

**ERRORES DE PREDICCIÓN Y RAÍCES
UNITARIAS EN SERIES TEMPORALES
UNIVARIANTES**

TESIS DOCTORAL

Autor: Ismael Sánchez Rodríguez-Morcillo

Director: Daniel Peña Sánchez de Rivera

UNIVERSIDAD CARLOS III DE MADRID

Departamento de Estadística y Econometría

Getafe, junio de 1996



tu/56

D. ISMAEL SÁNCHEZ RODRÍGUEZ-MORCILLO
con D.N.I.: 3.823.990

A U T O R I Z A :

Que su tesis doctoral con el título
"ERRORES DE PREDICCIÓN EN SERIES
TEMPORALES UNIVARIANTES"
pueda ser utilizada para fines de investigación por parte
de la Universidad Carlos III de Madrid.

Leganés, 17 de Septiembre de 1.996.

A handwritten signature in black ink, appearing to be "Ismael Sánchez Rodríguez-Morcillo".

Fdo.: Ismael Sánchez Rodríguez-Morcillo

R.1136278 .

TU/56

Universidad Carlos III de Madrid



1136278

ERRORES DE PREDICCIÓN Y RAÍCES UNITARIAS EN SERIES TEMPORALES UNIVARIANTES

TESIS DOCTORAL

Autor: Ismael Sánchez Rodríguez-Morcillo

Director: Daniel Peña Sánchez de Rivera



DONACION

UNIVERSIDAD CARLOS III DE MADRID

Departamento de Estadística y Econometría

Getafe, junio de 1996

A mis padres.

Índice

1	Introducción	6
1.1	Sumario	6
1.2	Raíces unitarias y el largo plazo en series temporales	11
1.2.1	Conceptos básicos	11
1.2.2	Raíces unitarias y cointegración	13
1.2.3	Raíces unitarias y predicción	14
1.2.4	Raíces unitarias y modelización	18
2	Sobrediferenciación de procesos autorregresivos casi no estacionarios	21
2.1	Introducción	21
2.2	Modelo y notación	25
2.3	Sobrediferenciación de procesos autorregresivos casi no estacionarios	28
2.3.1	Consideraciones generales	28
2.3.2	Consistencia de los estimadores	31
2.3.3	Sesgo y error cuadrático medio	34
2.4	ECMP a horizonte H	35
2.4.1	ECMP del predictor correcto AR(p+1)	36
2.4.2	ECMP del predictor sobrediferenciado ARIMA(p,1,0)	38
2.5	Comparación de la capacidad predictiva	40

2.6	Sobrediferenciación de procesos AR(1)	42
2.6.1	El caso AR(1)	42
2.6.2	El caso AR(1, μ)	44
2.7	Estudio de simulación	47
2.8	Conclusiones finales	51
	Apéndice A	54
A.1	Algunos lemas previos	54
A.2	Demostraciones de la sección 2.3	60
A.3	Demostraciones de la sección 2.4	62
A.4	Demostraciones de la sección 2.5	68
3	Contrastes de raíces unitarias en series temporales. Una revisión com- parada	72
3.1	Introducción	72
3.2	Series temporales no estacionarias	74
3.3	Introducción a la inferencia en presencia de raíz unitaria	78
3.4	Contrastes de raíces unitarias	83
3.4.1	Consideraciones generales	84
3.4.2	Contrastes de Dickey–Fuller	86
3.4.3	Contrastes de Sargan–Bhargava	92
3.4.4	Contrastes de Elliot, Rothenberg y Stock	94
3.4.5	Comparación en muestras finitas	97
3.4.6	Otros enfoques para contrastar raíces unitarias	111
4	Detección de raíces unitarias mediante los errores de predicción	120
4.1	Introducción	120
4.2	Errores de predicción y raíces unitarias	122

4.3	Contrastes basados en errores de predicción	124
4.3.1	Construcción de los estadísticos	124
4.3.2	Comparación con otros estadísticos de contraste	131
4.3.3	Distribuciones asintóticas	134
4.3.4	Valores críticos	139
4.3.5	Potencia en muestras finitas. Caso AR(1)	140
4.4	Extension al caso $\psi(B)$ general	151
4.4.1	Estimadores consistentes de ω^2	152
4.5	Contrastes modificados \overline{CS}	154
4.5.1	Estadísticos de contraste	154
4.5.2	Potencia en muestras finitas de \overline{CS}	155
4.6	Contrastes con predicciones eficientes. Contrastes ECS	164
4.6.1	Distribuciones asintóticas de los estadísticos ECS	168
4.6.2	Potencia en muestras finitas de ECS	174
4.7	Resumen y conclusiones	181
	Apéndice B	183
5	Conclusiones finales y extensiones	202
	Referencias	205

AGRADECIMIENTOS

Quiero agradecer en primer lugar a mi director de tesis, el profesor Daniel Peña, por toda la ayuda que me ha prestado, tanto en la elaboración de esta tesis doctoral como en todo lo que constituye mi carrera académica. Su dedicación y su confianza en mí han sido fuente de estímulo permanente.

Quiero agradecer la ayuda de todos mis compañeros del departamento de Estadística y Econometría de la Carlos III de Madrid. Sus contribuciones a mi formación han sido muy importantes y su ejemplo ha sido una excelente escuela. Estoy agradecido, también, a los miembros del Laboratorio de Estadística de la ETSII de la Universidad Politécnica de Madrid por la ayuda que me prestaron durante los años que compartí con ellos. Agradezco a Andrew Harvey su acogida durante mi estancia en la London School of Economics. Finalmente estoy agradecido a la financiación proporcionada por el programa de becas FPI.

Getafe, junio de 1996.

Capítulo 1

Introducción

1.1 Sumario

En esta tesis doctoral se desarrollan dos aspectos relacionados con la detección de raíces unitarias en series temporales univariantes. En primer lugar se analizan las consecuencias en predicción de considerar que existe raíz unitaria cuando el proceso es estacionario. En segundo lugar se proponen nuevos contrastes de raíces unitarias basados en errores de predicción multihorizonte.

Estas dos líneas de investigación se fundamentan en la estrecha relación existente entre las raíces unitarias y la evolución de una serie en el largo plazo. Es en el largo plazo donde más diferencias existen entre una serie estacionaria y otra con raíz unitaria. Por tanto, una detección incorrecta de una raíz unitaria tendrá consecuencias más graves a largo plazo que a corto plazo. Asimismo, el largo plazo de una serie contiene información relevante que permita detectar la presencia de raíces unitarias.

Las raíces unitarias han captado una gran atención de los investigadores. La mayor parte de la voluminosa literatura existente se ha centrado en el problema de su detección. En

toda esta literatura se remarca la dificultad de realizar una detección correcta cuando el proceso es estacionario pero tiene una raíz próxima al círculo unidad. A pesar de esta dificultad, son escasas las referencias donde se investigue sobre las consecuencias de una detección incorrecta. Asimismo, a pesar de ser en el largo plazo donde más se advierten las consecuencias de una raíz unitaria no existe ningún procedimiento de detección que esté basado en las predicciones a largo plazo.

El capítulo 2 se dedica a estudiar las consecuencias de una detección errónea de una raíz unitaria. Se demuestra en este capítulo que:

- Si un proceso autorregresivo estacionario tiene una raíz próxima al círculo unidad el modelo sobrediferenciado genera predicciones con menor error cuadrático medio de predicción que el modelo estacionario correctamente especificado.

La justificación de este resultado se encuentra en la mayor parsimonia del modelo sobrediferenciado. La proximidad de la raíz al círculo unidad ha de considerarse en función del tamaño muestral. Cuanto menor sea el tamaño muestral mayor es la vecindad del círculo unitario en la que un predictor sobrediferenciado proporciona menor error cuadrático medio. Este resultado se complementa con el ya conocido de que tomar una diferencia en un predictor con raíz unitaria conduce también a un menor error cuadrático medio de predicción. Se obtienen entonces las dos siguientes conclusiones

- Para predecir es mejor sobrediferenciar que infradiferenciar.
- La baja potencia de los contrastes de raíces unitarias en la proximidad al círculo unidad no constituye un problema si el objetivo es la predicción.

Estas dos conclusiones han sido sugeridas por numerosos analistas e investigadores basándose en su experiencia empírica. Sin embargo no habían sido demostradas.

Otro resultado que se obtiene en este segundo capítulo es:

- La superioridad del predictor sobrediferenciado puede ser pequeña en el corto plazo, pero se incrementa con el horizonte de predicción.

Este resultado advierte que procedimientos de diagnóstico basados en errores de predicción a horizonte uno, o en la varianza residual, pueden conducir a suboptimalidad en horizontes mayores.

Si el objetivo de la modelización no es la predicción sino que existe interés en averiguar cuál es el modelo que generó las observaciones la conclusión que se obtiene es la siguiente:

- Seleccionar modelos por su comportamiento predictivo puede llevar a elegir un proceso no estacionario en lugar del proceso estacionario real si existe una raíz próxima a la unidad.

Esta conclusión es especialmente aplicable si el tamaño muestral no es elevado y la comparación de los modelos alternativos se realiza reservando parte de la muestra para evaluar las predicciones.

Los capítulos 3 y 4 se dedican al tema de detección de raíces unitarias. En el capítulo 3 se realiza una revisión de los contrastes existentes en la literatura. En el capítulo 4 se desarrollan nuevos contrastes basados en errores de predicción.

El número de contrastes que existen en la literatura es muy extenso, aunque estos contrastes admiten una clasificación en un número reducido de grupos. El grupo de contrastes más popular es el basado en las propiedades del estimador de la raíz unitaria. A este grupo pertenecen los contrastes de Dickey & Fuller (1979), que son los primeros que aparecen en la literatura y los más utilizados. También en este grupo se encuentran los contrastes de Elliot, Rothenberg & Stock (1992), que son los más potentes en el caso AR(1).

La construcción de un contraste de raíces unitarias requiere la solución de tres problemas

1. Encontrar un estadístico que mida el alejamiento al círculo unitario.
2. Eliminar la influencia del componente determinista.
3. Eliminar la influencia del resto de las raíces.

La mayoría del esfuerzo investigador se ha concentrado en la solución del primer punto, subestimando la importancia de los otros dos. Respecto a la influencia del componente determinista basta mencionar que los contrastes de Elliot, Rothenberg & Stock (1992) utilizan los mismos estadísticos de contraste que Dickey & Fuller (1979), pero modifican la forma de eliminar el componente determinista. Esta modificación propuesta permite duplicar la potencia de los contrastes originales de Dickey & Fuller. La solución que se adopte para eliminar la influencia de las otras raíces también tiene un efecto considerable en las propiedades finales del contraste. Sin embargo los dos procedimientos más empleados han demostrado ser inadecuados en la práctica. Estos procedimientos son el método de Dickey-Fuller aumentado (Said & Dickey (1984)) y el método de Phillips-Perron (Phillips (1987a), Phillips & Perron (1988)). Estos procedimientos son inapropiados cuando existe un componente de media móvil con raíces positivas. Este tipo de componentes son, sin embargo, los más frecuentes en las series reales. El resultado es que, si la serie presenta componente de media móvil de raíz positiva, el nivel de significación del contraste puede incrementarse excesivamente e invalidar el resultado del contraste.

En el capítulo 4 se proponen tres tipos de contrastes basados en errores de predicción. El predictor empleado se obtiene de la modelización de la serie suponiendo que existe raíz unitaria. Una vez estimado el predictor se evalúan los errores cuadráticos de predicción multihorizonte dentro de la muestra. El estadístico de contraste se construye dividiendo

la suma de los errores cuadráticos obtenidos por su valor teórico esperado.

Los tres tipos de contrastes considerados se diferencian en el número de errores cuadráticos empleado. Brevemente estas diferencias se pueden describir de la siguiente forma:

1. En el primer tipo se consideran los errores cuadráticos de predecir desde la primera observación disponible el resto de la muestra.
2. En el segundo tipo se añaden, además, los errores cuadráticos de predecir la última observación desde cada elemento de la muestra.
3. En el tercer tipo se utilizan todos los errores cuadráticos que puedan obtenerse, promediándose los correspondientes a un mismo horizonte de predicción.

El primero de los contrastes mencionado es el más potente, si bien el segundo presenta resultados similares. Para cada tipo de estadístico se construyen versiones tanto para contrastes cuya alternativa sea estacionaria de media no nula como para alternativa con pendiente lineal determinista. Se obtiene, además, la distribución asintótica de cada estadístico. Los valores críticos para diversos tamaños muestrales se obtienen mediante simulación por Monte Carlo.

Para el caso sencillo de un $AR(1)$ se obtienen resultados de potencia similares a los contrastes de Elliot, Rothenberg & Stock (1992) salvo en zonas alejadas del círculo unitario. En el caso de un proceso más general se consiguen resultados mejores a Elliot, Rothenberg & Stock (1992). Esta ventaja es especialmente importante en procesos con componente de media móvil que tengan raíz positiva. En estas circunstancias los contrastes propuestos en esta tesis son estables en su nivel de significación mientras que los contrastes de Elliot, Rothenberg & Stock (1992) presentan distorsiones demasiado elevadas.

1.2 Raíces unitarias y el largo plazo en series temporales

En esta sección se introduce la importancia de las raíces unitarias y la relación que existe entre la presencia de una raíz unitaria y el comportamiento de una serie en el largo plazo. En primer lugar se presenta la definición de raíz unitaria y su correspondiente notación. Posteriormente se relacionará la presencia de una raíz unitaria con tres aspectos fundamentales de las series temporales: las relaciones de cointegración, la predicción y la modelización.

1.2.1 Conceptos básicos

Sea y_t una serie temporal que sigue un modelo ARMA(p, q),

$$\varphi(B)(y_t - \mu) = \theta(B)a_t, \quad (1.2.1)$$

donde μ es la media del proceso, B es el operador de retardo tal que $By_t = y_{t-1}$, $\varphi(B)$ y $\theta(B)$ son los polinomios

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

y a_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas (iid) que verifican $E(a_t) = 0$ y $E(a_t^2) = \sigma^2$.

La ecuación característica del polinomio autorregresivo es

$$1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p = 0. \quad (1.2.2)$$

y a sus raíces las denotará por $r_1^{-1}, r_2^{-1}, \dots, r_p^{-1}$. Si estas raíces están fuera del círculo unidad, $|r_i^{-1}| > 1$ $i = 1, 2, \dots, p$, el proceso es estacionario (en sentido débil), mientras que

si una o más raíces están en el círculo unidad, $|r_j^{-1}| = 1$ el proceso es no estacionario. Estas raíces unitarias pueden ser $+1$, -1 o complejas con módulo unidad. El interés en series reales reside fundamentalmente en la presencia de raíces $+1$. Típicamente se dice que una serie tiene una raíz unitaria, o que es $I(1)$, cuando su polinomio autorregresivo tiene una raíz de valor $r_j^{-1} = +1$ y el resto están fuera del círculo unidad.

En general, dadas las raíces $r_1^{-1}, r_2^{-1}, \dots, r_p^{-1}$, el polinomio autorregresivo puede factorizarse como

$$\varphi(B) = \Pi(1 - r_i B),$$

entonces si dicho proceso tiene d raíces de valor $+1$ puede escribirse como

$$\phi(B)(1 - B)^d(y_t - \mu) = \theta(B)a_t$$

y se dice que y_t sigue un modelo $ARIMA(p - d, d, q)$. El proceso

$$z_t = (1 - B)^d(y_t - \mu)$$

será, entonces, estacionario (en sentido débil). Se denomina $I(d)$ a un proceso no estacionario que se convierte en estacionario, o $I(0)$, al aplicarle d veces el operador diferencia $(1 - B)$.

La importancia de las raíces unitarias en series temporales proviene de sus implicaciones en los tres aspectos siguientes

1. Relaciones de cointegración
2. Predicción
3. Modelización

A continuación se comentará brevemente estos aspectos.

1.2.2 Raíces unitarias y cointegración

Un vector de series Y_t de dimensión $(k \times 1)$, donde cada serie es $I(1)$, se dice que está cointegrado si existe alguna combinación lineal $a'Y_t$ que es estacionaria ($I(0)$). La combinación $a'Y_t$ recibe el nombre de relación de cointegración y su existencia implica que entre dichas series existe una relación de equilibrio en el largo plazo.

El concepto de cointegración aparece por primera vez en Box & Tiao (1977) y fue desarrollado posteriormente por Granger (1981,1986) y Engle & Granger (1987). Estimar una relación de cointegración entre variables económicas permite contrastar empíricamente la teoría económica sobre su equilibrio a largo plazo. Se ha de resaltar que este equilibrio entre variables económicas es compatible con discrepancias en su evolución a corto plazo.

La figura 1.1 muestra un ejemplo de dos series cointegradas obtenidas mediante simulación. Estas series tienen la siguiente relación

$$\begin{aligned}x_t &= x_{t-1} + a_{1,t} - 0.5a_{1,t-1} \\y_t &= 0.7x_t + a_{2,t} - 0.8a_{2,t-1}.\end{aligned}$$

Puede apreciarse que individualmente no son estacionarias pero siguen sendas paralelas. Entre ambas series pueden existir diferencias puntuales, pero en el largo plazo existe un equilibrio entre las mismas.

Detectar relaciones de cointegración es, por tanto, de indudable interés y es un problema íntimamente ligado al de detección de raíces unitarias. Siguiendo la definición expresada anteriormente, la hipótesis de que *no* existe una relación de cointegración es equivalente a la hipótesis de que la combinación lineal $a'Y_t$ *sí* contiene una raíz unitaria.

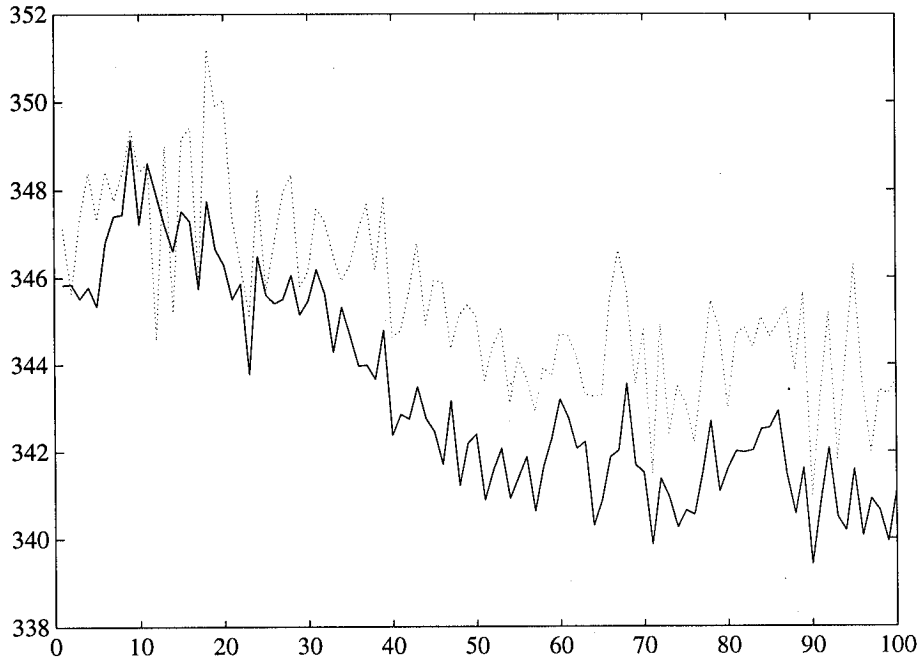


Figura 1.1: Dos series cointegradas

1.2.3 Raíces unitarias y predicción

La existencia de una raíz unitaria tiene importantes implicaciones sobre las predicciones en el largo plazo. Sea y_t la serie temporal observada que sigue el modelo ARMA(p, q) (1.2.1). La predicción que minimiza el error cuadrático medio de predicción (ECMP) es su esperanza condicionada a la información hasta $t = T$. Si se denomina $\hat{y}_{T+i} = E(y_{T+i}|y_T, y_{T-1}, \dots)$ y $\hat{a}_{T+i} = E(a_{T+i}|y_T, y_{T-1}, \dots)$ la predicción \hat{y}_{T+h} será

$$\hat{y}_{T+h} = \mu + \varphi_1 \hat{y}_{T+h-1} + \dots + \varphi_p \hat{y}_{T+h-p} + \hat{a}_{T+h} - \theta_1 \hat{a}_{T+h-1} - \dots - \theta_q \hat{a}_{T+h-q}.$$

Como $\hat{a}_{T+h-i} = 0$ para $i > h$ es bien conocido que las predicciones, después de q periodos, sólo dependen de la parte autorregresiva y por tanto satisfacen la ecuación en diferencias

$$\varphi(B)(\hat{y}_{T+j} - \mu) = 0 \quad j > q$$

donde B actúa sobre j . La solución de esta ecuación, si el origen es $t = T$ y las raíces de la ecuación característica r_i^{-1} , $i = 1, 2, \dots, p$ son distintas es

$$\hat{y}_{T+h} = \mu + A_1 r_1^h + A_2 r_2^h + \dots + A_p r_p^h, \quad (1.2.3)$$

donde los parámetros A_i son constantes que se determinan con las condiciones iniciales ($A_i \equiv A_i^{(T)}$).

La expresión (1.2.3) revela que, si todas las raíces están fuera del círculo unidad ($|r_i| < 1$), la predicción tiende al valor medio

$$\lim_{h \rightarrow \infty} \hat{y}_{T+h} = \mu$$

mientras que si una raíz i -ésima es unitaria

$$\lim_{h \rightarrow \infty} \hat{y}_{T+h} = \mu + A_i^{(T)}.$$

Como A_i depende de las condiciones iniciales y_T, y_{T-1}, \dots dicha predicción cambia con el origen de predicción. De esta forma un proceso estacionario presenta una inercia a retornar a la media mientras que en presencia de una raíz unitaria la predicción depende del origen seleccionado.

Una forma equivalente de observar esta diferencia en el comportamiento del largo plazo es a través de la persistencia de las innovaciones a_t . Si reescribimos el proceso ARMA(p, q) en forma de polinomio de medias móviles

$$(y_{T+h} - \mu) = \psi(B)a_{T+h} = a_{T+h} + \psi_1 a_{T+h-1} + \psi_2 a_{T+h-2} + \dots \quad (1.2.4)$$

puede observarse cómo la observación y_{T+h} es el resultado de la acción de las innovaciones pasadas, ponderadas por los pesos ψ_j . La influencia de la innovación a_T en y_{T+h} viene

determinada por el coeficiente ψ_h . Los pesos ψ_j pueden relacionarse con los parámetros del modelo mediante la igualación de términos de igual potencia en

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 + \psi_1 B + \dots) = (1 - \theta_1 B - \dots - \theta_q B^q).$$

Se observa que, para $j > \max(p, q)$, los pesos ψ_j cumplen también la ecuación en diferencias

$$\varphi(B)\psi_j = 0,$$

cuya solución es

$$\psi_j = C_1 r_1^j + C_2 r_2^j + \dots + C_p r_p^j. \quad (1.2.5)$$

La expresión (1.2.5) revela que si el proceso es estacionario los pesos ψ_j decaen exponencialmente y por tanto

$$\lim_{h \rightarrow \infty} \psi_h = 0.$$

Por tanto la innovación a_T tiene un efecto cada vez menor en las observaciones futuras hasta que éste acaba siendo nulo. El efecto de las innovaciones es pues transitorio. Sin embargo si existe una raíz unitaria

$$\lim_{h \rightarrow \infty} \psi_h = C_1 \neq 0,$$

y cada innovación produce una modificación permanente en la trayectoria de la serie.

La figura 1.2 ilustra esta idea mostrando los pesos ψ_h del proceso con raíz unitaria

$$(1 - 0.4B)(1 - B)y_t = (1 + 0.4B)a_t$$

y los del proceso estacionario

$$(1 - 0.4B)(1 - 0.6B)y_t = (1 + 0.4B)a_t.$$

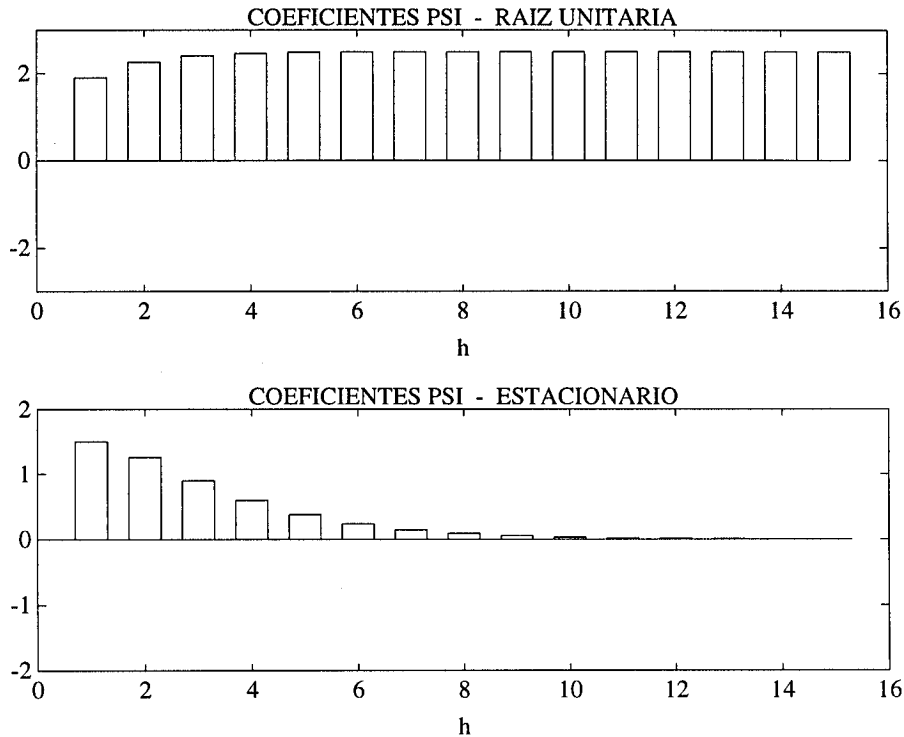


Figura 1.2: Coeficientes ψ_h de un proceso con raíz unitaria y estacionario

Comparando ambas figuras se aprecia que en el proceso con raíz unitaria las innovaciones tienen una influencia permanente en cualquier observación futura mientras que en un proceso estacionario esta influencia decae exponencialmente.

La influencia de las raíces unitarias en el largo plazo puede también observarse en la incertidumbre de las predicciones. En los procesos estacionarios el ECMP crece con el horizonte de predicción hasta que se alcanza una cota, a partir de dicha cota el ECMP permanece constante. En los procesos con una raíz unitaria el ECMP crece siempre de manera monótona y tiende a infinito con el horizonte de predicción.

Es bien conocido que el ECMP de un proceso ARMA puede escribirse en función de los pesos ψ_j de la siguiente forma

$$ECMP(\hat{y}_{t+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2,$$

donde, como se mostró en (1.2.5), los coeficientes ψ_j siguen, después de ciertos valores iniciales, la misma ecuación en diferencias que las raíces del polinomio característico autorregresivo. Por tanto, si el proceso es estacionario estos pesos tienden a cero, de manera que el ECMP tendrá un crecimiento acotado. Si el proceso es un MA(q) puro dicha cota será alcanzada después de q , periodos mientras que si tiene parte autorregresiva dicho límite superior se alcanzará gradualmente. Además puede comprobarse que

$$\lim_{h \rightarrow \infty} ECMP(\hat{y}_{t+h}) = \text{VAR}(y_t).$$

Sin embargo, si el proceso contiene una raíz unitaria los coeficientes ψ_j tienden a una constante no nula, que dependerá de los parámetros del modelo. Por tanto, a partir de cierto horizonte

$$ECMP(\hat{y}_{t+h}) = ECMP(\hat{y}_{t+(h-1)}) + \sigma^2 C_1^2$$

y el ECMP tendrá un crecimiento lineal. La figura 1.3 muestra distintas curvas del ECMP teórico de un AR(1) para distintos valores de su raíz. La diagonal representa el caso de raíz unitaria, donde se verifica que $ECMP(\hat{y}_{t+h}) = \sigma^2 h$. Las restantes curvas son, de arriba a abajo los casos de $\varphi = 0.9$ y 0.7 respectivamente.

1.2.4 Raíces unitarias y modelización

La modelización de una serie temporal con raíz unitaria se realiza habitualmente diferenciando la serie. Diferenciar consiste en aplicar el operador diferencia $(1 - B)$ a una serie

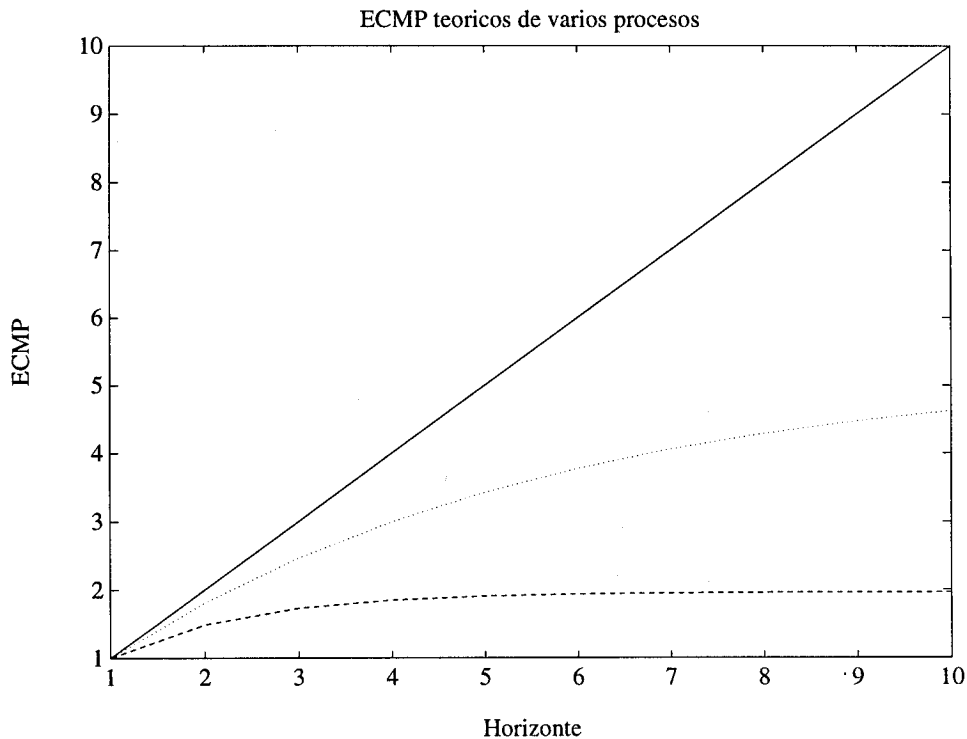


Figura 1.3: Curvas teóricas de ECMP. De abajo a arriba: $\varphi = 1$, $\varphi = 0.9$ y $\varphi = 0.7$.

temporal. Si dicha serie temporal es no estacionaria homogénea la serie resultante será estacionaria. Conseguir esta representación estacionaria ofrece ventajas en los tres puntos siguientes:

1. Identificación
2. Estimación
3. Eficiencia en predicción

Es muy frecuente la utilización de los correlogramas muestrales para identificar el modelo ARMA que sigue la serie. Sin embargo, cuando la serie contiene una raíz unitaria ésta domina la información que puede obtenerse de dichos correlogramas, siendo de escasa información sobre el resto de las raíces. Por tanto, en presencia de una raíz unitaria, un correlograma de la serie diferenciada proporciona una indicación más clara sobre la

estructura del proceso.

En la estimación, la incorporación de la información de la raíz unitaria mediante la diferenciación tiene dos ventajas. En primer lugar permite utilizar métodos de estimación que sólo son aplicables a series estacionarias. En segundo lugar se consigue una mayor eficiencia en la estimación del resto de los parámetros, debido a que diferenciar equivale a imponer una restricción cierta.

Esta misma idea de buscar eficiencia mediante el uso óptimo de la información puede extenderse al caso de predicción. Por tanto, si existen raíces unitarias, es más eficiente construir un predictor que incorpore ésta información que otro predictor que estime todos sus parámetros.

Capítulo 2

Sobrediferenciación de procesos autorregresivos casi no estacionarios

2.1 Introducción

En este capítulo se investigan las consecuencias en la estimación y en la predicción de sobrediferenciar un proceso $AR(p + 1)$ estacionario que tiene una raíz próxima al círculo unitario.

La diferenciación, como ya se mencionó en el anterior capítulo, se utiliza normalmente para transformar una serie temporal no estacionaria homogénea en un proceso estacionario. Dicho proceso estacionario puede ser modelizado como un modelo $ARMA(p, q)$. Se considera, entonces, que la serie original sigue un proceso $ARIMA(p, d, q)$. El valor d es el número de veces que ha sido necesario diferenciar para hacer la serie estacionaria y es igual al número de raíces unitarias de la ecuación característica. Cuando la ecuación característica de un proceso autorregresivo estacionario tiene una raíz muy próxima a la unidad se dice que dicho proceso es *casi no estacionario* (near nonstationary) o casi integrado (near integrated). Dada una muestra moderada de este proceso será muy probable concluir, dada la baja potencia de los contrastes de raíces unitarias en estos casos, que el

proceso es no estacionario. Por tanto, se aplicará el operador diferencia cuando realmente la serie es estacionaria. A la serie así diferenciada se le denominará *sobrediferenciada*.

Desde los trabajos de Fuller (1976) y Dickey & Fuller (1979) ha habido una extensa literatura sobre detección de raíces unitarias en polinomios autorregresivos. Asimismo, recientemente se ha prestado mucha atención a los contrastes de raíces unitarias en el polinomio de medias móviles (Tanaka 1990, Saikkonen & Luukkonen 1993, Tsay 1993). Sin embargo se ha escrito muy poco sobre las consecuencias de una detección incorrecta.

Las curvas de potencia de los contrastes de raíces unitarias indican que un proceso casi no estacionario será considerado no estacionario, y por tanto será sobrediferenciado, en un porcentaje muy alto de ocasiones. El contraste más popular y más utilizado en la práctica es debido a Dickey & Fuller (1979) y está basado en el estadístico pivote de regresión. Dicho contraste tiene una potencia de 0.10 para una alternativa de $\rho = 0.95$ y un tamaño muestral de $T = 100$ e la misma potencia con $\rho = 0.90$ y $T = 50$. Por tanto, si en dichas situaciones se sobrediferencia en el 90% de las ocasiones, resulta de interés conocer las consecuencias de dicha decisión incorrecta. Trabajos previos sobre las consecuencias de la sobrediferenciación se encuentran en Plosser & Schwert (1977, 1978), Harvey (1981) y Campbell & Perron (1991). Plosser & Schwert (1977) examinan, mediante simulación por Monte Carlo, el efecto de la sobrediferenciación en dos situaciones: procesos con tendencia lineal determinista y modelos de regresión estocásticos. Estos autores concluyen que, en estas situaciones, la pérdida de eficiencia tanto en la estimación de los parámetros como en la predicción no es importante si se incluye un término MA(1) en el modelo. Harvey (1981) propone un predictor para muestras finitas, basado en el filtro de Kalman, que es útil para la obtención de predicciones óptimas en procesos sobrediferenciados. Este autor concluye igualmente que la sobrediferenciación no afecta gravemente a las predicciones si se incluye

un parámetro MA y se usa el predictor propuesto. Campbell & Perron (1991) realizan un ejercicio de simulación en el que comparan el ECMP de un predictor estimado en niveles con el de un predictor estimado en diferencias en un proceso ARMA(1,1) estacionario. Ambos predictores se construyen mediante el ajuste de un modelo autorregresivo. El orden del polinomio autorregresivo se selecciona de manera que todos los retardos sean significativos (usando la teoría asintótica normal estandar). Estos autores encuentran que el predictor sobrediferenciado obtiene menor ECMP si el proceso se encuentra próximo a la no estacionariedad. El resultado de Campbell & Perron (1991) puede considerarse como un antecedente empírico del resultado que se demuestra en este capítulo.

Las consecuencias de los errores de especificación en el ECMP ha recibido mucho interés en la literatura (Berk 1974, Bhansali 1978, 1981, Davies & Newbold 1980, Tanaka & Maekawa 1984, Kunitomo & Yamamoto 1985 entre otros). Kunitomo & Yamamoto (1985) investigan las consecuencias en el ECMP de estimar un modelo autorregresivo de orden p a un proceso autorregresivo de orden m (m puede ser infinito). A diferencia del enfoque que se aplica en este capítulo, los trabajos de los autores arriba mencionados suponen que tanto el modelo correcto como el modelo mal especificado tienen el mismo orden de integración.

En este capítulo se asume que una raíz del polinomio autorregresivo está cercana a la unidad y que, por tanto, el modelo sobrediferenciado es ARIMA($p, 1, 0$), sin componente de medias móviles. Se analizará las propiedades de los estimadores de este modelo ARIMA($p, 1, 0$) y se comparará su ECMP respecto al modelo correctamente especificado AR($p + 1$). Se demuestra que el ECMP del modelo sobrediferenciado ARIMA($p, 1, 0$) es menor que el ECMP del modelo correcto AR($p + 1$) si la raíz más próxima a la unidad verifica $\rho = \exp(-c/T^\beta)$; $\beta > 1$. La causa de esta ventaja se encuentra en la mayor

parsimonia de la representación $ARIMA(p, 1, 0)$.

Por otra parte, es conocido que cuando el proceso tiene una raíz unitaria se consiguen predicciones más eficientes si se aplica el operador diferencia. Por tanto el resultado que se demuestra en este capítulo se complementa con este resultado clásico. De esta forma tanto si se tiene raíz unitaria como si se está próximo a ella diferenciar lleva a mejorar las predicciones.

Algunas consecuencias que se extraen de los resultados de este capítulo son:

1. A efectos de predicción es mejor sobrediferenciar que infradiferenciar.
2. La baja potencia de los contrastes de raíces unitarias no es tan importante en predicción como lo es en la identificación del modelo correcto.
3. La superioridad del predictor sobrediferenciado puede ser pequeña en el corto plazo, pero se incrementa con el horizonte de predicción.
4. Seleccionar modelos por su comportamiento predictivo puede llevar a elegir un modelo estacionario en lugar del modelo no estacionario real.

La estructura de este capítulo es la siguiente. La sección 2.2 introduce el modelo y la notación. Las consecuencias de la sobrediferenciación en la estimación se analizan en la sección 2.3 y las consecuencias en el ECMP en la sección 2.4. La sección 2.5 compara el ECMP del predictor correctamente especificado y el predictor sobrediferenciado y demuestra la ventaja del predictor sobrediferenciado. La sección 2.6 estudia el caso en que los datos se generan con un $AR(1)$ y se utiliza el paseo aleatorio como predictor alternativo. En la sección 2.7 se realiza un ejercicio de simulación que apoya e ilustra los resultados teóricos.

2.2 Modelo y notación

Sea $\{y_t\}$ una serie temporal discreta que sigue el siguiente modelo estacionario AR($p+1$)

$$\varphi(B)y_t = \alpha + a_t, \quad (2.2.1)$$

donde B es el operador de retardo; $\varphi(B) = (1 - \sum_{i=1}^{p+1} \varphi_i B^i)$ es un polinomio en el operador de retardo tal que $\varphi(B) = 0$ con todas sus raíces fuera del círculo unidad. Esta condición es equivalente a decir que las raíces del polinomio

$$z^{p+1} - \varphi_1 z^p - \dots - \varphi_{p+1} = 0$$

están dentro del círculo unidad. El término a_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas (iid) de media cero y varianza σ^2 . Sobre los momentos de orden superior de a_t se realiza la siguiente hipótesis

A1. Para algún $s_0 > 2$, $E\{|a_t|^{s_0}\} < \infty$.

Se denota por ρ^{-1} a la raíz de $\varphi(B) = 0$ más próxima al círculo unidad. El polinomio autorregresivo puede factorizarse como $\varphi(B) = \phi(B)(1 - \rho B)$, donde $\phi(B) = 1 - \sum_{j=1}^p \phi_j B^j$ y $\varphi_i = \phi_i - \rho \phi_{i-1}$, con $\phi_0 = -1$ y $\phi_{p+1} = 0$. Por conveniencia se expresará este modelo en forma de un vector autorregresivo de primer orden de la siguiente manera

$$Y_t = A_\alpha Y_{t-1} + U_{t,p+2}, \quad (2.2.2)$$

con $Y_t = (y_t, \dots, y_{t-p}, 1)'$, $U_{t,p+2} = (a_t, 0, \dots, 0)'$, donde el subíndice $(p+2)$ indica la di-

mensión del vector y

$$A_\alpha = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_p & \varphi_{p+1} & \alpha \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}.$$

Entonces $y_t = e'_{p+2} Y_t$ con $e_{p+2} = (1, 0, \dots, 0)'$. Se denota por $\Gamma_y = E(Y_t Y_t')$ y $\gamma_y = E(Y_t y_{t+1})$. Este proceso admite una representación en desviaciones respecto a la media de la siguiente forma

$$\tilde{Y}_t = A_o \tilde{Y}_{t-1} + U_{t,p+1}, \quad (2.2.3)$$

donde $\tilde{Y}_t = (\tilde{y}_t, \tilde{y}_{t-1}, \dots, \tilde{y}_{t-p})'$, $\tilde{y}_t = y_t - \mu$; $\mu = E(y_t) = \varphi(1)\alpha$; y

$$A_o = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_p & \varphi_{p+1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Se denota igualmente $\Gamma_{\tilde{y}} = E(\tilde{Y}_t \tilde{Y}_t')$. Si se aplica el operador diferencia a la serie y_t se obtiene $w_t = (1 - B)y_t$, que seguirá el modelo

$$\phi(B)(1 - \rho B)w_t = (1 - B)a_t, \quad (2.2.4)$$

que es no invertible. Este proceso w_t admite también una representación en forma de proceso vectorial autorregresivo de primer orden (Lütkepohl 1991, p. 223)

$$Z_t = A_1 Z_{t-1} + U_{t,p+1}^*, \quad (2.2.5)$$

con $Z_t = (W_t', a_t)'$, $W_t = (w_t, \dots, w_{t-p})'$, $U_{t,p+1}^* = (a_t, 0, \dots, 0, a_t)'$, y

$$A_1 = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_p & \varphi_{p+1} & -1 \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

con $w_t = e'_{p+1} Z_t$. Sea $\Gamma_w = E(W_t W_t')$ y $\gamma_w = E(W_t w_{t+1})$. En la presente exposición se empleará el símbolo $\hat{\phi}$ para hacer referencia a estimaciones procedentes de una muestra del proceso sobrediferenciado $\{w_t\}$ y el símbolo $\check{\phi}$ para estimaciones en una muestra del proceso original $\{y_t\}$. Las estimaciones del presente capítulo se realizan por el método de mínimos cuadrados. El estimador de los parámetros del modelo AR($p+1$), $\varphi = (\varphi_1, \dots, \varphi_{p+1}, \alpha)'$, utilizando una muestra de tamaño T del proceso original (2.2.1) es

$$\check{\phi} = \check{\Gamma}_y^{-1} \check{\gamma}_y, \quad (2.2.6)$$

donde $\check{\Gamma}_y = (T-p-1)^{-1} \sum_{j=p+1}^{T-1} Y_j Y_j'$, $\check{\gamma}_y = (T-p-1)^{-1} \sum_{j=p+1}^{T-1} Y_j y_{j+1}$. Análogamente el estimador de los parámetros $\phi = (\phi_1, \dots, \phi_p)'$ del modelo incorrecto AR(p) en una muestra de tamaño $T-1$ ($t = 2, 3, \dots, T$) del proceso sobrediferenciado (2.2.4) es

$$\hat{\phi} = \hat{\Gamma}_w^{-1} \hat{\gamma}_w, \quad (2.2.7)$$

donde $\hat{\Gamma}_w = (T-p-1)^{-1} \sum_{j=p+1}^{T-1} W_j W_j'$, $\hat{\gamma}_w = (T-p-1)^{-1} \sum_{j=p+1}^{T-1} W_j w_{j+1}$. Se añadirán, además, las siguientes hipótesis:

A2. $E\{\|\check{\Gamma}_y^{-1}\|^{2k}\}$ ($k = 1, 2, \dots, k_0$) está acotada para $T > T_0$ y algún k_0 .

A3. $E\{\|\hat{\Gamma}_w^{-1}\|^{2k}\}$ ($k = 1, 2, \dots, k_0$) está ecotada para $T > T_0$ y algún k_0 .

Estas hipótesis A2 y A3 son similares a la hipótesis A3 de Kunitomo & Yamamoto (1985) y ambas se cumplen si a_t es normal.

2.3 Sobrediferenciación de procesos autorregresivos casi no estacionarios

2.3.1 Consideraciones generales

En esta sección se analizan las propiedades del estimador $\hat{\phi} = \hat{\Gamma}_w^{-1} \hat{\gamma}_w$ del modelo mal especificado ARIMA($p, 1, 0$) cuando el proceso es casi no estacionario. En general, una serie temporal se dice que es casi no estacionaria si, siendo estacionaria, tiene una raíz muy próxima a la unidad, de forma que puede confundirse con un proceso no estacionario. Esta confusión entre ambas modelizaciones alternativas tiene una clara vinculación con el tamaño de la muestra que se posee. A mayor tamaño muestral más próxima debe estar dicha raíz de la unidad para que dicha confusión se produzca. Análogamente, con menores tamaños muestrales será más fácil confundir un proceso estacionario con otro que no lo sea. Esta idea de proximidad a la unidad dependiente del tamaño muestral ha sido formalizada en la literatura mediante la reparametrización de dicha raíz ρ^{-1} como (Phillips 1987)

$$\rho = \exp\left(-\frac{c}{T}\right) = 1 - \frac{c}{T} + o(T^{-1}), \quad (2.3.1)$$

donde c es una constante fija y T es el tamaño muestral. La limitación de (2.3.1) para la presente exposición es que el ratio de convergencia a la unidad se fija en $O(T^{-1})$. La razón de esta definición es que dicho ratio es el orden de consistencia del estimador mínimo cuadrático. Sin embargo en este capítulo será necesario emplear una definición más general para ρ mediante

$$\rho = \exp\left(-\frac{c}{T^\beta}\right), \quad (2.3.2)$$

con c, β constantes fijas. Se limitará el estudio al caso $c > 0$, por lo que se supone que la raíz más próxima a la unidad se encuentra fuera del círculo unidad y se aproxima a él con el ratio de convergencia $O(T^{-\beta})$.

Sea $\pi(B)w_t = a_t$ la representación autorregresiva del proceso sobrediferenciado (2.2.4).

Los coeficientes de $\pi(B)$ seguirían la siguiente expresión

$$\pi_j = \begin{cases} \phi_j + (\rho - 1)(1 - \sum_{k=1}^{j-1} \phi_k) & \text{if } j \leq p \\ (\rho - 1)(1 - \sum_{k=1}^p \phi_k) & \text{if } j > p. \end{cases} \quad (2.3.3)$$

Si ρ verifica (2.3.2) con β suficientemente grande, el término $(\rho - 1)$ será pequeño ($O(T^{-\beta})$) comparado con la variabilidad muestral de los correlogramas estimados (el error estandar de los coeficientes de autocorrelación estimados es $O(T^{-\frac{1}{2}})$). Por tanto, aunque el proceso sobrediferenciado w_t sea no invertible, un correlograma medio de w_t sugerirá la estimación de un $AR(p)$. La figura 2.1 muestra el resultado de un ejercicio de simulación que ilustra este punto. En cada replicación de la simulación se han estimado la función de autocorrelación simple (*fas*) y parcial (*fap*) tanto de la serie original como de la diferencia para un tamaño muestral $T = 100$. El modelo simulado es $(1 - 0.5B)(1 - 0.95B)y_t = 10 + a_t$ donde a_t es un proceso iid que sigue $a_t \sim N(0, 1)$. El gráfico es el resultado de promediar los correlogramas de 5000 replicaciones. Esta figura muestra que las modelizaciones más plausibles son un $AR(2)$ o un $ARIMA(1, 1, 0)$.

El ajuste de un $AR(p)$ en lugar de un $ARMA(p + 1, 1)$ es equivalente a realizar un truncamiento de un autorregresivo de orden infinito con coeficientes (2.3.3). Berk (1974) y Banshali (1978) analizaron las consecuencias de truncar un polinomio autorregresivo de orden no necesariamente finito cuando el proceso es tanto invertible como estacionario. Estos autores encuentran el orden del truncamiento que permite ignorar el sesgo de dicha

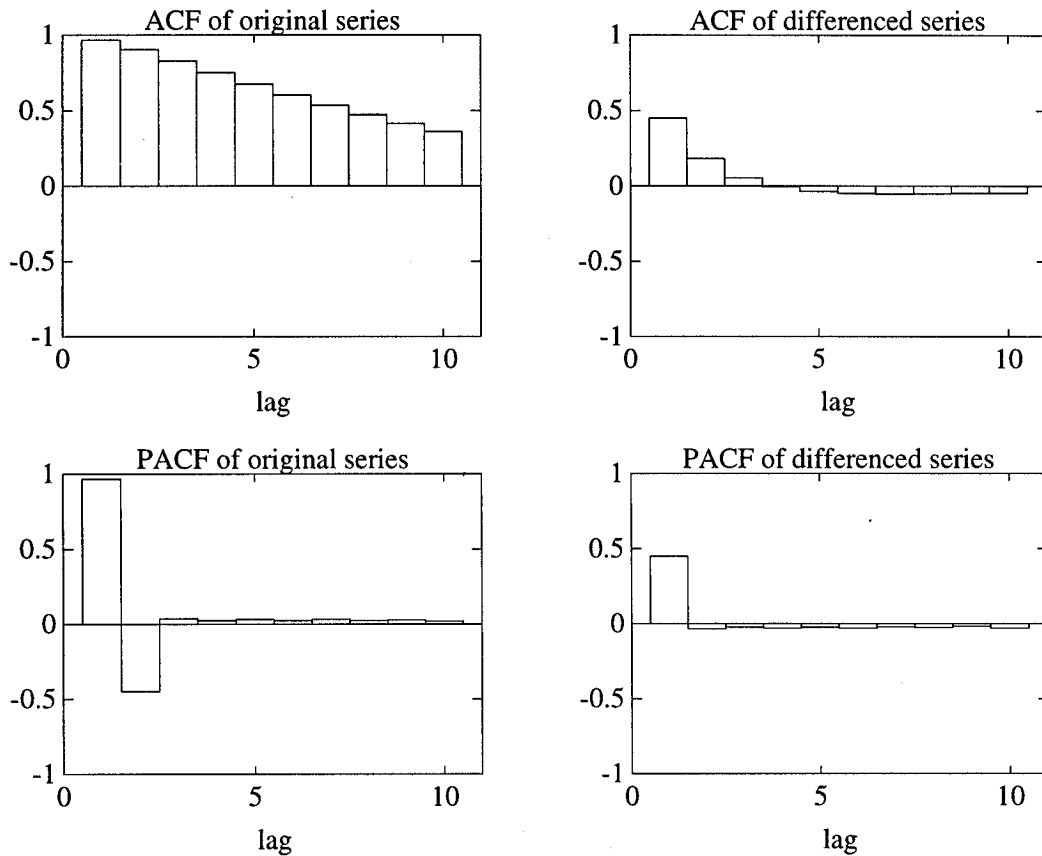


Figura 2.1: Correlogramas estimados del modelo $(1 - 0.5B)(1 - 0.95B)y_t = 10 + a_t$, a_t iid con distribución $N(0, 1)$, tamaño muestral $T = 100$. Promedio de 5000 replicaciones

simplificación. En la presente situación el proceso es no invertible y el truncamiento se realiza en una posición fija (p). Se estudia entonces qué propiedades debe cumplir el proceso para que el modelo propuesto sea estimado consistentemente y las predicciones realizadas con el mismo sean eficientes.

La expresión (2.3.3) muestra también la influencia de las restantes raíces. Si se denota por r_i^{-1} , $i = 1, \dots, p$ a las raíces del polinomio característico $\phi(B) = 0$ entonces, $\phi(B) =$

$\prod_{i=1}^p (1 - r_i B)$. Para $B = 1$ se puede escribir

$$(1 - \sum_{k=1}^p \phi_k) = \prod_{i=1}^p (1 - r_i). \quad (2.3.4)$$

Por tanto, aunque la diferencia entre π_j y ϕ_j depende principalmente de $(1 - \rho)$ existe también una influencia de las restantes raíces. Valores negativos de r_i aumentarán el valor de π_j , $j > p$ e incrementará el sesgo del truncamiento propuesto en $j = p$ en muestras pequeñas.

2.3.2 Consistencia de los estimadores

Sea $\{w_{t|p}\}$ el proceso límite de $\{w_t\}$ cuando $T \rightarrow \infty$ y por tanto $\rho \rightarrow 1$. Este proceso límite sigue un AR(p) puro cuya representación markoviana es

$$W_{t|p} = A_p W_{t-1|p} + U_{t,p}, \quad (2.3.5)$$

$$A_p = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

donde $W_{t|p} = (w_{t|p}, \dots, w_{t-p+1|p})'$. Se tiene entonces, de (2.2.4)

$$\begin{aligned} w_t &= \phi^{-1}(B)(1 - \rho B)^{-1}(1 - B)a_t \\ &= \phi^{-1}(B) [1 - (1 - \rho)(B + \rho B^2 + \cdots)] a_t \\ &= w_{t|p} - \sum_{j=0}^{\infty} \psi_j (1 - \rho) z_{t-1-j}, \end{aligned} \quad (2.3.6)$$

donde ψ_j ; $j = 0, 1, \dots$ son los coeficientes de $\phi^{-1}(B)$, y $(1 - \rho B)z_t = a_t$. Se denotará $\Gamma_{w|p} = E(W_{t|p}W'_{t|p})$ $\gamma_{w|p} = E(W_{t|p}w_{t+1|p})$. Se definen también las autocovarianzas muestrales como

$$\hat{\Gamma}_{w|p} = (T - p - 1)^{-1} \sum_{j=p+1}^{T-1} W_{j|p}W'_{j|p},$$

$$\hat{\gamma}_{w|p} = (T - p - 1)^{-1} \sum_{j=p+1}^{T-1} W_{j|p}w_{j+1|p}.$$

Se necesita también la siguiente hipótesis:

A3'. $E\{\|\hat{\Gamma}_{w|p}^{-1}\|^{2k}\}$ ($k = 1, 2, \dots, k_0$) está acotada para $T > T_0$ y algún k_0 .

Como los elementos de $\hat{\Gamma}_w$ y $\hat{\gamma}_w$ son autocovarianzas muestrales se obtiene que

$$\hat{\Gamma}_w = \hat{\Gamma}_{w|p} + O_p(r_t),$$

$$\hat{\gamma}_w = \hat{\gamma}_{w|p} + O_p(r_t).$$

donde $r_t = \sum_{j=0}^{\infty} \psi_j(1 - \rho)z_{t-1-j}$. La magnitud del término de error r_t queda determinada en el siguiente teorema.

Teorema 2.1 Sea $\{w_t\}$ una serie temporal generada por (2.2.4) y sea w_1, \dots, w_T una muestra de este proceso. Sea ρ^{-1} la mayor raíz de su polinomio característico que verifica

$$\rho = \exp\left(-\frac{c}{T^\beta}\right) ; \beta > 1. \quad (2.3.7)$$

Entonces,

$$w_t = w_{t|p} + o_p(T^{-\frac{1}{2}}) \quad (2.3.8)$$

y

$$\hat{\Gamma}_w = \hat{\Gamma}_{w|p} + o_p(T^{-\frac{1}{2}}), \quad (2.3.9)$$

$$\hat{\gamma}_w = \hat{\gamma}_{w|p} + o_p(T^{-\frac{1}{2}}). \quad (2.3.10)$$

Además, si $\beta = 1$, el orden de probabilidad en (2.3.8), (2.3.9) y (2.3.10) es $O_p(T^{-\frac{1}{2}})$.

Demostración: Al verificarse

$$E(z_t^2) = \frac{\sigma^2}{1 - \rho^2} = O\left((1 - \rho^2)^{-1}\right), \quad (2.3.11)$$

entonces, por la desigualdad de Chebyshev, $z_t = O_p\left((1 - \rho^2)^{-\frac{1}{2}}\right)$. Por tanto, como w_t es estacionario,

$$O_p(r_t) = O_p\left(\left[\frac{1 - \rho}{1 + \rho}\right]^{\frac{1}{2}}\right). \quad (2.3.12)$$

Aplicando que $\rho = \exp(-c/T^\beta) = 1 - c/T^\beta + O(T^{-2\beta})$, se obtiene

$$\frac{1 - \rho}{1 + \rho} = O(T^{-\beta}). \quad (2.3.13)$$

Puesto que $\beta \geq 1$ se cumple el teorema. \square

Aunque se ha impuesto la definición de ρ de la expresión (2.3.2) se verifica fácilmente que esta forma de ρ aparece de forma natural en el presente contexto. Sea $T^{-\beta} = (1 - \rho)/(1 + \rho)$.

Entonces,

$$\rho = \frac{T^\beta - 1}{T^\beta + 1} = 1 - \frac{2}{T^\beta + 1}.$$

Como $T^{-\beta} < 1$

$$\rho = 1 - \frac{2}{T^\beta} \left[\sum_{j=0}^{\infty} \left(\frac{-1}{T^\beta}\right)^j \right] = e^{-\frac{2}{T^\beta}} + o(T^{-3\beta}).$$

El término $O\left((1 - \rho)(1 + \rho)^{-1}\right)$ en (2.3.12) no se ve afectado por el término constante de la exponencial y por esta razón el número 2 se reemplaza por c en la definición de ρ .

Corolario 2.1 Sean las condiciones del teorema 2.1, con $\beta \geq 1$. Entonces,

$$\begin{aligned} \hat{\gamma}_w &= \gamma_{w|p} + O_p(T^{-\frac{1}{2}}), \\ \hat{\Gamma}_w &= \Gamma_{w|p} + O_p(T^{-\frac{1}{2}}). \end{aligned}$$

Demostración: Puesto que $w_{t|p}$ es un proceso estacionario se verifica que $\hat{\gamma}_{w|p} = \gamma_{w|p} + O_p(T^{-\frac{1}{2}})$. Si se aplica este resultado en el teorema 2.1 se cumple el resultado. \square

Puede probarse entonces la \sqrt{T} -consistencia de $\hat{\phi}$.

Teorema 2.2 Sean las condiciones del teorema 2.1 con $\beta \geq 1$. Entonces,

$$\hat{\phi} = \phi + O_p(T^{-\frac{1}{2}}).$$

Ver demostración en apéndice A.2.

2.3.3 Sesgo y error cuadrático medio

Sea $\hat{\phi}_{|p}$ el estimador mínimo cuadrático de ϕ obtenido en una muestra de un proceso ARIMA($p, 1, 0$) que ha sido correctamente especificado. El sesgo y error cuadrático medio (ECM) de este estimador han sido estudiados extensamente en la literatura y pueden ser encontrados en los trabajos de Marriot & Pope (1954), Kendall (1954), Whitte (1961), Shenton & Johnson (1965), Bhansali (1981), Hosoya & Taniguchi (1982), Tjøsteim & Paulsen (1983), Tanaka (1984), Yamamoto & Kunitomo (1984), Kunitomo & Yamamoto (1985) y Shaman & Stine (1988) entre otros. Puesto que las posibles diferencias entre los estimadores $\hat{\phi}$ y $\hat{\phi}_{|p}$ dependen de la condición de casi no estacionariedad se expresarán sus diferencias en términos de ρ . Los siguientes teoremas formulan los momentos de primer y segundo orden del estimador $\hat{\phi}$, de un proceso AR($p + 1$) casi no estacionario sobrediferenciado, alrededor de los verdaderos parámetros ϕ como los momentos de $\hat{\phi}_{|p}$ alrededor de ϕ más un término de error que depende de ρ .

Teorema 2.3 Sean las hipótesis A1, A2 (con $s_o = 16$), A3 y A3'. Entonces,

$$E(\hat{\phi} - \phi) = E(\hat{\phi}_{|p} - \phi) + O\left(\left[\frac{1 - \rho}{1 + \rho}\right]^{\frac{1}{2}}\right). \quad (2.3.14)$$

La demostración se encuentra en el apéndice A.2.

Como $((1 - \rho)(1 + \rho)^{-1}) = O(T^{-\beta})$ y puesto que $E(\hat{\phi}_{|p} - \phi) = O(T^{-1})$ (ver, por ejemplo, Bhansali 1981) se necesita de un valor de $\beta > 2$ para que ambos sesgos sean iguales hasta términos de orden $O(T^{-1})$, mientras que para tener \sqrt{T} -consistencia sólo se necesitaba $\beta \geq 1$.

Teorema 2.4 Sean las hipótesis A1 (con $s_o = 16$), A2, A3 y A3'. Entonces,

$$E[(\hat{\phi} - \phi)(\hat{\phi} - \phi)'] = E[(\hat{\phi}_{|p} - \phi)(\hat{\phi}_{|p} - \phi)'] + O\left(\max\left\{\left(\frac{1 - \rho}{1 + \rho}\right)^{\frac{1}{2}} T^{-\frac{1}{2}}, \frac{1 - \rho}{1 + \rho}\right\}\right)$$

Ver demostración en apéndice A.2.

Se comprueba con este teorema que el ECM de ambos estimadores están más próximos que los sesgos. Si ρ es tal que $\beta > 1$ ambas expresiones de ECM serán iguales hasta términos de orden $O(T^{-1})$.

2.4 ECMP a horizonte H

En esta sección se obtienen las expresiones del ECMP de predecir y_{T+H} desde $t = T$. Para poder comparar el ECMP del modelo $AR(p+1)$ con el ECMP del modelo sobrediferenciado $AR(p, 1, 0)$ es necesario expresar las predicciones del horizonte H (\check{y}_{T+H} y \hat{y}_{T+H}) en términos de los incrementos estimados (\check{w}_t y \hat{w}_t respectivamente). Para el modelo $AR(p+1)$ los incrementos estimados son $\check{w}_t = \check{y}_t - \check{y}_{t-1}$. Entonces,

$$\check{y}_{T+H} = y_T + \sum_{h=1}^H \check{w}_{T+h}. \quad (2.4.1)$$

Por tanto,

$$\begin{aligned} \text{PMSE}(\check{y}_{T+H}) &= E(y_{T+H} - \check{y}_{T+H})^2 \\ &= \sum_{h=1}^H \text{PMSE}(\check{w}_{T+h}) + 2 \sum_{h=1}^H \sum_{k=h+1}^H E[(w_{T+h} - \check{w}_{T+h})(w_{T+k} - \check{w}_{T+k})]. \end{aligned} \quad (2.4.2)$$

De la misma forma, para el modelo sobrediferenciado, el ECMP(\hat{y}_{T+H}) puede ser expresado como

$$\text{ECMP}(\hat{y}_{T+H}) = \sum_{h=1}^H \text{ECMP}(\hat{w}_{T+h}) + 2 \sum_{h=1}^H \sum_{k=h+1}^H E[(w_{T+h} - \hat{w}_{T+h})(w_{T+k} - \hat{w}_{T+k})]. \quad (2.4.3)$$

2.4.1 ECMP del predictor correcto AR(p+1)

Sea \check{A}_α la estimación mínimo cuadrática de A_α en la muestra y_1, y_2, \dots, y_T , utilizando el modelo correcto (2.2.2). El valor previsto \check{y}_{T+H} utilizando esta información es

$$\check{y}_{T+H} = e'_{p+2} \check{A}_\alpha^H Y_T. \quad (2.4.4)$$

De acuerdo con los resultados de Kunitomo & Yamamoto (1985), el ECMP de predecir y_{T+H} desde T , utilizando este predictor estimado insesgado, es

$$\begin{aligned} \text{ECMP}(\check{y}_{T+H}) &= \sigma^2 \sum_{h=0}^{H-1} (e'_{p+2} A_\alpha^h e_{p+2})^2 + \frac{\sigma^2}{T} \sum_{h=0}^{H-1} \sum_{k=0}^{H-1} (e'_{p+2} A_\alpha^h e_{p+2})(e'_{p+2} A_\alpha^k e_{p+2}) \\ &\quad \times \text{traza} \left(A_\alpha^{H-1-h} \Gamma_y A_\alpha' A_\alpha^{H-1-k} \Gamma_y^{-1} \right) + O(T^{-3/2}). \end{aligned} \quad (2.4.5)$$

Sin embargo, para poder comparar el comportamiento predictivo de esta modelización respecto a la del modelo sobrediferenciado ARIMA(p, 1, 0) se necesita expresar el ECMP en términos de la serie sobrediferenciada w_t como se muestra en (2.4.2). Los incrementos estimados \check{w}_{T+h} como función de los coeficientes estimados \check{A}_α son

$$\check{w}_{T+h} = \check{y}_{T+h} - \check{y}_{T+h-1} = e'_{p+2} \check{A}_\alpha^{h-1} (\check{A}_\alpha - I_{p+2}) Y_T \quad (2.4.6)$$

donde I_{p+2} es la matriz identidad de dimensión $(p+2) \times (p+2)$. El valor observado w_{T+h} es

$$w_{T+h} = e'_{p+2} A_\alpha^{h-1} (A_\alpha - I_{p+2}) Y_T + L_1 - L_2,$$

donde

$$L_1 = \sum_{k=0}^{h-1} e'_{p+2} A_\alpha^k U_{t+h-k, p+2},$$

$$L_2 = \sum_{k=0}^{h-2} e'_{p+2} A_\alpha^k U_{t+h-1-k, p+2} = \sum_{k=1}^{h-1} e'_{p+2} A_\alpha^{k-1} U_{t+h-k, p+2}.$$

El error de predicción es, por tanto,

$$\check{w}_{T+h} - w_{T+h} = e'_{p+2} (\check{A}_\alpha^h - A_\alpha^h) Y_T - e'_{p+2} (\check{A}_\alpha^{h-1} - A_\alpha^{h-1}) Y_T - L_1 + L_2 \quad (2.4.7)$$

El ECMP(\check{w}_{T+h}) y $E[(\check{w}_{T+h} - w_{T+h})(\check{w}_{T+k} - w_{T+k})]$ se obtienen en el siguiente teorema (ver demostración en apéndice A.3).

Teorema 2.5 Sea w_t el proceso (2.2.4) con parámetro $\rho = \exp(-c/T^\beta)$, y $\beta > 1$. Sean las hipótesis A2, A3, A3' y A1 con $s_0 = 32$ cuando $h = 1, 2$ y $s_0 = 16h$ cuando $h \geq 3$. Entonces, el ECMP h periodos por delante usando el predictor estimado (2.4.6) es

$$ECMP(\check{w}_{T+h}) = E(\check{w}_{T+h} - w_{T+h})^2 = \sigma^2 \sum_{j=0}^{h-1} (e'_{p+2} A_1^j c_{p+2})^2 + \frac{\sigma^2}{T} \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} (e'_p A_p^j e_p) (e'_p A_p^k e_p) \\ \times \text{traza} \left(A_\alpha^{h-1-j} \Gamma_y A_\alpha^{h-1-k} \Gamma_y^{-1} \right) + o(T^{-1}), \quad (2.4.8)$$

y, para $k \geq h$,

$$E[(\check{w}_{T+h} - w_{T+h})(\check{w}_{T+k} - w_{T+k})] = \sigma^2 \sum_{i=0}^{h-1} (e'_{p+2} A_1^i c_{p+2}) (e'_{p+2} A_1^{i+(k-h)} c_{p+2}) \\ + \frac{\sigma^2}{T} \sum_{n=0}^{k-1} \sum_{i=0}^{h-1} (e'_p A_p^n e_p) (e'_p A_p^i e_p) \times \text{traza} \left(A_\alpha^{h-1-i} \Gamma_y A_\alpha^{k-1-n} \Gamma_y^{-1} \right) + o(T^{-1}), \quad (2.4.9)$$

donde $c_{p+1} = (1, 0, \dots, 0, 1)'$.

Los términos a la derecha de (2.4.8) y (2.4.9) tienen dos componentes. La primera parte es la varianza de los errores de predicción del proceso y la covarianza entre errores de predicción de diferentes horizontes, respectivamente, del proceso ARIMA($p+1, 1, 1$) verdadero. La segunda parte es el error muestral debido a la estimación de los $p+2$ parámetros $\hat{\phi}$. Los términos $(e'_p A_p^v e_p)$; $v = j, k, n, i$, son aproximaciones a los términos $(e'_{p+2} A_1^v e_{p+2})$; $v = j, k, n, i$. Esta sustitución de términos, bajo las condiciones del teorema 2.5, provoca un error de orden de magnitud reducido, $o(T^{-1})$, y permite la comparación con las expresiones del ECMP del modelo sobrediferenciado.

2.4.2 ECMP del predictor sobrediferenciado ARIMA($p, 1, 0$)

El predictor de w_{T+h} si se considera, erróneamente, que el proceso contiene una raíz unitaria será el que se deriva del modelo estimado ARIMA($p, 1, 0$), es decir,

$$\hat{w}_{T+h} = e'_p \hat{A}_p^h W_T \quad (2.4.10)$$

donde \hat{A}_p es el estimador mínimo cuadrático de A_p .

Se puede reescribir (2.4.10) como

$$\begin{aligned} \hat{w}_{T+h} &= e'_p A_p^h W_T + e'_p (\hat{A}_p^h - A_p^h) W_T \\ &= E(w_{T+h}|T) + e'_p (\hat{A}_p^h - A_p^h) W_T \end{aligned}$$

El verdadero valor w_{T+h} es, utilizando (2.2.5),

$$w_{T+h} = e'_{p+2} A_1^h Z_T + L_h = E(w_{T+h}|T) + L_h,$$

donde $L_h = \sum_{j=0}^{h-1} e'_{p+2} A_1^j U_{T+h-j}^*$. Por (2.3.6)

$$E(w_{T+h}|T) = E(w_{T+h}|T) - \sum_{j=h-1}^{\infty} \psi_j (1-\rho) z_{T-1-j} - \sum_{j=0}^{h-2} \psi_j (1-\rho) \rho^{h-1-j} z_T \quad (2.4.11)$$

Por tanto, el error de predicción h periodos por delante del predictor (2.4.12) es

$$\begin{aligned}
 (w_{T+h} - \hat{w}_{T+h}) &= L_h - e'_p(\hat{A}_p^h - A_p^h)W_T \\
 &\quad - \sum_{j=h-1}^{\infty} \psi_j(1-\rho)z_{T-1-j} - \sum_{j=0}^{h-2} \psi_j(1-\rho)\rho^{h-1-j}z_T \quad (2.4.12)
 \end{aligned}$$

El siguiente teorema proporciona una aproximación hasta orden $o(T^{-1})$ del ECMP de este predictor (ver demostración en el apéndice A.3).

Teorema 2.6 *Sea w_t el proceso (2.2.4) con parámetro $\rho = \exp(-c/T^\beta)$, y $\beta > 1$. Sean las hipótesis A2, A3, A3' y A1 con $s_0 = 32$ cuando $h = 1, 2$ y $s_0 = 16h$ cuando $h \geq 3$. Entonces, el ECMP de predecir h periodos utilizando el predictor (2.4.10) es*

$$\begin{aligned}
 ECMP(\hat{w}_{T+h}) = E(\hat{w}_{T+h} - w_{T+h})^2 &= \sigma^2 \sum_{k=0}^{h-1} (e'_{p+2} A_1^k c_{p+2})^2 + \frac{\sigma^2}{T} \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} (e'_p A_p^j e_p)(e'_p A_p^k e_p) \\
 &\quad \times \text{traza} \left(A_p^{h-1-j} \Gamma_{w|p} A_p^{h-1-k} \Gamma_{w|p}^{-1} \right) + o(T^{-1}), \quad (2.4.13)
 \end{aligned}$$

y, para $k \geq h$,

$$\begin{aligned}
 E[(\hat{w}_{T+h} - w_{T+h})(\hat{w}_{T+k} - w_{T+k})] &= \sigma^2 \sum_{i=0}^{h-1} (e'_{p+2} A_1^i c_{p+2})(e'_{p+2} A_1^{i+(k-h)} c_{p+2}) \\
 &\quad + \frac{\sigma^2}{T} \sum_{n=0}^{k-1} \sum_{i=0}^{h-1} (e'_p A_p^n e_p)(e'_p A_p^i e_p) \times \text{traza} \left(A_p^{h-1-i} \Gamma_{w|p} A_p^{k-1-n} \Gamma_{w|p}^{-1} \right) + o(T^{-1}), \quad (2.4.14)
 \end{aligned}$$

donde $c_{p+2} = (1, 0, \dots, 0, 1)'$.

Como se mencionó en la anterior subsección, los términos a la derecha de (2.4.13) y (2.4.14) tienen dos componentes. El primero de ellos, la varianza de los errores de predicción y sus covarianzas entre distintos horizontes del proceso verdadero ARIMA($p+1, 1, 1$), es el mismo que en el teorema 2.5. El segundo componente es el error muestral debido a la estimación de los p parámetros $\hat{\phi}$, en contraste con los $p+2$ parámetros del modelo ARIMA($p+1, 1, 1$). Nótese que estos segundos componentes, en ambos predictores alternativos, difieren sólo en los elementos contenidos en el operador traza.

2.5 Comparación de la capacidad predictiva

En esta sección se comparan los ECMP de ambos predictores, calculados en la sección anterior. Esta comparación permite demostrar que, bajo la hipótesis de casi no estacionariedad formulada en (2.3.7), sobrediferenciar produce menor ECMP. Al comparar las expresiones que proporcionan los teoremas 2.5 y 2.6 puede observarse que la única diferencia entre el ECMP(\check{y}_{T+H}) y el ECMP(\hat{y}_{T+H}) se encuentra en los elementos contenidos en las respectivas trazas. Estas trazas pueden compararse aplicando los resultados de dos lemas. El lema 2.1 compara las trazas de procesos con y sin término constante. El lema 2.2 compara las trazas de un proceso casi no estacionario sin término constante y la de un proceso sobrediferenciado. Las demostraciones de estos lemas puede encontrarse en el apéndice A.4.

Lema 2.1 Sea y_t el proceso (2.2.1) con $\alpha \neq 0$. Entonces,

$$\text{traza}(A_\alpha^i \Gamma_y A_\alpha^j \Gamma_y^{-1}) = 1 + \text{traza}(A_o^i \Gamma_{\check{y}} A_o^j \Gamma_{\check{y}}^{-1}). \quad (2.5.1)$$

Lema 2.2 Sea y_t el proceso (2.2.1) con parámetro $\rho = \exp(-c/T^\beta)$; $\beta > 1$. Entonces,

$$\text{traza}(A_o^i \Gamma_{\check{y}} A_o^j \Gamma_{\check{y}}^{-1}) = \rho^{i+j} + \text{traza}(A_p^i \Gamma_{w|p} A_p^j \Gamma_{w|p}^{-1}) + o(T^{-1}). \quad (2.5.2)$$

Puede ahora demostrarse la ventaja de sobrediferenciar cuando el proceso es casi no estacionario.

Teorema 2.7 Sea y_t el proceso (2.2.1) y sean las condiciones de los teoremas 2.5 y 2.6. Entonces, para $H \geq 1$,

$$\text{ECMP}(\hat{y}_{T+H}) < \text{ECMP}(\check{y}_{T+H}). \quad (2.5.3)$$

Demostración: Mediante la aplicación directa de los lemas 2.1 y 2.2 a las diferencias entre (2.4.8) y (2.4.13) y entre (2.4.14) y (2.4.9) se obtiene (véase (2.4.2) y (2.4.3))

$$\text{ECMP}(\check{y}_{T+H}) - \text{ECMP}(\hat{y}_{T+H}) = \frac{\sigma^2}{T} \left(\sum_{h=1}^H \sum_{j=0}^{h-1} \psi_j \right)^2 + \frac{\sigma^2}{T} \left(\sum_{h=1}^H \sum_{j=0}^{h-1} \psi_j \rho^{h-1-j} \right)^2 > 0, \quad (2.5.4)$$

donde $\psi_j = (e'_p A_p^j e_p)$, $j = 1, \dots, H$. □

La expresión (2.5.4) muestra que la ventaja del modelo sobrediferenciado puede descomponerse en dos partes. La primera parte, el primer término a la derecha de (2.5.4) es el resultado de aplicar el lema 2.1 y por tanto es la diferencia procedente de la estimación del término constante α en el modelo $\text{AR}(p+1)$. La segunda parte es el resultado de aplicar el lema 2.2 y por tanto procede de la estimación de un parámetro adicional, en el polinomio autorregresivo, en el modelo $\text{AR}(p+1)$ respecto al modelo $\text{ARIMA}(p, 1, 0)$. Por tanto, el superior comportamiento predictivo del modelo $\text{ARIMA}(p, 1, 0)$ es debido a su mayor parsimonia.

La diferencia (2.5.4) aumenta monótonamente con el horizonte de predicción. Para $H = 1$ la diferencia es $2\sigma^2/T$ si el modelo contiene el término $\alpha \neq 0$, y σ^2/T si $\alpha = 0$ y no se estima término constante. Este resultado es similar al obtenido por Ledolter & Abraham (1981) al analizar la pérdida de eficiencia en predicción por utilizar más parámetros de los necesarios. Estos autores demuestran que cada parámetro innecesario incrementa el ECMP a un periodo en σ^2/T . En el presente contexto esto es equivalente a decir que tanto α como ρ son innecesarios y por tanto existe una pérdida de eficiencia (que tiende a cero asintóticamente) si se utiliza el modelo con todos los parámetros $\text{AR}(p+1)$. Para $H > 1$ la pérdida de eficiencia aumenta en una cantidad que depende de los pesos ψ_j .

Análogamente, si $\beta < 1$ el modelo sobrediferenciado tiene peor comportamiento predictivo que el verdadero modelo, como se demuestra en el siguiente teorema.

Teorema 2.8 *Sea y_t el proceso definido en (2.2.1). Sean las condiciones del teorema 2.1 con $\beta < 1$. Entonces, para $h > 0$,*

$$ECMP(\hat{y}_{T+H}) > ECMP(\check{y}_{T+H}). \quad (2.5.5)$$

Ver demostración en el apéndice A.4.

2.6 Sobrediferenciación de procesos AR(1)

Aunque los resultados de la anterior sección son aplicables para un proceso autorregresivo de orden $p + 1$ genérico resulta de interés analizar el caso AR(1). Una razón es que la formulación del ECMP es más simple y muchas de las aproximaciones asintóticas que se hicieron en la sección anterior no serán necesarias. Esto permitirá encontrar nuevos resultados. Una segunda razón es que no existen más raíces que puedan influir en el resultado. La sección anterior mostraba que raíces negativas aumentaban la distancia entre el modelo propuesto y el correcto mientras que las raíces positivas disminuían dicha distancia. En el caso de un AR(1) no existen dichas raíces y los resultados obtenidos pueden considerarse como una referencia neutra. En esta sección se analizarán los casos AR(1) de media nula (AR(1)) y media no nula (AR(1, μ)).

2.6.1 El caso AR(1)

Sea y_t , $t = 1, 2, \dots, T$, una muestra del proceso estacionario

$$y_t = \phi y_{t-1} + a_t, \quad \phi < 1, \quad (2.6.1)$$

y sea $\check{\phi}$ el estimador mínimo cuadrático de ϕ . La esperanza condicionada de y_{T+H} , utilizando información hasta T es $\check{y}_{T+H} = \check{\phi}^H y_T$, y el ECMP es (véase (2.4.4))

$$\text{ECMP}(y_{T+H}|\text{AR}(1)) = \sigma^2 \left[\frac{1 - \phi^{2H}}{1 - \phi^2} + \frac{H^2 \phi^{2H-2}}{T} \right] + O(T^{-\frac{3}{2}}). \quad (2.6.2)$$

El predictor sobrediferenciado en este caso es el paseo aleatorio (PA). Entonces, la predicción de y_{T+H} desde T es $\hat{y}_{T+H} = y_T$ y por tanto

$$\text{ECMP}(y_{T+H}|\text{PA}) = \sigma^2 \left(\frac{1 - \phi^{2H}}{1 - \phi^2} + \frac{(1 - \phi^H)^2}{1 - \phi^2} \right). \quad (2.6.3)$$

Comparando (2.6.2) con (2.6.3) se verifica que, si se cumple la desigualdad

$$\frac{H^2 \phi^{2(H-1)}}{T} - \frac{(1 - \phi^H)^2}{1 - \phi^2} > 0, \quad (2.6.4)$$

entonces, $\text{ECMP}(y_{T+H}|\text{AR}(1)) > \text{ECMP}(y_{T+H}|\text{PA})$. La tabla 2.1 muestra la relación entre ϕ , T y H que hacen que se cumpla la desigualdad (2.6.4). Esta tabla muestra que al aumentar el horizonte de predicción resulta cada vez más difícil superar al modelo AR(1) correcto, es decir, es necesario que ϕ esté más cerca de uno para que sobrediferenciar proporcione menor ECMP. Sin embargo, esta variación con el horizonte H es muy pequeña. Para analizar mejor este efecto puede observarse que, para $H = 1$, aplicando (2.6.4)

$$\frac{1}{T} > \frac{1 - \phi}{1 + \phi}, \quad (2.6.5)$$

y por tanto

$$\phi > 1 - \frac{2}{T+1} = \exp\left(-\frac{2}{T}\right) + O(T^{-3}). \quad (2.6.6)$$

Este resultado muestra que la convergencia a la unidad con el tamaño muestral ha de ser $O(T^{-1})$. Aplicando este resultado en la expansión de Taylor de ϕ^H alrededor de la unidad se obtiene

$$\phi^H = 1 - H(1 - \phi) + O(T^{-2}). \quad (2.6.7)$$

Tabla 2.1: Mínimos valores de ϕ para obtener $\text{ECMP}(y_{T+H}|\text{AR}(1)) > \text{ECMP}(y_{T+H}|\text{PA})$

T	Horizonte H							
	1	2	3	4	5	10	15	20
25	0.923	0.929	0.933	0.937	0.940	0.951	0.958	0.963
50	0.961	0.962	0.964	0.965	0.966	0.970	0.973	0.976
75	0.974	0.974	0.975	0.976	0.976	0.978	0.980	0.982
100	0.980	0.981	0.981	0.981	0.982	0.983	0.984	0.985
150	0.987	0.987	0.987	0.987	0.987	0.988	0.989	0.989
200	0.990	0.990	0.990	0.990	0.990	0.991	0.991	0.992
300	0.993	0.993	0.993	0.994	0.994	0.994	0.994	0.994
500	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996

Si se denomina $\phi(H)$ al mínimo valor de ϕ necesario para que se verifique la desigualdad (2.6.4) en el horizonte H se obtiene

$$\phi(H) = 1 - \frac{2}{T + 1 + 4(H - 1)} + O(T^{-2}) = \exp\left(-\frac{2}{T + 4(H - 1)}\right) + O(T^{-2}). \quad (2.6.8)$$

Por ejemplo, si $\phi = 0.97$ y $T = 50$ el paseo aleatorio tendrá menos ECMP que el AR(1) hasta $H = 10$. Desde este horizonte en adelante el PA tiene ligeramente mayor ECMP que el modelo correcto. Se verifica fácilmente que $\phi(H) - \phi(1) = O(HT^{-2})$ y $\phi(H+1) - \phi(H) = O(T^{-2})$. Por tanto, la dependencia de $\phi(H)$ con el horizonte de predicción es de magnitud $O(T^{-2})$ y sólo es apreciable en muestras pequeñas y horizontes muy alejados.

2.6.2 El caso AR(1, μ)

Sea y_t , $t = 1, 2, \dots, T$, una muestra del proceso estacionario

$$y_t = \alpha + \phi y_{t-1} + a_t, \phi < 1, \quad (2.6.9)$$

y sean $\check{\alpha}$ y $\check{\phi}$ los estimador mínimo cuadráticos de α y ϕ respectivamente. Siguiendo los mismos argumentos que en la subsección precedente se tiene, utilizando información hasta el periodo T ,

$$E(y_{T+H}|T) = \check{y}_{T+H} = \check{\alpha} \frac{1 - \check{\phi}^H}{1 - \check{\phi}} + \check{\phi}^H y_T, \quad (2.6.10)$$

y utilizando que el ECMP de este predictor, por (2.4.4), es

$$\text{ECMP}(y_{T+H}|\text{AR}(1|\mu)) = \sigma^2 \frac{1 - \phi^{2H}}{1 - \phi^2} + \frac{\sigma^2}{T} \left[H^2 \phi^{2H-2} + \left(\frac{1 - \phi^H}{1 - \phi} \right)^2 \right] + O(T^{-\frac{3}{2}}). \quad (2.6.11)$$

Por el contrario, si el predictor empleado es el PA el ECMP será

$$\begin{aligned} \text{ECMP}(y_{T+H}|\text{PA}) &= \sigma^2 \frac{1 - \phi^H}{1 - \phi^2} + \frac{\alpha^2 (1 - \phi^H)^2}{(1 - \phi)^2} + (1 - \phi^H)^2 E(y_T^2) \\ &\quad - 2 \frac{\alpha (1 - \phi^H)^2}{1 - \phi} E(y_T). \end{aligned} \quad (2.6.12)$$

Puesto que $E(y_T) = \frac{\alpha}{1 - \phi}$, y $E(y_T^2) = \frac{\sigma^2}{1 - \phi^2} + \left(\frac{\alpha}{1 - \phi} \right)^2$, se obtiene

$$\text{ECMP}(y_{T+H}|\text{PA}) = \sigma^2 \left(\frac{1 - \phi^{2H}}{1 - \phi^2} + \frac{(1 - \phi^H)^2}{1 - \phi^2} \right), \quad (2.6.13)$$

que coincide con (2.6.3). La comparación de (2.6.11) con (2.6.13) proporciona que, si se verifica la desigualdad

$$\frac{H^2 \phi^{2(H-1)}}{T} + \frac{(1 - \phi^H)^2}{T(1 - \phi)^2} - \frac{(1 - \phi^H)^2}{1 - \phi^2} > 0, \quad (2.6.14)$$

Tabla 2.2: Mínimos valores de ϕ para obtener $\text{ECMP}(y_{T+H}|\text{AR}(1 | \mu)) > \text{ECMP}(y_{T+H}|\text{PA})$

T	Horizonte H							
	1	2	3	4	5	10	15	20
25	0.852	0.862	0.869	0.876	0.881	0.898	0.907	0.913
50	0.923	0.926	0.928	0.931	0.932	0.940	0.945	0.948
75	0.948	0.949	0.951	0.952	0.953	0.957	0.960	0.962
100	0.961	0.962	0.962	0.963	0.964	0.966	0.968	0.970
150	0.974	0.974	0.974	0.975	0.975	0.976	0.977	0.978
200	0.980	0.980	0.981	0.981	0.981	0.982	0.982	0.983
300	0.987	0.987	0.987	0.987	0.987	0.988	0.988	0.988
500	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992

entonces, $\text{ECMP}(y_{T+H}|\text{AR}(1 | \mu)) > \text{ECMP}(y_{T+H}|\text{RW})$. La tabla 2.2 muestra los valores mínimos de ϕ necesarios para que se verifique dicha desigualdad. Utilizando (2.6.7) se obtiene, para $H \geq 1$

$$\phi(H) = 1 - \frac{4}{T + 2 + 4(H - 1)} + O(T^{-2}) = \exp\left(-\frac{4}{T + 4(H - 1)}\right) + O(T^{-2}). \quad (2.6.15)$$

Se observa en la tabla 2.2, así como en la expresión (2.6.15), que la necesidad de estimar una constante hace más fácil para un PA superar al autorregresivo. Al igual que se describía en la subsección anterior, tanto la tabla 2.2 como la expresión (2.6.15) muestran que, al aumentar el horizonte de predicción, es necesario estar cada vez más cerca de la no estacionariedad para que (2.6.14) se cumpla, aunque esta variación es pequeña.

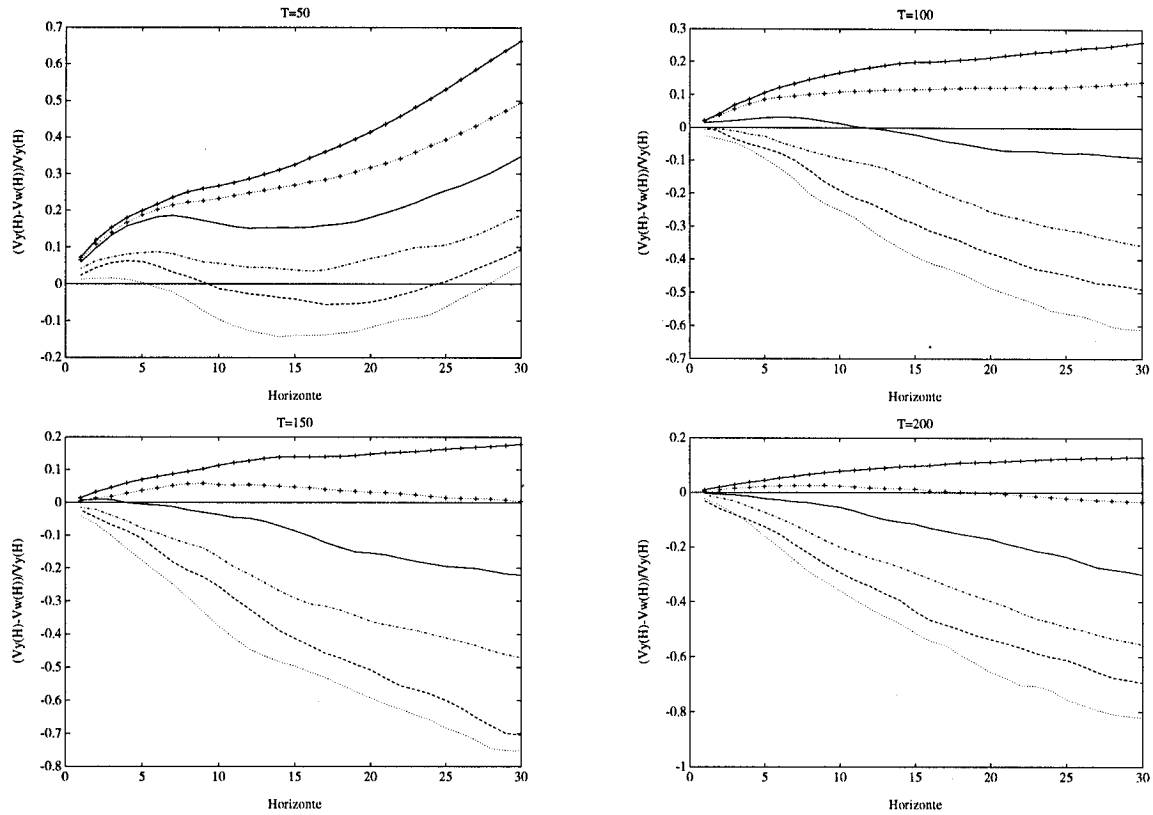


Figura 2.2: $(V_y(H) - V_w(H))/V_y(H)$ del modelo M1 para horizontes $H = 1, \dots, 30$ y tamaño muestral T . Los valores de ρ son (de abajo a arriba): 0.90, 0.92, 0.94, 0.96, 0.98, 0.99.

2.7 Estudio de simulación

Esta sección presenta un ejercicio de simulación que ilustra los anteriores resultados y los valida en muestras finitas. Se analizan tres modelos AR(2) diferentes,

$$\text{M1: } (1 - 0.5B)(1 - \rho B)y_t = 10 + a_t$$

$$\text{M2: } (1 - 0.5B)(1 - \rho B)y_t = a_t$$

$$\text{M3: } (1 + 0.8B)(1 - \rho B)y_t = 10 + a_t.$$

con $\rho = 0.9, 0.92, 0.94, 0.96, 0.98, 0.99$, y cuatro tamaños muestrales $T = 50, 100, 150, 200$. Las series reales usualmente tienen media no nula por lo que los modelos M1 y M3

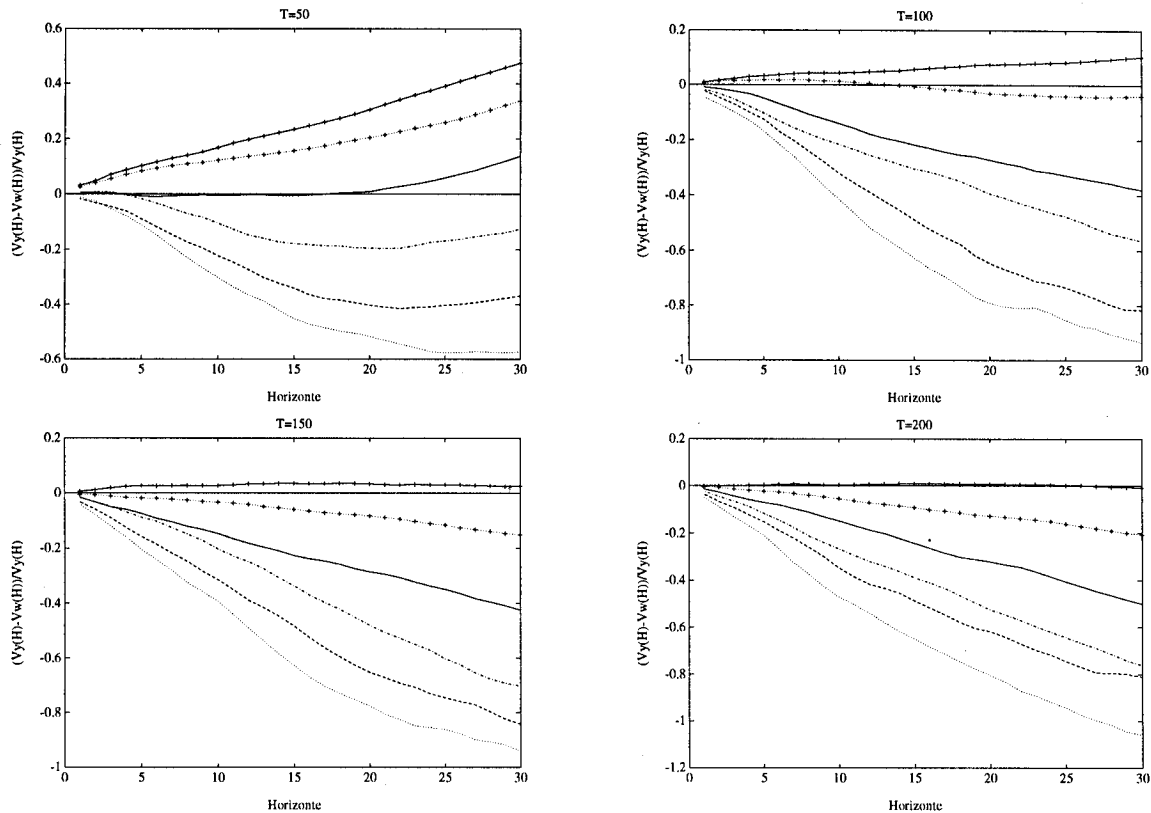


Figura 2.3: $(V_y(H) - V_w(H))/V_y(H)$ del modelo M2 para horizontes $H = 1, \dots, 30$ y tamaño muestral T . Los valores de ρ son (de abajo a arriba): 0.90, 0.92, 0.94, 0.96, 0.98, 0.99.

ilustran las consecuencias de sobrediferenciar en estas series. La diferencia entre estos dos modelos es en el valor de la segunda raíz. Es de esperar que la sobrediferenciación de M3 produzca peores resultados que en M2 al poseer una raíz negativa. También la mayoría de las series reales tienen su primera diferencia con media nula. Por tanto, si la decisión a tomar es sobre si se necesita una segunda diferencia pueden esperarse situaciones sin término constante como la del modelo M2.

En cada replicación se ha generado una muestra del modelo considerado de tamaño

$500 + T + 30$ con perturbaciones $a_t \sim N(0, 1)$. Las primeras 500 observaciones fueron eliminadas para evitar la influencia de los valores iniciales. Las últimas 30 fueron reservadas para evaluar las predicciones. En las T observaciones centrales se estimaron, por mínimos cuadrados, el predictor correcto y el sobrediferenciado. Se han realizado un total de 5000 repeticiones de cada modelo. Se ha denominado $V_y(H)$ al ECMP muestral resultante de promediar, en cada horizonte $H = 1, 2, \dots, 30$, los 5000 errores cuadráticos de predecir con el modelo AR(2) bien especificado. Asimismo se ha denominado $V_w(H)$ al ECMP muestral cuando las predicciones de y_{T+H} han sido generadas por el modelo sobrediferenciado ARIMA(1,1,0). Las figuras 2.2 a 2.4 muestran el ratio $(V_y(H) - V_w(H))/V_y(H)$ para M1 a M3 y diferentes valores de T y ρ . Este ratio representa la ganancia esperada (o pérdida si es negativo) de sobrediferenciar a cada horizonte.

Estas figuras revelan que, como se esperaba de los resultados teóricos, existen situaciones donde la sobrediferenciación supera al modelo correcto. La ganancia esperada se incrementa con la proximidad de ρ a la unidad y disminuye al aumentar T . También se observa que, congruente con la expresión (2.3.4), la ganancia es mayor en el modelo con segunda raíz positiva (M1) que en el modelo con segunda raíz negativa (M3). La ganancia se reduce sustancialmente si no es necesario estimar constante (M2).

La principal característica de estas curvas es su divergencia al aumentar el horizonte de predicción. Para $H = 1$ la diferencia entre ambos predictores es pequeña. Las ganancias (o pérdidas) son casi nulas. Sin embargo, en el largo plazo las ganancias o pérdidas son bastante importantes en la mayoría de los casos.

Una segunda característica es la asimetría de las consecuencias de sobrediferenciar. En muestras grandes la pérdida de eficiencia de la sobrediferenciación pueden ser muy

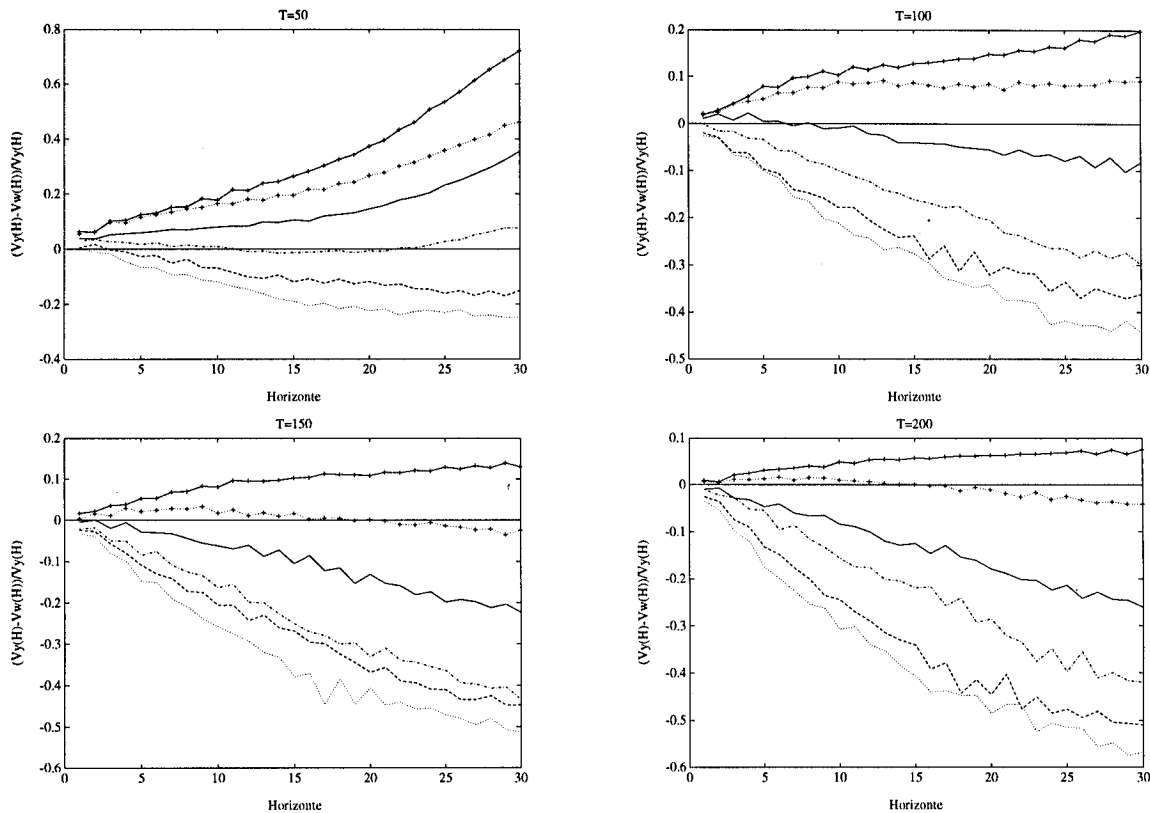


Figura 2.4: $(V_y(H) - V_w(H))/V_y(H)$ del modelo M3 para horizontes $H = 1, \dots, 30$ y tamaños muestrales T . Los valores de ρ son (de abajo a arriba): 0.90, 0.92, 0.94, 0.96, 0.98, 0.99.

grandes si no se encuentra ρ muy próximo a la unidad, mientras que las ventajas de sobrediferenciar son modestas (son de orden $O(T^{-1})$). Este resultado no debe, sin embargo, producir especial preocupación pues los contrastes de raíces unitarias existentes tienen una potencia alta en tamaños muestrales elevados. Sin embargo, en el caso de muestras reducidas, la ganancia de sobrediferenciar es elevada, mientras que la pérdida en el caso en que ρ no se encuentre cerca de la unidad es pequeña. Este resultado coincide con las conclusiones de los teoremas 2.7 a 2.8. Si ρ está suficientemente cerca de la unidad ($\beta > 1$) el ECMP del modelo sobrediferenciado es menor que el del modelo correcto, siendo dicha ventaja de magnitud $O(T^{-1})$, mientras que si $\beta < 1$ el predictor

sobrediferenciado es mayor que el del modelo correcto en un factor de magnitud mayor que $O(T^{-1})$.

Puede asimismo observarse que existen casos en los que cambia el signo de la ganancia esperada. Por ejemplo, en la figura 2.2, si $T = 100$ y $\rho = 0.96$ hay una ganancia esperada positiva si se sobrediferencia en horizontes bajos, pero para horizontes altos se produce una pérdida. El modelo óptimo para predecir varía, por tanto, con el horizonte de predicción. Este efecto se detectó previamente en la sección 2.6, donde se vió que, a medida que aumentaba el horizonte de predicción, era necesario un valor de ρ más próximo a la unidad para que la sobrediferenciación fuese ventajosa. En $T = 50$ se observa que el signo puede cambiar incluso dos veces, pero este segundo cambio se produce a horizontes muy alejados comparado con el tamaño muestral.

Puesto que los resultados dependen principalmente del tamaño de las restantes raíces más que de su número es razonable prever conclusiones similares en autorregresivos de orden mayor al empleado en este ejercicio de simulación.

2.8 Conclusiones finales

Se demuestra en este capítulo que si un proceso $AR(p+1)$ estacionario tiene una raíz de valor $\rho = \exp(-c/T^\beta)$, $c > 0$, $\beta > 1$, la estimación de un modelo sobrediferenciado $ARIMA(p, 1, 0)$ es \sqrt{T} -consistente y su ECMP es menor que el del predictor correcto debido a su parsimonia. Para un $AR(1)$ de media nula se obtiene que el valor de ρ a partir del cuál se obtiene este resultado tiene los valores $c = 2$ y $\beta = 1$, mientras que si es de media no nula es necesario $c = 4$ y $\beta = 1$. En autorregresivos de órdenes superiores las restantes raíces tienen influencia en muestras pequeñas. Raíces negativas disminuyen la

ventaja del predictor sobrediferenciado y serán necesarios valores de $c < 2$ (o $c < 4$ si la media es no nula) para que sobrediferenciar sea eficiente. Raíces positivas aumentarán el efecto de la sobrediferenciación y por tanto se tendrá $c > 2$ ($c > 4$). A estos procesos en los que $\rho = \exp(-c/T^\beta)$, $\beta > 1$ se les denomina *casi no estacionarios*.

La superioridad del predictor sobrediferenciado en los procesos casi no estacionarios es pequeña en el corto plazo pero aumenta, en general, con el horizonte de predicción. Sin embargo, si ρ no está suficientemente cerca de la unidad, el predictor sobrediferenciado puede aún tener similar comportamiento en el corto plazo que el predictor correcto, pero las predicciones a largo plazo se verán seriamente afectadas.

Este resultado disminuye la importancia de la baja potencia de los contrastes de raíces unitarias en la región cercana a la no estacionariedad. Aunque en estos casos el contraste detecte erróneamente, con muy alta probabilidad, una raíz unitaria el modelo $ARIMA(p, 1, 0)$ resultante tendrá mejor comportamiento predictivo que el modelo correctamente especificado.

Se ha observado que existen situaciones en las que el predictor óptimo a horizonte uno no es necesariamente el óptimo a mayores horizontes. En otras situaciones la optimalidad en horizonte uno es difícil de determinar, al ser los predictores alternativos muy similares. Sin embargo, en horizontes mayores la optimalidad de uno de ellos es clara. Este diferente comportamiento de la capacidad predictiva relativa en el corto y largo plazo revela que la diagnosis basada en los errores de predicción a horizonte uno es una herramienta poco eficaz en las proximidades de la no estacionariedad. Por tanto, si un predictor va a ser utilizado para predecir H periodos por delante es aconsejable analizar las propiedades de los diferentes predictores disponibles en dicho horizonte.

Ha de tenerse, no obstante, especial precaución con el procedimiento empleado para evaluar el comportamiento predictivo de un modelo. Es práctica frecuente en la modelización de series temporales el utilizar una parte de la muestra para la estimación y reservar así algunas observaciones para evaluar el error de predicción. Si la serie es casi no estacionaria esta práctica es bastante arriesgada, pues aumentará la probabilidad de detectar una raíz unitaria. La región de *casi no estacionariedad* aumenta exponencialmente al reducirse el tamaño muestral. Posteriormente, al validar dicho resultado en la parte de la muestra reservada para predicción, se volverá de nuevo a favorecer la elección del modelo no estacionario por tener menor ECMP.

La idea de utilizar diferentes predictores para cada horizonte de predicción, independientemente de cual sea el modelo correcto, ha sido ampliamente explorada en la literatura (Cox 1961, Gersh & Kitagawa 1983, Findley 1984, Weiss 1991, Tsay 1993, Tiao & Xu 1993, Tiao & Tsay 1994 entre otros). Estos autores justifican esta práctica, principalmente, bajo la hipótesis de que el modelo correcto no se conoce y por tanto los predictores disponibles son todos erróneos. En el presente caso la recomendación de buscar el predictor adecuado en cada horizonte sigue siendo válida aunque se conozca el modelo, pero no el valor de sus parámetros. La variabilidad muestral debida a la estimación de los parámetros hace que el modelo correcto pueda ser superado en algunas situaciones límite, como *casi no estacionariedad*. Se pueden obtener mejores predictores, especialmente en el largo plazo, usando un modelo más parsimonioso que el modelo correcto. Este resultado refuerza la importancia de la parsimonia en la predicción y extiende los resultados de Ledolter & Abraham (1981) señalando las ventajas, en algunas situaciones, de utilizar modelos con menos parámetros de los que posee el proceso generador de datos.

Apéndice

A.1 Algunos lemas previos

En esta sección se demuestran algunos lemas que serán necesarios para la posteriores demostraciones. Para un vector x arbitrario $r \times 1$ y una matriz M $r \times r$, se define la norma euclídea de x como $\|x\| = (x'x)^{1/2}$, y para M , $\|M\| = \sup_{\|x\| \leq 1} (x'M' Mx)^{1/2}$.

Lema A.1.1 Sean las hipótesis A1 y A2, con $s_0 = 2k$ y $k \geq 1$. Entonces si $T \rightarrow \infty$

$$E(\|\hat{\Gamma}_w - \hat{\Gamma}_{w|p}\|^k) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right), \quad (\text{A.1.1})$$

y

$$E(\|\hat{\gamma}_w - \hat{\gamma}_{w|p}\|^k) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right). \quad (\text{A.1.2})$$

Demostración:

La norma de una matriz es de la misma magnitud que la norma de su mayor elemento. Por tanto $E(\|\hat{\Gamma}_w - \hat{\Gamma}_{w|p}\|^k) = O\left(E\{[(w_t w_{t-s} - w_{t|p} w_{t-s|p})^2]^{\frac{k}{2}}\}$, y el mismo resultado puede aplicarse a (A.1.2). Utilizando la descomposición (2.3.6) se tiene que

$$\begin{aligned} w_t w_{t-s} - w_{t|p} w_{t-s|p} &= -\sum_{i=0}^{\infty} \psi_i (1-\rho) z_{t-s-1-i} w_{t|p} - \sum_{j=0}^{\infty} \psi_j (1-\rho) z_{t-1-j} w_{t-s|p} \\ &\quad + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i (1-\rho)^2 z_{t-1-j} z_{t-s-1-i} \\ &= -a - b + c, \end{aligned}$$

donde

$$\begin{aligned} a &= \sum_{i=0}^{\infty} \psi_i (1-\rho) z_{t-s-1-i} w_{t|p}, \\ b &= \sum_{j=0}^{\infty} \psi_j (1-\rho) z_{t-1-j} w_{t-s|p}, \end{aligned}$$

$$c = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i (1-\rho)^2 z_{t-1-j} z_{t-s-1-i}.$$

Haciendo uso de la desigualdad de Minkowski

$$E| -a - b + c|^k \leq \left\{ (E|a|^k)^{\frac{1}{k}} + (E|b|^k)^{\frac{1}{k}} + (E|c|^k)^{\frac{1}{k}} \right\}^k.$$

Se resuelve primeramente el término $E|a|^k$:

$$\begin{aligned} E|a|^k &= E \left[\left| \sum_{i=0}^{\infty} \psi_i (1-\rho) z_{t-s-1-i} w_{t|p} \right|^k \right] \\ &\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_k=0}^{\infty} |\psi_{i_1} \cdots \psi_{i_k} (1-\rho)^k| \end{aligned} \quad (\text{A.1.3})$$

$$\times E[|z_{t-s-1-i_1} \cdots z_{t-s-1-i_k} w_{t|p}^k|], \quad (\text{A.1.4})$$

donde

$$\begin{aligned} E[|z_{t-s-1-i_1} \cdots z_{t-s-1-i_k} w_{t|p}^k|] &\leq \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \sum_{m_1=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} \rho^{r_1} \cdots \rho^{r_k} |\psi_{m_1} \cdots \psi_{m_k}| \\ &\times E(|a_{t-(r_1+i_1+s+1)} \cdots a_{t-(r_k+i_k+s+1)} a_{t-m_1} \cdots a_{t-m_k}|). \end{aligned} \quad (\text{A.1.5})$$

Utilizando la hipótesis A1 se verifica que $E(|a_t|^{s_0}) < \infty$. Por tanto el término en la última línea de (A.1.5) es $O(1)$ si las innovaciones a_t 's coinciden por pares y es nulo si no lo hacen.

Para simplificar la exposición se considera el caso $k = 2$, entonces

$$\begin{aligned} E[|z_{t-s-1-i_1} z_{t-s-1-i_2} w_{t|p}^2|] &\leq \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \rho^{r_1} \rho^{r_2} |\psi_{m_1} \psi_{m_2}| \\ &\times E(|a_{t-(r_1+i_1+s+1)} a_{t-(r_2+i_2+s+1)} a_{t-m_1} a_{t-m_2}|). \end{aligned} \quad (\text{A.1.6})$$

Se denota por \mathcal{S}_1 a las situaciones en que las cuatro innovaciones a_t 's de (A.1.6) sean iguales. Entonces $\{(r_1 + i_1 + s + 1) = (r_2 + i_2 + s + 2) = m_1 = m_2\}$; y por tanto, suponiendo sin pérdida de generalidad el caso $i_1 > i_2$,

$$\begin{aligned} E[|z_{t-s-1-i_1} z_{t-s-1-i_2} w_{t|p}^2| | \mathcal{S}_1] &= \sum_{r_1=0}^{\infty} \rho^{r_1} \rho^{r_1+i_1+2s+2-i_2} \psi_{r_1+i_1+s+1}^2 O(1) \\ &\leq \sum_{r_1=0}^{\infty} \psi_{r_1}^2 O(1) = O(1). \end{aligned}$$

Una segunda posibilidad es que las innovaciones a_t 's de (A.1.6) coincidan por pares, pero que cada par contenga el producto cruzado de z 's y w 's simultáneamente, como en el caso $\{(r_1 + i_1 + s + 1 = m_1) \neq (r_2 + i_2 + s + 2 = m_2)\}$. Se denota por \mathcal{S}_2 a estas situaciones, entonces

$$\begin{aligned} E[|z_{t-s-1-i_1} z_{t-s-1-i_2} w_{t|p}^2| \mid \mathcal{S}_2] &= \left(\sum_{r_1=0}^{\infty} \rho^{r_1} |\psi_{r_1+i_1+s+1}| O(1) \right) \left(\sum_{r_2=0}^{\infty} \rho^{r_2} |\psi_{r_2+i_2+s+1}| O(1) \right) \\ &\leq \left(\sum_{r_1=0}^{\infty} |\psi_{r_1}| \right)^2 = O(1). \end{aligned}$$

La tercera posibilidad es que las innovaciones a_t 's de (A.1.6) coincidan por pares pero un par contenga el producto de dos términos z 's y el otro de w 's, entonces $\{(r_1 + i_1 + s + 1 = r_2 + i_2 + s + 2) \neq (m_1 = m_2)\}$. Si se supone, sin pérdida de generalidad, el caso $i_1 > i_2$ y se denota por \mathcal{S}_3 a estas situaciones

$$\begin{aligned} E[z_{t-s-1-i_1} z_{t-s-1-i_2} w_{t|p}^2 \mid \mathcal{S}_3] &= \left(\sum_{r_1=0}^{\infty} \rho^{r_1} \rho^{r_1+i_1+2s+2-i_2} O(1) \right) \left(\sum_{m_1=0}^{\infty} \psi_{m_1}^2 O(1) \right) \\ &\leq \left(\sum_{r_1=0}^{\infty} \rho^{2r_1} \right) \left(\sum_{m_1=0}^{\infty} \psi_{m_1}^2 \right) O(1) = O\left(\frac{1}{1-\rho^2}\right). \end{aligned}$$

Por tanto este tercer caso proporciona los términos de mayor orden de magnitud. De esta forma, en el caso de k general, será de interés sólo aquellas situaciones en las que las innovaciones a_t 's procedentes de las z 's coincidan por pares pero no lo hagan con las procedentes de los w 's. Si se supone primeramente que k es par y se utiliza que

$$\frac{1}{1-\rho^k} < \left(\frac{1}{1-\rho^{k-2}}\right) \left(\frac{1}{1-\rho^2}\right) < \dots < \frac{1}{(1-\rho^2)^{\frac{k}{2}}},$$

se puede verificar que los casos que proporcionan mayor orden de magnitud son aquellos en los que las innovaciones a_t 's procedentes z 's coinciden por pares pero que no existen coincidencias entre pares. Entonces

$$E[|z_{t-s-1-i_1} \cdots z_{t-s-1-i_k} w_{t|p}^k|] = O\left(\left[\frac{1}{1-\rho^2}\right]^{\frac{k}{2}}\right).$$

Análogamente, si k es impar, el término mayor se producirá si existen coincidencias en las innovaciones por pares, pero en este caso sólo se tienen $k - 1$ pares. Por tanto

$$E[|z_{t-s-1-i_1} \cdots z_{t-s-1-i_k} w_{t|p}^k|] = O\left(\left[\frac{1}{1-\rho^2}\right]^{\frac{k-1}{2}}\right).$$

Aplicando este resultado a (A.1.3) se obtiene

$$E|a|^k \leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_k=0}^{\infty} |\psi_{i_1} \cdots \psi_{i_k} (1-\rho)^k| O\left(\left[\frac{1}{1-\rho^2}\right]^{\frac{k}{2}}\right) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right).$$

Se verifica fácilmente, siguiendo los anteriores argumentos, que

$$E|b|^k = O(E|a|^k) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right).$$

Se resuelve a continuación el término $E|c|^k$.

$$\begin{aligned} E|c|^k &= E\left[\left|\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i (1-\rho)^2 z_{t-1-j} z_{t-s-1-i}\right|^k\right] \\ &\leq \sum_{j_1=0}^{\infty} \cdots \sum_{j_k=0}^{\infty} \sum_{i_1=0}^{\infty} \cdots \sum_{i_k=0}^{\infty} |\psi_{j_1} \cdots \psi_{j_k} \psi_{i_1} \cdots \psi_{i_k} (1-\rho)^{2k}| \\ &\quad \times E(|z_{t-1-j_1} \cdots z_{t-1-j_k} z_{t-s-1-i_1} \cdots z_{t-s-1-i_k}|), \end{aligned} \quad (\text{A.1.7})$$

donde

$$\begin{aligned} E(|z_{t-1-j_1} \cdots z_{t-1-j_k} z_{t-s-1-i_1} \cdots z_{t-s-1-i_k}|) &\leq \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \sum_{m_1=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} \rho^{r_1} \cdots \rho^{r_k} \rho^{m_1} \cdots \rho^{m_k} \\ &\quad \times E(|a_{t-(r_1+j_1+1)} \cdots a_{t-(r_k+j_k+1)} a_{t-(m_1+i_1+s+1)} \cdots a_{t-(m_k+i_k+s+1)}|). \end{aligned} \quad (\text{A.1.8})$$

De manera análoga a como se expuso en el análisis de $E(|a|^k)$, la magnitud de (A.1.8) viene determinada por los términos donde las innovaciones a_t 's coinciden por pares, pero los diferentes pares difieren entre si. Si k es par la magnitud de estos términos es $O((1-\rho^2)^{-k})$, mientras que si es impar dicha magnitud es $O((1-\rho^2)^{-(k-1)})$. Puesto que existe un número finito de tales términos se obtiene, utilizando (A.1.7),

$$E(|c|^k) = O\left(\frac{(1-\rho)^{2k}}{(1-\rho^2)^k}\right) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^k\right).$$

Como el orden de magnitud cuando k es impar es menor que en el caso en que es par pueden expresarse todas las situaciones con el orden de magnitud del caso par, para evitar una mayor complejidad en la notación. Por tanto se concluye que

$$E|-a - b + c|^k = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{-\frac{k}{2}}\right),$$

y el lema queda demostrado. \square

Lema A.1.2 Sean las hipótesis A1, A2, A3 and A3', con $s_0 = 2k$. Entonces cuando $T \rightarrow \infty$

$$E(\|\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}\|^k) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right).$$

Demostración:

Puede verificarse que

$$\|\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}\|^k = \|\hat{\Gamma}_w^{-1}(\hat{\Gamma}_w - \hat{\Gamma}_{w|p})\hat{\Gamma}_{w|p}^{-1}\|^k \leq \|\hat{\Gamma}_w^{-1}\|^k \|\hat{\Gamma}_w - \hat{\Gamma}_{w|p}\|^k \|\hat{\Gamma}_{w|p}^{-1}\|^k.$$

Utilizando la desigualdad de Hölder y los resultados del lema A.1.1,

$$\begin{aligned} E(\|\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}\|^k) &\leq (E(\|\hat{\Gamma}_w^{-1}\|^{4k})E(\|\hat{\Gamma}_{w|p}^{-1}\|^{4k})^{\frac{1}{4}})(E(\|\hat{\Gamma}_w - \hat{\Gamma}_{w|p}\|^{2k})^{\frac{1}{2}}) \\ &= O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right), \end{aligned}$$

y el lema queda así demostrado. \square

Lema A.1.3 Sean las hipótesis A1, A2, A3 y A3', con $s_0 = 2k$. Entonces, si $T \rightarrow \infty$

$$E(\|\hat{\phi} - \hat{\phi}_{|p}\|^k) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right), \quad (\text{A.1.9})$$

y si $s_0 = 4k$

$$E(\|\hat{\phi} - \phi\|^k) = O\left(\max\left\{\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}, T^{-\frac{k}{2}}\right\}\right). \quad (\text{A.1.10})$$

Demostración:

El estimador $\hat{\phi}$ puede expresarse como

$$\begin{aligned}\hat{\phi} &= \hat{\Gamma}_w^{-1} \hat{\gamma}_w = (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_p^{-1} + \hat{\Gamma}_p^{-1})(\hat{\gamma}_w - \hat{\gamma}_p + \hat{\gamma}_p) \\ &= (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_p^{-1})(\hat{\gamma}_w - \hat{\gamma}_{w|p}) + (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1})\hat{\gamma}_{w|p} + \hat{\Gamma}_{w|p}^{-1}(\hat{\gamma}_w - \hat{\gamma}_{w|p}) + \hat{\phi}_{|p},\end{aligned}$$

donde $\hat{\phi}_{|p} = \hat{\Gamma}_{w|p}^{-1} \hat{\gamma}_{w|p}$. Utilizando la desigualdad c_r se obtiene

$$\begin{aligned}E(\|\hat{\phi} - \hat{\phi}_{w|p}\|^k) &\leq 2^{k-1} E[\|(\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1})(\hat{\gamma}_w - \hat{\gamma}_{w|p})\|^k] \\ &\quad + 2^{k-1} \left\{ 2^{k-1} E[\|(\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1})\hat{\gamma}_{w|p}\|^k] + 2^{k-1} E[\|\hat{\Gamma}_{w|p}^{-1}(\hat{\gamma}_w - \hat{\gamma}_{w|p})\|^k] \right\}.\end{aligned}$$

Mediante la desigualdad de Hölder y aplicando el lema A.1.2 se tiene que

$$E[\|(\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1})(\hat{\gamma}_w - \hat{\gamma}_{w|p})\|^k] \leq \left\{ E(\|\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}\|^{2k}) E(\|\hat{\gamma}_w - \hat{\gamma}_{w|p}\|^{2k}) \right\}^{\frac{1}{2}} = O\left(\left[\frac{1-\rho}{1+\rho}\right]^k\right),$$

y también

$$E[\|(\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1})\hat{\gamma}_{w|p}\|^k] \leq \left\{ E(\|\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}\|^{2k}) E(\|\hat{\gamma}_{w|p}\|^{2k}) \right\}^{\frac{1}{2}} = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right).$$

Aplicando la hipótesis A3 se tiene que

$$E[\|\hat{\Gamma}_{w|p}^{-1}(\hat{\gamma}_w - \hat{\gamma}_{w|p})\|^k] \leq \left\{ E(\|\hat{\Gamma}_{w|p}^{-1}\|^{2k}) E(\|\hat{\gamma}_w - \hat{\gamma}_{w|p}\|^{2k}) \right\}^{\frac{1}{2}} = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}\right),$$

y por tanto se cumple la expresión (A.1.9). Para demostrar (A.1.10) se utilizará la descomposición

$$\hat{\phi} - \phi = \hat{\Gamma}_w^{-1} \hat{\gamma}_w - \Gamma_{w|p}^{-1} \gamma_{w|p} = \Gamma_{w|p}^{-1}(\hat{\gamma}_w - \gamma_{w|p}) + \hat{\gamma}_w(\hat{\Gamma}_w^{-1} - \Gamma_{w|p}^{-1}).$$

Entonces, mediante las desigualdades de c_r y Hölder

$$E[\|\hat{\phi} - \phi\|^k] \leq 2^{k-1} \left\{ E[\|\Gamma_{w|p}^{-1}\|^{2k}] E[\|\hat{\gamma}_w - \gamma_{w|p}\|^{2k}] \right\}^{\frac{1}{2}} + 2^{k-1} \left\{ E[\|\hat{\gamma}_w\|^{2k}] E[\|\hat{\Gamma}_w^{-1} - \Gamma_{w|p}^{-1}\|^{2k}] \right\}^{\frac{1}{2}},$$

donde, utilizando el lema A.1.1 y el lema 3.3 of Bhansali (1981) (con $\varepsilon_0 = 4k$),

$$\begin{aligned} E[\|\hat{\gamma}_w - \gamma_{w|p}\|^k] &\leq 2^{k-1} E[\|\hat{\gamma}_w - \hat{\gamma}_{w|p}\|^k] + 2^{k-1} E[\|\hat{\gamma}_{w|p} - \gamma_{w|p}\|^k] \\ &= O\left(\max\left\{\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}, T^{-\frac{k}{2}}\right\}\right), \end{aligned}$$

y, mediante el lema A.1.2,

$$\begin{aligned} E[\|\hat{\Gamma}_w^{-1} - \Gamma_{w|p}^{-1}\|^k] &\leq 2^{k-1} E[\|\hat{\Gamma}_w^{-1} - \hat{\gamma}_{w|p}^{-1}\|^k] + 2^{k-1} E[\|\hat{\Gamma}_{w|p}^{-1} - \Gamma_{w|p}^{-1}\|^k] \\ &= O\left(\max\left\{\left[\frac{1-\rho}{1+\rho}\right]^{\frac{k}{2}}, T^{-\frac{k}{2}}\right\}\right). \end{aligned}$$

□

A.2 Demostraciones de la sección 2.3

Demostración del teorema 2.2:

Utilizando (2.2.7)

$$\hat{\phi} - \phi = \hat{\Gamma}_w^{-1} \hat{\gamma}_w - \Gamma_{w|p}^{-1} \gamma_{w|p} = \Gamma_{w|p}^{-1} (\hat{\gamma}_w - \gamma_{w|p}) + \hat{\gamma}_w (\hat{\Gamma}_w^{-1} - \Gamma_{w|p}^{-1}).$$

Por la estacionariedad de $\{w_{t|p}\}$ se tiene que $\Gamma_{w|p}^{-1} = O(1)$. Del corolario 2.1 se desprende que $(\hat{\gamma}_w - \gamma_{w|p}) = O_p(T^{-1/2})$. Además, si $\hat{\Gamma}_w^{-1}$ existe, se tiene que $(\hat{\Gamma}_w^{-1} - \Gamma_{w|p}^{-1}) = \hat{\Gamma}_w^{-1} (\Gamma_{w|p} - \hat{\Gamma}_{w|p}^{-1}) \hat{\Gamma}_w^{-1} = O_p(T^{-1/2})$, donde se ha aplicado el corolario 2.1. Por tanto $\hat{\phi} - \phi = O_p(T^{-1/2})$. □

Demostración del teorema 2.3:

El estimador $\hat{\phi}$ se puede expresar como

$$\begin{aligned} \hat{\phi} &= \hat{\Gamma}_w^{-1} \hat{\gamma}_w = (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1} + \hat{\Gamma}_{w|p}^{-1}) (\hat{\gamma}_w - \hat{\gamma}_{w|p} + \hat{\gamma}_{w|p}) \\ &= (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}) (\hat{\gamma}_w - \hat{\gamma}_{w|p}) + (\hat{\Gamma}_w^{-1} - \hat{\Gamma}_{w|p}^{-1}) \hat{\gamma}_{w|p} + \hat{\Gamma}_{w|p}^{-1} (\hat{\gamma}_w - \hat{\gamma}_{w|p}) + \hat{\phi}_{|p}, \end{aligned}$$



donde $\hat{\phi}_{|p} = \hat{\Gamma}_{w|p}^{-1} \hat{\gamma}_{w|p}$. Por tanto $E(\hat{\phi}) = E(\hat{\phi}_{|p}) + O(R)$. Mediante la desigualdad de Hölder y aplicando los lemas A.1.2 y A.1.3

$$O(\|R\|) = O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{1}{2}}\right),$$

y

$$E(\hat{\phi} - \phi) = E(\hat{\phi}_{|p} - \phi) + O\left(\left[\frac{1-\rho}{1+\rho}\right]^{\frac{1}{2}}\right). \quad (\text{A.2.1})$$

En el caso del predictor correcto se tiene que (Bhansali (1981)), $E(\hat{\phi}_{|p} - \phi) = O(T^{-1})$ (si $s_0 = 16$). \square

Demostración del teorema 2.4:

Se realizará la siguiente descomposición

$$\begin{aligned} (\hat{\phi} - \phi)(\hat{\phi} - \phi)' &= (\hat{\phi} - \hat{\phi}_{|p})(\hat{\phi} - \hat{\phi}_{|p})' + 2(\hat{\phi} - \phi)(\hat{\phi}_{|p} - \phi)' \\ &\quad + (\hat{\phi}_{|p} - \phi)(\hat{\phi}_{|p} - \phi)'. \end{aligned}$$

Por tanto $E[(\hat{\phi} - \phi)(\hat{\phi} - \phi)'] = E[(\hat{\phi}_{|p} - \phi)(\hat{\phi}_{|p} - \phi)'] + O(R)$, donde

$$R = E[(\hat{\phi} - \hat{\phi}_{|p})(\hat{\phi} - \hat{\phi}_{|p})'] + 2E[(\hat{\phi} - \phi)(\hat{\phi}_{|p} - \phi)'].$$

Utilizando que la norma-2 de una matriz M es el mayor valor singular de $M'M$ entonces $\|M\| \leq \sqrt{\text{traza}(M'M)}$. Por tanto

$$\begin{aligned} E[\|(\hat{\phi} - \hat{\phi}_{|p})(\hat{\phi} - \hat{\phi}_{|p})'\|] &\leq E\left[\text{traza}\left((\hat{\phi} - \hat{\phi}_{|p})(\hat{\phi} - \hat{\phi}_{|p})'(\hat{\phi} - \hat{\phi}_{|p})(\hat{\phi} - \hat{\phi}_{|p})'\right)^{\frac{1}{2}}\right] \\ &= E[\|\hat{\phi} - \hat{\phi}_{|p}\|^2]. \end{aligned}$$

Análogamente, y utilizando la desigualdad de Hölder

$$E[\|(\hat{\phi} - \phi)(\hat{\phi}_{|p} - \phi)\|] \leq \left\{E[\|(\hat{\phi} - \phi)\|^2]E[\|(\hat{\phi}_{|p} - \phi)\|^2]\right\}^{\frac{1}{2}}.$$

Como $E(\|\hat{\phi}_{|p} - \phi\|^2) = O(T^{-1})$ (véase, por ejemplo, Bhansali 1981) y aplicando el lema A.1.3 se obtiene

$$O(\|R\|) = O\left(\max\left\{\left[\frac{1-\rho}{1+\rho}\right]^{\frac{1}{2}}, \frac{1-\rho}{1+\rho}\right\}\right),$$

□

A.3 Demostraciones de la sección 2.4

Demostración del teorema 2.5:

La expansión de Taylor de \check{A}_α^h y \check{A}_α^{h-1} alrededor de A_α^h y A_α^{h-1} , respectivamente, es

$$\check{A}_\alpha^k = A_\alpha^k + \sum_{j=0}^{k-1} A_\alpha^j (\check{A}_\alpha - A_\alpha) A_\alpha^{k-1-j} + O_p(T^{-1}), \quad k = h, h-1.$$

Entonces, utilizando que $\sum_{j=0}^{h-2} A_\alpha^j (\check{A}_\alpha - A_\alpha) A_\alpha^{h-2-j} = \sum_{j=1}^{h-1} A_\alpha^{j-1} (\check{A}_\alpha - A_\alpha) A_\alpha^{h-1-j}$,

$$\begin{aligned} \check{w}_{T+h} - w_{T+h} &= -L_1 + L_2 + e'_{p+2} A_\alpha^0 (\check{A}_\alpha - A_\alpha) A_\alpha^{h-1} Y_T \\ &+ e'_{p+2} \sum_{j=1}^{h-1} A_\alpha^{j-1} (A_\alpha - I_{p+2}) (\check{A}_\alpha - A_\alpha) A_\alpha^{h-1-j} Y_T + O_p(T^{-1}). \end{aligned} \quad (\text{A.3.1})$$

Como se verifica que $(\check{A}_\alpha - A_\alpha) = e_{p+2}(\check{\varphi} - \varphi)'$, se puede reescribir (A.3.1) como

$$\check{w}_{T+h} - w_{T+h} = (C'_{h,1} + C'_{h,2}) Y_T - L_1 + L_2 + O_p(T^{-1}),$$

donde

$$C'_{h,1} = e'_{p+2} A_\alpha^0 e_{p+2} (\check{\varphi} - \varphi)' A_\alpha^{h-1} \quad (\text{A.3.2})$$

$$C'_{h,2} = \sum_{j=1}^{h-1} e'_{p+2} A_\alpha^{j-1} (A_\alpha - I_{p+2}) e_{p+2} (\check{\varphi} - \varphi)' A_\alpha^{h-1-j}. \quad (\text{A.3.3})$$

Tomando esperanzas al cuadrado de (A.3.1) se obtiene que

$$\begin{aligned} E(\check{w}_{T+h} - w_{T+h})^2 &= E(L_1 - L_2)^2 + E(C'_{h,1} Y_T Y_T' C_{h,1}) \\ &\quad + E(C'_{h,2} Y_T Y_T' C_{h,2}) + 2E(C'_{h,1} Y_T Y_T' C_{h,2}) + o(T^{-1}). \end{aligned} \quad (\text{A.3.4})$$

El primer término a la derecha de (A.3.4) verifica que

$$\begin{aligned} L_1 - L_2 &= \sum_{k=0}^{h-1} e'_{p+2} A_\alpha^k U_{t+h-k, p+2} - \sum_{k=1}^{h-1} e'_{p+2} A_\alpha^k U_{t+h-k, p+2} \\ &= a_{T+h} + \sum_{k=1}^{h-1} e_{p+2} A_\alpha^{k-1} (A_\alpha - I_{p+2}) e_{p+2} a_{T+h-k}. \end{aligned}$$

Si se denota como $\psi_{h[\text{AR}(p+1)]}$ al coeficiente h -ésimo de $\varphi(B)^{-1}$ y como $\psi_{h[\text{ARIMA}(p+1,1,1)]}$ al coeficiente h -ésimo de $\varphi(B)^{-1}(1-B)$, se obtiene

$$\begin{aligned} e'_{p+2} A_\alpha^{h-1} (A_\alpha - I_{p+2}) e_{p+2} &= \psi_{h[\text{AR}(p+1)]} - \psi_{h-1[\text{AR}(p+1)]} \\ &= \psi_{h[\text{ARIMA}(p+1,1,1)]} = e'_{p+2} A_1^k c_{p+2}, \end{aligned}$$

y por tanto

$$L_h = L_1 - L_2 = \sum_{k=0}^{h-1} e'_{p+2} A_1^k c_{p+2} a_{T+h-k}. \quad (\text{A.3.5})$$

Se tiene entonces

$$E(L_h)^2 = \sigma^2 \sum_{k=0}^{h-1} (e_{p+2} A_1^k c_{p+2})^2.$$

Puede también verificarse que

$$(e'_{p+2} A_1^h c_{p+2}) = (e'_p A_p^h e_p) + O(1 - \rho) = (e'_p A_p^h e_p) + o(T^{-1}). \quad (\text{A.3.6})$$

Aplicando este resultado a $E(C'_{h,2} Y_T Y_T' C_{h,2})$ se obtiene

$$\begin{aligned} E(C'_{h,2} Y_T Y_T' C_{h,2}) &= \sum_{j=0}^{h-2} \sum_{k=0}^{h-2} (e'_p A_p^j e_p) (e'_p A_p^k e_p) \\ &\quad \times \text{traza} \left\{ E \left((\check{\varphi} - \varphi)' A_\alpha^{h-1-j} Y_T Y_T' A_\alpha^{h-1-k} (\check{\varphi} - \varphi) \right) \right\}. \end{aligned}$$

Además, puesto que la dependencia entre Y_T y $\check{\varphi}$ es de orden $O(T^{-\frac{3}{2}})$ (Kunitomo & Yamamoto 1985) y aplicando que

$$\text{MSE}(\check{\varphi}) = \frac{\sigma^2}{T} \Gamma_y^{-1} + O(T^{-\frac{3}{2}}),$$

se obtiene

$$\begin{aligned} E(C'_{h,2} Y_T Y_T' C_{h,2}) &= \frac{\sigma^2}{T} \sum_{j=0}^{h-2} \sum_{k=0}^{h-2} (e'_p A_p^j e_p) (e'_p A_p^k e_p) \\ &\quad \times \text{traza} \left\{ A_\alpha^{h-1-j} \Gamma_y A_\alpha^{h-1-k} \Gamma_y^{-1} \right\} + o(T^{-1}). \end{aligned} \quad (\text{A.3.7})$$

De manera análoga se puede demostrar que

$$\begin{aligned} E(C'_{h,1} Y_T Y_T' C_{h,1}) &= \frac{\sigma^2}{T} (e'_p A_p^0 e_p) (e'_p A_p^0 e_p) \\ &\quad \times \text{traza} \left(A_\alpha^{h-1} \Gamma_y A_\alpha^{h-1} \Gamma_y^{-1} \right) + o(T^{-1}), \end{aligned} \quad (\text{A.3.8})$$

donde se han conservado los términos A_α^0 para facilitar la comparación con (A.3.7). Se obtiene también que

$$\begin{aligned} E(C'_{h,1} Y_T Y_T' C_{h,2}) &= \frac{\sigma^2}{T} \sum_{j=1}^{h-1} (e'_p A_p^j e_p) (e'_p A_p^0 e_p) \\ &\quad \times \text{traza} \left(A_\alpha^{h-1-j} \Gamma_y A_\alpha^{h-1} \Gamma_y^{-1} \right) + o(T^{-1}). \end{aligned} \quad (\text{A.3.9})$$

Sumando (A.3.8), (A.3.7) y dos veces (A.3.9) se obtiene el segundo término que se encuentra en la parte derecha de (2.4.8).

Se demuestra a continuación la ecuación (2.4.9). Esta covarianza se puede desarrollar como:

$$\begin{aligned} E[(\check{w}_{T+h} - w_{T+h})(\check{w}_{T+k} - w_{T+k})] &= E(L_h L_k) + E(C'_{h,1} Y_T Y_T' C_{k,1}) + E(C'_{h,1} Y_T Y_T' C_{k,2}) \\ &\quad + E(C'_{h,2} Y_T Y_T' C_{k,1}) + E(C'_{h,2} Y_T Y_T' C_{k,2}) + o(T^{-1}), \end{aligned}$$

donde $C_{h,1}$, $C_{h,2}$, $C_{k,1}$ y $C_{k,2}$ son como en (A.3.2) y (A.3.3) y los términos L_h , L_k son como el término L_h en (A.3.5). Por tanto

$$\begin{aligned} E(L_h L_k) &= E \left[\left(\sum_{j=0}^{h-1} e'_{p+2} A_1^j c_{p+2} a_{T+h-j} \right) \left(\sum_{i=0}^{k-1} e'_{p+2} A_1^i c_{p+2} a_{T+k-i} \right) \right] \\ &= \sigma^2 \sum_{j=0}^{h-1} (e'_{p+2} A_1^j c_{p+2}) (e'_{p+2} A_1^{i+(k-h)} c_{p+2}). \end{aligned}$$

Los restantes términos se resuelven como en la demostración anterior para (2.4.8). Los valores de s_0 necesarios para la demostración de este teorema son más restrictivos que en teoremas previos para garantizar la aplicabilidad de los resultados de Yamamoto & Kunitomo (1985) sobre la dependencia entre Y_T y $\check{\varphi}$. \square

Demostración del teorema 2.6:

La esperanza del cuadrado de (2.4.12) es

$$\begin{aligned} E[(w_{T+h} - \hat{w}_{T+h})^2] &= E(L_h^2) + E \left[e'_p (\hat{A}_p^h - A_p^h) W_T W_T' (\hat{A}_p^h - A_p^h)' e_p \right] \\ &\quad + E \left[\left(\sum_{j=h-1}^{\infty} \psi_j (1 - \rho) z_{T-1-j} \right)^2 \right] + E \left[\left(\sum_{j=0}^{h-2} \psi_j (1 - \rho) \rho^{h-1-j} z_T \right)^2 \right] \\ &\quad + 2E \left[e'_p (\hat{A}_p^h - A_p^h) W_T \sum_{j=h-1}^{\infty} \psi_j (1 - \rho) z_{T-1-j} \right] \\ &\quad + 2E \left[e'_p (\hat{A}_p^h - A_p^h) W_T \sum_{j=0}^{h-2} \psi_j (1 - \rho) \rho^{h-1-j} z_T \right] \\ &\quad + 2E \left[\left(\sum_{j=h-1}^{\infty} \psi_j (1 - \rho) z_{T-1-j} \right) \left(\sum_{j=0}^{h-2} \psi_j (1 - \rho) \rho^{h-1-j} z_T \right) \right]. \end{aligned}$$

El término L_h es igual que en (A.3.5), por tanto

$$E(L_h)^2 = \sigma^2 \sum_{k=0}^{h-1} (e_{p+2} A_1^k c_{p+2})^2.$$

Aplicando (A.1.7) con $k = 1$

$$E \left[\left(\sum_{j=h-1}^{\infty} \psi_j (1 - \rho) z_{T-1-j} \right)^2 \right] = O \left(\frac{1 - \rho}{1 + \rho} \right) = o(T^{-1}).$$

Análogamente

$$E \left[\left(\sum_{j=0}^{h-2} \psi_j (1 - \rho) \rho^{h-1-j} z_T \right)^2 \right] = O \left(\frac{1 - \rho}{1 + \rho} \right) = o(T^{-1}),$$

y utilizando la desigualdad de Hölder,

$$E \left[\left(\sum_{j=h-1}^{\infty} \psi_j (1 - \rho) z_{T-1-j} \right) \left(\sum_{j=0}^{h-2} \psi_j (1 - \rho) \rho^{h-1-j} z_T \right) \right] = O \left(\frac{1 - \rho}{1 + \rho} \right) = o(T^{-1}).$$

Se utilizará a continuación una expansión de Taylor de \hat{A}_p alrededor de A_p . La magnitud del resto viene determinada por la \sqrt{T} -consistencia de \hat{A}_p

$$\begin{aligned} \hat{A}_p^h &= A_p^h + \sum_{j=0}^{h-1} A_p^j (\hat{A}_p - A_p) A_p^{h-1-j} \\ &\quad + \sum_{j=1}^{h-1} \left(\sum_{k=0}^{j-1} A_p^k (\hat{A}_p - A_p) A_p^{j-1-k} \right) \times (\hat{A}_p - A_p) A_p^{h-i-j} + O_p(T^{-\frac{3}{2}}). \end{aligned}$$

Por tanto,

$$e_p' (\hat{A}_p^h - A_p^h) W_T = (B'_{h,1} + B'_{h,2}) W_T, \quad (\text{A.3.10})$$

donde

$$\begin{aligned} B'_{1h,p} &= e_p' \sum_{j=0}^{h-1} A_p^j (\hat{A}_p - A_p) A_p^{h-1-j} \\ B'_{2h,p} &= e_p' \sum_{j=1}^{h-1} \left(\sum_{k=0}^{j-1} A_p^k (\hat{A}_p - A_p) A_p^{j-1-k} \right) \times (\hat{A}_p - A_p) A_p^{h-i-j}. \end{aligned}$$

Tomando esperanzas en el cuadrado de (A.3.10)

$$\begin{aligned} E[e_p' (\hat{A}_p^h - A_p^h) W_T W_T' (\hat{A}_p^h - A_p^h)' e_p] &= E(B'_{h,1} W_T W_T' B_{h,1}) + E(B'_{h,2} W_T W_T' B_{h,2}) \\ &\quad + 2E(B'_{h,1} W_T W_T' B_{h,2}). \end{aligned}$$

Aplicando el lema A.1.3 se obtiene

$$E(\|\hat{A}_p - A_p\|^k) \leq E(\|\hat{A}_p - A_p\|^2)^{\frac{k}{2}} = O\left(\left[\frac{1-\rho}{1+\rho}\right]^k 2\right) = o(T^{-\frac{k}{2}}). \quad (\text{A.3.11})$$

De esta forma, aplicando a la desigualdad de Hölder se obtiene

$$E(B'_{h,2} W_T W_T' B_{h,2}) = o(T^{-2}),$$

$$E(B'_{h,1} W_T W_T' B_{h,2}) = o(T^{-\frac{3}{2}}).$$

Análogamente se verifica que

$$E[e'_p(\hat{A}_p^h - A_p^h) W_T \sum_{j=h-1}^{\infty} \psi_j(1-\rho) z_{T-1-j}] = o(T^{-1}),$$

$$E[e'_p(\hat{A}_p^h - A_p^h) W_T \sum_{j=0}^{h-2} \psi_j(1-\rho) \rho^{h-1-j} z_T] = o(T^{-1}).$$

Además

$$\begin{aligned} E(B'_{h,1} W_T W_T' B_{h,1}) &= \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} (e'_p A_p^j e_p)(e'_p A_p^k e_p) \\ &\quad \times \text{traza} \left\{ E \left((\hat{\phi} - \phi)(\hat{\phi} - \phi)' A_p^{h-1-j} W_T W_T' A_p^{h-1-k} \right) \right\}. \end{aligned}$$

Aplicando el teorema 2.4 y la desigualdad de Hölder

$$\begin{aligned} E \left((\hat{\phi} - \phi)(\hat{\phi} - \phi)' A_p^{h-1-j} W_T W_T' A_p^{h-1-k} \right) &= \\ E \left((\hat{\phi}_{|p} - \phi)(\hat{\phi}_{|p} - \phi)' A_p^{h-1-j} W_T W_T' A_p^{h-1-k} \right) &= \\ + o(T^{-1}). \end{aligned}$$

Puede ahora aplicarse que la dependencia entre $\hat{\phi}_{|p}$ y W_T es de magnitud $O(T^{-\frac{3}{2}})$ (Kunitomo & Yamamoto (1985)) y que

$$\text{MSE}(\hat{\phi}_{|p}) = \frac{\sigma^2}{T-1} \Gamma_w^{-1} + O(T^{-3/2}),$$

Se obtiene entonces,

$$E \left((\hat{\phi} - \phi)(\hat{\phi} - \phi)' A_p^{h-1-j} W_T W_T' A_p'^{h-1-k} \right) = \frac{\sigma^2}{T-1} (A_p^{h-1-j} \Gamma_w A_p'^{h-1-k} \Gamma_w^{-1}) + o(T^{-1}).$$

Se concluye la demostración aplicando que $\sigma^2/(T-1) = \sigma^2/T + O(T^{-2})$. \square

La demostración de (2.4.14) es una aplicación directa de esta demostración y la demostración de (2.4.9). Los valores de s_0 necesarios en este teorema son, de nuevo, más restrictivos que en anteriores teoremas para preservar la aplicabilidad de los resultados de Yamamoto & Kunitomo (1985). \square

A.4 Demostraciones de la sección 2.5

Demostración del lema 2.1:

El vector Y_t admite la descomposición $Y_t = (\tilde{Y}_t', 0)' + \mu$, donde $\mu = (\mu, \mu, \dots, \mu, 1)'$. Como $\alpha = \mu - \sum_{i=1}^{p+1} \varphi_i \mu$ se verifica que $A_\alpha^i \mu A_\alpha'^j = \bar{\mu}$, donde $\bar{\mu} = \mu \mu'$. Entonces $A_\alpha^i \Gamma_y A_\alpha'^j = A_\alpha^i \Gamma_y^* A_\alpha'^j + \bar{\mu}$, donde Γ_y^* es una matriz $(p+2) \times (p+2)$ con Γ_y en la primera submatriz $(p+1) \times (p+1)$ y ceros en el resto. Además, la matriz de covarianzas Γ_y tiene la siguiente estructura por bloques

$$\Gamma_y = \begin{pmatrix} \Gamma_o & \mu_o \\ \mu_o' & 1 \end{pmatrix},$$

donde $\Gamma_o = E(Y_{ot} Y_{ot}')$, with $Y_{ot} = (y_t, y_{t-1}, \dots, y_{t-p})'$ y $\mu_o = E(Y_{ot})$. Utilizando las propiedades de las inversas de matrices en bloques se puede particionar Γ_y^{-1} como

$$\Gamma_y^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

donde $B_{11} = [\Gamma_o - \mu_o \mu_o']^{-1} = \Gamma_y^{-1}$. Puede entonces verificarse fácilmente que

$$\text{traza}(A_\alpha^i \Gamma_y^* A_\alpha^j \Gamma_y^{-1}) = \text{traza}(A_o^i \Gamma_y A_o^j \Gamma_y^{-1})$$

y entonces

$$\text{traza}(A_\alpha^i \Gamma_y A_\alpha^j \Gamma_y^{-1}) = \text{traza}(A_o^i \Gamma_y A_o^j \Gamma_y^{-1}) + \text{traza}(\mu \Gamma_y^{-1}).$$

Como se verifica que $\text{traza}(\bar{\mu} \Gamma_y^{-1}) = \mu' \Gamma_y^{-1} \mu$ y aplicando Searle (1984, pag 258) puede verse que

$$\mu' \Gamma_y^{-1} \mu = 1 - |\Gamma_y - \mu \mu'| / |\Gamma_y| = 1,$$

donde se ha aplicado que la última columna y fila de $\Gamma_y - \mu \mu'$ son ceros y la matriz Γ_y es invertible. \square

Demostración del lema 2.2:

En esta demostración se utilizan argumentos similares a los empleados en Box & Tiao (1977), donde los autores descomponen un proceso vectorial casi no estacionario en dos partes, la primera de ellas sigue un proceso estacionario y la segunda esta cerca de la no estacionariedad.

Sea C la siguiente matriz no singular

$$C = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\rho \\ 1 & -\phi_1 & -\phi_2 & \cdots & -\phi_{p-1} & -\phi_p \end{pmatrix}.$$

Entonces

$$CA_oC^{-1} = \begin{pmatrix} A_p & 0 \\ 0 & \rho \end{pmatrix} = D.$$

La no singularidad de C puede verse, por ejemplo, en Pearl (1973, p. 156). Puede entonces verificarse que

$$(CA_oC^{-1})^i = CA_o^iC^{-1} = \begin{pmatrix} A_p^i & 0 \\ 0 & \rho^i \end{pmatrix} = D^i. \quad (\text{A.4.1})$$

Sea λ_k un valor propio de la matriz $Q = \Gamma_{\bar{y}}^{-1}A_o^i\Gamma_{\bar{y}}A_o'^j$. Entonces λ_k es una raíz del polinomio del determinante

$$|Q - \lambda I| = |A_o^i\Gamma_{\bar{y}}A_o'^j - \lambda\Gamma_{\bar{y}}| = 0.$$

Utilizando (A.4.1) esta expresión es equivalente a

$$|D^i\Gamma_C D'^j - \lambda\Gamma_C| = 0, \quad (\text{A.4.2})$$

donde $\Gamma_C = C\Gamma_{\bar{y}}C'$. Esta matriz Γ_C puede ser considerada como la matriz de covarianzas de la transformación $Z_t = CY_t$, donde $Z_t = (z_{1,t}, z_{1,t-1}, \dots, z_{1,t-p+1}, z_{2,t})'$ y

$$Z_t = DZ_{t-1} + a_t c_{p+1}, \quad (\text{A.4.3})$$

con $c_{p+1} = (1, 0, \dots, 0, 1)'$. Por tanto, la primera submatriz $p \times p$ de Γ_C es la matriz de covarianzas de un proceso que sigue la matriz de coeficientes A_p y ruido a_t , y que se denotará como $\Gamma_{w|p}$. Si se denomina V_{12} , V_{21} y V_{22} a las otras submatrices de esta partición se puede reescribir (A.4.2) como

$$\begin{vmatrix} (A_p^i\Gamma_{w|p}A_p'^j - \lambda\Gamma_{w|p}) & (A_p^iV_{12}\rho^j - \lambda V_{12}) \\ (\rho^iV_{21}A_p'^j - \lambda V_{21}) & (\rho^{i+j}V_{22} - \lambda V_{22}) \end{vmatrix} = 0,$$

que puede también expresarse como

$$\begin{vmatrix} (A_p^i \Gamma_{w|p} A_p^{j'} - \lambda \Gamma_{w|p}) & (A_p^i V_{12} \rho^j - \lambda V_{12})(V_{22}^{-\frac{1}{2}}) \\ (\rho^i V_{21} A_p^{j'} - \lambda V_{21})(V_{22}^{-\frac{1}{2}}) & (\rho^{i+j} - \lambda) \end{vmatrix} = 0.$$

Siguiendo (A.4.3) el término V_{22} es la varianza de un proceso AR(1) con coeficiente ρ y por tanto $V_{22}^{-1} = O(1 - \rho)$. Por consiguiente, siguiendo las propiedades de los determinantes de matrices particionadas (véase, por ejemplo, Searle 1984)

$$|Q - \lambda I| = |A_p^i \Gamma_{w|p} A_p^{j'} - \lambda \Gamma_{w|p}| (\rho^{i+j} + O(1 - \rho) - \lambda) = 0.$$

Como la traza de una matriz es igual a la suma de sus valores propios se cumple el resultado. \square

Demostración del teorema 2.8:

Si $\beta = 1 - v$, con $v \geq 0$, no se puede realizar la sustitución de los términos $(e'_{p+2} A_1^v e_{p+2})$ por $(e'_p A_p^v e_p)$, como se hizo en (A.3.6). Esto hace que ambos ECMP incrementen el orden de magnitud de la diferencia que existe entre ellos. Además, la mayoría de los argumentos empleados en la demostración del teorema 2.6 se basa en la hipótesis de que el término $(1 - \rho)(1 + \rho)^{-1}$ es $o(T^{-1})$, lo cual es cierto si $\beta > 1$. Esto introduce un término de error en las expresiones (2.4.13) y (2.4.14) de magnitud $O(T^{-1+v})$. El lema 2.2 precisa igualmente de $\beta < 1$ para poder ser aplicado, en caso contrario el término de error es de mayor orden de magnitud. \square

Capítulo 3

Contrastes de raíces unitarias en series temporales.

Una revisión comparada

3.1 Introducción

En este capítulo se introduce el problema de contrastar la existencia de una raíz unitaria en el polinomio autorregresivo de una serie temporal. En el capítulo 1 se enumeraron tres aspectos en los que es importante saber si existen raíces unitarias: las relaciones de cointegración, la predicción y la modelización. También se mencionó la dificultad de dicha detección en zonas próximas al círculo unidad. En el capítulo 2 se demostró que existe una región próxima al círculo unidad (dependiente del tamaño muestral) en la cuál suponer que el proceso es $I(1)$ cuando el proceso verdadero es $I(0)$ no tiene repercusiones negativas en las predicciones. Este resultado disminuye la importancia de realizar un contraste si el objetivo es la predicción. No obstante, se puede desear contrastar la estacionariedad como contraste de una hipótesis de teoría económica. El resultado del contraste será si existe o no evidencia suficiente para considerar que cierta variable evoluciona, en el largo plazo, alrededor de un valor medio. Dicha variable puede ser una serie observada, su d -ésima diferencia o el término de error de una relación de cointegración.

Existen numerosas propuestas para realizar contrastes de raíces unitarias. El motivo de dicha proliferación está en que, de acuerdo con el lema de Neyman-Pearson, no existe un contraste que sea uniformemente más potente (UMP). Otra razón que ha contribuido a la existencia de abundantes referencias sobre raíces unitarias es que, en la construcción de dichos contrastes, hay que tomar una serie de decisiones que influyen en el resultado final. Por ejemplo, la decisión del método de estimación, o del modelo que se empleará para aproximar la estructura dinámica de la serie, influyen enormemente en la potencia de un mismo contraste.

La principal dificultad en la detección de una raíz unitaria reside en la continuidad de las propiedades de los procesos en $\rho = 1$. Esta continuidad se hace aún más evidente en las propiedades muestrales de una realización finita, de manera que existe una *equivalencia observacional* entre ciertos procesos estacionarios y no estacionarios (véase, entre otros, Campbell & Perron (1991, p. 157), Blough (1992) y Hamilton (1994, p. 515)). Por ejemplo, dada una muestra del proceso con raíz unitaria

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

puede encontrarse un proceso estacionario y_t^* con parámetro $\rho = 1 - \epsilon$, con ϵ positivo suficientemente pequeño, cuyos momentos muestrales sean indistinguibles de los de y_t . Análogamente, puede encontrarse un proceso no estacionario que sea indistinguible en la práctica de uno estacionario. Por ejemplo sea una muestra finita del proceso

$$(1 - B)y_t = (1 - \theta B)a_t$$

con $|\theta| < 1$. Este proceso es no estacionario. Sin embargo, si θ está suficientemente cerca de la unidad y_t será indistinguible de un ruido blanco. Esta dificultad de detección de una raíz unitaria puede extenderse también para el contraste de cualquier valor estacionario. Gonzalo & Lee (1994) muestran que los contrastes basados en el estadístico t para raíces

fuera del círculo unidad, en procesos autorregresivos, tienen una potencia tan baja, e incluso menor, que para raíces unitarias.

En este capítulo se revisan los contrastes existentes en la literatura. La estructura del capítulo es la siguiente. La sección 3.2 introduce el modelo empleado y describe los tipos de no estacionariedad que se van a analizar. La sección 3.3 introduce los conceptos de inferencia que se utilizan en los contrastes de raíces unitarias. La sección 3.4 hace una revisión de los contrastes que se encuentran en la literatura. En esta sección se describe con más detalle sólo los tres contrastes que, a juicio del autor, son más relevantes, con el fin de tener un marco adecuado donde comparar los contrastes que se proponen en el capítulo 4.

3.2 Series temporales no estacionarias

Sea y_t una serie temporal de la que se observa una realización y_1, y_2, \dots, y_T . Se considerará que esta serie, en su expresión más general, puede tener parte estocástica x_t así como parte determinista v_t ,

$$y_t = v_t + x_t.$$

Como componente determinista se admitirá media no nula μ y tendencia lineal. En general resulta difícil explicar la presencia de una tendencia lineal en una serie económica real. No obstante, la mayoría de los contrastes que se encuentran en la literatura contemplan la posibilidad de su detección. Por esta razón se estudiará este tipo de componente. El componente determinista podría extenderse al caso de un polinomio de cualquier orden, pero su utilidad práctica sería muy limitada. Por tanto, se considerará

$$v_t = \mu + \delta t.$$

La componente estocástica puede, a su vez, descomponerse en la parte asociada a la raíz más próxima al círculo unitario ρ^{-1} y un término estacionario e invertible u_t que sigue un proceso ARMA(p, q),

$$\begin{aligned}x_t &= \rho x_{t-1} + u_t \\ \phi(B)u_t &= \theta(B)a_t\end{aligned}$$

donde a_t es una secuencia de variables aleatorias iid que verifica $E(a_t) = 0$ y $E(a_t^2) = \sigma^2$. El proceso generador de datos (PGD) es, por tanto,

$$y_t = \mu + \delta t + x_t, \quad x_t = \rho x_{t-1} + u_t, \quad u_t = \psi(B)a_t, \quad (3.2.1)$$

donde $\psi(B) = \phi(B)^{-1}\theta(B)$. Una representación alternativa de (3.2.1) es

$$y_t = \mu(1 - \rho) + \delta\rho + \delta(1 - \rho)t + \rho y_{t-1} + \psi(B)a_t \quad (3.2.2)$$

En el modelo (3.2.2) la no estacionariedad puede ser debida a tres causas:

1. presencia de tendencia lineal determinista,
2. raíz unitaria
3. ambas simultáneamente.

Encontrar métodos que permitan distinguir entre una serie temporal estacionaria y estos tres tipos de no estacionariedad es lo que se conoce como el problema de detección de raíces unitarias.

El primer tipo de no estacionariedad considerado es el caso de parte estocástica estacionaria ($\rho < 1$) y parte determinista con tendencia lineal ($\delta \neq 0$). La representación de este proceso es

$$y_t = \mu + \delta t + \psi^*(B)a_t, \quad (3.2.3)$$

donde $\psi^*(B) = (1 - \rho B)^{-1} \psi(B)$. La media de este proceso es una tendencia lineal de pendiente δ . Alrededor de esta tendencia el proceso evoluciona de forma estacionaria según la forma de $\psi^*(B)$. Por esta razón se le denomina frecuentemente *tendencia-estacionario* (en inglés *trend-stationary*) pues se convierte en estacionario al sustraerle la tendencia.

El segundo tipo de no estacionariedad se obtiene en procesos sin tendencia determinista pero con raíz unitaria. Haciendo $\rho = 1$ y $\delta = 0$ en (3.2.2) resulta

$$(1 - B)y_t = \psi(B)a_t, \quad (3.2.4)$$

que incrementos estacionarios de media nula. En el caso particular de $\psi(B)$ igual a 1 se tiene un paseo aleatorio sin deriva.

El tercer tipo considerado de no estacionariedad corresponde a la unión de una raíz unitaria y una tendencia determinista, que equivale a tener los valores $\rho = 1$ y $\delta \neq 0$ en (3.2.2). El modelo que resulta es

$$(1 - B)(y_t - \mu - \delta t) = \psi(B)a_t,$$

o también,

$$(1 - B)y_t = \delta + \psi(B)a_t, \quad (3.2.5)$$

que en el caso particular en que $\psi(B)$ es igual a 1 es un paseo aleatorio con deriva.

Cualquiera de estos tres tipos de procesos no estacionarios puede reducirse a estacionario aplicando una diferencia a y_t . El problema surge si la no estacionariedad es debida sólo a una tendencia determinista. Si se toma una diferencia en (3.2.3) se obtiene el proceso no invertible

$$(1 - B)y_t = \delta + (1 - B)\psi^*(B)a_t.$$

Por tanto, si se desea hacer estacionario e invertible al proceso (3.2.3) habrá que sustraer la tendencia en lugar de tomar una diferencia. Esta es una de las razones por la que no es suficiente con saber que un proceso es no estacionario, sino que es importante distinguir entre no estacionariedad debida a tendencia determinista o a presencia de una raíz unitaria.

Otra razón que aconseja la distinción entre los dos tipos de no estacionariedad: tendencia determinista o raíz unitaria, es por las distintas implicaciones de sus predicciones a largo plazo. La predicción del proceso con tendencia determinista (3.2.3) es

$$\hat{y}_{t+h} = E(y_{t+h}|y_t) = \mu + \delta(t+h) + \psi_h^* a_t + \psi_{h+1}^* a_{t-1} + \dots \quad (3.2.6)$$

Para h suficientemente grande esta predicción es una línea recta independientemente del origen de la predicción. Esta característica hace que sean poco factibles en series reales.

Para el segundo caso, proceso con raíz unitaria, la predicción es

$$\hat{y}_{t+h} = y_t + (\psi_1 + \psi_2 + \dots + \psi_s) a_t + (\psi_2 + \psi_3 + \dots + \psi_{s+1}) a_{t+1} + \dots \quad (3.2.7)$$

En este caso la predicción a largo plazo cambia con cada origen considerado.

Otro aspecto que diferencia a los procesos con tendencia determinista y los procesos con raíz unitaria es la varianza de sus errores de predicción. La varianza de los errores de predicción de un proceso con tendencia determinista, suponiendo conocidos los parámetros, puede obtenerse a partir de (3.2.6), resultando

$$E(y_{t+h} - \hat{y}_{t+h})^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^{*2}, \quad (3.2.8)$$

que, con $\rho < 1$, corresponde con la varianza de los errores de predicción de un proceso estacionario y está, por tanto, acotada. Sin embargo la varianza de los errores de predicción

en presencia de raíz unitaria (con o sin tendencia determinista) es

$$E(y_{t+h} - \hat{y}_{t+h})^2 = \sigma^2 \sum_{j=0}^{h-1} \left(\sum_{i=0}^j \psi_j \right)^2 \quad (3.2.9)$$

que aumenta en cada horizonte de predicción sin alcanzar ninguna cota.

3.3 Introducción a la inferencia en presencia de raíz unitaria

En esta sección se introducen los elementos de inferencia que se presentan en el problema de detección de raíces unitarias y que difieren de la inferencia en procesos estacionarios. Se considerará inicialmente el caso sencillo de $\mu = \delta = 0$ y $\psi(B) = 1$ con innovaciones normales,

$$y_t = \rho y_{t-1} + a_t, \quad (3.3.1)$$

con $y_0 = 0$ y por tanto $y_1 \sim N(0, \sigma^2)$. Dadas T observaciones y_1, y_2, \dots, y_T , el estimador de mínimos cuadrados (MCO) de ρ es

$$\hat{\rho} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2}. \quad (3.3.2)$$

Rubin (1950) demostró que $\hat{\rho}$ es un estimador consistente para todo valor de ρ . En el caso $|\rho| < 1$ Mann & Wald (1943) demostraron que este estimador es, además, asintóticamente normal (en condiciones algo más generales que las aquí expuestas) y por tanto

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

donde el símbolo \xrightarrow{d} significa convergencia en distribución. La dificultad del caso $|\rho| = 1$ estriba en que $\sqrt{T}(\hat{\rho} - 1)$ converge a cero en probabilidad y, por tanto, no es un estadístico útil para la realización de contrastes asintóticos. Sin embargo, $T(\hat{\rho} - 1)$ sí converge a una variable aleatoria, aunque de distribución no estandar. Por tener $\hat{\rho}$, en este caso,

un orden de convergencia $O(T^{-1})$ en lugar de $O(T^{-1/2})$, se dice que el estimador es *superconsistente*. White (1958) demostró que esta distribución límite puede expresarse como un funcional de movimientos Brownianos aplicando el teorema de Donsker (Donsker, 1951) a la sumas parciales de las innovaciones a_t . Donsker (1951) demuestra que la distribución límite de cualquier función de sumas de variables aleatorias iid con varianzas finitas es la distribución del correspondiente funcional del proceso de Wiener. Para aplicar este resultado se define el proceso de sumas parciales S_t como

$$S_t = \sum_{i=1}^t a_i.$$

El interés de estas sumas parciales se encuentra en que, haciendo $\rho = 1$ en (3.3.1), $y_t = S_t$.

Para estudiar la distribución asintótica de estas sumas parciales se define el proceso $X_T(r)$, con $r \in [0, 1]$ de la siguiente manera

$$X_T(r) = (1/T)S_{[Tr]}, \quad (3.3.3)$$

donde $[Tr]$ es el mayor entero que es menor o igual a Tr . La variable aleatoria $X_T(r)$ es una función en escalones de r

$$X_T(r) = \begin{cases} 0 & 0 \leq r < 1/T \\ a_1/T & 1/T \leq r < 2/T \\ (a_1 + a_2)/T & 2/T \leq r < 3/T \\ \vdots & \\ (\sum_{t=1}^T a_t)/T & r = 1. \end{cases}$$

La estandarización $\sqrt{T}X_T(r)/\sigma$, para un r fijo, es un elemento aleatorio en el espacio de funciones D , el espacio de las funciones reales en $[0, 1]$ que son continuas por la derecha en cada punto de $[0, 1]$ y tienen límites finitos por la izquierda. Por tanto, en el espacio D pueden existir escalones o discontinuidades de primer orden (para una discusión más

profunda véase Billingsley (1968, sección 14)). Bajo ciertas condiciones, que cumplen las innovaciones a_t , se demuestra que

$$\sqrt{T}X_T(r)/\sigma \xrightarrow{d} W(r), \quad (3.3.4)$$

donde $W(r)$ es un movimiento Browniano estandar o proceso de Wiener. Además, para un r fijo, $W(r) \sim N(0, r)$. Este resultado se conoce como el *teorema central del límite de funcionales* (TCLF) y se verifica en condiciones más generales de las aquí expuestas. Por ejemplo, a_t podría presentar heterocedasticidad condicionada y cierto grado de dependencia temporal, pudiendo seguir un proceso ARMA (véase Herrndorf (1984) y Phillips (1987a, p. 280)). Aplicando este resultado para $r = 1$ se obtiene el clásico resultado

$$\sqrt{T}X_T(1)/\sigma = \frac{\sum_{t=1}^T a_t}{\sigma\sqrt{T}} \xrightarrow{d} W(1).$$

Otro resultado importante que, junto con el TCLF, es necesario aplicar en la inferencia de procesos con raíz unitaria es el llamado *teorema de la aplicación continua* (TAC) (véase, por ejemplo, Billingsley (1968, teorema 5.1)) según el cual si $\sqrt{T}X_T(r)/\sigma \xrightarrow{d} W(r)$, y $g(\cdot)$ es un funcional continuo en el espacio D , entonces

$$g\left(\sqrt{T}X_T(r)\right)/\sigma \xrightarrow{d} g(W(r)). \quad (3.3.5)$$

Ejemplos de funciones $g(\cdot)$ son $\int_0^1 X_T(r)dr$, $\int_0^1 X_T^2(r)dr$ o multiplicación por una constante. La aplicación del TCLF y del TAC permite encontrar las distribuciones asintóticas de los diferentes momentos muestrales que intervienen en los contrastes. En la siguiente proposición se recopilan los resultados básicos que son de frecuente aplicación tanto en este capítulo como en el capítulo 4. La demostración puede encontrarse en los trabajos de Phillips (1987a), Phillips & Perron (1988) y Chan & Wei (1988) entre otros.

Proposición 3.1 Sea y_t el siguiente proceso

$$y_t = y_{t-1} + a_t$$

donde $y_0 = 0$, a_t es una secuencia iid con media cero y varianza σ^2 . Entonces

$$\begin{aligned} T^{-1/2} \sum_{t=1}^T a_t &\xrightarrow{d} \sigma W(1) && \sim N(0, \sigma^2) \\ T^{-3/2} \sum_{t=1}^T y_{t-1} &\xrightarrow{d} \sigma \int_0^1 W(r) dr && \sim N(0, \frac{1}{3} \sigma^2) \\ T^{-2} \sum_{t=1}^T y_{t-1}^2 &\xrightarrow{d} \sigma^2 \int_0^1 W^2(r) dr \\ T^{-\frac{k+2}{2}} \sum_{t=1}^T y_{t-1}^k &\xrightarrow{d} \sigma^k \int_0^1 W^k(r) dr \\ T^{-1} \sum_{t=1}^T y_{t-1} a_t &\xrightarrow{d} \frac{\sigma^2}{2} (W^2(1) - 1) && \sim \frac{1}{2} (\chi_1^2 - 1) \\ T^{-3/2} \sum_{t=1}^T t a_t &\xrightarrow{d} \sigma W(1) - \sigma \int_0^1 W(r) dr && \sim N(0, \frac{1}{3} \sigma^2) \\ T^{-5/2} \sum_{t=1}^T t y_{t-1} &\xrightarrow{d} \sigma \int_0^1 r W(r) dr && \sim N(0, \frac{2}{15} \sigma^2) \\ T^{-3} \sum_{t=1}^T t y_{t-1}^2 &\xrightarrow{d} \sigma^2 \int_0^1 r W^2(r) dr \end{aligned}$$

Aplicando estos resultados al estimador $\hat{\rho}$ definido en (3.3.2) se obtiene, si $\rho = 1$,

$$T(\hat{\rho} - 1) = \frac{T^{-1} \sum_{t=1}^T y_{t-1} a_t}{T^{-2} \sum_{t=1}^T y_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2} [W^2(1) - 1]}{\int_0^1 W^2(r) dr}. \quad (3.3.6)$$

Se denotará por $\hat{\tau}$ al estadístico pivote t de regresión. Su distribución límite, si $\rho = 1$, es

$$\hat{\tau} = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}} \xrightarrow{d} \frac{\frac{1}{2} [W^2(1) - 1]}{\left\{ \int_0^1 W^2(r) dr \right\}^{1/2}}. \quad (3.3.7)$$

donde $\hat{\sigma}_{\hat{\rho}}$ es el estimador MCO del error estandar de los coeficientes estimados, es decir

$$\hat{\sigma}_{\hat{\rho}} = \left\{ \frac{\sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2}{(T-1) \sum_{t=1}^T y_{t-1}^2} \right\}^{\frac{1}{2}}$$

El modelo (3.3.1) supone que, bajo la hipótesis nula $H_0 : \rho = 1$, el proceso es un paseo aleatorio mientras que bajo la alternativa $H_1 : \rho < 1$ es un proceso estacionario de media nula. Si se desea contemplar la situación más general de que bajo la alternativa el proceso pueda tener media no nula se tendrá el modelo, particularizando (3.2.2) al caso $\psi(B) = 1$, $\delta = 0$,

$$y_t = \mu(1 - \rho) + \rho y_{t-1} + a_t. \quad (3.3.8)$$

En este modelo, se sigue obteniendo un paseo aleatorio bajo la hipótesis nula de raíz unitaria. El estimador por MCO de ρ en este modelo, $\hat{\rho}^\mu$, tiene la expresión,

$$\hat{\rho}^\mu = \frac{\sum_{t=2}^T (y_t - \bar{y}_{(0)})(y_{t-1} - \bar{y}_{(-1)})}{\sum_{t=2}^T (y_{t-1} - \bar{y}_{(-1)})^2} \quad (3.3.9)$$

donde $\bar{y}_{(i)} = (T-1)^{-1} \sum_{t=2}^T y_{t+i}$. La distribución del estadístico $T(\hat{\rho}^\mu - 1)$ es, si $\rho = 1$,

$$T(\hat{\rho}^\mu - 1) \xrightarrow{d} \frac{\frac{1}{2}[W^2(1) - 1] - W(1) \int_0^1 W(r) dr}{\int_0^1 W^2(r) dr - \left[\int_0^1 W(r) dr \right]^2}, \quad (3.3.10)$$

y para el estadístico pivote t , que se denotará por $\hat{\tau}^\mu$,

$$\hat{\tau}^\mu = \frac{\hat{\rho}^\mu - 1}{\hat{\sigma}_{\hat{\rho}^\mu}} \xrightarrow{d} \frac{\frac{1}{2}[W^2(1) - 1] - W(1) \int_0^1 W(r) dr}{\left\{ \int_0^1 W^2(r) dr - \left[\int_0^1 W(r) dr \right]^2 \right\}^{1/2}}, \quad (3.3.11)$$

que difieren de (3.3.6) y (3.3.7). Nótese que estas distribuciones asintóticas no dependen de μ .

Análogamente, un modelo que bajo la alternativa pueda tener pendiente lineal determinista será, particularizando (3.2.2) para $\psi(B) = 1$,

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + a_t, \quad (3.3.12)$$

donde

$$\beta_0 = (1 - \rho)\mu + \delta\rho, \quad \beta_1 = \delta(1 - \rho).$$

Si se considera que bajo la hipótesis nula el proceso es paseo aleatorio, $H_0 : \rho = 1, \delta = 0$, el estimador MCO de ρ , $\hat{\rho}^\tau$ del modelo (3.3.12) tiene la distribución límite (Phillips & Perron, 1988)

$$T(\hat{\rho}^\tau - 1) \xrightarrow{d} \frac{\frac{1}{2}[W_\tau^2(1) - W_\tau^2(0)]}{\int_0^1 W_\tau^2(r) dr}, \quad (3.3.13)$$

donde

$$W_\tau(r) = W(r) - (4 - 6r) \int_0^1 W(s) ds - (12r - 6) \int_0^1 sW(s) ds,$$

y para el estadístico t , que se denotará por $\hat{\tau}$,

$$\hat{\tau} = \frac{\hat{\rho}^\tau - 1}{\hat{\sigma}_{\hat{\rho}^\tau}} \xrightarrow{d} \frac{\frac{1}{2} [W_\tau^2(1) - W_\tau^2(0)]}{\left\{ \int_0^1 W_\tau^2(r) dr \right\}^{1/2}}. \quad (3.3.14)$$

Finalmente se considera el caso en que el término de error en (3.3.1) sea u_t en lugar de a_t , donde, de acuerdo al modelo definido en (3.2.2), tiene que

$$u_t = \psi(B)a_t.$$

Para el estudio de este caso es de gran utilidad la aplicación de la descomposición de Beveridge & Nelson (1981). Estos autores encuentran una descomposición de $x_t = x_0 + u_1 + u_2 + \dots + u_t$ en parte no estacionaria más parte estacionaria. Para ello reescriben $\psi(B) = \psi(1) + \psi^*(B)(1 - B)$, donde $\psi^*(B) = -\sum_{i=j+1}^{\infty} \psi_i$. Entonces

$$x_t = \psi(1) \sum_{j=1}^t a_t + \psi^*(B)a_t + \tilde{x}_0, \quad (3.3.15)$$

donde $\tilde{x}_0 = x_0 - \psi^*(B)a_0$ representa las condiciones iniciales y el término $\psi^*(B)a_t$ es estacionario. La aplicación de esta descomposición junto con el TCLF al proceso de sumas parciales

$$X_T(r) = (1/T)S_{[Tr]}, \quad S_t = \sum_{i=0}^t u_i, \quad (3.3.16)$$

lleva a

$$\sqrt{T}X_T(r) \xrightarrow{d} \sigma\psi(1)W(r), \quad (3.3.17)$$

que es $\psi(1)$ veces el resultado (3.3.4). Este resultado permite generalizar fácilmente los resultados obtenidos para el caso $\psi(B) = 1$.

3.4 Contrastes de raíces unitarias

En esta sección se describen los contrastes de raíces unitarias más importantes que se encuentran en la literatura. El apartado 3.4.1 introduce el problema de la construcción

de estos contrastes. Los apartados 3.4.2 a 3.4.4 están dedicados a los tres contrastes más significativos que se encuentran publicados. Estos contrastes son:

1. contrastes de Dickey-Fuller (Dickey & Fuller, 1979),
2. contrastes de Sargan-Bhargava (Sargan & Bhargava (1983), Bhargava (1986)),
3. contrastes de Elliot-Rothenberg-Stock (Elliot, Rothenberg & Stock 1992).

En el apartado 3.4.5 se comparan sus potencias en un ejercicio de simulación. Finalmente en el apartado 3.4.6 se describen otros contrastes que se encuentran en la literatura.

El modelo sobre el que se realizan los contrastes es el modelo general (3.2.2) y la exposición se centra en las siguientes hipótesis a contrastar

- Contraste 1: Bajo la hipótesis nula el proceso tiene una raíz unitaria, bajo la alternativa el proceso es estacionario de media nula: $H_0 : \rho = 1, H_1 : \rho < 1, \mu = \delta = 0$.
- Contraste 2: Bajo la hipótesis nula el proceso tiene raíz unitaria y bajo la alternativa es estacionario con media no necesariamente nula: $H_0 : \rho = 1, H_{1\mu} : \rho < 1, \delta = 0$.
- Contraste 3: Bajo la hipótesis nula el proceso tiene una raíz unitaria y bajo la alternativa no tiene raíz unitaria pero presenta pendiente lineal determinista: $H_0 : \rho = 1, H_{1\tau} : \rho < 1, \delta \neq 0$.

3.4.1 Consideraciones generales

La aplicación del lema de Neyman-Pearson al problema de contrastes de raíces unitarias lleva a la conclusión de que, como sucede en otras situaciones, no existe un contraste que sea uniformemente más potente para $H_0 : \rho = 1, H_1 : \rho < 1$. Si se considera el caso más sencillo de un proceso AR(1) con errores normales

$$y_t = \rho y_{t-1} + a_t, \quad a_t \text{ iid } N(0, \sigma^2)$$

con $y_0 = 0$ y se supone, para facilitar la exposición, que $\sigma^2 = 1$ conocido, se puede utilizar el lema de Neyman-Pearson para contrastar la hipótesis nula $H_0 : \rho = 1$ frente

a la alternativa simple $H_1 : \rho = \rho_o$. La función de verosimilitud, dada una muestra de tamaño T , es

$$L(\rho_o) = \left(\frac{1}{\sqrt{2\pi}} \right)^T \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (y_t - \rho_o y_{t-1})^2 \right\}.$$

El contraste de la razón de verosimilitudes de $H_0 : \rho = 1$ frente a $H_1 : \rho = \rho_o$ propociona el contraste óptimo en el sentido de ser uniformemente más potente (UMP). Este contraste rechaza la hipótesis nula si $L(\rho_o)/L(1)$ es suficientemente grande. En la presente situación la construcción de este contraste lleva a la siguiente región crítica

$$(\rho_o^2 - 1) \sum_{t=1}^T y_{t-1}^2 - 2(\rho_o - 1) \sum_{t=1}^T y_t y_{t-1} < c$$

donde c es una constante. Esta región crítica es una combinación lineal de dos estadísticos con ponderaciones que dependen de la alternativa ρ_o , por lo que $c = c(\rho_o)$. Por tanto, en el caso de alternativa compuesta $H_1 : \rho < 1$ no existirá un contraste que sea UMP. Esta ausencia de un contraste óptimo ha llevado a una gran proliferación de contrastes de raíces unitarias.

La construcción del contraste en el caso I(1) general ha de evitar la distorsión que introduce la estructura dinamica de u_t y el componente determinista (μ, δ) . De esta forma, la construcción de un contraste de raíces unitarias presenta tres grandes líneas de investigación:

1. Encontrar formas mejores de estimar el valor verdadero ρ .
2. Encontrar procedimientos que disminuyan la distorsión que introduce u_t .
3. Encontrar procedimientos para eliminar la componente determinista.

Los resultados de estas tres grandes líneas de investigación pueden considerarse intercambiables en el sentido de que, en general, a un mismo procedimiento que mida la proximidad de ρ a la unidad se le pueden aplicar distintas correcciones para disminuir (eliminar asintóticamente) la influencia de u_t y también distintas formas de extraer el componente

determinista. Tradicionalmente el esfuerzo se ha concentrado en las dos primeras líneas expuestas. Sin embargo, las investigaciones de Elliot, Rothenberg & Stock (1992) muestran un procedimiento de eliminación del componente determinista que, aplicado sobre los contrastes originales de Dickey & Fuller (1979), duplican su potencia y los sitúa a la cabeza de los contrastes de raíces unitarias.

3.4.2 Contrastes de Dickey–Fuller

El problema general de detectar autocorrelación es muy antiguo y una primera referencia podría encontrarse en Yule (1921), siendo muy importantes las posteriores contribuciones de von Neumann et al. (1941) y Anderson (1948). No obstante, y según conocimientos del autor, es en la tesis doctoral de Dickey (1976) donde se encuentra la primera tabulación de la distribución de un estadístico que contraste la presencia de una raíz unitaria. Esta tabulación se encuentra publicada en Fuller (1976, p. 371–373). Han surgido desde entonces muchas propuestas que intentan, y consiguen, mejorar el comportamiento de estos contrastes. A pesar de dicho esfuerzo investigador, los contrastes de Dickey–Fuller siguen siendo los más utilizados.

Dickey (1976), Fuller (1976) y Dickey & Fuller (1979) construyen contrastes basados en los estadísticos de MCO $T(\hat{\rho} - 1)$ y el estadístico pivote t de regresión expuestos en la sección anterior. Estos autores fueron los primeros en tabular las distribuciones asintóticas de estos estadísticos cuando $\rho = 1$. Para ello utilizaron argumentos distintos a los empleados en la sección 3.3, basados en funcionales de movimientos brownianos. Dickey & Fuller calculan las distribuciones asintóticas basándose en la distribución de ratios de formas cuadráticas. Por ejemplo, en el caso más sencillo en que bajo la alternativa el proceso

tiene media nula obtuvieron

$$T(\hat{\rho} - 1) \xrightarrow{d} \frac{1}{2}\Gamma^{-1}(\xi^2 - 1),$$

$$\hat{\tau} = \frac{(\hat{\rho} - 1)}{\hat{\sigma}_\rho^2} \xrightarrow{d} \frac{1}{2}\Gamma^{-1/2}(\xi^2 - 1),$$

donde

$$\xi^2 = \sqrt{2} \sum_{i=0}^{\infty} \gamma_i Z_i, \quad \Gamma = \sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \quad \gamma_i = \frac{2(-1)^{i+1}}{(2i-1)\pi},$$

y Z_i son *iid* $N(0, 1)$. En el caso sencillo en que $u_t = a_t$ los contrastes utilizan la estimación por MCO de ρ , ρ^μ y ρ^τ en las regresiones

$$y_t = \rho y_{t-1} + a_t, \quad (3.4.1)$$

$$y_t = \alpha_0 + \rho^\mu y_{t-1} + a_t, \quad (3.4.2)$$

$$y_t = \alpha_0 + \alpha_1 t + \rho^\tau y_{t-1} + a_t. \quad (3.4.3)$$

Asímismo, los estadísticos $\hat{\tau}$ son los respectivos estadísticos t .

Dado un tamaño muestral, cuanto menor sea el valor del parámetro ρ más desplazada hacia la izquierda estará la distribución de los estadísticos. Por tanto se rechaza la hipótesis nula de raíz unitaria para valores inferiores a cierto valor crítico.

La tabla 3.1 muestra los valores críticos al 5% para distintos tamaños muestrales. Estos valores se encuentran en Fuller (1976) y han sido calculados con $y_0 = 0$ y $\psi(B) = 1$ con $a_t \sim N(0, 1)$.

Eliminación de la influencia de u_t

Los trabajos de Dickey & Fuller (1979) se basan en el caso sencillo AR(1) pudiéndose extender a un autorregresivo de cualquier orden aplicando los resultados de Fuller (1976)

Tabla 3.1: Valores críticos al 5% para diferentes contrastes de Dickey-Fuller (Fuller, 1976)

Contraste	T			
	25	50	100	250
$T(\hat{\rho} - 1)$	-7.3	-7.7	-7.9	-8.0
$T(\hat{\rho}^\mu - 1)$	-12.5	-13.3	-13.7	-14.0
$T(\hat{\rho}^\tau - 1)$	-17.9	-19.8	-20.7	-21.3
$\hat{\tau}$	-1.95	-1.95	-1.95	-1.95
$\hat{\tau}^\mu$	-3.00	-2.93	-2.89	-2.88
$\hat{\tau}^\tau$	-3.60	-3.50	-3.45	-3.43

p. 373). Esta extensión se basa en la estimación de un modelo en la forma denominada de *corrección de error*. Sea el caso en que y_t sigue un proceso autorregresivo $AR(p+1)$, donde se quiere contrastar si existe una raíz unitaria. Este modelo puede escribirse como

$$\left(1 - \sum_{j=1}^{p+1} \pi_j b^j\right)(y_t - \mu - \delta t) = \left(1 - \sum_{i=1}^p \phi_i B^i\right)(1 - \rho B)(y_t - \mu - \delta t) = a_t. \quad (3.4.4)$$

La forma de corrección de error se basa en que cualquier polinomio $\pi(B)$ puede descomponerse como:

$$\pi(B) = \pi(1)B + \pi^*(B)(1 - B), \quad (3.4.5)$$

donde si $\pi(B)$ es de orden infinito $\pi^*(B)$ también lo es, y si $\pi(B)$ es de orden $(p+1)$ $\pi^*(B)$ es de orden p . Es inmediato comprobar, igualando potencias, que

$$\pi_j^* = \sum_{i=j+1}^{p+1} \pi_i, \quad j = 1, \dots, p.$$

Aplicando la descomposición (3.4.5) a (3.4.4) se obtiene el modelo de corrección de error

$$\phi(B)(1 - \rho B) = \phi(1)(1 - \rho)B + \phi^*(B)(1 - B). \quad (3.4.6)$$

Dado que $\pi(B)(y_t - \mu - \delta t) = a_t$, $(1 - B)(y_t - \mu - \delta t) = \Delta y_t - \delta$ y llamando $c = \phi(1)$ se obtiene

$$a_t = c(1 - \rho)y_{t-1} - c(1 - \rho)(\mu + \delta) - c(1 - \rho)\delta t \quad (3.4.7)$$

$$+ \Delta y_t + \sum_{j=1}^p \phi_j^* \Delta y_{t-j} - (1 + \sum_{j=1}^p \phi_j^*)\delta. \quad (3.4.8)$$

Reagrupando términos se obtiene

$$\Delta y_t = k_1(1 - \rho) + k_2\delta + c(1 - \rho)\delta t + c(\rho - 1)y_{t-1} + \sum_{j=1}^p \phi_j^* \Delta y_{t-j}, \quad (3.4.9)$$

donde $k_1 = (\mu + \delta)$ y $k_2 = (1 + \sum_{j=1}^p \phi_j^*)$.

Los contrastes se realizan estimando por MCO las siguientes regresiones

Contraste 1: $H_0 : \rho = 1$, $H_1 : \rho < 1$, $\mu = \delta = 0$.

$$\Delta y_t = \alpha y_{t-1} + \sum_{j=1}^p \beta_j^* \Delta y_{t-j} + a_t.$$

Contraste 2: $H_0 : \rho = 1$, $H_{1\mu} : \rho < 1$, $\delta = 0$.

$$\Delta y_t = \alpha_0 + \alpha_\mu y_{t-1} + \sum_{j=1}^p \phi_j^* \Delta y_{t-j} + a_t.$$

Contraste 3: $H_0 : \rho = 1$, $H_{1\tau} : \rho < 1$, $\delta \neq 0$.

$$\Delta y_t = \alpha_0 + \beta t + \alpha_\tau y_{t-1} + \sum_{j=1}^p \phi_j^* \Delta y_{t-j} + a_t.$$

Donde $\alpha = c(\rho - 1)$, $\alpha_\mu = c(\rho^\mu - 1)$ y $\alpha_\tau = c(\rho^\tau - 1)$. Dickey y Fuller demuestran que $\hat{c}^{-1}T\hat{\alpha}$, $\hat{c}^{-1}T\hat{\alpha}_\mu$ y $\hat{c}^{-1}T\hat{\alpha}_\tau$, donde \hat{c} es un estimador consistente de c , tienen las mismas distribuciones asintóticas que en el caso $p = 0$. Igualmente demuestran que los estadísticos pivote de regresión $\hat{\tau}$, $\hat{\tau}^\mu$ y $\hat{\tau}^\tau$ mantienen las mismas propiedades asintóticas que en el caso $p = 0$ sin necesidad de correcciones adicionales. Como estimador de c pueden emplearse las estimaciones de los parámetros de los modelos estimados:

$$\hat{c} = 1 - \sum_{j=1}^p \hat{\phi}_j^*$$

Este resultado puede obtenerse igualmente aplicando el TCLF. Nótese que $c = \psi(1)^{-1}$ y por tanto puede aplicarse el resultado (3.3.17) por el que la convergencia de sumas parciales del proceso $u_t = \psi(B)a_t$ es equivalente a la del proceso a_t si se corrige por el término $\psi(1)$.

Said & Dickey (1984) (Said-Dickey) extienden estos resultados a procesos ARMA($p, 1, q$) de cualquier orden, pudiendo ser p y q desconocidos. Para ello Said-Dickey utilizan los resultados de Berck (1974). Berck (1974) demuestra que cualquier autorregresivo de orden desconocido puede aproximarse por un autorregresivo de orden finito si el número de retardos aumenta con el tamaño muestral T a una velocidad menor que $T^{1/3}$. Utilizando MCO para estimar los coeficientes de este modelo autorregresivo se obtienen los mismos resultados asintóticos que los tabulados por Dickey. A los contrastes de Dickey-Fuller basados en esta aproximación mediante una regresión en forma de corrección de error se les denomina contrastes de *Dickey-Fuller aumentado*.

Phillips (1987a) y Phillips & Perron (1988) (Phillips-Perron) proponen un procedimiento alternativo para recoger la estructura dinámica de $\psi(B)$. El método es no paramétrico con respecto a los parámetros que no son de interés (σ^2 y ϕ_j), por lo que puede adaptarse a una amplia clase de modelos. Los estadísticos construidos tienen la misma representación asintótica que los contrastes de Dickey-Fuller para el caso AR(1).

Tanto el procedimiento de Said-Dickey como el de Phillips-Perron son correcciones asintóticas del efecto de parámetros perturbadores. En muestras moderadas o pequeñas puede, sin embargo, existir influencia de estos parámetros. Esta influencia afectará al nivel de significación del contraste, y a su potencia. Estudios empíricos han mostrado que tal distorsión es elevada cuando el proceso tiene componente MA, especialmente si dicho

componente tiene raíces cercanas al círculo unidad que enmascaren la raíz unitaria autorregresiva (excelentes trabajos dedicados a estudiar este problema son Schwert (1987) y Agiakloglou & Newbold (1992)). De las dos propuestas anteriores, el estimador no paramétrico de Phillips-Perron es el que presenta mayores distorsiones haciendolo poco recomendable en el caso en que exista componente MA con raíces positivas. No obstante, las distorsiones del enfoque de Said-Dickey no son despreciables. Trabajos interesantes sobre modificaciones de estas dos propuestas se encuentran en Perron & Ng (1994), Hall (1994) y Ng & Perron (1995).

Existe en la literatura una tercera alternativa para extender los contrastes de Dickey-Fuller a procesos con componente $MA(q)$ y que presenta menos distorsiones en el nivel y la potencia que las propuestas anteriores. Este enfoque ha sido denominado de *variables instrumentales* y es debido a Hall (1989). Hall (1989) considera el modelo $ARIMA(0,1,q)$ y propone un estimador basado en variables instrumentales usando y_{t-k} , con $k > q$, como instrumento de y_{t-1} . Se demuestra que el estadístico del sesgo asintótico y el estadístico t , convenientemente modificado, convergen a las distribuciones de los estadísticos de Dickey-Fuller. Hall (1989) encuentra, mediante un ejercicio de simulación, que para el contraste basado en el sesgo asintótico las distorsiones en el nivel son menores que las del procedimiento de Phillips-Perron. No se ofrecen simulaciones sobre el estadístico t . Pantula & Hall (1991) extienden dicho enfoque al caso $ARIMA(p,1,q)$ y encuentran, para el sesgo asintótico, que el contraste presenta mejor comportamiento que Phillips-Perron pero similar a Said-Dickey. Li (1995) señala que la estimación del error estandar de los parámetros utilizada en el estadístico t del contraste de variables instrumentales puede ser negativo. Este autor propone una modificación de este contraste que evita tal problema. Además muestra un ejercicio de simulación tanto para el nivel como la potencia de los contrastes de variables instrumentales. En dichas simulaciones encuentra que las curvas de potencia

no son monótonas crecientes al alejarse de la unidad el verdadero valor de la raíz autorregresiva, presentándose mínimos locales cuando una raíz de polinomio AR es similar a la del polinomio MA.

3.4.3 Contrastes de Sargan–Bhargava

Sargan & Bhargava (1983) y Bhargava (1986) extienden los fundamentos de los contrastes de correlación serial (Anderson (1948), Durbin & Watson (1950, 1951, 1971)) al caso de raíces unitarias.

Estos autores utilizan la aproximación propuesta por Anderson (1948), y utilizada también por Durbin & Watson (1950, 1971), para la inversa de la matriz de covarianzas normalizada de y_t . Dicha aproximación permite construir contrastes que son, aproximadamente, invariantes uniformemente más potentes (UMPI) frente a la alternativa estacionaria $\rho < 1$. Los estadísticos que resultan son ratios del tipo Von Neumann. Originalmente han sido contruidos para el caso $\psi(B) = 1$ y sus distribuciones exactas pueden obtenerse aplicando el método de Imhof (1961). Stock (1994) propone un método para extenderlos al caso general $\psi(B) \neq 1$ y obtiene las expresiones asintóticas en funcionales de movimientos Brownianos. Por conveniencia, y siguiendo la formulación de Stock (1994), se escribirán las inversas de los estadísticos originales. Los estadísticos que se obtienen para contrastes con hipótesis alternativa de media nula, media no nula y tendencia determinista respectivamente son,

$$SB^o = \frac{T^{-2} \sum_{t=1}^T y_t^2}{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2} \xrightarrow{d} \kappa^{-2} \int_0^1 W^2(r) dr, \quad (3.4.10)$$

$$SB^\mu = \frac{T^{-2} \sum_{t=1}^T (y_t - \bar{y})^2}{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2} \xrightarrow{d} \kappa^{-2} \int_0^1 W^2(r) dr - \left\{ \int_0^1 W(r) dr \right\}^2, \quad (3.4.11)$$

$$SB^r = \frac{T^{-2} \sum_{t=1}^T (y_t - \beta_0^r - \beta_1^r t)^2}{T^{-1} \sum_{t=2}^T (\Delta y_t - \overline{\Delta y})^2} \xrightarrow{d} \kappa^{-2} \int_0^1 (W^B)^2(r) dr, \quad (3.4.12)$$

donde

$$\begin{aligned} \bar{y} &= T^{-1} \sum_{t=1}^T y_t, \\ \overline{\Delta y} &= T^{-1} \sum_{t=2}^T (y_t - y_{t-1}) = T^{-1} (y_T - y_1), \\ \beta_0^r &= \bar{y} - (y_T - y_1) \frac{T+1}{2(T-1)}, \\ \beta_1^r &= \frac{y_T - y_1}{T-1}, \\ W^B(r) &= W(r) - (r-0.5)W(1) - \int_0^1 W(r) dr \\ \kappa^2 &= \frac{E(u_t^2)}{\omega^2}, \quad \omega = \sigma\psi(1). \end{aligned} \quad (3.4.13)$$

En el caso $\psi(B) = 1$ se tiene que $E(u_t^2) = \sigma^2$ y por tanto $\kappa^2 = 1$. En el caso más general las distribuciones asintóticas dependen, a través de κ^2 , de parámetros perturbadores. Se necesita, pues, un estimador consistente $\hat{\kappa}^2$. Un estimador consistente de $E(u_t^2)$ es, en cada caso, el denominador de (3.4.10), (3.4.11) y (3.4.12) respectivamente. Un estimador consistente de ω puede obtenerse utilizando el procedimiento propuesto por Phillips (1987a) y utilizado también en Phillips y Perron (1988) consistente en la construcción de un estimador no paramétrico. Un segundo procedimiento consiste en aplicar los resultados de Said & Dickey (1984) para la estimación de $\psi(1)$ mediante la estimación de un autorregresivo de orden adecuado. En este segundo caso el estimador $\hat{\omega}_{AR}^2$ es

$$\hat{\omega}_{AR}^2 = \frac{\hat{\sigma}^2}{(1 - \sum_{i=1}^p \hat{\phi}_i^*)^2}, \quad (3.4.14)$$

donde $\hat{\phi}_j^*$, $j = 1, \dots, p$ son los estimadores MCO del modelo autorregresivo al que se ha ajustado la muestra. Al multiplicar por $\hat{\kappa}^2$ a los estadísticos \overline{SB} se obtienen unos estadísticos modificados cuyas distribuciones asintóticas no dependen de parámetros perturbadores. Estos estadísticos modificados son:

$$\overline{SB}^o = \hat{\kappa}^2 SB^o = \hat{\omega}^{-2} T^{-2} \sum_{t=1}^T y_t^2, \quad (3.4.15)$$

$$\overline{SB}^\mu = \hat{\kappa}^2 SB^\mu = \hat{\omega}^{-2} T^{-2} \sum_{t=1}^T (y_t - \bar{y})^2, \quad (3.4.16)$$

$$\overline{SB}^\tau = \hat{\kappa}^2 SB^\tau = \hat{\omega}^{-2} T^{-2} \sum_{t=1}^T (y_t - \beta_0^\tau - \beta_1^\tau t)^2. \quad (3.4.17)$$

En la tabla 3.2 se muestran los valores críticos al 5% para distintos tamaños muestrales. Estos valores se han obtenido por el método de Monte Carlo (20.000 replicaciones) y han sido calculados con $y_0 = 0$ y $\psi(B) = 1$ con $a_t \sim N(0, 1)$. Para un nivel de significación del 5% el contraste rechaza la hipótesis nula de raíz unitaria para valores inferiores a estos valores críticos.

Tabla 3.2: Valores críticos al 5% para diferentes contrastes de Sargan-Bhargava

Contraste	T			
	25	50	100	250
SB ^o	0.068	0.061	0.059	0.057
SB ^μ	0.043	0.040	0.038	0.037
SB ^τ	0.033	0.030	0.029	0.028

3.4.4 Contrastes de Elliot, Rothenberg y Stock

Elliot, Rothenberg & Stock (1992) construyen un contraste, que denominan τ_{GLS} , basado en el estadístico pivote $\hat{\tau}$ del contraste de Dickey-Fuller. La diferencia respecto del contraste de Dickey-Fuller es en la forma en que se elimina el componente determinista. En la

realización del contraste de Dickey-Fuller esta componente es eliminada estimando los respectivos parámetros en la regresión en la forma de corrección de error. Este procedimiento es equivalente a eliminar este componente bajo la alternativa estimada. El procedimiento propuesto por Elliot, Rothenberg & Stock (1992) consiste en eliminar dicha componente determinista bajo una alternativa local concreta

$$\rho_c = 1 + \frac{c}{T}, \quad (3.4.18)$$

donde c es un valor fijo. La nueva serie y_t^τ procede del filtro

$$y_t^\tau = y_t - z_t \tilde{\beta}$$

donde $z_t = 1$ o $z_t = (1, t)$ dependiendo de cómo sea dicha componente determinista. El coeficiente $\tilde{\beta}$ procede de la estimación por mínimos cuadrados generalizados suponiendo que y_t sigue un modelo AR(1) de parámetro la alternativa local (3.4.18). Este estimador $\tilde{\beta}$ puede obtenerse, de forma equivalente, estimando por MCO la regresión entre \tilde{y}_t y \tilde{z}_t , donde

$$\tilde{y}_t = (y_1, (1 - \rho_c B)y_2, \dots, (1 - \rho_c B)y_T)',$$

$$\tilde{z}_t = (z_1, (1 - \rho_c B)z_2, \dots, (1 - \rho_c B)z_T)'$$

Una vez conseguida la serie y_t^τ se obtiene el estadístico pivote $\hat{\tau}$ usual de la regresión, por MCO,

$$\Delta y_t^\tau = \alpha_0 y_{t-1}^\tau + \sum_{j=1}^p \alpha_j \Delta y_{t-j}^\tau + a_t \quad (3.4.19)$$

Elliot, Rothenber & Stock recomiendan en el caso de $z_t = 1$ utilizar el valor $c = -7$ y en el caso de $z_t = (1, t)$ utilizar el valor $c = -13.5$. Estos valores proceden de un estudio numérico realizado por los autores.

Se denotará con τ_{GLS}^μ al estadístico para el caso en que se corrija sólo por la media local ($z_t = 1$) y τ_{GLS}^τ para el caso $z_t = (1, t)$. La tabla 3.3 muestra los valores críticos para

Tabla 3.3: Valores críticos al 5% para diferentes contrastes τ_{GLS} y $\hat{\tau}$

Contraste	T			
	25	50	100	250
τ_{GLS}^{μ}	-2.56	-2.30	-2.14	-2.03
τ_{GLS}^{τ}	-3.57	-3.17	-3.03	-2.93
$\hat{\tau}$	-1.95	-1.95	-1.95	-1.95

un nivel de significación del 5% obtenidos mediante simulación por Monte Carlo. Las simulaciones se han realizado con $a_t \sim N(0,1)$, $y_1 = a_1$ y 20.000 replicaciones. Para el estadístico τ_{GLS}^{τ} los valores para $T = 50$ y 100 coinciden con los publicados por los autores. Los valores para $T = 25$ y 250 no estaban tabulados anteriormente. Los contrastes rechazan la hipótesis nula de raíz unitaria para valores inferiores a los críticos.

Al ser la distribución asintótica de τ_{GLS}^{μ} la misma que la del estadístico $\hat{\tau}$ de la tabla 3.1 los autores no obtienen ninguna tabulación de sus valores críticos en muestras finitas. No obstante la tabla 3.3 muestra que, en los tamaños muestrales indicados, existen diferencias entre los valores críticos de $\hat{\tau}^{\mu}$ y los obtenidos en la simulación para τ_{GLS}^{μ} . Estos valores críticos simulados coinciden, además, con los tabulados por Pantula et al. (1994) para este mismo contraste. En la comparación de los contrastes que se realiza en la próxima sección se utilizarán estos valores críticos simulados en lugar de los asintóticos para asegurar un nivel de significación nominal del 5%. No obstante, se incluirá una comparación de las potencias obtenidas utilizando ambos grupos de valores críticos.

3.4.5 Comparación en muestras finitas

En esta sección se comparan las potencias empíricas de los tres tipos de contrastes descritos en los anteriores apartados. Las potencias empíricas se han obtenido mediante simulación por Monte Carlo. En cada experimento se ha simulado el proceso

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

con $a_t, iid \sim N(0,1)$, $y_0 = 0$. En la elaboración de los contrastes se ha supuesto que el modelo sigue un AR(1). Se han realizado simulaciones para los valores de $\rho = 0.99, 0.97, 0.95, 0.90, 0.85, 0.80$ y 0.70 . Los tamaños muestrales analizados son $T = 25, 50, 100$ y 250 . En cada experimento se han obtenido 10.000 replicaciones del modelo y se ha contabilizado el porcentaje de veces que, para cada contraste, se rechaza la hipótesis nula, es decir, que el valor del estadístico sea inferior al valor crítico. Para el estadístico τ_{GLS}^{μ} se han utilizado los valores críticos obtenidos por simulación de la tabla 3.3. Al final de esta sección se incluye una comparación del nivel y potencia de este contraste utilizando los valores críticos del contraste $\hat{\tau}$

La tabla 3.4 y la figura 3.1 muestran las potencias de los contrastes que consideran bajo la alternativa que el proceso es estocástico puro. Puede observarse que los contrastes son prácticamente iguales.

Las tablas 3.5 a 3.12 y las figuras 4.2 a 4.9 muestran las potencias de los contrastes que permite media no nula. La comparación de estos contrastes es sencilla al no cortarse ninguna de las curvas de potencia. Los contrastes τ_{GLS} superan en potencia de forma uniforme en todas las situaciones. Esta ventaja es ligeramente superior en los contrastes sin tendencia lineal. Los contrastes de Sargan-Bhargava son inferiores a τ_{GLS} pero superan a los contrastes de Dickey-Fuller. El contraste de menor potencia es el basado en

el estadístico t de regresión.

En tamaños pequeños como $T = 25$ todos los contrastes son altamente conservadores, en el sentido de no rechazo de la hipótesis nula. Por el contrario la potencia es muy elevada en $T = 250$, donde es fácil conseguir una potencia cercana al 100% para alternativas no muy alejadas de uno.

Los resultados de este ejercicio de simulación serán completados en el próximo capítulo, donde se permitirá que el proceso generador de datos tenga componente MA. Ya se mencionó anteriormente que la presencia de dicha componente afecta negativamente al comportamiento de un contraste. Valores positivos del coeficiente θ aumentan el nivel de significación por encima del nivel nominal. Este efecto es especialmente importante en los contrastes basados en el sesgo asintótico $T(\hat{\rho} - 1)$, $T(\hat{\rho}^\mu - 1)$ y $T(\hat{\rho}^\tau - 1)$. Por contra los contrastes $\hat{\tau}$, $\hat{\tau}^\mu$ y $\hat{\tau}^\tau$ basados en el estadístico t presentan menores distorsiones en el nivel y mantienen buenos valores de potencia. Este resultado ha hecho que el contraste basado en el estadístico t sea el más recomendable de los dos.

Tabla 3.4: Potencia empírica para el nivel 5%. Sin componentes deterministas bajo la alternativa. 10.000 replicaciones por Monte Carlo. Proceso simulado: $y_t = \rho y_{t-1} + a_t$; $a_t \sim N(0, \sigma^2)$; $y_0 = 0$.

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T = 25$							
$T(\hat{\rho} - 1)$	0.06	0.07	0.09	0.15	0.23	0.33	0.58
$\hat{\tau}$	0.06	0.07	0.09	0.15	0.23	0.34	0.58
SB°	0.05	0.07	0.09	0.15	0.23	0.33	0.57
$T = 50$							
$T(\hat{\rho} - 1)$	0.06	0.09	0.14	0.32	0.56	0.78	0.97
$\hat{\tau}$	0.06	0.10	0.14	0.32	0.56	0.78	0.97
SB°	0.06	0.09	0.13	0.31	0.54	0.76	0.97
$T = 100$							
$T(\hat{\rho} - 1)$	0.08	0.18	0.32	0.76	0.98	1.00	1.00
$\hat{\tau}$	0.08	0.18	0.32	0.76	0.97	1.00	1.00
SB°	0.07	0.16	0.31	0.74	0.97	1.00	1.00
$T = 250$							
$T(\hat{\rho} - 1)$	0.15	0.55	0.91	1.00	1.00	1.00	1.00
$\hat{\tau}$	0.15	0.55	0.90	1.00	1.00	1.00	1.00
SB°	0.14	0.52	0.89	1.00	1.00	1.00	1.00

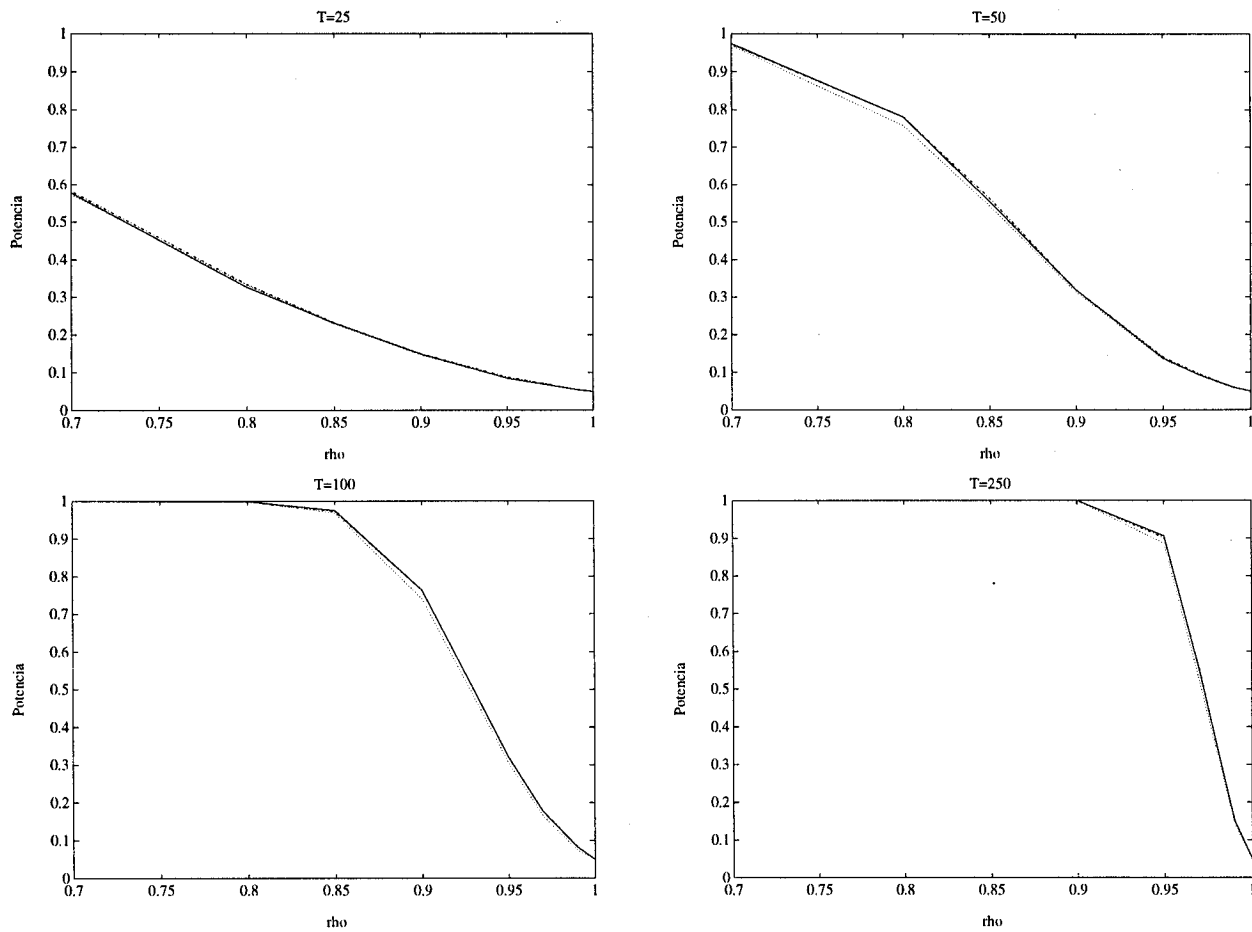


Figura 3.1: Potencia empírica de los contrastes de raíz unitaria. Alternativa sin componentes deterministas. (10.000 replicaciones, $y_0 = 0$). $-\cdot-\hat{\tau}$, $-T(\hat{\rho} - 1)$, $\cdots SB^o$.

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\mu - 1)$	0.05	0.06	0.07	0.10	0.14	0.18	0.31
$\hat{\tau}^\mu$	0.05	0.05	0.06	0.06	0.09	0.11	0.18
SB^μ	0.05	0.07	0.08	0.11	0.16	0.21	0.35
τ_{GLS}^μ	0.06	0.07	0.09	0.12	0.17	0.22	0.39

Tabla 3.5: Potencia empírica para el nivel 5%. $T = 25$, 10.000 replicaciones

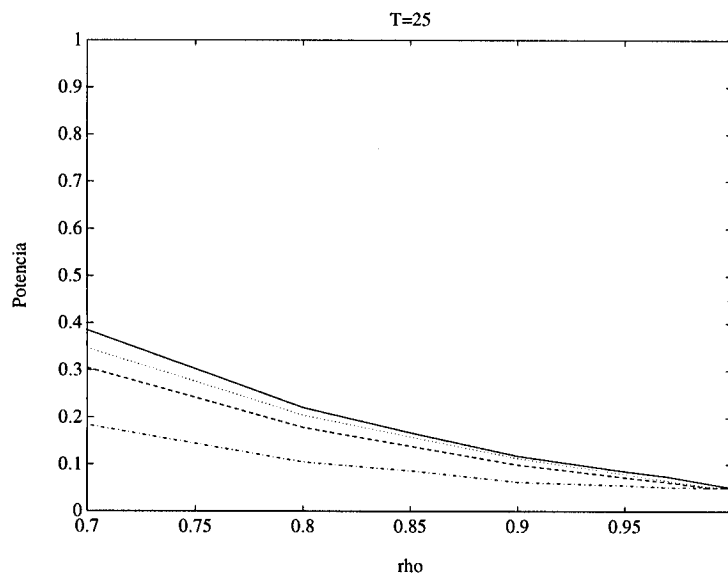


Figura 3.2: Potencia empírica de los contrastes de raíz unitaria. Alternativa con media no nula.

(10.000 replicaciones, $y_0 = 0$). $T = 25$. $-\cdot-\cdot-$ $\hat{\tau}^\mu$, $- - -$ $T(\hat{\rho}^\mu - 1)$, \dots SB^μ , $-$ τ_{GLS}^μ .

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\mu - 1)$	0.06	0.08	0.11	0.18	0.31	0.47	0.79
$\hat{\tau}^\mu$	0.05	0.06	0.07	0.11	0.20	0.32	0.64
SB^μ	0.06	0.10	0.12	0.22	0.37	0.55	0.86
τ_{GLS}^μ	0.06	0.09	0.13	0.26	0.44	0.62	0.89

Tabla 3.6: Potencia empírica para el nivel 5%. $T = 50$, 10.000 replicaciones

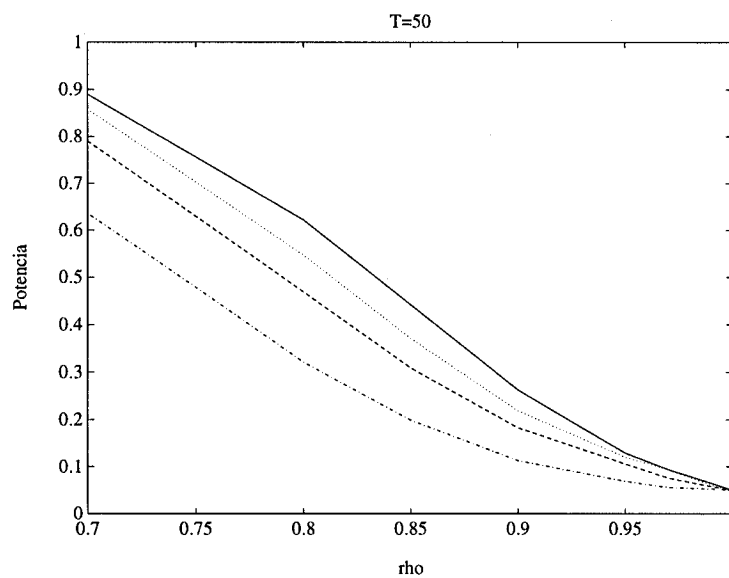


Figura 3.3: Potencia empírica de los contrastes de raíz unitaria. Alternativa con media no nula. (10.000 replicaciones, $y_0 = 0$). $T = 50$. $-\cdot-\hat{\tau}^\mu$, $--T(\hat{\rho}^\mu - 1)$, $\cdots SB^\mu$, $—\tau_{GLS}^\mu$.

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\mu - 1)$	0.07	0.12	0.19	0.47	0.79	0.95	1.00
$\hat{\tau}^\mu$	0.06	0.08	0.11	0.29	0.61	0.86	1.00
SB^μ	0.07	0.13	0.22	0.52	0.83	0.97	1.00
τ_{GLS}^μ	0.08	0.17	0.28	0.66	0.91	0.98	1.00

Tabla 3.7: Potencia empírica para el nivel 5%. $T = 100$, 10.000 replicaciones

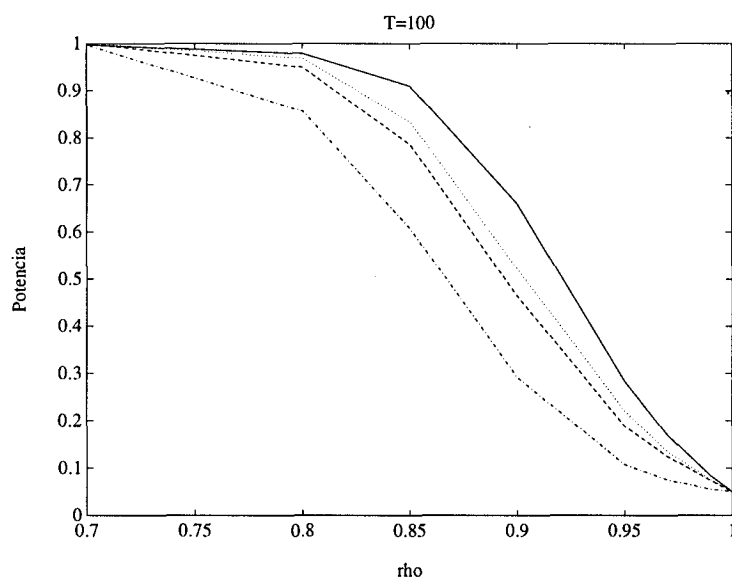


Figura 3.4: Potencia empírica de los contrastes de raíz unitaria. Alternativa con media no nula.

(10.000 replicaciones, $y_0 = 0$). $T = 100$. $-\cdot-\cdot-$ $\hat{\tau}^\mu$, $--$ $T(\hat{\rho}^\mu - 1)$, \cdots SB^μ , $-$ τ_{GLS}^μ .

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\mu - 1)$	0.11	0.32	0.64	0.99	1.00	1.00	1.00
$\hat{\tau}^\mu$	0.07	0.19	0.46	0.97	1.00	1.00	1.00
SB^μ	0.12	0.36	0.69	1.00	1.00	1.00	1.00
τ_{GLS}^μ	0.14	0.52	0.86	1.00	1.00	1.00	1.00

Tabla 3.8: Potencia empírica para el nivel 5%. $T = 250$, 10.000 replicaciones

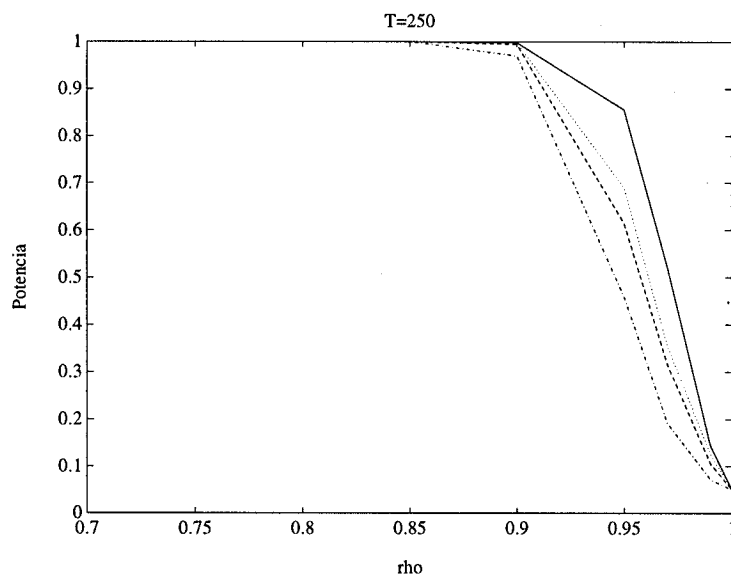


Figura 3.5: Potencia empírica de los contrastes de raíz unitaria. Alternativa con media no nula.

(10.000 replicaciones, $y_0 = 0$). $T = 250$. $\cdots \hat{\tau}^\mu$, $-- T(\hat{\rho}^\mu - 1)$, $\dots SB^\mu$, $— \tau_{GLS}^\mu$.

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\tau - 1)$	0.05	0.05	0.06	0.06	0.08	0.10	0.16
$\hat{\tau}^\tau$	0.05	0.05	0.05	0.05	0.07	0.08	0.12
SB^τ	0.05	0.05	0.05	0.06	0.08	0.10	0.16
τ_{GLS}^τ	0.05	0.06	0.06	0.06	0.08	0.10	0.16

Tabla 3.9: Potencia empírica para el nivel 5%. $T = 25$, 10.000 replicaciones

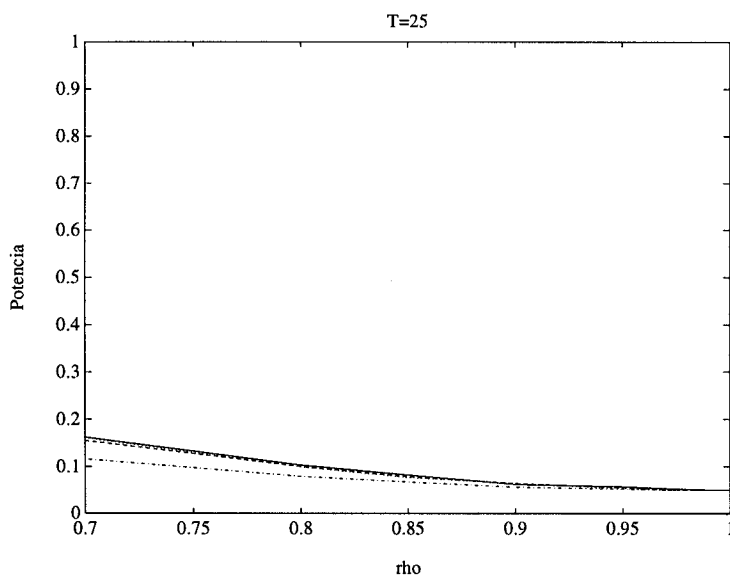


Figura 3.6: Potencia empírica de los contrastes de raíz unitaria. Alternativa con pendiente determinista. (10.000 replicaciones, $y_0 = 0$). $T = 25$. $\cdots \hat{\tau}^\tau$, $-- T(\hat{\rho}^\tau - 1)$, $\cdots SB^\tau$, $— \tau_{GLS}^\tau$

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\tau - 1)$	0.05	0.06	0.07	0.09	0.15	0.24	0.49
$\hat{\tau}^\tau$	0.05	0.06	0.06	0.08	0.12	0.19	0.40
SB^τ	0.05	0.06	0.07	0.10	0.17	0.27	0.53
τ_{GLS}^τ	0.05	0.06	0.07	0.11	0.18	0.29	0.58

Tabla 3.10: Potencia empírica para el nivel 5%. $T = 50$, 10.000 replicaciones

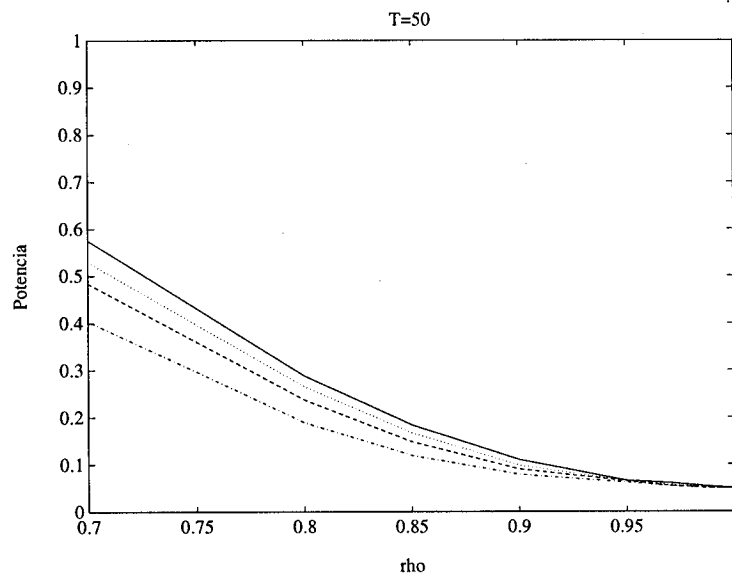


Figura 3.7: Potencia empírica de los contrastes de raíz unitaria. Alternativa con pendiente determinista. (10.000 replicaciones, $y_0 = 0$). $T = 50$. $-\cdot-\cdot-$ $\hat{\tau}^\tau$, $--$ $T(\hat{\rho}^\tau - 1)$, \cdots SB^τ , $—$ τ_{GLS}^τ

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\tau - 1)$	0.05	0.07	0.10	0.24	0.48	0.74	0.99
$\hat{\tau}^\tau$	0.05	0.07	0.08	0.19	0.40	0.64	0.96
SB^τ	0.06	0.07	0.11	0.27	0.53	0.77	0.97
τ_{GLS}^τ	0.06	0.07	0.11	0.30	0.58	0.83	0.99

Tabla 3.11: Potencia empírica para el nivel 5%. . $T = 100$, 10.000 replicaciones

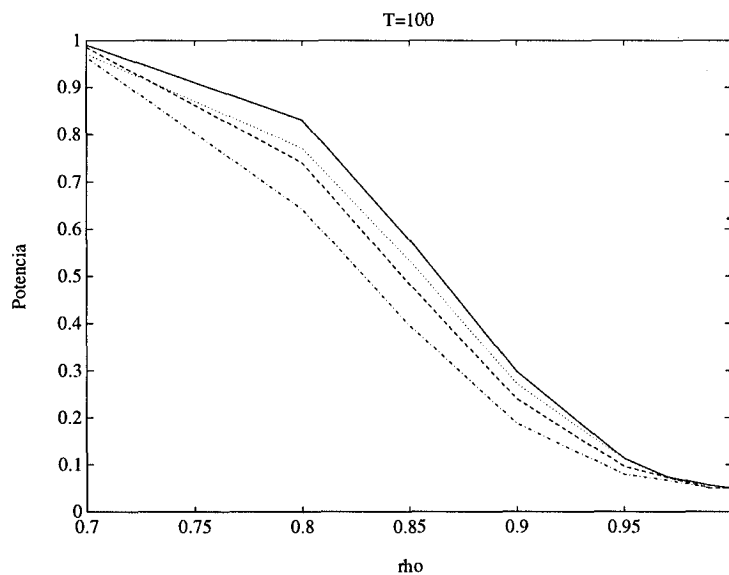


Figura 3.8: Potencia empírica de los contrastes de raíz unitaria. Alternativa con pendiente determinista. (10.000 replicaciones, $y_0 = 0$). $T = 100$. \cdots $\hat{\tau}^\tau$, $---$ $T(\hat{\rho}^\tau - 1)$, \dots SB^τ , $—$ τ_{GLS}^τ

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T(\hat{\rho}^\tau - 1)$	0.06	0.16	0.34	0.90	1.00	1.00	1.00
$\hat{\tau}^\tau$	0.06	0.13	0.27	0.83	1.00	1.00	1.00
SB^τ	0.07	0.18	0.39	0.92	1.00	1.00	1.00
τ_{GLS}^τ	0.07	0.19	0.45	0.96	1.00	1.00	1.00

Tabla 3.12: Potencia empírica para el nivel 5%. $T = 250$, 10.000 replicaciones

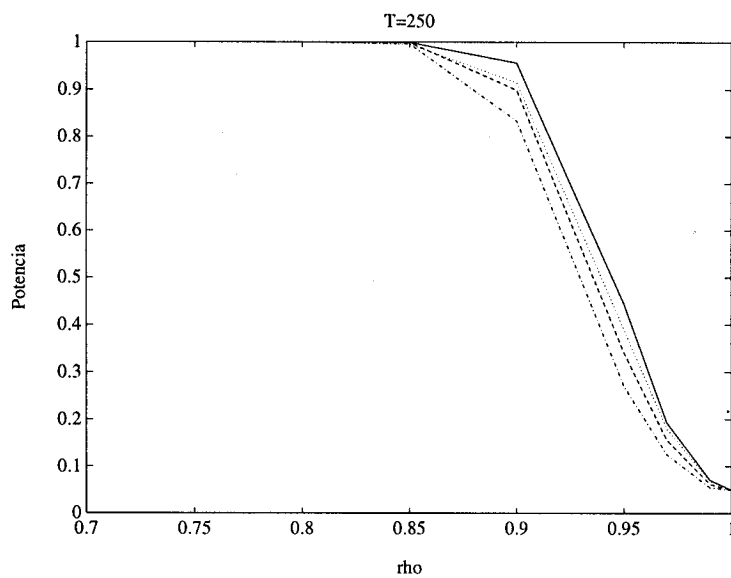


Figura 3.9: Potencia empírica de los contrastes de raíz unitaria. Alternativa con pendiente determinista. (10.000 replicaciones, $y_0 = 0$). $T = 250$. $-\cdot-\hat{\tau}^\tau$, $--T(\hat{\rho}^\tau - 1)$, $\cdots SB^\tau$, $—\tau_{GLS}^\tau$

Nivel y potencia de τ_{GLS}^{μ} con diferentes valores críticos

Elliot, Rothenberg & Stock (1992) demuestran que el estadístico del contrastes τ_{GLS}^{μ} tiene la misma distribución asintótica que el estadístico $\hat{\tau}$ de Dickey-Fuller para el caso no determinista. Basándose en este resultado los autores proponen la utilización de las tablas tabuladas por Dickey (publicadas en Fuller (1976, p. 373)) para el estadístico $\hat{\tau}$. La inspección de dichas tablas permite comprobar que dicho estadístico $\hat{\tau}$ converge muy rápido a su distribución asintótica, siendo el valor crítico para un nivel de significación del 5% igual a -1.95 para todos los tamaños muestrales desde $T = 25$. No obstante la tabla 3.3 muestra que, en los tamaños analizados, existen diferencias entre los valores críticos correspondientes al nivel 5% de $\hat{\tau}$ y τ_{GLS}^{μ} . Esta discrepancia indica que la convergencia de la distribución de τ_{GLS}^{μ} a su distribución asintótica no es tan rápida como la de $\hat{\tau}$. Por consiguiente, la utilización de los valores críticos de $\hat{\tau}$ para τ_{GLS}^{μ} no proporcionará un contraste con el nivel de significación nominal en muestras moderadas. Existirán, asimismo, diferencias en los valores de potencia.

En el apartado anterior se realizó la comparación entre los diferentes contrastes utilizando los valores críticos obtenidos por simulación de manera que el nivel de significación del 5% estaba asegurado por construcción. En este apartado se comparan los resultados de nivel y de potencia para el contraste τ_{GLS}^{μ} con ambos valores críticos. Para ello se ha realizado un ejercicio de simulación similar al anterior pero utilizando el valor crítico -1.95 . Los resultados obtenidos se encuentran en la tabla 3.13. Se denomina $\tau_{GLS}^{\mu} - MF$ al contraste que utiliza los valores críticos simulados para muestras finitas y $\tau_{GLS}^{\mu} - AS$ al contraste que utiliza los valores asintóticos (-1.95) de las tablas de Fuller (1976, p. 373). Puede comprobarse que sólo en el tamaño muestral $T = 250$ los valores obtenidos son similares. En el resto de los tamaños muestrales las diferencias entre ambos valores críticos son importantes. El contraste con valores críticos asintóticos tiene un nivel de significación

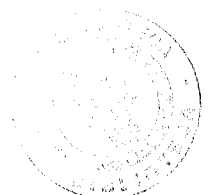


Tabla 3.13: Nivel y Potencia empíricos para el nivel nominal 5% del contraste τ_{GLS}^μ con diferentes valores críticos. $\tau_{GLS}^\mu - AS$: valores críticos asintóticos; $\tau_{GLS}^\mu - MF$ valores críticos para muestras finitas. 10.000 replicaciones por Monte Carlo. Proceso simulado: $y_t = \rho y_{t-1} + a_t$; $a_t \sim N(0, \sigma^2)$; $y_0 = 0$.

Contraste	ρ							
	1.00	0.99	0.97	0.95	0.90	0.85	0.80	0.70
$T = 25$								
$\tau_{GLS}^\mu - MF$	0.05	0.06	0.07	0.09	0.12	0.17	0.22	0.39
$\tau_{GLS}^\mu - AS$	0.16	0.18	0.21	0.23	0.31	0.41	0.50	0.70
$T = 50$								
$\tau_{GLS}^\mu - MF$	0.05	0.06	0.09	0.13	0.26	0.44	0.62	0.89
$\tau_{GLS}^\mu - AS$	0.10	0.12	0.18	0.25	0.45	0.65	0.81	0.97
$T = 100$								
$\tau_{GLS}^\mu - MF$	0.05	0.08	0.17	0.28	0.66	0.91	0.98	1.00
$\tau_{GLS}^\mu - AS$	0.08	0.11	0.23	0.40	0.77	0.95	0.99	1.00
$T = 250$								
$\tau_{GLS}^\mu - MF$	0.05	0.14	0.52	0.86	1.00	1.00	1.00	1.00
$\tau_{GLS}^\mu - AS$	0.06	0.17	0.57	0.88	1.00	1.00	1.00	1.00

mayor que el nominal lo que a su vez proporciona mayores valores de potencia.

3.4.6 Otros enfoques para contrastar raíces unitarias

Extensiones de los contrastes de Dickey-Fuller

Los contrastes de Dickey-Fuller se basan en las propiedades del estimador de MCO. Sobre la aplicación de estos contrastes con diferentes métodos de estimación son interesantes los trabajos de Pantula, Gonzalez-Farias & Fuller (1994) (así como las referencias que contiene) y Park & Fuller (1994). En estos trabajos se encuentra que, en general, métodos de estimación como máxima verosimilitud o mínimos cuadrados simétricos ponderados proporcionan mayores potencias que MCO.

Leybourne (1995) construye una variación de los contrastes de Dickey-Fuller basado en la aplicación de los estadísticos tanto a la serie observada (y_1, y_2, \dots, y_T) como a su secuencia inversa $(y_T, y_{T-1}, \dots, y_1)$. El contraste se basa en la distribución del máximo de los dos valores obtenidos (uno por cada ordenación de los datos). De esta forma se consiguen valores de potencia superiores a los de Dickey-Fuller, si bien no supera a los contrastes de Sargan-Bhargava ni Elliot-Rothenberg-Stock.

Se mencionó anteriormente que las distribuciones asintóticas de los contrastes de Dickey-Fuller no dependen del valor de la pendiente deterministas. No obstante Evans & Savin (1984) demuestran que, en muestras finitas dicha influencia existe y proponen contrastes que dependen de dicho valor, si bien las mejoras obtenidas son muy pequeñas (véase también Nankervis & Savin (1985)). Guilkey & Schmidt (1991) realizan un estudio de simulación donde encuentran que la utilización de valores críticos que dependan del valor de la constante que se incluya en la regresión (estandarizada por σ) mejora sustancialmente las propiedades de los contrastes de Dickey-Fuller. Dichos valores críticos pueden encontrarse en Schmidt (1989).

Ferreti & Romo (1995) proponen contrastes análogos a los contrastes de Dickey-Fuller cuyas distribuciones asintóticas son obtenidas mediante *bootstrap*. Ferreti & Romo (1995) demuestran un teorema central del límite funcional para un proceso estocástico construido a partir de la sucesión de sumas parciales correspondientes a las remuestras *bootstrap*. La técnica propuesta permite la extensión del contraste al caso $I(1)$ general sin necesidad de introducir modificaciones tipo Phillips-Perron. Las potencias conseguidas mediante este procedimiento son mejores que la de los contrastes de Dickey-Fuller en muestras pequeñas y similares en muestras grandes.

Ratio de verosimilitudes

Contrastes basados en el ratio de verosimilitudes pueden encontrarse en Dickey & Fuller (1981) para procesos autorregresivos y estimadores MCO, en Yap & Reinsel (1993) para procesos ARMA mediante máxima verosimilitud condicionada (método de Gauss-Newton) y en Perron (1990) para MCO y procesos generales utilizando el enfoque de Phillips-Perron.

Dickey & Fuller (1981) encuentran que los contrastes desarrollados presentan peores propiedades que sus anteriores contrastes basados en el sesgo asintótico y el estadístico t . En este trabajo, además, Dickey & Fuller estudian el comportamiento del ratio von Neumann para la detección de raíces unitarias y encuentran que su potencia supera a cualquiera de sus anteriores contrastes (para el caso de autorregresivos de primer orden correctamente especificado e $y_0 = 0$). Este ratio von Neumann será la base de los contrastes de Sargan-Bhargava. En adelante se sopondrá, cuando se haga referencia a los contrastes de Dickey & Fuller, que se hace referencia a los publicados en Dickey & Fuller

(1979) (sesgo asintótico y estadístico t de regresión).

Yap & Reinsel (1993) encuentran que, estimando por máxima verosimilitud condicionada, el ratio de verosimilitudes presenta mejores propiedades de potencia y nivel empíricos que el contraste basado en el estadístico t de regresión. Estos resultados proceden de un ejercicio de simulación donde el proceso simulado es ARMA(1,1) con diferentes valores de ϕ y θ . Los resultados son poco ilustrativos al estar basados en un tamaño muestral muy elevado ($T = 200$).

Multiplicadores de Lagrange

Los contrastes de multiplicadores de Lagrange para procesos ARIMA han sido desarrollados por Solo (1984), si bien estos contrastes han captado muy limitado interés. Saikkonen (1993) demuestra que este contraste es inconsistente si el modelo estimado es no invertible.

Said & Dickey (1985) construyen contrastes para procesos ARIMA($p,1,q$) muy similares al de los multiplicadores de Lagrange, estimando los parámetros mediante una iteración del método de Gauss-Newton. Los valores iniciales son $\rho = 1$ y un estimador consistente de θ . Al estimar los parámetros bajo la alternativa se obtienen estimaciones consistentes tanto bajo la nula como bajo la alternativa, no produciéndose el problema de inconsistencia señalado por Saikkonen (1993). El inconveniente de este contraste es que, al basarse en una sola iteración, es muy sensible a la estimación inicial de θ .

Utilizando también el enfoque de multiplicadores de Lagrange, Schmidt & Phillips (1992) desarrollan contrastes de raíces unitarias para un AR(1) cuando la alternativa es tendencia lineal determinista. Estos autores encuentran que el estadístico del contraste planteado

es proporcional al del contraste de Sargan-Bhargava (Bhargava (1986) para dicha alternativa. La potencia del contraste ($T = 100, 200$ y 500) es superior a la de Dickey-Fuller si $y_0 = 0$. Si se considera que y_0 procede de su distribución incondicional la potencia de este contraste es ligeramente superior a la de los contrastes de Dickey & Fuller (1976) para alternativas próximas a la unidad y ligeramente inferior en alternativas más alejadas de círculo unidad. Si se considera que las alternativas de mayor interés son las más próximas a la unidad puede decirse que este nuevo contraste mejora a Dickey-Fuller aunque de forma modesta. Un aspecto importante de este contraste es que la única diferencia respecto de los contrastes de Dickey-Fuller es en cómo se elimina la tendencia de los datos. Mientras que en Dickey-Fuller se elimina mediante una regresión en niveles Schmidt & Phillips (1992) la eliminan en una regresión en diferencias. Por tanto, un diferente tratamiento de la componente determinista puede producir una mejora de la potencia. Esta puntualización es relevante pues los contrastes de Elliot, Rothenberg & Stock (1992) se diferencian de los de Dickey-Fuller sólo en la forma en que se elimina el componente determinista y sin embargo sus valores de potencia aumentan sustancialmente. Schmidt & Phillips (1992) amplían la aplicabilidad de sus contrastes a procesos más generales utilizando el enfoque de Phillips-Perron.

Durbin-Hausman

Choi (1992) desarrolla contrastes de Durbin-Hausman para contrastar raíces unitarias. El principio de los contrastes de Durbin-Hausman (Durbin (1954), Hausman (1978)) consiste en considerar la diferencia de dos estimadores, ambos consistentes bajo la hipótesis nula y uno de ellos inconsistente bajo la alternativa. La diferencia de ambos estimadores, convenientemente estandarizada, debe tener una distribución límite bajo la nula y diverger bajo la alternativa. Choi (1992) utiliza la parametrización empleada por Dickey-Fuller y

los estimadores que emplea son MCO y variables instrumentales, donde los instrumentos son y_t , $(1, y_t)$ o $(1, t, y_t)$ dependiendo de la hipótesis alternativa (estrictamente son pseudo variables instrumentales al no utilizar variables retardadas). Las potencias obtenidas en un ejercicio de simulación, para el caso u_t iid, son superiores a los contrastes de Dickey-Fuller y similares a los de Sargan-Bhargava. Son, por tanto, inferiores a Elliot-Rothenberg-Stock. La extensión al caso más general en que u_t no sea iid se desarrolla en Choi (1990).

Contrastes sobre la función de autocorrelación

La inspección visual de los correlogramas muestrales puede considerarse el método de detección de raíces unitarias más empleado, lo que sugiere que puedan aprovecharse sus propiedades para la realización de un contraste de raíces unitarias. Hassler (1994) construye un contraste basado en el coeficiente de autocorrelación simple muestral de retardo k , \hat{r}_k , cuyos valores críticos son los del contraste de Dickey-Fuller del sesgo asintótico. El contraste propuesto disminuye su potencia con el retardo k . Por tanto el contraste tiene menos potencia que el contraste del sesgo asintótico de Dickey-Fuller, que puede considerarse como el caso $k = 1$.

Bierens (1993) construye contrastes basados en funciones de los coeficientes de autocorrelación muestrales. En dichas funciones interviene la diferencia entre dos coeficientes de autocorrelación muestral consecutivos y el retardo seleccionado se hace función del tamaño muestral. Los resultados de potencia, obtenidos en un ejercicio de simulación, son superiores a los del contraste de Phillips & Perron (1988). El diseño empleado en la simulación no permite compararlo fácilmente con resultados publicados sobre los contrastes de Dickey-Fuller aumentado.

Contrastes con hipótesis nula estacionaria

La especificación habitual de los contrastes de raíces unitarias es que bajo la hipótesis nula existe raíz unitaria. Esta formulación es lógica pues permite al analista disminuir la posibilidad de incurrir en infradiferenciación. Como se demostró en el capítulo 2, considerar que existe raíz unitaria cuando el proceso es estacionario, puede llevar a mejorar las predicciones. Por el contrario, considerar que el proceso es estacionario cuando tiene una raíz unitaria produce predicciones ineficientes. No aparece, por tanto, justificada, la realización de contrastes en los que la hipótesis nula sea la estacionariedad. La aplicación de estos contrastes conducirán, con alta probabilidad, a ineficiencias en la predicción. Existen, sin embargo, varios contrastes en la literatura con hipótesis nula estacionaria. El fundamento de estos contrastes es el siguiente. Se plantea un modelo que bajo la hipótesis nula es estacionario (alrededor de una media o una tendencia) y de coeficientes constantes, mientras que bajo la alternativa dichos coeficientes varían según un paseo aleatorio que lo hacen no estacionario. Los contrastes se pueden, así, basar en la varianza de dicho paseo aleatorio que, bajo la hipótesis nula de estacionariedad será cero.

Nyblom & Mäkeläinen (1983) y LaMotte & McWhorter (1978) desarrollan contrastes sobre presencia de paseo aleatorio en los coeficientes de un modelo lineal. Estos contrastes pueden aplicarse a un IMA(1,1) donde la hipótesis nula sea que $\theta = 1$, y por tanto se tenga una secuencia iid, frente a la alternativa $|\theta| < 1$ y se tenga un proceso no estacionario con componente MA. Kwiatkowski et al. (1992) aplican estos resultados para la construcción de contrastes de estacionariedad alrededor de una tendencia lineal donde el término de error se modeliza según el enfoque de Phillips-Perron. Leybourne & McCabe (1994) utilizan argumentos similares a Kwiatkowski et al (1992) para desarrollar un contraste donde la hipótesis nula es AR(p) y la alternativa ARIMA($p, 1, 1$).

Contraste del ratio de varianzas

Este contraste ha sido desarrollado por Lo & MacKinlay (1988), aunque la idea ha sido utilizada por otros autores como Cochrane (1988) y Fama & French (1988).

El contraste se basa en la idea de que, en un paseo aleatorio, la varianza de la diferencia de dos observaciones separadas k periodos ($E(y_{t+k} - y_t)^2$) es k veces la varianza de la diferencia de dos observaciones consecutivas σ^2 ($E(y_{t+k} - y_t)^2 = k\sigma^2$). Por el contrario, si el proceso es estacionario alrededor de una tendencia esta varianza tenderá a una constante. Esta idea ha sido utilizada por algunos autores como Cochrane (1988) y Fama & French (1988) para estudiar si ciertas variables económicas tienen un componente que sea paseo aleatorio. Para ello analizan el comportamiento del ratio de varianzas

$$\frac{Var(y_{t+k} - y_t)}{k\sigma^2}$$

como función de k . Si el proceso es paseo aleatorio puro dicho ratio debe ser constante. Si el proceso es estacionario debe decaer hasta cero al aumentar k . Si el proceso tiene un componente estacionario y otro que sea paseo aleatorio dicho ratio debe decaer hasta una constante no nula. El valor de dicha constante puede servir de indicación para conocer el tamaño del componente no estacionario.

Esta idea ha sido formalizada en un contraste por Lo & MacKinlay (1988, 1989). Estos autores encuentran que si se elige k adecuadamente el ratio de varianzas puede llegar a ser tan potente como el contraste de Dickey-Fuller. No existe una regla para la elección óptima de k , aunque los autores sugieren $k < T/2$.

Aunque este contraste ha sido típicamente interpretado como un ratio de varianzas puede interpretarse como un ratio de errores cuadráticos medios de predicción pues, en un paseo

aleatorio, la predicción de y_{t+k} desde y_t es precisamente y_t . El contraste puede así interpretarse como un contraste donde se evalúa la capacidad predictiva de un paseo aleatorio para predecir a horizonte k . Los contrastes que se proponen en el capítulo siguiente se basan explícitamente en errores de predicción, diferenciándose del presente contraste del ratio de varianzas en que la capacidad predictiva se evalúa en todos los horizontes que permite la longitud de la muestra en lugar de en un horizonte k .

Capítulo 4

Detección de raíces unitarias mediante los errores de predicción

4.1 Introducción

En este capítulo se construyen contrastes de raíces unitarias basados en errores de predicción multihorizonte. En el capítulo 1 se mencionó la relación existente entre la presencia de una raíz unitaria y el largo plazo de la variable considerada. Es en el largo plazo donde más diferencias existen al comparar un proceso estacionario y otro con una raíz unitaria. Los contrastes propuestos se basan en las diferencias en el ECMP. El ECMP de un proceso con raíz unitaria presenta la característica de no estar acotado con el horizonte de predicción. Esta particularidad los distingue tanto de los procesos estacionarios como de los procesos con tendencia lineal determinista.

Los resultados obtenidos en este capítulo confirman que el largo plazo contiene información para detectar una raíz unitaria. En el caso más sencillo de un $AR(1)$ los contrastes propuestos tienen potencias similares a los contrastes de Elliot-Rothenberg-Stock en zonas próximas al círculo unidad. Para estructuras dinámicas más generales los contrastes propuestos aventajan a sus competidores si se emplea un predictor eficiente así

como un estimador eficiente de ω^2 .

La razón fundamental de la ventaja de los contrastes propuestos en el caso I(1) general reside en su casi nula distorsión del nivel de significación y potencia frente a componentes de medias móviles. Por el contrario, los contrastes de Elliot-Rothenberg-Stock, así como los de Dickey-Fuller, pueden experimentar, en presencia de medias móviles, un aumento excesivo del nivel de significación.

Si se adopta la hipótesis nula de raíz unitaria será importante asegurar que el nivel de significación seleccionado se cumple. Aumentos en el nivel de significación de un contraste respecto al nivel nominal elegido por el analista hacen que el resultado sea poco fiable. Por otra parte, cuando un contraste presenta mayores niveles de significación que el nominal presentará también mayores potencias. Por tanto, es arriesgado valorar positivamente potencias elevadas si éstas van acompañadas de niveles mayores al nominal.

Las distorsiones que los componentes de medias móviles introducen en el nivel de significación de los contrastes de Dickey-Fuller y los de Elliot-Rothenberg-Stock pueden disminuirse si se modeliza un autorregresivo de orden elevado. Este procedimiento tiene el inconveniente de proporcionar menores valores de potencia. Por el contrario, los contrastes que se proponen en este capítulo mantienen las propiedades de potencia que se obtuvieron para el caso AR(1).

La organización del capítulo es la siguiente. La sección 4.2 introduce los fundamentos en que se basan los contrastes propuestos. La sección 4.3 presenta los contrastes, sus distribuciones asintóticas y un ejercicio de simulación para el caso AR(1). En la sección 4.4 se extienden los resultados al caso más general mediante la estimación de un término

corrector que elimine asintóticamente la influencia de parámetros perturbadores. En la sección 4.6 se presenta un segundo enfoque a esta extensión donde, además de dicho término corrector, se utilizan predictores más eficientes. En la sección 4.7 se resumen los resultados obtenidos.

4.2 Errores de predicción y raíces unitarias

Sea $\{y_t\}$ un proceso estocástico puro AR(1). El valor de y_{T+h} puede escribirse como

$$y_{T+h} = \rho^h y_T + \sum_{j=0}^{h-1} \rho^j a_{T+h-j}.$$

Se considera el siguiente predictor

$$\tilde{y}_{T+h} = y_T. \quad (4.2.1)$$

Este predictor no será centrado ya que

$$E(y_{T+h}|y_T) = \rho^h y_T$$

y tendrá un sesgo de predicción de $(1 - \rho^h)y_T$. Su ECMP es

$$\begin{aligned} E(y_{T+h} - y_T)^2 &= \sigma^2 \left(\sum_{j=0}^{h-1} \rho^{2j} + \frac{(1 - \rho^h)^2}{1 - \rho^2} \right) \\ &= 2\sigma^2 \frac{1 - \rho^h}{1 - \rho^2}, \end{aligned} \quad (4.2.2)$$

donde se ha aplicado que $E(y_T^2) = \sigma^2/(1 - \rho^2)$. El primer sumando a la derecha de (4.2.2) es el ECMP si se hubiese utilizado un predictor centrado, por lo que el segundo sumando es el error cometido por utilizar el predictor (4.2.1).

En el caso en que $\rho = 1$ el predictor (4.2.1) será centrado y su ECMP es

$$E(y_{T+h} - y_T | \rho = 1)^2 = \sigma^2 h.$$

Para permitir la comparación del ECMP del predictor (4.2.1) cuando $\rho = 1$ y cuando $\rho < 1$ se obtiene, operando en (4.2.2),

$$E(y_{T+h} - y_T)^2 = \sigma^2(h + d(h, \rho)),$$

donde

$$d(h, \rho) = 2 \frac{1 - \rho^h}{1 - \rho^2} - h. \quad (4.2.3)$$

Esta representación permite analizar las diferencias entre el ECMP teórico si existiese raíz unitaria respecto al que realmente se obtiene si el proceso es estacionario. Para el horizonte de predicción $h = 1$ se obtiene que $d(1, \rho) = (1 - \rho)/(1 + \rho) > 0$. Es decir, si se utiliza el predictor de un paseo aleatorio el ECMP a un periodo, cuando el proceso es un AR(1) estacionario, es mayor que cuando el proceso es un paseo aleatorio correctamente especificado. Para horizonte de predicción $h = 2$ se obtiene que $d(2, \rho) = 0$ y ambos ECMP coinciden para cualquier valor de ρ . Para $h > 2$ se obtiene $d(h, \rho) < 0$, disminuyendo de forma exponencial con el horizonte al ser la curva de ECMP de crecimiento amortiguado. Por tanto, el ECMP del predictor (4.2.1) es menor cuando el proceso es estacionario que cuando tiene una raíz unitaria. Además se tiene que

$$E(y_{T+h} - y_T | \rho = 0)^2 = 2\sigma^2$$

por lo que para $\rho = 0$ el ECMP es una horizontal de altura $2\sigma^2$.

La figura 4.1 ilustra estos resultados. En esta figura están representadas varias curvas de ECMP obtenidas a partir de la expresión (4.2.2), con $\sigma^2 = 1$, para distintos valores de ρ . También está representada la línea $\sigma^2 h$, que es el ECMP teórico de un paseo aleatorio. En este gráfico $d(h, \rho)$ es la distancia vertical, para un horizonte h , entre la línea de puntos y la curva de ECMP de cada valor de ρ . Se comprueba que el comportamiento de esta

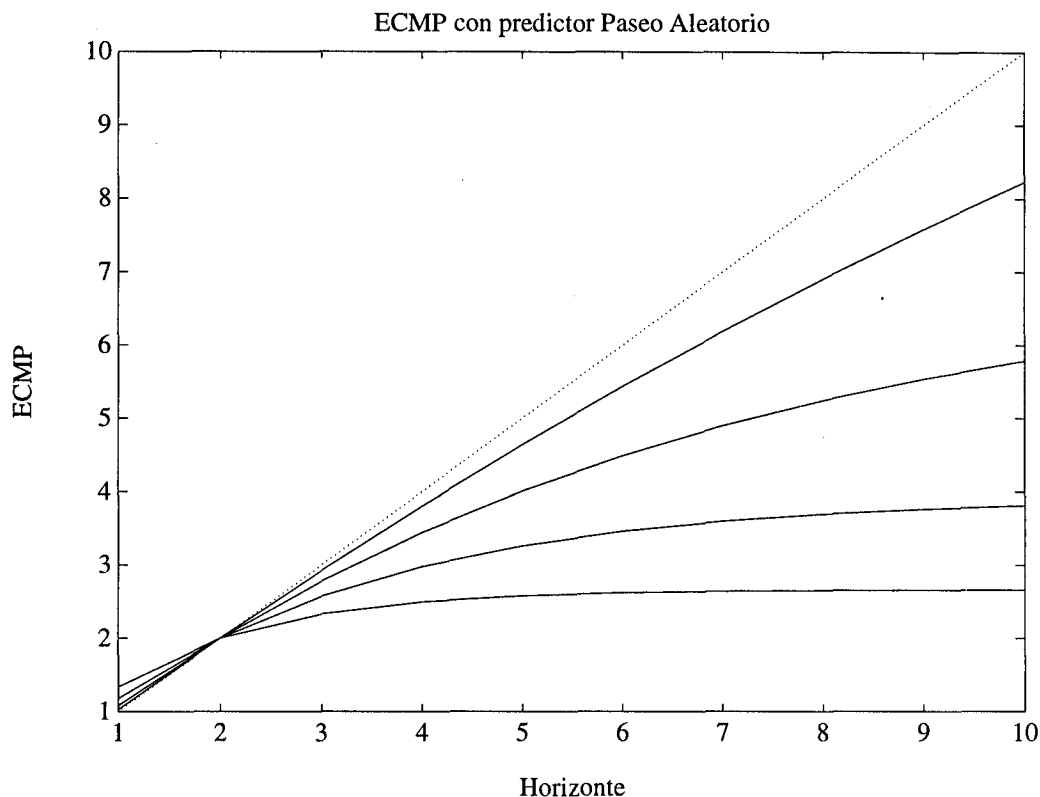


Figura 4.1: ECMP del modelo $y_t = \rho y_{t-1} + a_t$ utilizando como predictor paseo aleatorio.

Los valores de ρ son, de abajo a arriba en $h > 2$, 0.50, 0.70, 0.85 y 0.95.

distancia corresponde con lo antes mencionado. En la sección siguiente se construyen estadísticos basados en las propiedades de esta distancia.

4.3 Contrastes basados en errores de predicción

4.3.1 Construcción de los estadísticos

Sea el proceso generador de datos (PGD) un paseo aleatorio. Se define $e_{i|j}$ como el error de predicción empírico al predecir la observación y_i desde $t = j$ ($i > j$) con el predictor

(4.2.1), es decir,

$$e_{i|j} = y_i - y_j. \quad (4.3.1)$$

En esta sección se construyen contrastes, para la hipótesis nula de raíz unitaria, basados en la comparación entre estos errores de predicción empíricos y los teóricos bajo la hipótesis de raíz unitaria. Para ello se evalúa la suma acumulada de errores cuadráticos de predicción teóricos y empíricos desde $h = 1$ hasta $h = T - 1$. Esto es equivalente, en una aproximación gráfica, a comparar el área bajo la diagonal de la figura 4.1 y el área bajo las líneas continuas. Debido a que este error de predicción es la diferencia entre dos observaciones el resultado obtenido es independiente de la media del proceso cuando éste es estacionario. Por tanto es aplicable a procesos, que bajo la alternativa estacionaria, tengan media no nula.

Los errores de predicción empíricos desde y_1 son:

$$\begin{aligned} e_{2|1} &= y_2 - y_1 = a_2 \\ e_{3|1} &= y_3 - y_1 = a_2 + a_3 \\ &\vdots \\ e_{T|1} &= y_T - y_1 = \sum_{j=2}^T a_j. \end{aligned}$$

La suma acumulada de los cuadrados de éstos errores es

$$\sum_{t=2}^T e_{t|1}^2 = \sum_{t=2}^T (y_t - y_1)^2 \quad (4.3.2)$$

que tendrá un valor esperado de

$$\sigma^2(1 + 2 + \cdots + (T - 1)) = \sigma^2 \frac{T(T - 1)}{2}. \quad (4.3.3)$$

Como σ^2 es desconocido se estimará consistentemente con

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T - 1}. \quad (4.3.4)$$

El ratio entre la suma acumulada (4.3.2) y su esperanza, bajo la hipótesis nula, (4.3.3) con la estimación (4.3.4) es

$$\frac{\sum_{t=2}^T (y_t - y_1)^2}{\frac{T(T-1)}{2} \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T-1}}$$

Para hacer esta expresión comparable con los otros contrastes se utilizará $T-1$ en lugar de T y se omitirá el factor 2. Al estadístico que se obtiene se le denominará CS_1^μ ,

$$CS_1^\mu = \frac{(T-1)^{-2} \sum_{t=2}^T (y_t - y_1)^2}{(T-1)^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (4.3.5)$$

donde CS denota *Cumulated Sum*, pues éste estadístico es una suma acumulada, estandarizada, de los errores cuadráticos de predicción a cada horizonte. Como el origen de predicción es la primera observación se le denota por CS_1 . El superíndice μ representa la situación de media no necesariamente nula bajo la alternativa estacionaria.

Bajo la hipótesis nula de raíz unitaria este ratio es una variable aleatoria cuya distribución límite se derivará en la próxima sección. No obstante puede observarse que $\hat{\sigma}^2$ es un estimador consistente de σ^2 y por tanto es $O_p(1)$. Asimismo se verifica (ver proposición 3.1) que

$$\begin{aligned} \frac{\sum_{t=2}^T (y_t - y_1)^2}{(T-1)^2} &= \frac{\sum_{t=2}^T y_t^2}{(T-1)^2} + \frac{y_1^2}{(T-1)} - 2y_1 \frac{\sum_{t=2}^T y_t}{(T-1)^2} \\ &= O_p(1) + O_p(T^{-1}) + O_p(T^{-\frac{1}{2}}) = O_p(1), \end{aligned}$$

siendo $(T-1)^2 \sum_{t=2}^T y_t^2$ el término dominante de esta suma. Por tanto $CS_1^\mu = O_p(1)$.

Si el proceso es un $AR(1)$ de parámetro $\rho < 1$ el denominador sigue siendo $O_p(1)$ al verificarse que

$$\begin{aligned} \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T-1} &= (\rho-1)^2 \frac{\sum_{t=2}^T y_{t-1}^2}{T-1} + \frac{\sum_{t=2}^T a_t^2}{T-1} + 2(\rho-1) \frac{\sum_{t=2}^T a_t y_{t-1}}{T-1} \\ &= O_p(1) + O_p(1) + O_p(T^{-\frac{1}{2}}), \end{aligned}$$

cumpléndose en este caso que

$$\hat{\sigma}^2 \xrightarrow{p} \frac{2\sigma^2}{1+\rho}.$$

Sin embargo, en esta situación estacionaria el numerador es de menor orden de magnitud que en el caso de paseo aleatorio, comprobándose que

$$\begin{aligned} \frac{\sum_{t=2}^T (y_t - y_1)^2}{(T-1)^2} &= \frac{\sum_{t=2}^T y_t^2}{(T-1)^2} + \frac{y_1^2}{(T-1)} - 2y_1 \frac{\sum_{t=2}^T y_t}{(T-1)^2} \\ &= O_p(T^{-1}) + O_p(T^{-1}) + O_p(T^{-1}) \end{aligned}$$

y se obtiene entonces que, en el caso estacionario,

$$CS_1^\mu \xrightarrow{p} 0,$$

debido, básicamente, a que ha disminuido el orden de magnitud del término $\sum_{t=2}^T y_t^2$.

Asímismo, dado un tamaño muestral T , se observa que, de acuerdo a las propiedades de la distancia $d(h, \rho)$ en (4.2.3), el numerador del estadístico CS_1^μ tendrá, por término medio, valores menores cuanto menor sea el valor de ρ , mientras que el denominador aumentará. Por tanto un contraste para la detección de una raíz unitaria frente a la alternativa estacionaria tendrá la región de rechazo

$$CS_1^\mu < c_1^\mu,$$

donde el valor c_1^μ se elige en función del nivel de significación deseado.

En el caso en que el proceso sea un paseo aleatorio con deriva $y_t = \delta + y_{t-1} + a_t$ los errores de predicción $e_{t|1}$ son

$$\begin{aligned} e_{2|1} &= y_2 - y_1 - \delta = a_2 \\ e_{3|1} &= y_3 - y_1 - 2\delta = a_2 + a_3 \end{aligned}$$

$$\begin{aligned} & \vdots \\ e_{t|1} &= y_t - y_1 - (t-1)\delta = a_2 + a_3 + \cdots + a_t \end{aligned} \quad (4.3.6)$$

$$\begin{aligned} & \vdots \\ e_{T|1} &= y_T - y_1 - (T-1)\delta = \sum_{j=2}^T a_j. \end{aligned} \quad (4.3.7)$$

Un estimador eficiente de δ puede obtenerse de la regresión $\Delta y_t = \delta + a_t$, que lleva a

$$\hat{\delta} = \frac{\sum_{t=2}^T (y_t - y_{t-1})}{T-1} = \frac{y_T - y_1}{T-1}. \quad (4.3.8)$$

Igualmente un estimador consistente de σ^2 será

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}{T-2}, \quad (4.3.9)$$

lo que conduce al estadístico CS_1^τ ,

$$CS_1^\tau = \frac{(T-1)^{-2} \sum_{t=1}^T (y_t - y_1 - \hat{\delta}(t-1))^2}{(T-2)^{-1} \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}, \quad (4.3.10)$$

donde el superíndice τ hace referencia a que, tanto bajo la hipótesis nula como bajo la alternativa, el proceso tiene tendencia lineal determinista. Al igual que en el caso de paseo aleatorio sin deriva, este estadístico tiene, bajo la hipótesis nula, cierta distribución límite. En el caso en que el proceso sea estacionario, con o sin tendencia determinista, este estadístico converge a cero. Un contraste para detectar raíz unitaria rechazará dicha hipótesis para valores

$$CS_1^\tau < c_1^\tau.$$

Los dos estadísticos CS_1^μ y CS_1^τ presentan el inconveniente de que los dos primeros sumandos del denominador presentan, por término medio, comportamientos distintos al resto de los sumandos. Este resultado procede de los distintos valores que la distancia $d(h, \rho)$ en (4.2.3) presentaba con h . Como se vió en la sección 4.2

$$d(1, \rho) < 0,$$

$$d(2, \rho) = 0,$$

$$d(h, \rho) > 0, \quad h > 2.$$

Por tanto, por término medio y a medida que ρ se aleje de la unidad, el primer sumando del numerador en CS_1^μ y CS_1^τ tenderá a aumentar, el segundo sumando permanecerá constante y el resto tenderá a disminuir. Una alternativa que evita este problema es considerar solamente los errores de predicción a partir de cierto horizonte h , $h > 1$. Sin embargo, los ejercicios de simulación realizados con estos estadísticos no llevaron a ninguna mejora de la potencia del contraste respecto a CS_1^μ y CS_1^τ .

Una segunda característica de los estadísticos propuestos es su dependencia de la primera observación, al evaluarse todos los errores de predicción desde y_1 . Se proponen a continuación dos soluciones para disminuir dicha dependencia. En primer lugar puede observarse que los estadísticos CS_1^μ y CS_1^τ solamente utilizan un error de predicción para cada horizonte. Sin embargo es inmediato comprobar que existen $T - 1$ errores de predicción a horizonte $h = 1$ que son $(y_2 - y_1)$, $(y_3 - y_2)$, ..., $(y_T - y_{T-1})$. Asimismo existen $T - 2$ para horizonte $h = 2$ y, en general, $T - k$ para horizonte $h = k$. Si se desea utilizar toda la información contenida en la muestra puede construirse un estadístico donde, para cada horizonte de predicción, se promedien todos los errores cuadráticos que puedan calcularse en la muestra. Este método conduce a los siguientes estadísticos

$$CS_A^\mu = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T-(s-t))^{-1} (y_s - y_t)^2}{(T-1)^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (4.3.11)$$

$$CS_A^\tau = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T-(s-t))^{-1} (y_s - y_t - \hat{\delta}(s-t))^2}{(T-2)^{-1} \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}, \quad (4.3.12)$$

donde el subíndice A denota *Average*. Una segunda solución para disminuir la dependencia de la primera observación consiste en considerar también los errores de predicción desde cada observación hasta la observación y_T . Un estadístico basado únicamente en estos

errores de predicción sería

$$CS_T^\mu = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} (y_t - y_T)^2}{(T-1)^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (4.3.13)$$

$$CS_T^\tau = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} (y_t - y_T - \hat{\delta}(t-T))^2}{(T-1)^{-1} \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \quad (4.3.14)$$

Se comprueba que $CS_T^\tau = CS_1^\tau$. Esta igualdad se verifica al cumplirse que

$$\begin{aligned} y_T + \frac{y_T - y_1}{T-1}(t-T) &= y_T + \frac{y_T - y_1}{T-1}((t-1) - (T-1)) \\ &= y_1 + \frac{y_T - y_1}{T-1}(t-T)(t-1). \end{aligned}$$

Si el proceso es estacionario los estadísticos CS_T^μ y CS_1^μ proporcionarán resultados similares. Este resultado se justifica por la propiedad de reversibilidad de los procesos débilmente estacionarios (Box & Jenkins, 1976, p. 197). Según esta propiedad los procesos $\phi(B)w_t = \theta(B)a_t$ y $\phi(F)w_t = \theta(F)e_t$, donde $Fw_t = w_{t+1}$ y e_t es *iid* de media cero y varianza σ^2 , tienen la misma estructura de covarianzas. Sin embargo, si el proceso tiene raíz unitaria ambos estadísticos tendrán distintas distribuciones asintóticas. No obstante, los valores críticos al 5% obtenidos para CS_1^μ y CS_T^μ en muestras finitas son muy similares (ver tabla 4.1, la desviación típica de estos valores es, aproximadamente, 0.0007). Es de esperar, por tanto, que ambos contrastes, CS_1^μ y CS_T^μ , tengan idéntico comportamiento. Este resultado ha sido confirmado mediante un ejercicio de simulación. Un aspecto a destacar de esta simulación es la importancia de que no exista influencia de los valores iniciales con que se generan las muestras. Para evitar esta influencia se generaron 100 observaciones adicionales al comienzo de cada serie. Una posible influencia de los valores iniciales haría que las series generadas no fuesen estacionarias. Esta no estacionariedad provocaría la falsa impresión de que CS_1^μ y CS_T^μ fuesen distintos.

Al ser CS_1^μ y CS_T^μ claramente dependientes resulta de interés construir un estadístico que promedie CS_1^μ y CS_T^μ . Este estadístico utilizará, en cada observación, el error de predicción

desde ambos extremos de la muestra. A este estadístico se le denotará por CS_{1T}^μ y tiene la expresión

$$CS_{1T}^\mu = \frac{(T-1)^{-2} \left\{ \sum_{t=2}^T (y_t - y_1)^2 + \sum_{t=1}^{T-1} (y_t - y_T)^2 \right\}}{2(T-1)^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}. \quad (4.3.15)$$

La propiedad de reversibilidad de los procesos estacionarios ha sido utilizada por varios autores para modificar los contrastes de Dickey-Fuller y conseguir mayores potencias (Pantula et al. (1994), Leybourne (1995) entre otros). Es interesante analizar si un contraste basado en CS_{1T}^μ proporciona mejores resultados que CS_1^μ .

La construcción del estadístico análogo para el caso de tendencia determinista, es decir CS_{1T}^τ , no tiene especial interés al verificarse $CS_1^\tau = CS_T^\tau$, y por tanto $CS_{1T}^\tau = CS_1^\tau$.

4.3.2 Comparación con otros estadísticos de contraste

Para facilitar la comparación de los estadísticos propuestos con otros enfoques presentados en la literatura se centrará la atención en los estadísticos CS^μ , siendo las conclusiones extensibles a CS^τ .

Si bien no existe en la literatura, según conocimientos del autor, ningún enfoque para detectar raíces unitarias que se base explícitamente en errores de predicción pueden encontrarse similitudes con el contraste del ratio de varianzas (Lo & MacKinlay, 1988). El contraste del ratio de varianzas compara la varianza de la diferencia de orden k , $\text{VAR}(y_t - y_{t-k})$, con k veces la diferencia de orden 1, $k\text{VAR}(y_t - y_{t-1})$. Si el proceso es paseo aleatorio ambas varianzas serán iguales. El estadístico de dicho contraste puede escribirse como

$$M(k) = \frac{(T-k)^{-1} \sum_{t=1}^{T-k} (y_{t+k} - y_t - k\hat{\delta})^2}{k(T-1)^{-1} \sum_{t=1}^{T-1} (y_{t+1} - y_t - \hat{\delta})^2}.$$

Este estadístico puede interpretarse como el ratio entre el ECMP a horizonte k , estimado en la muestra, de un paseo aleatorio con deriva (o sin ella si se omite el término $\hat{\delta}$) y la estimación del valor teórico que se espera si la especificación es correcta. La diferencia entre este contraste y los contrastes que se proponen en este capítulo reside, básicamente, en el horizonte de predicción empleado. Mientras en el ratio de varianzas se utiliza un horizonte k seleccionado según el tamaño muestral, en los estadísticos CS se emplean todos los horizontes que permite la longitud de la muestra. Como ya se mencionó en la sección 3.4.6 este contraste puede tener la misma potencia que los contrastes de Dickey-Fuller si se elige k de manera adecuada. No existe un criterio óptimo para esta elección, aunque sus autores recomiendan elegir $k < T/2$.

Un segundo estadístico que se asemeja a CS_1^μ es el estadístico utilizado en el contraste de Sargan & Bhargava (1983). Este estadístico es un ratio tipo von Neumann,

$$SB^\mu = \frac{T^{-2} \sum_{t=1}^T (y_t - \bar{y})^2}{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}.$$

Ambos estadísticos, CS_1^μ y SB^μ comparten la propiedad asintótica de converger a cero si $\rho < 1$. Esta característica procede del término $\sum_{t=1}^T y_t^2$ del numerador (que en CS_1^μ es $\sum_{t=2}^T y_t^2$).

Existe, no obstante, una diferencia básica entre CS_1^μ y SB^μ que consiste en que en el numerador de SB^μ se toman diferencias respecto a la media incondicional, mientras que en CS_1^μ se toman diferencias respecto a la media condicionada a la primera observación. Esta característica tiene dos implicaciones. En primer lugar CS_1^μ tiene una interpretación en términos de predicción que SB^μ no posee. En segundo lugar, y más importante, en CS_1^μ se da más importancia a la primera observación. Un razonamiento análogo puede hacerse con CS_T^μ en (4.3.13), donde el mayor peso recae en la última observación.

No existe en la literatura una justificación teórica a favor de dar mayor importancia a la primera observación para detectar raíces unitarias. Sin embargo, implícitamente ha sido aplicado, con éxito, por varios autores, como se comenta a continuación.

El estimador por MCO de ρ utilizado, entre otros, en los contrastes de Dickey & Fuller (1979) pondera igualmente a todas las observaciones. Sin embargo la estimación por máxima verosimilitud exacta, bajo normalidad, proporciona los siguientes estimadores de μ y ρ ,

$$\hat{\mu}_{ML} = \frac{y_1 + (1 - \hat{\rho}_{ML}) \sum_{t=2}^T (y_t - \hat{\rho}_{ML} y_{t-1})^2}{1 + (T-1)(1 - \hat{\rho}_{ML})^2} = y_1 + O_p(T^{-\frac{1}{2}}) \quad (4.3.16)$$

$$\hat{\rho}_{ML} = \frac{\sum_{t=2}^T (y_t - \hat{\mu}_{ML})(y_{t-1} - \hat{\mu}_{ML})}{\sum_{t=2}^T (y_{t-1} - \hat{\mu}_{ML})^2}, \quad (4.3.17)$$

donde la primera observación tiene un peso mayor. Las simulaciones realizadas por Pantula et al (1994) muestran que un contraste basado en $\hat{\rho}_{ML}$ tiene mayor potencia que los contrastes de Dickey-Fuller.

La corrección que Elliot & Rothenberg & Stock (1992) realizan al contraste de Dickey-Fuller, y que se expuso en el capítulo 3, supone igualmente un mayor peso de la primera observación. Si se denota $\rho_c = 1 + c/T$ se tiene que la media local con que se corrigen las observaciones tiene la misma expresión que (4.3.16) pero sustituyendo el valor de ρ_c .

Puede observarse que estas dos modificaciones de los contrastes de Dickey-Fuller (máxima verosimilitud exacta en Pantula et al (1994) y mínimos cuadrados generalizados en Elliot, Rothenberg & Stock (1992)) son semejantes a la modificación que CS_1^μ supone respecto a SB^μ . En SB^μ se resta la media realizada con todas las observaciones y en CS^μ se resta sólo la primera observación.

4.3.3 Distribuciones asintóticas

En esta sección se calculan las distribuciones asintóticas de los estadísticos CS_1^μ , CS_T^μ , CS_{1T}^μ y CS_A^μ cuando el proceso es

$$y_t = \mu(1 - \rho) + \rho y_{t-1} + u_t, \quad u_t = \psi(B)a_t$$

y se verifica la hipótesis nula $H_o : \rho = 1$. También se obtendrán las distribuciones asintóticas de los estadísticos CS_1^τ y CS_A^τ cuando el proceso es

$$y_t = \mu(1 - \rho) + \delta\rho + \delta(1 - \rho)t + \rho y_{t-1} + u_t, \quad u_t = \psi(B)a_t$$

y se verifica la hipótesis nula $H_o : \rho = 1$. El símbolo \xrightarrow{d} significa convergencia débil en distribución de las medidas de probabilidad asociadas.

Teorema 4.1 *Sea y_t el proceso (3.2.2) con $\rho = 1$. Entonces*

$$CS_1^\mu \xrightarrow{d} \kappa^{-2} \int_0^1 W^2(r) dr, \quad (4.3.18)$$

$$CS_1^\tau \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 W^2(r) dr + \frac{1}{3} W^2(1) - 2W(1) \int_0^1 r W^2(r) dr \right\}, \quad (4.3.19)$$

$$CS_T^\mu \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 W^2(r) dr + W^2(1) - 2W(1) \int_0^1 W(r) dr \right\}, \quad (4.3.20)$$

$$CS_{1T}^\mu \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 W^2(r) dr + \frac{1}{2} W^2(1) - W(1) \int_0^1 W(r) dr \right\}, \quad (4.3.21)$$

$$CS_A^\mu \xrightarrow{d} \kappa^{-2} \int_0^1 \int_0^r (1 - (r - \lambda))^{-1} W^2(r - \lambda) d\lambda dr \quad (4.3.22)$$

$$CS_A^\tau \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 \int_0^r (1 - (r - \lambda))^{-1} W^2(r - \lambda) d\lambda dr + \frac{1}{3} W^2(1) \right. \quad (4.3.23)$$

$$\left. - 2W(1) \int_0^1 \int_0^r (1 - (r - \lambda))^{-1} (r - \lambda) W(r - \lambda) d\lambda dr \right\}, \quad (4.3.24)$$

donde

$$\kappa^2 = \frac{\sigma_u^2}{\omega^2},$$

$$\sigma_u^2 = E(u^2)$$

$$\omega^2 = \sigma^2 \psi^2(1).$$

Demostración:

La distribución de CS_1^μ es una aplicación directa del TCLF al proceso de sumas parciales $S_t = \sum_{j=2}^t u_j$. El estadístico puede reescribirse como

$$CS_1^\mu = \frac{(T-1)^{-2} \sum_{t=2}^T S_t^2}{\hat{\sigma}_u^2},$$

puede aplicarse, entonces, que (Phillips, 1987a)

$$(T-1)^{-2} \sum_{t=2}^T S_t^2 \xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 W^2(r) dr. \quad (4.3.25)$$

Dado que $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$ y que asintóticamente puede despreciarse la diferencia entre el factor $T-1$ y T se obtiene (4.3.18).

Para el cálculo de la distribución asintótica de CS_1^r se puede escribir

$$\begin{aligned} \frac{\sum_{t=2}^T (y_t - y_1 - \hat{\delta}(t-1))^2}{(T-1)^2} &= \frac{\sum_{t=2}^T (y_t - y_1 - \delta(t-1) - (\hat{\delta} - \delta)(t-1))^2}{(T-1)^2} \\ &= \frac{\sum_{t=2}^T (y_t - y_1 - \delta(t-1))^2}{(T-1)^2} \end{aligned} \quad (4.3.26)$$

$$+ \frac{(\hat{\delta} - \delta)^2 \sum_{t=2}^T (t-1)^2}{(T-1)^2} \quad (4.3.27)$$

$$- 2 \frac{(\hat{\delta} - \delta) \sum_{t=2}^T (y_t - y_1 - \delta(t-1))(t-1)}{(T-1)^2}. \quad (4.3.28)$$

El numerador del término (4.3.26) son sumas parciales de los términos u_t (véase (4.3.6)).

Por tanto, aplicando (4.3.25) se obtiene,

$$\frac{\sum_{t=2}^T (y_t - y_1 - \delta(t-1))^2}{(T-1)^2} \xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 W^2(r) dr. \quad (4.3.29)$$

Para resolver (4.3.27) se ha de aplicar que, por (4.3.7),

$$\hat{\delta} = \frac{y_T - y_1}{T-1} = \delta + \frac{\sum_{t=2}^T u_j}{T-1}$$

y por tanto

$$(\hat{\delta} - \delta) \xrightarrow{d} \frac{\sigma\psi(1)W(1)}{\sqrt{T}} \quad (4.3.30)$$

Además, aplicando que, cuando $T \rightarrow \infty$,

$$\frac{\sum_{t=1}^T t^k}{T^{(k+1)}} \rightarrow \frac{1}{k+1}, \quad k = 0, 1, \dots \quad (4.3.31)$$

se obtiene

$$\frac{(\hat{\delta} - \delta)^2 \sum_{t=2}^T (t-1)^2}{(T-1)^2} \xrightarrow{d} \frac{1}{3} \sigma^2 \psi(1)^2 W^2(1). \quad (4.3.32)$$

El término (4.3.28) puede reescribirse como

$$\begin{aligned} -2 \frac{(\hat{\delta} - \delta) \sum_{t=2}^T (y_t - y_1 - \delta(t-1)(t-1))}{(T-1)^2} &= \frac{\sum_{t=2}^T u_j \sum_{t=2}^T S_t(t-1)}{\sqrt{T-1} (T-1)^{5/2}} \\ &\quad - 2 \frac{(y_T - y_1) \sum_{t=2}^T (y_t - y_1)(t-1)}{(T-1)^{5/2}}. \end{aligned}$$

Aplicando (4.3.30) y que (Phillips & Perron 1988)

$$\frac{\sum_{t=2}^T S_t(t-1)}{(T-1)^{5/2}} \xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 rW(r)dr$$

se obtiene

$$-2 \frac{(\hat{\delta} - \delta) \sum_{t=2}^T (y_t - y_1 - \delta(t-1)(t-1))}{(T-1)^2} \xrightarrow{d} -2\sigma^2 \psi^2(1) W(1) \int_0^1 rW(r)dr. \quad (4.3.33)$$

Por (4.3.29), (4.3.32) y (4.3.33) y dado que $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$ se obtiene (4.3.19).

Para la obtención de la distribución asintótica de CS_T^μ se descompondrá el estadístico de la siguiente forma

$$\begin{aligned} CS_T^\mu &= \frac{(T-1)^{-2} \sum_{t=1}^{T-1} [(y_t - y_1) - (y_T - y_1)]^2}{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2} \\ &= \hat{\sigma}_u^{-2} (T-1)^{-2} \left\{ \sum_{t=1}^{T-1} (y_t - y_1)^2 + (T-1)(y_T - y_1)^2 - 2(y_T - y_1) \sum_{t=2}^T (y_t - y_1) \right\} \end{aligned}$$

La distribución del primer sumando se obtiene aplicando (4.3.25). Del segundo sumando resulta

$$\frac{(y_T - y_1)^2}{T - 1} \xrightarrow{d} \sigma^2 \psi^2(1) W^2(1).$$

La distribución del tercer sumando se obtiene al verificarse

$$2 \frac{(y_T - y_1) \sum_{t=2}^T (y_t - y_1)}{(T - 1)^{\frac{1}{2}} (T - 1)^{\frac{3}{2}}} \xrightarrow{d} -2\sigma^2 \psi^2(1) W(1) \int_0^1 W(r) dr, \quad (4.3.34)$$

obteniéndose la distribución (4.3.20). La distribución (4.3.21) se obtiene de forma inmediata de (4.3.18) y (4.3.20).

Para obtener la distribución asintótica de CS_A^μ se define el proceso $Z(r, \lambda)$, $\lambda < r$, como

$$Z(r, \lambda) = \frac{1}{T - 1} \frac{\sum_{[T\lambda]}^{[Tr]} u_t}{\sqrt{T - (r - \lambda)}} \quad (4.3.35)$$

donde $[\cdot]$ representa la parte entera del argumento. La suma parcial $\sum_{[T\lambda]}^{[Tr]} u_t$ es el error de predicción de la observación $y_{[Tr]}$ desde $y_{[T\lambda]}$ si se toma como predictor un paseo aleatorio. El término $T - (r - \lambda)$ es el número de sumas parciales con el mismo número de sumandos que pueden encontrarse en la muestra. Por ejemplo, existen $T - 1$ sumas parciales con un solo término, $T - 2$ sumas parciales con dos términos, etc. Aplicando el TCLF se obtiene

$$\begin{aligned} \int_0^1 \int_0^r Z^2(r, \lambda) d\lambda dr &= \frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s - t))^{-1} (y_s - y_t)^2}{(T - 1)^2} \\ &\xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 \int_0^r (1 - (r - \lambda))^{-1} W^2(r - \lambda) d\lambda dr. \end{aligned} \quad (4.3.36)$$

Para el estadístico CS_A^r se realizará una descomposición similar a la realizada para CS_1^r

$$\begin{aligned} \frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s - t))^{-1} (y_s - y_t - \hat{\delta}(s - t))^2}{(T - 1)^2} &= \\ \frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s - t))^{-1} (y_s - y_t - \delta(s - t))^2}{(T - 1)^2} &\quad (4.3.37) \end{aligned}$$

$$+ \frac{(\hat{\delta} - \delta)^2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T (s-t)^2}{(T-1)^2} \quad (4.3.38)$$

$$- 2 \frac{(\hat{\delta} - \delta) \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s-t))^{-1} (y_s - y_t - \delta(s-t))(s-t)}{(T-1)^2} \quad (4.3.39)$$

El término (4.3.37) es, aplicando (4.3.36),

$$\frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s-t))^{-1} (y_s - y_t - \delta(s-t))^2}{(T-1)^2} \xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 \int_0^r (1 - (r-\lambda))^{-1} W^2(r-\lambda) d\lambda dr. \quad (4.3.40)$$

En el término (4.3.38) hay que observar que la expresión $(T - (s-t))$ es el número de veces que aparece repetido el factor $(s-t)$, por lo que resulta

$$\sum_{t=1}^{T-1} \sum_{s=t+1}^T \frac{(s-t)^2}{T - (s-t)} = \sum_{t=1}^{T-1} \sum_{s=t+1}^T (s-t) = \sum_{t=1}^{T-1} t^2,$$

y, por tanto, aplicando (4.3.31) y (4.3.30), se obtiene

$$\frac{(\hat{\delta} - \delta)^2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T (s-t)^2}{(T-1)^2} \xrightarrow{d} \frac{\sigma^2 \psi^2(1)}{3} W^2(1). \quad (4.3.41)$$

Finalmente, utilizando los mismos argumentos que para CS_1^T , se obtiene

$$- 2 \frac{(\hat{\delta} - \delta) \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s-t))^{-1} (y_s - y_t - \delta(s-t))(s-t)}{(T-1)^2} \xrightarrow{d} - 2\sigma^2 \psi^2(1) W(1) \int_0^1 \int_0^r (1 - (r-\lambda))^{-1} (r-\lambda) W(r-\lambda) d\lambda dr. \quad (4.3.42)$$

Agregando (4.3.40), (4.3.41) y (4.3.42) se deriva (4.3.24). \square

Puede observarse que las distribuciones asintóticas obtenidas para CS_1^T y CS_A^T son independientes del valor de δ y del valor de y_0 .

Cuando la componente estacionaria u_t es independiente e idénticamente distribuida se tiene que $\sigma_u^2 = \sigma^2$ y por tanto $\psi(1) = 1$ y $\kappa^2 = 1$. En este caso las distribuciones

asintóticas no dependen de parámetros perturbadores. En el caso I(1) general, en que $u_t = \psi(B)a_t$, las distribuciones asintóticas dependen de los parámetros ψ_j , σ_u^2 y σ^2 . En las secciones 4.4 y 4.6 se proponen modificaciones en los contrastes CS para que, en el caso I(1), sus distribuciones asintóticas no dependan de parámetros perturbadores y puedan emplearse los mismos valores críticos que para el caso AR(1).

El que u_t sea independiente e idénticamente distribuido es condición suficiente pero no necesaria para que se verifique que $\sigma_u^2 = \sigma^2$. Esta igualdad se cumple aún en el caso en que u_t sea una diferencia de martingalas con ciertas restricciones en sus momentos (Phillips & Perron 1988).

4.3.4 Valores críticos

Para la obtención de los valores críticos se ha realizado una simulación por Monte Carlo. Al no depender las distribuciones asintóticas del valor de δ se han realizado las simulaciones considerando que $\delta = 0$, por lo que el proceso simulado es, para todos los contrastes,

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

con $\rho = 1$, $a_t \sim N(0,1)$ y $y_1 = a_1$. Los tamaños muestrales empleados han sido $T = 25, 50, 100, 250$. Se han realizado 20.000 replicaciones de cada experimento y se ha obtenido el percentil 0.05 de la distribución empírica. Estos valores críticos se recogen en la tabla 4.1.

Las distribuciones asintóticas de los estadísticos para los contrastes de Dickey-Fuller no dependen tampoco de δ . No obstante Evans & Savin (1984) demuestran que sí existe una influencia de dicho parámetro en muestras finitas. Estos autores proponen un contraste

Tabla 4.1: Valores críticos al 5% para los contrastes CS^μ y CS^τ

Contraste	T			
	25	50	100	250
CS_1^μ	0.0657	0.0624	0.0578	0.0583
CS_T^μ	0.0661	0.0614	0.0589	0.0572
CS_{1T}^μ	0.0799	0.0745	0.0743	0.0702
CS_A^μ	0.0930	0.0873	0.0826	0.0801
CS_1^τ	0.0414	0.0394	0.0376	0.0374
CS_A^τ	0.0616	0.0573	0.0561	0.0549

exacto que incorpora la información de dicho parámetro. Las mejoras que se producen son modestas. Es razonable suponer que en el presente contraste puedan esperarse similares conclusiones.

4.3.5 Potencia en muestras finitas. Caso AR(1)

En esta sección se muestra un ejercicio de simulación para comparar la potencia de los contrastes CS_1 , CS_{1T} y CS_A con los existentes en la literatura. El proceso simulado es $y_t = \rho y_{t-1} + a_t$ con $a_t \sim N(0,1)$ y $y_1 = a_1$. Los valores críticos empleados en el contraste son los de la tabla 4.1. La potencia se ha evaluado para las alternativas $\rho = 0.99, 0.97, 0.95, 0.90, 0.85, 0.80, 0.70$ mediante 10.000 replicaciones. Los tamaños muestrales considerados son $T = 25, 50, 100, 250$. Las tablas 4.2 a 4.9 contienen las potencias observadas. Estas tablas contienen además los valores de las potencias empíricas para los contrastes de Dickey-Fuller, Sargan-Bhargava y Elliot-Rothenberg-Stock expuestos en

el anterior capítulo. Para facilitar la comparación entre los diferentes contrastes se han construido los gráficos 4.2 a 4.9. Ha de decirse que el contraste de Elliot Rothenberg & Stock es el más potente que puede encontrarse en la literatura.

Una primera conclusión que se extrae de la comparación de los diferentes contrastes es que la potencia para una alternativa estacionaria ha de interpretarse en relación al tamaño muestral. Por ejemplo, se obtienen potencias similares para las siguientes combinaciones de ρ y T : $\rho = 0.7$ y tamaño muestral de $T = 25$, $\rho = 0.85$ con $T = 50$, ρ entre 0.95 y 0.90 con $T = 100$ y $\rho = 0.97$ en $T = 250$. De esta forma a medida que aumenta el tamaño muestral pierde interés la zona más alejada a la unidad. Por esta razón en la literatura se utiliza el término *localidad* para describir a las alternativas que, en función del tamaño muestral, estén próximas a la unidad. Lejos de la localidad la potencia de todos los contrastes es muy alta.

La comparación de las curvas de potencia muestra que la potencia de CS_A tiene un perfil semejante a los contrastes de los otros autores. Estos perfiles son, a su vez, diferentes a los que presentan CS_1 y CS_{1T} , que son muy similares.

Los valores de potencia del contraste CS_A^μ se sitúan entre los contrastes de Sargan-Bhargava (SB^μ) y el contraste de Elliot-Rothenberg-Stock (τ_{GLS}^μ) y siempre se posiciona por encima de los contrastes de Dickey-Fuller. En el caso de tendencia determinista, el estadístico CS_A^τ tiene un comportamiento similar al contraste de Sargan-Bhargava (SB^τ) salvo en zonas alejadas de la localidad. En estas zonas el comportamiento relativo del contraste empeora al aumentar el tamaño muestral.

Los contrastes CS_1^μ y CS_{1T}^μ tienen mejor comportamiento que CS_A^μ . En general también

puede decirse que CS_1^μ tiene mejor comportamiento que CS_{1T}^μ . Ambos contrastes son muy similares pero en muestras grandes presentan diferencias. CS_1^μ tiene mejor comportamiento que los contrastes de los otros autores en $T = 25$, salvo en $\rho = 0.7$. A medida que aumenta el tamaño muestral va empeorando su comportamiento relativo en zonas alejadas de la localidad, mientras que en la localidad se sitúa en valores semejantes al contraste de Elliot-Rothenberg-Stock.

La conclusión más importante que se extrae del presente ejercicio de simulación es que los contrastes CS_1^μ y CS_1^T tienen, en la localidad, comportamientos muy similares, e incluso superiores, al contraste de Elliot-Rothenberg-Stock.

Tabla 4.2: Potencia empírica para el nivel 5%. $T = 25$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^μ	0.06	0.07	0.09	0.13	0.18	0.24	0.37
CS_{1T}^μ	0.06	0.07	0.09	0.13	0.18	0.24	0.38
CS_A^μ	0.06	0.07	0.08	0.12	0.17	0.22	0.38
$T(\hat{\rho}^\mu - 1)$	0.05	0.06	0.07	0.10	0.14	0.18	0.31
$\hat{\tau}^\mu$	0.05	0.05	0.06	0.06	0.09	0.11	0.18
SB^μ	0.05	0.07	0.08	0.11	0.16	0.21	0.35
τ_{GLS}^μ	0.06	0.07	0.09	0.12	0.17	0.22	0.39

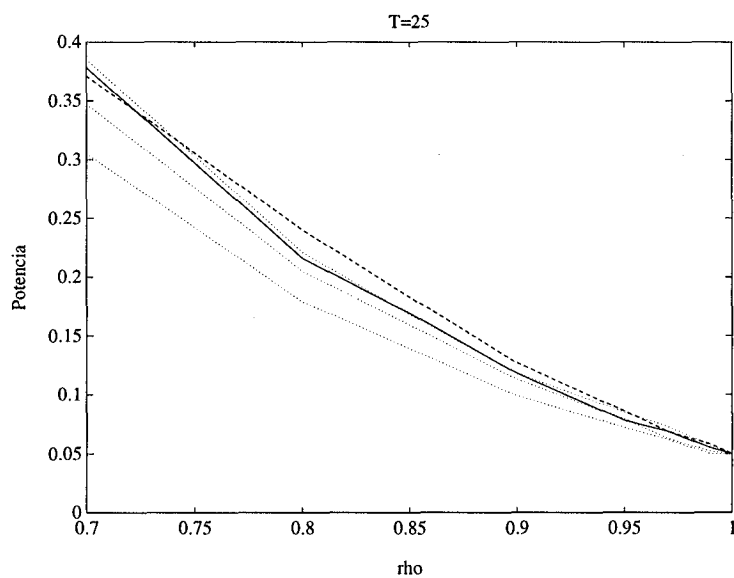


Figura 4.2: Potencia empírica: $-- CS_1^\mu$, $- CS_A^\mu$. Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^\mu - 1)$, SB^μ , τ_{GLS}^μ .

Tabla 4.3: Potencia empírica para el nivel 5%. $T = 50$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^μ	0.07	0.10	0.14	0.27	0.45	0.59	0.77
CS_{1T}^μ	0.06	0.10	0.14	0.27	0.43	0.56	0.77
CS_A^μ	0.06	0.08	0.12	0.23	0.38	0.56	0.84
$T(\hat{\rho}^\mu - 1)$	0.06	0.08	0.11	0.18	0.31	0.47	0.79
$\hat{\tau}^\mu$	0.05	0.06	0.07	0.11	0.20	0.32	0.64
SB^μ	0.06	0.10	0.12	0.22	0.37	0.55	0.86
τ_{GLS}^μ	0.06	0.09	0.13	0.26	0.44	0.62	0.89

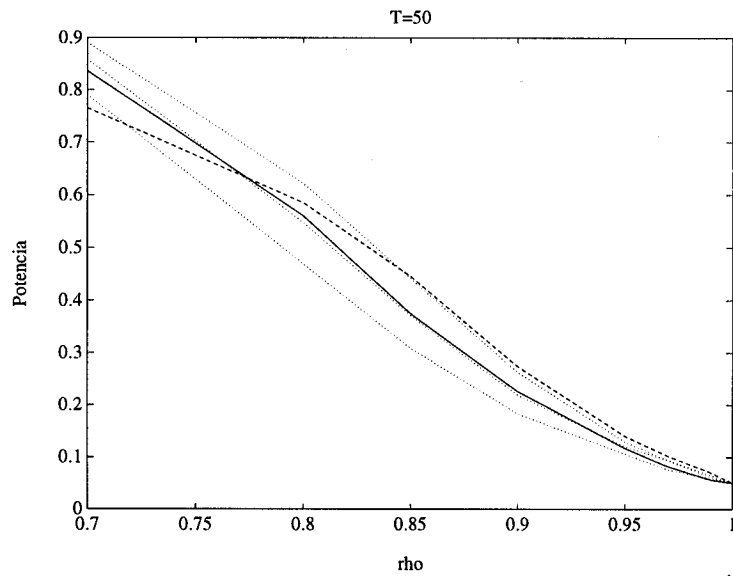
**Figura 4.3:** Potencia empírica: -- CS_1^μ , — CS_A^μ . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^\mu - 1)$, SB^μ , τ_{GLS}^μ .

Tabla 4.4: Potencia empírica para el nivel 5%. $T = 100$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^μ	0.08	0.15	0.28	0.63	0.84	0.92	0.97
CS_{1T}^μ	0.08	0.16	0.28	0.62	0.83	0.92	0.98
CS_A^μ	0.08	0.15	0.24	0.58	0.85	0.97	1.00
$T(\hat{\rho}^\mu - 1)$	0.07	0.12	0.19	0.47	0.79	0.95	1.00
$\hat{\tau}^\mu$	0.06	0.08	0.11	0.29	0.61	0.86	1.00
SB^μ	0.07	0.13	0.22	0.52	0.83	0.97	1.00
τ_{GLS}^μ	0.08	0.17	0.28	0.66	0.91	0.98	1.00

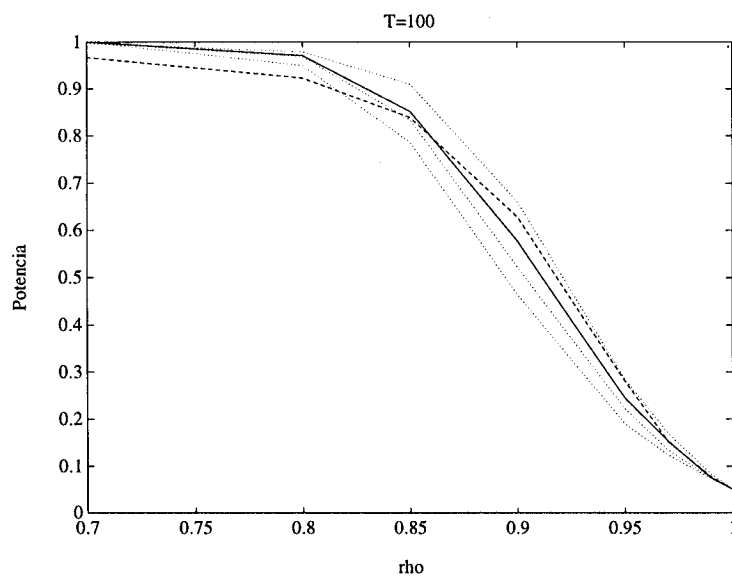


Figura 4.4: Potencia empírica: $--$ CS_1^μ , $-$ CS_A^μ . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^\mu - 1)$, SB^μ , τ_{GLS}^μ .

Tabla 4.5: Potencia empírica para el nivel 5%. $T = 250$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^μ	0.14	0.51	0.83	0.99	1.00	1.00	1.00
CS_{1T}^μ	0.14	0.47	0.77	0.97	1.00	1.00	1.00
CS_A^μ	0.12	0.39	0.72	1.00	1.00	1.00	1.00
$T(\hat{\rho}^\mu - 1)$	0.11	0.32	0.64	0.99	1.00	1.00	1.00
$\hat{\tau}^\mu$	0.07	0.19	0.46	0.97	1.00	1.00	1.00
SB^μ	0.12	0.36	0.69	1.00	1.00	1.00	1.00
τ_{GLS}^μ	0.14	0.52	0.86	1.00	1.00	1.00	1.00

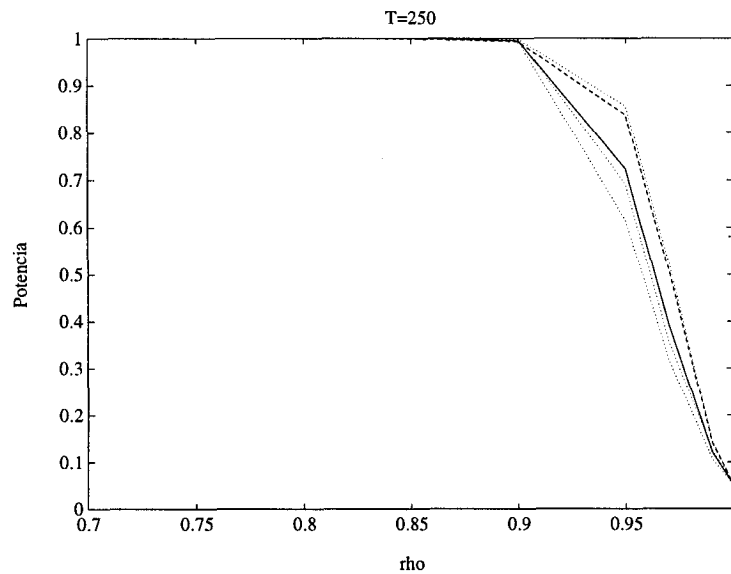
**Figura 4.5:** Potencia empírica: -- CS_1^μ , — CS_A^μ . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^\mu - 1)$, SB^μ , τ_{GLS}^μ .

Tabla 4.6: Potencia empírica para el nivel 5%. $T = 25$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^T	0.05	0.06	0.06	0.07	0.09	0.12	0.18
CS_A^T	0.05	0.05	0.06	0.07	0.08	0.10	0.17
$T(\hat{\rho}^T - 1)$	0.05	0.05	0.06	0.06	0.08	0.10	0.16
$\hat{\tau}^T$	0.05	0.05	0.05	0.05	0.07	0.08	0.12
SB^T	0.05	0.05	0.05	0.06	0.08	0.10	0.16
τ_{GLS}^T	0.05	0.06	0.06	0.06	0.08	0.10	0.16

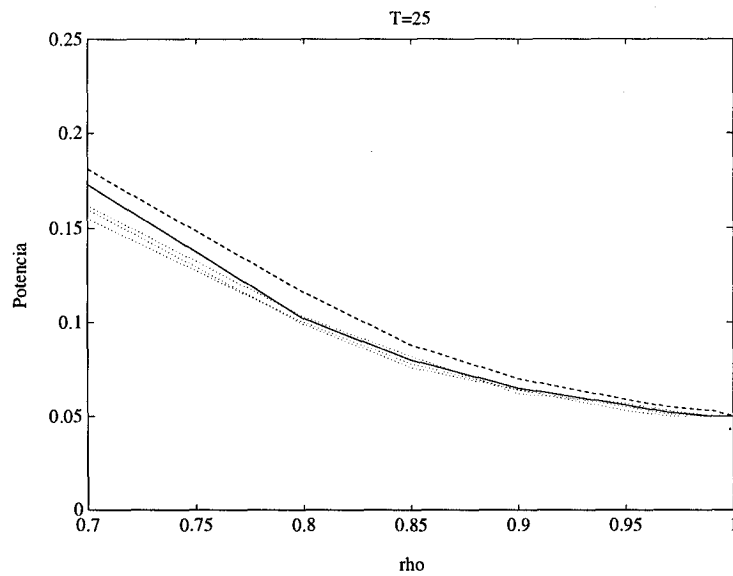


Figura 4.6: Potencia empírica: $-- CS_1^T$, $— CS_A^T$. Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^T - 1)$, SB^T , τ_{GLS}^T .

Tabla 4.7: Potencia empírica para el nivel 5%. $T = 50$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^T	0.05	0.05	0.07	0.11	0.19	0.28	0.51
CS_A^T	0.05	0.05	0.07	0.10	0.17	0.26	0.51
$T(\hat{\rho}^T - 1)$	0.05	0.06	0.07	0.09	0.15	0.24	0.49
$\hat{\tau}^T$	0.05	0.06	0.06	0.08	0.12	0.19	0.40
SB^T	0.05	0.06	0.07	0.10	0.17	0.27	0.53
τ_{GLS}^T	0.05	0.06	0.07	0.11	0.18	0.29	0.58

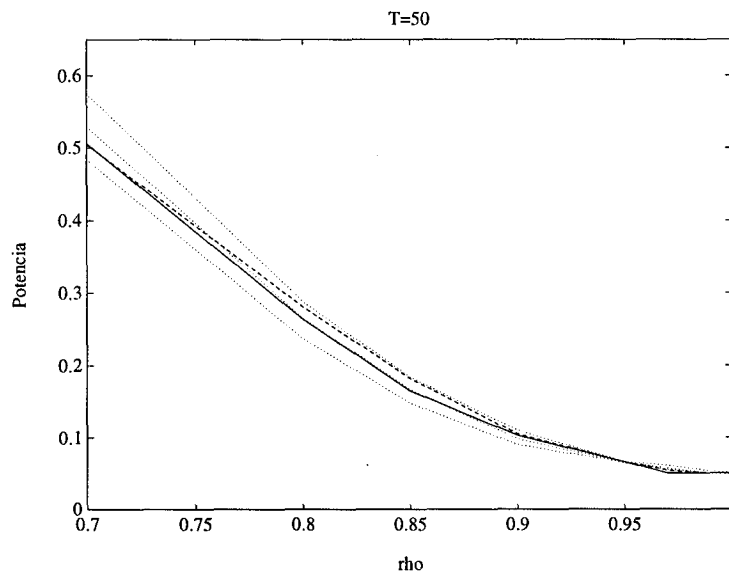


Figura 4.7: Potencia empírica: $--$ CS_1^T , $-$ CS_A^T . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^T - 1)$, SB^T , τ_{GLS}^T .

Tabla 4.8: Potencia empírica para el nivel 5%. . $T = 100$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^T	0.06	0.07	0.12	0.30	0.53	0.72	0.91
CS_A^T	0.06	0.08	0.11	0.27	0.52	0.73	0.93
$T(\hat{\rho}^T - 1)$	0.05	0.07	0.10	0.24	0.48	0.74	0.99
$\hat{\tau}^T$	0.05	0.07	0.08	0.19	0.40	0.64	0.96
SB_T^B	0.06	0.07	0.11	0.27	0.53	0.77	0.97
τ_{GLS}^T	0.06	0.07	0.11	0.30	0.58	0.83	0.99

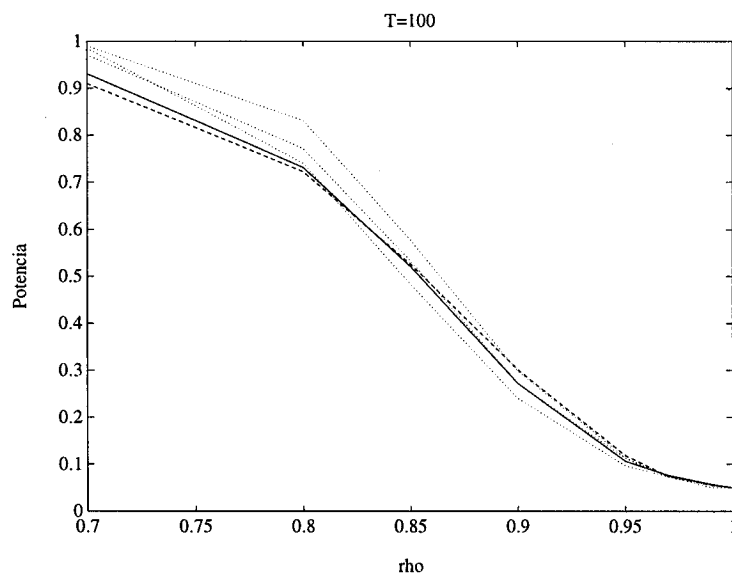


Figura 4.8: Potencia empírica: -- CS_1^T , — CS_A^T . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^T - 1)$, SB_T^B , τ_{GLS}^T .

Tabla 4.9: Potencia empírica para el nivel 5%. $T = 250$, 10.000 replicaciones

Contraste	ρ						
	0.99	0.97	0.95	0.90	0.85	0.80	0.70
CS_1^T	0.07	0.19	0.43	0.87	0.97	0.99	1.00
CS_A^T	0.06	0.17	0.38	0.87	0.98	1.00	1.00
$T(\hat{\rho}^T - 1)$	0.06	0.16	0.34	0.90	1.00	1.00	1.00
$\hat{\tau}^T$	0.06	0.13	0.27	0.83	1.00	1.00	1.00
SB^T	0.07	0.18	0.39	0.92	1.00	1.00	1.00
τ_{GLS}^T	0.07	0.19	0.45	0.96	1.00	1.00	1.00

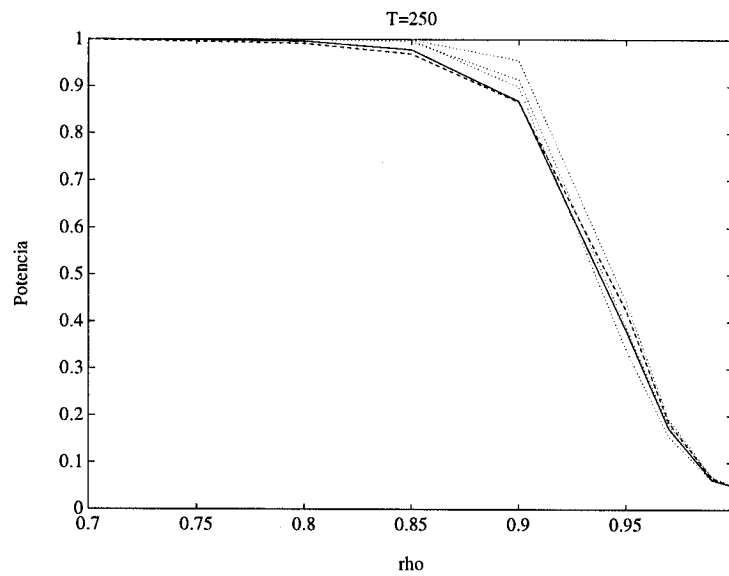


Figura 4.9: Potencia empírica: $--$ CS_1^T , $-$ CS_A^T . Las líneas de puntos son, de abajo a arriba, $T(\hat{\rho}^T - 1)$, SB^T , τ_{GLS}^T .

4.4 Extension al caso $\psi(B)$ general

En esta sección se presentan los dos procedimientos que se emplearán para extender los contrastes propuestos al caso I(1) general. En ambos procedimientos se proponen modificaciones de los contrastes CS que incorporen la información de la estructura dinámica de u_t . Los nuevos estadísticos que resultan tienen las mismas distribuciones asintóticas que en el caso AR(1). El primer procedimiento se basa en la utilización de un estimador consistente de κ^2 con el que se multiplicará a cada estadístico, resultando los estadísticos \overline{CS} . El comportamiento de estos contrastes se analiza en la sección 4.5. En el segundo procedimiento, además de un estimador consistente de κ^2 , se modeliza el componente u_t . La modelización de u_t proporciona predicciones más eficientes que las generadas por el predictor paseo aleatorio empleado en CS. A los estadísticos de contraste que incorporan dichas predicciones se les denomina ECS, donde E denota *Efficient*. Su comportamiento se analiza en 4.6.

Los dos procedimientos que se proponen para el caso I(1) general tienen diferente justificación. Los contrastes \overline{CS} son la extensión directa de los estadísticos CS al caso general y no incorporan más información adicional que la necesaria y suficiente para poder utilizar sus mismas distribuciones asintóticas. Puesto que la evolución de la serie temporal aparece dominada, en el largo plazo, por la raíz unitaria, es deseable que un contraste de raíces unitarias no necesite de la información referente al corto plazo. Por tanto bastaría con utilizar un estimador consistente de κ^2 , aunque no sea necesariamente eficiente. Además la modelización de u_t no estará exenta de incertidumbre y puede suceder que no se consiga un modelo satisfactorio.

La justificación para emplear los contrastes ECS es clara. Si u_t tiene dependencia temporal el predictor paseo aleatorio no será óptimo. Es de esperar, entonces, que la modelización

de u_t proporcione predicciones que mejoren el comportamiento del contraste en muestras finitas. Esta mejora será tanto mayor cuanto más alejada esté la raíz de interés del círculo unitario, al dejar de dominar sobre la estructura de u_t .

4.4.1 Estimadores consistentes de ω^2

La extensión de los contrastes CS al caso I(1) general precisa de la estimación consistente del término $\omega^2 = \sigma^2\psi(1)$. Para la estimación de ω^2 existen dos alternativas en la literatura. La primera de ellas, propuesta por Phillips (1987a), consiste en una estimación no paramétrica de la covarianza mediante la expresión

$$\hat{\omega}_c^2 = \sum_{m=-l_T}^{l_T} k\left(\frac{m}{l_T}\right) \hat{\gamma}_a(m), \quad (4.4.1)$$

donde

$$\hat{\gamma}_x(m) = (T - m)^{-1} \sum_{t=|m|+1}^T (x_t - \bar{x})(x_{t-m} - \bar{x}),$$

$k(\cdot)$ es una función kernel y \hat{a}_t son los residuos de una regresión de y_t sobre $(1, y_{t-1})$ o $(1, t, y_{t-1})$. La elección apropiada de la función kernel (Newey & West, 1987) junto a la adecuada elección de l_T en función del tamaño muestral aseguran que

$$\hat{\omega}_c^2 \xrightarrow{p} \omega^2.$$

Un segundo procedimiento consiste en el estimador autorregresivo de la densidad espectral (Fuller 1976, Dickey & Savin 1984)

$$\hat{\omega}_{AR}^2 = \frac{\hat{\sigma}^2}{\left(1 - \sum_{j=1}^p \hat{\phi}_j^2\right)^2} \quad (4.4.2)$$

donde $\hat{\phi}_j$ son los coeficientes estimados por MCO de la regresión en forma de corrección de error

$$\Delta y_t = \delta_0 + (\rho - 1)y_{t-1} + \sum_{j=1}^p \hat{\phi}_j^2 + a_t \quad (4.4.3)$$

con $\delta_0 = \beta_0$ o $\delta_0 = \beta_0 + \beta_1 t$ según sea el componente determinista y $\hat{\sigma}^2$ la varianza residual de esta regresión. Si la elección de p en función del tamaño muestral es tal que $p \rightarrow \infty$ cuando $T \rightarrow \infty$ pero $p^3/T \rightarrow 0$ (Berck, 1984) este estimador es consistente,

$$\hat{\omega}_{AR} \xrightarrow{p} \omega^2.$$

La elección del estimador de ω tiene una gran influencia en la distorsión que la estructura dinámica de u_t introduce en la distribución de los estadísticos de los contrastes. Resultados empíricos (Schwert 1989, Elliot et al. 1992,) muestran que, en presencia de componente de medias móviles, el estimador $\hat{\omega}_{AR}^2$ proporciona mejores resultados que $\hat{\omega}_c^2$, en el sentido de que la distribución de los estadísticos de contraste se aleja menos de la que se tiene en el caso $u_t = a_t$.

El origen de estas dos propuestas es la extensión de los contrastes originales de Dickey-Fuller a procesos I(1). En particular $\hat{\omega}_{AR}^2$ se obtiene en la misma regresión de MCO que los estadísticos de Dickey-Fuller. Como los contrastes CS no están basados en las propiedades de los estimadores de MCO puede utilizarse un estimador de ω^2 que se aproxime más a la estructura de u_t . Puede así utilizarse un tercer estimador, $\hat{\omega}_{arma}^2$, resultante de la estimación del modelo

$$\phi(B)(1 - \rho B)x_t = \theta(B)a_t, \quad (4.4.4)$$

donde $x_t = y_t - \mu - \delta t$. El estimador $\hat{\omega}_{arma}^2$ vendrá dado por la expresión

$$\hat{\omega}_{arma}^2 = \hat{\sigma}^2 \frac{(1 - \sum_{i=1}^q \hat{\theta}_i)^2}{(1 - \sum_{j=1}^p \hat{\phi}_j)^2}. \quad (4.4.5)$$

Este estimador será consistente si lo son los estimadores de los coeficientes y la varianza residual.

Este estimador $\hat{\omega}_{arma}^2$ no es más complejo de obtener que $\hat{\omega}_c^2$ o $\hat{\omega}_{AR}^2$. Aunque aparentemente en $\hat{\omega}_c^2$ o $\hat{\omega}_{AR}^2$ no sea necesario especificar el modelo, su optimalidad depende de una

adecuada selección del número de retardos empleados en cada uno. Las propiedades de los contrastes basados en $\hat{\omega}_c^2$ o $\hat{\omega}_{AR}^2$ son, además, muy sensibles a esta elección. Asimismo, no se ha encontrado aún una solución satisfactoria para decidir éste número de retardos (Perron & Ng, 1995).

4.5 Contrastes modificados $\overline{\text{CS}}$

4.5.1 Estadísticos de contraste

En esta sección se proponen modificaciones de los contrastes CS mediante la aplicación de un estimador consistente de κ^2 .

Un estimador de κ^2 puede obtenerse de

$$\hat{\kappa}^2 = \frac{\hat{\sigma}_u^2}{\hat{\omega}^2} \quad (4.5.1)$$

donde $\hat{\sigma}_u^2$

$$\hat{\sigma}_u^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T-1} \xrightarrow{p} \sigma_u^2,$$

y en el caso con tendencia lineal determinista, utilizando el estimador $\hat{\delta}$ mostrado en (4.3.8)

$$\hat{\sigma}_u^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}{T-2} \xrightarrow{p} \sigma_u^2.$$

Nótese que este estimador de σ_u^2 corresponde al denominador de los contrastes CS propuestos. Si se denota $\overline{\text{CS}}$ a los contrastes modificados tal que $\overline{\text{CS}} = \hat{\kappa}^2 \text{CS}$, se obtiene

$$\overline{\text{CS}}_1^\mu = \frac{(T-1)^{-2} \sum_{t=1}^T (y_t - y_1)^2}{\hat{\omega}^2}, \quad (4.5.2)$$

$$\overline{\text{CS}}_1^r = \frac{(T-1)^{-2} \sum_{t=1}^T (y_t - y_1 - \hat{\delta}(t-1))^2}{\hat{\omega}^2}, \quad (4.5.3)$$

$$\overline{CS}_{1T}^{\mu} = \frac{(T-1)^{-2} \left\{ \sum_{t=2}^T (y_t - y_1)^2 + \sum_{t=1}^{T-1} (y_t - y_T)^2 \right\}}{2\hat{\omega}^2}, \quad (4.5.4)$$

$$\overline{CS}_A^{\mu} = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s-t))^{-1} (y_s - y_t)^2}{\hat{\omega}^2}, \quad (4.5.5)$$

$$\overline{CS}_A^{\tau} = \frac{(T-1)^{-2} \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T - (s-t))^{-1} (y_s - y_t - \hat{\delta}(s-t))^2}{\hat{\omega}^2}. \quad (4.5.6)$$

Al verificarse que $\hat{\kappa}^2 \xrightarrow{p} \kappa^2$ las distribuciones de los estadísticos \overline{CS} serán las mismas que se obtuvieron en el teorema (4.1) pero sin el término κ^{-2} y por tanto los contrastes basados en estos estadísticos modificados son similares. Este resultado se muestra en el siguiente corolario.

Corolario 4.1 *Sea CS el conjunto de estadísticos CS_1^{μ} , CS_{1T}^{μ} , CS_A^{μ} , CS_1^{τ} , CS_A^{τ} . Se denota por $\kappa^{-2} dCS$ a las distribuciones asintóticas de los mismos obtenidas en el teorema 4.1. Sea \overline{CS} el conjunto de los estadísticos $\overline{CS} = \hat{\kappa}^2 CS$ y $\hat{\kappa}^2$ un estimador consistente de κ^2 . Entonces*

$$\overline{CS} \xrightarrow{d} dCS$$

4.5.2 Potencia en muestras finitas de \overline{CS}

Para la evaluación de la potencia en muestras finitas se ha realizado un experimento por Monte Carlo en el que el proceso generador de datos sigue el modelo

$$y_t = \rho y_{t-1} + a_t - \theta a_{t-1}, \quad (4.5.7)$$

con $a_t \sim N(0,1)$ y $y_1 = a_1$. Este diseño es habitual para el estudio del comportamiento de un contraste en muestras finitas, pues permite observar, con un modelo sencillo, cómo varia el nivel de significación y la potencia según el valor de $\psi(1)$. En este caso $\psi(1) = 1 - \theta$. Valores altos, en valor absoluto, de θ proporcionan valores de $\psi(1)$ alejados de la unidad y por tanto cambiará sustancialmente la distribución asintótica de los estadísticos CS . La

modificación introducida en los estadísticos \overline{CS} elimina esta influencia asintóticamente, pero en muestras moderadas o pequeñas puede aún existir dicho efecto y afectar negativamente al contraste.

Se han planteado dos ejercicios de simulación. Para el primer ejercicio de simulación se supone que existe una identificación correcta del modelo ARMA(1,1) y se ha utilizado el estimador $\hat{\omega}_{arma}^2$. Las estimaciones se han realizado utilizando máxima verosimilitud condicionada (mínimos cuadrados no lineales). La minimización de las respectivas funciones objetivo se ha realizado utilizando las funciones FMIN y FMINS de MATLAB. El punto óptimo se ha buscado alrededor de unos valores iniciales. Se ha utilizado la media muestral como valor inicial de la media. Asimismo se ha utilizado el estimador (4.3.8) como valor inicial de la pendiente. Para la obtención de valores iniciales del parámetro autorregresivo ρ y de media móvil θ se ha utilizado la relación que existe entre estos parámetros y los pesos π_j de su representación autorregresiva. Esta representación cumple que

$$\pi(B)x_t = a_t, \quad \pi(B) = (1 - \rho B)(1 - \theta B)^{-1}$$

y por tanto

$$\begin{aligned}\pi_1 &= \rho - \theta, \\ \pi_2 &= \rho(\rho - \theta) - (\rho - \theta)^2.\end{aligned}$$

Se verifica entonces que

$$\theta = \frac{\pi_2}{\pi_1}, \quad \rho = \pi_1 - \theta.$$

Para la aplicación de este resultado se ha estimado un AR(6), por MCO, a la serie corregida de componentes deterministas.

En el segundo ejercicio de simulación se utiliza el estimador $\hat{\omega}_{AR}^2$. Este estimador se obtendrá seleccionando el valor de p en (4.4.3) mediante el criterio BIC (Bayesian information criteria) mediante el cuál p es el valor que minimiza

$$\text{BIC}(p) = \log \hat{\sigma}_p^2 + p \log T/T, \quad (4.5.8)$$

siendo $\hat{\sigma}_p^2$ la varianza residual de la regresión estimada por MCO,

$$\hat{\sigma}_p^2 = \frac{\sum_{j=k+1}^T \hat{a}_j^2}{T - k - 1}, \quad (4.5.9)$$

donde \hat{a}_j son los residuos de la regresión y k es el número de parámetros de la regresión ($k = p + 2$ si el modelo tiene constante y $k = p + 3$ si contiene además tendencia lineal).

El criterio BIC fue introducido independientemente por Schwartz (1978), Akaike (1977) y Rissanen (1978). Con la estimación de σ^2 mencionada en (4.5.9) Hannan & Quinn (1979) demostraron que el criterio BIC es fuertemente consistente en la modelización de procesos autorregresivos estacionarios. Tsay (1984) demostró que este criterio, con (4.5.9), es débilmente consistente en la modelización de procesos autorregresivos no estacionarios. Para la presente simulación se ha utilizado el criterio BIC utilizando dos restricciones distintas sobre la elección de p . En un primer caso se ha utilizado el criterio con $0 \leq p \leq 8$ para evitar valores excesivamente elevados de p . Algunos autores han comprobado que, cuando el valor de θ es elevado y positivo el criterio BIC (y también el AIC -Akaike information criteria-) proporciona valores de p más bajos de los necesarios para una correcta modelización. Por esta razón se ha incluido un segundo caso donde la selección de p cumple $3 \leq p \leq 8$. Este segundo criterio de selección ha sido utilizado previamente en Elliot, Rothenberg & Stock (1992).

Resultados de la simulación 1: estimador $\hat{\omega}_{arma}^2$

Este ejercicio de simulación ha sido realizado, por su coste computacional, sólo para el contraste \overline{CS}_1^μ y \overline{CS}_1^τ . El tamaño muestral es $T = 100$ y se han utilizado los valores $\rho = 1, 0.95, 0.90, 0.80$, y $\theta = -0.8, -0.5, 0, 0.5, 0.8$. Para cada combinación de parámetros se realizan 5000 replicaciones en \overline{CS}_1^μ y 1000 replicaciones en \overline{CS}_1^τ y se contabiliza el porcentaje de veces que se rechaza la hipótesis nula de raíz unitaria. Los valores críticos utilizados son los de la tabla 4.1. El resultado se muestra en las tablas 4.10 y 4.11.

En estas tablas se ha incluido los resultados del contraste $\hat{\tau}_{SD}$ de Said & Dickey (1985), basado en el estadístico t de regresión. Este contraste se construye a partir de la estimación de un modelo ARMA(1,1) mediante una iteración del método Gauss-Newton. Los datos que aparecen en las tablas corresponden a potencias empíricas publicadas por diversos autores. Los resultados de $\hat{\tau}_{SD}^\mu$ corresponden al trabajo original de Said & Dickey (1985), donde los valores iniciales (a partir de los cuales se realiza una iteración por Gauss-Newton) son $\rho_o = 1$ y la estimación por el método de Durbin (1959) para θ_o . Los resultados de $\hat{\tau}_{SD-S}^\mu$ y $\hat{\tau}_{SD-S}^\tau$ son los publicados en Schwert (1989) donde, en lugar de una sola iteración del método Gauss-Newton, se itera hasta conseguir convergencia en las estimaciones. Otro aspecto en el que los resultados de $\hat{\tau}_{SD-S}^\mu$ y $\hat{\tau}_{SD-S}^\tau$ difieren respecto al contraste de Said & Dickey (1985) es en el valor de la primera observación. Mientras en Said & Dickey (1985) se considera $y_0 = 0$ en Schwert (1989) se generan series de tamaño $20 + T$ y se eliminan las primeras 20 observaciones.

Los valores de $\hat{\tau}_{SD-AN}^\mu$ son los resultados del trabajo de Agiakloglou & Newbold (1991), donde emplean una iteración del método Gauss-Newton con valores iniciales $\rho_o = 1$ y el resultado del método de máxima verosimilitud exacta en el modelo ARIMA(0,1,1) como estimación inicial de θ_o .

Tabla 4.10: Potencia empírica para el nivel 5% del contraste \overline{CS}_1^μ , con $\hat{\omega}_{arma}^2$, y otros contrastes. $T = 100$.

Contraste	ρ	θ				
		-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.06	0.04	0.05	0.05	0.11
	0.95	0.31	0.31	0.27	0.17	0.31
	0.90	0.71	0.70	0.63	0.31	0.54
	0.80	0.99	0.97	0.90	0.48	0.67
$\hat{\tau}_{SD}^\mu$	1.00	nd	0.05	nd	0.05	0.11
	0.95	nd	0.11	nd	0.12	0.27
	0.90	nd	0.30	nd	0.28	0.46
	0.80	nd	0.83	nd	0.63	0.71
$\hat{\tau}_{SD-S}^\mu$	1.00	0.06	0.04	0.04	0.02	0.05
$\hat{\tau}_{SD-AN}^\mu$	1.00	nd	nd	nd	0.05	0.09
	0.95	nd	nd	nd	nd	nd
	0.90	nd	nd	0.29	nd	nd
	0.80	nd	nd	0.74	nd	nd
τ_{GLS}^μ ($3 \leq p \leq 8$)	1.00	0.07	0.05	0.05	0.08	0.38
	0.95	0.33	0.30	0.26	0.31	0.71
	0.90	0.63	0.60	0.52	0.55	0.79
	0.80	0.86	0.89	0.84	0.79	0.83
τ_{GLS}^μ ($0 \leq p \leq 8$)	1.00	0.07	0.08	0.06	0.19	0.63
	0.95	0.32	0.35	0.27	0.61	0.93
	0.90	0.57	0.66	0.66	0.84	0.98
	0.80	0.83	0.90	0.98	0.96	1.00

Nota: El número de replicaciones es \overline{CS}_1^μ (5000 rep.), $\hat{\tau}_{SD}^\mu$ (2000 rep.), $\hat{\tau}_{SD-S}^\mu$ (10000 rep.) y $\hat{\tau}_{SD-AN}^\mu$ (1000). Condiciones iniciales: $y_0 = 0$ excepto en $\hat{\tau}_{SD-S}^\mu$ (ver texto). El tamaño muestral en $\hat{\tau}_{SD}^\mu$ es $T = 99$. nd: no disponible.

Tabla 4.11: Potencia empírica para el nivel 5% del contraste \overline{CS}_1^τ , con $\hat{\omega}_{arma}^2$, y otros contrastes. $T = 100$.

Contraste	ρ	θ				
		-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^τ	1.00	0.06	0.05	0.05	0.05	0.10
	0.95	0.13	0.13	0.13	0.10	0.20
	0.90	0.30	0.31	0.30	0.18	0.34
	0.80	0.78	0.74	0.67	0.33	0.35
$\hat{\tau}_{SD-S}^\tau$	1.00	0.05	0.03	0.02	0.01	0.05
τ_{GLS}^τ ($3 \leq p \leq 8$)	1.00	0.09	0.06	0.05	0.08	0.51
	0.95	0.17	0.12	0.10	0.15	0.69
	0.90	0.32	0.25	0.22	0.33	0.82
	0.80	0.60	0.57	0.54	0.68	0.91
τ_{GLS}^τ ($0 \leq p \leq 8$)	1.00	0.09	0.09	0.06	0.27	0.84
	0.95	0.16	0.17	0.11	0.47	0.94
	0.90	0.31	0.37	0.29	0.72	0.98
	0.80	0.57	0.71	0.81	0.95	1.00

Nota: El número de replicaciones es \overline{CS}_1^τ (1000 rep.), $\hat{\tau}_{SD-S}^\tau$ (10000 rep.) y τ_{GLS}^τ (5000).
Condiciones iniciales: $y_0 = 0$ excepto en $\hat{\tau}_{SD-S}^\tau$.

Se ha incluido también en las tablas 4.10 y 4.11 los resultados del contraste $\hat{\tau}_{GLS}$, donde se ha utilizado el estimador $\hat{\omega}_{AR}$ con $(3 \leq p \leq 8)$ y $(0 \leq p \leq 8)$.

Las conclusiones que se obtienen de las tablas 4.10 y 4.11 son:

- Con carácter general se observa que los contrastes que no permiten la modelización del componente de medias móviles tienen distorsiones en su nivel de significación muy altas. Para $\theta = 0.8$ la distorsión es excesiva. Sin embargo, los contrastes que permiten la modelización de dicho componente tienen una distorsión muy baja, por lo que son más fiables.
- El contraste \overline{CS}_1^μ es superior al contraste $\hat{\tau}_{SD}^\mu$ salvo para $\rho = 0.8$ y $\theta = 0.5, 0.8$. En el caso $\rho = \theta = 0.8$ el proceso simulado es ruido blanco, por lo que la varianza de las estimaciones del modelo ARMA(1,1) será muy elevada. Por esta razón las estimaciones de $\hat{\omega}_{arma}^2$ serán, en este caso, poco eficientes. No obstante, en una situación real, serán muy escasas las situaciones en las que un analista decida ajustar un modelo ARMA(1,1) a un ruido blanco. Puede considerarse, por tanto, que el contraste \overline{CS}_1^μ es más recomendable que $\hat{\tau}_{SD}^\mu$.
- La comparación de los dos criterios de selección de p para el contraste $\hat{\tau}_{GLS}$ muestran que el criterio $(3 \leq p \leq 8)$ proporciona menos distorsiones del nivel de significación. No obstante, las distorsiones en este caso son también muy elevadas. Esta menor distorsión en valores de $\theta > 0$ produce, sin embargo, una menor potencia.
- La comparación de $\hat{\tau}_{GLS}$, $(3 \leq p \leq 8)$, con \overline{CS}_1 lleva a la conclusión de que, \overline{CS}_1 supera en potencia a $\hat{\tau}_{GLS}$ salvo en las situaciones en que $\hat{\tau}_{GLS}$ presenta distorsiones en su nivel de significación.

- El caso $\theta = 0.5$ presenta un aspecto que merece destacarse. En este caso $\hat{\tau}_{GLS}$ presenta un pequeño aumento en su nivel de significación, lo que provoca un aumento de su potencia. Por el contrario, en esta situación, \overline{CS}_1 disminuye ligeramente su nivel de significación (también lo hace $\hat{\tau}_{SD-S}$), disminuyendo también su potencia respecto a $\theta = 0$. Sería difícil, por tanto, saber cuál de los dos contrastes es preferible en esta situación (esta disminución de la potencia en $\theta = 0.5$ respecto a $\theta = 0$ puede también apreciarse en $\hat{\tau}_{GLS}^\mu$ con $\rho = 0.80$ aunque de forma menos acusada). Puesto que para valores inferiores de θ $\hat{\tau}_{GLS}$ tiene menos potencia y para valores superiores tiene una distorsión de su nivel de significación muy elevada puede aconsejarse el uso del contraste \overline{CS}_1 con independencia de θ .

De este ejercicio de simulación puede concluirse que el contraste \overline{CS}_1 , utilizando el estimador $\hat{\omega}_{arma}^2$, es el más recomendable por su combinación de potencia y escasa distorsión del nivel de significación. Con los contrastes que se han comparado, sólo puede superarse en potencia a \overline{CS}_1 si se está dispuesto a asumir aumentos del nivel de significación muy por encima del nominal. Estas distorsiones son especialmente importantes para valores de $\theta > 0$, que es una situación muy frecuente en series reales.

Resultados de la simulación 2: estimador $\hat{\omega}_{AR}^2$

En este ejercicio de simulación se compara el comportamiento de los contrastes propuestos (\overline{CS}_1 , \overline{CS}_{1T} , \overline{CS}_A) y los contrastes de Elliot-Rothenberg-Stock ($\hat{\tau}_{GLS}$) y Dickey-Fuller ($\hat{\tau}$). Los contrastes se realizan mediante la modelización de un autorregresivo de orden p . Para evaluar la influencia de sobreparametrizar el polinomio autorregresivo se han analizado dos casos. En el primero, el orden de p se selecciona con el criterio BIC con $0 \leq p \leq 8$.

En el segundo caso, el rango sobre el que se aplica el criterio BIC es $3 \leq p \leq 8$. Puesto que este ejercicio de simulación tiene un menor coste computacional se ha realizado para los tamaños muestrales $T = 25, 50, 100$. Los valores de ρ han sido 1.00, 0.97, 0.95, 0.90, 0.80, 0.70. Para el componente de medias móviles se ha empleado $\theta = -0.8, -0.5, 0, 0.5, 0.8$.

Los contrastes \overline{CS} utilizan el estimador $\hat{\omega}_{AR}^2$. Los contrastes τ_{GLS} y $\hat{\tau}$ se basan en estadísticos pivote por lo que no necesitan de ninguna corrección adicional (si el orden p seleccionado es adecuado). Los valores críticos empleados para el contraste τ_{GLS} son los obtenidos mediante simulación en la tabla 3.3.

Las tablas 2.15 a 2.20 muestran los resultados del ejercicio de simulación donde el criterio de selección del orden autorregresivo es $BIC(p)$, con $0 \leq p \leq 8$. Las tablas 2.21 a 2.26 muestran los resultados para $3 \leq p \leq 8$.

El principal resultado de este segundo ejercicio de simulación es que todos los contrastes presentan distorsiones muy elevadas en su nivel de significación para $\theta > 0$. Se obtiene, entonces, que si el proceso tiene componente de medias móviles y el modelo empleado no admite dicho componente el nivel de significación del contraste puede ser muy superior al nominal. Sin embargo, las tablas 4.10 y 4.11 mostraban que si el proceso no tiene componente de medias móviles (columna $\theta = 0$) y el modelo estima dicho componente, no se producen efectos negativos ni en el nivel de significación ni en la potencia. La comparación de las tablas 4.10 y 4.11 con las que se obtuvieron para el caso AR(1) muestran que sólo se produce pérdida de potencia en $\rho = 0.8$. Puede, por tanto, concluirse que la realización de contrastes de raíces unitarias mediante la aproximación de un autorregresivo debe evitarse en la práctica.

Otras conclusiones que se obtienen de este ejercicio de simulación son:

- Para $T = 25$ todos los contrastes tienen una escasa capacidad de detección.
- Al igual que en el caso AR(1), los contrastes \overline{CS}_1^μ y \overline{CS}_{1T}^μ son similares. En general, \overline{CS}_1^μ tiene menos distorsiones en el nivel que \overline{CS}_{1T}^μ . Estos dos contrastes son superiores a \overline{CS}_A^μ salvo en alternativas muy alejadas del círculo unidad.
- De la comparación de los diversos contrastes se obtiene que el contraste $\hat{\tau}_{GLS}$ sobreparametrizado ($3 \leq p \leq 8$) es el que proporciona mejores resultados.
- Esta sobreparametrización también mejora los resultados de $\hat{\tau}$. Para \overline{CS}_1^μ la sobreparametrización disminuye la distorsión en el nivel para $\theta > 0$ pero la aumenta en $\theta = 0$.
- Si no se sobreparametriza, el contraste \overline{CS}_1^μ es el que presenta menos distorsiones en el nivel para $\theta > 0$. Para $\theta < 0$ presentan, en general, más distorsión en el nivel que sus competidores $\hat{\tau}$ y $\hat{\tau}_{GLS}$, salvo en $T = 25$.

4.6 Contrastes con predicciones eficientes. Contrastes ECS

En esta sección se proponen alternativas a los contrastes \overline{CS} mediante el uso de predictores eficientes. Para no introducir complejidad en la notación se plantea primeramente, sin pérdida de generalidad, el caso estocástico puro. Posteriormente se amplía la discusión al caso de tendencia lineal determinista.

El procedimiento consiste en la incorporación de la estructura dinámica de u_t en los contrastes. Se supondrá que u_t tiene una representación ARMA(p, q) estacionaria e invertible,

$$\phi(B)u_t = \theta(B)a_t, \quad (4.6.1)$$

donde $\psi(B) = \phi(B)^{-1}\theta(B)$. Se denotará por $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$ y $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)'$ a los estimadores de $\phi = (\phi_1, \dots, \phi_p)'$ y $\theta = (\theta_1, \dots, \theta_q)'$, obtenidos en la muestra.

Al igual que en el caso AR(1) se construirá un predictor bajo la hipótesis de raíz unitaria. Esto implica la modelización de la serie en diferencias. A partir de este modelo se generan las predicciones mediante la suma de las predicciones de los incrementos de la serie. El valor y_t puede escribirse como

$$y_t = y_1 + \sum_{j=2}^t \Delta y_j,$$

con $\Delta y_{j+1} = y_{j+1} - y_j = u_{j+1}$. El valor previsto desde y_1 es, suponiendo conocidos los parámetros,

$$\tilde{y}_{t|1} = y_1 + \sum_{j=2}^t \Delta \tilde{y}_{j|1}$$

con $\Delta \tilde{y}_{j+1|1} = E(u_t | u_1)$. Al ser los parámetros ϕ y θ desconocidos habrá que emplear sus estimaciones. Esto provoca que la primera observación disponible sobre la que realizar predicciones sea y_{p+1} , al necesitarse las p primeras observaciones para generar la predicciones. La previsión de y_t será,

$$\hat{y}_{t|p+1} = y_{p+1} + \sum_{j=p+2}^t \Delta \hat{y}_{j|p+1},$$

donde $\Delta \hat{y}_{j|p+1} = \hat{u}_{j|p+1} = E(u_j | y_{p+1}, \hat{\phi}, \hat{\theta})$. El error cuadrático de predicción acumulado, empírico, será

$$\sum_{t=p+2}^T e_{t|p+1}^2 = \sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2 \quad (4.6.2)$$

Para la obtención del ECMP teórico bajo la hipótesis nula de raíz unitaria se utilizará que, al ser el proceso u_t estacionario,

$$u_{j+1} - \tilde{u}_{j+1|p+1} = \Delta y_{j+1} - \Delta \tilde{y}_{j+1|1} = a_{j+1} + \psi_1 a_{j-1} + \cdots + \psi_j a_2.$$

Operando recursivamente se obtiene que

$$\begin{aligned} y_t - \tilde{y}_{t|p+1} &= \sum_{j=p+2}^t (\Delta y_{j|p+1} - \Delta \tilde{y}_{j|p+1}) \\ &= a_t + (1 + \psi_1) a_{t-1} + \cdots + (1 + \psi_1 + \cdots + \psi_{t-2}) a_2. \end{aligned}$$

Por lo tanto, el error cuadrático medio de predicción es, bajo la hipótesis de raíz unitaria,

$$E(y_t - \tilde{y}_{t|p+1})^2 = \sigma^2 \sum_{j=0}^{t-p-2} \left(\sum_{i=0}^j \psi_i \right)^2. \quad (4.6.3)$$

Al ser u_t estacionario los coeficientes ψ_j decaerán exponencialmente al aumentar j . Según esto, a medida que aumente el valor de t en (4.6.3) el término $\sum_{i=0}^{t-p-2} \psi_i$ se va aproximando a un valor constante y el ECMP se aproximará a la recta $\omega^2(t-p-2)$, donde $\omega^2 = \sigma^2 \psi^2(1)$. Por tanto los ECMP acumulados en el caso $u_t = \psi(B)a_t$ y $\rho = 1$ son, aproximadamente,

$$\sum_{t=p+2}^T E(y_t - \tilde{y}_{t|p+1})^2 \approx \omega^2(1 + 2 + \cdots + (T-p-1)) = \omega^2 \frac{(T-p)(T-p-1)}{2}. \quad (4.6.4)$$

El parametro ω^2 es desconocido por lo que se necesitará un estimador consistente del mismo, $\hat{\omega}^2$. El ratio entre los errores cuadráticos acumulados empíricos (4.6.2) y los esperados (4.6.4) sugieren el estadístico ECS_1^μ ,

$$ECS_1^\mu = \frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2}{\hat{\omega}^2 (T-p-1)^2}, \quad (4.6.5)$$

que es semejante al estadístico \overline{CS}_1^μ , salvo que se reemplaza y_1 por la predicción $\hat{y}_{t|p+1}$. Asimismo, se divide por $(T-p-1)^2$ en lugar de $(T-1)^2$ al tener el numerador $(T-p-1)$ términos.

En el caso en que se quiera contrastar la existencia de tendencia determinista se tiene que $\Delta y_t = \delta + u_t$. El parámetro δ es desconocido pero su estimador $\hat{\delta}$ puede obtenerse conjuntamente con $\hat{\phi}$ y $\hat{\theta}$ (o bien utilizar el estimador (4.3.8), que es asintóticamente equivalente). Si se denomina $\hat{y}_{t|p+1}^\tau$ a las predicciones realizadas utilizando incrementos estimados de esta forma se obtiene el estadístico ECS_1^τ ,

$$ECS_1^\tau = \frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1}^\tau)^2}{\hat{\omega}^2 (T - p - 1)^2}. \quad (4.6.6)$$

Análogamente se puede construir el estadístico que promedie los errores de predecir cada observación desde la primera disponible y los errores de predecir la observación más alejada desde cada uno de los datos.

$$ECS_{1T}^\mu = \frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2 + \sum_{t=p+1}^{T-1} (y_T - \hat{y}_{T|t})^2}{2\hat{\omega}^2 (T - p - 1)^2}. \quad (4.6.7)$$

Igualmente, si se promedian los errores cuadráticos que, para cada horizonte, puedan extraerse de la muestra obtendremos los estadísticos ECS_A , que serán los análogos de \overline{CS}_A ,

$$ECS_A^\mu = \frac{\sum_{t=p+1}^{T-1} \sum_{s=t+1}^T (T - (s - t))^{-1} (y_s - \hat{y}_{s|t})^2}{\hat{\omega}^2 (T - p - 1)^2} \quad (4.6.8)$$

$$ECS_A^\tau = \frac{\sum_{t=p+1}^{T-1} \sum_{s=t+1}^T (T - (s - t))^{-1} (y_s - \hat{y}_{s|t}^\tau)^2}{\hat{\omega}^2 (T - p - 1)^2} \quad (4.6.9)$$

De esta forma cada contraste \overline{CS} tiene su análogo ECS sustituyendo y_1 por la predicción que se obtiene del modelo considerado y dividiendo por el cuadrado del número de horizontes de predicción considerados.

Como estimador de ω^2 puede utilizarse cualquiera de los comentados anteriormente. No obstante, debido al ejercicio de modelización realizado, la decisión más natural será elegir ω_{arma}^2 . De esta forma, el procedimiento para calcular los estadísticos de los contrastes será el siguiente. Se estima el mejor modelo que explique la serie diferenciada y se evalúan

con él las predicciones. Posteriormente se reestima dicho modelo sin imponer la noestacionariedad y se calcula $\hat{\omega}_{arma}^2$.

Esta forma de construir el contraste, utilizando el mejor predictor que el analista considere, permite interpretar los resultados en términos de predicción. En el caso en que no se rechaze la hipótesis de raíz unitaria puede interpretarse que, sea o no cierta esta hipótesis, la incertidumbre que se posee del futuro, con el predictor empleado, es la misma que si hubiese realmente tal raíz. En el caso en que se rechaze la hipótesis de raíz unitaria puede interpretarse que, aunque exista realmente, sus efectos aún no son perceptibles. Se ha de recordar que los efectos de una raíz unitaria están relacionados con las predicciones de la serie y ésta es precisamente la información empleada en el contraste.

4.6.1 Distribuciones asintóticas de los estadísticos ECS

Para los estadísticos ECS el siguiente teorema demuestra que las distribuciones asintóticas de los estadísticos \overline{CS} siguen siendo válidas.

Teorema 4.2 *Sea y_t el proceso (3.2.2) con $\rho = 1$. Sea u_t un proceso estacionario e invertible con representación ARMA(p, q) (4.6.1). Sean $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$ y $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)'$ estimadores de $\phi = (\phi_1, \dots, \phi_p)'$ y $\theta = (\theta_1, \dots, \theta_q)'$, respectivamente, tales que*

$$\hat{\phi} - \phi = O_p(T^{-1/2})$$

y

$$\hat{\theta} - \theta = O_p(T^{-1/2})$$

entonces

$$ECS \xrightarrow{d} dCS \tag{4.6.10}$$

Demostración:

Estadístico ECS₁^μ:

Como $\hat{\theta} - \theta = O_p(T^{-1/2})$ y $\hat{\phi} - \phi = O_p(T^{-1/2})$ se tiene que

$$\hat{y}_{t|p+1} - \tilde{y}_{t|p+1} = O_p(T^{-1/2}) \quad (4.6.11)$$

(véase, por ejemplo, Fuller 1976, p. 384) y por tanto

$$\frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2}{(T-p-1)^2} = \frac{\sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})^2}{(T-p-1)^2} + \frac{\sum_{t=p+2}^T (\tilde{y}_{t|p+1} - \hat{y}_{t|p+1})^2}{(T-p-1)^2} \quad (4.6.12)$$

$$\frac{\sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})(\tilde{y}_{t|p+1} - \hat{y}_{t|p+1})}{(T-p-1)^2} \quad (4.6.13)$$

Si en el término (4.6.12) se aplica la propiedad (4.6.11) se obtiene

$$\frac{\sum_{t=p+2}^T (\tilde{y}_{t|p+1} - \hat{y}_{t|p+1})^2}{(T-p-1)^2} \xrightarrow{p} 0.$$

Para ver que el término (4.6.13) también converge en probabilidad a cero se plantea por simplicidad, y sin pérdida de generalidad, el caso en que u_t siga un AR(1), es decir $u_t = \phi u_{t-1} + a_t$. Entonces

$$\begin{aligned} y_t - \tilde{y}_{t|2} &= \sum_{j=3}^t u_j - \tilde{u}_{j|3} = \sum_{j=3}^t \left(\sum_{i=3}^j \phi^{j-i} a_i \right) \\ &= \frac{1}{1-\phi^2} \sum_{j=3}^t a_j + \sum_{j=0}^t \alpha_j a_{t-j} \end{aligned}$$

donde $\alpha_j = \sum_{i=1}^{\infty} \phi^{j+i}$ y además $\sum_{j=0}^{\infty} |\alpha_j| < \infty$. Por otra parte

$$\tilde{y}_{t|3} - \hat{y}_{t|3} = \sum_{j=3}^t (\phi^{j-2} - \hat{\phi}^{j-2}) u_1.$$

donde tanto $\sum_{j=3}^t \phi^{j-2}$ como $\sum_{j=3}^t \hat{\phi}^{j-2}$ están acotadas por una función de decrecimiento exponencial con t y por tanto, si se denomina γ_t a la diferencia entre ambos términos,

$$\gamma_t = \sum_{j=3}^t (\phi^{j-2} - \hat{\phi}^{j-2}) = O_p(T^{-1/2}).$$

De esta forma (4.6.13) es el promedio de sumandos que verifican

$$T^{-1} \left(\frac{\gamma_t}{1 - \phi^2} \sum_{j=3}^t a_j u_1 + \gamma_t \sum_{j=0}^t \alpha_j a_{t-j} u_1 \right) \xrightarrow{p} 0$$

y por consiguiente

$$\frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})(\tilde{y}_{t|p+1} - \hat{y}_{t|p+1})}{(T - p - 1)^2} \xrightarrow{p} 0.$$

Se verifica entonces que

$$\frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2}{(T - p - 1)^2} - \frac{\sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})^2}{(T - p - 1)^2} \xrightarrow{p} 0. \quad (4.6.14)$$

Se puede, pues, estudiar la distribución asintótica del estadístico considerando que los parámetros ϕ y θ son conocidos.

Los términos del numerador de este estadístico admiten la siguiente representación:

$$\begin{aligned} y_t - \tilde{y}_{t|p+1} &= y_t - y_{p+1} - \sum_{j=p+2}^t \Delta \tilde{y}_{j|p+1} \\ &= \sum_{j=p+2}^t u_j - \tilde{u}_{j|p+1} = \sum_{j=p+2}^t \left(\sum_{i=p+2}^j \psi_{j-i} a_i \right). \end{aligned} \quad (4.6.15)$$

Si se analiza cada sumando se observa que

$$\begin{aligned} y_t - \tilde{y}_{t|p+1} &= a_{p+2} + \\ &\quad (a_{p+3} + \psi_1 a_{p+2}) + \\ &\quad (a_{p+4} + \psi_1 a_{p+3} + \psi_2 a_{p+2}) + \cdots + \\ &\quad (a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + \psi_{t-p-2} a_{p+2}) \end{aligned} \quad (4.6.16)$$

$$= a_{p+2}\psi(1) - a_{p+2} \sum_{j=1}^{\infty} \psi_{t-p-1+j} + \quad (4.6.17)$$

$$a_{p+3}\psi(1) - a_{p+3} \sum_{j=1}^{\infty} \psi_{t-p-2+j} + \cdots +$$

$$a_t\psi(1) - a_t \sum_{j=1}^{\infty} \psi_j \quad (4.6.18)$$

$$= \psi(1)(a_{p+2} + a_{p+3} + \cdots + a_t) + \nu_t,$$

donde

$$\nu_t = \sum_{j=0}^{t-p-2} \alpha_j a_{t-j}, \quad \alpha_j = \sum_{i=1}^{\infty} \psi_{j+i},$$

cumpléndose que $\sum_{j=0}^{t-p-2} |\alpha_j| < \infty$. Se puede escribir, entonces,

$$e_{t|p+1} = \psi(1) \sum_{j=1}^t a_j - \psi(1) \sum_{j=1}^{p+1} a_j + \nu_t.$$

Se define el proceso de sumas parciales $X_T(r)$, con $r \in [0, 1]$

$$X_T(r) = \frac{e_{[Tr]|p+1}}{T} = \psi(1) \frac{\sum_{j=1}^{[Tr]} a_j}{T} - \psi(1) \frac{\sum_{j=1}^{p+1} a_j}{T} + \frac{\nu_{[Tr]}}{T}.$$

Si se multiplica \sqrt{T} se obtiene, aplicando el TCLF,

$$\psi(1) \frac{\sum_{j=1}^{[Tr]} a_j}{\sqrt{T}} \xrightarrow{d} \sigma\psi(1)W(r)$$

Además, por ser p fijo respecto a T se obtiene que

$$\psi(1) \frac{\sum_{j=1}^{p+1} a_j}{\sqrt{T}} \xrightarrow{p} 0.$$

Como ν_t es estacionario con varianza finita se tiene que

$$\frac{\nu_{[Tr]}}{\sqrt{T}} \xrightarrow{p} 0, \quad (4.6.19)$$

(véase, por ejemplo, Hamilton 1994, p. 482). Por tanto, aplicando el teorema de la aplicación continua se obtiene

$$\frac{\sum_{t=p+2}^T e_{t|p+1}^2}{T^2} \xrightarrow{d} \sigma^2\psi(1)^2 \int_0^1 W^2(r) dr$$

como el denominador del estadístico ECS es un estimador consistente de $\sigma^2\psi(1)^2$ se concluye el resultado.

Estadístico ECS_1^T :

Si se denomina $\hat{\delta}$ al estimador de δ

$$\hat{\delta} = \frac{\sum_{t=2}^T (y_t - y_{t-1})}{T-1} = \frac{y_T - y_1}{T-1}$$

y $\hat{\delta}^*$ al que se obtiene de estimar el modelo

$$\phi(B)(\Delta y_t - \delta) = \theta(B)a_t,$$

y si se utiliza $\tilde{y}_{t|p+1}$ para denotar la previsión de y_t desde y_{p+1} cuando los parámetros ϕ y θ son conocidos y se utiliza el estimador $\hat{\delta}$ se obtiene, como se demostró anteriormente, que

$$\frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2}{(T-p-1)^2} - \frac{\sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})^2}{(T-p-1)^2} \xrightarrow{p} 0. \quad (4.6.20)$$

Se puede, así, estudiar la distribución asintótica suponiendo que los parámetros que definen la estructura dinámica de u_t son conocidos y que el estimador de δ es el utilizado en los estadísticos \overline{CS}_1^T . Por tanto

$$\begin{aligned} y_t - \tilde{y}_{t|p+1} &= y_t - y_{p+1} - \sum_{j=p+2}^t (\hat{\delta} + \tilde{u}_{j|p+1}). \\ &= y_t - y_{p+1} - \hat{\delta}(t-p-1) \sum_{j=p+2}^t \tilde{u}_{j|p+1} \\ &= \sum_{j=p+2}^t u_j - \hat{\delta}(t-p-1) \\ &= \psi(1)(a_{p+2} + a_{p+3} + \dots + a_t) + \nu_t - \hat{\delta}(t-p-1). \end{aligned} \quad (4.6.21)$$

Se verifica, entonces, que

$$e_{t|p+1} = \psi(1) \sum_{j=1}^t a_j - \psi(1) \sum_{j=1}^{p+1} a_j + \nu_t - \hat{\delta}(t-p-1)$$

Siguiendo los mismos argumentos que en el caso anterior puede demostrarse que la influencia de ν_t y $\sum_{j=1}^{p+1} a_j$ no influye en la distribución asintótica del estadístico, siendo los restantes términos iguales al caso del estadístico CS_T^r . Puesto que el denominador del estadístico es un estimador consistente de $\sigma^2\psi^2(1)$ se cumple el resultado.

Estadístico ECS_{1T}^μ :

En este caso es suficiente con demostrar que

$$\frac{(T-p-1)^{-2} \sum_{t=p+1}^{T-1} (y_T - \hat{y}_{T|t}^\mu)^2}{(T-p-1)^{-1} \sum_{t=p+2}^T (y_t - \hat{y}_{t|t-1}^\mu)^2}$$

tiene la misma distribución asintótica que \overline{CS}_T^μ . Al igual que en los anteriores contrastes puede suponerse que los parámetros ϕ y θ son conocidos al ser sus estimadores \sqrt{T} -consistentes. Se tiene entonces

$$\begin{aligned} y_T - \tilde{y}_{T|t}^\mu &= y_T - y_t - \sum_{j=t+1}^T \Delta \tilde{y}_{j|t} \\ &= \sum_{j=t+1}^T u_j - \tilde{u}_{j|t} = \sum_{j=t+1}^T \left(\sum_{i=t+1}^j \psi_{j-i} a_i \right) \\ &= a_{t+1} \psi(1) - a_{t+1} (\psi_{T-t} + \psi_{T-t+1} + \dots) + \\ &\quad a_{t+2} \psi(1) - a_{t+2} (\psi_{T-t-1} + \psi_{T-t} + \dots) + \dots \\ &= \psi(1) \sum_{j=t+1}^T a_j + \eta_t \end{aligned}$$

donde $\eta_t = \sum_{j=1}^{T-t} \beta_j a_{t+j}$, con $\beta_j = \sum_{i=1}^{\infty} \psi_{T-t+1-j}$ y $\sum_{j=1}^{T-t} |\beta_j| < \infty$. Los errores de predicción son entonces

$$e_{t|p+1} = \psi(1) \sum_{j=t+1}^T a_j + \eta_t.$$

Usando los mismos argumentos que anteriormente se demuestra que η_t no tiene influencia en la distribución asintótica del estadístico, siendo los restantes términos análogos al caso

del estadístico CS_T^μ . De nuevo, al ser el denominador un estimador consistente de $\sigma^2\psi^2(1)$ se verifica el resultado.

Estadísticos ECS_A^μ y ECS_A^τ :

La demostración para estos dos estadísticos es inmediata con los resultados obtenidos de los anteriores estadísticos. El término $(y_s - \hat{y}_{s|t})$ puede descomponerse de manera análoga a (4.6.15) con t en lugar de $p + 1$ y por tanto demostrar que el incorporar la información sobre las predicciones de u_t no afecta a su distribución asintótica. De forma similar el término $(y_s - \hat{y}_{s|t}^\tau)$ se resuelve utilizando los mismos argumentos que en (4.6.21). \square

4.6.2 Potencia en muestras finitas de ECS

En este apartado se realiza un estudio de la potencia de los contrastes ECS mediante un ejercicio de simulación por Monte Carlo. El proceso que se ha empleado en las simulaciones es de nuevo un ARMA(1,1), con innovaciones normales, en el que se contrasta si el parámetro autorregresivo es uno.

La simulación de este tipo de contrastes requeriría de un analista que modelizase cada serie simulada con el fin de encontrar el predictor que sea más adecuado. Este procedimiento no es factible si se desea un elevado número de replicaciones. Como alternativa se han planteado dos experimentos. En el primero de ellos se analiza el comportamiento del contraste en una situación en la que el modelo se especifica correctamente. En el segundo experimento se sitúa al contraste en una situación desfavorable, construyéndose el predictor mediante una aproximación autorregresiva.

Para el primer ejercicio de simulación se supone que existe una identificación correcta del modelo ARMA(1,1). Para el contraste ECS^μ se ha estimado un modelo IMA(1,1) para generar las predicciones y se ha obtenido $\hat{\omega}_{arma}^2$ estimando un ARMA(1,1) de media no nula. Para el contraste ECS^τ se ha estimado un IMA(1,1) con constante para generar las predicciones y un ARMA(1,1) con media no nula y tendencia lineal para estimar ω^2 . Estas estimaciones se han realizado utilizando máxima verosimilitud condicionada (mínimos cuadrados no lineales) salvo en los modelos IMA(1,1) sin constante, donde se ha empleado máxima verosimilitud exacta. Para la estimación del parámetro θ en IMA(1,1) por máxima verosimilitud se ha utilizado la restricción de que la búsqueda del valor estimado se realice entre $-1 < \theta < 1$. Para el resto de los casos los valores iniciales se han obtenido de manera similar a la sección 4.5.2.

En el segundo ejercicio de simulación se han realizado las predicciones mediante la aproximación a un autorregresivo, tal y como se hizo en la sección 4.5.2. El criterio para seleccionar el orden ha sido de nuevo BIC, restringiendo los valores en $0 \leq p \leq 8$. Asimismo el estimador de ω^2 ha sido $\hat{\omega}_{AR}^2$. De esta forma la regresión en forma de corrección del error proporciona un estimador consistente de ω y esa misma regresión, pero estimada sin el regresor y_{t-1} , genera las predicciones. La diferencia respecto a las simulaciones realizadas para los contrastes \overline{CS} es en la utilización de estas predicciones. Mediante este ejercicio se pretende estudiar el comportamiento del contraste en una situación muy desfavorable. La construcción de un predictor mediante una aproximación autorregresiva produce, en general, predictores poco eficientes si el verdadero proceso tiene componente de medias móviles (Ledolter & Abraham 1981).

Resultados de la simulación 1: predictor ARMA(1,1)

Este ejercicio de simulación ha sido realizado, debido a su elevado coste computacional, para $T = 100$ y contraste ECS_1^μ . Los resultados de nivel y potencia empíricos se encuentran en las tablas 4.12 y 4.13. En estas tablas se ha incluido, para facilitar la comparación, los resultados de \overline{CS}_1^μ . También se ha añadido una columna correspondiente a los resultados del caso AR(1) (tablas 4.8 y 4.4).

Tabla 4.12: Potencia empírica para el nivel 5% del contraste ECS_1^μ y \overline{CS}_1^μ con $\hat{\omega}_{arma}^2$.

$T = 100$.

Contraste	ρ	AR(1)	θ				
			-0.8	-0.5	0.0	0.5	0.8
ECS_1^μ	1.00	0.05	0.06	0.05	0.05	0.04	0.10
	0.95	0.28	0.29	0.29	0.28	0.20	0.32
	0.90	0.63	0.65	0.63	0.61	0.40	0.55
	0.80	0.92	0.94	0.93	0.92	0.64	0.74
\overline{CS}_1^μ	1.00	0.05	0.06	0.04	0.05	0.05	0.11
	0.95	0.28	0.31	0.31	0.27	0.17	0.31
	0.90	0.63	0.71	0.70	0.63	0.31	0.54
	0.80	0.92	0.99	0.97	0.90	0.48	0.67

Nota: Los contrastes se basan en 5000 replicaciones por Monte Carlo. Condiciones iniciales $y_0 = 0$.

Los contrastes ECS_1^μ y \overline{CS}_1^μ son muy similares en niveles de significación. También son similares en potencia para $\theta = 0$. Existen diferencias en valores alejados del círculo

unidad. El contraste ECS_1^μ presenta, en estos casos, valores de potencia más parecidos a los obtenidos en el caso AR(1), por lo que es más robusto a la presencia de componentes de medias móviles.

Los contrastes ECS_1^T y \overline{CS}_1^T son bastante similares. Sus principales diferencias están en $\rho = 0.8$. En esta zona las potencias de ECS_1^T son más cercanas a las obtenidas para el caso AR(1). Es de destacar, sin embargo, la baja potencia, comparativamente al caso AR(1), de ambos contrastes en los casos $\rho = 0.8$ con $\theta = 0.5$ y 0.8 .

Puede concluirse, por tanto, que tanto ECS_1 como \overline{CS}^μ presentan excelentes propiedades tanto de potencia como en su nivel de significación. La principal diferencia entre ambos es que la incorporación de un predictor eficiente en ECS_1^μ aumenta la robustez frente a componentes de medias móviles, por lo que los hace más recomendables.

Resultados de la simulación 2: aproximación AR(p)

Las tablas 2.27 a 2.32 contienen los resultados de las simulaciones realizadas mediante el ajuste de un AR(p), donde p se selecciona con el criterio BIC sujeto a la restricción $0 \leq p \leq 8$.

De la comparación de los contrastes ECS^μ con \overline{CS}^μ se observa que, para tamaños muestrales pequeños, los nuevos contrastes ECS^μ disminuyen ligeramente la distorsión del nivel de significación pero son menos potentes. Sin embargo estas diferencias disminuyen a medida que aumenta el tamaño muestral, siendo para $T = 100$ casi idénticos.

En el caso de tendencia determinista se observa que en los contrastes ECS^T hay situaciones

Tabla 4.13: Potencia empírica para el nivel 5% de los contrastes ECS_1^T y \overline{CS}_1^T con $\hat{\omega}_{arma}^2$.
 $T = 100$.

Contraste	ρ	AR(1)	θ				
			-0.8	-0.5	0.0	0.5	0.8
ECS_1^T	1.00	0.05	0.04	0.05	0.05	0.06	0.12
	0.95	0.12	0.10	0.13	0.13	0.12	0.25
	0.90	0.30	0.23	0.29	0.29	0.22	0.34
	0.80	0.72	0.69	0.68	0.69	0.34	0.40
\overline{CS}_1^T	1.00	0.05	0.06	0.05	0.05	0.05	0.10
	0.95	0.12	0.13	0.13	0.13	0.10	0.20
	0.90	0.30	0.30	0.31	0.30	0.18	0.34
	0.80	0.72	0.78	0.74	0.67	0.33	0.35

Nota: Los contrastes se basan en 1000 replicaciones por Monte Carlo. Condiciones iniciales $y_0 = 0$.

en las que el contraste disminuye su potencia al disminuir ρ . Para ECS_1^T este efecto sólo ocurre en $T = 25$ y $\theta > 0$ pero en ECS_A^T esta situación se repite en más ocasiones. Este efecto se detectó previamente en la tabla 2.18 en algunas zonas de los contrastes \overline{CS}^T y τ_{GLS}^T aunque de forma muy moderada. El contraste ECS_1^T sólo supera, ligeramente, a \overline{CS}_1^T en los casos $\theta < 0$.

Por tanto la incorporación de las predicciones en el contraste no mejora su comportamiento si el predictor empleado no es eficiente. En estas circunstancias un predictor más parsimonioso puede ofrecer mejores resultados.

Este ejercicio de simulación confirma el resultado obtenido anteriormente sobre la poca conveniencia de contrastar raíces unitarias mediante aproximaciones autorregresivas.

Robustez a las condiciones iniciales

Las distribuciones asintóticas de los contrastes propuestos no dependen del valor de y_0 . Sin embargo, los estudios de Evans & Savin (1984) y Schmidt & Phillips (1992), entre otros, señalan que la potencia en muestras finitas de los contrastes de raíces unitarias dependen de dicha condición inicial. Para investigar esta dependencia se ha realizado un ejercicio de simulación para los contrastes \overline{CS}_1^μ y ECS_1^μ utilizando $\hat{\omega}_{arma}^2$ en ambos y predicciones eficientes en ECS_1^μ . En este ejercicio se generan, para $\rho < 1$, series de tamaño 200 donde se suprimen las primeras 100 observaciones para eliminar el efecto de las observaciones iniciales y permitir que la primera observación proceda de su distribución incondicional. De esta forma la serie es estacionaria. El número de replicaciones es 1000. Los resultados de esta simulación se muestran en la tabla 4.14. En esta tabla se ha incluido el resultado de realizar el mismo ejercicio con el contraste τ_{GLS}^μ . La fila correspondiente a $\rho = 1$

procede de las tablas 4.10 y 4.12.

Tabla 4.14: Potencia empírica para el nivel 5% de los contrastes ECS_1^μ y \overline{CS}_1^T con $\hat{\omega}_{arma}^2$.

$T = 100$.

Contraste	ρ	θ				
		-0.8	-0.5	0.0	0.5	0.8
ECS_1^μ	1.00	0.06	0.05	0.05	0.04	0.10
	0.95	0.12	0.13	0.21	0.20	0.34
	0.90	0.30	0.32	0.42	0.40	0.56
	0.80	0.53	0.61	0.77	0.65	0.74
\overline{CS}_1^μ	1.00	0.06	0.04	0.05	0.05	0.11
	0.95	0.19	0.21	0.19	0.17	0.28
	0.90	0.44	0.45	0.41	0.25	0.51
	0.80	0.78	0.77	0.70	0.48	0.69
τ_{GLS}^μ ($3 \leq p \leq 8$)	1.00	0.07	0.05	0.05	0.08	0.38
	0.95	0.23	0.19	0.18	0.26	0.67
	0.90	0.41	0.39	0.38	0.48	0.78
	0.80	0.65	0.69	0.69	0.75	0.84
τ_{GLS}^μ ($0 \leq p \leq 8$)	1.00	0.07	0.08	0.06	0.19	0.63
	0.95	0.21	0.24	0.20	0.50	0.91
	0.90	0.39	0.47	0.48	0.78	0.98
	0.80	0.62	0.78	0.87	0.96	1.00

Nota: Los contrastes se basan en 1000 replicaciones por Monte Carlo. y_1 procede de su distribución incondicional.

Los resultados obtenidos en la tabla 4.14 muestran una disminución de la potencia, en los

cuatro contrastes considerados, respecto al caso $y_0 = 0$. Esta disminución de la potencia cuando y_1 procede de su distribución incondicional ha sido previamente detectada por otros autores (Elliot, Rothenberg & Stock (1992), Pantula et al (1994)). Los cambios más destacables se encuentran, para los cuatro contrastes, en la zona donde ρ está alejado de la unidad y θ toma valores negativos.

El análisis de esta tabla no cambia, sin embargo, las principales conclusiones obtenidas anteriormente. Para $\theta = 0$ el contraste que proporciona mayores valores de potencia es $\hat{\tau}_{GLS}^\mu$ con $0 \leq p \leq 8$. Este contraste, sin embargo, tiene aumentos excesivos de su nivel de significación. El contraste $\hat{\tau}_{GLS}^\mu$ con $3 \leq p \leq 8$ es superado por \overline{CS}_1^μ salvo en $\theta > 0$, donde $\hat{\tau}_{GLS}^\mu$ aumenta su nivel de significación considerablemente. Por tanto, para obtener mayores potencias que los contrastes propuestos ha de ser a cambio de importantes aumentos en el nivel de significación en $\theta > 0$. La comparación entre ECS_1^μ y \overline{CS}_1^T es similar a la realizada para el caso $y_0 = 0$. ECS_1^μ supera a \overline{CS}_1^T en la zona $\rho \approx \theta$, mientras que \overline{CS}_1^T aventaja a ECS_1^μ en la zona $\rho \approx -\theta$.

4.7 Resumen y conclusiones

En este capítulo se proponen contrastes de raíces unitarias basados en errores de predicción a varios horizontes. Entre ellos, los basados en los errores de predicción desde la primera observación disponible (CS_1) son los que ofrecen mejor comportamiento global. En el caso AR(1) estos contrastes tienen una potencia similar, en la localidad, a los de Elliot-Rothenberg-Stock, que pueden considerarse los más potentes que existen en la literatura. En el caso más general en que el proceso siga un modelo ARIMA el comportamiento del contraste mejora con la utilización de un estimador eficiente de ω^2 y con la utilización de un predictor eficiente.

Los ejercicios de simulación realizados en tamaños muestrales $T = 100$ y procesos ARMA(1,1) muestran que los contrastes CS superan a las alternativas existentes si el modelo se especifica correctamente. Los niveles de significación obtenidos son, en todos los casos, iguales o muy próximos a los nominales. Los demás contrastes con los que se ha comparado a los contrastes CS sólo los superan en potencia en las situaciones en que presentan aumentos considerables en su nivel de significación ante presencia de medias móviles.

Las razones fundamentales por las que los contrastes CS muestran mejores resultados que los demás contrastes analizados son dos.

1. Están contruidos utilizando errores de predicción a varios horizontes. Es en estos errores de predicción donde se encuentra la información sobre el largo plazo de la serie, siendo el largo plazo donde las raíces unitarias tienen mayor efecto. Este aspecto le confiere buenas propiedades de potencia.
2. Permiten la modelización del modelo más apropiado que el analista considere, sin necesidad de utilizar MCO ni aproximaciones autorregresivas. Esto aumenta la eficiencia de las predicciones y de la estimación de ω^2 . Esta característica le proporciona robustez frente a la estructura dinámica de u_t .

Apéndice B: Tablas.

Tabla 2.15: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
CS_1^μ	1.00	0.14	0.13	0.07	0.18	0.35
	0.97	0.18	0.16	0.09	0.22	0.37
	0.95	0.20	0.18	0.11	0.24	0.39
	0.90	0.24	0.23	0.13	0.28	0.40
	0.80	0.30	0.29	0.20	0.33	0.42
	0.70	0.35	0.33	0.26	0.39	0.42
CS_{1T}^μ	1.00	0.14	0.13	0.07	0.17	0.36
	0.97	0.18	0.16	0.10	0.22	0.39
	0.95	0.21	0.18	0.11	0.24	0.40
	0.90	0.24	0.23	0.13	0.29	0.42
	0.80	0.30	0.30	0.20	0.35	0.45
	0.70	0.35	0.34	0.26	0.41	0.45
CS_A^μ	1.00	0.16	0.14	0.07	0.20	0.45
	0.97	0.20	0.19	0.09	0.24	0.49
	0.95	0.23	0.21	0.10	0.26	0.51
	0.90	0.26	0.25	0.12	0.32	0.55
	0.80	0.31	0.30	0.17	0.41	0.58
	0.70	0.34	0.33	0.22	0.50	0.58
τ_μ	1.00	0.32	0.29	0.24	0.41	0.72
	0.97	0.34	0.31	0.26	0.45	0.75
	0.95	0.34	0.32	0.26	0.47	0.77
	0.90	0.36	0.34	0.27	0.55	0.80
	0.80	0.42	0.36	0.31	0.65	0.84
	0.70	0.43	0.40	0.37	0.75	0.86
τ_{GLS}^μ	1.00	0.19	0.17	0.13	0.35	0.65
	0.97	0.22	0.20	0.16	0.41	0.68
	0.95	0.23	0.22	0.18	0.44	0.71
	0.90	0.28	0.25	0.22	0.52	0.74
	0.80	0.35	0.32	0.29	0.64	0.77
	0.70	0.39	0.37	0.37	0.72	0.77

Tabla 2.16: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.18	0.12	0.06	0.24	0.60
	0.97	0.27	0.21	0.11	0.38	0.71
	0.95	0.34	0.27	0.15	0.44	0.73
	0.90	0.50	0.46	0.27	0.58	0.79
	0.80	0.67	0.65	0.53	0.74	0.81
	0.70	0.72	0.75	0.69	0.79	0.83
\overline{CS}_{1T}^μ	1.00	0.17	0.12	0.06	0.24	0.65
	0.97	0.27	0.20	0.11	0.38	0.77
	0.95	0.34	0.27	0.14	0.45	0.81
	0.90	0.48	0.43	0.26	0.60	0.86
	0.80	0.65	0.63	0.51	0.79	0.90
	0.70	0.71	0.74	0.68	0.86	0.91
\overline{CS}_A^μ	1.00	0.20	0.15	0.06	0.29	0.73
	0.97	0.30	0.24	0.11	0.45	0.85
	0.95	0.37	0.30	0.14	0.53	0.89
	0.90	0.49	0.44	0.24	0.67	0.94
	0.80	0.62	0.63	0.51	0.87	0.97
	0.70	0.69	0.73	0.77	0.94	0.98
τ_μ	1.00	0.20	0.11	0.07	0.32	0.78
	0.97	0.28	0.13	0.09	0.45	0.87
	0.95	0.32	0.15	0.10	0.51	0.90
	0.90	0.43	0.21	0.15	0.66	0.95
	0.80	0.58	0.38	0.34	0.87	0.99
	0.70	0.67	0.55	0.64	0.91	1.00
τ_{GLS}^μ	1.00	0.10	0.08	0.06	0.32	0.77
	0.97	0.16	0.15	0.11	0.49	0.89
	0.95	0.21	0.21	0.15	0.57	0.91
	0.90	0.35	0.35	0.27	0.73	0.95
	0.80	0.53	0.59	0.59	0.90	0.98
	0.70	0.65	0.73	0.83	0.96	0.98

Tabla 2.17: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 repeticiones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.11	0.09	0.05	0.15	0.54
	0.97	0.28	0.24	0.15	0.40	0.79
	0.95	0.41	0.38	0.26	0.52	0.87
	0.90	0.66	0.69	0.60	0.76	0.94
	0.80	0.87	0.91	0.89	0.90	0.96
	0.70	0.93	0.96	0.94	0.95	0.97
\overline{CS}_{1T}^μ	1.00	0.11	0.09	0.05	0.17	0.59
	0.97	0.30	0.26	0.16	0.43	0.85
	0.95	0.43	0.39	0.28	0.57	0.91
	0.90	0.64	0.67	0.60	0.79	0.97
	0.80	0.84	0.88	0.90	0.95	1.00
	0.70	0.91	0.95	0.96	0.99	1.00
\overline{CS}_A^μ	1.00	0.13	0.11	0.05	0.19	0.60
	0.97	0.29	0.26	0.14	0.44	0.86
	0.95	0.40	0.37	0.23	0.56	0.92
	0.90	0.60	0.63	0.54	0.82	0.98
	0.80	0.81	0.87	0.96	0.97	1.00
	0.70	0.90	0.96	1.00	1.00	1.00
τ_μ	1.00	0.08	0.08	0.06	0.19	0.65
	0.97	0.12	0.13	0.09	0.36	0.87
	0.95	0.17	0.18	0.13	0.48	0.94
	0.90	0.30	0.37	0.32	0.75	0.99
	0.80	0.58	0.72	0.88	0.97	1.00
	0.70	0.75	0.88	0.99	1.00	1.00
τ_{GLS}^μ	1.00	0.07	0.08	0.06	0.19	0.63
	0.97	0.20	0.22	0.16	0.46	0.88
	0.95	0.32	0.35	0.27	0.61	0.93
	0.90	0.57	0.66	0.66	0.84	0.98
	0.80	0.83	0.90	0.98	0.96	1.00
	0.70	0.92	0.97	0.99	0.99	1.00

Tabla 2.18: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^r	1.00	0.12	0.12	0.10	0.11	0.18
	0.97	0.12	0.13	0.10	0.12	0.18
	0.95	0.12	0.13	0.09	0.12	0.18
	0.90	0.14	0.13	0.09	0.12	0.17
	0.80	0.16	0.16	0.11	0.15	0.18
	0.70	0.18	0.19	0.13	0.16	0.17
\overline{CS}_A^r	1.00	0.13	0.14	0.10	0.11	0.16
	0.97	0.13	0.14	0.10	0.11	0.15
	0.95	0.13	0.14	0.10	0.11	0.15
	0.90	0.15	0.15	0.10	0.11	0.13
	0.80	0.16	0.18	0.11	0.12	0.14
	0.70	0.19	0.20	0.13	0.13	0.13
τ_τ	1.00	0.45	0.41	0.38	0.44	0.64
	0.95	0.45	0.42	0.38	0.45	0.65
	0.97	0.46	0.41	0.39	0.45	0.66
	0.90	0.45	0.44	0.40	0.48	0.68
	0.80	0.47	0.44	0.40	0.55	0.72
	0.70	0.49	0.46	0.43	0.62	0.76
τ_{GLS}^T	1.00	0.13	0.12	0.11	0.17	0.38
	0.97	0.13	0.13	0.11	0.20	0.41
	0.95	0.12	0.12	0.11	0.20	0.41
	0.90	0.14	0.13	0.11	0.23	0.44
	0.80	0.15	0.14	0.12	0.29	0.49
	0.70	0.17	0.16	0.15	0.37	0.53

Tabla 2.19: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^r	1.00	0.30	0.20	0.11	0.23	0.63
	0.97	0.32	0.22	0.11	0.28	0.63
	0.95	0.35	0.25	0.12	0.31	0.65
	0.90	0.42	0.32	0.16	0.40	0.70
	0.80	0.57	0.49	0.26	0.58	0.74
	0.70	0.63	0.61	0.42	0.68	0.77
\overline{CS}_A^r	1.00	0.33	0.23	0.11	0.24	0.64
	0.97	0.35	0.23	0.11	0.29	0.66
	0.95	0.37	0.26	0.12	0.32	0.68
	0.90	0.45	0.34	0.15	0.42	0.72
	0.80	0.58	0.50	0.24	0.61	0.77
	0.70	0.63	0.62	0.38	0.70	0.78
τ_τ	1.00	0.18	0.15	0.10	0.33	0.85
	0.97	0.19	0.15	0.11	0.40	0.87
	0.95	0.21	0.17	0.11	0.43	0.89
	0.90	0.24	0.20	0.14	0.56	0.93
	0.80	0.32	0.32	0.24	0.80	0.98
	0.70	0.43	0.46	0.44	0.91	0.99
τ_{GLS}^r	1.00	0.13	0.12	0.07	0.31	0.84
	0.97	0.15	0.12	0.08	0.39	0.85
	0.95	0.16	0.15	0.09	0.43	0.88
	0.90	0.22	0.20	0.14	0.56	0.92
	0.80	0.35	0.37	0.28	0.79	0.96
	0.70	0.46	0.53	0.52	0.90	0.97

Tabla 2.20: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 repeticiones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
CS_1^T	1.00	0.20	0.12	0.06	0.20	0.70
	0.97	0.25	0.16	0.07	0.27	0.80
	0.95	0.33	0.22	0.11	0.37	0.85
	0.90	0.50	0.44	0.27	0.59	0.93
	0.80	0.70	0.73	0.66	0.85	0.97
	0.70	0.80	0.86	0.85	0.93	0.97
CS_A^T	1.00	0.22	0.14	0.07	0.22	0.76
	0.97	0.27	0.18	0.08	0.30	0.82
	0.95	0.35	0.25	0.12	0.38	0.87
	0.90	0.50	0.44	0.25	0.61	0.95
	0.80	0.69	0.73	0.66	0.88	0.98
	0.70	0.79	0.86	0.88	0.95	0.98
τ_T	1.00	0.10	0.11	0.07	0.28	0.86
	0.97	0.12	0.13	0.08	0.35	0.91
	0.95	0.15	0.16	0.10	0.44	0.95
	0.90	0.25	0.29	0.21	0.68	0.99
	0.80	0.46	0.61	0.65	0.94	1.00
	0.70	0.63	0.80	0.95	0.99	1.00
τ_{GLS}^T	1.00	0.09	0.09	0.06	0.27	0.84
	0.97	0.12	0.12	0.08	0.37	0.90
	0.95	0.16	0.17	0.11	0.47	0.94
	0.90	0.31	0.37	0.29	0.72	0.98
	0.80	0.57	0.71	0.81	0.95	1.00
	0.70	0.73	0.87	0.98	0.99	1.00

Tabla 2.21: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.17	0.18	0.19	0.20	0.34
	0.97	0.22	0.22	0.24	0.24	0.40
	0.95	0.25	0.26	0.26	0.28	0.41
	0.90	0.31	0.31	0.32	0.32	0.43
	0.80	0.37	0.39	0.39	0.38	0.48
	0.70	0.43	0.43	0.44	0.42	0.49
\overline{CS}_{1T}^μ	1.00	0.19	0.19	0.19	0.22	0.36
	0.97	0.23	0.23	0.25	0.27	0.42
	0.95	0.26	0.27	0.27	0.30	0.43
	0.90	0.31	0.31	0.33	0.34	0.45
	0.80	0.37	0.38	0.41	0.40	0.50
	0.70	0.42	0.43	0.45	0.45	0.52
\overline{CS}_A^μ	1.00	0.20	0.21	0.22	0.23	0.37
	0.97	0.24	0.25	0.27	0.27	0.43
	0.95	0.27	0.29	0.29	0.30	0.44
	0.90	0.32	0.33	0.35	0.35	0.46
	0.80	0.37	0.39	0.41	0.41	0.51
	0.70	0.42	0.43	0.45	0.46	0.53
τ_μ	1.00	0.31	0.25	0.22	0.20	0.27
	0.97	0.31	0.27	0.22	0.22	0.28
	0.95	0.32	0.28	0.22	0.22	0.29
	0.90	0.33	0.30	0.24	0.24	0.34
	0.80	0.34	0.30	0.27	0.26	0.37
	0.70	0.35	0.30	0.27	0.29	0.38
τ_{GLS}^μ	1.00	0.17	0.13	0.12	0.11	0.19
	0.97	0.19	0.16	0.13	0.13	0.20
	0.95	0.21	0.17	0.14	0.14	0.21
	0.90	0.25	0.20	0.17	0.17	0.24
	0.80	0.28	0.25	0.21	0.20	0.27
	0.70	0.30	0.26	0.23	0.22	0.32

Tabla 2.22: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.20	0.16	0.13	0.13	0.36
	0.97	0.32	0.26	0.23	0.21	0.47
	0.95	0.41	0.32	0.28	0.25	0.52
	0.90	0.55	0.49	0.42	0.36	0.63
	0.80	0.74	0.68	0.61	0.54	0.79
	0.70	0.80	0.76	0.70	0.66	0.87
\overline{CS}_{1T}^μ	1.00	0.22	0.17	0.14	0.15	0.40
	0.97	0.34	0.27	0.24	0.23	0.51
	0.95	0.43	0.34	0.30	0.28	0.58
	0.90	0.55	0.50	0.45	0.40	0.69
	0.80	0.73	0.68	0.64	0.60	0.84
	0.70	0.80	0.77	0.74	0.72	0.91
\overline{CS}_A^μ	1.00	0.25	0.19	0.15	0.14	0.39
	0.97	0.35	0.29	0.26	0.23	0.51
	0.95	0.44	0.34	0.30	0.27	0.57
	0.90	0.54	0.48	0.43	0.38	0.71
	0.80	0.69	0.65	0.61	0.60	0.86
	0.70	0.78	0.75	0.73	0.73	0.92
τ_μ	1.00	0.12	0.09	0.06	0.08	0.32
	0.97	0.13	0.09	0.08	0.10	0.40
	0.95	0.15	0.11	0.08	0.12	0.46
	0.90	0.18	0.13	0.12	0.17	0.60
	0.80	0.30	0.23	0.22	0.34	0.83
	0.70	0.41	0.35	0.34	0.56	0.94
τ_{GLS}^μ	1.00	0.09	0.06	0.05	0.07	0.32
	0.97	0.14	0.11	0.09	0.20	0.40
	0.95	0.20	0.15	0.12	0.31	0.45
	0.90	0.31	0.25	0.21	0.58	0.54
	0.80	0.50	0.43	0.38	0.79	0.65
	0.70	0.62	0.56	0.51	0.83	0.71

Tabla 2.23: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1^μ	1.00	0.12	0.10	0.08	0.09	0.35
	0.97	0.32	0.28	0.23	0.22	0.59
	0.95	0.46	0.43	0.38	0.33	0.69
	0.90	0.75	0.73	0.64	0.58	0.83
	0.80	0.91	0.93	0.88	0.79	0.94
	0.70	0.96	0.97	0.92	0.87	0.98
\overline{CS}_{1T}^μ	1.00	0.12	0.09	0.08	0.10	0.38
	0.97	0.33	0.28	0.23	0.25	0.64
	0.95	0.45	0.41	0.38	0.35	0.74
	0.90	0.72	0.69	0.63	0.61	0.89
	0.80	0.88	0.90	0.88	0.85	0.97
	0.70	0.94	0.96	0.94	0.93	0.99
\overline{CS}_A^μ	1.00	0.15	0.12	0.09	0.09	0.40
	0.97	0.34	0.29	0.23	0.23	0.68
	0.95	0.44	0.39	0.35	0.33	0.80
	0.90	0.69	0.65	0.60	0.62	0.95
	0.80	0.87	0.90	0.89	0.91	1.00
	0.70	0.94	0.98	0.97	0.98	1.00
τ_μ	1.00	0.09	0.06	0.05	0.08	0.45
	0.97	0.13	0.09	0.08	0.13	0.69
	0.95	0.17	0.13	0.11	0.18	0.82
	0.90	0.33	0.28	0.25	0.41	0.97
	0.80	0.63	0.62	0.63	0.85	1.00
	0.70	0.80	0.84	0.85	0.98	1.00
τ_{GLS}^μ	1.00	0.07	0.05	0.05	0.08	0.38
	0.97	0.22	0.17	0.15	0.20	0.63
	0.95	0.33	0.30	0.26	0.31	0.71
	0.90	0.63	0.60	0.52	0.55	0.79
	0.80	0.86	0.89	0.84	0.79	0.83
	0.70	0.94	0.96	0.90	0.83	0.85

Tabla 2.24: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1	1.00	0.15	0.19	0.22	0.26	0.31
	0.97	0.16	0.18	0.23	0.26	0.33
	0.95	0.16	0.19	0.23	0.26	0.33
	0.90	0.18	0.20	0.24	0.28	0.34
	0.80	0.20	0.23	0.28	0.29	0.34
	0.70	0.22	0.27	0.31	0.33	0.35
\overline{CS}_A	1.00	0.16	0.20	0.23	0.27	0.31
	0.97	0.17	0.20	0.24	0.27	0.33
	0.95	0.17	0.21	0.24	0.27	0.33
	0.90	0.18	0.21	0.25	0.28	0.34
	0.80	0.21	0.24	0.28	0.30	0.34
	0.70	0.23	0.28	0.32	0.33	0.35
τ_τ	1.00	0.41	0.39	0.35	0.31	0.35
	0.97	0.41	0.39	0.35	0.31	0.35
	0.95	0.42	0.39	0.35	0.33	0.35
	0.90	0.42	0.40	0.35	0.33	0.37
	0.80	0.42	0.40	0.37	0.34	0.38
	0.70	0.43	0.40	0.37	0.35	0.41
τ_{GLS}^T	1.00	0.11	0.10	0.08	0.08	0.09
	0.97	0.11	0.10	0.08	0.08	0.09
	0.95	0.11	0.10	0.09	0.08	0.09
	0.90	0.11	0.10	0.09	0.09	0.10
	0.80	0.11	0.11	0.10	0.09	0.10
	0.70	0.11	0.11	0.10	0.11	0.13

Tabla 2.25: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 repeticiones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
CS_1^T	1.00	0.38	0.31	0.27	0.26	0.54
	0.97	0.39	0.34	0.30	0.27	0.54
	0.95	0.45	0.37	0.31	0.30	0.58
	0.90	0.51	0.45	0.40	0.38	0.65
	0.80	0.66	0.59	0.56	0.52	0.77
	0.70	0.74	0.69	0.65	0.65	0.84
\overline{CS}_A^T	1.00	0.42	0.33	0.29	0.27	0.53
	0.97	0.44	0.37	0.33	0.28	0.54
	0.95	0.48	0.40	0.34	0.30	0.58
	0.90	0.55	0.48	0.41	0.38	0.64
	0.80	0.68	0.61	0.57	0.51	0.77
	0.70	0.74	0.70	0.65	0.65	0.84
τ_T	1.00	0.15	0.10	0.08	0.11	0.35
	0.97	0.15	0.10	0.09	0.11	0.35
	0.95	0.17	0.11	0.09	0.12	0.38
	0.90	0.19	0.13	0.11	0.14	0.46
	0.80	0.30	0.18	0.15	0.23	0.64
	0.70	0.41	0.24	0.23	0.38	0.79
τ_{GLS}^T	1.00	0.11	0.07	0.06	0.07	0.30
	0.97	0.11	0.07	0.07	0.08	0.30
	0.95	0.13	0.09	0.07	0.09	0.31
	0.90	0.16	0.11	0.09	0.12	0.40
	0.80	0.26	0.19	0.16	0.23	0.54
	0.70	0.36	0.27	0.25	0.35	0.65

Tabla 2.26: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $3 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
\overline{CS}_1	1.00	0.22	0.15	0.13	0.13	0.48
	0.97	0.30	0.21	0.18	0.17	0.57
	0.95	0.37	0.30	0.25	0.23	0.65
	0.90	0.58	0.50	0.43	0.43	0.80
	0.80	0.79	0.77	0.73	0.70	0.94
	0.70	0.88	0.89	0.85	0.84	0.98
\overline{CS}_A	1.00	0.26	0.18	0.15	0.12	0.48
	0.97	0.33	0.23	0.20	0.17	0.56
	0.95	0.39	0.32	0.26	0.22	0.64
	0.90	0.59	0.51	0.43	0.41	0.79
	0.80	0.79	0.77	0.73	0.69	0.94
	0.70	0.88	0.90	0.86	0.84	0.98
τ_τ	1.00	0.09	0.07	0.06	0.08	0.58
	0.97	0.12	0.08	0.07	0.11	0.66
	0.95	0.15	0.10	0.09	0.14	0.74
	0.90	0.25	0.18	0.16	0.27	0.91
	0.80	0.47	0.42	0.41	0.65	1.00
	0.70	0.66	0.65	0.65	0.90	1.00
τ_{GLS}^τ	1.00	0.09	0.06	0.05	0.08	0.51
	0.97	0.13	0.08	0.07	0.11	0.60
	0.95	0.17	0.12	0.10	0.15	0.69
	0.90	0.32	0.25	0.22	0.33	0.82
	0.80	0.60	0.57	0.54	0.68	0.91
	0.70	0.77	0.79	0.76	0.84	0.93

Tabla 2.27: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 repeticiones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(\hat{p})$ con $0 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^μ	1.00	0.11	0.10	0.06	0.11	0.24
	0.97	0.14	0.12	0.07	0.14	0.25
	0.95	0.16	0.14	0.09	0.16	0.26
	0.90	0.18	0.16	0.09	0.19	0.26
	0.80	0.21	0.19	0.12	0.22	0.28
	0.70	0.24	0.21	0.16	0.24	0.31
ECS_{1T}^μ	1.00	0.11	0.09	0.06	0.12	0.26
	0.97	0.12	0.11	0.08	0.14	0.27
	0.95	0.16	0.13	0.08	0.16	0.27
	0.90	0.18	0.15	0.08	0.20	0.29
	0.80	0.21	0.19	0.13	0.23	0.32
	0.70	0.24	0.22	0.18	0.27	0.34
ECS_A^μ	1.00	0.16	0.14	0.08	0.19	0.46
	0.97	0.19	0.16	0.09	0.24	0.52
	0.95	0.20	0.19	0.10	0.27	0.54
	0.90	0.25	0.22	0.13	0.33	0.56
	0.80	0.28	0.27	0.17	0.44	0.60
	0.70	0.32	0.30	0.22	0.52	0.60

Tabla 2.28: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^μ	1.00	0.18	0.13	0.06	0.21	0.53
	0.97	0.25	0.21	0.11	0.33	0.66
	0.95	0.30	0.26	0.14	0.39	0.68
	0.90	0.40	0.39	0.25	0.52	0.74
	0.80	0.53	0.52	0.49	0.68	0.77
	0.70	0.59	0.62	0.65	0.73	0.80
ECS_{1T}^μ	1.00	0.16	0.11	0.05	0.22	0.61
	0.97	0.24	0.18	0.10	0.35	0.73
	0.95	0.30	0.24	0.12	0.42	0.76
	0.90	0.40	0.37	0.24	0.56	0.83
	0.80	0.56	0.54	0.47	0.75	0.87
	0.70	0.64	0.65	0.67	0.81	0.88
ECS_A^μ	1.00	0.19	0.13	0.06	0.31	0.77
	0.97	0.28	0.22	0.10	0.46	0.88
	0.95	0.34	0.27	0.13	0.54	0.91
	0.90	0.43	0.39	0.24	0.71	0.96
	0.80	0.59	0.60	0.51	0.88	0.98
	0.70	0.68	0.71	0.80	0.96	0.98

Tabla 2.29: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^μ	1.00	0.11	0.08	0.05	0.15	0.50
	0.97	0.26	0.23	0.14	0.37	0.75
	0.95	0.35	0.34	0.25	0.48	0.81
	0.90	0.53	0.60	0.56	0.69	0.90
	0.80	0.69	0.79	0.86	0.85	0.94
	0.70	0.78	0.86	0.91	0.91	0.95
ECS_{1T}^μ	1.00	0.11	0.08	0.05	0.17	0.57
	0.97	0.27	0.23	0.16	0.42	0.83
	0.95	0.37	0.36	0.26	0.54	0.90
	0.90	0.57	0.61	0.57	0.77	0.97
	0.80	0.76	0.84	0.88	0.93	0.99
	0.70	0.85	0.91	0.95	0.97	0.99
ECS_A^μ	1.00	0.13	0.09	0.05	0.20	0.64
	0.97	0.27	0.22	0.15	0.46	0.90
	0.95	0.36	0.33	0.23	0.60	0.94
	0.90	0.56	0.61	0.54	0.85	0.99
	0.80	0.81	0.88	0.95	0.99	1.00
	0.70	0.92	0.97	1.00	1.00	1.00

Tabla 2.30: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 25$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^T	1.00	0.12	0.12	0.09	0.13	0.20
	0.97	0.12	0.13	0.09	0.12	0.19
	0.95	0.12	0.13	0.10	0.12	0.19
	0.90	0.13	0.13	0.10	0.13	0.18
	0.80	0.16	0.15	0.11	0.15	0.19
	0.70	0.17	0.17	0.12	0.16	0.20
ECS_A^T	1.00	0.13	0.13	0.11	0.14	0.19
	0.97	0.12	0.12	0.10	0.14	0.18
	0.95	0.13	0.13	0.11	0.13	0.18
	0.90	0.14	0.14	0.12	0.13	0.17
	0.80	0.16	0.16	0.12	0.14	0.16
	0.70	0.18	0.19	0.14	0.15	0.17

Tabla 2.31: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 replicaciones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $\text{BIC}(p)$ con $0 \leq p \leq 8$

$T = 50$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^T	1.00	0.28	0.18	0.09	0.24	0.64
	0.97	0.30	0.20	0.11	0.27	0.65
	0.95	0.32	0.21	0.12	0.30	0.66
	0.90	0.37	0.29	0.14	0.40	0.73
	0.80	0.48	0.41	0.26	0.59	0.77
	0.70	0.57	0.51	0.41	0.68	0.79
ECS_A^T	1.00	0.28	0.17	0.10	0.20	0.46
	0.97	0.30	0.19	0.09	0.18	0.38
	0.95	0.32	0.21	0.10	0.18	0.34
	0.90	0.38	0.27	0.12	0.20	0.29
	0.80	0.50	0.42	0.17	0.22	0.24
	0.70	0.58	0.53	0.20	0.23	0.27

Tabla 2.32: Nivel y potencia empíricos para el nivel nominal 5%. Basado en 5000 repeticiones del proceso $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$ con $y_1 = a_1$. Criterio $BIC(p)$ con $0 \leq p \leq 8$

$T = 100$		θ				
Contraste	ρ	-0.8	-0.5	0.0	0.5	0.8
ECS_1^T	1.00	0.19	0.10	0.06	0.20	0.75
	0.97	0.24	0.16	0.08	0.29	0.81
	0.95	0.28	0.20	0.11	0.38	0.88
	0.90	0.43	0.38	0.25	0.61	0.95
	0.80	0.63	0.65	0.64	0.87	0.97
	0.70	0.75	0.79	0.83	0.95	0.98
ECS_A^T	1.00	0.19	0.12	0.05	0.24	0.67
	0.97	0.24	0.15	0.06	0.28	0.58
	0.95	0.30	0.19	0.06	0.33	0.50
	0.90	0.45	0.37	0.11	0.45	0.32
	0.80	0.69	0.67	0.16	0.43	0.18
	0.70	0.81	0.83	0.17	0.29	0.20

Capítulo 5

Conclusiones finales y extensiones

En esta tesis doctoral se desarrollan dos líneas de investigación relacionadas con la identificación de raíces unitarias y sus implicaciones en el largo plazo de las series temporales.

La primera línea de investigación analiza las propiedades de los predictores sobrediferenciados. En el capítulo 2 de esta tesis doctoral se demuestra que, si un proceso autorregresivo tiene una raíz próxima al círculo unidad, sobrediferenciar produce menor error cuadrático medio de predicción. Esta ventaja del modelo sobrediferenciado aumenta, además, con el horizonte de predicción. Este resultado procede de la mayor parsimonia del modelo sobrediferenciado. Muchos analistas han sugerido que, cuando existen dudas sobre la presencia de una raíz unitaria, sobrediferenciar puede mejorar las predicciones. Este resultado, sin embargo, no había sido demostrado formalmente.

Esta línea de investigación tiene interesantes extensiones en las siguientes direcciones:

- Puesto que el resultado anterior se ha demostrado para un proceso autorregresivo, sería de interés la extensión al caso $\text{ARMA}(p + 1, q)$, donde una de las raíces del polinomio AR esté próxima al círculo unidad.

- Asimismo sería de interés la extensión al caso de raíces próximas al círculo unidad en el polinomio de medias móviles de procesos no estacionarios. La pregunta a resolver sería ¿Cómo se ve afectado el ECMP si se considera estacionario a un proceso no estacionario con una raíz en el polinomio de medias móviles próxima al círculo unidad? La respuesta de esta cuestión es relevante pues, en dicha situación, los contrastes de raíces unitarias presentan aumentos de su nivel de significación respecto al nominal.

- Otra extensión de los resultados del capítulo 2 es en la selección de los órdenes p y q de un modelo ARMA. Kunitomo & Yamamoto (1985) demostraron que un modelo AR(p) ajustado a un proceso AR($p+1$) puede proporcionar menor ECMP que el modelo correcto. La metodología empleada en el capítulo 2 puede emplearse para delimitar las situaciones en que dicho resultado es cierto. Podría responderse así a las siguientes cuestiones:
 - Dado un tamaño muestral ¿qué valor debe tener una raíz para poder prescindir de ella en la modelización?

 - Si se concluye que no se puede prescindir de una raíz para elaborar predicciones eficientes a horizonte $h = 1$ ¿puede existir un horizonte para el que dicha raíz sea prescindible? Una respuesta afirmativa llevaría a recomendar la técnica de predicción adaptativa incluso si el modelo es conocido.

En la segunda línea de investigación se proponen nuevos contrastes de raíces unitarias. Los contrastes propuestos están basados en los errores de predicción que se cometen bajo

la hipótesis nula de raíz unitaria. Los estadísticos de los contrastes son ratios entre los errores cuadráticos multihorizonte de un predictor que supone que existe raíz unitaria y el ECMP teórico si dicha hipótesis fuese cierta. Las simulaciones realizadas muestran que, si se emplean estimaciones eficientes, los estadísticos propuestos tienen una elevada potencia en las zonas próximas al círculo unidad y casi nula distorsión del nivel de significación en presencia de componentes de medias móviles. Estas dos características hacen que los contrastes que se proponen en esta tesis doctoral sean alternativas interesantes a los ya existentes. Los contrastes basados en aproximaciones autorregresivas pueden proporcionar mejores potencias que los contrastes propuestos en el caso AR(1), sin embargo presentan aumentos considerables de su nivel de significación ante componentes de medias móviles lo que los hace poco recomendables.

Los resultados obtenidos sugieren las siguientes investigaciones futuras:

- Extensión de los contrastes propuestos al caso de raíces unitarias estacionales. Asimismo, se podría extender al caso de varias raíces unitarias.
- Estimación por máxima verosimilitud exacta. Al no depender los contrastes propuestos del método de estimación debe ser posible beneficiarse de métodos de estimación más eficientes.
- Investigar si existe alguna justificación teórica que demuestre que un mayor peso de las observaciones de los extremos de la serie proporciona mayores potencias.
- Investigar la influencia de las condiciones iniciales en la potencia de los contrastes en muestras finitas

Las dos líneas de investigación de esta tesis doctoral pueden también tener extensiones al caso multivariante.

Referencias

- Akaike, H. (1977). "On Entropy Maximization Principle," . En *Applications of Statistics*. 27-41. Ed. P.R. Krishnaiah. North-Holland, Amsterdam.
- Agiakloglou, C. & Newbold, P. (1992). "Empirical Evidence on Dickey-Fuller-Type Tests," *Journal of Time Series Analysis*, 13, 471-483.
- Anderson, T. W. (1948). "On the theory of Testing Serial Correlation," *Skandinavisk Aktuarietidskrift*. 31. 88-116.
- Anderson, T. W. & Takemura, A. (1986). "Why Do Noninvertible Estimated Moving Averages Occur?," *Journal of Time Series Analysis*, 7, 235-254.
- Armstrong, J.S. & Collopy, F. (1992). "Error Measures for Generalizing About Forecasting Methods: Empirical comparisons," *International Journal of Forecasting*. 8. 69-80.
- Berk, K. N. (1974). "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489-502.
- Beveridge, S. & Nelson, C. R. (1981). "A New Approach to Decomposition of Economic

- Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151-174.
- Bhargava, A. (1986). "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, 369-384.
- Bierens, H. J. (1993). "Higher-Order Sample Autocorrelations and the Unit Root Hypothesis," *Journal of Econometrics*, 57, 137-160.
- Billingsley, P. (1968). *Convergence of Probability Measures*. New York. John Wiley.
- Bhansali, R.J. (1978). "Linear Prediction by Autoregressive Model Fitting in the Time Domain," *The Annals of Statistics*, 6, 224-231.
- Bhansali, R. J. (1981). "Effects of Not Knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction- I," *Journal of the American Statistical Association*, 76, 588-597.
- Blough, S. R. (1992). "The Relationship Between Power and Level for Generic Unit Root Test in Finite Samples," *Journal of Applied Econometrics*, 7, 295-308.
- Box, G. E. P. & Tiao, G. C. (1976). "Comparison of Forecast and Actuality," *Applied Statistics*, 25, 195-200.
- Box, G. E. P. & Tiao, G. C. (1977). "A Canonical Analysis of Multiple Time Series," *Biometrika*, 6, 355-365.
- Campbell, J. H. (1988). "How Big is the Random Walk in GNP," *Journal of Political*

Economy, 96, 893–920.

Campbell, J. Y. & Perron, P. (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," *NBER Macroeconomics Annual*, 141–201.

Chan, N. H. & Wei, C. Z. (1988). "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *The Annals of Statistics*, 16, 367–401.

Choi, I. (1990). "The Durbin–Hausman Tests for a Unit Root When Errors Are Serially Correlated," *Working Paper 91–10*. Dep. of Economics. The Ohio State University.

Choi, I. (1992). "Durbin–Hausman Tests for a Unit Root," *Oxford Bulletin of Economics and Statistics*, 54, 289–304.

Clements, M.P. & Hendry, D.F. (1993). "On the limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting*, 12, 617–637.

Cox, D. R. (1961). "Prediction by Exponentially Weighted Moving Averages and Related Methods," *Journal of the Royal Statistical Society, Serie B*, 23, 414–422.

Cryer, J. D. & Ledolter, J. (1981). "Small-Sample Properties of the Maximum–Likelihood Estimator in the First Order Moving Average Model," *Biometrika*, 68, 691–694.

Davies, N. & Newbold, P. (1980). "Forecasting with Misspecified Models," *Applied Statistics*, 29, 87–92.

DeJong, D. N., & Whiteman, C. H. (1993). "Estimating Moving Average Parameters: Classical Pitfalls and Bayesian Posteriors," *Journal of Business & Economic Statistics*, 11, 311–317.

- Dickey, D. A. (1976). "Estimation and Hypotesis Testing in Nonstationary Time Series," Ph. D. dissertation, Iowa State University.
- Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.
- Dickey, D. A., & Fuller, W. A. (1981). "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Roots," *Econometrica*, 49, 1057-1072.
- Dickey, D. A. & Pantula, S. G. (1987). "Determining the Order of Differencing in Autor-regressive Processes," *Journal of Business & Economic Statistics*, 5, 455-461.
- Donsker, M. D. (1951). "An Invariance Principle for Certain Probability Limit Theorems," *Memoirs of the American Mathematical Society*. 6. 1-12.
- Dufour, J. M. & King, M. L. (1991). "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115-143.
- Durbin, J. (1954). "Errors in Variables," *Review of the International Statistical Institute*. 22. 23-32.
- Durbin, J. (1959). "Efficient Estimation of Parameters in Moving-Average Models," *Biometrika*, 46, 306-316.
- Durbin, J. & Watson, G. S. (1950). "Testing for Serial Correlation in Least Squares Regression I," *Biometrika*, 37, 409-428.

- Durbin, J. & Watson, G. S. (1951). "Testing for Serial Correlation in Least Squares Regression II," *Biometrika*, 38, 159-178.
- Durbin, J. & Watson, G. S. (1951). "Testing for Serial Correlation in Least Squares Regression III," *Biometrika*, 58, 1-19.
- Elliot, G., Rothenberg, T. J. & Stock, J. H. (1992). "Efficient Tests for an Autorregresive Unit Root," NBER Technical Working Paper No. 130.
- Engle, R. F. & Granger, C. W. J. (1987). "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.
- Evans, G.B.A. & Savin, N.E. (1984). "Testing for Unit Roots: 2," *Econometrica*, 52, 1241-1269.
- Fama, E. F. & French, K. R. (1988). "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246-273.
- Ferreti, N. & Romo, J. (1995). "Unit root bootstrap tests for AR(1) models," Aceptado en *Biometrika*.
- Fildes, R. (1992). "The Evaluation of Extrapolative Forecasting Methods," *International Journal of Forecasting*, 8, 81-98.
- Findley, D. F. (1984). "On Some Ambiguities Associated With the Fitting of ARMA Models to Time Series," *Journal of Time Series Analysis*, 5, 213-225.
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*. New York: John Wiley.
- Gersh, W. & Kitagawa, G. (1983). "The Prediction of Time Series With Trends and

Seasonalities," *Journal of Business & Economic Statistics*, 1, 253–263.

Gonzalo, J. & Lee, T (1994). "Relative Power of t Type Tests for Stationary and Unit Root Processes" *Journal of Time Series Analysis*, 17, 37–47.

Granger, C. W. J. (1981). "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121– 130.

Granger, C. W. J. (1986). "Developments in the study of cointegrated economic variables," *Journal of Business & Economic Statistics*, 48, 213–228.

Granger, C. W. J. & Newbold, P. (1973). "Some comments on the evaluation of economic forecasts," *Applied Economics*. 5. 35- 47.

Guilkey, D. K. & Schmidt, P. (1991). "An Interpolation Test for a Unit Root in the Presence of Drift," *Biometrika*, 78, 809–815.

Hall, A. (1989). "Testing for a Unit Root in the Presence of Moving Average Errors," *Biometrika*, 76, 49–56.

Hall, A. (1994). "Testing for a Unit Root in Time Series With Pretest Data-Based Model Selection," *Journal of Business & Economic Statistics*, 12, 461–470.

Hall, P. & Heide, C. C. (1980). *Martingale Limit Theory and its Applications*, Academic Press. London.

Hamilton, J. D. (1994). *Time Series Analysis*. New York: Princeton University Press.

Hannan, E.J. & Quinn, B.G. (1979). "The determination of the order of an autoregression," *Journal of the Royal Statistical Society, Serie B*, 41, 190–195.

- Harvey, A. C. (1981), "Finite Sample Prediction and Overdifferencing," *Journal of Time Series Analysis*, 2, 221– 232.
- Hassler, U. (1994). "The Sample Autocorrelation Function of I(1) Processes," *Statistical Papers*, 35, 1–16.
- Hasza, D. P. (1980). "The Asymptotic Distribution of the Sample Autocorrelation for an Integrated ARMA Process," *Journal of the American Statistical Association*, 75, 349– 352.
- Hausman, J. A. (1978). "Specification Tests in Econometrics," *Econometrica*, 6, 1251– 1271.
- Herrndorf, N. (1984). "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," em *Annals of Probability*. 12. 141–153.
- Hosoya, Y. and Taniguchi, M. (1982). "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," *The Annals of Statistics*, 10, 132–153.
- Imhof, P. J. (1961). "Computing the distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.
- Kendall, M. G. (1954). "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403–404.
- King, M. L. (1980). "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271.

- King, M. L. & Hillier, G. H. (1985). "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society, Serie B*, 47, 98-102.
- Kunitomo, N. & Yamamoto, T. (1985). "Properties of Predictors in Misspecified Autoregressive Time Series Models," *Journal of the American Statistical Association*, 80, 941-950.
- Kwiatkowski, D., Phillips, P. C. B. Schmidt, P. & Shin, Y. (1992). "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 54, 159-178.
- LaMotte, L. R. & McWhorter, A. Jr. (1978). "An Exact Test for the Presence of Random Walk Coefficients in Linear Regression Model," *Journal of the American Statistical Association*, 73, 816-820.
- Ledolter, J. & Abraham, B. (1981). "Parsimony and Its Importance in Time Series Forecasting," *Technometrics*, 23, 411-414.
- Leybourne, S. J. (1995). "Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions," *Oxford Bulletin of Economics and Statistics*, 57, 559-571.
- Leybourne, S. J. & McCabe, B. P. M. (1994). "A consistent Test for a Unit Root," *Journal of Business & Economic Statistics*, 12, 157-166.
- Li, H. (1995). "The Power of Modified Instrumental Variables Tests for Unit Roots," *Biometrika*, 82, 660-666.

- Lo, A. W. & MacKinlay A. C. (1988). "The Size and Power of the Variance Ratio Test in Finite Samples," *Journal of Econometrics*, 40, 203-238.
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- Mann, H. B. & Wald, A. (1943). "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, 11, 172-200.
- Marriot, F. H. C. and Pope, J. A. (1954). "Bias in Estimation of Autocorrelations," *Biometrika*, 41, 390-402.
- Nankervis, J. C. & Savin, N. E. (1985). "Testing the Autoregressive Parameter With the t Statistics," *Journal of Econometrics*, 27, 143-161.
- Newey, W. & West, K.D. (1987). "A Simple Positive Semi Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
- Ng, S. & Perron, P. (1994). "Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *manuscrito*. Université de Montréal.
- Nyblom, J. & Mäkeläinen, T. (1983). "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association*, 78, 856-864.
- Pantula, S. G. & Hall, A. (1991). "Testing for Unit Roots in Autoregressive Moving Average Models," *Journal of Econometrics*, 48, 325-353.

- Pantula, S. G., Gonzalez-Farias, G. & Fuller, W. A. (1994). "A Comparison of Unit-Root Test Criteria," *Journal of Business & Economic Statistics*, 12, 449-459.
- Park, H. J. & Fuller, W. A. (1994). "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis*, 16, 415-429.
- Park, J. Y. & Phillips, P. C. B. (1988). "Statistical Inference in Regressions with Integrated Processes. Part 1," *Econometric Theory*, 4, 468-497.
- Park, J. Y. & Phillips, P. C. B. (1989). "Statistical Inference in Regressions with Integrated Processes. Part 2," *Econometric Theory*, 5, 95-131.
- Perron, P. (1990). "Tests of Joint Hypotheses for Time Series Regression with a Unit Root," en *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, volumen 8. Ed. T Fomby & G. Rhodes. pag. 135-159. Greenwich, Connecticut. JAI Press.
- Perron, P. & Ng S. (1995). "Unit Roots Tests in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268-281.
- Phillips, P. C. B. (1987a). "Time Series Regression With Unit Root," *Econometrica*, 55, 277-301.
- Phillips, P. C. B. (1987b). "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535-547.
- Phillips, P. C. B. & Perron P. (1988). "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.

- Plosser, C. I. & Schwert, G. W. (1977), "Estimation of a Non-invertible Moving Average Process: The Case of Overdifferencing," *Journal of Econometrics*, 6, 199-224.
- Plosser, C. I. & Schwert, G. W. (1978), "Money, Income and Sunspots: Measuring Economic Relationships and the Effects of Differencing," *Journal of Monetary Economics*, 4, 637-660.
- Rissanen, J. (1978). "Modeling by shortest data description," *Automatica*, 14, 465-471.
- Rubin, H. (1950). "Consistency of Maximum-Likelihood Estimates in the Explosive Case," en *Statistical Inference in Dynamic Economic Models*, Ed. T.C. Koopmans, New York. John Wiley & Sons.
- Said, S.E. & Dickey, D.A. (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.
- Saikkonen, P. (1993). "A Note on a Lagrange Multiplier Test for Testing an Autoregressive Unit Root" *Econometric Theory*, 9, 494-498.
- Saikkonen, P. & Luukkonen, R. (1993). "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association*, 88, 596-601.
- Sargan, J. D. (1973). "The Durbin-Watson Ratio of the Gaussian Random Walk," . Manuscripto. Department of Economics. London School of Economics. C.R.D.E.
- Sargan, J. D. & Bhargava, A. (1983). "Testing the Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 41, 153-174.

- Shaman, P. & Stine, R. (1988). "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842-848.
- Schmidt, P. (1990). "Dickey-Fuller Tests With Drift," en *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, volumen 8. Ed. T Fomby & G. Rhodes. pag. 161-200. Greenwich, Connecticut. JAI Press.
- Schmidt, P. & Phillips, P. C. B. (1992). "LM Test for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257-287.
- Schwartz, G. (1978). "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461-464.
- Schwert, G. W. (1989). "Testing for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7, 147-160.
- Shenton, L. R. & Johnson, W. L. (1965). "Moments of a Serial Correlation Coefficient," *Journal of the Royal Statistical Society, Serie B*, 27, 308-320.
- Solo, V. (1984). "The Order of Differencing in ARIMA Modelling," *Journal of the American Statistical Association*, 79, 916-920.
- Stock, J. H. (1994). "Unit Roots, Structural Breaks and Trends," en R. F. Engle & D. L. McFadden, eds, *Handbook of Econometrics*. Vol. IV. Elsevier. Francia.
- Tanaka, K. (1984). "An Asymptotic Expansion Associated with the Maximum Likelihood Estimators in ARMA Models," *Journal of the Royal Statistical Society, Serie B*, 46, 58-67.

- Tanaka, K. (1990). "Testing for a Moving Average Unit Root," *Econometric Theory*, 6, 433-444.
- Tanaka, K. & Maekawa, K. (1984). "The Sampling Distributions of the Predictor for an Autoregression Model under Misspecifications," *Journal of Econometrics*, 25, 327-351.
- Tanaka, K. & Satchell, S. E. (1989). "Asymptotic Properties of the Maximum Likelihood and Non-linear Least Squares Estimators for Non-invertible Moving Average Models," *Econometric Theory*, 5, 333-353.
- Tiao G. C. & Tsay, R. S. (1994). "Some Advances in Nonlinear and Adaptive Modelling in Time Series Analysis," *Journal of Forecasting*, 13, 109-131
- Tiao, G. C. & Xu, D. (1993). "Robustness of Maximum Likelihood Estimates for Multi-step Predictions: The Exponential Smoothing Case," *Biometrika*, 80, 623-641.
- Tjøsteim, D. & Paulsen, J. (1983). "Bias of Some Commonly-Used Time Series Estimates," *Biometrika*, 70,, 389-399.
- Tsay, R. S. (1984). "Order Selection in Nonstationary Autoregressive Models," *The Annals of Statistics*, 12, 1425-1433.
- Tsay, R. S. (1993). "Adaptive Forecasting, A Comment on 'Calculating Interval Forecasts' by Chatfield, C.," *Journal of Business & Economic Statistics*, 11, 140-142.
- Tsay, R. S. (1993), "Testing for Noninvertible Models With Applications," *Journal of Business & Economic Statistics*, 11, 225-233.

- von Neumann, J., Kent, R. H., Bellinson, H. R. & Hart, B. I. (1941). "The Mean Squares Successive Difference," *Annals of Mathematical Statistics*, 12, 153–162.
- Weiss, A. A. (1991). "Multi-step Estimation and Forecasting in Dynamic Models," *Journal of Econometrics*, 48, 135–149.
- Weiss, A. A. (1984). "Estimating Time Series Models Using the Relevant Forecast Evaluation Criterion," *Journal of the Royal Statistical Society, Serie A*, 147, 484–487.
- White, J. S. (1958). "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188– 1197.
- White, J. S. (1961). "Asymptotic Expansions for the Mean and Variance of Serial Correlation Coefficient," *Biometrika*, 48, 85–94.
- Wichern, D. W. (1973). "The Behaviour of the Sample Autocorrelation Function for an Integrated Moving Average Process," *Biometrika*, 60, 235–239.
- Yap, S. F. & Reinsel, G. C. (1993). "Results on Estimation and Testing for a Unit Root in the Nonstationary Autoregressive Moving- -Average Model," *Journal of Time Series Analysis*, 16, 339–353.
- Yule, G. U. (1921). "On the Time-Correlation Problem," *Journal of the Royal Statistical Society*. 84. 469–537.