y la pendiente de t , que no forman parte del proceso generador de datos, llevándonos a una situación de sobreajuste. A pesar de que estos parámetros pueden ser estimados consistentemente y en muestras grandes se aproximen a sus verdaderos valores (de cero), el error muestral que introducen en la estimación contamina la prueba, reduciendo su potencia. Ante ello, es siempre recomendable evitar, en la medida de lo posible, estimar coeficientes que, *a priori*, se espere no sean estadísticamente significativos.

Esta conclusión permite brindar pautas prácticas para el uso de las pruebas de raíz unitaria. Primero, se sugiere implementar el contraste C, que incluye intercepto y tendencia. Si es que H_0 es rechazada, hemos llegado a un resultado categórico. Recuerde que el contraste C tiene, en general, baja potencia, por lo que un rechazo no es algo que ocurra frecuentemente por casualidad. Si H_0 no es rechazada, es recomendable inspeccionar los datos. Si la tendencia en estos es muy marcada, entonces concluimos y no rechazamos H_0 ; pero si la presencia de una tendencia no es muy obvia, es posible que el no rechazo sea simplemente el reflejo de una baja potencia. Quizá pueda aumentarse la potencia si en lugar de estimar la pendiente de t , le imponemos el valor de cero. En este caso, pasamos al contraste B. Si ahora H_0 es rechazada, hemos llegado a un resultado categórico. El contraste B tiene la menor potencia entre los contrastes sin tendencia, por lo que un rechazo no es algo que ocurra al azar con alta probabilidad. Si H_0 no es rechazada, es recomendable observar los datos. Si la media es obviamente distinta de cero, entonces concluimos y no rechazamos H_0 ; pero si no es muy fácil concluir que la media sea distinta de cero, es posible que el no rechazo sea producto de una baja potencia. Podrían conseguirse incrementos en la potencia al imponer una media de cero, en lugar de estimarla. Así, finalmente, pasamos al contraste A y tomamos las decisiones finales.

III.7 Incrementando la potencia con una prueba LM

La baja potencia del contraste de Dickey y Fuller en muestras finitas, sobre todo cuando la regresión de contraste incluye una tendencia lineal, ha motivado mucha investigación que busca maneras, muchas veces creativas, de incrementarla. Un ejemplo bastante citado es el estudio de Schmidt y Phillips (1992)⁴. Ellos adoptan un enfoque de **multiplicadores de Lagrange (LM)**, por las siglas de *Lagrange multipliers*, que es un principio de inferencia ampliamente aceptado. La idea detrás del enfoque LM es obtener estadísticos imponiendo varias de las consecuencias detrás de la hipótesis nula.

Considere el caso en que $\Delta y_t = \delta + \varepsilon_t$, es decir y_t es un paseo aleatorio con tendencia. Por el álgebra de la estimación por MCO, es simple verificar que la ecuación de contraste $\Delta y_t = c + d t + \phi y_{t-1} + \text{error}_t$ produce un estimador $\hat{\phi}$ y un ratio $\tau_{\hat{\phi}}$ idénticos a los que se obtendrían de la regresión

$$\Delta y_t = \text{intercepto} + \phi \hat{S}_{t-1} + \text{error}_t,$$

donde \hat{S}_{t-1} es el residuo de una regresión MCO de y_{t-1} sobre un intercepto y una tendencia lineal. Es decir, \hat{S}_t se obtiene al remover una tendencia lineal de y_t , estimando los parámetros de esta tendencia con un modelo **en niveles**. La propuesta de Schmidt y Phillips (1992) es conducir una prueba de raíz unitaria similar, es decir basada en el ratio t de $H_0 : \phi = 0$, pero en el modelo

$$\Delta y_t = \text{intercepto} + \phi \tilde{S}_{t-1} + \text{error}_t,$$

donde \tilde{S}_t también se obtiene al remover una tendencia lineal de y_t , pero estimando los parámetros de esta tendencia con un modelo **en diferencias**. En el modelo en diferencias la pendiente de la tendencia lineal puede ser estimada como el promedio de Δy_t ,

$$\tilde{\delta} = \frac{1}{T-1} \sum_{t=2}^T \Delta y_t = \frac{y_T - y_1}{T-1},$$

de modo que

$$\tilde{S}_t = y_t - (y_1 - \tilde{\delta}) - \tilde{\delta} t,$$

donde $y_1 - \tilde{\delta}$ es un estimador de la condición inicial y_0 . Nótese que, por construcción, $\tilde{S}_1 = \tilde{S}_T = 0$.

El argumento detrás de esta idea se basa en resultados que hemos obtenido en la sección III.4. Si y_t es un paseo aleatorio con tendencia, la manera más eficiente de estimar los parámetros de la tendencia es mediante un modelo en diferencias (el modelo DS). La

⁴ Otro enfoque bastante celebrado es el denominado *GLS detrending* de Elliott, Rothenberg y Stock (1996). Este procedimiento se encuentra completamente automatizado en EViews. Incorporamos el enfoque en nuestros estudios de Monte Carlo, pero no conseguimos diferencias sustanciales respecto a los resultados del enfoque de Schmidt y Phillips (1992). Por este motivo, decidimos (por ahora), no discutir acerca de *GLS detrending* en este texto.

manera alternativa, un modelo en niveles (el modelo TS), produce resultados ineficientes y propensos a distorsionar la inferencia. Esta aleatoriedad adicional hace que \hat{S}_t , de la prueba de Dickey y Fuller, tenga propiedades más complicadas y oscuras que \tilde{S}_t , de la prueba LM propuesta. En este sentido, las propiedades más simples de \hat{S}_t , argumentan Schmidt y Phillips, nos llevan a suponer que contrastes basados en este residuo ostentarán una mayor potencia. Intuitivamente, al ser el modelo en diferencias el correcto bajo H_0 , se alivia la carga del sobreajuste al utilizar $\tilde{\delta}$ como estimador de la pendiente de la tendencia, y $y_1 - \tilde{\delta}$ como estimador del intercepto.

El programa Prog36_SchmidtPhillips.prg implementa un estudio de Monte Carlo que investiga las bondades de esta propuesta. El proceso generador de datos es el de un modelo con tendencial lineal

```
GENR u = !rho*u(-1) + NRND
SERIES y = !alpha + !delta*time + u,
```

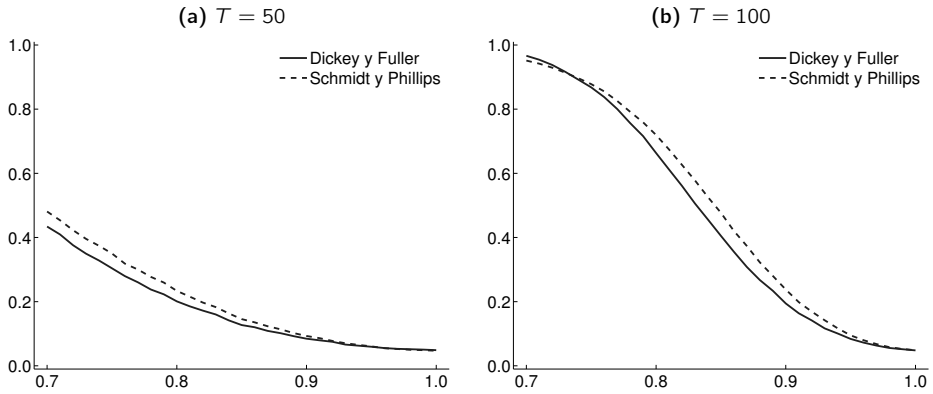
donde !alpha y !delta contienen los parámetros de $\mathbb{E}(y_t)$ y !rho almacena valores cambiantes del coeficiente de persistencia ρ . Estos valores varían con un contador llamado !irho. La lógica de programación es muy similar a la del programa Prog35_DFPotencia.prg de la sección anterior, aunque se han alterado intencionalmente algunos detalles para explorar maneras alternativas, y quizá más eficientes computacionalmente, de llegar a los mismos resultados. En particular, se define previamente una matriz R{!T} llena de ceros, con tantas filas como valores de ρ y tres columnas. En la primera columna, se coloca el valor de ρ ; en la segunda irá la proporción de rechazos en la prueba de Dickey y Fuller; y en la tercera, la proporción de rechazos en la prueba LM.

Luego, en una iteración dada, para un valor de ρ asociado con el contador !irho y un tamaño muestral !T, una primera parte del código relevante es:

```
SMPL !T0+1 !T0+!T
EQUATION eq.LS D(y) y(-1) C time
IF eq.@TSTATS(1) < (-3.4126 - 4.039/!T - 17.83/!T^2) THEN
    R{!T}(!irho, 2) = R{!T}(!irho, 2) + 1/!R
ENDIF .
```

Claramente, estas líneas evalúan la potencia de la prueba de Dickey y Fuller, como en el programa Prog35_DFPotencia.prg. La ecuación de contraste C se corre en el objeto eq. Luego, se identifica un rechazo al comparar el estadístico t asociado con $y(-1)$, eq.@TSTATS(1), con el valor crítico \hat{a} la MacKinnon (1996). En caso de detectar un rechazo, se le agrega 1/!R a la !irho-ésima fila de la segunda columna de R{!T}. Al término del experimento, esta fila contendrá la suma de rechazos dividida entre !R, es decir la proporción que aproxima $\Pr(\text{rechazo de } H_0)$.

Gráfico III.7 Probabilidad de rechazo en las pruebas DF y LM



Notas: (Prog36_SchmidtPhillips.prg) Curvas de potencia con $R = 100,000$ repeticiones. El proceso generador de datos es el modelo AR(1) con tendencia, $y_t = \alpha + \delta t + u_t$, donde $u_t = \rho u_{t-1} + \varepsilon_t$ y $\varepsilon_t \sim N(0, 1)$. Las figuras muestran la proporción de rechazos de $H_0 : \rho = 1$ (eje vertical) ante distintos valores de ρ (eje horizontal).

Del mismo modo, en lo referente a la prueba LM, se tiene que:

```
!deltatilde = (y(!T0+!T) - y(!T0+1))/(!T - 1)
SERIES S = y - (y(!T0+1) - !deltatilde) - !deltatilde*time
EQUATION eq.LS D(y) S(-1) C
IF eq.@TSTATS(1) < crit_lm(!it) THEN
    R{!T}(!irho, 3) = R{!T}(!irho, 3) + 1/!R
ENDIF.
```

Primero, se calcula el residuo \tilde{S}_t , tal y como se ha detallado previamente, y se registra en la serie S. Seguidamente, se estima una ecuación de contraste que regresa a Δy_t sobre \tilde{S}_{t-1} y una constante para luego identificar los rechazos de H_0 , comparando los estadísticos t asociados con $y(-1)$, $\text{eq.}@TSTATS(1)$, con los valores críticos provistos en Schmidt y Phillips (1992: 264; cuadro 1A). Estos valores han sido previamente almacenados en un vector crit_lm y se asocian con tamaños muestrales que varían con el contador $!it$. Por ejemplo, para $T = 50$ el valor crítico, al 5% de significación, es -3.11 , mientras que asciende a -3.06 para $T = 100$. Ante un rechazo, se agrega $1/!R$ a la $!irho$ -ésima fila de la tercera columna de $R\{!T\}$. Al término del experimento, esta fila contendrá la proporción que aproxima $\text{Pr}(\text{rechazo de } H_0)$.

El gráfico III.7 muestra los resultados para $R = 100,000$ repeticiones. Ciertamente, el contraste LM es más potente que el de Dickey y Fuller. Las diferencias, sin embargo, son bastante moderadas. Si bien es cierto que implementar el contraste LM es sencillo, y permite las generalizaciones que veremos más adelante para la prueba de Dickey y Fuller,

las ganancias en términos de potencia no parecen ser muy sustanciales. Por este motivo, seguimos en adelante explorando el enfoque más tradicional de Dickey y Fuller.

III.8 ¿Retorno a la normalidad?

En secciones anteriores se ha determinado que las pruebas de raíz unitaria se pueden ver distorsionadas si es que la ecuación de contraste incluye regresores determinísticos que no forman parte del proceso generador de datos. Cabe preguntarse: ¿qué ocurriría en una situación contraria, en donde más bien se excluyen aspectos importantes del modelo poblacional en la regresión de contraste? Para responder esta pregunta, analizaremos, y luego simularemos, un caso propuesto por West (1988).

Suponga que el proceso generador de datos es un paseo aleatorio con tendencia, $\Delta y_t = \delta + \varepsilon_t$ o, equivalentemente, $y_t = \delta t + z_t$, donde $\Delta z_t = \varepsilon_t$. Se parte de una condición inicial de cero, sin pérdida de generalidad. Como se discutió en la sección III.3, el **término dominante** en el comportamiento de y_t es la tendencia (no aleatoria) lineal.

Considere el contraste de Dickey y Fuller que incluye únicamente un intercepto (contraste B). Es decir, se estima $\Delta y_t = c + \phi y_{t-1} + \text{error}_t$. Note que se llega al proceso generador de datos si $c = \delta$ y $\phi = 0$. Sin ahondar en detalles, que son simples pero algo engorrosos, la dominancia de la tendencia lineal implica que

$$T^{3/2}(\hat{\rho} - 1) = T^{3/2}\hat{\phi} \xrightarrow{d} N\left(0, 12\frac{\sigma^2}{\delta^2}\right),$$

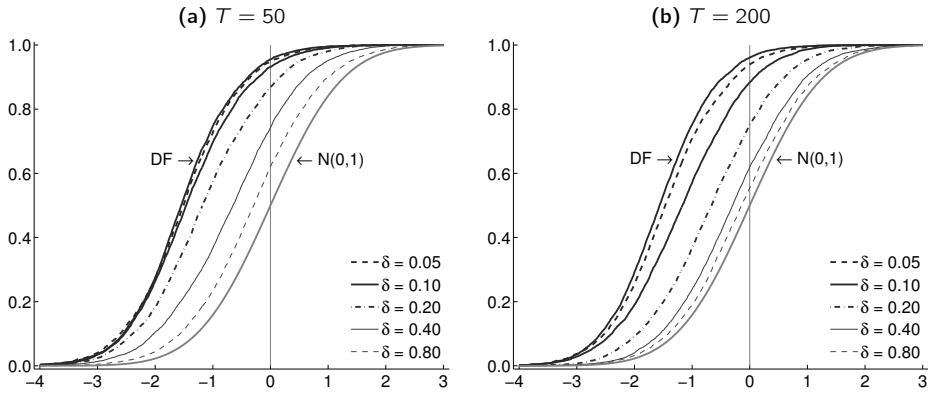
donde σ^2 es la varianza de ε_t . Note que la tasa que estabiliza la varianza del estimador es $T^{3/2}$, como en el caso estudiado en la sección III.1. Más importante aún, el estimador MCO ya no sigue una distribución de Dickey y Fuller sino que es **¡asintóticamente normal!** Este es un resultado notable que, entre otras cosas, conlleva que

$$\tau_{\hat{\phi}} \xrightarrow{d} N(0, 1).$$

Es bueno mencionar que si (i) $\delta = 0$ o (ii) $\delta \neq 0$ e incluimos una tendencia lineal en la regresión de contraste, entonces regresamos al esquema de Dickey y Fuller estudiado en secciones previas.

Este resultado implica, en la práctica, una pérdida de potencia. Recuerde que se ha estimado el contraste B de Dickey y Fuller, cuyo valor crítico es aproximadamente -2.90 , que es el que muy probablemente considere el investigador. El percentil correspondiente de una distribución normal estándar es el conocido -1.645 . Considere una muestra tal que el ratio t calculado es $\hat{\tau}_{\hat{\phi}} = -2$. Si consideráramos el valor crítico correcto, proveniente de una distribución normal estándar, rechazaríamos $H_0 : \rho = 1$. No obstante, si consideramos el valor crítico de DF_B , no podríamos rechazar (y sobradamente) $H_0 : \rho = 1$. Utilizar,

Gráfico III.8 Funciones de distribución empíricas de τ_ϕ para un proceso con tendencia



Notas: (Prog37_HaldrupHylleberg .prg) Distribuciones de probabilidad acumuladas del ratio τ_ϕ de contraste de raíz unitaria, para $R = 100,000$ repeticiones. El proceso generador de datos es un paseo aleatorio con tendencia, $\Delta y_t = \delta + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$.

equivocada pero inadvertidamente, el valor crítico DF_B podría implicar un incremento notable en la probabilidad de no rechazo y, con ello, una reducción de la potencia.

Una conclusión lógica de este análisis es que, en un mundo donde muchas series de interés presentan algún tipo de tendencia, las distribuciones DF podrían ser poco relevantes. Sin embargo, todo es una cuestión de grado. Resulta que la dominancia de la tendencia lineal, que es el motor detrás de la normalidad asintótica de τ_ϕ , es bastante menos obvia en muestras finitas. Así, a pesar del resultado normal asintótico, en la práctica las distribuciones DF siguen siendo de utilidad.

Para ilustrar este punto, el programa Prog37_HaldrupHylleberg.prg implementa un estudio de Monte Carlo similar al considerado en Haldrup y Hylleberg (1995). El ejercicio es simple. En la repetición $!i$, se genera un paseo aleatorio con tendencia, $GENR D(y) = !delta + NRND$, se estima la regresión de Dickey y Fuller que incluye únicamente un intercepto, $EQUATION eq.LS D(y) y(-1) C$, y se almacena el ratio τ_ϕ , $\tau\{!id\}_{!T}\{!i\} = eq.@TSTATS(1)$. El experimento consiste en variar el tamaño muestral $!T$ y el valor del *drift* $!delta$ que se asocia con el contador $!id$. Por ello, los resultados se almacenan en series que dependen de estas dos dimensiones.

El gráfico III.8 muestra las funciones de distribución empíricas de τ_ϕ para varios casos. En el panel (a) el tamaño muestral es de $T = 50$, mientras que en el panel (b) este es $T = 200$. Dentro de cada panel, se consideran valores de δ de 0.05, 0.10, 0.20, 0.40 y 0.80, además de mostrar las distribuciones DF_B (que corresponde a $\delta = 0$) y normal estándar. Se aprecia que, en la práctica, la distribución de τ_ϕ es un intermedio entre DF_B y la distribución normal. Para un valor dado de δ , se observa que incrementar el tamaño de la

muestra acerca a τ_ϕ a la distribución normal, que es el resultado asintótico. No obstante, en muchas ocasiones la distribución de τ_ϕ es más cercana a DF_B que a la distribución normal. Esto es particularmente cierto para valores pequeños de δ , digamos $\delta \leq 0.30$.

Es bueno mencionar que este rango de valores es sumamente relevante en la práctica. En estricto, lo que altera la distribución de τ_ϕ es δ/σ , es decir el valor de la *drift* relativo a la desviación estándar de Δy_t . En el gráfico III.8 se utiliza la normalización $\sigma = 1$ en toda instancia. Muchas series macroeconómicas, sobre todo de frecuencias trimestral o mensual, cumplen que promedio(Δy_t)/desviación estándar(Δy_t) ≤ 0.30 , por lo que los valores críticos de las distribuciones DF son relevantes en aplicaciones.

III.9 Proyecciones y raíces unitarias

¿Es importante saber si un proceso es estacionario en diferencias o estacionario en tendencia? En general, sí. Como se ha discutido, conocer si los datos corresponden a un proceso integrado puede ser una guía útil para un análisis más adecuado de sus autocovarianzas, y para estar alertas frente a ciertas anomalías que puedan aparecer en problemas de estimación. Sin embargo, donde las diferencias entre estas dos formas de modelar una serie de tiempo son más elocuentes es en la proyección (es decir, predicción fuera de la muestra), sobre todo de mediano y largo plazo. Considere un modelo AR(1) con tendencia

$$y_t = \alpha + \delta t + u_t \quad \text{donde} \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

y $\varepsilon_t \sim iid(0, \sigma^2)$. Note que $y_{t+h} = \alpha + \delta(t+h) + u_{t+h}$, así que la proyección óptima de y_{t+h} con información hasta el momento t es

$$\mathbb{E}(y_{t+h} | y_t) = \alpha + \delta(t+h) + \mathbb{E}(u_{t+h} | u_t).$$

En general, para $\rho \leq 1$, $\mathbb{E}(u_{t+h} | u_t) = \rho^h u_t$, por lo que

$$\mathbb{E}(y_{t+h} | y_t) = (1 - \rho^h) [\alpha + \delta(t+h)] + \rho^h (y_t + \delta h) \quad \text{y} \quad \mathbb{V}(y_{t+h} | y_t) = \sigma^2 \left(\frac{1 - \rho^{2h}}{1 - \rho^2} \right).$$

Se aprecia que cuando el proceso es estacionario en tendencia, $\rho < 1$, la predicción óptima converge a la tendencia lineal $\alpha + \delta(t+h)$ conforme h se incrementa. En otras palabras, la influencia de la observación en t se diluye conforme aumenta el horizonte de proyección, por lo que u_t es un **componente transitorio**. Asimismo, el error cuadrático medio de esta predicción es tal que $\mathbb{V}(y_{t+h} | y_t) \rightarrow \sigma^2/(1 - \rho^2)$ conforme $h \rightarrow \infty$. La incertidumbre del pronóstico de largo plazo es acotada.

El caso de raíz unitaria se obtiene como el límite $\rho \rightarrow 1$, dando como resultado

$$\mathbb{E}(y_{t+h} | y_t) = y_t + \delta h \quad \text{y} \quad \mathbb{V}(y_{t+h} | y_t) = \sigma^2 h.$$

Se espera que un paseo aleatorio con tendencia se incremente a una tasa constante δ por período partiendo del valor final y_t , sea cual sea este. A diferencia del caso estacionario en tendencia, esta proyección no se estabiliza alrededor de la tendencia, incrementándose la incertidumbre en torno a esta. Ello ocurre porque por construcción el paseo aleatorio es igual a la acumulación de choques, sin que estos sean descontados. Dado que el valor de y_t tiene un efecto permanente sobre todos los horizontes de proyección, u_t es una **tendencia estocástica** en este caso. La incertidumbre del pronóstico de largo plazo no tiene cota, $V(y_{t+h} | y_t) \rightarrow \infty$ conforme $h \rightarrow \infty$.

En la práctica, para horizontes de predicción finitos (y no tan grandes), los resultados obtenidos para un modelo DS pueden ser muy similares a los obtenidos con un modelo TS con ρ cercano a uno. Puede incluso darse el caso de que los pronósticos de un modelo DS sean mejores que los de un modelo TS, a pesar de que el modelo TS sea verdadero. Ello ocurre porque el modelo DS impone el valor de $\rho = 1$, en lugar de estimarlo. Así, si $\rho \simeq 1$, puede darse el caso de que el modelo DS, aunque falso, presente menor variabilidad muestral (por imponer una restricción “aproximadamente” correcta) y, por consiguiente, pronósticos con mejores propiedades. Curiosamente, estamos describiendo una situación en donde los contrastes de raíz unitaria tienen baja potencia y tenderían a favorecer al modelo DS.

Siguiendo esta motivación, Diebold y Kilian (2000) investigan hasta qué punto las pruebas de raíz unitaria pueden ser una herramienta útil para la predicción. El programa Prog38_DieboldKilian.prg implementa un estudio de Monte Carlo muy similar al propuesto por estos autores. El diseño consiste en generar y_t como en el modelo TS para distintos valores de ρ . Luego se siguen tres estrategias de predicción de valores futuros de y_t . Primero, se estima el modelo DS y se obtiene su predicción. La segunda estrategia consiste en estimar y utilizar la predicción del modelo TS. Finalmente, en la tercera estrategia se ejecuta una prueba de Dickey y Fuller con tendencia; dependiendo del resultado de esta prueba, se opta por la predicción del modelo TS (ante un rechazo) o del modelo DS (ante un no rechazo). Esta es la predicción PT, por las siglas del inglés *pre-testing*.

Varios detalles del programa Prog38_DieboldKilian.prg no han sido repetidos con mucha frecuencia en programas anteriores, por lo que conviene detenerse en ellos. Tenga en cuenta que antes de iniciar las repeticiones se define un vector Tvals con !nTvals valores de distintos tamaños muestrales T , cada uno de ellos asociado con un contador !it. Se considera $Tvals = (50, 100)'$, por lo que !nTvals = 2. Asimismo, se define un vector Hvals con !nHvals valores de distintos horizontes de proyección h , cada uno de ellos asociado con un contador !ih. Se considera $Hvals = (1, 10, 20)'$ por lo que !nHvals = 3. Finalmente, se define un vector rhovec de !nrho elementos con posibles valores de ρ , asociados con un contador !irho. Se considera $rhovec = (0.50, 0.75, 0.90, 0.95, 1.00)'$, por lo que !nrho = 5.

Los resultados de los ejercicios de proyección se almacenan en las series error2DS{!T}_{!h}, error2TS{!T}_{!h} y error2PT{!T}_{!h}, que contendrán los cuadrados de los errores de predicción de cada estrategia (DS, TS y PT), para un tamaño

muestral de T y un horizonte de proyección de h . El promedio de estas series a lo largo de un gran número de R repeticiones da una aproximación del error de proyección cuadrático medio, $\mathbb{E}((y_{t+h} - \hat{y}_{t+h})^2)$, que es la métrica de desempeño elegida para evaluar la bondad predictiva de cada estrategia.

Considerando que se descartan T_0 observaciones iniciales, los datos se generan según

```
SMPL 2 !T0+!T+!Hmax
GENR u = !rho*u(-1) + NRND
SERIES y = !alpha + !delta*time + u
```

para valores predeterminados de α y δ , y dado un valor cambiante ρ . La serie `time`, como es costumbre, es una tendencia lineal. Luego, dentro de un bucle donde el tamaño muestral es determinado según $T = Tvals(!it)$, se tiene que:

```
SMPL !T0 !T0+!T
EQUATION TS.LS y C time y(-1)
EQUATION DS.LS D(y) C
EQUATION DF.LS D(y) y(-1) time C
!TSmodel = 0
IF DF.@TSTATS(1) < (-3.4126 - 4.039/!T - 17.83/!T^2) THEN
    !TSmodel = 1
ENDIF
SMPL !T0+!T+1 !T0+!T+!Hmax
TS.FORECAST yTS_hat
DS.FORECAST yDS_hat
FOR !ih = 1 TO !nHvals
    !h = Hvals(!ih)
    !ytrue = y(!T0+!T+!h)
    !yfor = yTS_hat(!T0+!T+!h)
    error2TS{!T}_{!h}(!i) = (!yfor - !ytrue)^2
    !yfor = yDS_hat(!T0+!T+!h)
    error2DS{!T}_{!h}(!i) = (!yfor - !ytrue)^2
    error2PT{!T}_{!h}(!i) = error2DS{!T}_{!h}(!i)
    IF !TSmodel = 1 THEN
        error2PT{!T}_{!h}(!i) = error2TS{!T}_{!h}(!i)
    ENDIF
NEXT !ih
```

Primero, el análisis dentro de la muestra. Se restringe la muestra de estimación con `SMPL !T0 !T0+!T`, que contiene T observaciones, y se estiman tres ecuaciones: la ecuación

TS con el modelo TS, la ecuación DS con el modelo DS y la ecuación DF con la regresión de contraste de Dickey y Fuller que incluye una tendencia lineal. Luego, se procede al contraste de raíz unitaria. Si el valor del ratio τ_ϕ estimado en la ecuación DF resulta ser menor que el valor crítico \hat{a} la MacKinnon (1996), se rechaza la hipótesis de raíz unitaria, un evento que se marca asignando el valor de 1 al escalar `!TSmodel`, que por defecto toma el valor de 0.

Segundo, se calculan las proyecciones. Ahora, la muestra consiste en las `!Hmax` observaciones siguientes al término de la muestra de estimación, lo que se designa con `SMPL !T0+!T+1 !T0+!T+!Hmax`. Bajo esa muestra se generan, con el comando `FORECAST`, las proyecciones de los modelos TS y DS que quedan almacenadas, respectivamente, en las series `yTS_hat` e `yDS_hat`.

Tercero, se almacenan los errores de proyección. Para un horizonte variable `!h`, se tiene que el valor verdadero de y_{t+h} se encuentra en `!ytrue = y(!T0+!T+!h)`, mientras que los valores proyectados ocupan la misma ubicación en las series `yTS_hat` e `yDS_hat`, por ejemplo `!yfor = yDS_hat(!T0+!T+!h)`. Acto seguido, se almacena el error de proyección al cuadrado, por ejemplo `error2DS{!T}_{!h}(!i) = (!yfor - !ytrue)^2`. Finalmente, se determina el error de proyección de la estrategia PT. Si `!TSmodel = 1`, quiere decir que la raíz unitaria fue previamente rechazada y, por tanto, la estrategia manda utilizar la proyección TS. Para esa repetición, `error2PT{!T}_{!h}(!i) = error2TS{!T}_{!h}(!i)`. Esta instrucción sobrescribe `error2PT{!T}_{!h}(!i) = error2DS{!T}_{!h}(!i)`, que es la asignación por defecto y que se mantiene si es que `!TSmodel = 0` y, por tanto, la estrategia PT indica utilizar el modelo DS.

El cuadro III.3 reporta los resultados para $R = 100,000$ repeticiones. Además de la probabilidad de rechazo del contraste de Dickey y Fuller (comparable con lo visto en el panel (c) del gráfico III.6), se muestran los errores de proyección cuadráticos medios de cada estrategia. Más importante aún, las últimas columnas del cuadro muestran los errores cuadráticos **relativos**. Cuando estos números son superiores a 1, la estrategia del numerador es superior, mientras que cuando son menores de 1, la estrategia del denominador es dominante.

Analicemos primero las estrategias DS y TS, en particular los resultados de la columna DS / TS. Como se esperaba, ninguna de estas estrategias es dominante para todos los escenarios planteados. No es sorprendente verificar que cuando $\rho = 1$, la estrategia DS es uniformemente dominante, para todo h y T , ya que en este caso el modelo DS coincide con el proceso generador de datos. Ello ocurre, además, por el sesgo de estimación de ρ en el modelo TS, cuyos efectos se ven magnificados conforme h se incrementa. Este sesgo es importante en casos donde $\rho < 1$ pero $\rho \simeq 1$. Por ejemplo, para $\rho = 0.95$, a pesar de que el modelo verdadero corresponde a TS, DS sigue dominando. Las diferencias se ven reducidas con un mayor valor de T . Cuando $T = 50$, el efecto del sesgo en la estimación del modelo TS se aprecia incluso para $\rho = 0.90$, hecho que ya no se verifica para $T = 100$. Es interesante notar que, para los tamaños de muestra relativamente pequeños considerados, la razón de errores cuadráticos medios es **decreciente** en h para valores grandes de ρ . La

CAPÍTULO III. NO ESTACIONARIEDAD

Cuadro III.3 Desempeño predictivo de tres estrategias de proyección

T	ρ	Pr(TS)	h	DS	TS	PT	DS / TS	DS / PT	PT / TS
50	0.50	0.998	1	1.366	1.109	1.151	1.231	1.186	1.038
			10	3.342	1.844	2.078	1.813	1.608	1.127
			20	4.141	2.116	2.432	1.957	1.703	1.150
	0.75	0.319	1	1.158	1.109	1.168	1.044	0.992	1.053
			10	5.327	3.931	5.094	1.355	1.046	1.296
			20	7.014	5.072	6.680	1.383	1.050	1.317
	0.90	0.086	1	1.082	1.142	1.097	0.947	0.986	0.961
			10	8.641	9.287	8.708	0.930	0.992	0.938
			20	14.478	16.510	14.606	0.877	0.991	0.885
	0.95	0.059	1	1.051	1.134	1.064	0.927	0.989	0.938
			10	10.647	12.768	10.745	0.834	0.991	0.842
			20	21.604	28.062	21.768	0.770	0.992	0.776
	1.00	0.047	1	1.021	1.110	1.030	0.920	0.992	0.927
			10	12.032	15.334	12.129	0.785	0.992	0.791
			20	28.014	39.063	28.211	0.717	0.993	0.722
100	0.50	1.000	1	1.335	1.043	1.043	1.281	1.281	1.000
			10	2.956	1.516	1.516	1.951	1.951	1.000
			20	3.274	1.600	1.600	2.047	2.047	1.000
	0.75	0.876	1	1.141	1.043	1.067	1.094	1.069	1.023
			10	4.768	2.961	3.389	1.610	1.407	1.144
			20	5.607	3.261	3.803	1.720	1.475	1.166
	0.90	0.200	1	1.055	1.054	1.067	1.001	0.989	1.012
			10	7.598	6.777	7.605	1.121	0.999	1.122
			20	11.474	9.570	11.349	1.199	1.011	1.186
	0.95	0.086	1	1.042	1.072	1.050	0.972	0.992	0.980
			10	9.269	9.875	9.360	0.939	0.990	0.948
			20	16.807	17.776	16.899	0.945	0.995	0.951
	1.00	0.049	1	0.999	1.046	1.006	0.955	0.994	0.961
			10	11.041	13.294	11.158	0.831	0.990	0.839
			20	23.840	30.131	24.038	0.791	0.992	0.798

Notas: (Prog38_DieboldKilian.prg) Errores cuadráticos medios de proyección para $R = 100,000$ repeticiones. El proceso generador de datos es $y_t = \alpha + \delta t + u_t$, donde $u_t = \rho u_{t-1} + \varepsilon_t$ y $\varepsilon_t \sim N(0, 1)$. La estrategia PT consiste en el elegir la proyección del modelo DS o el modelo TS después de efectuar un contraste de raíz unitaria.

baja calidad del estimador TS se ve exacerbada en horizontes de proyección muy largos, introduciendo distorsiones incluso mayores que las implicadas en el uso de un modelo falso como DS.

Respecto a los resultados de la columna DS / PT, se aprecia que esta razón es muy cercana a uno para valores de $\rho \geq 0.90$, dando un resultado esencialmente equivalente para ambas estrategias. Por supuesto, ello ocurre porque para esos valores de ρ , la prueba de Dickey y Fuller tiene baja potencia y muy infrecuentemente rechaza al modelo DS. Por el contrario, cuando $\rho = 0.5$, la estrategia PT es superior a la DS porque la prueba de Dickey

y Fuller tiene alta potencia y distingue bien entre el verdadero modelo TS y el modelo DS, mal especificado. En el caso intermedio $\rho = 0.75$, la estrategia PT es superior a DS cuando $T = 100$ y la potencia es alta, no así cuando $T = 50$ y la potencia se ve alterada por el error muestral de estimación.

Finalmente, los resultados de la columna PT / TS indican que la estrategia PT es dominante cuando $\rho \geq 0.95$, ya que brinda resultados similares a la estrategia DS. Cuando $T = 50$, la zona de dominancia se extiende incluso a $\rho = 0.90$. La estrategia TS domina, no obstante, para valores más reducidos de ρ , lo que es reflejo de la baja potencia de la prueba de Dickey y Fuller con muestras tan reducidas. Cuando T se incrementa a $T = 100$, la dominancia de TS ocurre únicamente para $\rho = 0.75$. Para $\rho = 0.50$, la prueba es lo suficientemente potente como para identificar a TS como el modelo dominante.

En líneas generales, se tiene que para situaciones en donde la prueba de Dickey y Fuller tiene baja potencia, la estrategia PT favorece al modelo DS. En estas circunstancias, la estrategia DS es dominante. Por su parte, cuando la prueba de Dickey y Fuller es muy potente, la estrategia PT es similar a la dominante TS. Existe un intervalo de transición que es notorio cuando $T = 50$, en donde el modelo TS es el mejor predictor y, a pesar de ello, la estrategia PT favorece al modelo DS. Este intervalo se reduce sustancialmente para $T = 100$.

III.10 Parámetros fastidiosos y posibles soluciones

El contraste básico de Dickey y Fuller, $H_0 : y_t$ es un paseo aleatorio *versus* $H_1 : y_t$ es un proceso AR(1) estacionario con media cero, se basa en supuestos fuertes. Su valor es más pedagógico que práctico o realista. Es por este motivo que en secciones previas se ensayaron otras versiones del contraste, orientadas a formular modelos empíricamente más relevantes, por lo menos bajo H_1 . En todos los casos, no obstante, bajo H_0 la serie en cuestión se comporta como un paseo aleatorio (con o sin tendencia). Como se ha visto, el paseo aleatorio es un caso sumamente estilizado de procesos $I(1)$ más generales, que acumulan choques posiblemente autocorrelacionados.

Así, es de interés formular contrastes de raíz unitaria más amplios, en donde la presencia de una raíz unitaria se asocia con un proceso $I(1)$ más general $H_0 : y_t \sim I(1)$, mientras que la hipótesis alternativa también contemple casos más generales, $H_1 : y_t \sim I(0)$, no necesariamente un proceso autorregresivo. Al incorporar tendencias en las regresiones de contraste, se tendrían contrastes más generales aún, basados en el par $H_0 : y_t$ es DS *versus* $H_1 : y_t$ es TS. Para tales propósitos, primero debe analizarse cómo la presencia de choques autocorrelacionados altera el marco de inferencia *à la* Dickey y Fuller, para luego buscar maneras adecuadas de generalización.

Suponga que $y_t = y_{t-1} + u_t$, donde $u_t \sim I(0)$ es un proceso estacionario arbitrario, por lo que y_t es un proceso integrado general. En esta sección, llamaremos γ_0 a la varianza no

condicional de u_t , $\gamma_0 = \mathbb{V}(\Delta y_t)$, y denotamos como λ a la varianza de largo plazo de u_t ,

$$\lambda = \lim_{T \rightarrow \infty} T \cdot \mathbb{V} \left(\frac{1}{T} \sum_{t=1}^T \Delta y_t \right).$$

Dado y_t , suponga que estima ρ en la regresión $y_t = \rho y_{t-1} + \text{error}_t$. El resultado concreto es que

$$T(\hat{\rho} - 1) \xrightarrow{d} \frac{\mathcal{D}_3}{\mathcal{D}_2} + \left(1 - \frac{\gamma_0}{\lambda}\right) (\text{Variable aleatoria}).$$

Es decir, en el límite $\hat{\rho}$ se comporta como una variable aleatoria cuya distribución depende de los parámetros que gobiernan los patrones de correlación de Δy_t . El interés, recuerde, es inferir sobre ρ , o de modo más general, sobre si y_t es un proceso estacionario o no. Parámetros como γ_0 o λ son de segunda importancia para estos fines y, sin embargo, aparecen en las distribuciones asintóticas relevantes. Por ello, estos son denominados **parámetros fastidiosos** (*nuisance parameters*), ya que complican sobremanera las posibilidades de inferencia. Note que si $\Delta y_t = \varepsilon_t$, es decir, si y_t fuera un paseo aleatorio (y no un proceso $I(1)$ más general), se daría que $\gamma_0 = \lambda$ y el segundo término de la ecuación líneas arriba sería igual a cero, tal y como se vio en la sección III.5. En este caso, la distribución límite de $T(\hat{\rho} - 1)$ es $\mathcal{D}_3/\mathcal{D}_2$, que es la distribución simulada en el panel (a) del gráfico III.5. Aunque esta función no tiene una forma analítica, tampoco contiene parámetros fastidiosos.

El programa Prog39_Nuisance.prg explora estas extensiones. Partiendo de un ruido blanco normalmente distribuido, SERIES e = NRND, se generan tres series de tiempo según

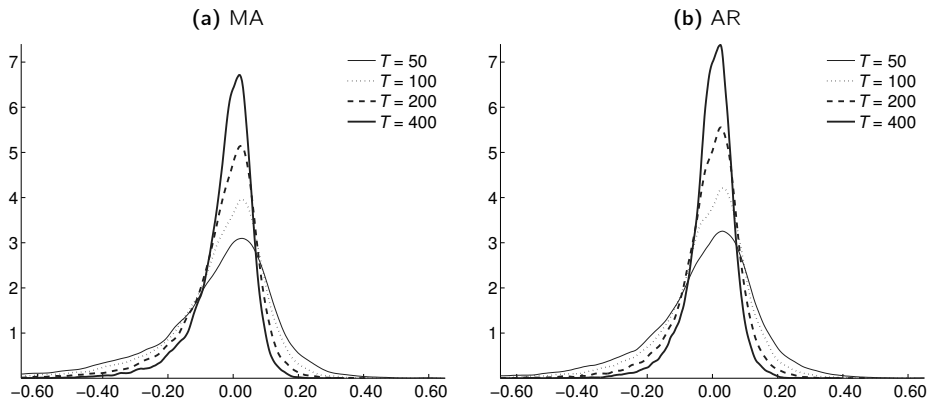
```
GENR D(yRW) = e
GENR D(yAR) - !zeta*D(yAR(-1)) = e
GENR D(yMA) = e + !theta*e(-1),
```

La serie yW es un paseo aleatorio; la serie yAR es un proceso integrado tal que sus primeras diferencias siguen un proceso AR(1) de media cero y parámetro de persistencia ζ , almacenado en el escalar !zeta; finalmente, la serie yMA es un proceso integrado tal que sus primeras diferencias siguen un proceso MA(1) de media cero y parámetro de persistencia θ , almacenado en el escalar !theta.

Luego, se procede a estimar la regresión de contraste básica $\Delta y_t = \phi y_{t-1} + \text{error}_t$, y se almacenan sus resultados:

```
FOR %a RW AR MA
    EQUATION eq.LS D(y{%a}) y{%a}(-1)
    tDF{%a}_{!T}(!i) = eq.@TSTATS(1)
    rho{%a}_{!T}(!i) = @SQRT(!T)*eq.C(1)
NEXT %a.
```

Gráfico III.9 Superconsistencia de ρ en la regresión básica de Dickey y Fuller



Notas: (Prog39_Nuisance.prg) Densidades kernel (automáticas) para $R = 100,000$ repeticiones. El proceso generador de datos es $\Delta y_t = u_t$ con $u_t = \varepsilon_t + \theta \varepsilon_{t-1}$ y $\theta = 0.95$ en el panel (a) y $u_t = \zeta u_{t-1} + \varepsilon_t$ y $\zeta = 0.95$ en el panel (b). En ambos casos, $\varepsilon_t \sim N(0, 1)$.

En la repetición i y con un tamaño muestral igual a T , para el proceso y_t la regresión se corre en el objeto `eq`, y se almacena el ratio τ_ϕ de contraste de $H_0 : \phi = \rho - 1 = 0$ en la serie `tDF{a}_{!T}` y el estadístico $\sqrt{T}(\hat{\rho} - 1)$ en la serie `rho{a}_{!T}`.

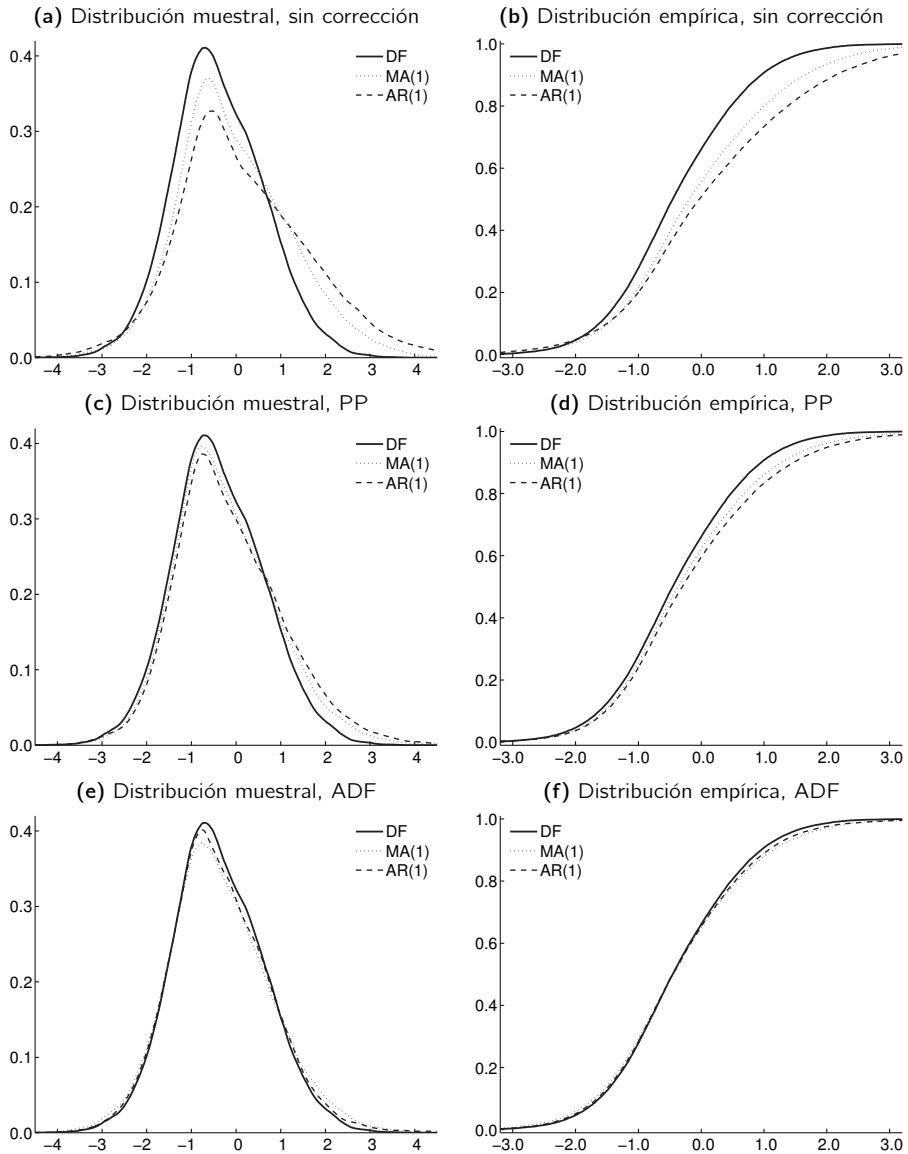
Es oportuno notar que, dado que $T(\hat{\rho} - 1)$ converge en distribución a una variable aleatoria, ocurre nuevamente que $\sqrt{T}(\hat{\rho} - 1)$ converge en probabilidad a cero. Es decir, $\hat{\rho}$ es un estimador **superconsistente**. Estos resultados se muestran en el gráfico III.9, que presenta las distribuciones muestrales de `rhoMA_{!T}` y `rhoAR_{!T}` para diversos tamaños muestrales. La convergencia hacia cero es bastante evidente, a pesar de que el error muestral $(\hat{\rho} - 1)$ es multiplicado por una función de T divergente.

Este hallazgo es sumamente interesante y contrasta con lo que ocurre bajo estacionariedad. Recuerde que se está estimando el modelo dinámico $\Delta y_t = \phi y_{t-1} + \text{error}_t$, donde el término de error presenta correlación serial. Es decir, se trata de un modelo dinámico mal especificado. Estos errores de especificación son muy perjudiciales en un mundo estacionario. Si $y_t \sim I(0)$, el estimador MCO de ϕ (o de ρ) sería inconsistente. En un mundo no estacionario, no obstante, el estimador MCO de ϕ (o de ρ) mantiene la propiedad de superconsistencia. La raíz unitaria contenida en y_{t-1} domina la varianza de la regresión, más allá del aporte de las autocorrelaciones en el término de error.

No hay lonche gratis. La mala especificación dinámica de la regresión de contraste contamina las distribuciones asintóticas de interés con parámetros fastidiosos. En el caso del ratio t vinculado con $H_0 : \phi = 0$, es posible establecer que

$$\tau_\phi \xrightarrow{d} DF_A + \left(1 - \frac{\gamma_0}{\lambda}\right) (\text{Variable aleatoria}),$$

Gráfico III.10 *Parámetros fastidiosos y posibles soluciones*



Notas: (Prog39_Nuisance.prg) Los paneles (a), (c) y (e) presentan densidades kernel (automáticas) del ratio τ_ϕ de contraste de raíz unitaria, mientras que los paneles (b), (d) y (f) muestran las respectivas funciones de probabilidad acumuladas. Véanse las notas al gráfico III.9.

donde DF_A es la distribución de Dickey y Fuller. Los paneles (a) y (b) del gráfico III.10 muestran, respectivamente, las densidades y funciones de distribución empíricas de estos ratios (almacenados en las series $\tau_{DF\{a\}}_{50}$), cuando el proceso generador es y_{RW} (esta es la distribución DF que no contiene parámetros fastidiosos), cuando es y_{MA} y cuando es y_{AR} . Las diferencias son bastante claras y resulta evidente que los parámetros fastidiosos distorsionan los resultados. Se sabe que el efecto de los parámetros fastidiosos sobre la distribución de τ_ϕ será más pronunciado cuanto más lejano se encuentre γ_0/λ de 1. Las simulaciones reportadas en el gráfico III.10 utilizan los valores $\theta = 0.95$ y $\zeta = 0.50$. Con ello, la razón de varianzas es

$$(MA) \frac{\gamma_0}{\lambda} = \frac{1 + \theta^2}{(1 + \theta)^2} \simeq 0.5 \quad \text{y} \quad (AR) \frac{\gamma_0}{\lambda} = \frac{(1 - \zeta)^2}{1 - \zeta^2} \simeq 0.3,$$

lo que explica que el caso AR simulado muestre mayores distorsiones (la distribución es más lejana de DF) que el caso MA.

De haber reparametrizado el ejercicio de simulación de manera distinta, por ejemplo con valores diferentes de ζ o θ , la razón γ_0/λ habría sido distinta, y habríamos obtenido como consecuencia distribuciones del ratio τ_ϕ distintas a las mostradas. Es elocuente, luego, lo fastidioso que resulta γ_0/λ para la inferencia.

¿Cómo lidiar con la presencia de parámetros fastidiosos? Dos son los enfoques dominantes en la literatura. Ambos son propuestas flexibles que tienen la virtud de que el contraste de raíz unitaria puede realizarse admitiendo patrones de autocorrelación muy generales en u_t , pero sin necesidad de modelar explícitamente estos patrones. Es decir, en ambos casos no es necesario determinar el mejor modelo ARMA que describa a u_t , sino tan solo aproximar sus efectos sobre la prueba de hipótesis.

El primer enfoque es el de **Phillips y Perron (1989)** y consiste en incorporar una **corrección no paramétrica** (es decir, que no dependa explícitamente de un modelo ARMA en particular) al ratio τ_ϕ obtenido de la regresión de contraste mal especificada. En concreto, se trata de agregarle una corrección a este ratio, tal que asintóticamente desaparezca el término que contiene parámetros fastidiosos. Esquemáticamente,

$$\tau_{PP} = \tau_\phi + \text{Corrección} \xrightarrow{d} DF_A.$$

La corrección de Phillips y Perron requiere estimar el ratio γ_0/λ . Para ello, se almacenan los residuos de la ecuación de contraste e_t , que son predictores de u_t bajo $H_0 : \phi = 0$. El numerador γ_0 es estimado con la varianza muestral de e_t , mientras que el denominador λ corresponde a un estimador HAC de la varianza de largo plazo de e_t , como aquellos discutidos en la sección II.3.

EViews implementa esta prueba como un VIEW del objeto SERIES. En particular, el comando `y.URROOT(PP,NONE)` produce un cuadro con los resultados de la prueba de Phillips y Perron como el presentado en el gráfico III.11. Se presenta, en primer

Gráfico III.11 Prueba de Phillips y Perron

	A	B	C	D	E
1	Null Hypothesis: Y has a unit root				
2	Exogenous: None				
3	Bandwidth: 6 (Newey-West automatic) using Bartlett kernel				
4					
5				Adj. t-Stat	Prob.*
6	Phillips-Perron test statistic				
7				-1.324734	0.1704
8	Test critical values: 1% level				
9				-2.588292	
10	5% level				
11				-1.944072	
12	10% level				
13				-1.614616	
14	*MacKinnon (1996) one-sided p-values.				
15	Residual variance (no correction)				
16				1.279167	
17	HAC corrected variance (Bartlett kernel)				
18				3.959733	
19	Phillips-Perron Test Equation				
20	Dependent Variable: D(Y)				
21	Method: Least Squares				
22	...				
23	Sample: 51 150				
24	Included observations: 100				
25					
26	Variable	Coefficient	Std. Error	t-Statistic	Prob.
27					
28	Y(-1)	-0.012911	0.012235	-1.055256	0.2939
29					
30					

Nota: Aspecto del cuadro producido por EViews 8 tras el comando `y.URROOT(PP,NONE)`.

lugar, una línea que informa al usuario que H_0 es la presencia de raíz unitaria. Luego, se brinda información sobre los regresores determinísticos utilizados en la regresión de contraste (Exogenous). En el ejemplo mostrado se excluye este tipo de regresores (Exogenous: None). La instrucción `y.URROOT(PP,CONST)` implementa el contraste que incluye un intercepto (Exogenous: Constant), mientras que `y.URROOT(PP,TREND)` incluye, además, una tendencia lineal (Exogenous: Constat, Linear Trend). La especificación por defecto, `y.URROOT(PP)`, incluye un intercepto.

Seguidamente, se provee información del número de rezagos utilizados en la estimación de la varianza de largo plazo HAC a la Newey y West. Esta es una selección automática e informada que EViews hace por defecto. El usuario puede imponer este valor con la opción BAND, `y.URROOT(PP,CONST,BAND=q)`. Se muestran, además, los estimados de γ_0 (reportado como Residual variance) y de λ (reportado como HAC corrected variance).

Gráfico III.12 Prueba de Dickey y Fuller Aumentada

	A	B	C	D	E	
1	Null Hypothesis: Y has a unit root					
2	Exogenous: None					
3	Lag Length: 1 (Automatic - based on SIC, maxlag=12)					
4						
5					t-Statistic	Prob.*
6						
7	Augmented Dickey-Fuller test statistic			-1.429995	0.1416	
8	Test critical values: 1% level		-2.588292			
9	5% level		-1.944072			
10	10% level		-1.614616			
11						
12	*MacKinnon (1996) one-sided p-values.					
13						
14	Augmented Dickey-Fuller Test Equation					
15	Dependent Variable: D(Y)					
16	Method: Least Squares					
17	...					
18	Sample: 51 150					
19	Included observations: 100					
20						
21	Variable	Coefficient	Std. Error	t-Statistic	Prob.	
22						
23	Y(-1)	-0.013155	0.009199	-1.429995	0.1559	
24	D(Y(-1))	0.662934	0.075487	8.782129	0.0000	
25						
26						

Nota: Aspecto del cuadro producido por EViews 8 tras el comando `y.URROOT(ADF,NONE)`.

Luego se presentan los estadísticos involucrados en la prueba de hipótesis. En particular, el número bajo el nombre Phillips-Perron test statistic es el ratio t ajustado por la posible presencia de residuos serialmente correlacionados. Bajo la hipótesis nula, la distribución de este estadístico es la de Dickey y Fuller, y por ello se hace referencia a sus valores críticos calculados a la MacKinnon (1996). En el caso mostrado en el gráfico III.11, el ratio t calculado es de -1.32 y es mayor que el menos conservador de los valores críticos (-1.61 , al 10% de significación), por lo que no puede rechazarse H_0 . Alternativamente, EViews reporta valores p , y se alcanza la misma conclusión al notar que el valor p asociado con el valor de -1.32 del estadístico t es de $0.17 > 0.10$.

Para terminar, EViews reporta el detalle de la regresión de contraste. Como se comentó, esta es la más simple de todas las posibilidades, en donde Δy_t se regresa sobre y_{t-1} únicamente. Por supuesto, las regresiones de contraste bajo los comandos `y.URROOT(PP,CONST)` o `y.URROOT(PP,TREND)` mostrarán los regresores determinísticos adicionales. Note que el valor de τ_ϕ obtenido en el gráfico III.11 es de -1.05 y es, por supuesto, distinto al ratio t ajustado con el que finalmente se implementa el contraste.

El segundo enfoque, propuesto por Said y Dickey (1984), es conocido como la prueba de **Dickey y Fuller Aumentada** y consiste en incorporar una **corrección paramétrica**, pero no al ratio τ_ϕ sino a la regresión de contraste. En lugar de estimar $\Delta y_t = \phi y_{t-1} + \text{error}_t$, se estima la ecuación **aumentada**

$$\Delta y_t = \phi y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \text{error}_t.$$

Incluir rezagos de la variable dependiente en una regresión dinámica es la solución habitual a la presencia de autocorrelación en los residuos. Said y Dickey (1984) muestran que si p es elegido de una manera apropiada, el residuo de la regresión de contraste aumentada tendrá las características de un ruido blanco. Luego, se calcula el ratio t vinculado con $H_0 : \phi = 0$ y se restablece el resultado

$$\tau_\phi \xrightarrow{d} DF_A.$$

Es bueno notar que la regresión aumentada bajo $H_0 : \phi = 0$ indica que $\Delta y_t \sim \text{AR}(p)$. Esta es, en términos generales, una aproximación al verdadero modelo de Δy_t . No obstante, esta aproximación autorregresiva, que es en espíritu similar a la estudiada en la sección II.6, podría ser arbitrariamente buena, por lo menos para el propósito de “blanquear” el error de la regresión de contraste. El resultado más interesante de Said y Dickey (1984) es que si p es elegido de tal manera que $p/T^{1/3} \rightarrow 0$ conforme $T \rightarrow \infty$ (una idea muy similar a la discutida en la estimación de varianzas HAC, véase la sección II.3), el **aumentar la regresión de contraste** con rezagos de Δy_t corrige adecuadamente las distorsiones traídas por la autocorrelación serial de Δy_t , **sin importar cuál sea el verdadero modelo de Δy_t .**

EViews implementa esta prueba también como un VIEW del objeto SERIES. El comando `y.ROOT(ADF,NONE)` produce un cuadro como el presentado en el gráfico III.12. Se presenta una línea que informa al usuario que H_0 es la presencia de raíz unitaria y se brinda información sobre los regresores determinísticos utilizados en la regresión de contraste (`Exogenous`), opción que puede modificarse con los comandos `y.ROOT(ADF,CONST)` o `y.ROOT(ADF,TREND)`. La especificación por defecto, `y.ROOT(ADF)`, incluye un intercepto. Por defecto, EViews elige p tras minimizar el criterio de información de Schwarz en una búsqueda que cubre los casos $p = 0, 1, \dots, p_{\max}$. El criterio de información puede ser modificado por el usuario con la opción `INFO`, por ejemplo `y.ROOT(ADF,NONE,INFO=AIC)` utiliza el criterio de Akaike, mientras que el máximo número de rezagos sobre el que se realiza la búsqueda se calibra con la opción `MAXLAG`, `y.ROOT(ADF,NONE,MAXLAG=p_{\max})`. Finalmente, `y.ROOT(ADF,NONE,LAG=p)` implementa el contraste con exactamente p rezagos, sin realizar búsqueda alguna.

A continuación, se muestran los estadísticos involucrados en la prueba de hipótesis. El número bajo el nombre *Augmented Dickey-Fuller test statistic* es el ratio t de la regresión aumentada. Bajo la hipótesis nula, la distribución de este estadístico es la de Dickey y Fuller, y por ello se hace referencia a los valores críticos de esta distribución (que son exactamente los mismos que aparecen en el contraste de Phillips y Perron, véase el gráfico III.11). En el caso mostrado en el gráfico III.12, el ratio t calculado es de -1.43 y

APUNTES DE ESTUDIO

es mayor que -1.61 , el valor crítico al 10 % de significación, por lo que no puede rechazarse H_0 . Alternativamente, el valor p asociado con este ratio t calculado es $0.14 > 0.10$.

Finalmente, EViews reporta el detalle de la regresión aumentada, en donde se regresa Δy_t sobre y_{t-1} y los rezagos de Δy_t . En este caso, $p^* = 1$. Las regresiones de contraste bajo los comandos `y.URROOT(ADF,CONST)` o `y.URROOT(ADF,TREND)` mostrarán los regresores determinísticos adicionales. Note que el valor de τ_ϕ obtenido en el gráfico III.12 es de -1.43 , idéntico al ratio t con el que finalmente se implementa el contraste.

Además de los resultados ya mostrados, el programa `Prog39_Nuisance.prg` también almacena los estadísticos t corregidos asociados con estos procedimientos. El código relevante es:

```
FOR %a AR MA
    y{%a}.URROOT(PP,NONE,SAVE=_temp)
    tPP{%a}_{!T}(!i) = _temp(3)
    D _temp
    y{%a}.URROOT(ADF,NONE,SAVE=_temp)
    tADF{%a}_{!T}(!i) = _temp(3)
    D _temp
NEXT %a.
```

Nótese que en el comando `URROOT` se ha utilizado la opción `SAVE`. Esta opción guarda en un vector (al que llamamos `_temp`) los principales escalares asociados con la aplicación del comando. En particular, el ratio τ_ϕ es el tercer elemento de este vector que es, para la repetición `!i` y con un tamaño muestral de `!T`, almacenado en las series `tPP{%a}_{!T}` y `tADF{%a}_{!T}`, respectivamente, para las pruebas de Phillips y Perron y de Dickey y Fuller Aumentada. El comando `D _temp` borra al vector `_temp` del *workfile*, ya que el objeto en donde el comando `URROOT` almacena los resultados con la opción `SAVE` debe corresponder a un objeto inexistente en el *workfile*.

Los paneles (c) y (d) del gráfico III.10 muestran las distribuciones del estadístico corregido con el método de Phillips y Perron para un tamaño muestral dado ($T = 100$), y bajo los dos procesos de generación de datos considerados en `Prog39_Nuisance.prg` (MA y AR). Asimismo, estos paneles incluyen la distribución DF para fines comparativos. Los paneles (e) y (f) hacen lo propio para el contraste de Dickey y Fuller Aumentado. Se aprecia que, en efecto, ambos enfoques cumplen con lo prometido y amortiguan considerablemente las distorsiones observadas en los paneles (a) y (b), como consecuencia de un error de regresión autocorrelacionado.

Dado que ambos enfoques atacan al mismo problema y tienen como objeto restablecer las condiciones de inferencia que el investigador enfrentaría considerando un paseo aleatorio, pero en un contexto de series integradas más generales, surge inevitablemente la siguiente pregunta: ¿cuál es el enfoque preferido en la práctica? Los resultados del gráfico III.10 sugieren que el enfoque de Dickey y Fuller Aumentado es más eficiente al remover las

distorsiones producidas por la autocorrelación residual. No obstante, este es un resultado muy particular que precisa ser generalizado.

Schwert (1989) investiga el tema con profundidad, y el programa Prog310_Schwert.prg replica parte de sus simulaciones de Monte Carlo. En esta ocasión, se considera el proceso generador de datos

$$y_t - y_{t-1} = \varepsilon_t + \theta\varepsilon_{t-1},$$

que es el caso MA(1) del programa Prog39_Nuisance.prg, pero para distintos valores de θ . Ante este proceso generador de datos, y para valores de T cambiantes, se realizan varios contrastes de raíz unitaria y se registra, para cada caso, la proporción de rechazos. Los contrastes son realizados considerando valores críticos al 5% de significación y, por tanto, cuanto más alejada se encuentre la probabilidad de rechazo de este 5%, más distorsionado y menos útil será el contraste en cuestión. Es decir, la métrica en este estudio es la probabilidad real de cometer Error Tipo I (rechazar una hipótesis verdadera).

Para un tamaño de muestra T , Schwert (1989) considera dos maneras de determinar los rezagos involucrados tanto en el estimador HAC utilizado por Phillips y Perron, como en la regresión de Dickey y Fuller aumentada. Estas son reglas "al ojo" (*rules of thumb*) muy utilizadas en la literatura: una conservadora, $!lag1 = @FLOOR(4*(T/100)^{(1/4)})$, y otra más laxa, $!lag2 = @FLOOR(12*(T/100)^{(1/4)})$. En general, $!lag2 \simeq 3*!lag1$.

Considere que el valor crítico relevante de la distribución DF ha sido almacenado en el escalar $!tcrit$. Asimismo, considere una matriz de nombre $R\{T\}$ inicialmente llena de ceros que registrará la proporción de rechazos de cada método. Esta matriz tiene tantas filas como valores de θ considerados en el proceso generador de datos, los que varían con el contador $!itheta$. La primera columna de $R\{T\}$ contiene el valor de θ , mientras que el resto contiene los resultados de los contrastes. Se consideran las siguientes siete alternativas:

DF La prueba de Dickey y Fuller básica, sin ningún tipo de corrección por la presencia de residuos autocorrelacionados.

```
y.UROOT(ADF,LAG=0,SAVE=_temp0)
IF _temp0(3) < !tcrit THEN
    R{!T}(!itheta, 2) = R{!T}(!itheta, 2) + 1/!R
ENDIF
```

PP₁ La prueba de Phillips y Perron, utilizando el método conservador de determinación de rezagos.

```
y.UROOT(PP,BAND=!lag1,SAVE=_temp1)
IF _temp1(3) < !tcrit THEN
    R{!T}(!itheta, 3) = R{!T}(!itheta, 3) + 1/!R
ENDIF
```

APUNTES DE ESTUDIO

PP₂ La prueba de Phillips y Perron, utilizando el método laxo de determinación de rezagos.

```
y.URROOT(PP,BAND=!lag2,SAVE=_temp2)
IF _temp2(3) < !tcrit THEN
    R{!T}(!itheta, 4) = R{!T}(!itheta, 4) + 1/!R
ENDIF
```

PP* La prueba de Phillips y Perron, utilizando un método de determinación automático, el que EViews utiliza como opción por defecto.

```
y.URROOT(PP,SAVE=_temp3)
IF _temp3(3) < !tcrit THEN
    R{!T}(!itheta, 5) = R{!T}(!itheta, 5) + 1/!R
ENDIF
```

ADF₁ La prueba de Dickey y Fuller Aumentada, utilizando el método conservador de determinación de rezagos.

```
y.URROOT(ADF,LAG=!lag1,SAVE=_temp4)
IF _temp4(3) < !tcrit THEN
    R{!T}(!itheta, 6) = R{!T}(!itheta, 6) + 1/!R
ENDIF
```

ADF₂ La prueba de Dickey y Fuller Aumentada, utilizando el método laxo de determinación de rezagos.

```
y.URROOT(ADF,LAG=!lag2,SAVE=_temp5)
IF _temp5(3) < !tcrit THEN
    R{!T}(!itheta, 7) = R{!T}(!itheta, 7) + 1/!R
ENDIF
```

ADF* La prueba de Dickey y Fuller Aumentada, utilizando un método de determinación automático, el que EViews utiliza como opción por defecto.

```
y.URROOT(ADF,SAVE=_temp6)
IF _temp6(3) < !tcrit THEN
    R{!T}(!itheta, 8) = R{!T}(!itheta, 8) + 1/!R
ENDIF
```

Se utilizan $R = 25,000$ replicaciones, en lugar de las $R = 100,000$ habituales, ya que cada repetición es computacionalmente exigente por la multitud de pruebas involucradas. Los resultados para los valores de $\theta = \{-0.8, -0.5, 0.0, 0.5, 0.8\}$ y tamaños muestrales de $T = \{25, 50, 100, 250, 500\}$ se reportan en el cuadro III.4. Note, en primer lugar, que cuando

CAPÍTULO III. NO ESTACIONARIEDAD

Cuadro III.4 Probabilidad de rechazo real (probabilidad nominal de 5%)

T	τ_{crit}	θ	DF	ARMA	PP ₁	PP ₂	PP*	ADF ₁	ADF ₂	ADF*
25	-2.985	-0.8	0.923	0.303	0.919	0.930	0.918	0.242	0.033	0.841
		-0.5	0.412	0.150	0.404	0.431	0.404	0.065	0.031	0.354
		0.0	0.053	0.064	0.055	0.061	0.058	0.047	0.039	0.074
		0.5	0.028	0.046	0.025	0.040	0.029	0.038	0.047	0.098
		0.8	0.029	0.069	0.025	0.039	0.029	0.033	0.043	0.116
50	-2.920	-0.8	0.990	0.193	0.983	0.994	0.985	0.296	0.040	0.789
		-0.5	0.520	0.056	0.473	0.562	0.479	0.063	0.033	0.320
		0.0	0.049	0.020	0.054	0.051	0.054	0.046	0.038	0.055
		0.5	0.025	0.034	0.025	0.027	0.026	0.055	0.040	0.085
		0.8	0.025	0.052	0.022	0.025	0.024	0.070	0.036	0.086
100	-2.890	-0.8	0.998	0.085	0.988	0.997	0.990	0.288	0.045	0.646
		-0.5	0.571	0.013	0.458	0.572	0.475	0.052	0.037	0.187
		0.0	0.050	0.025	0.055	0.053	0.054	0.046	0.041	0.052
		0.5	0.024	0.038	0.029	0.023	0.027	0.044	0.042	0.071
		0.8	0.026	0.047	0.029	0.024	0.028	0.036	0.041	0.069
250	-2.873	-0.8	0.998	0.011	0.980	0.993	0.984	0.267	0.050	0.440
		-0.5	0.607	0.018	0.389	0.501	0.411	0.053	0.045	0.131
		0.0	0.051	0.039	0.056	0.059	0.056	0.050	0.048	0.052
		0.5	0.024	0.046	0.039	0.034	0.037	0.053	0.046	0.056
		0.8	0.023	0.048	0.034	0.030	0.032	0.063	0.047	0.057
500	-2.868	-0.8	0.999	0.011	0.962	0.985	0.972	0.302	0.053	0.314
		-0.5	0.614	0.032	0.318	0.415	0.350	0.054	0.048	0.101
		0.0	0.049	0.042	0.052	0.057	0.052	0.048	0.047	0.050
		0.5	0.023	0.044	0.037	0.036	0.036	0.049	0.045	0.047
		0.8	0.025	0.048	0.039	0.038	0.039	0.065	0.050	0.055

Notas: (Prog310_Schwert.prg) Probabilidad real de rechazo de $H_0 : y_t \sim I(1)$, para valores críticos al 5% de significación y con $R = 25,000$ replicaciones. El proceso generador de datos es $\Delta y_t = \varepsilon_t + \theta \varepsilon_{t-1}$, $\varepsilon_t \sim N(0, 1)$.

$\theta = 0$, el proceso generador de datos de y_t corresponde a un paseo aleatorio. Dado que se utilizan valores críticos provenientes de la distribución DF al 5% de significación, se tiene que la tasa de rechazo del contraste DF es, precisamente, 5 por ciento. Curiosamente, la tasa de rechazo del resto de contrastes es similar, aunque con pequeñas diferencias producidas por los errores muestrales inducidos por las estimaciones adicionales detrás de estas propuestas. En segundo lugar, y como ocurre con todo resultado asintótico, el desempeño de todos los contrastes generalmente mejora cuando T se incrementa.

Los resultados más llamativos del cuadro III.4 se relacionan con las fuertes distorsiones asociadas con los casos donde $\theta < 0$. Recuerde que el parámetro fastidioso γ_0/λ aparece en las distribuciones de los ratios t sin correcciones, y esta razón asciende a 4.25 cuando $\theta = -0.5$ y a ¡32.24! cuando $\theta = -0.8$, ambos valores muy alejados de 1. Para los valores de $\theta > 0$ considerados, se tiene que $\gamma_0/\lambda \simeq 0.4$ implica una distorsión relativamente menor, que es atendida satisfactoriamente por todos los contrastes considerados.

Llama en especial la atención la incapacidad de las pruebas de Phillips y Perron de proveer una corrección adecuada para la autocorrelación generada por $\theta < 0$. La tasa de rechazo se mantiene en el orden de 30 % a 40 % cuando $\theta = -0.5$ y de 90 % cuando $\theta = -0.8$, incluso con muestras tan grandes como $T = 500$. Schwert (1989) anota que el fuerte sesgo en la estimación MCO de la ecuación de contraste podría hacer que utilizar los residuos para implementar la corrección no sea tan buena idea. Debería considerarse, sobre todo cuando $\theta < -0.5$, el residuo bajo H_0 , es decir simplemente Δy_t (una idea parecida a la de Schmidt y Phillips, sección III.7). Si bien este refinamiento produce mejoras en el desempeño de las pruebas de Phillips y Perron, estas no son suficientes en comparación con otras alternativas.

Se observa que las pruebas de Dickey y Fuller Aumentadas tienen un mejor desempeño que las pruebas de Phillips y Perron, que es la conclusión alcanzada por Schwert (1989). Cuando $\theta = -0.8$, no obstante, estas pruebas muestran algunas dificultades, dependiendo de cómo se determina el número de rezagos en la regresión aumentada. Como se discutió en la sección II.6, si $\theta \neq 0$, el proceso MA(1) que sigue Δy_t equivale a un proceso AR(∞), y si el valor de θ es cercano a -1 , entonces su mejor aproximación AR(p), con p finito, necesariamente tendrá un valor elevado de p . Es por ello que la versión más parametrizada del contraste (ADF₂) es sistemáticamente superior cuando $\theta = -0.8$. Note que la versión por defecto de EViews (ADF_S) utiliza el criterio de información de Schwarz en la elección del número de rezagos. Posiblemente, y a juzgar por los resultados obtenidos en la sección II.6 (en particular, el cuadro II.2), podríamos encontrar un mejor desempeño utilizando criterios de información alternativos.

Para concluir, ¿el caso $\theta < 0$ es relevante en la práctica? Para responder, note que la primera correlación de Δy_t es $\rho_1 = \theta/(1 + \theta^2) < 0$, igual a $\rho_1 \simeq -0.40$ cuando $\theta = -0.5$ y aproximadamente $\rho_1 \simeq -0.48$ cuando $\theta = -0.8$. Considere ahora una serie de tiempo formada como $x_t = z_t + \kappa_t$, donde z_t es un paseo aleatorio $\Delta z_t = \varepsilon_t$ con $\mathbb{V}(\varepsilon_t) = \sigma^2$, y κ_t es un ruido blanco con varianza $\mathbb{V}(\kappa_t) = \varphi\sigma^2$, independiente de ε_t . Es decir, x_t es un paseo aleatorio pero observado con un error de medición. El coeficiente $\varphi \geq 0$ indica cuán importante es este error de medición al determinar x_t y, aunque no hay un consenso, es común hablar de valores del orden de $\varphi \simeq 40$ en macroeconomía.

Note que $\Delta x_t = \varepsilon_t + \Delta \kappa_t$. Se tiene que, $\mathbb{V}(\Delta x_t) = \sigma^2(1 + 2\varphi)$, $\mathbb{C}(\Delta x_t, \Delta x_{t-1}) = -\sigma^2\varphi$ y $\mathbb{C}(\Delta x_t, \Delta x_{t-s}) = 0$ para $s > 0$. Es decir, Δx_t , así como Δy_t , se comporta como un MA(1). La primera correlación de este MA(1) es $\rho_1 = -\varphi/(1 + 2\varphi)$, que siempre es negativa para $\varphi > 0$. Es decir, si observamos una serie no estacionaria (en este caso un paseo aleatorio) aunque sea mínimamente contaminada por un error de medición tipo κ_t , entonces Δx_t se comportará como un MA(1) con coeficiente $\theta < 0$. Más aún, un valor de $\varphi = 2$, que es ciertamente reducido, corresponde a una primera correlación consistente con el caso $\theta = -0.5$, mientras que un valor de $\varphi = 20$, muy frecuente en la práctica, reproduce los patrones de un proceso MA(1) con el caso $\theta = -0.8$. En resumen, la respuesta a la pregunta de si el caso $\theta < 0$ es relevante, es un rotundo **sí**.

III.11 Estimación e inferencia con series no estacionarias

Para concluir este capítulo, ilustraremos resultados sumamente importantes sobre la estimación con series no estacionarias a través de un último ejercicio de simulación.

Suponga que el proceso generador de datos de y_t es

$$\Delta y_t = \zeta \Delta y_{t-1} + \varepsilon_t,$$

donde $\varepsilon_t \sim iid(0, \sigma^2)$. Es decir, y_t es un proceso $I(1)$ cuyas primeras diferencias son gobernadas por un proceso AR(1). Considere la regresión

$$\Delta y_t = c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \text{error}_t. \quad (\text{III.1})$$

Esta es una regresión que involucra exclusivamente series estacionarias, como las estudiadas en la sección II.4. Se trata de un modelo sobreparametrizado, ya que $c_2 = 0$ y $c_1 = \zeta$ en el modelo poblacional. No obstante, la estimación por MCO de esta regresión proveerá estimadores consistentes y asintóticamente normales (resultados que se ilustran, por ejemplo, en el gráfico II.4). Así,

$$\sqrt{T}(\hat{c}_1 - \zeta) \xrightarrow{d} N(0, v_1) \quad \text{y} \quad \sqrt{T}\hat{c}_2 \xrightarrow{d} N(0, v_2),$$

donde \hat{c}_i es el estimador MCO de c_i en (III.1) y v_i es su varianza asintótica (cuya expresión no es importante). A partir de este resultado, las pruebas de hipótesis que involucren a los coeficientes c_i pueden realizarse con el instrumental clásico de inferencia.

Considere ahora un modelo de regresión alternativo:

$$\Delta y_t = \phi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \text{error}_t. \quad (\text{III.2})$$

El lector reconocerá que esta es una regresión de contraste *à la* Dickey y Fuller Aumentada. Se aprecia del proceso generador de datos que $\phi = 0$. Cabe preguntarse, ¿cuáles son las propiedades de los estimadores MCO de ϕ , c_1 y c_2 en (III.2)? Ocurre lo siguiente:

- $\hat{\phi}$ es **superconsistente**, $T^n \hat{\phi} \xrightarrow{p} 0$ para cualquier $0 \leq n < 1$. Este es el mismo comportamiento que el estudiado en torno a las regresiones tipo Dickey y Fuller. De hecho, el ratio τ_ϕ converge a la distribución DF_A .
- \hat{c}_1 y \hat{c}_2 son **asintóticamente equivalentes** a \tilde{c}_1 y \tilde{c}_2 , respectivamente, que son los estimadores MCO en la ecuación (III.1).
- Por tanto, \hat{c}_1 y \hat{c}_2 son consistentes y **asintóticamente normales**:

$$\sqrt{T}(\hat{c}_1 - \zeta) \xrightarrow{d} N(0, v_1) \quad \text{y} \quad \sqrt{T}\hat{c}_2 \xrightarrow{d} N(0, v_2),$$

que son las mismas propiedades asintóticas de \tilde{c}_1 y \tilde{c}_2 .

APUNTES DE ESTUDIO

Antes de analizar las implicancias de estos resultados, describimos brevemente el ejercicio de simulación que los ilustra, implementado en el programa Prog311_ADF.prg. Para la repetición i , se generan los datos según

$$\text{GENR } D(y) - \zeta * D(y(-1)) = @RTDIST(3) / @SQRT(3) ,$$

donde el ruido blanco ε_t tiene media cero, varianza igual a uno y colas anchas. El valor de ζ se almacena en el escalar ζ . Luego, dado un tamaño muestral cambiante T :

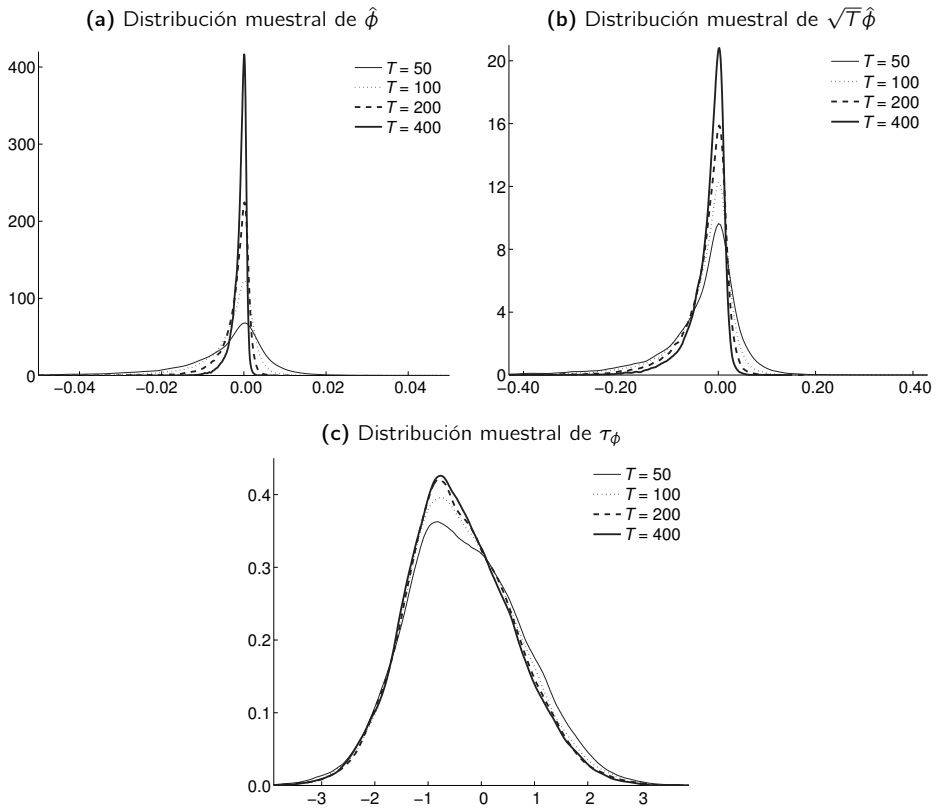
```
EQUATION eq1.LS D(y) D(y(-1)) D(y(-2))
EQUATION eq2.LS D(y) D(y(-1)) D(y(-2)) y(-1)
Dc1_{!T}(!i) = @SQR(!T)*(eq1.C(1) - eq2.C(1))
Dc2_{!T}(!i) = @SQR(!T)*(eq1.C(2) - eq2.C(2))
!phi = eq2.C(3)
D0phi_{!T}(!i) = !phi
D1phi_{!T}(!i) = @SQR(!T)*!phi
D2phi_{!T}(!i) = eq2.@TSTATS(3) .
```

Se estiman dos ecuaciones. La primera, eq1, es la regresión (III.1) de Δy_t sobre Δy_{t-1} y Δy_{t-2} , mientras que la segunda, eq2, es la regresión de Dickey y Fuller Aumentada (III.2). Las diferencias de los estimados de ambas regresiones y para ambos coeficientes se almacenan en las series $Dc1_{!T}$ y $Dc2_{!T}$. Las distribuciones de estas dos series a lo largo de las repeticiones de nuestro experimento nos permitirán verificar la equivalencia asintótica de ambos grupos de estimadores. Por otro lado, la serie $D0phi_{!T}$ almacena el estimado de ϕ en la regresión de Dickey y Fuller Aumentada y la serie $D1phi_{!T}$ almacena el estimado de ϕ multiplicado por la raíz de T . Las distribuciones de estas series permiten verificar la superconsistencia de $\hat{\phi}$. Finalmente, la serie $D3phi_{!T}$ almacena τ_ϕ con el propósito de evaluar si es que efectivamente emerge la distribución DF_A .

El gráfico III.13 presenta un primer grupo de resultados, asociado con la superconsistencia de $\hat{\phi}$. En los paneles (a) y (b) se verifica que, en efecto, $T^n \hat{\phi}$ converge (en probabilidad) a cero para valores de n tales que $0 \leq n < 1$. Por su parte, el panel (c) muestra el comportamiento asintótico del ratio τ_ϕ , que converge en distribución a una variable aleatoria. Aunque posiblemente no es fácil identificar visualmente cuál es la distribución límite (con $T = 400$), al estudiar sus percentiles se concluye que se trata, en efecto, de la distribución de Dickey y Fuller.

Por su parte, el gráfico III.14 presenta los resultados asociados con la equivalencia asintótica de los estimadores de c_i en ambas regresiones. Se aprecia de forma muy clara que $\sqrt{T}(\hat{c}_i - \tilde{c}_i)$ va colapsando hacia cero conforme T va incrementándose. En el límite, convergerá en probabilidad a cero, confirmando la equivalencia asintótica.

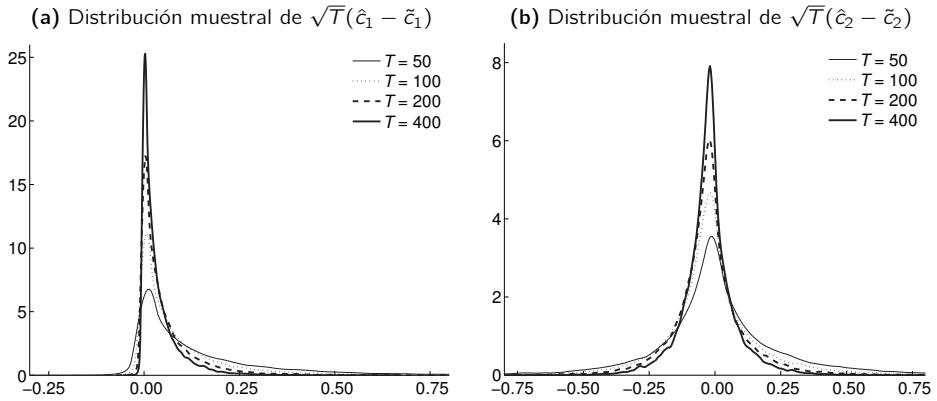
Gráfico III.13 Superconsistencia en la regresión de Dickey y Fuller Aumentada



Notas: (Prog311_ADF.prg) Distribuciones kernel (automáticas) con $R = 100,000$ repeticiones. El proceso generador de datos es el modelo AR(1) en diferencias, $\Delta y_t = \zeta \Delta y_{t-1} + \varepsilon_t$, donde ε_t es un ruido blanco de colas anchas y varianza igual a 1, y $\zeta = 0.5$. La escala de los ejes difiere entre paneles.

Note que la ecuación de las diferencias de y_t (III.2) es la ecuación de Dickey y Fuller Aumentada (III.1) **tras imponer el verdadero valor $\phi = 0$** , y recuerde que los estimadores de c_i en ambos casos tienen **exactamente la misma** distribución asintótica. Claramente, la fuente de variabilidad muestral adicional en la segunda ecuación por estimar ϕ se disipa conforme T se incrementa y, a la larga, desaparece. La fuerza detrás de este fenómeno es la superconsistencia de $\hat{\phi}$. Si $\hat{\phi}$ no fuera superconsistente, el error de estimación de $\hat{\phi}$ se desvanecería a la tasa habitual, y esta incertidumbre afectaría la distribución asintótica de \hat{c}_i , alejándola de la distribución asintótica del estimador \tilde{c}_i . Desde la perspectiva de la inferencia de los coeficientes c_i , para muestras lo suficientemente grandes, da lo mismo conocer el verdadero valor de ϕ que reemplazarlo por un estimador superconsistente.

Gráfico III.14 Equivalencia asintótica en la regresión de Dickey y Fuller Aumentada



Notas: (Prog311_ADF.prg) Distribuciones kernel (automáticas) con $R = 100,000$ repeticiones. Véanse las notas al gráfico III.13. La escala de los ejes difiere entre paneles.

El resultado formal detrás de este fenómeno es el siguiente: además de ser asintóticamente normales, los estimadores de c_1 y c_2 , debido a las distintas tasas de convergencia, son **asintóticamente independientes** del estimador de ϕ . Este fenómeno implica que el marco de inferencia clásico, basado en aproximaciones normales, es completamente válido para el constraite de muchas hipótesis con series no estacionarias. La única hipótesis que escapa de este marco es $H_0 : \phi = 0$.

Ilustraremos esta última proposición con un ejemplo. Considere la misma serie y_t y suponga que el interés se centra en estimar la siguiente ecuación **en niveles**,

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \text{error}_t,$$

y realizar pruebas de hipótesis sobre los coeficientes a_i . Es simple verificar que esta es tan solo una reparametrización de la ecuación (III.2). Se cumple que

$$a_1 = 1 + \phi + c_1, \quad a_2 = c_2 - c_1 \quad \text{y} \quad a_3 = -c_2.$$

Dada la linealidad de estas relaciones, los estimadores MCO de los coeficientes a_i serán idénticos a los estimadores MCO de ϕ y de c_i , cuyas propiedades conocemos: $\hat{a}_1 = 1 + \hat{\phi} + \hat{c}_1$, $\hat{a}_2 = \hat{c}_2 - \hat{c}_1$ y $\hat{a}_3 = -\hat{c}_2$. Así,

$$\begin{aligned} \sqrt{T}(\hat{a}_3 - a_3) &= -\sqrt{T}(\hat{c}_2 - c_2) \xrightarrow{d} N(0, v_2), \\ \sqrt{T}(\hat{a}_2 - a_2) &= \sqrt{T}(\hat{c}_2 - c_2) - \sqrt{T}(\hat{c}_1 - c_1) \xrightarrow{d} N(0, v_2 + v_1 - 2v_{12}), \\ \sqrt{T}(\hat{a}_1 - a_1) &= \sqrt{T}\hat{\phi} + \sqrt{T}(\hat{c}_1 - c_1) \xrightarrow{d} 0 + N(0, v_1) \equiv N(0, v_1). \end{aligned}$$

El escalar v_{12} es la covarianza asintótica entre \hat{c}_1 y \hat{c}_2 , cuya forma no es relevante para nuestros propósitos. Note lo que ocurre con la distribución asintótica de a_1 , donde se ha utilizado el hecho de que $\sqrt{T}\hat{\phi}$ converge en probabilidad a cero. Dada la superconsistencia de $\hat{\phi}$, la incertidumbre de estimar a_1 está completamente dominada por la incertidumbre de estimar c_1 , al menos asintóticamente.

Es notable que los estimadores de a_i sean asintóticamente normales, incluso cuando $y_t \sim I(1)$. La implicancia es, como se mencionó, que casi todas las pruebas de hipótesis sobre estos coeficientes se pueden realizar, por ejemplo, con estadísticos t que serán asintóticamente normales, o pruebas F que serán asintóticamente χ^2 . Una hipótesis interesante en el contexto del modelo en niveles es $H_0 : a_3 = 0$ para investigar si el modelo AR(3) es preferido sobre una especificación AR(2). Análogamente, se puede pensar en $H_0 : a_2 = a_3 = 0$ para averiguar si un modelo AR(1) es suficiente para modelar a $y_t \dots$ y así sucesivamente. La única hipótesis que necesitará atención especial, en el sentido de que las aproximaciones normales no serán válidas, es $H_0 : a_1 + a_2 + a_3 = 1$, ya que esto equivale a $H_0 : \phi = 0$.

El ejemplo desarrollado puede enmarcarse en un contexto de inferencia más general, en donde emerge un resultado muy importante de Sims, Stock y Watson (1990). Considere un modelo de regresión que contiene potencialmente una mezcla de variables $I(0)$ y variables $I(1)$. Después de simples manipulaciones, este modelo puede escribirse como

$$y_t = C + [\text{Variables } I(1)]'\Psi + [\text{Variables } I(0)]'\Upsilon + \epsilon_t,$$

donde ϵ_t es un **ruido blanco**. El resultado es:

- Los estimadores MCO de Ψ y Υ serán **asintóticamente independientes**.
- El estimador MCO del vector Ψ será **superconsistente** y, posiblemente, convergerá en distribución a una variable aleatoria "inusual". La teoría clásica de inferencia no siempre aplica a este grupo.
- El estimador MCO del vector Υ será **asintóticamente normal**. **La teoría clásica de inferencia aplica a este grupo de coeficientes**.

Finalizamos este capítulo con dos ejemplos de las "simples manipulaciones" para llegar a la forma de Sims, Stock y Watson (1990). Otros ejemplos aparecerán en el capítulo IV.

En el modelo AR(3) discutido anteriormente,

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \epsilon_t,$$

sume y reste $a_3 y_{t-2}$ para conseguir

$$y_t = a_1 y_{t-1} + (a_2 + a_3) y_{t-2} - a_3 \Delta y_{t-2} + \epsilon_t.$$

APUNTES DE ESTUDIO

Vemos que a_3 multiplica a la variable estacionaria Δy_{t-2} y, por tanto, pertenece al grupo Υ . Siguiendo el mismo argumento, se puede determinar que a_2 y a_1 pertenecen también, individualmente, al grupo Υ . Tomando la última ecuación, y restándole y sumándole $(a_2 + a_3)y_{t-1}$, se consigue

$$y_t = (a_1 + a_2 + a_3)y_{t-1} - (a_2 + a_3)\Delta y_{t-1} - a_3\Delta y_{t-2} + \epsilon_t.$$

Concluimos, entonces, que $a_2 + a_3$ pertenece al grupo Υ , pero que $a_1 + a_2 + a_3$ pertenece al grupo "superconsistente" Ψ . Esta es una conclusión que habíamos alcanzado anteriormente. Note que hemos llegado a la forma de una ecuación de Dickey y Fuller Aumentada, lo que nos permite concluir que pruebas de hipótesis en la selección de rezagos de Δy_t en esta ecuación se pueden efectuar de la manera tradicional, así como las pruebas de hipótesis que determinan la estructura de rezagos del modelo AR en niveles.

Por otro lado, considere el modelo de tendencia lineal

$$y_t = a_0 + a_1 t + a_2 y_{t-1} + \epsilon_t,$$

donde $|a_2| < 1$. Se tiene que $y_t - \mathbb{E}(y_t) \sim I(0)$, ya que y_t es estacionario en tendencia.

Se sabe que la esperanza de y_t tiene la forma $b_0 + b_1 t$, por lo que la esperanza de y_{t-1} es $b_0 + b_1(t-1)$ ⁵. Así, al restar y sumar $a_2 \mathbb{E}(y_{t-1})$ al lado derecho de la ecuación se consigue

$$\begin{aligned} y_t &= (a_0 + a_2 b_0 - a_2 b_1) + (a_1 + a_2 b_1)t + a_2(y_{t-1} - \mathbb{E}(y_{t-1})) + \epsilon_t \\ &\equiv C + \Psi t + \Upsilon(y_{t-1} - \mathbb{E}(y_{t-1})) + \epsilon_t. \end{aligned}$$

Se concluye que el estimador MCO de a_2 será asintóticamente normal. Su significación, por ejemplo $H_0 : a_2 = 0$, podría ser evaluada con una prueba t estándar. Podemos concluir, además, que no hay manera de aislar al coeficiente de a_1 de la tendencia t . Este coeficiente, por tanto, será del grupo Ψ .

Compare estas conclusiones con el análisis de la sección III.1, al inicio de este capítulo ¿Todos los caminos llevan a Roma?

⁵ Los valores de b_0 y b_1 pueden determinarse a partir del método de coeficientes indeterminados. Al tomar esperanzas $\mathbb{E}(y_t) = a_0 + a_1 t + a_2 \mathbb{E}(y_{t-1})$, reemplazar $b_0 + b_1 t = a_0 + a_1 t + a_2(b_0 - b_1 + b_1 t)$, agrupar $b_0 + b_1 t = a_0 + a_2(b_0 - b_1) + (a_1 + a_2 b_1)t$, e igualar coeficientes, se concluye que:

$$b_1 = \frac{a_1}{1 - a_2} \quad \text{y} \quad b_0 = \frac{a_1 a_2 - a_0(1 - a_2)}{(1 - a_2)^2}.$$

IV Cointegración

Este capítulo provee una introducción al importante concepto de **cointegración**, y sus implicancias para la estimación e inferencia de modelos dinámicos. Varios de los principales hallazgos del capítulo III, o resultados muy similares, aparecerán en este análisis. Una diferencia fundamental, no obstante, es que el análisis previo ha sido univariado, mientras que la idea de cointegración es necesariamente multivariada. Así pues, en el capítulo III la conexión entre un contexto $I(1)$ y uno $I(0)$ implicaba tomar diferencias a una serie integrada (el modelo DS) o remover una tendencia determinística a una serie estacionaria en tendencia (el modelo TS), todas ellas transformaciones univariadas de los datos. En cambio, bajo la óptica de cointegración, el paso de un mundo $I(1)$ a uno $I(0)$ implica la combinación lineal de varias variables, que es una operación multivariada.

Considere dos series integradas $y_t \sim I(1)$ y $x_t \sim I(1)$. Se dice que estas series cointegran si es que existe entre ellas una combinación lineal estacionaria $e_t = y_t - \gamma x_t \sim I(0)$. El vector $(1, -\gamma)'$ se conoce como **vector de cointegración**. Esta definición se extiende a casos más allá del bivariado. Se puede partir de un vector de dimensión $n \times 1$, $y_t \sim I(1)$ cuyos elementos cointegran si es que existe un vector β , también de dimensión $n \times 1$, tal que $e_t = \beta' y_t \sim I(0)$. No obstante, nos centraremos en el caso bivariado, ya que muchos de los resultados más interesantes de la teoría de cointegración emergen en este marco analítico simple.

IV.1 Representaciones

Considere el modelo dinámico

$$y_t = c + \phi y_{t-1} + \theta_0 x_t + \theta_1 x_{t-1} + \varepsilon_t, \quad (\text{IV.1})$$

donde ε_t es un ruido blanco y $|\phi| < 1$. Esta ecuación es un caso particular de un **modelo autorregresivo de rezagos distribuidos** (ARDL). Es posible incorporar rezagos de x_t y rezagos adicionales de y_t , pero las conclusiones principales del análisis no se verían afectadas. Igualmente, x_t y los coeficientes θ podrían ser vectores de dimensión $k \times 1$.

La ecuación (IV.1) establece una relación dinámica entre las variables y_t y x_t . Los coeficientes ϕ , θ_0 y θ_1 , por supuesto, contienen mucha información sobre los aspectos más relevantes de esta relación. Por ejemplo, al conocer estos coeficientes puede estudiarse cómo un cambio en x_t se transfiere a y_t , y cuánto tiempo tarda esta transferencia en manifestarse, entre otros.

Una de las cantidades más informativas sobre la relación entre y_t y x_t es el denominado **multiplicador de largo plazo** (MLP). Imagine una situación de estado estacionario, en donde $x_t = x$, $y_t = y$ y $\varepsilon_t = \mathbb{E}(\varepsilon_t) = 0$, para todo t . Al evaluar (IV.1) en este estado, se consigue

$$y = c + \phi y + \theta_0 x + \theta_1 x \quad \rightarrow \quad y = \frac{c}{1 - \phi} + \left(\frac{\theta_0 + \theta_1}{1 - \phi} \right) x.$$

El modelo dinámico colapsa a una ecuación estática. La pendiente entre y y x es el MLP,

$$\gamma = \frac{\theta_0 + \theta_1}{1 - \phi},$$

y mide el efecto que un cambio **permanente** en x_t tendrá sobre y_t **en el largo plazo**. Este es un análisis intencionalmente simple, por ser determinístico, y puramente motivacional. En un mundo estocástico, la idea de estado estacionario no es muy atractiva. No obstante, la noción del MLP sigue siendo muy relevante en casos más complejos.

Para elaborar, (IV.1) puede reescribirse como

$$y_t - \phi y_{t-1} = c + \frac{\theta_0 + \theta_1}{1 - \phi} (x_t - \phi x_{t-1}) - \frac{\theta_1 + \phi \theta_0}{1 - \phi} \Delta x_t + \varepsilon_t,$$

donde se ha sumado y restado $\gamma(x_t - \phi x_{t-1})$ a la ecuación original. De modo más compacto,

$$y_t - \phi y_{t-1} = \mu(1 - \phi) + \gamma(x_t - \phi x_{t-1}) + \vartheta \Delta x_t + \varepsilon_t,$$

donde los parámetros μ y ϑ se encuentran implícitamente definidos. Definamos una variable e_t tal que $e_t - \phi e_{t-1} = \vartheta \Delta x_t + \varepsilon_t$. Con ello, (IV.1) pasa a ser

$$y_t - \phi y_{t-1} = \mu(1 - \phi) + \gamma(x_t - \phi x_{t-1}) + e_t - \phi e_{t-1}.$$

Al resolver esta "cuasidiferencia" para y_t , se consigue

$$y_t = \mu + \gamma x_t + e_t. \tag{IV.2}$$

Esta ecuación, donde el MLP aparece explícitamente, es la versión estocástica de la relación de estado estacionario encontrada previamente. Se conoce como **ecuación de cointegración**, ya que nos permite explorar la noción de cointegración de una manera muy transparente.

Suponga que $x_t \sim I(1)$. Ello implica que, por definición, $\Delta x_t \sim I(0)$. Dado que $|\phi| < 1$ y que $\varepsilon_t \sim I(0)$, se concluye que $e_t \sim I(0)$. Por ejemplo, si x_t fuera un paseo aleatorio tal que $\Delta x_t =$ ruido blanco, se tendría que $e_t = \phi e_{t-1} +$ ruido blanco, que corresponde a un proceso AR(1) estacionario. En la práctica, e_t puede ser caracterizado con un proceso ARMA mucho más complejo que el proceso AR(1) de nuestro ejemplo, pero es bueno enfatizar que la conclusión relevante hasta el momento es que e_t **es estacionario** incluso cuando x_t no lo es.

Se concluye, luego, que $y_t \sim I(1)$, ya que es igual a la suma de un término $I(1)$, dado por γx_t , más un término $I(0)$, dado por $\mu + e_t$. Es claro que la no estacionariedad de y_t proviene de la relación dinámica que guarda con x_t . Más aún, se verifica que

$$e_t = y_t - \mu - \gamma x_t \sim I(0),$$

por lo que y_t y x_t **cointegran** con un vector de cointegración $(1, -\gamma)'$. Enfatizamos nuevamente que es **el MLP el parámetro relevante en la ecuación de cointegración**¹.

La ecuación de cointegración nos recuerda, inevitablemente, a un modelo de regresión, donde y_t es regresada sobre una constante y x_t . El término e_t hace las veces del error de esta regresión y, de hecho, así es como se le conoce: e_t es el “**error**”. Por ello, usualmente se interpreta a $\bar{y}_t = \mu + \gamma x_t$, que es la predicción sobre y_t hecha exclusivamente con la información en x_t , como el valor de largo plazo o “de equilibrio” de y_t , por lo que $e_t = y_t - \bar{y}_t$ resulta ser un desvío entre el valor efectivamente observado y el valor de largo plazo: un error o un “desequilibrio”. La cointegración indica que estos desvíos son necesariamente transitorios, ya que $e_t \sim I(0)$ con media cero.

El modelo dinámico (IV.1) puede ser representado de una manera alternativa a la ecuación de cointegración. Tome la ecuación (IV.1). Sume $-y_{t-1}$ a ambos lados. Luego, sume y reste $\theta_0 x_{t-1}$ al lado derecho. Con ello,

$$y_t - y_{t-1} = c + \phi y_{t-1} - y_{t-1} + \theta_0 x_{t-1} + \theta_0 x_t - \theta_0 x_{t-1} + \theta_1 x_{t-1} + \varepsilon_t$$

$$\Delta y_t = (\phi - 1) \left(y_{t-1} - \frac{c}{1 - \phi} - \frac{\theta_0 + \theta_1}{1 - \phi} x_{t-1} \right) - \theta_0 \Delta x_t + \varepsilon_t.$$

De manera más compacta,

$$\Delta y_t = \alpha (y_{t-1} - \mu - \gamma x_{t-1}) + \pi \Delta x_t + \varepsilon_t, \quad (\text{MCE})$$

donde $\alpha = -(1 - \phi)$, $c = (1 - \phi)\mu$, γ es el MLP y $\pi = \theta$.

Este es el **modelo de corrección de errores** (MCE) y es una ecuación que expresa a Δy_t (y sus rezagos en caso de que aparezcan más rezagos de y_t en el modelo ARDL original) en función de Δx_t (y sus rezagos en caso de que aparezcan rezagos de x_t en el modelo ARDL original) y de $e_{t-1} = y_{t-1} - \mu - \gamma x_{t-1}$. Una ventaja de esta representación es que describe cómo y_t responde y evoluciona para preservar la cointegración con x_t .

Sea \mathcal{I}_t el conjunto de información disponible hasta el período t , inclusive. Suponga que se parte de una situación en donde $\mathbb{E}(\Delta x_{t+1} | \mathcal{I}_t) = 0$; es decir, se espera que x_t no cambie

¹ Dado que $x_t \sim I(1)$, esta variable podría representarse (utilizando la descomposición de Beveridge y Nelson) como $x_t = z_t + v_{xt}$, donde z_t es un paseo aleatorio y v_{xt} es un término estacionario. Así, $y_t = \mu + \gamma z_t + v_{yt}$, donde $v_{yt} = e_t + \gamma v_{xt} \sim I(0)$. Ocurre que x_t e y_t **comparten una tendencia común**, el paseo aleatorio z_t , que es la causa de la no estacionariedad de ambas variables. La cointegración **elimina la tendencia común**: $e_t = y_t - \mu - \gamma x_t$ ya no contiene a z_t .

de valor. Además, considere que $y_t > \bar{y}_t = \mu + \gamma x_t$: y_t se encuentra por encima de su valor de largo plazo, lo que genera un “desequilibrio” positivo $e_t > 0$. Luego, la proyección de y_t hecha por (MCE) es

$$\mathbb{E}(\Delta y_{t+1} | \mathcal{I}_t) = \alpha e_t < 0,$$

ya que $\alpha = -(1 - \phi) < 0$, por ser $|\phi| < 1$. Se espera que el desequilibrio positivo genere una reducción en el valor de y_t , de modo que la brecha $e_t = y_t - \bar{y}_t$ tienda a cerrarse. Este es el **mecanismo de corrección de errores**: $\mathbb{E}(e_{t+h} | \mathcal{I}_t) \rightarrow 0$ conforme $h \rightarrow \infty$, un comportamiento típico de $e_t \sim I(0)$ con media cero.

No obstante, quizá la característica más conveniente de la representación (MCE) es que, a diferencia de las representaciones alternativas (IV.1) y (IV.2) en donde aparecen y_t y x_t , el modelo de corrección de errores **solo involucra variables estacionarias**: Δy_t (y sus rezagos), Δx_t (y sus rezagos) y e_t . Así, en un mundo donde algunas series son $I(1)$, el MCE es un vehículo de inferencia $I(0)$.

El análisis de esta sección aplica a casos más generales. Considere un modelo ARDL más completo,

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=0}^q \theta_i x_{t-i} + \varepsilon_t,$$

donde $\phi_1 + \phi_2 + \dots + \phi_q < 1$. En esta ocasión, la ecuación de cointegración tiene exactamente la misma forma que (IV.2), pero con μ y γ redefinidos como

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_q} \quad \text{y} \quad \gamma = \frac{\theta_0 + \theta_1 + \theta_2 + \dots + \theta_q}{1 - \phi_1 - \phi_2 - \dots - \phi_q},$$

y con $e_t \sim I(0)$ pero con una dinámica ARMA bastante más compleja que la considerada anteriormente. Por su parte, el modelo de corrección de errores adopta la forma

$$\Delta y_t = \alpha(y_{t-1} - \mu - \gamma x_{t-1}) + \sum_{i=0}^q \pi_i \Delta x_{t-i} + \sum_{i=1}^{p-1} \zeta_i \Delta y_{t-i} + \varepsilon_t,$$

donde $\alpha = -(1 - \phi_1 - \phi_2 - \dots - \phi_q) < 0$ y los coeficientes π_i y ζ_i dependen de los coeficientes originales ϕ_i y θ_i . Lo importante es siempre tener en cuenta que la representación ARDL, la ecuación de cointegración y el modelo de corrección de errores son reparametrizaciones de la misma relación dinámica entre y_t y x_t .

IV.2 Ausencia de cointegración y regresiones espurias

La sección anterior ha discutido maneras alternativas de representar la dinámica de dos variables cointegradas. No obstante, no todas las variables integradas cointegran. Una condición suficiente para la existencia de cointegración, en el modelo más simple analizado (IV.1), es que $|\phi| < 1$. Si $\phi = 1$, entonces el MLP no se encontrará bien definido e y_t y

APUNTES DE ESTUDIO

x_t no cointegrarán. La otra cara de la misma moneda es que el modelo de corrección de errores tendrá $\alpha = 0$ y, por tanto, el error “no se corregirá”. Puede, por supuesto, existir una relación entre Δy_t y Δx_t , pero no entre sus niveles.

No obstante, la ausencia de cointegración puede generar distorsiones en la inferencia sobre la relación entre dos variables integradas. Emerge el fenómeno conocido como **regresión espuria** (regresión falsa, regresión engañosa), que ha sido conocido desde hace mucho tiempo y es rigurosamente estudiado en Phillips (1986). A continuación, se presentan simulaciones de Monte Carlo que ilustran los principales hallazgos de este estudio.

Suponga que tenemos dos series estacionarias, y_{1t} e y_{2t} , generadas como un proceso AR(1) $y_{it} = \rho y_{i,t-1} + \varepsilon_{it}$ ($i = 1, 2$), donde ε_{1t} y ε_{2t} son dos ruidos blancos **independientes**. No existe relación entre y_{1t} e y_{2t} . Luego, ¿qué esperaríamos de los resultados de una regresión $y_{1t} = \beta y_{2t} + \text{error}_t$? La respuesta es, obviamente, que los estadísticos de regresión revelen que no existe una relación entre estas variables, $\beta = 0$. En particular:

- El coeficiente de determinación R^2 debería ser muy bajo. Conforme T se incrementa, R^2 debería converger a cero.
- Análogamente, el estadístico F de significación conjunta (que en la regresión simple propuesta es el cuadrado del ratio t para $H_0 : \beta = 0$), debería ser también bajo. Este estadístico, por supuesto, podría rechazar $H_0 : \beta = 0$ por la existencia de Error Tipo I, pero este debería ser un evento infrecuente, digamos el 5% de las veces.

El ejercicio se implementa en el programa Prog41_RegresionEspuria.prg. En la repetición !i, los datos se generan como:

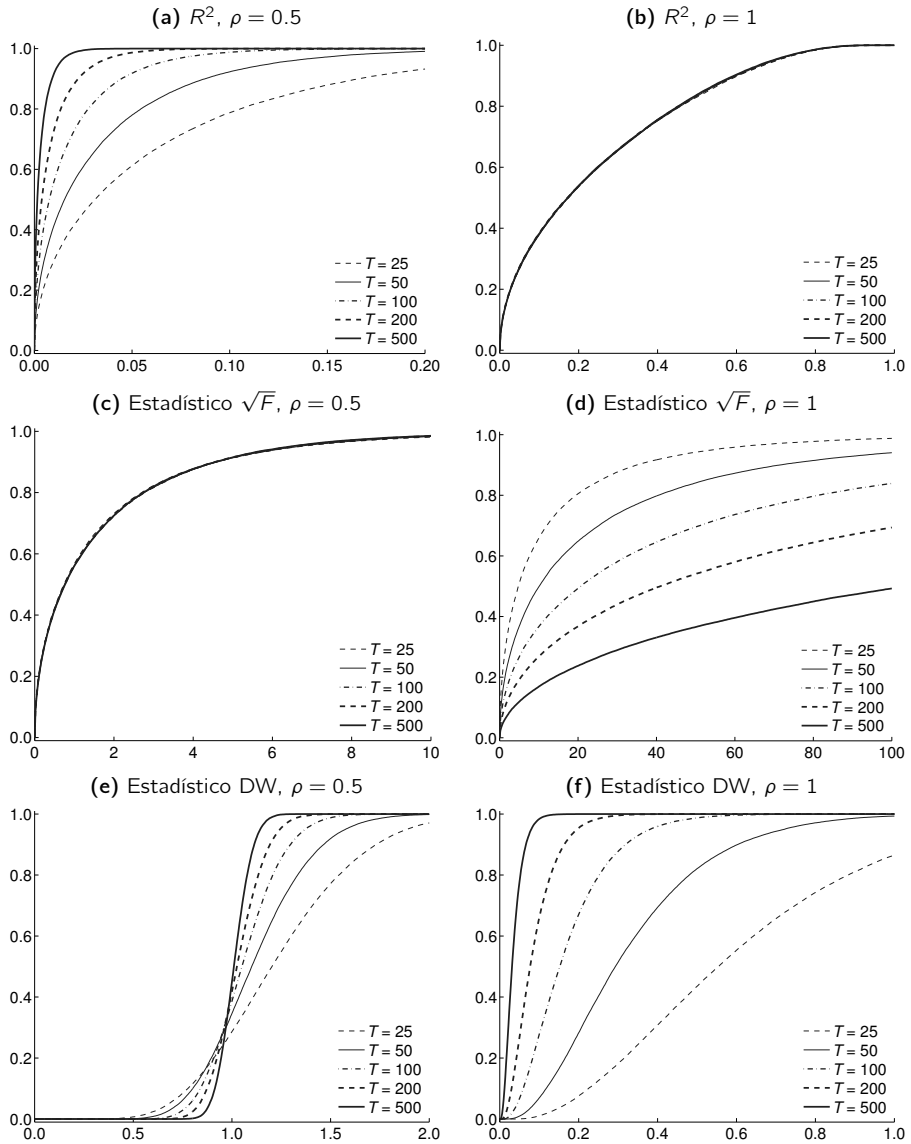
```
SERIES e1 = NRND
SERIES e2 = NRND
GENR y1 - !rho*y1(-1) = e1
GENR y2 - !rho*y2(-1) = e2,
```

donde !rho contiene el valor de ρ . Luego, para un tamaño de muestra variante !T, se corre la regresión (ecuación eq) y se almacenan los estadísticos de interés: el R^2 (serie Rcuad_{!T}), la raíz cuadrada del estadístico F (series Fstat_{!T}) y, luego explicaremos por qué, el estadístico de Durbin y Watson (serie DW_{!T}):

```
EQUATION eq.LS y1 C y2
Rcuad_{!T}(!i) = eq.@R2
Fstat_{!T}(!i) = @SQRT(eq.@F)
DW_{!T}(!i) = eq.@DW.
```

Los paneles (a) y (c) del gráfico IV.1 muestran las distribuciones empíricas de R^2 y \sqrt{F} resultantes de un experimento con $\rho = 0.5$ y $R = 100,000$ repeticiones, para diversos

Gráfico IV.1 Función de distribución empírica de estadísticos de regresión



Notas: (Prog41_RegresionEspuria.prg) Distribuciones empíricas para $R = 100,000$ replicaciones. El proceso generador de datos es $y_{it} = \rho y_{it-1} + \varepsilon_{it}$, donde $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ son ruidos blancos independientes.

tamaños muestrales. Se confirman nuestras sospechas. Se tiene que $R^2 \rightarrow 0$ conforme $T \rightarrow \infty$ y que el estadístico F sigue una distribución concentrada en valores relativamente pequeños. Al repetir exactamente el mismo experimento con $\rho = 1$, que es el caso donde $y_{it} \sim I(1)$, se percibe un comportamiento muy distinto, reportado en los paneles (b) y (d) del gráfico IV.1. A saber:

- El R^2 de la regresión **no converge a cero** conforme T se incrementa, sino que converge a una variable aleatoria. Es decir, incluso asintóticamente, se tiene una probabilidad positiva de que el coeficiente R^2 tome valores elevados.
- El estadístico F **no converge a una variable aleatoria**, sino que **diverge** con T .

En pocas palabras, a pesar de que no existe una relación entre y_{1t} e y_{2t} , cuando estas series no son estacionarias, los diagnósticos usuales de regresión se ven distorsionados, al punto que podrían encontrarse valores de R^2 y del estadístico F elevados, dando así la impresión de que la regresión ha descubierto una relación significativa entre las variables.

Suponga que efectivamente se estima $\hat{\beta} \simeq 0$. Luego, el residuo de la regresión es $\hat{e}_t = y_{1t} - \hat{\beta}y_{2t} \simeq y_{1t}$. Así, cuando $\rho = 0.5$, este residuo, que es muy similar a y_{1t} , es estacionario. Por su parte, cuando $\rho = 1$, **el residuo de la regresión no es estacionario** (una clara manifestación de que ambas series **no cointegran**) y esta es la fuente de las distorsiones en los diagnósticos de la regresión.

El estadístico de Durbin y Watson es utilizado para estudiar la presencia de correlación serial de primer orden en los residuos de una regresión. Se sabe que $DW \simeq 2(1 - r_e)$, donde r_e es la primera autocorrelación muestral del término de error. Este estadístico, además, puede proveer información rápida sobre la estacionariedad de \hat{e}_t . En el caso estacionario, luego se espera que DW converja a $2(1 - \rho) = 1$ dado que $\rho = 0.5$, que es el comportamiento que se verifica en el panel (d) del gráfico IV.1. En la regresión espuria, **DW converge a cero**, lo que se confirma en el panel (e) del mismo gráfico.

En conclusión: si al regresar y_{1t} sobre y_{2t} los residuos son $I(1)$, entonces no hay cointegración y la regresión es espuria ($DW \simeq 0$); en cambio, si los residuos son $I(0)$, entonces la regresión es una ecuación de cointegración ($DW > 0$).

Es bueno mencionar que no es la primera vez que enfrentamos una regresión espuria. Un fenómeno similar emergió en el caso estudiado en la sección III.4 (programa Prog33_DurlaufPhillips.prg, gráfico III.4), donde se tenía un modelo de regresión bien especificado (el modelo DS), cuya inferencia respondía a lo esperado, y un modelo alternativo mal especificado (el modelo TS), cuyos estadísticos de regresión presentaban un comportamiento inusual. La mala especificación tenía como consecuencia, precisamente, la presencia de residuos no estacionarios que distorsionaban la inferencia, llevando posiblemente (al investigador desinformado) a conclusiones equivocadas sobre la verdadera naturaleza de los datos. El lector puede verificar que el estadístico DW es también de utilidad como alerta de regresión espuria en el caso de la sección III.4 (para el modelo TS, DW converge a cero).

IV.3 Estimación e inferencia

Nos centraremos en el estudio de la ecuación (IV.1) con $c = \theta_1 = 0$, por simplicidad. La ecuación de cointegración asociada es $y_t = \gamma x_t + e_t$. El estimador MCO de γ satisface

$$T^n(\hat{\gamma} - \gamma) \xrightarrow{p} 0 \text{ para } 0 \leq n < 1 \quad \text{y} \quad T(\hat{\gamma} - \gamma) \xrightarrow{d} \text{Variable aleatoria.}$$

Esto es, $\hat{\gamma}$ es **superconsistente**. Conviene recordar la discusión más general desarrollada en la sección III.11 y notar que γ pertenece al grupo Ψ de Sims, Stock y Watson (1990).

La superconsistencia de $\hat{\gamma}$ tiene dos implicancias relevantes en la práctica. La primera tiene que ver con el contraste de cointegración conocido como la **prueba de Engle y Granger (1987)**. Como se ha discutido, los residuos de la regresión de y_t sobre x_t , \hat{e}_t , deberían ser estacionarios para concluir que estas series cointegran. Note que $\hat{e}_t = e_t - (\hat{\gamma} - \gamma)x_t$. Así, como $(\hat{\gamma} - \gamma)$ tiende a cero rápidamente, entonces si $e_t \sim I(0)$, debe ocurrir que $\hat{e}_t \sim I(0)$. La prueba de Engle y Granger consiste en contrastar la presencia de una raíz unitaria en \hat{e}_t con una prueba de Dickey y Fuller Aumentada. No obstante, a pesar de la superconsistencia de $\hat{\gamma}$, el error muestral de reemplazar γ por $\hat{\gamma}$ podría alterar los valores críticos de esta prueba. Los valores críticos son simulados por MacKinnon (1996) y difieren de los de las distribuciones DF, entre otras cosas, porque dependen del número de parámetros estimados en la ecuación de cointegración. Si γ fuera un coeficiente conocido (por ejemplo, provisto por la teoría económica), la prueba de Engle y Granger no sería más que una prueba de Dickey y Fuller Aumentada estándar (con los valores críticos de DF) para e_t .

La segunda implicancia se asocia con la inferencia en la relación dinámica entre y_t y x_t . Considere el MCE

$$\Delta y_t = \alpha(y_{t-1} - \gamma x_{t-1}) + \pi \Delta x_t + \varepsilon_t,$$

que involucra series $I(0)$. Si γ fuera conocido, MCO redituaría estimadores consistentes y asintóticamente normales de α y π (estos parámetros pertenecen al grupo grupo Υ de Sims, Stock y Watson [1990]). Pero ¿qué ocurre cuando γ precisa ser estimado? Damos una respuesta a través de un ejercicio de Monte Carlo.

En el programa Prog42_EngleGranger.prg se generan datos según

```
GENR D(x) = NRND^2 - 1
GENR y - !phi*y(-1) = !theta*x + @RTDIST(3)/@SQRT(3) .
```

Claramente, x_t es un paseo aleatorio, mientras que y_t se genera siguiendo la ecuación (IV.1) con $c = \theta_1 = 0$, siendo ε_t un ruido blanco de media cero, varianza igual a uno y colas anchas. El valor de ϕ se almacena en el escalar `!phi` y el valor de θ_0 , en el escalar `!theta`. Además, se definen los parámetros del MCE como `!alpha = !phi - 1`, `!gamma = !theta/(1 - !phi)` y `!pi = !theta`.

APUNTES DE ESTUDIO

Luego, para un tamaño muestral T y en la repetición i , se estima el MCE asumiendo que γ es conocido:

```
EQUATION eq1.LS D(y) (y(-1) - !gamma*x(-1)) D(x) C
MD1alpha(!i, !it) = (eq1.C(1) - !alpha)/eq1.@STDERRS(1)
MD1pi(!i, !it)    = (eq1.C(2) - !pi)/eq1.@STDERRS(2) .
```

Note que el ratio t correctamente centrado del estimador de α se almacena en una matriz de nombre MD1alpha, mientras que el ratio t centrado del estimador de π se almacena en la matriz MD1pi. Estas matrices tienen tantas filas como repeticiones R (por lo que el contador i se asocia con una fila) y tantas columnas como distintos valores de T (por lo que el contador it se asocia con una columna). Esta es una forma distinta (y ligeramente más eficiente en términos de tiempo de ejecución) de almacenar los resultados de la propuesta en simulaciones similares (por ejemplo, en los programas Prog31_Tendencia.prg o Prog311_ADF.prg), donde el almacenamiento ocurría sobre series y no matrices. Al final del programa, el contenido de estas matrices es asignado a un conjunto de series para graficar las funciones de distribución a lo largo de repeticiones. Ello se consigue con el comando MTOS(MD1alpha, D1alpha), donde D1alpha es un grupo de series.

Luego, se procede a estudiar la situación en donde γ precisa ser estimado. Se tiene que:

```
EQUATION coint.LS y x
!gammahat = coint.C(1)
MD0gamma(!i, !it) = !gammahat - !gamma
MD1gamma(!i, !it) = @SQR(!T)*(!gammahat - !gamma)
MD2gamma(!i, !it) = !T*(!gammahat - !gamma) .
```

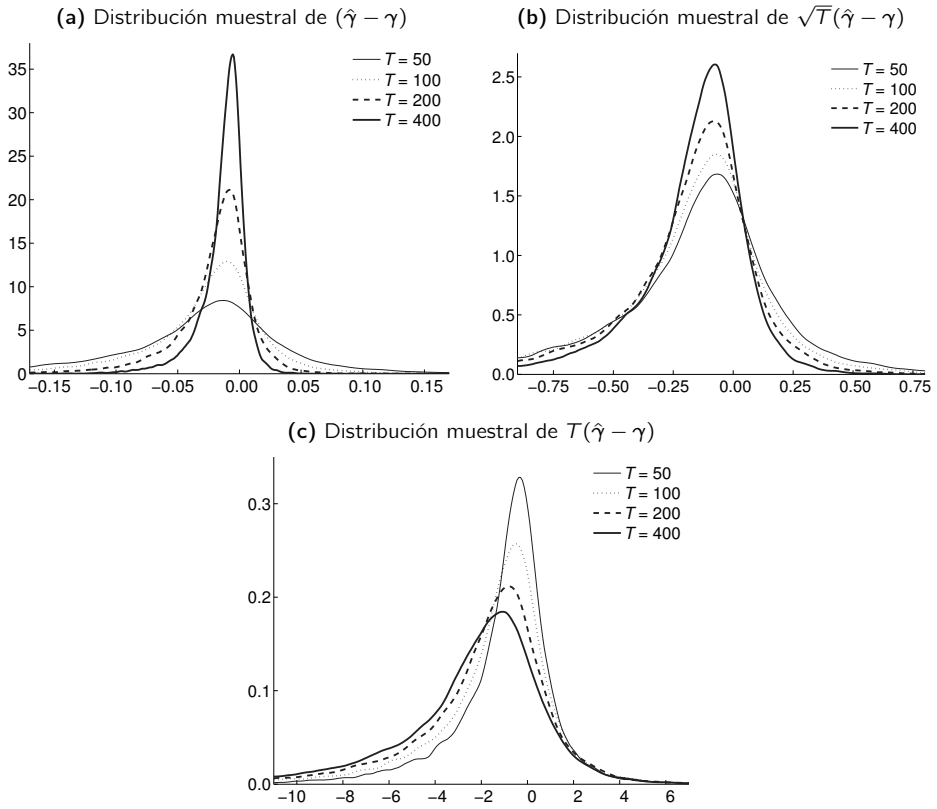
El objeto coint contiene la ecuación de cointegración, y los estadísticos $\sqrt{T^{1a}}(\hat{\gamma} - \gamma)$ son almacenados en las matrices MD{!a}gamma para $a = 0, 1, 2$, para así evaluar la superconsistencia de $\hat{\gamma}$. Asimismo, este estimador es utilizado para generar un nuevo grupo de estimaciones del MCE. Este enfoque se conoce como **procedimiento en dos etapas de Engle y Granger**. A saber,

```
EQUATION eq2.LS D(y) (y(-1) - !gammahat*x(-1)) D(x) C
MD2alpha(!i, !it) = (eq2.C(1) - !alpha)/eq2.@STDERRS(1)
MD2pi(!i, !it)    = (eq2.C(2) - !pi)/eq2.@STDERRS(2) ,
```

donde la única diferencia entre eq1 y eq2 es que la segunda ecuación utiliza !gammahat en lugar del conocido !gamma.

En el gráfico IV.2 se muestran las distribuciones muestrales contenidas en las matrices MD{!a}gamma, para diversos valores de T . En el panel (a) se observa la convergencia en

Gráfico IV.2 Superconsistencia del MLP en el procedimiento de Engle y Granger

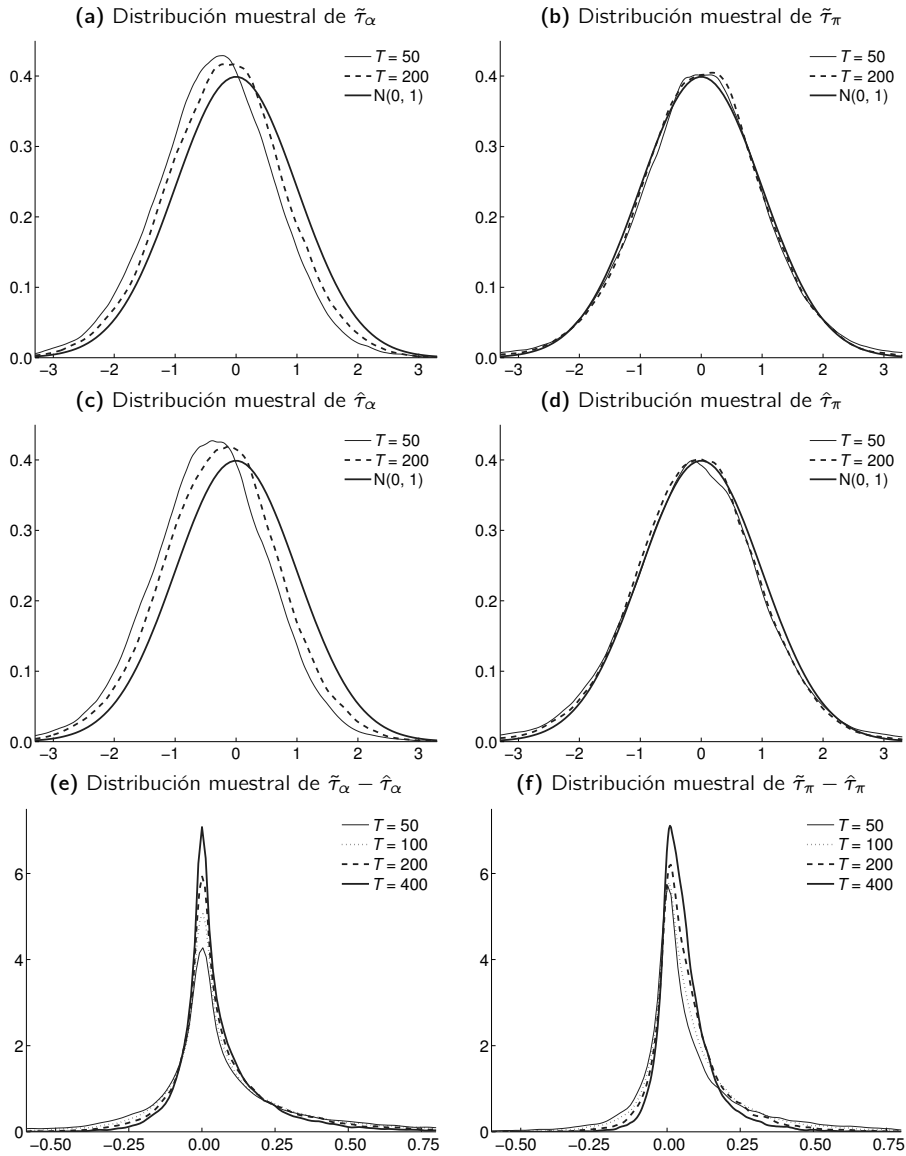


Notas: (Prog42_EngleGranger.prg) Densidades kernel (automáticas) con $R = 100,000$ repeticiones. El proceso generador de datos es el modelo ARDL $y_t = \phi y_{t-1} + \theta x_t + \varepsilon_t$, donde x_t es un paseo aleatorio y ε_t es un ruido blanco de colas anchas y varianza igual a 1. El MLP se define como $\gamma = \theta / (1 - \phi)$. La escala de los ejes difiere entre paneles.

probabilidad a cero de $\hat{\gamma} - \gamma$, mientras que en el panel (b) se verifica que esta convergencia es más rápida que la tasa de divergencia de $T^{1/2}$. En otras palabras, se confirma la superconsistencia de $\hat{\gamma}$. El panel (c) muestra que $T(\hat{\gamma} - \gamma)$ converge a una variable aleatoria.

Por su parte, el gráfico IV.3 muestra las distribuciones muestrales de los ratios t centrados almacenados en las matrices $MD\{!a\}alpha$, en los paneles (a) y (c), y $MD\{!a\}pi$, en los paneles (b) y (d). Los ratios $\tilde{\tau}$ asumen que el MLP es conocido, mientras que los ratios $\hat{\tau}$ utilizan el estimador superconsistente $\hat{\gamma}$. Es bastante claro que, como se esperaba, $\tilde{\tau}_\alpha$ y $\tilde{\tau}_\pi$ convergen en distribución a variables normales estándares. Este también parece ser el comportamiento de $\hat{\tau}_\alpha$ y $\hat{\tau}_\pi$. De hecho, el panel (e) muestra que la diferencia $\tilde{\tau}_\alpha - \hat{\tau}_\alpha$ va

Gráfico IV.3 Equivalencias asintóticas en el procedimiento de Engle y Granger



Notas: (Prog42_EngleGranger.prg) Densidades kernel (automáticas) para $R = 100,000$ repeticiones, véanse las notas al gráfico IV.2. Los ratios $\tilde{\tau}$ asumen que el MLP es conocido, mientras que los ratios $\hat{\tau}$ utilizan el estimador $\hat{\gamma}$.

aproximándose a cero conforme T se incrementa. El panel (f) reporta el mismo fenómeno para la diferencia $\tilde{\tau}_\pi - \hat{\tau}_\pi$. Se tiene, en resumen, que los estimadores $\tilde{\alpha}$ y $\tilde{\pi}$ del MCE con γ conocido son **asintóticamente equivalentes** a los estimadores $\hat{\alpha}$ y $\hat{\pi}$ del MCE con γ estimado. Así, estos hallazgos sugieren que se puede proceder en dos etapas y reemplazar γ por $\hat{\gamma}$ sin alterar las propiedades asintóticas de la estimación MCO del MCE.

No es la primera vez que aparecen resultados de equivalencia asintótica para un grupo de parámetros, una vez que se reemplazan ciertos valores desconocidos por estimadores superconsistentes. Este fenómeno apareció también en nuestros estudios de las secciones III.1 (programa Prog31_Tendencia.prg, gráfico III.2) y III.11 (programa Prog311_ADF.prg, gráfico III.14). Las fuerzas detrás de este fenómeno son exactamente las mismas, tanto en el caso univariado del capítulo anterior, como en el caso multivariado ilustrado a través del procedimiento de Engle y Granger.

IV.4 Inferencia sobre el multiplicador de largo plazo

La ecuación de cointegración (IV.2) es una regresión, en general, mal especificada. Salvo en casos muy rebuscados, el término de error e_t es autocorrelacionado y podría estar correlacionado con el regresor x_t . En un mundo $I(0)$, el hecho de que $\mathbb{E}(x_t e_t) \neq 0$ nos impide obtener un estimador MCO consistente para γ . Cuando $x_t \sim I(1)$, no obstante, el estimador MCO es superconsistente incluso si la regresión se encuentra mal especificada. Este resultado ha sido documentado en el gráfico IV.2 y es un fenómeno muy similar al reportado en la sección III.10 (programa Prog39_Nuisance.prg, gráfico III.9), cuando se analizó la regresión de Dickey y Fuller con errores autocorrelacionados. La superconsistencia de $\hat{\gamma}$ facilita la inferencia en los parámetros del MCE.

No obstante, la noción de superconsistencia de $\hat{\gamma}$ **no debe exagerarse si el objetivo es realizar inferencias sobre γ** . La mala especificación de la ecuación de cointegración no impide obtener un estimador consistente pero sí afecta su distribución, que se contamina con **parámetros fastidiosos**, tal y como ocurría con las regresiones de Dickey y Fuller. Los parámetros fastidiosos incluyen, en esta ocasión, la varianza de largo plazo de Δx_t y la covarianza $\mathbb{E}(x_t e_t)$. La distribución a la que converge $\hat{\gamma}$ en el panel (c) del gráfico IV.2 contiene estos parámetros.

Exploraremos, a través de simulaciones, las propiedades de varios estimadores de γ , con especial atención en las posibilidades de inferencia que estos nos brindan, más allá de la superconsistencia. El ejercicio se implementa en el programa Prog43_MLP.prg que considera un proceso generador de datos muy similar al del programa Prog42_EngleGranger.prg, GENR D(x) = NRND y GENR y - !phi*y(-1) = !theta*x + NRND, pero para un tamaño muestral fijo en $T = 150$ observaciones. Recuerde que el MLP se almacena en el escalar !gamma = !theta/(1 - !phi).

APUNTES DE ESTUDIO

En todos los casos, el interés se centra en el ratio t correctamente centrado,

$$\tau_\gamma = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_\gamma},$$

donde $\hat{\sigma}_\gamma$ es un estimador de la desviación estándar de $\hat{\gamma}$.

En primer lugar, se tiene el estimador ya estudiado que resulta de aplicar MCO a la ecuación de cointegración (IV.2). Este se conoce como el estimador **MCO estático**, ya que ignora la dinámica “de corto plazo” entre y_t y x_t . Para la repetición i ,

```
EQUATION eq1.LS y x C
!num = eq1.C(1) - !gamma
!den = eq1.@STDERRS(1)
tstat1(!i) = !num!/den,
```

donde `tstat1` es una serie de R observaciones previamente definida. Asimismo, se considera un refinamiento a esta regresión estática. Una práctica bastante común en aplicaciones es la de utilizar errores estándar HAC (*à la* Newey y West) cuando se sospecha que el error de regresión puede presentar correlación serial. EViews implementa esta corrección de una manera muy sencilla, simplemente utilizando la opción adecuada en el momento de ejecutar el comando LS:

```
EQUATION eq1c.LS(n) y x C
!num = eq1c.C(1) - !gamma
!den = eq1c.@STDERRS(1)
tstat1c(!i) = !num!/den.
```

La diferencia entre `tstat1` y `tstat1c` es el estimador $\hat{\sigma}_\gamma$ en el denominador del ratio t .

En segundo lugar, podría pensarse en aumentar la regresión estática con términos dinámicos, siguiendo el razonamiento detrás de la regresión de Dickey y Fuller Aumentada:

```
EQUATION eq2.LS y x D(y(-1)) D(x) D(x(-1)) C
!num = eq2.C(1) - !gamma
!den = eq2.@STDERRS(1)
tstat2(!i) = !num!/den.
```

Esta es una versión rudimentaria del método conocido como **MCO dinámico** propuesto por Stock y Watson (1993), en donde la selección de los términos aumentados se hace de una manera más cuidadosa, basada, por ejemplo, en criterios de información. Una segunda versión del método considera el estimador HAC de los errores estándar, utilizando LS(n) en lugar de LS, y almacena los ratios t en `tstat2c`.

El tercer estimador ya no se basa en la regresión de cointegración, sino en el MCE:

```
EQUATION eq3.LS D(y) = C(1)*(y(-1) - C(2)*x(-1)) + C(3)*D(x) + C(4)
!num = eq3.C(2) - !gamma
!den = eq3.@STDERRS(2)
tstat3(!i) = !num!/den.
```

Note que el MCE es lineal **condicional en** γ , pero si γ se incluye dentro de los parámetros por estimar, entonces el MCE es un modelo no lineal. Luego, se precisa estimar γ por MC no lineales. Para estimar modelos no lineales en EViews, simplemente hay que especificar completamente la forma de la ecuación por estimar, y sus coeficientes. El segundo coeficiente de la ecuación eq3 contendrá el estimado de γ de interés. Note que no se incorporan correcciones HAC. Es importante mencionar que el MCE estimado en la ecuación eq3 es un **modelo bien especificado**, ya que es simplemente una reparametrización del proceso generador de datos.

El cuarto estimador considerado también se deriva del MCE. Considere la ecuación **lineal**

$$\Delta y_t = c_1 y_{t-1} + c_2 x_{t-1} + c_3 \Delta x_t + \text{error}_t.$$

Se aprecia que $\alpha = c_1$ y $\theta = -c_2$. Así, $\gamma = -\theta/\alpha = -c_2/c_1$. Con ello, el estimador de γ es una función no lineal (una razón) de dos estimadores estimados por MCO. Para estimar su desviación estándar, se debe utilizar el **método delta**. Si $\gamma = g(c_1, c_2)$ y, por tanto, $\hat{\gamma} = g(\hat{c}_1, \hat{c}_2)$, el estimador de la varianza de $\hat{\gamma}$ del método delta es

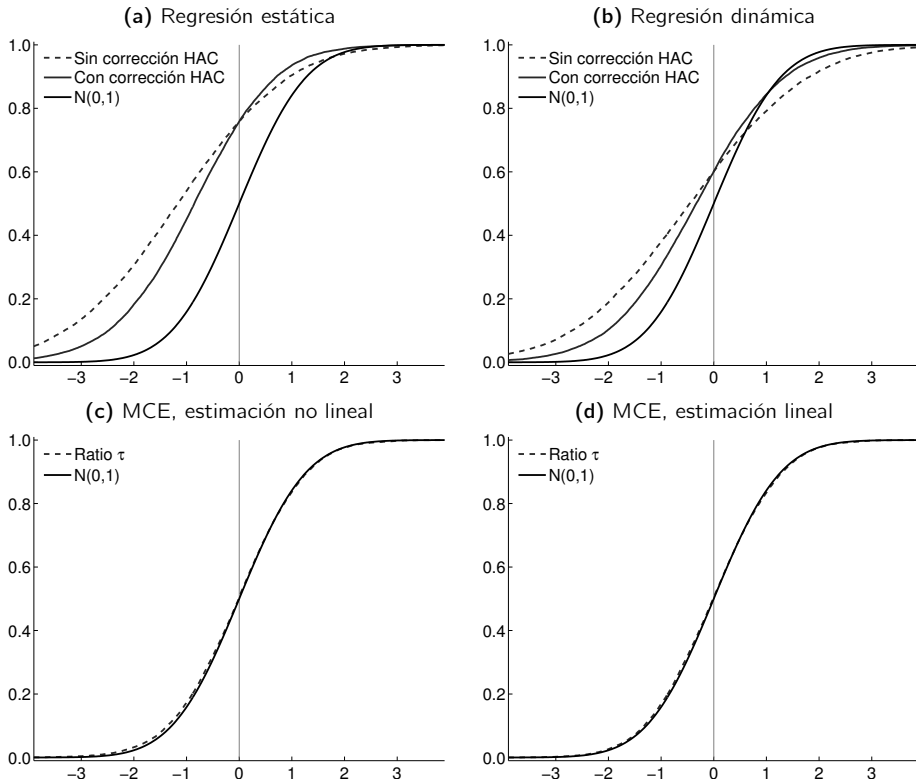
$$\hat{V}(\hat{\gamma}) = \left(\frac{\partial g(\cdot)}{\partial c_1} \right)^2 \hat{V}(\hat{c}_1) + \left(\frac{\partial g(\cdot)}{\partial c_2} \right)^2 \hat{V}(\hat{c}_2) + 2 \left(\frac{\partial g(\cdot)}{\partial c_1} \right) \left(\frac{\partial g(\cdot)}{\partial c_2} \right) \hat{C}(\hat{c}_1, \hat{c}_2).$$

Así:

```
EQUATION eq4.LS D(y) y(-1) x(-1) D( x ) C
!gammahat = -eq4.C(2)/eq4.C(1)
!dgd1      = -!gammahat/eq4.C(1)
!dgd2      = -1/eq4.C(1)
!num       = !gammahat - !gamma
!den2      = (!dgd1^2)*eq4.@COEFCOV(1,1) + (!dgd2^2)*eq4.@COEFCOV(2,2) + ...
            ... + 2*(!dgd1*!dgd2)*eq4.@COEFCOV(1,2)
tstat4(!i) = !num/@SQR(!den2).
```

Note que eq.@COEFCOV(i, j) recupera el (i, j)-ésimo elemento de la matriz de covarianzas del vector de coeficientes estimados en la ecuación eq. Esta matriz no incorpora una corrección tipo HAC.

Gráfico IV.4 Inferencia sobre el multiplicador de largo plazo



Notas: (Prog43_MLP) Funciones de distribución empírica del ratio τ_γ centrado, con $R = 100,000$ repeticiones. El proceso generador de datos es $y_t = \phi y_{t-1} + \theta x_t + \varepsilon_t$, donde $\Delta x_t \sim N(0, 1)$, $\varepsilon_t \sim N(0, 1)$ y $\gamma = \theta / (1 - \phi)$.

El gráfico IV.4 presenta las distribuciones de probabilidad acumuladas de los ratios τ_γ , para $R = 100,000$ repeticiones. Cada panel muestra, además, la función de distribución normal como referencia. Los resultados revelan que $\tau_\gamma \xrightarrow{d} N(0, 1)$ únicamente con los métodos que atacan directamente al MCE (posiblemente, se alcanzaría un resultado similar con MCO dinámico de aplicar un mejor método de selección del modelo). El MCE consigue este resultado deseable al tratar **simultáneamente** al MLP γ con los parámetros de corto plazo, y así garantizar una correcta especificación de la relación dinámica entre y_t y x_t .

Esta es la principal motivación para el uso del celebrado método de Johansen (1995), que es en espíritu similar al cuarto enfoque considerado en nuestro estudio y promete el mismo resultado: **a pesar de que la distribución de $\hat{\gamma}$ sea inusual**, una correcta especificación dinámica conlleva que τ_γ sea asintóticamente normal. La inferencia clásica se restablece.

Resumen de programas

CAPÍTULO 1: INTEGRACIÓN DE MONTE CARLO

Prog11_pi.prg

- Objetivo : Aproximar π mediante Integración de Monte Carlo.
- Descripción : Se toma un gran número de observaciones $x \sim U(0, 1)$ y se evalúa $g(x)$ tal que su esperanza, aproximada por el promedio de $g(x)$ a lo largo de observaciones aleatoriamente generadas, sea π . Se sigue un enfoque similar con una función bivariada $g(x, y)$.
- Resultados : Aproximaciones en el gráfico l.1 (p. 6) y cálculos en el texto.

Prog12_MC.prg

- Objetivo : Ilustrar el funcionamiento de la Integración de Monte Carlo, aproximando momentos de distribuciones de probabilidad conocidas.
- Descripción : Se toma un gran número de observaciones $x \sim F$, donde F es alguna distribución conocida (uniforme, normal, t de Student, chi cuadrado, exponencial o beta), y se compara el promedio de $g(x)$ con $\mathbb{E}(g(x))$ que es una cantidad también conocida, para varios tamaños muestrales.
- Resultados : Las cifras comparativas del cuadro l.1 (p. 10).

Prog13_Eficiencia.prg

- Objetivo : Explorar la eficiencia relativa del promedio y mediana muestrales como estimadores de locación de una variable aleatoria cuya distribución pueda presentar colas anchas.
- Descripción : Para varios valores de ν , se generan datos $y \sim t_\nu$. Se almacenan el promedio y la mediana de y , y se comparan sus varianzas a lo largo de repeticiones.
- Resultados : La eficiencia relativa mostrada en el gráfico l.2(b) (p. 14).

Prog14_ PreTesting.prg

- Objetivo : Analizar las propiedades de estimadores en un modelo de regresión simple.
- Descripción : Para varios valores de α , se generan datos $y = \alpha + \beta x + \varepsilon$, donde $\varepsilon \sim N(0, \sigma^2)$, y se calculan los estimadores de mínimos cuadrados ordinarios, mínimos cuadrados restringidos y un híbrido que elige entre estos estimadores según los resultados de una prueba de significación estadística.
- Resultados : Los momentos mostrados el gráfico I.3 (p. 22) y las distribuciones muestrales del gráfico I.4 (p. 23).

Prog15_ LeyesAsintoticas.prg

- Objetivo : Ilustrar el funcionamiento de la Ley Débil de Grandes Números y del Teorema del Límite Central en muestras aleatorias.
- Descripción : Para varios valores de T y de a , se generan datos y después de estandarizar u^a , donde $u \sim N(0, 1)$, de modo que la media y varianza de y son siempre 0 y 1, respectivamente. Luego, se almacenan los promedios muestrales \bar{y} y de $\sqrt{T}\bar{y}$.
- Resultados : Las distribuciones muestrales del gráfico I.5 (p. 26).

CAPÍTULO 2: SERIES DE TIEMPO ESTACIONARIAS**Prog21_ LeyesRevisadas.prg**

- Objetivo : Ilustrar el funcionamiento del Teorema Ergódico y del Teorema del Límite Central aplicado a procesos dependientes.
- Descripción : Para varios valores de T , se generan datos según el modelo ARMA(1,1) $y_t = (1 - \phi)\mu + \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$, donde $\varepsilon_t \sim iid(0, 1)$, utilizando varias combinaciones de los parámetros ϕ y θ . Luego, se almacenan los promedios muestrales \bar{y} y el estadístico $\sqrt{T/\Omega}(\bar{y} - \mu)$, donde Ω es la varianza de largo plazo de y_t . Se almacenan, además, las primeras correlaciones muestrales de estos procesos.
- Resultados : Las distribuciones muestrales de los gráficos II.1 (p. 40) y II.2 (p. 42).

Prog22_Cobertura.prg

- Objetivo : Evaluar las propiedades asintóticas de estimadores HAC de la varianza de largo plazo $\hat{\alpha}$ a la Newey y West.
- Descripción : Para varios tamaños muestrales T , se generan datos según el proceso MA(1) $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ y el proceso AR(1) $y_t = (1 - \phi)\mu + \phi y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$, y se calculan dos estimadores HAC de sus varianzas de largo plazo. Luego, se procede a evaluar si el intervalo de confianza construido con estas varianzas estimadas contiene efectivamente a la media poblacional μ .
- Resultados : Las probabilidades de cobertura reales del cuadro II.1 (p. 46).

Prog23_SesgoAR.prg

- Objetivo : Estimación del sesgo en muestras finitas del estimador MCO de ϕ en el modelo AR(1) $y_t = \phi y_{t-1} + \text{error}_t$.
- Descripción : Para varios tamaños muestrales T y varios valores del coeficiente autorregresivo ϕ , se generan datos según el proceso AR(1) $y_t = \phi y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$, y se calcula el estimador MCO de una regresión de y_t sobre y_{t-1} . Se almacena $\hat{\phi} - \phi$, cuyo promedio a lo largo de repeticiones converge al sesgo.
- Resultados : Las estimaciones del gráfico II.3 (p. 49).

Prog24_LeyesARMA.prg

- Objetivo : Estudiar la consistencia y normalidad asintótica del estimador MCO o MC no lineal de modelos de la familia ARMA.
- Descripción : Para varios valores de T , se generan datos según el modelo ARMA(1,1) $y_t = (1 - \phi)\mu + \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, donde $\varepsilon_t \sim iid(0, 1)$, utilizando varias combinaciones de los parámetros ϕ y θ . Luego, se estiman por mínimos cuadrados los parámetros ϕ y θ , que son almacenados junto con sus ratios t correctamente centados.
- Resultados : Las distribuciones muestrales del gráfico II.4 (p. 53).

Prog25_ Autocorrelacion

- Objetivo : Estudiar la potencia de las pruebas de autocorrelación serial Q de Ljung y Box (1978) y LM de Godfrey (1978) y Breusch (1979).
- Descripción : Para varios valores de T y de ρ , se generan datos según el modelo AR(2) $y_t = (\phi + \rho)y_{t-1} - \phi\rho y_{t-2} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1 - \rho^2)$, se estima un modelo AR(1) mal especificado y se registra la proporción de veces que las pruebas Q y LM detectan esta mala especificación.
- Resultados : Curvas de potencia, que presentan la proporción de rechazos del modelo AR(1) como función de ρ y de T , mostradas en el gráfico II.7 (p. 59).

Prog26_ CritInfo

- Objetivo : Evaluar el desempeño de varios criterios de información (Akaike, Akaike corregido y Schwarz) en la selección de modelos dinámicos.
- Descripción : Para varios valores de tamaños muestrales, se genera un proceso ARMA(3,1) con varias parametrizaciones y se estiman modelos AR(p), donde p es elegido tal que se minimizan los criterios de información mencionados. Luego, se pasa a evaluar las bondades de la elección de p , sobre todo en términos de predicción.
- Resultados : Indicadores de desempeño comparado de criterios de información alternativos, reportados en el cuadro II.2 (p. 65).

CAPÍTULO 3: NO ESTACIONARIEDAD**Prog31_ Tendencia.prg**

- Objetivo : Analizar las propiedades (superconsistencia y equivalencias asintóticas) del estimador MCO de un modelo con tendencia lineal.
- Descripción : Para varios valores de T , se generan datos según $y_t = \alpha + \delta t + u_t$, donde $u_t = \phi u_{t-1} + \varepsilon_t$, $|\phi| < 1$ y $\varepsilon_t \sim iid(0, 1)$. Se estima δ al regresar y_t sobre un intercepto y t , y se estima ϕ de dos formas: regresando u_t sobre u_{t-1} , y e_t sobre e_{t-1} , siendo e_t el residuo de la regresión que estimó δ .
- Resultados : Las distribuciones muestrales del estimador de δ del gráfico III.1 (p. 73) y las distribuciones muestrales de los estimadores de ϕ del gráfico III.2 (p. 75).

Prog32_ RandomWalk.prg

- Objetivo : Comparar el comportamiento asintótico de los promedios de un paseo aleatorio con el de los promedios de procesos persistentes pero estacionarios.
- Descripción : Para varios valores de T , se generan paseos aleatorios $\Delta z_t = \varepsilon_t$ y procesos AR(1) con raíz cerca a la unidad, $u_t = 0.975u_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$. Luego se almacenan los promedios de estas series, adecuadamente estabilizados.
- Resultados : Distribuciones muestrales presentadas en el gráfico III.3 (p. 83).

Prog33_ DurlaufPhillips.prg

- Objetivo : Comparar las propiedades de los estimadores del *drift* de una serie estacionaria en diferencias, con un modelo DS y un modelo TS. Basado en Durlauf y Phillips (1988).
- Descripción : Para varios valores de T , se generan paseos aleatorios con tendencia $\Delta y_t \sim N(\delta, 1)$ y se estima el coeficiente δ y ratios t para su inferencia, con el modelo DS ($\Delta y_t = \delta + \text{error}_t$) y con el modelo TS ($y_t = \delta t + \text{error}_t$).
- Resultados : Funciones de distribución empíricas de los ratios t del gráfico III.4 (p. 89).

Prog34_ DickeyFuller.prg

- Objetivo : Determinar la distribución del estimador MCO de ρ en regresiones del tipo $y_t = c + dt + \rho y_{t-1} + \text{error}_t$, así como del ratio t que contrasta $H_0 : \rho = 1$, cuando el proceso generador de datos de y_t corresponde a un paseo aleatorio. Basado en Dickey y Fuller (1979).
- Descripción : Para un valor dado de T , se generan paseos aleatorios $\Delta y_t \sim N(0, 1)$ y se procede a estimar las ecuaciones de contraste sin términos determinísticos ($c = 0, d = 0$), con intercepto ($c \neq 0, d = 0$) y con tendencia ($c \neq 0, d \neq 0$).
- Resultados : Distribuciones muestrales de $\hat{\rho}$ (el estimador MCO) y de τ_ϕ (el ratio t) que se muestran en el gráfico III.5 (p. 93). El programa, además, calcula los percentiles de estas distribuciones utilizados en pruebas de hipótesis.

Prog35_ DFPotencia.prg

- Objetivo : Estudiar la potencia, $\Pr(\text{rechazar } H_0 | H_0 \text{ falsa})$, en muestras finitas de las pruebas de Dickey y Fuller, ante diversas hipótesis alternativas.
- Descripción : Para diversos valores de ρ y del tamaño muestral T , se generan datos según el modelo AR(1) $y_t = \rho y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$, se estiman las ecuaciones de contraste de Dickey y Fuller y se registra el número de veces que estas pruebas rechazan $H_0 : \rho = 1$ (que es, para $\rho < 1$, una hipótesis falsa).
- Resultados : Curvas de potencia, que presentan la proporción de rechazos de H_0 como función de ρ y de T , mostradas en el gráfico III.6 (p. 99).

Prog36_SchmidtPhillips.prg

- Objetivo : Comparar la potencia, $\Pr(\text{rechazar } H_0 | H_0 \text{ falsa})$, en muestras finitas de la prueba de Dickey y Fuller, *versus* la prueba LM de Schmidt y Phillips (1992).
- Descripción : Para diversos valores de ρ y del tamaño muestral T , se generan datos según $y_t = \alpha + \delta t + u_t$, donde $u_t = \rho u_{t-1} + \varepsilon_t$ y $\varepsilon_t \sim N(0, 1)$, se estiman las ecuaciones de contraste de Dickey y Fuller y de Schmidt y Phillips, y se registra el número de veces que estas pruebas rechazan $H_0 : \rho = 1$.
- Resultados : Curvas de potencia, que presentan la proporción de rechazos de H_0 como función de ρ y de T , mostradas en el gráfico III.7 (p. 103).

Prog37_HaldrupHylleberg.prg

- Objetivo : Estudiar cómo se restablece la normalidad asintótica en estimaciones con series no estacionarias, cuando la regresión estimada omite tendencias dominantes en los datos. Basado en West (1988) y en Haldrup y Hylleberg (1995).
- Descripción : Para diversos valores de δ y de T , se generan datos como $\Delta y_t = \delta + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$, y se estima la ecuación $y_t = c + \rho y_{t-1} + \text{error}_t$. En teoría, $\hat{\rho}$ y el ratio τ_ϕ son asintóticamente normales. No obstante, para valores realistas de δ , sus distribuciones siguen pareciéndose a las de Dickey y Fuller.
- Resultados : Funciones de distribución empíricas de τ_ϕ mostradas en el gráfico III.8 (p. 105).

Prog38_DieboldKilian.prg

- Objetivo : Estudiar la utilidad de los contrastes de raíz unitaria para fines predictivos. Basado en Diebold y Kilian (2000).
- Descripción : Para diversos valores de ρ , de T y del horizonte de proyección h , se generan datos según $y_t = \alpha + \delta t + u_t$, donde $u_t = \rho u_{t-1} + \varepsilon_t$ y $\varepsilon_t \sim N(0, 1)$, y se proyectan valores futuros y_{t+h} con tres estrategias: el modelo DS ($\Delta y_t = c + \text{error}_t$), el modelo TS ($y_t = c + dt + \rho y_{t-1} + \text{error}_t$) y *pre-testing*, en donde primero se implementa una prueba de raíz unitaria y se elige el modelo DS ante no rechazos, y el modelo TS ante rechazos.
- Resultados : Indicadores de desempeño predictivo (errores cuadráticos medios de proyección) mostrados en el cuadro III.3 (p. 110).

Prog39_Nuisance.prg

- Objetivo : Estudiar el efecto de errores autocorrelacionados sobre las pruebas de raíz unitaria. Asimismo, evaluar el comportamiento de contrastes que corrigen por estos efectos. Basado en Phillips y Perron (1989) y Said y Dickey (1984).
- Descripción : Para diversos tamaños muestrales T , se genera el proceso $I(1)$ $\Delta y_t = u_t$, donde u_t puede ser un ruido blanco, un proceso MA(1) o un proceso AR(1). Luego, se implementa la corrección de Phillips y Perron y se estima la ecuación aumentada de Dickey y Fuller para almacenar el estimador MCO de ρ y los ratios τ_ϕ .
- Resultados : Distribuciones muestrales de $\sqrt{T}(\hat{\rho} - 1)$, mostradas en el gráfico III.9 (p. 113), que ilustran la prevalencia de la superconsistencia, y distribuciones muestrales de los diversos ratios τ_ϕ mostradas en el gráfico III.10 (p. 114).

Prog310_Schwert.prg

- Objetivo : Explorar el desempeño comparado de las pruebas de Phillips y Perron y de Dickey y Fuller Aumentada para el contraste de raíz unitaria. Basado en Schwert (1989).
- Descripción : Para diversos tamaños muestrales T y sendos valores de θ , se genera el proceso $I(1)$ $\Delta y_t = \varepsilon_t + \theta \varepsilon_{t-1}$, donde $\varepsilon_t \sim N(0, 1)$, y se registra el número de veces que diversos contrastes rechazan la hipótesis (verdadera) de raíz unitaria.
- Resultados : Probabilidades reales de rechazo reportadas en el cuadro III.4 (p. 122).

Prog311_ADF.prg

- Objetivo : Analizar las propiedades (superconsistencia y equivalencias asintóticas) del estimador MCO de la regresión de Dickey y Fuller Aumentada.
- Descripción : Para varios valores de T , se generan datos según $\Delta y_t = \zeta \Delta y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim iid(0, 1)$. Se estiman las ecuaciones $\Delta y_t = c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \text{error}_t$ y $\Delta y_t = \phi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \text{error}_t$ y se comparan sus resultados.
- Resultados : Las distribuciones muestrales del estimador de ϕ del gráfico III.13 (p. 126) y las distribuciones muestrales de los estimadores de c_i del gráfico III.14 (p. 127).

CAPÍTULO 4: COINTEGRACIÓN**Prog41_RegresionEspuria.prg**

- Objetivo : Analizar el comportamiento de regresiones espurias. Basado en Phillips (1986).
- Descripción : Para varios tamaños de muestra T , se generan dos paseos aleatorios independientes $\Delta z_{it} = \varepsilon_{it}$, donde $\varepsilon_{it} \sim N(0, 1)$, se estima una regresión de z_{1t} sobre z_{2t} y se almacenan estadísticos de diagnóstico de esta regresión. Para fines comparativos, se repite el procedimiento pero para dos procesos AR(1) estacionarios independientes, $y_{it} = 0.5y_{it-1} + \varepsilon_{it}$.
- Resultados : Las distribuciones empíricas de los diagnósticos del gráfico IV.1 (p. 136).

Prog42_EngleGranger.prg

- Objetivo : Analizar las propiedades (superconsistencia y equivalencias asintóticas) del procedimiento de Engle y Granger (1987).
- Descripción : Para varios valores de T , se generan datos siguiendo el modelo dinámico $y_t = \phi y_{t-1} + \theta x_t + \varepsilon_t$, donde $\varepsilon_t \sim iid(0, 1)$ e independiente de $x_t \sim I(1)$. Luego, se estima el multiplicador de largo plazo γ directamente de la ecuación de cointegración (la "regresión estática") y se estima el modelo de corrección de errores asociado bajo dos métodos: asumiendo que γ es conocido, y utilizando el γ estimado en la primera etapa.
- Resultados : Las distribuciones muestrales del estimador de γ del gráfico IV.2 (p. 140) y las distribuciones muestrales de los estimadores de α y π (coeficientes de corto plazo del modelo de corrección de errores) del gráfico IV.3 (p. 141).

Prog43_MLP.prg

- Objetivo : Analizar las propiedades de estimadores alternativos del coeficiente de cointegración (multiplicador de largo plazo), sobre todo el comportamiento de sus ratios t para la inferencia.
- Descripción : Para un tamaño muestral dado, se generan datos según el modelo dinámico $y_t = \phi y_{t-1} + \theta x_t + \varepsilon_t$, donde $\varepsilon_t \sim N(0, 1)$ e independiente de $\Delta x_t \sim N(0, 1)$. Luego, se estima el multiplicador de largo plazo $\gamma = \theta/(1 - \phi)$ bajo varios métodos y se registran los ratios t centrados τ_γ .
- Resultados : Las distribuciones empíricas de τ_γ del gráfico IV.4 (p. 145).

MATERIAL DISPONIBLE PARA DESCARGA

Los programas de EViews listados y los gráficos en alta resolución (en formato .pdf) se encuentran en:

<https://datasets.up.edu.pe/dataset.xhtml?persistentId=hdl:20.500.14139/Q2WZUA>

Bibliografía

ANDREWS, D. W. K.

1991 "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation". En: *Econometrica*, 59(3), pp. 817-858.

BANERJEE, A.; J. J. DOLADO, J. W. GALBRAITH y D. HENDRY

1993 *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Advanced Texts in Econometrics. Oxford: Oxford University Press.

BEVERIDGE, S. y C. R. NELSON

1981 "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'". En: *Journal of Monetary Economics*, 7(2), pp. 151-174.

BHARGAVA, A.

1986 "On the Theory of Testing for Unit Roots in Observed Time Series". En: *Review of Economic Studies*, 53(3), pp. 369-384.

BREUSCH, T. S.

1979 "Testing for Autocorrelation in Dynamic Linear Models". En: *Australian Economic Papers*, 17(31), pp. 334-355.

CASTRO, J. F. y R. RIVAS LLOSA

2007 *Econometría Aplicada*. Biblioteca Universitaria. Lima: Centro de Investigación de la Universidad del Pacífico.

DAVIDSON, R. y J. G. MACKINNON

1993 *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

DEZHBAKHSH, H.

1990 "The Inappropriate Use of Serial Correlation Tests in Dynamic Linear Models". En: *Review of Economics and Statistics*, 72(1), pp. 126-132.

DICKEY, D. A.; W. R. BELL y W. A. FULLER

1986 "Unit Roots in Time Series Models: Tests and Implications". En: *The American Statistician*, 40(1), pp. 12-36.

DICKEY, D. A. y W. A. FULLER

1979 "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". En: *Journal of the American Statistical Association*, 74(366), pp. 427-431.

DIEBOLD, F. y L. KILIAN

2000 "Unit-Root Tests Are Useful for Selecting Forecasting Models". En: *Journal of Business and Economic Statistics*, 18(3), pp. 265-273.

- DURLAUF, S. N. y P. C. B. PHILLIPS
1988 "Trends versus Random Walks in Time Series Analysis". En: *Econometrica*, 56(6), pp. 1333-1354.
- ELLIOTT, G.; T. J. ROTHENBERG y J. H. STOCK
1996 "Efficient Tests for an Autoregressive Unit Root". En: *Econometrica*, 64(4), pp. 813-836.
- ENDERS, W.
2010 *Applied Econometric Time Series*. 3ª edición. Nueva York: Wiley.
- ENGLE, R. F. y C. W. J. GRANGER
1987 "Co-Integration and Error Correction: Representation, Estimation, and Testing". En: *Econometrica*, 55(2), pp. 251-276.
- GODFREY, L. G.
1978 "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables". En: *Econometrica*, 46(6), pp. 1293-1302.
- HALDRUP, N. y S. HYLLEBERG
1995 "A Note on the Distribution of the Least Squares Estimator of a Random Walk with Drift: Some Analytical Evidence". En: *Economics Letters*, 48(3-4), pp. 221-228.
- HAMILTON, J. D.
1994 *Time Series Analysis*. Princeton: Princeton University Press.
- HAYASHI, F.
2000 *Econometrics*. Princeton: Princeton University Press.
- HURVICH, C. M. y C. L. TSAI
1991 "Bias of the Corrected AIC Criterion for Underfitted Regression". En: *Biometrika*, 78(3), pp. 499-509.
1989 "Regression and Time Series Model Selection in Small Samples". En: *Biometrika*, 76(2), pp. 297-307.
- JOHANSEN, S.
1995 *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Advanced Texts in Econometrics. Oxford: Oxford University Press.
- LJUNG, G. M. y G. E. P. BOX
1978 "On a Measure of a Lack of Fit in Time Series Models". En: *Biometrika*, 65(2), pp. 297-303.
- MACKINNON, J. G.
1996 "Numerical Distribution Functions for Unit Root and Cointegration Tests". En: *Journal of Applied Econometrics*, 11(6), pp. 601-618.
- MADDALA, G. S. e I. M. KIM
1998 *Unit Roots, Cointegration, and Structural Change*. Cambridge: Cambridge University Press.

NEWKEY, W. y K. D. WEST

1994 "Automatic Lag Selection in Covariance Matrix Estimation". En: *Review of Economic Studies*, 61(4), pp. 631-654.

1987 "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". En: *Econometrica*, 55(3), pp. 703-708.

NG, S. y P. PERRON

2005 "A Note on the Selection of Time Series Models". En: *Oxford Bulletin of Economics and Statistics*, 67(1), pp. 115-134.

2001 "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power". En: *Econometrica*, 69(6), pp. 1519-1554.

PHILLIPS, P. C. B.

1986 "Understanding Spurious Regressions in Econometrics". En: *Journal of Econometrics*, 33(3), pp. 311-340.

PHILLIPS, P. C. B. y P. PERRON

1989 "Testing for a Unit Root in Time Series Regression". En: *Biometrika*, 75(2), pp. 335-346.

SAID, S. E. y D. A. DICKEY

1984 "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order". En: *Biometrika*, 71(3), pp. 599-607.

SCHMIDT, P. y P. C. B. PHILLIPS

1992 "LM Tests for a Unit Root in the Presence of Deterministic Trends". En: *Oxford Bulletin of Economics and Statistics*, 54(3), pp. 257-287.

SCHWERT, W.

1989 "Tests for Unit Roots: A Monte Carlo Investigation". En: *Journal of Business and Economic Statistics*, 7(2), pp. 147-159.

SIMS, C.; J. STOCK y M. W. WATSON

1990 "Inference in Linear Time Series Models with Some Unit Roots". En: *Econometrica*, 56(6), pp. 1397-1417.

STOCK, J. y M. W. WATSON

1993 "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems". En: *Econometrica*, 61(4), pp. 783-820.

ULLAH, A.

2004 *Finite Sample Econometrics*. Advanced Texts in Econometrics. Oxford: Oxford University Press.

WEST, K. D.

1988 "Asymptotic Normality, when Regressors Have a Unit Root". En: *Econometrica*, 56(6), pp. 1397-1417.

SE TERMINÓ DE IMPRIMIR EN LOS TALLERES GRÁFICOS DE
TAREA ASOCIACIÓN GRÁFICA EDUCATIVA
PASAJE MARÍA AUXILIADORA 156 - BREÑA
CORREO E.: tareagrafica@tareagrafica.com
PÁGINA WEB: www.tareagrafica.com
TELÉF. 332-3229 FAX: 424-1582
MARZO 2016 LIMA - PERÚ