

## CAPÍTULO 5

### MODELOS DE SERIES DE TIEMPO UNIVARIADAS NO ESTACIONARIAS

En los capítulos anteriores se ha supuesto que las variables eran estacionarias o, alternativamente, que el econometrista sabía qué tipo de transformación resultaba pertinente para volver estacionaria una variable que no lo era. El objetivo de este capítulo es introducir al lector en la econometría de variables no estacionarias. Para ello, lo primero que se hace es discutir las profundas diferencias que tienen procesos estacionarios y no estacionarios. Posteriormente se desarrollan los elementos centrales de la teoría asintótica para procesos no estacionarios. Finalmente se discuten varios tests que permiten dilucidar empíricamente la naturaleza de las variables y el tratamiento adecuado que se debe darle.

#### 5.01 Representaciones de Tendencias en Análisis Univariado

Muchas de las variables económicas más comúnmente usadas tienen tendencia (p.e., PIB, nivel de precios). Cuando estudiamos la metodología Box-Jenkins o los modelos VAR, el problema de remover la tendencia de las series fue soslayado, señalando que bastaba con hacer una regresión contra una tendencia lineal o tomar primeras diferencias. Exactamente cuándo usar cada método es lo que nos preocupa en esta sección.

Hay dos maneras de describir “tendencias”. Como de costumbre, consideramos las variables en logaritmos, de modo tal que sus cambios corresponden a su tasa de crecimiento.

Aquellas variables que tienen tendencia determinística se pueden modelar como:

$$y_t = \alpha + \beta t + \psi(L) \epsilon_t \quad (5.1)$$

donde  $\alpha$  y  $\beta$  son dos parámetros y  $\psi(L)$  es la representación media móvil de un modelo ARMA(p,q) para las innovaciones  $\epsilon_t$ . La remoción de tendencia correspondería a obtener una serie filtrada del tipo  $y_t - \hat{\alpha} - \hat{\beta}t$ .

Aquellas variables que tienen tendencia estocástica<sup>1</sup> ( $\delta$ ) se pueden modelar usando primeras diferencias, es decir como

$$(1 - L) y_t = \delta + \psi^*(L) \epsilon_t \quad (5.2)$$

Vamos a estudiar las diferencias entre el modelo de la ecuación (5.1) y el de la ecuación (5.2), para concluir que éstos son radicalmente diferentes en tres dimensiones: la naturaleza de los errores de predicción asociados a cada proceso, la varianza de dichos errores de predicción y la persistencia de las innovaciones de cada proceso.

Consideremos primero el error de predicción de una variable descrita por la ecuación (5.1) y re-escribamos el proceso ARMA de manera extensiva:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) \mu_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \epsilon_t \quad (5.3)$$

es decir,

$$\mu_t = \frac{(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)} \epsilon_t \quad (5.4)$$

o sea, el modelo ARMA( $p, q$ ) incluido en la ecuación (5.1) es perfectamente general y corresponde a:

$$\mu_t = \frac{(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)}{(1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)} \epsilon_t \equiv \psi(L) \epsilon_t \quad (5.5)$$

Cuando se cumple sumabilidad absoluta (*absolute summability*)  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , entonces  $|\lambda_i| < 1$  y la variable es estacionaria alrededor de la tendencia determinística  $\beta$ .

Supongamos por otro lado que  $\lambda_1 = 1$  y  $|\lambda_i| < 1$  con  $i = 2, \dots, p$ . Entonces:

$$(1 - L) \mu_t = \frac{(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)}{(1 - \lambda_2 L) \dots (1 - \lambda_p L)} \epsilon_t \equiv \psi^*(L) \epsilon_t \quad (5.6)$$

---

1 Usaremos *deriva* o tendencia estocástica, equivalente al término inglés “drift”, reservando el término *tendencia* para la tendencia determinística. El diccionario de la Real Academia de la Lengua señala que deriva es una “variación lenta y continua de una propiedad que puede ser medida respecto de un valor, dirección o punto preestablecido”.

donde ahora se cumple sumabilidad absoluta para la secuencia de  $\psi^*$ , es decir,

$$\sum_{j=0}^{\infty} |\psi_j^*| < \infty. \text{ Así,}$$

$$\begin{aligned} (1-L) y_t &= (1-L) \alpha + (1-L) \delta t + (1-L) \mu_t \\ &= \delta + \psi^*(L) \epsilon_t \end{aligned} \quad (5.7)$$

El proceso descrito en la ecuación (5.7) para la primera diferencia de la variable es invertible, en tanto que el proceso implícito en la ecuación (5.6) para el nivel de la variable no es estacionario pues posee una raíz unitaria. También se le llama proceso integrado. Lo que señala la ecuación (5.7), entonces, es que un proceso integrado de orden 1 (sólo un  $\lambda_i = 1$ ) se vuelve estacionario al tomar la primera diferencia de la serie.

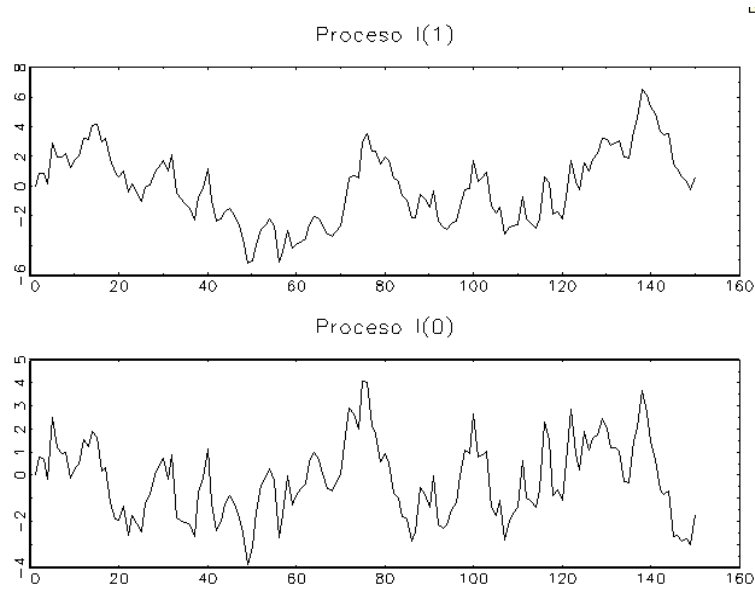
Obviamente si tenemos una serie que cumple  $\lambda_1 = \lambda_2 = 1$  y  $|\lambda_i| < 1$  con  $i=3, \dots, p$  habrán dos raíces unitarias. En tal caso, la segunda diferencia de la variable  $(1-L)^2 y_t = \delta' + \psi'(L) \epsilon_t$  es estacionaria, en tanto que  $y_t$  y  $\Delta y_t$  no lo son. Resulta común denominar a las series de tiempo por su nivel de integración de forma tal que el proceso anterior es I(2) y el proceso de la ecuación (5.2) es I(1), en tanto que el proceso de la ecuación (5.1) es llamado I(0).

¿Qué tan distintos son estos procesos? En términos visuales usualmente resulta prácticamente imposible diferenciar una variable integrada de una estacionaria, en particular cuando la muestra es pequeña. La Figura 5.1 presenta una realización completamente aleatoria de un proceso estacionario AR(1) y un proceso no estacionario paseo aleatorio (*random walk*) con una pequeña tendencia estocástica de tamaño 0.05.<sup>2</sup>

---

2 Reflexionando sobre el tema de la persistencia de las tendencias, el Vicepresidente de Estados Unidos, Dan Quayle señaló: *I believe we are on an irreversible trend toward more freedom and democracy - but that could change.*

**Figura 5.1**  
Simulación de procesos artificiales estacionarios y no estacionarios



### Predicciones<sup>3</sup>

Para el proceso estacionario en tendencia podemos usar la representación media móvil del error para computar:

$$\hat{y}_{t+s|t} = \alpha + \beta(t+s) + \psi_s \epsilon_t + \psi_{s+1} \epsilon_{t-1} + \dots \quad (5.8)$$

es decir, la predicción incluye la constante -  $\alpha$  -, el crecimiento secular de la variable -  $\beta(t+s)$  - y el efecto de las innovaciones que aún están presentes en el instante  $t$  (es decir,  $\epsilon_t, \epsilon_{t-1}, \dots$ ) pero ponderados por el impacto que tienen en cada instante de tiempo  $s$  ( $\psi_s, \psi_{s+1}, \dots$ ).

Debido a que se cumple sumabilidad absoluta, se obtiene

$$E[\hat{y}_{t+s|t} - \alpha - \beta(t+s)]^2 \rightarrow 0 \quad \text{si} \quad s \rightarrow \infty \quad (5.9)$$

<sup>3</sup> Those who have knowledge, don't predict. Those who predict, don't have knowledge, Lao Tzu, poeta chino, siglo VI A.C.

es decir, en el largo plazo las innovaciones van a desaparecer y lo único que queda activo en la predicción es la parte determinística.

Para el proceso estacionario en diferencia podemos usar la representación media móvil del proceso en primera diferencia y luego restar  $y_t$ , para obtener:

$$\begin{aligned}\hat{y}_{t+s|t} = & y_t + s\delta + (\psi_s + \psi_{s-1} + \dots + \psi_1)\epsilon_t \\ & + (\psi_{s+1} + \psi_s + \dots + \psi_2)\epsilon_{t-1} \\ & + (\psi_{s+2} + \psi_{s+1} + \dots + \psi_3)\epsilon_{t-2} \\ & \dots\end{aligned}\tag{5.10}$$

Prueba: Primero predecimos la tasa de crecimiento del periodo “ $t+s$ ”  $\Delta \hat{y}_{t+s|t} = \delta + \psi_s \epsilon_t + \psi_{s+1} \epsilon_{t-1} + \psi_{s+2} \epsilon_{t-2} + \dots$ . El nivel de la serie es la suma de las tasas de crecimiento más el nivel inicial  $y_{t+s|t} = \Delta y_{t+s} + \Delta y_{t+s-1} + \dots + y_t$ . Así,  $\hat{y}_{t+s|t} = \Delta \hat{y}_{t+s} + \Delta \hat{y}_{t+s-1} + \dots + y_t$ , de donde se deduce la ecuación (5.10).■

Es decir, vemos que la predicción en  $t+s$  en el caso de las variables integradas depende de la posición inicial de la serie,  $y_t$ , en tanto que en los modelos de variables estacionarias no.<sup>4</sup> El intercepto en modelos de variables estacionarias es fijo ( $\alpha$ ), en tanto que en series integradas depende de la historia de la variable. Para cada instate  $t$  en el que se haga la predicción habrá un distinto punto de partida,  $y_t$ . En segundo lugar, si  $\delta$  no es cero, la serie tiene tasa de crecimiento pero, a diferencia de los variables estacionarias, ésta no es determinística (constante) sino estocástica: para deducirlo basta con tomar el valor esperado de la ecuación (5.2). En tercer lugar, en la predicción del proceso I(1) la parte estocástica de la tasa de crecimiento –representada por la historia de las innovaciones  $\epsilon_t$ – afecta permanentemente el nivel de la variable  $y_t$  en tanto que en las variables estacionarias no. En resumen, el intercepto en variables estacionarias en tendencia es “fijo”, en cambio en variables estacionarias en diferencia depende de la historia y la innovación contemporánea.

---

4 Esto implica que la ubicación de la variable integrada es, entonces, dependiente de la historia. En algún sentido a las series de tiempo no estacionarias se le aplica la célebre frase de Heráclito “*Nadie se baña dos veces en el mismo río*”.

## Errores de Predicción

De las ecuaciones (5.1) y (5.8) resulta directo demostrar que el error de predicción de una variable estacionaria en tendencia es la suma ponderada de las innovaciones futuras:

$$\hat{y}_{t+s|t} - y_{t+s} = \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \psi_2 \epsilon_{t+s-2} + \dots + \psi_{s-1} \epsilon_{t+1} \quad (5.11)$$

Si las innovaciones son i.i.d., el error cuadrático medio (ECM) de la predicción es:

$$E[\hat{y}_{t+s|t} - y_{t+s}]^2 = [1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2] \sigma^2 \quad (5.12)$$

Evidentemente, el ECM aumenta con  $s$  pero debido a la propiedad de sumabilidad absoluta converge a un valor finito. En particular:

$$\lim_{s \rightarrow \infty} E[\hat{y}_{t+s|t} - y_{t+s}]^2 = [1 + \psi_1^2 + \psi_2^2 + \dots] \sigma^2 \quad (5.13)$$

De las ecuaciones (5.2) y (5.10) resulta directo –pero más largo– demostrar que el error de predicción de una variable estacionaria en diferencia es:

$$y_{t+s} - \hat{y}_{t+s|t} = \epsilon_{t+s} + (1 + \psi_1) \epsilon_{t+s-1} + (1 + \psi_1 + \psi_2) \epsilon_{t+s-2} + \dots \quad (5.14)$$

Nuevamente, si las innovaciones son i.i.d., el ECM de la predicción asociado es:

$$E[\hat{y}_{t+s|t} - y_{t+s}]^2 = [1 + (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + \dots] \sigma^2 \quad (5.15)$$

ahora el ECM aumenta con  $s$  pero no converge.

Notemos que, como el ECM crece linealmente con  $s$ , la desviación estándar de la predicción crece con  $\sqrt{s}$ . Si  $\delta > 0$ , la serie crece linealmente con  $s$ . Así, un intervalo de confianza estándar de la predicción hecho al 95% crece más lento que la serie.<sup>5</sup> La conclusión es que:

- una serie integrada va a crecer en el largo plazo.
- la tendencia estocástica domina la variabilidad proveniente de la raíz unitaria.

---

5 Any forecast can be made to look good by an appropriate measure, Hendry's Forecasting Theorem.

La figura siguiente grafica las dos características diferenciales entre los procesos estacionarios en tendencia y los procesos estacionarios en diferencias.

Figura 5.2a

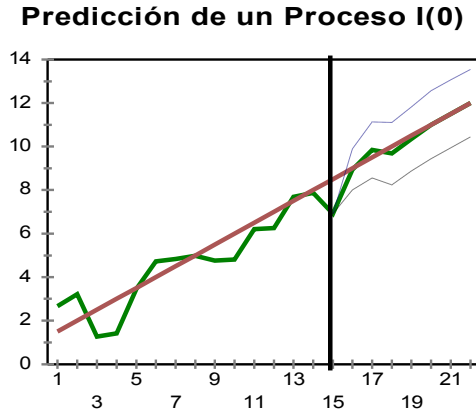
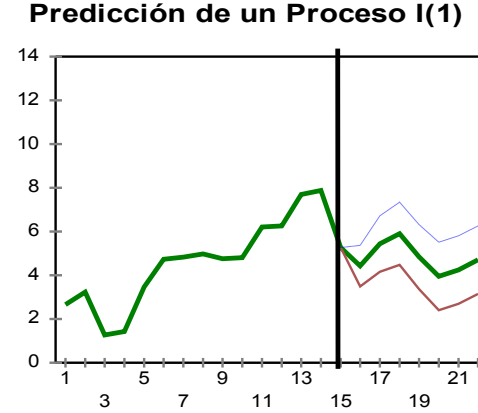


Figura 5.2b



## Persistencia de las Innovaciones

Otra manera de entender la diferencia entre procesos integrados y estacionarios es en términos de la persistencia de sus innovaciones. Como se desprende de la ecuación (5.8), en una variable estacionaria en tendencia los multiplicadores dinámicos son:

$$\frac{\partial y_{t+s}}{\partial \epsilon_t} = \psi_s \quad (5.16)$$

El hecho que se cumpla sumabilidad absoluta implica que  $\lim_{s \rightarrow \infty} \frac{\partial y_{t+s}}{\partial \epsilon_t} = 0$ .

Por otro lado, en una variable estacionaria en diferencia los multiplicadores dinámicos se derivan de la ecuación (5.10) y son:

$$\frac{\partial y_{t+s}}{\partial \epsilon_t} = \frac{\partial y_t}{\partial \epsilon_t} + \psi_s + \psi_{s-1} + \dots + \psi_1 = 1 + \psi_1 + \psi_2 + \dots + \psi_s \quad (5.17)$$

por lo tanto tiene efectos permanentes (note el valor 1 en la ecuación 5.17).

## Función de autocorrelación del proceso integrado

Resulta interesante obtener el correlograma del proceso I(1). Evidentemente, la varianza del proceso no es estacionaria pero puede ser representada por:

$$V(y_t) = V\left(\sum_i \epsilon_{t-i}\right) = t \sigma^2$$

La covarianza, por otro lado, puede ser obtenida como:

$$\begin{aligned} E[y_t, y_{t-k}] &= E\left[\sum \epsilon_t \sum \epsilon_{t-k}\right] \\ &= E[(\epsilon_1 + \epsilon_2 + \dots + \epsilon_{t-k} + \dots + \epsilon_t)(\epsilon_1 + \epsilon_2 + \dots + \epsilon_{t-k})] \\ &= E[\epsilon_1^2 + \epsilon_2^2 + \dots + \epsilon_{t-k}^2 + \epsilon_1 \epsilon_2 + \epsilon_1 \epsilon_3 + \dots] \\ &= (t-k) \sigma^2 \end{aligned}$$

Así, el correlograma es  $\rho_t = \sqrt{\frac{t-k}{t}}$ . En este caso, a diferencia de los procesos estacionarios, la función de autocorrelación decae muy lentamente, tomando valores cercanos a 1 incluso en intervalos de tiempo amplios. En concreto, ocurre que  $\lim_{t \rightarrow \infty} \rho_t \rightarrow 1$ . Este resultado produce una fuerte correlación de cada periodo con los anteriores. La representación gráfica de estos procesos es una serie que evoluciona suavemente, sin grandes altibajos. Además, puede producir localmente el efecto de una pendiente lineal, lo que puede llevar a concluir que la fuente de no estacionariedad es un componente determinista en lugar de una pendiente de naturaleza estocástica.

## Remoción de Tendencias

Como hemos visto, la forma que tome la tendencia de una variable –determinística o estocástica– implica notables diferencias para el proceso generador de los datos. No resulta sorprendente que la manera de remover las tendencias de las series de tiempo sea un tema central en el análisis de los modelos de series de tiempo. Consideremos la aplicación de lo que hemos aprendido:

- La manera correcta de quitar la tendencia en series estacionarias en tendencia es quitar la tendencia determinística. Esto se hace mediante el pre-filtrado de las series usando una regresión lineal de la variable contra una secuencia aritmética lineal.



- La manera correcta de remover la tendencia en series estacionarias en diferencias es tomar la diferencia de orden  $d$ , tal que  $(1-L)^d y_t$  sea estacionaria.

El problema radica en que resulta imposible saber si una variable aleatoria es  $I(0)$  o  $I(1)$  *a priori* o por observación de una muestra particular. ¿Qué sucede si nos equivocamos en el método de remoción de la tendencia?

- Si una variable estacionaria en diferencias es pre-filtrada para remover la tendencia determinística, la(s) raíz(es) unitarias(s) no ha(n) sido removida(s).
- Si diferenciamos una serie que es estacionaria en tendencia determinística obtendremos

$$(1-L) y_t = \delta + (1-L) \psi(L) \epsilon_t \quad (5.18)$$

la variable resultante tiene una representación que no es invertible. Un primer problema de una representación no invertible es que posiblemente ésta será no estacionaria. Un segundo problema es que los algoritmos de estimación requieren el uso de la representación invertible del proceso (ver Hamilton (1994) para los detalles).

Como puede verse, el problema de remoción de tendencia es importante. En realidad, la correcta caracterización de una variable como estacionaria o no estacionaria es fundamental por la presencia de dos fenómenos que estudiaremos un poco más adelante. En primer lugar, los estimadores clásicos (mínimos cuadrados o máxima verosimilitud) no conservan sus propiedades en modelos que incluyen variables no estacionarias, sean como regresores o variables de interés. Esto implica que los tests de hipótesis estarán distorsionados. En segundo lugar, y más importante, la presencia de variables integradas suele llevar al problema de correlación espuria, es decir, a obtener como resultado parámetros aparentemente significativos y altas bondades de ajuste cuando las variables no están efectivamente relacionadas entre sí.

Para derivar tests que nos permitan discriminar la naturaleza de las variables económicas estudiaremos primero las propiedades asintóticas de los estimadores clásicos de mínimos cuadrados cuando las variables son estacionarias y luego sus contrapartes cuando las variables son no estacionarias. De esta comparación se derivarán los principales resultados respecto del impacto del impacto de las variables integradas sobre los tests de hipótesis y se demostrará la necesidad de usar distribuciones particulares para los diferentes tests de raíces unitarias que se utilizan frecuentemente en la práctica.

## 5.02 Teoría Asintótica para Procesos con Tendencias Determinísticas

Hasta el momento sólo hemos señalado que los procesos estacionarios y no-estacionarios son de naturaleza muy distinta respecto de su representación, memoria, y predicciones. Pero, en realidad, las diferencias son mucho más profundas.

Consideremos un modelo simple:

$$y_t = \alpha + \delta z_t + \epsilon_t \quad (5.19)$$

si  $\epsilon_t$  es un proceso Gaussiano y  $z_t$  es una (o más) variable determinística, entonces los estimadores de mínimos cuadrados de los parámetros en una muestra de tamaño T son:

$$\hat{\beta}_T = \begin{bmatrix} \hat{\alpha}_T \\ \hat{\delta}_T \end{bmatrix} = \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \left[ \sum_{t=1}^T x_t y_t \right] \quad (5.20)$$

con  $x_t = [1 \ z_t]$ , de donde se desprende que

$$\hat{\beta}_T - \beta = \begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\delta}_T - \delta \end{bmatrix} = \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \left[ \sum_{t=1}^T x_t \epsilon_t \right] \quad (5.21)$$

La distribución límite se obtiene de manera directa

$$\sqrt{T} (\hat{\beta}_T - \beta) = \left[ \frac{1}{T} \sum_{t=1}^T x_t x_t' \right]^{-1} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t \epsilon_t \right] \quad (5.22)$$

- el primer término converge en probabilidad a  $Q^{-1}$  (condiciones de Grenander).
- el segundo término converge en distribución a  $N(0, \sigma^2 Q^{-1})$  por el teorema del límite central.<sup>6</sup>

Por lo que  $\sqrt{T} (\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$ .

---

<sup>6</sup> Un tratamiento simple de teoría asintótica para procesos estacionarios se encuentra en W. Greene *Análisis Económico* (1997), cap. 6.

Suponga ahora que  $z$  incluye una variable estacionaria alrededor de una tendencia determinística. La manera más simple de ver el efecto de una variable no estacionaria es suponer que hay una tendencia lineal  $z=t$ , entonces  $x=[1 \ t]$  y por lo tanto:

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\delta}_T - \delta \end{bmatrix} = \begin{bmatrix} \sum 1 & \sum t \\ \sum t & \sum t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \epsilon_t \\ \sum t \epsilon_t \end{bmatrix} \quad (5.23)$$

es directo demostrar que  $\sum_{t=1}^T t = \frac{T(T+1)}{2}$  en tanto que  $\sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$ .

Por lo tanto:

$$\sum_{t=1}^T x_t x_t' = \begin{bmatrix} \sum 1 & \sum t \\ \sum t & \sum t^2 \end{bmatrix}^{-1} = \begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix} \quad (5.24)$$

entonces cuando  $T \rightarrow \infty$ ,  $\frac{1}{T} \sum x_t x_t'$  diverge. De esta manera, no se cumple el supuesto que el primer término de la ecuación (5.22) converge en probabilidad a una matriz positiva definida ( $Q$ ).

Para que el denominador converja podemos pensar en usar alguna transformación del estimador ("transformación estabilizadora") que evite que la matriz de momentos de los regresores diverja. Por ejemplo, podríamos usar  $\frac{1}{T^3} \sum x_t x_t'$ . El problema es que tal transformación, si bien evita la divergencia, produce una matriz que no es invertible porque converge a  $Q = \begin{bmatrix} 0 & 0 \\ 0 & 1/3 \end{bmatrix}$ .

Otra cosa que resulta evidente de la inspección de la ecuación (5.23) es que  $\hat{\alpha}$  y  $\hat{\delta}$  convergen a sus parámetros poblacionales a distinta "tasa" en términos del tamaño de muestra.

Necesitamos una manera alternativa de tratar el problema. Para encontrar una distribución límite no degenerada, usamos una transformación distinta para cada estimador. En particular:

$$\Psi_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix} \quad (5.25)$$

para premultiplicar (transformar) la ecuación (5.21). Así los nuevos estimadores son:

$$\begin{aligned} \begin{bmatrix} \sqrt{T} (\hat{\alpha}_T - \alpha) \\ T^{3/2} (\hat{\delta}_T - \delta) \end{bmatrix} &= \Psi_T \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \begin{bmatrix} \sum_{t=1}^T x_t \epsilon_t \\ \sum_{t=1}^T x_t \epsilon_t \end{bmatrix} \\ &= \Psi_T^{-1} \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \Psi_T \Psi_T^{-1} \begin{bmatrix} \sum_{t=1}^T x_t \epsilon_t \\ \sum_{t=1}^T x_t \epsilon_t \end{bmatrix} \end{aligned} \quad (5.26)$$

Ahora estudiamos el inverso del primer término de la segunda igualdad:

$$\Psi_T \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \Psi_T = \left[ \Psi_T^{-1} \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \Psi_T^{-1} \right]^{-1} = \begin{bmatrix} T^{-1} \sum 1 & T^{-2} \sum t \\ T^{-2} \sum t & T^{-3} \sum t^2 \end{bmatrix} \quad (5.27)$$

que obviamente converge a  $Q = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}$ .

Por otro lado, al estudiar el segundo término de la ecuación (5.26) descubrimos que el teorema de límite central es aplicable, por lo que éste es asintóticamente normal:

$$\Psi_T^{-1} \begin{bmatrix} \sum_{t=1}^T x_t \epsilon_t \\ \sum_{t=1}^T x_t \epsilon_t \end{bmatrix} = \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{bmatrix} \begin{bmatrix} \sum \epsilon_t \\ \sum t \epsilon_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{T}} \sum \epsilon_t \\ \frac{1}{\sqrt{T}} \sum \left( \frac{t}{T} \right) \epsilon_t \end{bmatrix} \quad (5.28)$$

Finalmente de las ecuaciones (5.27) y (5.28) obtenemos que:

$$\begin{bmatrix} \sqrt{T} (\hat{\alpha}_T - \alpha) \\ T^{3/2} (\hat{\delta}_T - \delta) \end{bmatrix} \xrightarrow{d} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} \right) \quad (5.29)$$

el estimador de la constante converge como siempre que se trate de variables estacionarias a tasa  $\sqrt{T}$ , pero el estimador del parámetro de la variable con tendencia (determinística) es *superconsistente* porque converge a tasa  $T^{3/2}$ .

En general para referirnos al tipo de convergencia de las variables usamos el concepto de orden de probabilidad. Suponga que existen dos secuencias de números reales,  $a$  y  $b$ . La notación  $a_n = O(b_n)$  corresponde a “ $a$  es al menos del orden de  $b_n$ ”, en tanto que  $a_n = o(b_n)$  corresponde a “ $a$  es de un orden menor a  $b_n$ ”. Formalmente,

- $a_n = O(b_n)$  cuando  $n \rightarrow \infty$  si  $\lim_{n \rightarrow \infty} \left( \frac{|a_n|}{b_n} \right) < K$  y  $K > 0$  (5.30)
- $a_n = o(b_n)$  cuando  $n \rightarrow \infty$  si  $\lim_{n \rightarrow \infty} \left( \frac{|a_n|}{b_n} \right) = 0$

Por ejemplo, en la ecuación (5.29) el orden de probabilidad del estimador de  $\hat{\alpha}$  es  $O(T^{1/2})$  en tanto que el orden del estimador de la tendencia es  $O(T^{3/2})$ .

### 5.03 Teoría Asintótica para Procesos con Tendencias Estocásticas

Hemos aprendido que si las variables son no estacionarias, los estimadores de mínimos cuadrados del parámetro  $\rho$  no se distribuyen de acuerdo a la teoría asintótica convencional. Para encontrar cómo y a qué converge un proceso con tendencia estocástica debemos introducir dos conceptos adicionales: procesos Brownianos y el funcional del teorema de límite central. La razón de usar este tipo de matemática radica en que la existencia de procesos integrados nos obliga a trabajar con secuencias infinitas y con funciones de éstas.

Tomemos un modelo AR(1) con parámetro  $\rho$ , es decir,  $y_t = \rho y_{t-1} + \epsilon_t$  con  $y_0 = 0$  e innovaciones  $\epsilon_t \rightarrow N(0, \sigma^2)$ . El estimador de mínimos cuadrados de  $\rho$  en una muestra de tamaño  $T$  es:

$$\hat{\rho}_T = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} \quad (5.31)$$

donde hemos aprovechado el hecho que  $y_0$  es conocida e igual a cero. Se puede demostrar que si  $|\rho| < 1$ , entonces

$$\sqrt{T}(\hat{\rho}_T - \rho) \rightarrow N(0, (1 - \rho^2) \sigma^2) \quad (5.32)$$

Cuando  $\rho = 1$  entonces la ecuación (5.32) converge a una distribución degenerada con varianza cero, lo que es un resultado válido pero impide la inferencia.

Para encontrar la distribución asintótica no degenerada multiplicamos la ecuación (5.31) por  $T$  en vez de  $\sqrt{T}$ , para el caso en que el verdadero parámetro  $\rho$  es 1.

$$T(\hat{\rho}_t - 1) = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t}{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2} \quad (5.33)$$

Necesitamos determinar a qué converge el numerador y el denominador de la ecuación (5.33). Consideremos primero el numerador. Si  $\rho=1$ , entonces  $y_t$  es un proceso paseo aleatorio, es decir,  $y_t = y_{t-1} + \epsilon_t = y_{t-2} + \epsilon_t + \epsilon_{t-1} = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots$ , por lo que  $y_t \rightarrow N(0, \sigma^2 t)$ . Más aún, para este tipo de proceso se cumple que  $y_t^2 = (y_{t-1} + \epsilon_t)^2 = y_{t-1}^2 + 2y_{t-1} \epsilon_t + \epsilon_t^2$ , por lo que:

$$y_{t-1} \epsilon_t = \frac{1}{2} [y_t^2 - y_{t-1}^2 - \epsilon_t^2] \quad (5.34)$$

sumando para todo  $t$  obtenemos:

$$\sum_{t=1}^T y_{t-1} \epsilon_t = \frac{1}{2} [y_T^2 - y_0^2] - \frac{1}{2} \sum_{t=1}^T \epsilon_t^2 \quad (5.35)$$

pero  $y_0 = 0$ , por lo que:

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} \epsilon_t = (1/2)(1/T) y_T^2 - (1/2)(1/T) \sum_{t=1}^T \epsilon_t^2 \quad (5.36)$$

dividiendo la ecuación (5.36) por  $\sigma^2$  obtenemos

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T y_{t-1} \epsilon_t = (1/2) \left[ \frac{y_T}{\sigma \sqrt{T}} \right]^2 - (1/2)(1/\sigma^2 T) \sum_{t=1}^T \epsilon_t^2 \quad (5.37)$$

El primer término del lado derecho de la ecuación (5.37) es una  $N(0,1)$  al cuadrado, es decir  $\chi^2(1)$ . El segundo término es una suma de cuadrados de variables i.i.d.

con media cero y varianza  $\sigma^2$ , por lo que  $\frac{1}{T} \sum \epsilon_t^2 \xrightarrow{p} \sigma^2$ . Así, la ecuación (5.36) converge a  $\frac{1}{2}[\chi^2(1)-1]$ .

Ahora estudiamos el denominador de la ecuación (5.30). Como  $y_{t-1} \rightarrow N(0, \sigma^2(t-1))$ , entonces  $E[y_{t-1}^2] = \sigma^2(t-1)$  lo que implica  $E \sum_{t=1}^T [y_{t-1}^2] = \sigma^2 \sum_{t=1}^T (t-1) = \sigma^2(T-1) \frac{T}{2}$  para que converja tiene que ser dividido por  $1/T^2$ .

En resumen, si el verdadero proceso es un paseo aleatorio, la desviación del estimador de mínimos cuadrados de  $\rho$  respecto de su verdadero valor, es decir  $(\hat{\rho}_T - 1)$ , debe ser multiplicada por  $T$  (no  $\sqrt{T}$ ) de modo que tenga una distribución no degenerada.

Dicha distribución es, por lo demás, una razón entre una  $\chi^2(1)$  y una distribución no estándar. Para caracterizarla necesitamos dos elementos más de análisis: movimientos Brownianos y el Funcional del Teorema de Límite Central.

## 5.04 Movimientos Brownianos<sup>7</sup> y Procesos de Wiener

Tomemos un proceso tipo paseo aleatorio  $y_t = y_{t-1} + \epsilon_t$ . Consideremos la innovación  $\epsilon_t = y_t - y_{t-1}$  como proviniendo de una distribución  $N(0,1)$ . Entonces,  $y_t = \epsilon_t + \epsilon_{t-1} + \dots + \epsilon_1$  y por tanto,  $y_t \rightarrow N(0, t)$ . Consideremos  $y_t - y_s = \epsilon_{t+1} + \epsilon_{t+2} \dots + \epsilon_s$ , entonces se cumplen las siguientes propiedades:

- $[y_t - y_s] \rightarrow N(0, s-t)$
- $[y_t - y_s]$  es independiente de  $[y_r - y_q]$  si  $r < q < s < t$ .

<sup>7</sup> En 1827 el botánico escocés Robert Brown descubrió que los granos de polen suspendidos en agua se movían a brincos siguiendo un patrón de zigzag. El primero en observar este fenómeno fue Jan Ingenhousz en 1785 pero pasó desapercibido hasta que fué redescubierto por Brown (providencialmente, el olvido nos evitó hablar de “procesos Ingenhouszianos”). En 1889 G.L. Gouy descubrió que los movimientos Brownianos ocurren en partículas pequeñas (no en autos o personas) y son más rápidos mientras más pequeña es la partícula. La primera explicación del fenómeno es la de Desaulx (1877): *In my way of thinking, the phenomenon is a result of thermal molecular motion in the liquid environment (of the particles)*. Efectivamente, una partícula suspendida está constantemente sujeta a innovaciones aleatorias de todos lados por las moléculas del líquido. Si la partícula es relativamente pequeña, las innovaciones aleatorias harán saltar a la variable instantáneamente. El primer modelo matemático the *Brownian motion* fue desarrollado por A. Einstein en 1905 (por lo que recibió el premio Nobel).

Supongamos, ahora, que  $\varepsilon_t$  está compuesto de dos procesos Gaussianos independientes, donde cada uno es una realización de una distribución  $N(0, 1/2)$ :

$$\varepsilon_t = e_t^1 + e_t^2$$

donde podemos pensar que  $e_t^1$  corresponde a una parte de  $\Delta y$ , digamos  $e_t^1 = y_{t-1/2} - y_{t-1}$ . Obviamente,  $e_t^2 = y_t - y_{t-1/2}$ . Es decir, el incremento del paseo aleatorio podría ser pensado como el incremento de medio  $\varepsilon_t$  primero y otro medio  $\varepsilon_t$  después, tal que cada “pedazo” es en sí un paseo aleatorio. Note que estos procesos  $e_t$  también verifican las propiedades descritas.

Si hacemos el mismo ejercicio fraccionando el intervalo en  $k$  trozos, se obtienen los mismos resultados. Si hacemos tender  $k \rightarrow \infty$ , obtenemos un proceso continuo, llamado proceso Browniano estándar (*standard Brownian motion*) o proceso de Wiener<sup>8</sup> y lo denominamos como  $W(t)$ . Podemos pensar que es un paseo aleatorio, pero haciendo el intervalo infinitamente pequeño.<sup>9</sup>

Una manera de visualizar qué es lo que implica este “fraccionamiento” se muestra en la Figura 5.2. En el panel izquierdo se ha graficado un paseo aleatorio y se ha cortado la curva en ocho trozos. Nótese que cada trozo es de igual ancho, pero de distinto alto. En el panel derecho se ha cortado la curva en 24 trozos. Como resulta evidente, el ajuste en este caso es muy superior. Si hacemos un fraccionamiento infinitamente angosto, entonces podremos replicar la evolución de la curva completamente y habremos pasado de tiempo discreto a continuo.

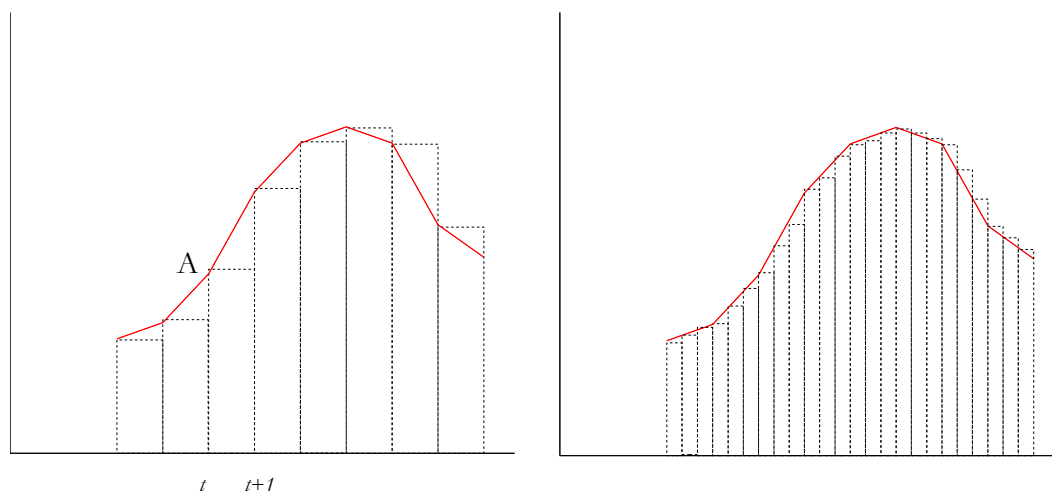
---

8 Norbert Wiener fue un matemático norteamericano considerado el padre de la cibernética. A la edad de 11 años entró a la universidad a estudiar matemáticas, recibiendo su Ph.D en lógica matemática a la edad de 18 años.

9 Este análisis recuerda la célebre paradoja de Zenón de Elea (490 AC- 430 AC) según la cual el más rápido de los hombres, Aquiles, no podrá alcanzar nunca al más lento de los animales, la tortuga, si se da a ésta una ventaja inicial en una carrera. Pues, mientras Aquiles recorre el camino que la tortuga llevaba por la mencionada ventaja inicial, la tortuga habrá recorrido otra porción, aunque más pequeña. Cuando Aquiles haya llegado a recorrer esta última porción de camino, la tortuga habrá avanzado otra porción más pequeña, y así la tortuga llevará siempre la ventaja hasta en espacios infinitamente pequeños, con lo cual, Aquiles no podrá alcanzarla nunca. Zenón concluía que al sumar infinitos números se obtendría infinito, sin importar el tipo de números que se sume. Sin embargo, al reducir la suma a “límites de secuencias” se observa que este argumento es falso en el caso de la paradoja.



**Figura 5.2**  
**Paseo aleatorio y función peldaño**



Este proceso tiene tres características. Primero, es un proceso Markoviano. Es decir, la historia futura sólo depende la posición actual de  $y$ , pero no del pasado. Segundo, los incrementos son independientes; en principio, ellos pueden ser i.i.d. Tercero, los incrementos se distribuyen normal.

En términos de la Figura 5.2, considere que la variable se ubica en el punto  $A$  en el instante  $t$ . Su ubicación en el instante  $t+1$  sólo dependerá de su posición inicial (el punto  $A$ ) y de la magnitud de la innovación que recibe en el instante  $t+1$ . Que el incremento sea normal sólo dice que, *sobre  $t+1$* , los valores más alejados de la altura de  $A$  son menos probables.

Matemáticamente, se puede escribir el proceso de Wiener como:

$$\text{si } y_t \text{ es } W(t) \text{ entonces } \Delta y_t = \epsilon \sqrt{\Delta t}$$

Las siguientes características de este proceso son importantes:

- Cuando  $\Delta t \rightarrow 0$ , el proceso es  $dy = \epsilon \sqrt{dt}$ .
- Este proceso continuo no tiene derivada porque  $dy/dt = \epsilon dt^{1/2}$ .<sup>10</sup>

<sup>10</sup> Si este tipo de funciones lo deja estupefacto, usted no está solo. El famoso matemático francés Charles Hermite señaló en 1893 *I recoil in fear and loathing from that deplorable evil: continuous functions with no derivatives*.

- Nótese que  $E[dy]=0$  y  $V[dy]=dt$ .
- Sea  $z1(t) = \sigma W(t)$ . Entonces esta transformación del proceso original es también un movimiento Browniano, sólo que sus incrementos son i.i.d.  $N(0, \sigma^2)$  en vez de  $N(0,1)$ .

De esta manera hemos pasado de procesos discretos a continuos. Los movimientos Brownianos son procesos de Wiener, pues éstos se definen como un proceso estocástico Gaussiano en tiempo continuo  $\{W_t\}$  que satisface las siguientes propiedades (1)  $W_t - W_s \rightarrow N[\mu, \sigma^2(t-s)]$  y (2)  $(W_t - W_s)$  y  $(W_u - W_v)$  son variables aleatorias independientes.<sup>11</sup>

## 5.05 Funcional del Teorema del Límite Central

Para identificar las distribuciones límites de procesos estocásticos que son funciones acumuladas de perturbaciones de orden potencialmente infinito necesitamos una generalización del teorema de límite central que se aplique a funciones y no a variables aleatorias. Recordemos la versión más simple del Teorema del Límite Central (TLC): Si  $\mu_t$  es i.i.d. con media 0 y varianza  $\sigma^2$ , la media muestral satisface:

$$\sqrt{T} \bar{\mu}_T \rightarrow N(0, \sigma^2) \quad (5.39)$$

Es importante notar que el TLC se refiere a la distribución de la media muestral que, en el caso de procesos estacionarios es constante. En el caso de procesos no estacionarios, la media va cambiando en el tiempo, por lo que se enfrenta el problema de descubrir la distribución límite de una secuencia de medias, no de una única media. La solución consiste en mapear las medias de su intervalo creciente  $[0, T]$  a un intervalo fijo  $[0, 1]$ .

Supongamos un estimador curioso: “calcule la media muestral con la mitad de los datos, bote el resto”. El estimador  $\bar{\mu}_{T/2}$  también cumple el TLC. Y, de hecho, es independiente del mismo tipo de estimador hecho con la otra mitad de los datos.

Podemos generalizar esto para la fracción  $r$  de los datos, con  $r \in [0, 1]$  para definir una función tipo pedazo (*step*), es decir, la “media calculada con una fracción  $r$  de la muestra de  $T$  datos”:

<sup>11</sup> Otros procesos de Wiener son el proceso de Ito  $dx = a(x, t)dt + b(x, t)dW$  donde los parámetros son función de la posición de la serie y el Wiener geométrico  $\Delta x/x = a \Delta t + b \Delta W$  para tasas de crecimiento.

$$X_T(r) = \frac{1}{T} \sum_{i=1}^{Tr} \mu_i \quad (5.40)$$

Nótese que la función  $X_T(r)$  es una función en  $r$ , que mapea la información en el intervalo  $[0,T]$  en la recta  $[0,1]$  de la siguiente forma:

$$X_T(r) = \begin{bmatrix} 0 & 0 \leq r < \frac{1}{T} \\ \frac{\mu_1}{T} & \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{(\mu_1 + \mu_2)}{T} & \frac{2}{T} \leq r < \frac{3}{T} \\ \vdots & \vdots \\ \frac{(\mu_1 + \mu_2 + \dots + \mu_T)}{T} & r = 1 \end{bmatrix} \quad (5.41)$$

entonces, aún cuando la muestra va creciendo de tamaño ( $T \rightarrow \infty$ ),  $X_T$  continua definida en el intervalo  $[0,1]$ . Así

$$\sqrt{T} X_T(r) = \frac{1}{\sqrt{T}} \sum_{i=1}^{Tr} \mu_i = \frac{\sqrt{Tr}}{\sqrt{T}} \left[ \frac{1}{\sqrt{Tr}} \sum_{i=1}^{Tr} \mu_i \right] \quad (5.42)$$

aplicando el TLC al paréntesis de la ecuación (5.42) se obtiene que éste converge a  $N(0, \sigma^2)$ , en tanto que la fracción converge a  $\sqrt{r}$ .<sup>12</sup> Así:

$$\sqrt{T} X_T(r) \rightarrow N(0, r \sigma^2) \quad (5.43)$$

y, por lo tanto,

$$\sqrt{T} \frac{X_T(r)}{\sigma} \rightarrow N(0, r) \quad (5.44)$$

es decir, converge a un proceso Browniano.<sup>13</sup> Este resultado es llamado teorema del funcional del límite central. Consideremos ahora, la diferencia entre las ecuaciones (5.39) y (5.44). La ecuación (5.39) es para variables aleatorias, en tanto que la ecuación (5.44) es

<sup>12</sup> Note que la fracción es una razón de desviaciones estándares.

<sup>13</sup> Este resultado utiliza el Teorema de Donsker (1951).

para funciones de variables aleatorias. Nótese que la ecuación (5.39) es un caso particular de la ecuación (5.44) cuando  $r=1$ .

Este teorema del funcional del límite central se puede usar para obtener “convergencia en probabilidad” de funciones, etc. En particular, vamos a usar estos resultados para obtener las distribuciones de los distintos tests de raíces unitarias.

Consideremos de nuevo el paseo aleatorio  $y_t = y_{t-1} + \epsilon_t$ . Si  $y_0 = 0$ , entonces  $y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} \dots$ . Definimos la función estocástica:

$$X_T(r) = \begin{bmatrix} 0 & 0 \leq r < 1/T \\ y_1/T & 1/T \leq r < 2/T \\ y_2/T & 2/T \leq r < 3/T \\ \vdots & \vdots \\ y_T/T & r = 1 \end{bmatrix} \quad (5.45)$$

la integral de esta función es

$$\int_0^1 X_T(r) dr = \frac{y_1}{T^2} + \frac{y_2}{T^2} + \dots + \frac{y_T}{T^2} \quad (5.46)$$

Prueba: La función peldaño que definimos en la ecuación (5.45) señala que los incrementos en  $r$  son del tipo  $y_i/T$  y ocurren a distancia  $1/T$ . Entonces,

$$\int_0^1 X_T(r) dr = \sum_{i=1}^T \int_{(i-1)/T}^{i/T} X_T(r) dr, \text{ pero en el intervalo } [(i-1)/T, i/T] \text{ la función } X_T(r) \text{ no depende de } r. \text{ Así, } \sum_{i=1}^T \int_{(i-1)/T}^{i/T} X_T(r) dr = \sum_{i=1}^T \frac{y_i}{T} \int_{(i-1)/T}^{i/T} dr. \text{ Evaluamos la integral que, trivialmente, es } 1/T. \text{ Por lo tanto, } \int_0^1 X_T(r) dr = \frac{1}{T} \sum_{i=1}^T \frac{y_i}{T}. \blacksquare$$

Multiplicando por  $\sqrt{T}$  a cada lado la ecuación (5.46) obtenemos

$$\int_0^1 \sqrt{T} X_T(r) dr = T^{-3/2} \sum_{i=1}^T y_{i-1} \quad (5.47)$$

pero, ya hemos visto en la ecuación (5.44) que el lado izquierdo converge a  $\sigma \int_0^1 W(r) dr$ . Por ello,

$$T^{-\frac{3}{2}} \sum_{t=1}^T y_{t-1} \rightarrow \sigma \int_0^1 W(r) dr \quad (5.48)$$

¿Cuál es la intuición? Escribamos primero la ecuación anterior como:

$$\begin{aligned} T^{-\frac{3}{2}} \sum_{t=1}^T y_{t-1} &= T^{-\frac{3}{2}} [\epsilon_1 + (\epsilon_1 + \epsilon_2) + (\epsilon_1 + \epsilon_2 + \epsilon_3) + \dots] \\ &= T^{-\frac{3}{2}} [\epsilon_1(T-1) + \epsilon_2(T-2) + \epsilon_3(T-3) + \dots] \\ &= T^{-\frac{3}{2}} \sum_{t=1}^T (T-t) \epsilon_t = T^{-\frac{1}{2}} \sum_{t=1}^T \epsilon_t - T^{-\frac{3}{2}} \sum_{t=1}^T t \epsilon_t \end{aligned} \quad (5.49)$$

Usemos ahora un resultado que ya vimos en la ecuación (5.29):

$$\begin{bmatrix} T^{-1/2} \sum_{t=1}^T \epsilon_t \\ T^{-3/2} \sum_{t=1}^T t \epsilon_t \end{bmatrix} \rightarrow N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} \right) \quad (5.50)$$

entonces, lo que la ecuación (5.50) dice es que  $T^{-3/2} \sum_{t=1}^T y_{t-1}$  es asintóticamente Gaussiano con media cero y varianza,  $\sigma^2 [1 - 2 \cdot 1/2 + 1/3] = \sigma^2/3$ . Naturalmente,  $\sigma \int_0^1 W(r) dr$  en la ecuación (5.48) describe una variable aleatoria paseo aleatorio con distribución  $N(0, \sigma^2/3)$ .

En resumen, si  $y_t$  es un paseo aleatorio sin deriva, la media muestral diverge, pero  $T^{-3/2} \sum_{t=1}^T y_t$  converge a una variable aleatoria que se describe por la integral de un proceso Browniano con varianza  $\sigma^2$ . Hay un número grande de resultados adicionales que se derivan de este tipo de análisis para funciones de  $y_t$  que usted puede estudiar por su cuenta. Un resumen práctico de los principales resultados se encuentra en el Cuadro 5.1.

Sobre la base de los resultados obtenidos en el Cuadro 5.1 vamos a derivar la distribución del test de raíces unitarias para distintos tipos de procesos estocásticos. Es decir, vamos a estudiar qué le sucede al test  $t$  hecho sobre los estimadores de los parámetros de una regresión del tipo  $y_t = \alpha + \beta t + \rho y_{t-1} + \epsilon_t$  cuando el verdadero proceso que genera  $y_t$  tiene o no tiene una raíz unitaria.

Para derivar los resultados de convergencia del Cuadro 5.1 se utiliza el Teorema del Mapeo de Continuos (*cointinuous mapping theorem*)<sup>14</sup>, que señala que si  $f(\cdot)$  es un funcional continuo en el intervalo  $[0,1]$ , entonces:

$$f(X_T(r)) \rightarrow f(W(r))$$

En este sentido, el “mapeo” toma la siguiente forma:

$$\Sigma \Rightarrow \int$$

$$t \Rightarrow r$$

$$\epsilon_t \Rightarrow dW(r)$$

$$y_t \Rightarrow W(r)$$

---

14 Para una demostración ver P. Billingsley (1968): *Probability and Measure*, John Wiley and Sons.

**Cuadro 5.1**  
**Convergencia en procesos estocásticos**

Modelo:  $y_t = y_{t-1} + \epsilon_t$

Variable	Converge a
$T^{\frac{-1}{2}} \sum_{t=1}^T \epsilon_t$	$\sigma \cdot W(1)$
$T^{-5/2} \sum_{t=1}^T t y_{t-1}$	$\sigma \int_0^1 r W(r) dr$
$T^{-1} \sum_{t=1}^T \epsilon_t y_{t-1}$	$\frac{1}{2} \sigma^2 \cdot [W(1)^2 - 1]$
$T^{-2} \sum_{t=1}^T y_{t-1}^2$	$\sigma^2 \int_0^1 [W(r)]^2 dr$
$T^{\frac{-3}{2}} \sum_{t=1}^T y_{t-1}$	$\sigma \int_0^1 [W(r)]^2 dr$
$T^{\frac{-3}{2}} \sum_{t=1}^T t \epsilon_t$	$\sigma W(1) - \sigma \int_0^1 W(r) dr$
$T^{-3} \sum_{t=1}^T t y_{t-1}^2$	$\sigma \int_0^1 r [W(r)]^2 dr$
$T^{\nu+1} \sum_{t=1}^T t^\nu$	$\frac{1}{1+\nu}$

**Caso 1. Modelo sin constante ni tendencia determinística y el verdadero proceso es un paseo aleatorio.**

Supongamos que hacemos un AR(1),  $y_t = \rho y_{t-1} + \epsilon_t$ , y estimamos  $\rho$  por mínimos cuadrados ordinarios. ¿Qué pasa con el estimador mínimos cuadrados ordinarios de  $\rho$  si el verdadero  $\rho$  es 1?

$$\hat{\rho}_T = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} \quad (5.51)$$

y se desvía del verdadero valor

$$T(\hat{\rho}_T - 1) = \frac{\frac{1}{T} \sum y_{t-1} \epsilon_t}{\frac{1}{T^2} \sum y_{t-1}^2} \quad (5.52)$$

El numerador de la ecuación (5.52) converge a  $1/2 \sigma^2 \cdot [\mathcal{W}(1)^2 - 1]$  en tanto que el denominador converge a  $\sigma^2 \cdot \int_0^1 [\mathcal{W}(r)]^2 dr$ . Este último es un valor dado que no interviene de manera importante en la distribución del estimador  $\hat{\rho}$ . La distribución de  $T(\hat{\rho}_T - 1)$  y por ende la distribución de  $\hat{\rho}$ , depende de  $\mathcal{W}(1)^2$ . Como  $\mathcal{W}(1)$  es  $N(0,1)$ , entonces  $\mathcal{W}(1)^2$  es  $\chi^2(1)$ . Esto implica dos resultados interesantes:

1. La probabilidad de obtener un valor menor que 1 en una  $\chi^2(1)$  es de 68%. Es decir, en un 68% de muestras generadas por un paseo aleatorio, el  $\rho$  estimado por MCO será menor que 1 aún si el verdadero proceso tiene  $\rho = 1$ .
2. La distribución no es simétrica, sino que está sesgada hacia la izquierda.

Por ello, el test “t” sobre el estimador de  $\rho$  no es adecuado y se debe usar  $T(\hat{\rho}_T - 1)$ . Además, hay que usar valores críticos derivados de tablas especiales (obtenidas por Monte Carlo o exactas por simulación).

El estimador de mínimos cuadrados ordinarios de  $\rho$  es, sin embargo, superconsistente.<sup>15</sup>

---

<sup>15</sup> Ver demostración en Hamilton, capítulo 15.



Otro test común usa el estadígrafo  $t$ .

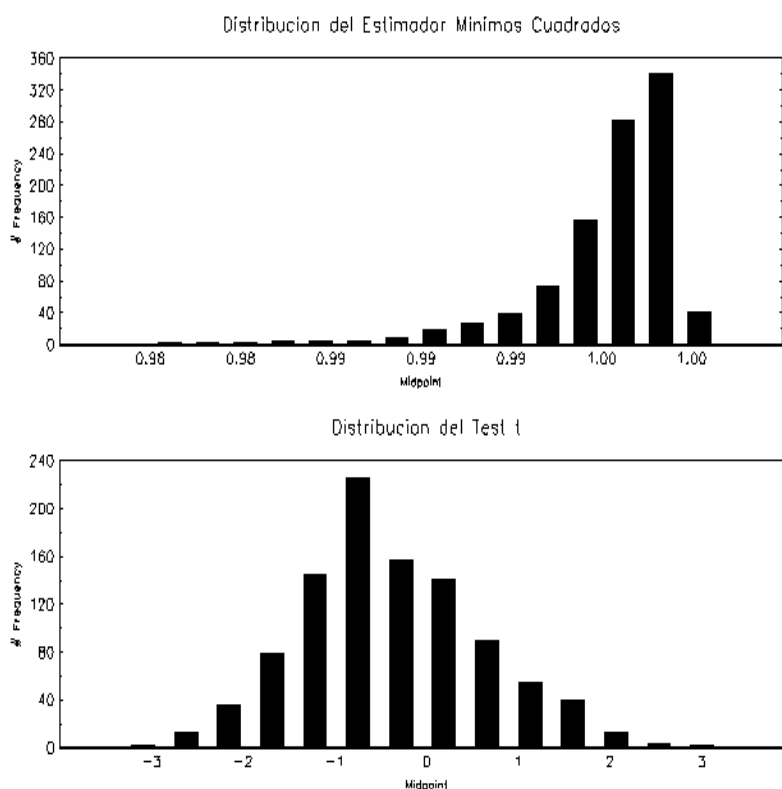
$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{T^{-1} \sum_{i=1}^T y_{t-1} \epsilon_t}{\left[ T^{-2} \sum_{i=1}^T y_{t-1}^2 \right]^{1/2} (s_T^2)^{1/2}} \quad (5.53)$$

Cuya distribución límite cuando  $T \rightarrow \infty$  es:

$$t_T \rightarrow \frac{1/2 \cdot [W(1)^2 - 1]}{\left[ \int_0^1 [W(r)]^2 dr \right]^{1/2}} \quad (5.54)$$

La Figura 5.3 muestra empíricamente cómo se distribuye el estimador de mínimos cuadrados y cómo se distorsiona el test "t".

**Figura 5.3**  
**Distribución del estimador de  $\rho$  y test  $t$**   
**cuando se estima un AR(1) y el verdadero modelo es paseo aleatorio**  
 Tamaño de Muestra = 1000, Número de Replicaciones = 1000



**Caso 2. Modelo con constante pero sin tendencia determinística y el verdadero proceso es un paseo aleatorio.**

Supongamos que hacemos un AR(1),  $y_t = \alpha + \rho y_{t-1} + \epsilon_t$ , y estimamos  $\rho$  por mínimos cuadrados ordinarios. Queremos describir,

$$\hat{\beta}_T = \begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix} = \begin{bmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{bmatrix} \quad (5.55)$$

La hipótesis nula es  $H_0: \alpha=0, \rho=1$ , entonces aplicamos (5.25) y obtenemos

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{bmatrix} \quad (5.56)$$

Se puede demostrar que:

$$T(\hat{\rho}_T - 1) \rightarrow \frac{1/2 \cdot [W(1)^2 - 1] - W(1) \int_0^1 W(r) dr}{\int_0^1 [W(r)]^2 dr - \left( \int_0^1 W(r) dr \right)^2} \quad (5.57)$$

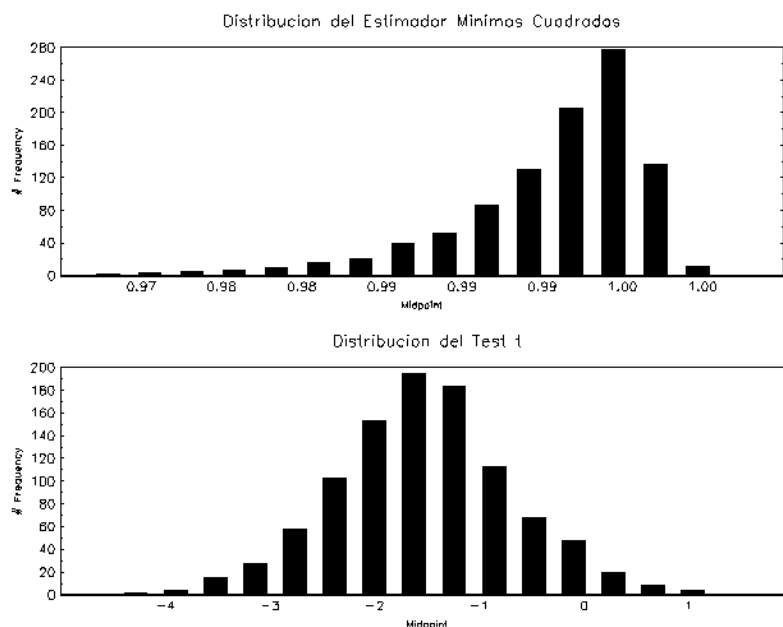
es decir, el estimador mínimos cuadrados ordinarios no tiene como distribución límite un proceso Gaussiano. Más aún, tampoco se distribuye como el caso 1. Por ello, no se puede usar los valores críticos de la tabla anterior, sino una nueva tabla. La distribución del test es aún más *skewed* hacia la izquierda.

Dickey y Fuller (1979) proponen un test tipo  $t$  como ya vimos. Este test se distribuye asintóticamente como:

$$\frac{T(\hat{\rho}_T - 1)}{T \hat{\sigma}_{\rho_T}} \rightarrow \frac{1/2 \cdot [W(1)^2 - 1] - W(1) \int_0^1 W(r) dr}{\left[ \int_0^1 [W(r)]^2 dr - \left( \int_0^1 W(r) dr \right)^2 \right]^{1/2}} \quad (5.58)$$

**Figura 5.4**  
**Distribución del estimador de  $\rho$  y test  $t$  cuando se estima un AR(1) con constante y el verdadero modelo es paseo aleatorio**

Tamaño de Muestra = 1000, Número de Replicaciones = 1000



**Caso 3. Modelo con constante pero sin tendencia determinística y el verdadero proceso es un paseo aleatorio con tendencia estocástica.**

Supongamos que el verdadero proceso es  $y_t = \alpha + y_{t-1} + \epsilon_t$ , y estimamos  $\rho$  en  $y_t = \alpha + \rho y_{t-1} + \epsilon_t$  por mínimos cuadrados ordinarios cuando  $\alpha$  no es cero. Demostraremos que la distribución de los estimadores de  $\alpha$  y  $\rho$  cambia bastante con respecto a los casos anteriores

El proceso puede ser descrito como

$$\begin{aligned} y_t &= \alpha t + y_0 + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots \\ &= \alpha t + y_0 + \psi_t \end{aligned} \tag{5.59}$$

con  $\psi = \sum \epsilon$ .

Consideremos  $\sum_{t=1}^T y_{t-1} = \sum_{t=1}^T [y_0 + \alpha(t-1) + \psi_{t-1}]$  el primer término es  $T y_0$  y converge si se divide por  $T$ ; el segundo converge a  $\alpha/2$  si se divide por  $T^2$ , en tanto que el tercero converge a  $\sigma \int W(r) dr$  si se divide por  $T^{3/2}$ .

Entonces, el segundo término domina asintóticamente a los otros dos. Así:

$$\frac{1}{T^2} \sum_{t=1}^T y_{t-1} = \frac{1}{T} y_0 + \frac{1}{T^2} \sum_{t=1}^T \alpha(t-1) + \frac{1}{T^2} \sum_{t=1}^T \psi_{t-1} \rightarrow 0 + \frac{\alpha}{2} + 0 \quad (5.60)$$

Estudiamos ahora  $\sum_{t=1}^T y_{t-1}^2$

$$\begin{aligned} \sum y_{t-1}^2 &= \sum y_0^2 + \sum \alpha^2(t-1)^2 + \sum \psi_{t-1}^2 \\ &+ \sum 2 y_0 \alpha(t-1) + \sum 2 y_0 \alpha \psi_{t-1} + \sum 2 \alpha(t-1) \psi_{t-1} \end{aligned} \quad (5.61)$$

el orden de probabilidad es  $O(T)$ ,  $O(T^3)$ ,  $O(T^2)$ ,  $O(T^2)$ ,  $O(T^{3/2})$  y  $O(T^{5/2})$  respectivamente. Por ello,

$$\frac{1}{T^3} \sum_{t=1}^T y_{t-1}^2 \rightarrow \frac{\alpha^2}{3} \quad (5.62)$$

Con el mismo raciocinio se deriva que

$$\frac{1}{T^{3/2}} \sum_{t=1}^T y_{t-1} \epsilon_t \rightarrow \frac{1}{T^{3/2}} \sum_{t=1}^T \alpha(t-1) \epsilon_t \quad (5.63)$$

es decir

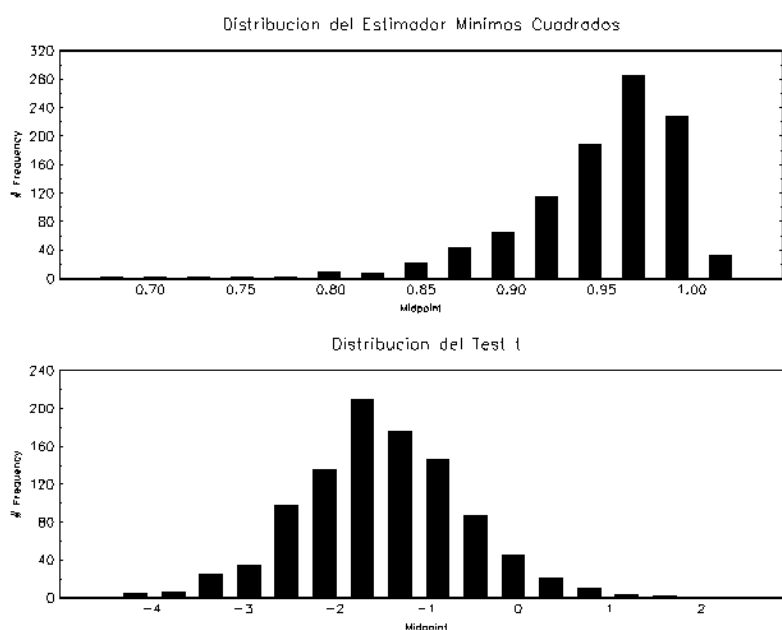
$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} O(T) & O(T^2) \\ O(T^2) & O(T^3) \end{bmatrix}^{-1} \begin{bmatrix} O(T^{1/2}) \\ O(T^{3/2}) \end{bmatrix} \quad (5.64)$$

ahora basta con definir  $\Psi_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix}$  para obtener que:

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} \rightarrow N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha/3 \end{bmatrix} \right) \quad (5.65)$$

es decir, obtuvimos lo mismo que en el caso del modelo con tendencia determinística. Este es un resultado notable. Por lo tanto se aplican los tests  $t$  y  $F$  directamente.

**Figura 5.5**  
**Distribución del estimador de  $\rho$  y test  $t$  cuando se estima un AR(1) con constante y el verdadero modelo es paseo aleatorio con tendencia aleatoria**  
 Tamaño de Muestra = 1000, Número de Replicaciones = 1000



**Caso 4. Modelo con constante y tendencia determinística y el verdadero proceso es un paseo aleatorio con tendencia estocástica.**

Se puede demostrar que el proceso  $y_t = \alpha + \rho y_{t-1} + \delta t + \epsilon_t$  se puede escribir como

$$\begin{aligned} y_t &= (1-\rho)\alpha + \rho[y_{t-1} - \alpha(t-1)] + (\delta + \rho\alpha)t + \epsilon_t \\ &= \alpha^* + \rho^* \psi_{t-1} + \delta^* t + \epsilon_t \end{aligned} \quad (5.66)$$

Bajo la hipótesis nula  $\psi_t = y_0 + \epsilon_1 + \epsilon_2 + \dots$ . Por lo tanto, el estimador mínimos cuadrados ordinarios es:

$$\begin{bmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* - 1 \\ \hat{\delta}_T^* - \delta_0 \end{bmatrix} = \begin{bmatrix} T & \sum \psi_{t-1} & \sum t \\ \sum \psi_{t-1} & \sum \psi_{t-1}^2 & \sum \psi_{t-1} t \\ \sum t & \sum t \psi_{t-1} & \sum t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \epsilon_t \\ \sum \psi_{t-1} \epsilon_t \\ \sum t \epsilon_t \end{bmatrix} \quad (5.67)$$

nótese que redefinimos la hipótesis nula del test. Entonces

$$\Psi_T = \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix} \quad (5.68)$$

Se obtiene la distribución asintótica de

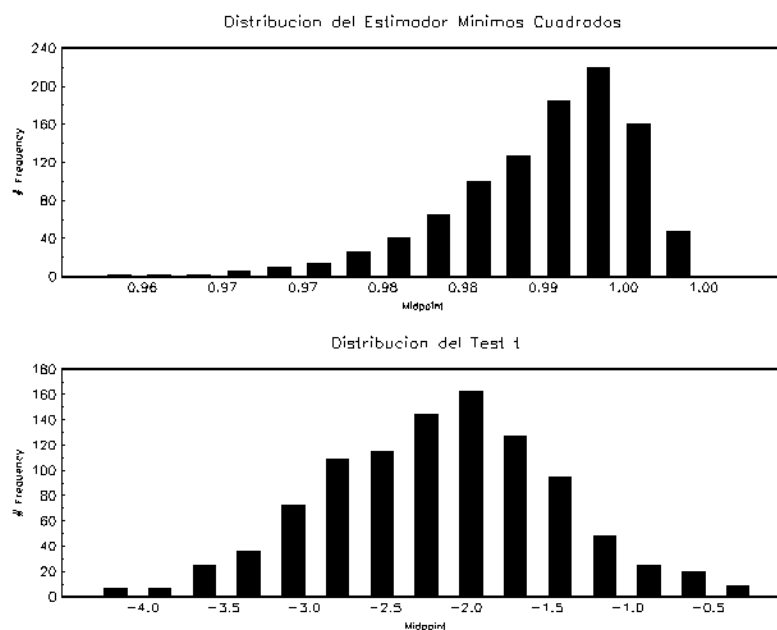
$$\begin{bmatrix} T^{1/2} \hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \delta_0) \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \int W(r) dr & 1/2 \\ \int W(r) dr & \int W(r)^2 dr & \int rW(r) dr \\ 1/2 & \int rW(r) dr & 1/3 \end{bmatrix}^{-1} \cdot \begin{bmatrix} W(1) \\ 1/2(W(1)^2 - 1) \\ W(1) - \int W(r) dr \end{bmatrix} \quad (5.69)$$

La distribución de  $T(\hat{\rho}_T - 1)$  no depende de  $\alpha$  o  $\sigma$ , en particular no depende de que  $\alpha = 0$ . La distribución de  $(\hat{\rho}_T - 1)/\sigma_{\hat{\rho}_T}$  es más ancha que todas las anteriores.

Figura 5.5

Distribución del estimador de  $\rho$  y test  $t$  cuando se estima un AR(1) con tendencia determinística y el verdadero modelo es paseo aleatorio con tendencia aleatoria

Tamaño de Muestra = 1000, Número de Replicaciones = 1000



## 5.06 Tests Empíricos Estándares de Raíces Unitarias

Existe una enorme variedad de tests de raíces unitarias, los que difieren en enfoque, poder y robustez frente a desvíos en las condiciones supuestas para derivarlos. En esta sección nos ocupamos de los tests paramétricos de raíces unitarias, en tanto que en una sección posterior estudiaremos tests no paramétricos.

Para el modelo  $y_t = \alpha + \rho y_{t-1} + \epsilon_t$  la hipótesis nula de una raíz unitaria sería  $H_0: \rho = 1$ . Un test de Wald de dicha hipótesis exigiría computar  $(\hat{\rho} - 1)/\sigma_{\hat{\rho}}$ . Resulta conveniente modificar ligeramente la hipótesis nula de modo que se pueda leer el resultado del test directamente de la regresión (todos los programas econométricos entregan la razón entre el parámetro estimado y su desviación estándar). Para ello, restamos  $y_{t-1}$  de ambos lados del modelo anterior para obtener:  $\Delta y_{t-1} = (\rho - 1)y_{t-1} + \epsilon_t$ . Ahora la hipótesis nula es simplemente  $H_0: \phi = 0$ , donde  $\phi = (\rho - 1)$ .

La hipótesis alternativa del test es que el modelo es estacionario. Como hay varias alternativas de modelos estacionarios, habrán varias especificaciones de tests de raíces unitarias que debemos considerar. En principio, cuando la variable no tiene una raíz unitaria, ésta podría oscilar alrededor de cero, alrededor de una constante, o alrededor de una tendencia determinística lineal. Cada una de estos casos es una distinta hipótesis alternativa.

### Familia de Tests de Dickey-Fuller<sup>16</sup>

Podemos anidar ambas hipótesis (nula y alternativa) en un solo modelo, como se describe en la ecuación (5.71). Habrá varias hipótesis alternativas, dependiendo del tipo de proceso que caracteriza a la serie cuando ésta es estacionaria. La especificación (a) corresponde al caso en que la hipótesis alternativa es que el modelo es autoregresivo -  $\rho$  será menor que 1- con media cero y sin tendencia determinística. En la especificación (b) la hipótesis alternativa corresponde a un proceso con media distinta de cero pero sin tendencia determinística. La última especificación es el modelo más general, en el sentido que permite que tanto la media como la tendencia determinística sea libres.

<sup>16</sup> Dickey, D.A. and W.A. Fuller (1979) "Distribution of Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427-431.



$$\begin{aligned}
(a) \Delta y_t &= \phi y_{t-1} + \epsilon_t \\
(b) \Delta y_t &= \alpha + \phi y_{t-1} + \epsilon_t \\
(c) \Delta y_t &= \alpha + \beta t + \phi y_{t-1} + \epsilon_t
\end{aligned}
\tag{5.70}$$

Como es la sana práctica en econometría, el testeo de raíces unitarias debe hacerse de “general a particular”. Es decir, se empieza por con el modelo (c) y se verifica que el estimador del parámetro  $\beta$  sea estadísticamente distinto de cero. Si no se rechaza dicha hipótesis nula, se estima el modelo (b), verificándose que el parámetro el estimador del  $\alpha$  sea cero. Si no se rechaza dicha hipótesis nula, se estima la especificación (a).

Nótese que de acuerdo a lo descrito en las secciones anteriores, la distribución de los estimadores de  $\alpha$  y  $\beta$  es estándar (asintóticamente normal), en tanto que la del estimador de  $\rho$  es no estándar. Además, al cambiar de hipótesis alternativa, cambia la forma de la distribución de este último debido a la presencia de parámetros incidentales. Así, habrá que construir tablas para cada tipo de test y nivel de significancia.

La subrutina del Apéndice B permite computar de manera sencilla los valores críticos, tomando en consideración los tres tipos de modelos y el tamaño de muestra. Los resultados están en el Cuadro 5.2. Como puede verse, el tamaño del test depende del número de datos de la muestra y de la presencia de parámetros incidentales (*nuisance parameters*) los que afectan la ubicación de la distribución Dickey-Fuller y el ancho de los intervalos de confianza del test sobre el parámetro de interés.

**Cuadro 5.2**  
**Valores críticos de la distribución Dickey Fuller**  
 Basado en 100.000 replicaciones

	<i>Tamaño Muestra</i>	<i>Valor crítico al 95%</i>	<i>Valor crítico al 90%</i>
Modelo (a)	50	-1,969	-1,621
	100	-1,950	-1,616
	1000	-1,953	-1,621
Modelo (b)	50	-2,936	-2,602
	100	-2,892	-2,581
	1000	-2,897	-2,589
Modelo (c)	50	-3,510	-3,187
	100	-3,460	-3,155
	1000	-3,452	-3,149

## Familia de Tests Dickey-Fuller Aumentado

Este test es exactamente igual al anterior pero controla por la posible autocorrelación de los residuos que en el caso anterior se supone -implícitamente- son ruido blanco. Como vimos en el Capítulo 3, la inclusión de rezagos de la variable endógena permite controlar la correlación serial.

$$\begin{aligned}
 (a) \Delta y_t &= \phi y_{t-1} + \epsilon_t + \sum_{p=1}^k \theta_p \Delta y_{t-p} \\
 (b) \Delta y_t &= \alpha + \phi y_{t-1} + \epsilon_t + \sum_{p=1}^k \theta_p \Delta y_{t-p} \\
 (c) \Delta y_t &= \alpha + \beta t + \phi y_{t-1} + \epsilon_t + \sum_{p=1}^k \theta_p \Delta y_{t-p}
 \end{aligned} \tag{5.71}$$

Nuevamente, la distribución de los estimadores de  $\alpha$ ,  $\beta$  y  $\theta$  se distribuyen asintóticamente normales, en tanto que el estimador de  $\rho$  se distribuye de manera no estándar. Del mismo modo, al cambiar de hipótesis alternativa, cambia la forma de la distribución de este último debido a la presencia de parámetros incidentales. McKinnon (1993), entre otros, provee tablas para cada tipo de test y nivel de significancia.

## Familia de Tests Phillips-Perron<sup>17</sup>

Phillips y Perron (1988) utilizan exactamente el material de convergencia que vimos anteriormente para desarrollar un test paramétrico de raíces unitarias para el caso que el residuo sigue cualquier tipo de correlación serial (es decir, no necesariamente  $AR(p)$ ).

Primero, si la primera diferencia de la serie es estacionaria,  $\mu_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ , y  $\epsilon_t$  es i.i.d.  $(0, \sigma^2)$ , entonces la serie puede ser descrita como:

$$\begin{aligned}
 y_t &= \mu_1 + \dots + \mu_t + y_0 \\
 &= \psi(1)(\epsilon_1 + \dots + \epsilon_t) + \eta_t - \eta_0 + y_0
 \end{aligned} \tag{5.72}$$

es decir, la combinación de un paseo aleatorio  $\psi(1)(\epsilon_1 + \dots + \epsilon_t)$ , condiciones iniciales  $\eta_0$  y  $y_0$  y un proceso estocástico,  $\eta_t$ .

---

17 Phillips, P.C.B. and P. Perron (1988). "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75: 335-346.

Esta decomposición (desarrollada por Beveridge y Nelson) se usa para obtener el equivalente de las ecuaciones (5.50), pero en el caso general de correlación serial.

Para hacer el test,

1. se estima un modelo AR(1) típico ( $y_t = \alpha + \rho y_{t-1} + \epsilon_t$ ) y se calculan los estimadores de  $\rho$  y  $\sigma$  y el error estándar de la regresión,  $s$ .
2. se estima la función de autocorrelación de los residuos muestrales, la que se usa para construir el estimador de Newey-West de la persistencia de la serie  $\lambda = \psi(1)\sigma$  y de la varianza de los residuos,  $\gamma_0$ .
3. se transforman los estimadores obtenidos en 1. para obtener tests que tienen la distribución del caso (2). Las transformaciones son:

$$[\gamma_0 \lambda^2] t_T - 1/2 (\lambda^2 - \gamma_0) / \lambda (T \sigma_{\rho_t} / s_T) \quad (5.73)$$

$$T(\hat{\rho}_T - 1) - 1/2 (T^2 \hat{\sigma} \rho_T^2 / s_T) (\lambda^2 - \gamma_0) \quad (5.74)$$

## Extensiones del Test de Dickey-Fuller con mayor poder

Elliott et al. (1996) sugiere usar una variación del estimador de mínimos cuadrados generalizados que consiste en pre-filtrar la variable de interés, es decir usa los residuos de la regresión de  $[y_1, y_2 - \alpha y_1, \dots, y_T - \alpha y_{T-1}]$  en  $[\tilde{z}_1, \tilde{z}_2 - \alpha \tilde{z}_1, \dots, \tilde{z}_T - \alpha \tilde{z}_{T-1}]$ , donde  $z_t$  puede incluir una constante o una constante y una tendencia determinística. Note que la variable es cuasi-diferenciada usando  $\alpha = 1 + \bar{c} T^{-1}$  donde  $\bar{c} \in (-\infty, 0)$  es un parámetro que se elige considerando el poder de la envolvente (típicamente -7 si sólo se incluye una constante en  $z_t$  y -13.5 si el test incluye constante y tendencia). Se hace un test de Dickey-Fuller como el que se describe en la ecuación (5.70a) sobre la variable cuasi-diferenciada.

Elliott (1999) señala que en el test anterior se supone que la primera observación es fija, siendo preferible suponer que hay estacionariedad en covarianzas. Por ello, si el modelo es de la forma  $y_t = \alpha \tilde{z}_t + \rho y_{t-1} + \epsilon_t$ , las desviaciones con respecto a la tendencia son una variable aleatoria con media cero y varianza no-condicional  $\sigma^2(1 - \rho^2)^{-1}$ . Así, el test se hará usando los residuos de la regresión de  $[(1 - \alpha^2)^{1/2} y_1, y_2 - \alpha y_1, \dots, y_T - \alpha y_{T-1}]$  en  $[(1 - \alpha^2)^{1/2} \tilde{z}_1, \tilde{z}_2 - \alpha \tilde{z}_1, \dots, \tilde{z}_T - \alpha \tilde{z}_{T-1}]$ .<sup>18</sup>

18 Otras extensiones del tipo DF-GLS son el test ERS (Elliott, Rothenberg and Stock (1996): "Efficient Tests

## Familia de Tests de Perron para quiebre estructural

Perron (1989) hace la observación que la hipótesis nula de los tests anteriores supone que el proceso determinístico no cambia. Pero puede haber quiebres estructurales en la media de la serie y/o en su tendencia, que pueden “engañar” al test de raíces unitarias, sesgando los resultados hacia la hipótesis nula de no estacionariedad.

Un test general considera la hipótesis alternativa con variables mudas (*dummies*) para capturar los posibles quiebres. El test con ambas hipótesis anidadas es:

$$\Delta y_t = \alpha + \alpha_1 D_t^1 + \beta t + \beta_1 D_t^2 t + \rho y_{t-1} + \epsilon_t + \sum_{p=1}^k \theta_p \Delta y_{t-p} \quad (5.75)$$

la distribución de los estimadores, en este caso, depende además de la ubicación del quiebre en la muestra. Por ello, hay tablas Perron para encontrar los valores críticos adecuados.

## Endogeneizando la ubicación del quiebre

Naturalmente, el test de Perron para quiebre estructural requiere conocer la fecha en que ocurre el quiebre. Ello no es común. Banerjee et al. (1992) y Zivot y Andrews (1992) proponen tests en que, aparte de determinar la presencia de raíces unitarias, se determina la existencia y fecha de quiebres estructurales.<sup>19</sup>

La lógica de estos tests es simple y consiste en hacer un “barrido” de la muestra con tests de raíces unitarias que van cambiando la ubicación de los quiebres. Como este procedimiento es efectivamente “data mining”, el test no tiene la distribución de Perron sino una distribución derivada *ad-hoc*, que precisamente reconoce el problema de data mining.

Zivot y Andrews (1992) observan que los cambios estructurales observados en ciertas muestras pueden ser eventos endógenos al DGP, por lo que el test de raíces unitarias debe incorporar quiebres endógenos. Por lo tanto, Zivot y Andrews consideran

---

for an Autoregressive Unit Root," *Econometrica* 64, 813-836) y el test Ng-Perron (Ng and Perron (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69(6), 1519-1555.)

19 Banerjee, A., R. Lumsdaine and J. Stock (1992) "Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence." *Journal of Business & Economic Statistics*, 10: 271-287. Zivot, E. and D. Andrews (1992) "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Journal of Business & Economic Statistics*, 10(3): 251-70.

una hipótesis nula que es un paseo aleatorio con drift sin cambio estructural, mientras que construyen la hipótesis alternativa tomando en consideración la existencia de quiebre en el intercepto, la tendencia o ambos. La ubicación del test se determina seleccionando la posición del mismo que entrega el test  $t$  menos favorable para la hipótesis nula. Los autores, a la Perron, utilizan tres especificaciones para el cambio estructural en la hipótesis alternativa:

$$\begin{aligned}
 \text{Caso A} \quad \Delta y_t &= \alpha^A + \alpha_1^A DU_t(\lambda) + \beta^A t + \rho^A y_{t-1} + \sum_{p=1}^k \theta_p^A \Delta y_{t-p} + \epsilon_t^A \\
 \text{Caso B} \quad \Delta y_t &= \alpha^B + \beta^B t + \beta_1^B DT_t(\lambda) + \rho^B y_{t-1} + \sum_{p=1}^k \theta_p^B \Delta y_{t-p} + \epsilon_t^B \\
 \text{Caso C} \quad \Delta y_t &= \alpha^C + \alpha_1^C DU_t(\lambda) + \beta^C t + \beta_1^C DT_t(\lambda) + \rho^C y_{t-1} + \sum_{p=1}^k \theta_p^C \Delta y_{t-p} + \epsilon_t^C
 \end{aligned} \tag{5.76}$$

donde  $\lambda$  es la posición del quiebre que surge del criterio antes mencionado, y  $k$  es el número de rezagos, el cual es determinado mediante un mecanismo que se discute más adelante. Así, el estadígrafo que se construye es  $\inf_{\lambda \in \Lambda} t_\alpha(\lambda)$ . Por supuesto, dicho estadígrafo tiene una distribución distinta al de Perron, pues es condicional a la ubicación del quiebre. Finalmente, los autores presentan tablas para las tres especificaciones antes descritas, las cuales utilizan para argumentar que los valores críticos utilizados por Perron son demasiado chicos, por lo que están sesgados hacia rechazar la nula.

## Determinación del número de rezagos

La determinación del número de rezagos ( $k$ ) es menos inocente de lo que parece. En principio bastaría con hacer que el residuo de la regresión que se hace para el test de raíces unitarias fuese ruido blanco. El problema radica en que el número de rezagos incluidos en la regresión afecta el poder del test. Alternativamente, el tamaño del test no es 5% o 10% si el número de rezagos es el incorrecto.

Una alternativa es fijar *ex-ante* el número de rezagos. Ng y Perron (1995) demuestran que para residuos que siguen un  $AR(p)$  el test tiene el tamaño correcto si el número de rezagos incluidos es mayor que  $p$ . Desafortunadamente, para residuos que siguen un  $MA(q)$  no hay una regla útil. Schwert (1989) propone usar una regla

dependiente del tamaño de muestra, del tipo  $k = \text{Trunc}(c(T/100)^{1/d})$  y sugiere usar  $c = 12$  y  $d = 5$ . El problema de este tipo de regla es que no hay una razón clara para esperar que esta regla sea óptima para todo  $\text{AR}(p)$  o  $\text{MA}(q)$ .

Es común fijar  $k$  usando criterios informacionales como Akaike o Schwartz. Alternativamente se puede partir de un número máximo de rezagos  $-k_{\max}$  y luego ir reduciendo el orden del rezago si el marginal no es estadísticamente significativo. De esta manera se procede de *general a particular*.

Ng y P.Perron (2001)<sup>20</sup> proponen dos estrategias para obtener mayor poder y tamaño que los tests anteriores:

1. Usar datos cuasidiferenciados empleando GLS, dado que se ha demostrado (en otros trabajos) que este tipo de cuasidiferenciación entrega ganancias de poder.
2. Escoger el adecuado número de rezagos ( $k$ ), penalizando el underfitting dado que cuando se penaliza el overfitting se está subestimando el costo de usar un modelo de orden bajo cuando el proceso de raíz unitaria tiene una raíz negativa en el MA o cuando el proceso MA tiene un coeficiente alto.

La estructura básica del test, propuesta originalmente por Stock (1990), se basa en el hecho de que las series convergen a una tasa de normalización distinta bajo la nula que bajo la alternativa. La suma de los cuadrados de una serie  $I(1)$  es  $O(T^2)$ , en cambio, la de una serie  $I(0)$  es  $O(T)$ . Es decir,

$$\sum_{t=1}^T y_{t-1}^2 \quad \text{converge a tasa} \quad \begin{matrix} T^2 & \text{si la variable es } I(1) \\ T & \text{si la variable es } I(0) \end{matrix}$$

Con este principio se elaboran los siguientes estadígrafos para el caso univariado, además de  $MZ_t = MZ_{\#} \times MSB$ :

$$\begin{aligned} MZ_{\#} &= (T^{-1} y_T^2 - s_{AR}^2) (2T^{-2} \sum_{t=1}^T y_{t-1}^2)^{-1} \\ MSB &= (T^{-2} \sum_{t=1}^T y_{t-1}^2 / s_{AR}^2)^{-1/2} \end{aligned} \tag{5.77}$$

<sup>20</sup> Ng, S. y P. Perron, "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica* 69(6), 1519-1554, 2001.

donde  $s_{AR}^2 = \hat{\sigma}_k^2 / (1 - \hat{\beta}(1))^2$  es corresponde a una corrección de la varianza estimada de los residuos de la ecuación del test de DF  $\hat{\sigma}_k^2$  cuando la estimación incluye  $k$  rezagos, y  $\hat{\beta}(1) = \sum_{i=1}^k \hat{\beta}_i$  es la suma de los coeficientes del término de rezagos en el test ADF. La varianza se va haciendo más pequeña a medida que  $k$  es mayor, por lo que cada vez hay menos probabilidad de rechazar la hipótesis nula de raíz unitaria. La corrección permite “recuperar” el tamaño de la varianza mejorando el poder del test.

La cuasidiferenciación por GLS consiste en definir un  $\Delta^d y_t \equiv y_t - \alpha y_{t-1}$ . Así se diferencia la serie, la constante y/o la tendencia. A partir de  $\Delta^d y_t = \Delta_{\#}^d + \Delta^d \delta t$  se estiman y/o para eliminar el componente determinístico de la serie:  $\tilde{y}_t = y_t - (\hat{\alpha}(a) + \hat{\delta}t(a))$ . El objetivo es que al aplicar el test a la serie cuasidiferenciada ( $\tilde{y}_t$ ), éste permita rechazar la hipótesis nula de raíz unitaria, incluso en presencia de series muy persistentes. El  $k$  óptimo permite controlar la persistencia de la serie cuando se observa un componente MA(1). Un modelo parsimonioso con un  $k$  muy bajo, pero con un parámetro  $\theta$  alto en  $\theta \epsilon_{t-j}$ , tenderá producir que el test se “confunda” interpretando que la serie tiene una raíz unitaria.

## 5.07 Tests con hipótesis nula de estacionariedad

Debido a los problemas de poder y tamaño de los test de raíces unitarias que tienen como hipótesis nula no estacionariedad, ha habido interés en desarrollar test con la hipótesis nula opuesta, es decir, estacionariedad. Este tipo de test es menos común pero aún así es útil usarlo con el objetivo de comprobar las conclusiones del primer tipo de test.

Uno de los tests más populares es el de Kwiatkovsky et al. (1992)<sup>21</sup>, usualmente llamado KPSS. Considere el siguiente modelo:

$$y_t = \delta t + \varsigma_t + \epsilon_t \quad (5.78)$$

donde  $\varsigma$  es un paseo aleatorio del tipo  $\varsigma_t = \varsigma_{t-1} + \mu_t$  con  $\mu_t \rightarrow N(0, \sigma_\mu^2)$ . La hipótesis nula de estacionariedad corresponde a  $H_0: \sigma_\mu^2 = 0$  que equivale a decir que  $\varsigma$  es constante. Este modelo es una versión del modelo de parámetros cambiantes de Nabeya y Tanaka (1988)<sup>22</sup>

21 Kwiatkowski, D., P. C. B. Phillips, P. Schmidt y Y. Shin (1992): "Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159-178.

22 Nabeya, E. y K. Tanaka (1988): "Asymptotic theory of a test for the constancy of regression coefficients

$$\begin{aligned} y_t &= \beta_t x_t + \gamma z_t + \epsilon_t \\ \beta_t &= \beta_{t-1} + \mu_t \end{aligned} \quad (5.79)$$

Nabeya y Tanaka proponen el test  $LM = \sum_{t=1}^T S_t^2 / \hat{\sigma}_e^2$ , donde  $e$  son los residuos de una regresión de  $y_t$  en una constante y tendencia lineal<sup>23</sup>,  $\hat{\sigma}_e^2$  es la varianza residual de dicha regresión, y  $S_t = \sum_{i=1}^t e_i$  es la suma acumulada de residuos. Este estadígrafo fue derivado suponiendo que las innovaciones son i.i.d. KPSS extiende el test para el caso en que las innovaciones no son i.i.d., lo que requiere usar un estimador de largo plazo de la varianza residual  $\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} E(S_T^2)$ . Un estimador consistente de  $S_T^2$  es:

$$S_{TL}^2 = T^{-1} \sum_{t=1}^T e_t^2 + 2T^{-1} \sum_{\tau=1}^l w_{\tau l} \sum_{t=\tau+1}^T e_t e_{t-\tau} \quad (5.80)$$

donde  $w_{\tau l}$  es una función de ponderación (kernel) que, en el caso de KPSS, equivale a

$$w_{\tau l} = 1 - \frac{\tau}{l+1}.^{24}$$

## 5.08 Equivalencia Observacional, Poder y Tests de Raíces Unitarias

Un problema común a cualquier test de raíz unitaria es el bajo poder derivado de la equivalencia observacional. Supongamos que el verdadero modelo es

$$y_t = y_{t-1} + \epsilon_t \quad (5.81)$$

pero estimamos en una muestra finita de tamaño  $T$  un proceso AR(1).

$$y_t = \rho y_{t-1} + \epsilon_t \quad (5.82)$$

Consideremos ahora la predicción y ECM de ambos modelos. Para la ecuación (5.81)

---

against the random walk hypothesis", *Annals of Statistics*, 16:218-235.

23 Para testear la hipótesis nula de estacionariedad alrededor de una constante, se omite la tendencia lineal.

24 Esta es una ventana de Bartlett similar a la de Newey-West.



$$\begin{aligned}\hat{y}_{t+s|t} &= y_t \\ E[y_{t+s} - \hat{y}_{t+s|t}]^2 &= s\sigma^2\end{aligned}\tag{5.83}$$

en tanto que para la ecuación (5.82)

$$\begin{aligned}\hat{y}_{t+s|s} &= \rho^s y_t \\ E[y_{t+s} - \hat{y}_{t+s|t}]^2 &= (1 + \rho^2 + \rho^4 + \dots)\sigma^2\end{aligned}\tag{5.84}$$

Así, para un  $\phi$  suficientemente grande, los dos procesos se ven casi iguales en una muestra finita. A esto se le llama *equivalencia observacional*.

Igualmente, si el proceso verdadero es estacionario  $y_t = \epsilon_t$  y estimamos el modelo en primeras diferencias

$$(1-L)y_t = (1+\theta)\epsilon_t\tag{5.85}$$

con  $\epsilon_0 = y_0 = 0$ . Las proyecciones y el ECM  $s$  periodos adelante para el verdadero modelo son:

$$\begin{aligned}\hat{y}_{t+s|s} &= 0 \\ E[y_{t+s} - \hat{y}_{t+s|t}]^2 &= \sigma^2\end{aligned}\tag{5.86}$$

en tanto que para el modelo de la ecuación (5.85) se obtiene

$$\begin{aligned}\hat{y}_{t+s|s} &= (1+\theta)(\epsilon_t + \epsilon_{t-1} + \dots + \epsilon_1) \\ E[y_{t+s} - \hat{y}_{t+s|t}]^2 &= (1 + (s-1)(1+\theta)^2)\sigma^2\end{aligned}\tag{5.87}$$

de nuevo, si el  $\theta$  es cercano a -1, los procesos se parecen bastante en una muestra finita.

La razón de la equivalencia observacional es que en una muestra finita los procesos que tienen mucha memoria suelen asemejarse a aquellos que tienen raíces unitarias porque no hay suficiente información para determinar si las innovaciones son transitorias o permanentes.

¿Qué efecto tiene esto sobre los tests de raíces unitarias?

- Pérdida de poder cuando la  $H_1$  es “cercana” a  $H_0$ .
- Búsqueda de tests donde  $H_0$  sea estacionariedad (ver Kwiatkosvky et al., 1992)
- Búsqueda de tests no paramétricos.

### 5.09 Tests No Paramétricos de Raíces Unitarias

Uno de los tests no paramétricos de persistencia de una serie de tiempo más comunes es el de Cochrane (1988). El argumento es bastante simple y se basa en evaluar qué tan persistentes son las innovaciones del proceso estocástico. Si ellas son completamente persistentes, entonces existe una raíz unitaria. Si ellas desaparecen en el tiempo, el proceso es estacionario en tendencia.

Medimos la persistencia de las innovaciones usando el hecho que toda variable cuya primera diferencia es un proceso estacionario lineal, puede ser representada por una combinación de procesos estacionarios ( $c_t$ ) y camino aleatorio ( $z_t$ ).

$$\begin{aligned}
 y_t &= z_t + c_t \\
 z_t &= \mu + z_{t-1} + a(1)\epsilon_t \\
 a(1) &= \sum_{j=0}^{\infty} a(j) \\
 c_t &= a^*(L)\epsilon_t a_k^* = - \sum_{j=k+1}^{\infty} a(j)
 \end{aligned} \tag{5.88}$$

donde  $\epsilon_t$  es ruido blanco,  $a(1)$  es la parte del innovaciones que no se disipa en el largo plazo y  $\mu$  es el deriva de la variable  $y_t$ .

Cualquiera sea la forma en que se separen componentes transitorios y permanentes, la varianza de este último es la misma y corresponde a:

$$\sigma_{\Delta z}^2 = \left( \sum a_j \right)^2 \sigma_{\epsilon}^2 \tag{5.89}$$

Así, se testea usando la razón entre la varianza de  $z$  y aquella de la primera diferencia de  $y_t$ :

$$V = \frac{\sigma_{\Delta z_t}^2}{\sigma_{\Delta y_t}^2} \tag{5.90}$$

- Si  $y_t$  es TS,  $\forall \varepsilon_t$  no hay efecto permanente y  $\sigma^2_{\Delta z}$  y  $V$  tenderán a cero.
- Si  $y_t$  es RW,  $\forall \varepsilon_t$  hay efectos permanentes que  $\sigma^2_{\Delta z} \approx \sigma^2_{\Delta y}$  y  $V=1$ .

En términos prácticos, se estima una secuencia de  $V$ , variando el largo del intervalo en el numerador:

$$V_k = \frac{1}{k} \frac{\sigma^2_{(y_t - y_{t-k})}}{\sigma^2_{(y_t - y_{t-1})}} \quad (5.91)$$

Se puede demostrar que  $V$  es equivalente a la suma de las autocorrelaciones de la serie más uno. En muestras finitas se usa el estimador de la autocorrelación de la primera diferencia de la serie:

$$V^m = 1 + 2 \left[ \sum_{j=1}^{m-1} \frac{m-j}{m} \hat{\rho}_{\Delta Y_j} \right] \quad (5.92)$$

$$\hat{\rho}_{\Delta Y_j} = \frac{T}{T-j} \frac{\text{cov}(\Delta Y_t, \Delta Y_{t-j})}{\text{var}(\Delta Y_t)}$$

Los términos  $\frac{m-j}{m}$  y  $\frac{T-j}{T}$  son correcciones por tamaño de muestra.

Prueba: Sea el modelo  $(1-L)y_t = \beta' + A(L)\mu_t = \beta' + \sum_{j=0}^{\infty} \theta_j \mu_{t-j}$ . Usando la siguiente identidad  $(1-L^k)(1-L)^{-1} = (1+L+L^2+\dots+L^{k-1})$  se puede obtener la diferencia de orden  $k$  de  $y_t$ :  $y_t - y_{t-k} = k\beta' + \sum_{j=0}^{k-1} \left( \sum_{l=0}^j \theta_l \right) \mu_{t-j} + \sum_{j=k}^{\infty} \left( \sum_{l=j-k+1}^j \theta_l \right) \mu_{t-j}$

La varianza es  $\sigma_k^2 = k^{-1} \text{var}(y_t - y_{t-k}) = k^{-1} \left[ \sum_{j=0}^{k-1} \left( \sum_{l=0}^j \theta_l \right)^2 + \sum_{j=k}^{\infty} \left( \sum_{l=j-k+1}^j \theta_l \right)^2 \right] \sigma_{\mu}^2$ .

Expresamos la varianza como una ecuación en diferencias  $k\sigma_k^2 - (k-1)\sigma_{k-1}^2 = \left[ \sum_{j=0}^{\infty} (\theta_j^2 + 2\theta_j \sum_{l=1}^{k-1} \theta_{j+l}) \right] \sigma_{\mu}^2$ . Sea  $\sigma_l^2 = \left( \sum_{j=0}^{\infty} \theta_j^2 \right) \sigma_{\mu}^2$ , se deduce

que  $1 + 2 \sum_{j=1}^{k-1} \rho_j = \frac{k\sigma_k^2 - (k-1)\sigma_{k-1}^2}{\sigma_k^2}$  donde  $\rho_j$  es la autocorrelación de orden  $j$

de  $(1-L)y_t$ . Así,  $\frac{\sigma_k^2}{\sigma_l^2} = k^{-1} [1 + (1+2\rho_1) + (1+2\rho_1+2\rho_2) + \dots] = 1 + 2 \sum_{j=1}^{k-1} \frac{k-j}{k} \rho_j$  ■.

## 5.10 Tests de Raíces Unitarias Estacionales

Es común encontrar en variables económicas el fenómeno de estacionalidad, es decir, el movimiento intra-anual sistemático –aunque no necesariamente regular– de la serie de tiempo que representa una variable económica<sup>25</sup>. Existe un largo, y aún inconcluso, debate sobre la conveniencia de modelar la economía usando variables originales o desestacionalizadas. Hansen y Sargent (1993)<sup>26</sup> defienden la idea que agentes racionales toman decisiones en base a datos sin desestacionalizar y que remover la estacionalidad implica, en el mejor de los casos, eliminar información que podría ser útil para estimar los parámetros de un modelo. Sims (1993)<sup>27</sup>, por el contrario, apoya la idea opuesta; resulta preferible usar datos desestacionalizados por cuanto en el proceso de ajustar un modelo a los datos originales, se corre el riesgo de darle demasiada importancia a capturar efectos estacionales –que en sí son secundarios– en desmedro de una mejor parametrización de las características principales del modelo. A pesar de lo importante que resulta este debate, la mayoría de los investigadores prefiere usar datos desestacionalizados. La justificación sería que, si bien estos movimientos son anticipables, típicamente éstos no están directamente relacionados al objetivo del estudio y, por lo tanto, no resulta necesario modelarlos explícitamente.

Entre los métodos más comunes para remover la estacionalidad están el uso de variables mudas estacionales y el método de promedios móviles, que incluye desde la “variación en x-periodos” hasta el ARIMA X-11. Esta última es la metodología más popular tanto porque su nivel de desarrollo analítico es superior, como porque viene implementada como una opción sencilla de usar en muchos programas econométricos.

Este tipo de filtro asume que la estacionalidad es estocástica pero estacionaria. No obstante, si los componentes estacionales de una variable son estocásticos, es posible que ellos tengan una raíz unitaria no sólo en su comportamiento de largo plazo, sino que también en los efectos estacionales. No reconocer la existencia raíces unitarias estacionales puede originar problemas de correlaciones espúreas y parametrizaciones inestables.

Más aún, cuando las series macroeconómicas son estacionalmente integradas (es decir, en el caso que los shocks estacionales tienen componentes permanentes), desestacionalizarlas con los métodos clásicos resulta inadecuado. La remoción de estacionalidad, además, afecta el poder de los tests de raíces unitarias y de cointegración sobre cuyos resultados los econométricos se basan para determinar la forma funcional de los modelos y juzgar la calidad de las estimaciones.

---

25 Hylleberg, S. (1992). *Modelling Seasonality*. Oxford: Oxford University Press.

26 Hansen, L.P. and T. Sargent (1993). Seasonality and approximation errors in rational expectations models, *Journal of Econometrics* 55(1-2): 21-56.

27 Sims, C.A. (1993). Rational expectations modeling with seasonally adjusted data, *Journal of Econometrics* 55(1-2): 9-20.

Si la estacionalidad es determinística, removerla con la ayuda de variables estacionales no tiene efecto alguno sobre el desempeño o resultado de los tests de raíz unitaria. Sin embargo, en presencia de efectos estacionales estocásticos, los filtros tradicionales pueden afectar de manera significativa el poder de los tests de raíz unitaria. Ghysels (1990)<sup>28</sup> muestra que la remoción de estacionalidad con el método X-11 o una “variación en X períodos” introduce persistencia excesiva en la serie, lo cual reduce el poder de los tests para rechazar la hipótesis de no estacionariedad. Olekalns (1994) extiende este resultado a los casos en que usan dummies o filtros band-pass para remover la estacionalidad.

Al igual que en el caso que existen raíces unitarias en el componente secular de una variable, la presencia de raíces unitarias en el componente estacional es capaz de producir correlación espúrea en regresiones que utilizan dichas variables. Abeyasinghe (1991) realiza un análisis de Monte Carlo estudiando el efecto en regresiones de mínimos cuadrados que incluyen variables que por construcción están no correlacionadas pero cuyos componentes estacionales son no estacionarios. Utilizando 500 replicaciones encuentra que se rechazó la hipótesis nula de no-correlación entre las variables en más del 50% de los casos pese a que se controló el test con variables mudas estacionales.

Este tipo de observaciones son las que han llevado, desde mediados de los años 1980, al desarrollo de un gran número de tests de raíces unitarias en el componente estacional. Tests paramétricos incluyen los de Dickey et al. (1984), Hylleberg, Engle, Granger y Yoo (1990) y Johansen y Schaumburg (1999). Geweke y Porter-Hudak (1983) proponen tests no-paramétricos, mientras que Franses et al. (1997) proponen una metodología Bayesiana para testear por raíces unitarias estacionales simultáneamente con cambio de medias estacionales, lo que incrementa el poder de los tests basados en la hipótesis nula de no estacionariedad.<sup>29</sup>

Para testear la existencia de raíces unitarias estacionales, usualmente se emplea la extensión del test de Dickey-Fuller desarrollada para el caso de datos trimestrales por Hylleberg et al. (1990) y llamado popularmente test HEGY. Existe una versión para datos mensuales desarrollada por Beaulieu y Miron (1993).<sup>30</sup> Si bien existen métodos alternativos (como el de Canova y Hansen, 1995<sup>31</sup>), el test HEGY es preferido porque

28 Ghysels, E. (1990). Unit-root tests and the statistical pitfalls of seasonal adjustment: the case of U.S. postwar real gross national product, *Journal of Business and Economic Statistics*, 8, 145-152

29 Dickey, D. A., Hasza, D. P., and W. A. Fuller (1984). Testing for unit roots in seasonal time series, *Journal of the American Statistical Association*, 79: 355-367. Franses Ph.H.B.F., Hans, H. Hoek R. Paap (1997), Bayesian analysis of seasonal unit roots and seasonal mean shifts, *Journal of Econometrics* 78(2): 359-380. Geweke, J. and Porter-Hudak, S. (1983), The estimation and application of long memory time series models, *Journal of Time Series Analysis*, 4: 221-238. Hylleberg, S., R. Engle, C. W. J. Granger and B. S. Yoo (1990). Seasonal integration and co-integration, *Journal of Econometrics*, 44: 215-238. Johansen, S. and E. Schaumburg (1999). Likelihood analysis of seasonal cointegration, *Journal of Econometrics* 88(2): 301-339.

30 Beaulieu, J. J. and J. A. Miron (1995). What have macroeconomists learned about business cycles from the study of seasonal cycles?, NBER Working Papers # 5258.

31 Canova, F. and B. Hansen (1995). Are seasonal patterns constant over time? A test for seasonal stability,

procede de general a específico, tiende a ser más robusto en presencia de raíces unitarias no estacionales en las variables, y tiene mayor poder que especificaciones alternativas. El test HEGY parte de la base que el crecimiento anual de cualquier serie de tiempo en frecuencia trimestral se puede descomponer de la siguiente forma:

$$(1-L^4) = (1-L)(1+L)(1-iL)(1+iL) \quad (5.93)$$

donde  $L$  es el operador lineal de rezagos e  $i = \sqrt{-1}$ . El término del lado izquierdo es la tasa de crecimiento anual o diferencia en logaritmos en cuatro periodos. Los términos del lado derecho corresponden respectivamente a la diferencia anual, la diferencia semestral y, los últimos dos términos, al componente trimestral de la serie.

La especificación es muy útil ya que permite obtener un test de forma general para evaluar en forma simultánea diversas hipótesis sobre el proceso de comportamiento estacional de la serie. La forma general del test de la descomposición anterior es:

$$(1-L^4) = (1-\alpha_1 L)(1+\alpha_2 L)(1-\alpha_3 iL)(1+\alpha_4 iL) \quad (5.94)$$

donde  $\alpha_1, \alpha_2, \alpha_3$ , y  $\alpha_4$  son parámetros. La distinta conformación de parámetros da origen a variables que tienen raíces unitarias en distintas frecuencias. En particular:

- Si  $\alpha_1=1$ , la variable tiene una raíz unitaria no estacional (la raíz unitaria tradicional).
- Si  $\alpha_2=1$ , la variable tiene una raíz unitaria de frecuencia semestral.
- Si  $\alpha_3=1$  y  $\alpha_4=1$ , la variable tiene raíz unitaria trimestral.

La implementación del test requiere definir las siguientes variables auxiliares:

$$\begin{aligned} x_t^1 &= (1+L+L^2+L^3)x_{t-1} = x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4} \\ x_t^2 &= (1-L+L^2-L^3)x_{t-1} = x_{t-1} - x_{t-2} + x_{t-3} - x_{t-4} \\ x_t^3 &= x_{t-1} - x_{t-3} \end{aligned} \quad (5.95)$$

Las variables se usan para estimar esta ecuación por mínimos cuadrados ordinarios:

$$(1-L^4)x_t = \pi_1 x_{t-1}^1 - \pi_2 x_{t-1}^2 + \pi_3 x_{t-1}^3 - \pi_4 x_{t-2}^2 + \epsilon_t \quad (5.96)$$

Al igual que en el test de Dickey-Fuller, la hipótesis nula del test es que existen raíces unitarias en los componentes anuales, semestrales y trimestrales. La hipótesis alternativa es que las variables son estacionarias. Opcionalmente, se puede incorporar una tendencia, la constante y dummies estacionales. Estas últimas capturan la presencia de estacionalidad no estocástica. La interpretación de los resultados es la siguiente:

- (a) Si no se puede rechazar  $H_0: \pi_1=0$ , existe una raíz unitaria no estacional en  $x_t$ .
- (b) Si no se puede rechazar  $H_0: \pi_2=0$ , existe una raíz unitaria semestral (bianual) en  $x_t$ .
- (c) Si no se puede rechazar  $H_0: \pi_3=\pi_4=0$ , existe una raíz unitaria estacional trimestral en  $x_t$ .

La interpretación de la hipótesis (a) es la estándar del test de Dickey-Fuller. Nótese que la hipótesis (c) requiere un test conjunto. En este caso, basta usar un test F. Bajo la hipótesis nula que hay raíces unitarias en la variable, los tests de hipótesis de los parámetros obtenidos no tienen la distribución estándar normal, por lo que los resultados del test deben contrastarse con los valores críticos tabulados por HEGY. Éstos, a su vez, dependen de si la hipótesis nula incorpora regresores de intercepto, tendencia determinística y estacionalidad determinística.

Además, debido a que el test se basa en la hipótesis de no estacionariedad en los distintos componentes, su poder es limitado para distinguir procesos estacionarios con autocorrelación cercana, pero distinta, a uno. La inclusión de rezagos de la variable dependiente, sin embargo, permite controlar la eventual correlación de residuos y aumenta el poder del test. Además, la inclusión de variables mudas estacionales, tendencia y constante permite enriquecer la hipótesis alternativa, que en ausencia de estas adiciones es que la serie es un ruido blanco.

Este modelo puede ser extendido para acomodar una hipótesis alternativa más rica que la que hemos considerado implícitamente hasta el momento (es decir, que  $y_t$  es ruido blanco). De manera equivalente a como se extiende el test de Dickey-Fuller se pueden incluir un intercepto, tendencia determinística, y estacionalidad determinística de la siguiente manera:

$$(1-L^4)x_t = \pi_1 x_{t-1}^1 - \pi_2 x_{t-1}^2 + \pi_3 x_{t-1}^3 - \pi_4 x_{t-2}^2 + \pi_0 + \beta t + \theta_1 D_1 + \theta_2 D_2 + \theta_3 D_3 + \epsilon_t \quad (5.97)$$

## 5.11 Integración Fraccional

La distinción entre variables de orden 0 (estacionarias) y variables de orden 1 (integradas) es un tanto arbitraria, ya que el concepto relevante es la propiedad de reversión a la media, más que el orden de integración de la variable. De hecho, abandonar la hipótesis de un orden de integración entero por uno fraccional permite captar la dinámica de aquellos de procesos que revierten muy lentamente a la media.<sup>32</sup>

Una variable integrada fraccionalmente presenta un lento ajuste hacia el equilibrio, lo que hace que este tipo de proceso sea atractivo al momento de modelar algunas variables económicas. Al permitir que el orden de integración de las series asuma cualquier valor real podemos estudiar modelos de memoria larga. Una de las características de este tipo de proceso es que al observar el correlograma de las series originales, éstas parecen ser no estacionarias, mientras que las series diferenciadas parecen estar sobrediferenciadas. Evidencia de larga memoria ha sido encontrada en el consumo (Diebold y Rudebusch, 1991)<sup>33</sup>, tasas de interés (Shea, 1991) e inflación (Baum et al., 1999). Por otro lado, Granger y Joyeux (1980)<sup>34</sup> y Lin (1991) muestran que esto puede ocurrir debido a la agregación de datos.

El proceso ARMA fraccionalmente integrado –*ARFIMA*– es una generalización del modelo ARIMA para el caso en que el orden de diferenciación puede asumir un valor no entero. La serie  $y = \{y_1, \dots, y_T\}$  es un proceso *ARFIMA*( $p, d, q$ ) con media  $\mu$  si:

$$\phi(L)(1-L)^d(y_t - \mu) = \theta(L)\epsilon_t, \text{ donde } \epsilon \sim i.i.d.(0, \sigma_\epsilon^2) \quad (5.98)$$

donde el orden de diferenciación de la variable  $-d-$  puede asumir cualquier valor real,  $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ ,  $\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$  y

$$(1-L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \dots \quad (5.99)$$

El proceso estocástico  $y_t$  es estacionario e invertible si todas las raíces de  $\Phi(L)$  y de  $\Theta(L)$  se encuentran fuera del círculo unitario y si  $|d| \leq 0.5$ . El proceso es no estacionario si  $|d| > 0.5$ , teniendo varianza infinita.

32 La teoría sobre integración fraccional se inicia con Granger y Joyeux (1980) y Hosking (1981) y es considerada otra modelación del concepto de persistencia en series de tiempo.

33 Diebold, F. X. y Rudebusch, G. D. (1991). Is consumption too smooth? Long memory and the Deaton paradox, *Review of Economics and Statistics*, 73:1-9.

34 Granger, C. W. J. y Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1:15-39.



Asumiendo que  $0 < d < 0.5$  y que  $d \neq 0$ , Hosking (1981)<sup>35</sup> mostró que la función de autocorrelación,  $\rho(k)$ , de un proceso ARFIMA es proporcional a  $k^{2d-1}$  cuando  $k \rightarrow \infty$ . Consecuentemente, las autocorrelaciones de un proceso ARFIMA, para  $0 < d < 0.5$ , decaen hiperbólicamente a cero cuando  $k \rightarrow \infty$ . Lo contrario sucede en un proceso ARMA donde decaen exponencialmente cuando  $k \rightarrow \infty$ .

Para  $0 < d < 0.5$  el proceso ARFIMA se dice que tiene memoria larga. Se dice que el proceso tiene memoria intermedia si  $-0.5 < d < 0$ . Exhibe memoria corta cuando  $d = 0$ , lo que corresponde al modelo ARMA estacionario e invertible. Para  $d < 1$  el proceso presenta reversión a la media y es estacionario en covarianza cuando  $-0.5 < d < 0.5$  y es no estacionario en covarianza para  $0.5 < d < 1$ .

## 5.12 Tests de Integración Fraccional

Los tests de raíces unitarias convencionales no son, en general, indicados para testear integración fraccional. Diebold y Rudebusch (1991) y Hassler y Wolters (1994)<sup>36</sup> mostraron, respectivamente, que el test de Dickey-Fuller y el test de Phillip-Perron son poco potentes para distinguir entre un proceso  $I(1)$  y un proceso  $I(d)$ . Lee y Schmidt (1996) afirman, sin embargo, que el test KPSS puede ser utilizado para distinguir un proceso  $I(0)$  de un proceso  $I(d)$ .

**Test GPH:** Geweke y Porter-Hudak (1983)<sup>37</sup> propusieron un estimador semi-paramétrico para el parámetro  $d$ . La función de densidad espectral de  $y_t$  es

$$f_y(w) = \left( 4 \sin^2\left(\frac{w}{2}\right) \right)^{-d} f_u(w) \quad (5.100)$$

donde  $f_u(w)$  es la función de densidad espectral de  $u_t = (1-L)^d y_t$ . Aplicando logaritmo y permitiendo la inclusión de frecuencias armónicas  $w_j = \frac{2\pi j}{T}$ , donde  $T$  es el tamaño de muestra, se obtiene:

35 Hosking, J.R.M. (1981): "Fractional differencing", *Biometrika*, 68: 165-176.

36 Diebold, F.X. and G.D. Rudebusch (1989): "Long memory and persistence in aggregate output", *Journal of Monetary Economics*, 24: 189-209. Hassler, U. y J. Wolters (1994) "On the power of unit root tests against fractional alternatives", *Economics Letters*, 45: 1-5.

37 Geweke, J. y S. Porter-Hudak (1983) "Estimation and application of long memory time series models", *Journal of Time Series Analysis*, 4: 221-238.

$$\log(f_y(w)) = \log(f_u(0)) - d \log\left(4 \sin^2\left(\frac{w_j}{2}\right)\right) + \log\left(\frac{f_u(w_j)}{f_u(0)}\right) \quad (5.101)$$

Para frecuencias  $w_j$  próximas a cero, el último término es despreciable en relación al resto. Sumando  $I_y(w_j)$ , el periodograma de  $w_j$ , a ambos lados de la ecuación (5.101) resulta

$$\log(I_y(w)) = \log(f_u(0)) - d \log\left(4 \sin^2\left(\frac{w_j}{2}\right)\right) + \log\left(\frac{I_y(w_j)}{f_u(0)}\right) \quad (5.102)$$

Geweke y Porter-Hudak sugieren la siguiente regresión:

$$\log(I_y(w)) = \alpha + \beta \log\left(4 \sin^2\left(\frac{w_j}{2}\right)\right) + v_j \quad j = 1, 2, \dots, k \quad (5.103)$$

en que  $\beta = -d$  y  $v_j$  es i.i.d. con media cero y varianza asintótica  $\pi^2/6$ . Si el número de frecuencias  $k$  fue escogido de forma que  $k = g(T)$ , donde  $g(T)$  cumple

$$\lim_{T \rightarrow \infty} g(T) = \infty, \quad \lim_{T \rightarrow \infty} \frac{g(T)}{T} = 0, \quad \lim_{T \rightarrow \infty} \log \frac{(T)^2}{g(T)} = 0$$

entonces el estimador obtenido por el método de mínimos cuadrados ordinarios de  $d$  es consistente y la distribución de  $(\hat{d} - d)/SE(\hat{d})$  es asintóticamente Normal.

Una de las ventajas de este método es que permite estimar  $d$  sin conocer los parámetros  $p$  y  $q$  del proceso  $ARFIMA(p, q, d)$ . Las propiedades teóricas del estimador GPH han sido investigadas sólo en procesos con  $-0.5 < d < 0.5$ , esto es, en procesos estacionarios e invertibles. Robinson (1991)<sup>38</sup> sugiere que para una serie con un  $d$  fuera de aquel intervalo, ésta sería diferenciada un número entero de veces hasta que se obtenga un  $d$  dentro de dicho intervalo. En la práctica el método de estimación GPH es frecuentemente aplicado a series no estacionarias o no invertibles antes de proceder a su diferenciación. Sería deseable tener un estimador de  $d$  que no estuviese sujeto a la restricción  $|d| < 0.5$ , como sucede con el estimador GPH. Hurvick y Ray (1995), mostraron que el estimador GPH puede ser fuertemente sesgado en procesos no estacionarios o no invertibles y además, generalmente no es invariante en primera

38 Robinson, P.M. (1991) "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression", *Journal of Econometrics* 47: 67-84.

diferencia, esto es, la estimación de  $d$  obtenida a partir de la serie original no es igual a la suma de las estimaciones de  $d$  obtenida a partir de la serie diferenciada.

Un importante problema práctico en la implementación del estimador GPH es la elección de  $k$ . Geweke y Porter-Hudak (1983) sugieren utilizar  $k = T^{1/2}$ , sin embargo, en la práctica, dicha estrategia parece no ser la más indicada. Por un lado, reducir  $k$  permite reducir de cierta forma el sesgo, pero por otro, esto sólo es posible a costa de un aumento de la varianza. Hurvich et al. (1998)<sup>39</sup> sugieren  $k = o(T^{4/5})$ , que demuestran ser el  $k$  óptimo que resulta de la minimización del error cuadrático medio y cuyo desempeño es superior a  $k = T^{1/2}$ , incluso en muestras pequeñas. A pesar de que el estimador GPH es fácil de aplicar, razón por la cual es el más utilizado, y ser robusto a la no normalidad, su comportamiento en presencia de alta autocorrelación en  $u_t$  reduce su atractivo.

**Test REG:** Agiakloglou y Newbold (1994)<sup>40</sup> propusieron un test LM (Lagrange Multiplier), denominado REG, basado en la regresión auxiliar

$$\hat{\epsilon}_t = \sum_{i=1}^p \beta_i W_{t-i} + \sum_{j=1}^q \gamma_j Z_{t-j} + \delta K_m + \mu_t \quad (5.104)$$

donde  $K_m = \sum_{j=1}^m j^{-1} \hat{\epsilon}_{t-j}$ ,  $\hat{\Theta}(L)W_t = y_t$ ,  $\hat{\Theta}(L)Z_t = \hat{\epsilon}_t$ ,  $u_t$  es i.i.d. con distribución Normal, y  $\hat{\epsilon}_t$  y  $\hat{\Theta}(L)$  son respectivamente el residuo estimado y el polinomio MA estimado del modelo ARFIMA bajo la hipótesis nula ( $d=0$ ).

Los parámetros  $p$  y  $q$  de los componentes AR y MA son elegidos utilizando el criterio informacional de Schwartz (1978). La ecuación (104) debe ser estimada por mínimos cuadrados ordinarios para el período  $t = m+1, \dots, T$ . El test  $t$  de la hipótesis  $\delta=0$ , conjuntamente con los valores críticos asintóticos de la distribución Normal estándar constituyen el test REG.

**Test AUTO:** Una vez que se pierden  $m$  observaciones en la estimación del modelo (7), Agiakloglou y Newbold (1994) sugieren en el mismo artículo, un test alternativo llamado AUTO. Se asume que una serie fue suficientemente diferenciada hasta tener una apariencia estacionaria, a la cual le fue ajustado un modelo ARMA ( $p,q$ ). Bajo la

39 Hurvich, C.M., R. Deo y J. Brodsky (1998): "The mean square error of Geweke and Porter-Hudak estimator of the memory parameter in a long memory time series", *Journal of Time Series Analysis* 19: 19-46.

40 Agiakloglou, C. y P. Newbold (1994): "Lagrange multiplier tests for fractional difference", *Journal of Time Series Analysis*, 15: 253-262.

hipótesis nula de que la especificación ARMA (p,q) es correcta,  $\hat{r}' = (\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m)$ , en que  $\hat{r}_j$  es la autocorrelación de residuos de orden  $j$ , es asintóticamente normal con media cero y matriz varianza-covarianza  $C = T^{-1}V$ , donde los elementos de la matriz  $V$  son función de los parámetros de los componente AR y MA.

Davis et al (1977) y Ansley y Newbold (1979) sugieren la siguiente expresión para la matriz varianza-covarianza  $W = T^{-1}(T+2)^{-1}LV$ , donde  $L$  es una matriz diagonal  $m \times m$  en que el  $i$ -ésimo elemento de la diagonal es  $(T-i)^{1/2}$ . Si  $S_m = \sum_{j=1}^m j^{-1} \hat{r}_j$ , entonces, bajo la hipótesis nula,  $\text{var}(S_m) = h'Wh$ , donde el vector  $h$  con  $m$  elementos tiene como  $j$ -ésimo elemento a  $j^{-1}$ . El test está basado en

$$Z = (h' \hat{W} h)^{-1/2} S_m$$

que posee una distribución asintóticamente normal estándar bajo la hipótesis nula y donde los parámetros de los componentes AR y MA son sustituidos por sus estimadores en el cálculo de  $\hat{W}$ .

Agiakloglou y Newbold (1994) compararon los tests AUTO y REG y concluyeron que este último era el que verificaba una mayor proximidad entre el nivel de significancia real y el nivel de significancia nominal. Además demostraron que la elección de  $m$  tiene poco impacto en la potencia de ambos tests y que mientras el test REG es mejor en la detección de procesos con  $d$  negativo, el test AUTO es más potente cuando  $d$  es positivo.

**Test MRR:** El test R/S fue propuesto por Hurst (1951) y refinado por Mandelbrot (1972) y McLeod y Hipel (1978). Lo (1991) propuso una versión del estadístico robusta frente a la presencia de memoria corta en la serie y de heterocedasticidad. El test MRR (Modified Rescaled Range) está definido por

$$Q_T = \frac{R_T}{\hat{\sigma}_T(k)}$$

donde  $R_T = \max_{0 \leq i \leq T} \sum_{t=1}^i (y_t - \bar{y}) - \min_{0 \leq i \leq T} \sum_{t=1}^i (y_t - \bar{y})$  y

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + 2 \sum_{j=1}^k \sum_{i=j+1}^T \left(1 - \frac{j}{k+1}\right) (y_i - \bar{y})(y_{i-j} - \bar{y})$$

En que  $\bar{y}$  es la media muestral y  $\hat{\sigma}^2$  es el estimador de máxima verosimilitud de la varianza. De acuerdo con Andrews (1991),  $k$  debe ser la parte entera de  $(3T/2)^{1/3} (2\hat{\rho}/(1-\hat{\rho}^2))^{2/3}$ , en que  $\hat{\rho}$  es el coeficiente de autocorrelación de primer orden. Lo (1991) provee los valores críticos asintóticos para el test MRR. Cheung (1993) muestra que el test MRR es conservador en procesos con componente autoregresivo, es decir, el nivel de significancia real es inferior al nivel de significancia nominal y que tiene dificultades en detectar integración fraccional positiva, especialmente, en muestras pequeñas.

**Método de Sowell:** Al contrario de los procedimientos anteriores, el estimador de máxima verosimilitud propuesto por Sowell (1992a) permite obtener simultáneamente estimadores de todos los parámetros del modelo. Sea  $Y_T$  un vector  $T \times 1$  de observaciones generadas por el modelo (1) cuyos residuos tienen distribución Normal. En ese caso, el logaritmo de la función de verosimilitud está dado por

$$L(\gamma, Y_T) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_T| - \frac{1}{2} (Y_T' \Sigma_T^{-1} Y_T)$$

El estimador de máxima verosimilitud es obtenido maximizando (12) respecto al vector  $\gamma = (\Phi, \Theta, D)$  y es consistente y asintóticamente Normal. A pesar del atractivo teórico del estimador propuesto por Sowell, en cada iteración de maximización es necesaria la inversa de una matriz  $T \times T$ ,  $\Sigma_T$ , de ahí que Sowell haya utilizado una estructura Toeplitz de  $\Sigma_T$  para reducir el costo de cálculo. Además, este método exige que las raíces del polinomio autorregresivo sean distintos. En tanto, Sowell (1992a) argumenta que dicha hipótesis en la práctica no es una limitación dado que la probabilidad de que existan raíces repetidas en un polinomio es muy pequeña.

Sowell sugiere que a través de la estimación de un modelo ARFIMA para una serie en primeras diferencias, es posible distinguir un modelo con tendencia determinística de un paseo aleatorio con drift. Si la estimación de  $d$  estuviese próximo a -1 entonces los datos soportan un modelo con tendencia determinística. Si la estimación de  $d$  estuviese próxima a 0, entonces los datos soportan un proceso integrado de orden 1 con drift.

### 5.13 Test de Integración Fraccional con Estacionalidad

A pesar que la importancia de los procesos estacionales fraccionalmente integrados ya ha sido reconocida hace un tiempo por Porter-Hudak (1990), la investigación en esta área ha sido bastante escasa. El modelo ARFIMA estacional o ARFISMA (Autoregressive Fractionary Integrated Seasonal Moving Average) está dado por

$$\Phi(L)(1-L^s)^d y_t = \Theta(L)\epsilon_t, \quad \epsilon_t \rightarrow i.i.d(0, \sigma_\epsilon^2)$$

donde  $L$  es el operador de rezagos,  $d$  es el orden de diferenciación,  $\Phi(L)=1-\phi_1L-\dots-\phi_pL^p$ ,  $\Theta(L)=1+\theta_1L+\dots+\theta_qL^q$  y  $s$  es el número de períodos intra-anales. El proceso ARFISMA tiene un comportamiento de frecuencias estacionales idéntico al de un proceso ARFIMA en frecuencia cero. Supongamos  $s=4$  (datos trimestrales), entonces el modelo ARFISMA se reduce simplemente a  $\Phi(L)(1-L^4)^d y_t = \Theta(L)\epsilon_t$ . Hassler (1994) argumenta que el filtro  $(1-L^4)^d$  es un filtro rígido, en la medida en que las contribuciones de las oscilaciones semestrales y anuales y del comportamiento de largo plazo de la varianza del respectivo proceso son todas gobernadas por  $d$  común. De ahí que haya propuesto la utilización de un filtro flexible,  $(1-L)^{d_0}(1+L)^{d_1}(1+L^2)^{d_2}$ , para la modelación independiente de la importancia de las frecuencias estacionales. Así, el modelo ARFISMA está dado por

$$\Phi(L)(1-L)^{d_0}(1+L)^{d_1}(1+L^2)^{d_2} y_t = \Theta(L)\epsilon_t$$

en que  $d_0$ ,  $d_1$  y  $d_2$  son los órdenes de integración fraccional de las frecuencias  $0$ ,  $\pi/2$  y  $\pi$  respectivamente. El proceso estocástico  $y$  es estacionario e invertible si todas las raíces de  $\Phi(L)$  y de  $\Theta(L)$  se encuentran fuera del círculo unitario y si  $-0,5 < d_i < 0,5$  para todo  $i$ . Además, el proceso ARFISMA es un proceso de memoria larga cuando  $0 < d_i < 0,5$  para todo  $i$ .

**Test GPH Generalizado:** Hassler (1994) generaliza el test GPH para testear el parámetro fraccional en frecuencia cero y en frecuencias estacionales. Consideremos la siguiente regresión:

$$\ln(I_y(W_{i,j})) = \alpha_i - d_i \ln(4\sin^2(w_j/2)) + v_{i,j}, \quad i = 0, 1, 2 \quad y \quad j = 1, 2, \dots, k$$

donde  $I_y(W_{i,j})$  es el periodograma de  $y_t$  en frecuencia  $w_{i,j}$ .  $d_i$  son los parámetros de integración fraccional en frecuencias 0,  $\pi/2$ ,  $\pi$  respectivamente, pertenecientes a las frecuencias armónicas  $w_{i,j}$ :  $w_{0,j}=2\pi j/T$ ,  $w_{1,j}=\pi/2+2\pi j/T$ ,  $w_{2,j}=\pi-2\pi j/T$ , en que  $k$  es el número de frecuencias y  $T$  el tamaño de muestra.

Para testear  $H_0:d_i=0$  contra  $H_1:d_i \neq 0$ ,  $i=0,1,2$  en (15), el estadístico del test está definido por

$$\frac{\hat{d}_i}{\sqrt{\hat{\sigma}^2(\hat{d}_i)}}$$

donde  $\hat{d}_i$  es el estimador de mínimos cuadrados ordinarios de  $d_i$  y  $\hat{\sigma}^2(\hat{d}_i)$  es la varianza estimada de  $\hat{d}_i$ , definida por  $\pi^2/6S_R^2$  para las frecuencias 0 y  $\pi$ , y  $\pi^2/12S_R^2$  para la

frecuencia  $\pi/2$ , en que  $S_R^2 = \sum_{j=1}^k (R_j - \bar{R})^2$  y  $R_j = -\ln(4\sin^2(w_j/2))$ . La distribución asintótica de este test es Normal estándar.

**Test TDS:** El test TDS (Time Domain Store) fue derivado por Silvapulle (1996) para testear la existencia de integración fraccional en frecuencia cero y en frecuencias estacionales en series trimestrales. Consideremos el modelo (14) en que  $y_t$  es una serie observable o un error inobservable de la regresión  $x_t = \beta'z_t + y_t$ ,  $t=1,2,\dots,T$ ; donde  $z_t$  es un vector  $k \times 1$  de variables estocásticas o no estocásticas y  $\beta$  es un vector  $k \times 1$  de parámetros desconocidos. Sea  $\gamma=(d,\eta)$  con  $d=(d_0, d_1, d_2)$ ,  $\eta=(\Phi,\Theta,\beta,\sigma^2)$  y  $0=(0,0,0)$ . Para testear  $H_0:d=0$  contra  $H_1:d \neq 0$ , el estadístico está definido por

$$T_0 = n^{-1} s_d' l_{dd}^{-1} s_d$$

en que  $n=(T-2m)$ .  $T$  es el número de observaciones,  $s_d$  es la gradiente del logaritmo de la función de verosimilitud  $L(\gamma)$  para  $T$  observaciones. Esto es

$$s_d = \left( \frac{\partial L(\gamma)}{\partial d_0}, \frac{\partial L(\gamma)}{\partial d_1}, \frac{\partial L(\gamma)}{\partial d_2} \right)$$

con

$$\begin{aligned}\frac{\partial L(y)}{\partial d_0} &= \frac{1}{\hat{\sigma}^2} \sum_{t=1}^T \hat{\epsilon}_t \sum_{j=1}^m j^{-1} \hat{\epsilon}_{t-j} \\ \frac{\partial L(y)}{\partial d_1} &= \frac{1}{\hat{\sigma}^2} \sum_{t=1}^T \hat{\epsilon}_t \sum_{j=1}^m j^{-1} \hat{\epsilon}_{t-j} (-1)^{j-1} \\ \frac{\partial L(y)}{\partial d} &= \frac{1}{\hat{\sigma}^2} \sum_{t=1}^T \hat{\epsilon}_t \sum_{j=1}^m j^{-1} \hat{\epsilon}_{t-2j} (-1)^{j-1}\end{aligned}$$

en que  $\hat{\epsilon}_t$  son los residuos obtenidos de la estimación del modelo (14), bajo la hipótesis nula, por el método de mínimos cuadrados ordinarios,  $m$  es escogido de forma que  $\ln(1+L)$  pueda ser aproximado por los primeros  $m$  términos de su expansión. Además,  $t_{dd}$  es la matriz de información de  $s_d$  bajo  $H_0$ , en que la parte triangular inferior de  $t_{dd}$  está dada por

$$\begin{bmatrix} \sum_{j=1}^m j^{-2} & & \\ \sum_{j=1}^m (-1)^{j-1} j^{-2} & \sum_{j=1}^m j^{-2} & \\ \sum_{j=1}^{[m/2]} (-1)^{j-1} (2j^2)^{-1} & \sum_{j=1}^{[m/2]} (-1)^j (2j^2)^{-1} & \sum_{j=1}^m j^{-2} \end{bmatrix}$$

donde  $[m/2]$  es el mayor entero de  $m/2$  el estadístico  $T_0$  tiene como distribución asintótica una distribución  $\chi^2$  con 3 grados de libertad. Silvapulle (1996) muestra que  $T_0$  tiene propiedades deseables, tanto en términos de nivel de significancia como en términos de potencia, y que ésta aumenta a medida que  $d_1$  y/o  $T$  aumentan. En el caso particular en que se quiera testear  $H_0: d_1=0$  versus  $H_1: d_1 \neq 0$ , es decir, testear la existencia de integración fraccional en frecuencia cero con  $d_1=d_2=0$ , entonces el estadístico  $T_0$  vendrá dado por

$$T'_0 = \frac{\left( \frac{\partial L(y)}{\partial d_0} \right)}{\left[ \left( \sum_{j=1}^m j^{-2} \right) (T - m) \right]}$$

que permite una forma simple de testear la existencia de integración fraccional.



### 5.13 Filtros para Series de Tiempo

Como hemos discutido, muchas de las variables económicas más comúnmente usadas tienen tendencia. El problema de remover la tendencia de las series se abordó históricamente mediante una regresión contra una tendencia lineal. Cuando la variable es estacionaria, basta con usar técnicas de mínimos cuadrados para obtener estimadores consistentes de los parámetros. Se puede usar también tendencias cuadráticas, cúbicas, etc. El modelo más simple es:

$$y_t = \alpha + \beta t + \mu_t = \alpha + \beta t + \psi(L) \epsilon_t$$

donde  $\alpha$  y  $\beta$  son dos parámetros y  $\mu_t$  es una innovación que admite una representación media móvil para un modelo ARMA(p,q) sobre las innovaciones  $\epsilon_t$  sintetizado en  $\psi(L)$ . La remoción de tendencia correspondería a obtener una serie filtrada del tipo  $y_t - \hat{\alpha} - \hat{\beta}t$ .

Esta metodología es adecuada sólo si la tendencia es determinística, cosa que no sucede frecuentemente con los datos macroeconómicos. Es por ello que se han desarrollado diversos filtros que conviene revisar, a la luz de la discusión sobre raíces unitarias.

#### *Filtro de Hodrick y Prescott<sup>41</sup>*

Los autores del filtro parten de la definición de ciclo económico propuesta por Lucas (1976), es decir, lo definen como las fluctuaciones recurrentes en la actividad real respecto a una tendencia. Las fluctuaciones son por definición desviaciones respecto a un *camino suave pero variable* (tendencia), el cual es posible estimar mediante un proceso computacional que ajuste una curva suave a los datos. Matemáticamente:

$$\min_{\tau_t} \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2$$

El primer término es la suma de las desviaciones de la serie respecto a la tendencia al cuadrado, y es una medida del grado de ajuste. El segundo término es la suma de cuadrados de las segundas diferencias de los componentes de tendencia y es una medida del grado de suavidad. En otras palabras, el primer término recoge

41 Hodrick, R., and E.P. Prescott (1997), "Postwar Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29:1-16.

variaciones de la serie alrededor de su tendencia estocástica, en tanto que el segundo recoge variaciones en la tasa de cambio de dicha tendencia estocástica.

El parámetro  $\lambda$  controla cuánto de cada componente quedará en la variable filtrada y se le llama parámetro de suavidad. De esta manera, si el parámetro de suavidad  $\lambda \rightarrow 0$ , la tendencia será muy parecida a la serie original, en tanto que si  $\lambda \rightarrow \infty$ , la tendencia será lineal. Puede demostrarse que el valor de  $\lambda$  es igual a la ratio de la varianza del componente tendencial respecto a la varianza del componente cíclico.

La derivación del filtro es sencilla. Considere que se desea separar la serie  $y_t$  en dos componentes: tendencia ( $\tau$ ) y un componente irregular ( $\mu$ ). Naturalmente,  $y_t = \tau_t + \mu_t$ . Defina ahora la siguiente perturbación de la tendencia:

$$v_t = ((\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})) \quad \forall t=3, 4, \dots, T$$

o  $v_t = P\tau_t$  con

$$P = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

La descomposición de la serie original se obtiene minimizando la suma de cuadrados  $v'v + \lambda \mu' \mu = (y_t - \tau_t)'(y_t - \tau_t) + \lambda \tau' P' P \tau$  con respecto a  $\tau$ . Se obtiene la siguiente condición de primer orden  $(I + \lambda P' P)\tau = y$ . Como  $(I + \lambda P' P)$  es positiva definida, la condición de segundo orden se satisface siempre. La solución es, además, única y corresponde a  $(I + \lambda P' P)^{-1} y = \tau$ .<sup>42</sup>

### *El problema de la elección de $\lambda$*

Hodrick y Prescott propusieron un valor de  $\lambda = 1600$  para series trimestrales bajo el supuesto de que cualquier perturbación que tiene efectos durante 8 o más años tiene carácter permanente. En Estados Unidos los ciclos económicos tienen una duración media de 5 años. Para series mensuales se suele utilizar 14.400 y para series anuales se recomienda un valor igual a 10.

<sup>42</sup> Algunos autores encuentran que el filtro induce ciclos espúreos en las series filtradas (Harvey, A.C., Jaeger, A., 1993. "Detrending, stylized facts, and the business cycle". *Journal of Applied Econometrics* 8:231-247.). Esto es conocido como "efecto Slutsky".

El siguiente ejemplo ayuda a entender lo que hace el filtro de Hodrick y Prescott. Considere el caso en que  $T=5$ , entonces:

$$(I + \lambda P'P) = \begin{pmatrix} 1+\lambda & -2\lambda & -\lambda & 0 & 0 \\ -2\lambda & 1+5\lambda & 0 & -\lambda & 0 \\ -\lambda & 0 & 1+6\lambda & 0 & -\lambda \\ 0 & -\lambda & 0 & 1+5\lambda & 2\lambda \\ 0 & 0 & -\lambda & 2\lambda & 1+\lambda \end{pmatrix}$$

Si  $\lambda=7$ , entonces  $[I + 7P'P]^{-1}$  es:

$$\begin{pmatrix} \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \hat{\tau}_4 \\ \hat{\tau}_5 \end{pmatrix} = \begin{pmatrix} 0.644 & 0.375 & 0.156 & -0.014 & -0.161 \\ 0.375 & 0.322 & 0.216 & 0.100 & -0.014 \\ 0.156 & 0.216 & 0.254 & 0.216 & 0.156 \\ -0.014 & 0.100 & 0.216 & 0.322 & 0.375 \\ -0.161 & -0.014 & 0.156 & 0.375 & 0.644 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

Note los siguientes resultados interesantes:

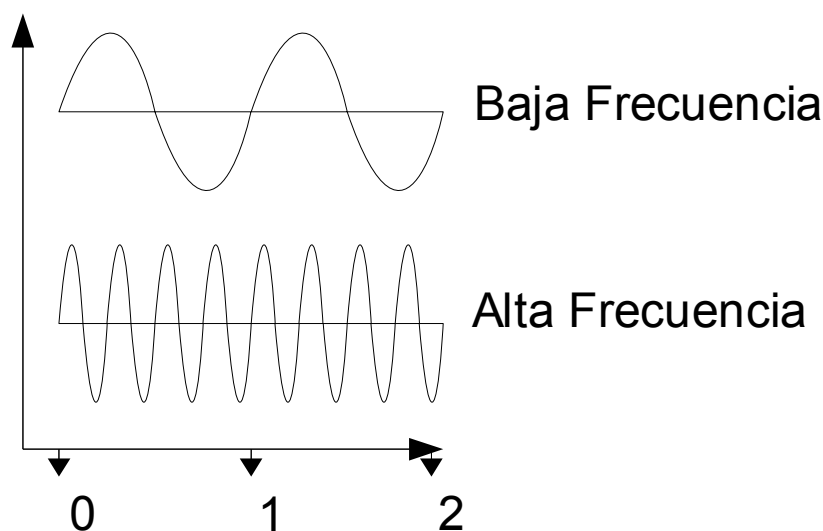
1. Para cada  $t$ , los pesos suman uno y no dependen de los datos.
2. Hay pesos negativos para algunas observaciones
3. Solo la tercera observación es filtrada de manera simétrica
4. Las observaciones efectivas de los puntos de partida y final reciben un gran peso (0.644) en la tendencia computada para las puntas de la serie.
5. La tendencia para las observaciones cerca de los extremos, sin embargo, le darán más peso a la primera y última observación (0.375) que a sí mismos (0.322).

Los puntos 4 y 5 ilustran el origen de los problemas del filtro de Hodrick-Prescott con los extremos. Cuando cambia el tamaño de muestra o se revisan los datos, cambia mucho la estimación de la tendencia en  $T-1$  y  $T$ .

### Filtro de Baxter y King

Utiliza el análisis de dominio de frecuencia y dos ideas subyacentes. La primera idea es que es posible descomponer toda la variación de una serie en tendencia (ciclos mayores a 8 años), ciclo de negocios (entre 1.5 y 8 años), y componentes irregulares (menos de 1.5 años). La segunda idea es que posible realizar una aproximación de estos componentes cíclicos usando funciones periódicas (p.e., trigonométricas tipo coseno).

Una función periódica tiene la característica de que  $F(x + \lambda) = F(x)$ , con  $\lambda \neq 0$ ,  $\lambda$  conocido como el período de la función. Sea  $\omega$  la frecuencia de la función. El periodo  $\lambda$  está relacionado con la frecuencia por  $\lambda = 2\pi/\omega$ . La cantidad de tiempo que le toma a la función para completar un ciclo es igual a  $(2\lambda/\omega)$ . Entonces existe una relación inversa entre la frecuencia y el período. Entre menor sea la frecuencia, mayor será la cantidad de períodos que abarque un ciclo. Si la frecuencia tiende a cero, el período tiende a infinito.



Entonces, si se quiere estudiar un ciclo que dura 1.5 año, fije  $\lambda=6$ , la frecuencia será  $\pi/3$ . Por otro lado, si se quiere estudiar un ciclo que dura 8 años, fije  $\lambda=32$ , la frecuencia será  $\pi/16$ .

Baxter y King definen el ciclo de negocios entre 1.5 y 8 años, entonces hay que promediar las frecuencias entre  $\pi/16$  y  $\pi/3$  y eliminar aquellas entre 0 y  $\pi/16$  (tendencia) y  $\pi/16$  y  $\pi$  (irregulares). Por eso se le llama, band-pass filter. El componente cíclico corresponde a un promedio móvil de la forma  $x_t^c = \sum_{j=-\infty}^{\infty} \beta_j (x_{t+j} - \mu_x)$  con

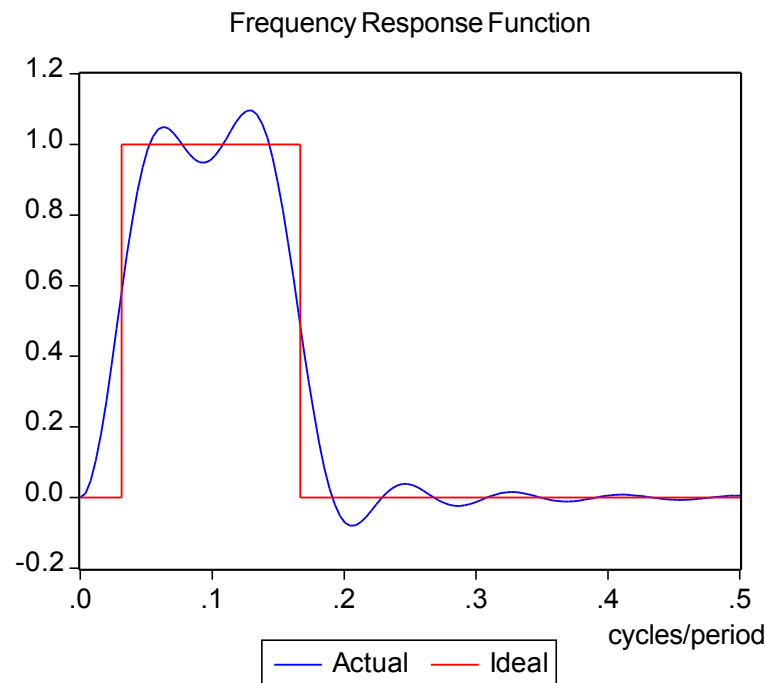
ponderadores  $\beta_0 = \frac{\sin \omega_{\pi/3} - \sin \omega_{\pi/16}}{\pi}$  si  $j=0$  y  $\beta_j = \frac{\sin j \omega_{\pi/3} - \sin j \omega_{\pi/16}}{\pi j}$  si  $j \neq 0$ .

En muestra finita hay que seleccionar el máximo “j”. No existe un óptimo, pero entre más rezagos se incorporen en el promedio móvil, mejor será la aproximación con el filtro ideal ... a costa de una mayor pérdida de datos por encima y por debajo del valor de interés.

Según Baxter y King, un método óptimo de extracción de ciclos económicos debe cumplir con 6 objetivos:

- El filtro debe extraer un rango específico de periodicidades, sin variar sus propiedades inherentes (varianza, correlaciones y otras medidas descriptivas de los datos).
- No debe producir un movimiento de fase (es decir, que no altere las relaciones temporales de las series a ninguna frecuencia). Este aspecto y el anterior definen un promedio móvil ideal, con ponderaciones simétricas para rezagos y adelantos.
- El método debe ser una aproximación óptima (porque es muestra finita) de un filtro ideal. Esto se puede determinar midiendo la diferencia de los resultados obtenidos con un filtro ideal y uno aproximado.
- La aplicación de un filtro debe producir una serie de tiempo estacionaria cuando se aplica a cifras que presentan tendencia.
- El método debe ser independiente de la longitud de la serie.
- El método debe ser operacional, esto es, de fácil aplicación y uso.

### Filtro de Baxter-King para frecuencias entre 6 y 32 trimestres



*Christiano y Fitzgerald*<sup>43</sup>

- Desarrollan un filtro band-pass pero permiten que la elección de leads y lags sea asimétrica y en función de los datos.
- Si se desea filtrar la serie  $x$  con periodos de oscilación entre  $2 \leq p_l \leq p_u \leq \infty$ , el filtro CF es

<sup>43</sup> The Band Pass Filter, Lawrence J. Christiano and Terry J. Fitzgerald, Working Paper 9906, Federal Reserve Bank of Cleveland.

$$\hat{y}_t = B_0 x_t + B_1 x_{t+1} + \dots + B_{T-1-t} x_{T-1} + \tilde{B}_{T-t} x_T \\ + B_1 x_{t-1} + \dots + B_{t-2} x_2 + \tilde{B}_{t-1} x_1, \quad (1.2)$$

for  $t = 3, 4, \dots, T-2$ . In (1.2),

$$B_j = \frac{\sin(jb) - \sin(ja)}{\pi j}, \quad j \geq 1 \quad (1.3)$$

$$B_0 = \frac{b-a}{\pi}, \quad a = \frac{2\pi}{p_u}, \quad b = \frac{2\pi}{p_l}.$$

con  $\bar{B}_{T-t} = \sum_{j=T-t}^1 B_j$  y  $\bar{B}_{T-1} = \sum_{j=1}^t B_j$

- Note que el filtro es asimétrico y que depende del tiempo, en el sentido que los datos determinan los pesos.

### Refinamientos

- Muchos autores proponen refinamientos del filtro de Hodrick y Prescott para corregir sus potenciales deficiencias.
- Para datos anuales, Hodrick Prescott sugieren usar  $\lambda=100$ 
  - Baxter y King prefieren usar  $\lambda=10$ .
  - Ravn y Uhlig (2002)<sup>44</sup> indican que una mejor manera de escoger  $\lambda$  es  $\left(\frac{T}{4}\right)^k * 1600$  con T el largo de la muestra (Si usa  $k=2$  se obtiene HP). Esto implica que óptimo es  $\lambda=6.25$
- Schlicht (2004)<sup>45</sup> propone un método de momentos para estimar el parámetro  $\lambda$ , sobre la base que éste es la razón entre la varianza del componente de tendencia y la varianza del componente cíclico

44 Ravn, M. O., and H. Uhlig (2002). "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations, *Review of Economics and Statistics*, 84: 371-376.

45 Schlicht, E. "Estimating the Smoothing Parameter in the So-Called Hodrick-Prescott Filter", Discussion paper 2004-2, Department of Economics, University of Munich

- Mise et al (2005)<sup>46</sup> demuestran que el HP no funciona bien en los extremos de la muestra y proponen como solución realizar una proyección fuera de ambos lados de la muestra usando un modelo ARIMA.

---

<sup>46</sup> Mise, E. T.H Kim and P. Newbold "On suboptimality of the Hodrick-Prescott filter at time series endpoints" *Journal of Macroeconomics* 27 (2005) 53-67



## Apéndice A: Códigos Gauss

```

/*      Figura 5.1 Construye y grafica un proceso TS y uno DS      */
new;
cls;
library pgraph;
len=150;
shock=rndn(len,1);          /* Crea data ruido blanco */
vara=zeros(len,1);         /* Define variable */
varb=vara;
drft=0.05;                 /* Define drift */
vara[1,1]=0;               /* Define punto de partida */
i=2;
tita="Proceso I(1)";       /* Define titulos */
titb="Proceso I(0)";
do while i<=len;
    vara[i,.]=drft+vara[i-1,.]+shock[i,1];          /* Construye el random-walk */
    varb[i,1]=vara[1,1]+0.85*varb[i-1,1]+shock[i,1]; /* Construye el TS */
    i=i+1;
endo;

mono(vara,varb);
end;
proc(0) = mono(data,rho);
graphset;
    _pnumht={0.25};          /* tamaño de numeros */
    _pdate="";              /* omite fecha */
    _plwidth = 3;           /* ancho de lineas */
    _pltype={6};           /* tipo de lineas */
    _ptitlht={0.35};       /* tamaño titulo */
begwind;
    window(2,1,0);
    title(tita);
    xy(seqa(1,1,rows(data)),data);
nextwind;
    title(titb);
    xy(seqa(1,1,rows(rho)),rho);
endwind;
ret;
endp;

/*      Figura 5.2 Computa la distribución del estimador de mínimos cuadrados del AR(1) cuando los datos
fueron generados por un paseo aleatorio */

new;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1);              /* pide largo de datos desde la consola */
result=zeros(1000,2);
i=1;
do while i<=1000;
    shock=rndn(len,1);      /* crea data white noise */
    var1=zeros(len,1);      /* crea variables */
    j=2;
    do while j<=len;
        var1[j,.]=var1[j-1,.]+shock[j,1];          /* crea random walks con drift */
        j=j+1;
    endo;
    {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a}=myols(var1[2:rows(var1),1],var1[1:rows(
var1)-1,1]);
    result[i,1]=bhat;

```

```

    result[i,2]=(bhat-1)/sqrt(covb);
    i=i+1;
  endo;
  mono(result);
end;

```

*/\* Figura 5.3 Computa la distribución del estimador de mínimos cuadrados del AR(1) con constante cuando los datos fueron generados por un paseo aleatorio \*/*

```

new;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1); /* pide largo de datos desde la consola */
result=zeros(1000,2);
i=1;
do while i<=1000;
    shock=rndn(len,1); /* crea data white noise */
    var1=zeros(len,1); /* crea variables */
    j=2;
    do while j<=len;
        var1[j,.]=var1[j-1,.]+shock[j,1]; /* crea random walks con drift */
        j=j+1;
    endo;
    {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a}=myols(var1[2:rows(var1),1],ones(rows(var1)-1,1)~var1[1:rows(var1)-1,1]);
    result[i,1]=bhat[2,1];
    result[i,2]=(bhat[2,1]-1)/sqrt(covb[2,2]);
    i=i+1;
  endo;
  mono(result);
end;

```

*/\* Figura 5.4 Computa la distribución del estimador de mínimos cuadrados del AR(1) con constante cuando los datos fueron generados por un paseo aleatorio con drift\*/*

```

new;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1); /* pide largo de datos desde la consola */
result=zeros(1000,2);
i=1;
do while i<=1000;
    shock=rndn(len,1); /* crea data white noise */
    var1=zeros(len,1); /* crea variables */
    j=2;
    do while j<=len;
        var1[j,.]=0.05+var1[j-1,.]+shock[j,1]; /* crea random walks con drift */
        j=j+1;
    endo;
    {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a}=myols(var1[2:rows(var1),1],ones(rows(var1)-1,1)~var1[1:rows(var1)-1,1]~seqa(1,1,rows(var1)-1));
    result[i,1]=bhat[2,1];
    result[i,2]=(bhat[2,1]-1)/sqrt(covb[2,2]);
    i=i+1;
  endo;
  mono(result);
end;

```

*/\* Cuadro 5.2 Computa los valores críticos de la distribución Dickey-Fuller \*/*

```

new;

```

```

cls;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1); /* pide largo de datos desde la consola */
"Dame el numero de replicaciones " ;
replic=con(1,1);
result=zeros(replic,3);
i=1;
do while i<=replic;
  shock=rndn(len,1); /* crea data white noise */
  var1=zeros(len,1); /* crea variables */
  j=2;
  do while j<=len;
    var1[j,.]=var1[j-1,.]+shock[j,1]; /* crea random walks con drift */
    j=j+1;
  endo;
  deltay=var1[2:rows(var1),1]-var1[1:rows(var1)-1,1];
  {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a} = myols(deltay,var1[1:rows(var1)-
1,1]);
  result[i,1]=that;
  {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a}=myols(deltay,ones(rows(var1)-
1,1)~var1[1:rows(var1)-1,1]);
  result[i,2]=that[2,1];
  {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a}=myols(deltay,ones(rows(var1)-
1,1)~var1[1:rows(var1)-1,1]~sega(1,1,rows(var1)-1));
  result[i,3]=that[2,1];
  i=i+1;
endo;
result=sortc(result,1);
index1=trunc(0.05*rows(result));
index2=trunc(0.10*rows(result));
"Valores criticos del Test de Dickey-Fuller Modelo 1
95% 90%";
result[index1,1]~result[index2,1];
result=sortc(result,2);
index1=trunc(0.05*rows(result));
index2=trunc(0.10*rows(result));
"Valores criticos del Test de Dickey-Fuller Modelo 2
95% 90%";
result[index1,2]~result[index2,2];
result=sortc(result,3);
index1=trunc(0.05*rows(result));
index2=trunc(0.10*rows(result));
"Valores criticos del Test de Dickey-Fuller Modelo 3
95% 90%";
result[index1,3]~result[index2,3];
end;

/* Computa Test de Zivot y Andrews */

proc zivot(var,model);
local dvar,lvar,i,var,criter,const,trend,right,sum,brk,linea,du,dt,dtu,x,cval1,
cval2,dvar,right;

dvar=trimr((var-lag1(var)),1,0);
lvar=trimr(lag1(var),1,0);
if lags >0;
  i=1;
  do while i<=lags; @ generates lags @
    dvar=dvar~shiftr(dvar[.,cols(dvar)]',1,(miss(0,0)))';
    i=i+1;
  endo;

```

```

endif;

dvar=trimr(dvar,lags,0);           @ trims matrices @
var=trimr(var,lags+1,0);
lvar=trimr(lvar,lags,0);

criter=ones(rows(dvar),2);        @ counters @
const=ones(rows(dvar),1);
trend=seqa(1,1,rows(dvar));

if lags==0;
    right=const~trend~lvar;        @ generates RHS @
else;
    right=const~trend~dvar[:,2:cols(dvar)]~lvar;
endif;

sum=olss(var,right);              @ are lags OK? @
"Durbin-Watson of Total Sample = " sum[1,2];
if sum[1,2]<=1.66;
    "Need more lags.";
end;
endif;

brk=trm+lags;                     @ moving break point @
linea=ones(rows(trend)-trm-lags,1);

do while brk<=rows(dvar)-1;
    du=trimr((zeros(brk,1)|ones(rows(dvar)-brk+1,1)),1,0); @ generates dummies @
    dt=trimr((zeros(brk,1)|seqa(1,1,rows(dvar)-brk+1)),1,0);
    dtu=du~dt;

    if model==1;x=du~right;cval1=linea*(-3.45);cval2=linea*(-5.58);
    elseif model==2;x=dt~right;cval1=linea*(-3.68);cval2=linea*(-5.11);
    elseif model==3;x=dtu~right;cval1=linea*(-3.95);cval2=linea*(-5.82);
    endif;

    criter[brk,:]=olss(var,x);
    brk=brk+1;
end;
retp(criter);

endp;

/* Computa el test KPSS */

new;
cls;
load data[86,38] = datagauss.txt;

"Tipo de Test? (0=constante, 1=constante y tendencia deterministica ";;tipo=con(1,1);

y=(data[:,13]);

if tipo .eq 0;
    x1=ones(rows(data),1);
else;
    x1=ones(rows(data),1)~seqa(1,1,rows(data));
endif;

res=y-x1*(inv(x1'x1)*x1'y);        /* genera residuos */
res_sq=res.^2;
s=cumsumc(res);                    /* Genera residuos acumulados */

```

```

s_sq=s.^2;

bartlett_win=10;                                /* Genera estimador del espectro residual en frecuencia cero
*/
integers=seqa(1,1,bartlett_win);
w=ones(bartlett_win,1)-integers./(bartlett_win+1);

sum=0;
nobs=rows(res);

for l (1,bartlett_win,1);
    row=l;
    res_lag=lagn(res,l);
    res_unlag=trimr(res,row,0);
    res_lag=trimr(res_lag,row,0);
    sum_temp=w[l]*(res_unlag'*res_lag);
    sum=sum_temp+sum;
endfor;
sum_dub=sum;

sig_sq=inv(nobs)*sumc(res_sq)+2*inv(nobs)*sum_dub;
LM=1/(nobs*nobs)*sumc(s_sq)/sig_sq;

"El valor de LM es: " LM;
end;

```

*/\* Computa el estimador no paramétrico V de Cochran \*/*

```

proc vvv(yv);
library auto;
local rhoat,v,kv,t,m ;
t = rows(yv);                                /* NUMERO DE DATOS USADOS */
m = round(t/2);                                /* MAX DE AUTOCORRELATIONS */
yv = yv-li(yv,1) ;
yv = yv[2:rows(yv),1] ;
v = ones(1,3)|zeros(m-1,3) ;
kv = 1 ;
do while kv < m ;
    rhoat = ((ONES(kv,1)-(1/(rows(yv)-1))*SEQA(1,1,kv))^(1))
        .*AUTOCOR(yv,1,kv) ;
    v[1+kv,1] = 1+2*SUMC((ONES(kv,1)-(1/(kv+1))*SEQA(1,1,kv)).*rhoat) ;
    v[1+kv,2] = v[1+kv,1]*(1+(1/SQRT(0.75*(rows(yv)-1)/(kv+1)))) ;
    v[1+kv,3] = v[1+kv,1]*(1-(1/SQRT(0.75*(rows(yv)-1)/(kv+1)))) ;
    kv = kv+1 ;
enddo ;
retp(v) ;
endp ;
proc li(z,i);
local z,j;
j=1;do while j<=i;z=lagn(z);j=j+1;enddo;
retp(z);
endp;

```

## Apéndice B: Ejercicios

1. En general, hay dos modelos rivales para series macroeconómicas:

$$\begin{aligned}(a) \quad y_t &= \alpha + \beta t + \psi(L)\epsilon_t \\ (b) \quad (1-L)y_t &= \delta + \psi(L)\epsilon_t\end{aligned}$$

donde  $t$  es una tendencia determinística,  $\psi(L)$  es un modelo ARMA( $p, q$ ) y  $\epsilon$  es ruido blanco. Obtenga el error cuadrático medio de la predicción de  $y_t$  para el período  $t+s$  en estos modelos. ¿Cuál es la diferencia fundamental entre ellos? ¿Qué tan distintos son estos procesos cuando  $s \rightarrow \infty$ ?

2. Suponga que el modelo es  $y_t = \alpha + \delta z_t + \epsilon_t$ , donde  $z_t$  contiene dos variables: una variable estacionaria ( $w_t$ ) y una variable que crece con una tendencia exactamente lineal ( $x_t$ ). Demuestre que la distribución de los estimadores de  $\delta$  degenera cuando  $t \rightarrow \infty$ . ¿por qué en este modelo  $T^{3/2} \hat{\delta}$  es superconsistente?

3. Suponga que  $y_t$  es un paseo aleatorio pero Ud. estima un AR(1),  $y_t = \rho y_{t-1} + \epsilon_t$ , por mínimos cuadrados ordinarios. La desviación límite del estimador mínimos cuadrados ordinarios de su verdadero valor es

$$\lim_{T \rightarrow \infty} T(\hat{\rho}_T - 1) = \frac{\frac{1}{T} \sum y_{t-1} \mu_t}{\frac{1}{T^2} \sum y_{t-1}^2} = \frac{1/2 \sigma^2 \cdot [W(1)^2 - 1]}{\sigma^2 \cdot \int_0^1 [W(r)]^2 dr}$$

donde  $W(1)$  es un proceso Browniano. ¿Cómo se afecta el test del parámetro?

## Contenidos

Capítulo 5.....	1
Modelos de Series de Tiempo .....	1
Univariadas No Estacionarias.....	1
5.01 Representaciones de Tendencias en Análisis Univariado.....	1
5.02 Teoría Asintótica para Procesos con Tendencias Determinísticas.....	10
5.03 Teoría Asintótica para Procesos con Tendencias Estocásticas.....	13
5.04 Movimientos Brownianos y Procesos de Wiener.....	15
5.05 Funcional del Teorema del Límite Central.....	18
5.06 Tests Empíricos Estándares de Raíces Unitarias.....	32
5.07 Tests con hipótesis nula de estacionariedad.....	39
5.08 Equivalencia Observacional, Poder y Tests de Raíces Unitarias.....	40
5.09 Tests No Paramétricos de Raíces Unitarias.....	42
5.10 Tests de Raíces Unitarias Estacionales.....	44
5.11 Integración Fraccional.....	48
5.12 Tests de Integración Fraccional.....	49
5.13 Filtros para Series de Tiempo.....	57
Apéndice A: Códigos Gauss.....	62
Apéndice B: Ejercicios.....	67

Ultima corrección: 06/03/10