

Introducción al

Análisis de Series Temporales

José Alberto Mauricio

Universidad Complutense de Madrid

OBSERVACIONES PRELIMINARES

El origen de este libro es una colección de transparencias que preparé durante los años 2002 a 2005 para impartir unas 30 horas de clase de la asignatura "Series Temporales", correspondiente al último curso de la Licenciatura en Economía (Especialidad en Análisis Económico) de la Universidad Complutense de Madrid.

En la página viii (después del Contenido) figura la bibliografía (actualizada) que utilicé para elaborar el programa de mi parte de la asignatura y las transparencias.

Todos los cálculos y las figuras que aparecen en este libro se han elaborado con la versión 4.1 del programa EViews (Jul 29 2004 build). Los datos (archivos *.wfl) y los archivos de instrucciones (*.prg) utilizados se encuentran en el archivo IAST-DATOS.ZIP que se distribuye junto con este libro.

Agradezco sinceramente la colaboración en muchos sentidos de Sonia Sotoca, con quien compartí la docencia de "Series Temporales" durante los años mencionados y con quien sigo compartiendo despacho y multitud de tareas docentes y académicas, espero que por mucho tiempo ...

Última revisión: Marzo de 2007

CONTENIDO

1	Introducción	1
1.1	Series temporales	1
1.2	Procesos estocásticos	9
	Procesos estacionarios.....	11
	Procesos no estacionarios.....	12
1.3	Modelos.....	13
1.4	Aplicaciones	33
2	Modelos Univariantes	42
2.1	Introducción	42
2.2	Procesos estocásticos estacionarios.....	47
	ACF – PACF teóricas.....	49
	ACF – PACF muestrales.....	54

2.3	Modelos ARMA	63
	Estacionariedad – Invertibilidad	65
	ACF – PACF teóricas en modelos ARMA.....	67
2.4	Procesos estocásticos no estacionarios	80
	No estacionariedad en varianza – Transformación de Box-Cox	81
	No estacionariedad en media – Tendencias.....	86
2.5	Modelos ARIMA	95
	Tendencias en modelos invertibles.....	96
	Tendencias en modelos no invertibles.....	98
2.6	Elaboración de modelos ARIMA	100
	Identificación	100
	Estimación.....	102
	Diagnosís	103
	Criterios de información	106
	Contraste de no estacionariedad de Shin-Fuller	108
	Contrastes de no invertibilidad de Davis-Chen-Dunsmuir	109

2.7	Previsión con modelos ARIMA	120
	Previsiones puntuales – Intervalos de confianza.....	130
	Criterios de evaluación de previsiones	132
2.8	Heteroscedasticidad condicional autorregresiva	138
	Modelos ARCH – GARCH	139
	Contrastes de heteroscedasticidad condicional	142
	Previsión de la varianza condicional con un modelo GARCH(1,1).....	145
3	Modelos Multivariantes Estacionarios	146
3.1	Introducción	146
3.2	Modelos de regresión con perturbaciones ARMA	159
	Análisis de intervención	160
3.3	Modelos de retardos distribuidos en Econometría	170
	Modelos ADL.....	171
	Modelos de corrección de error	172

3.4	Modelos VAR estacionarios.....	178
	Identificación – Estimación – Contrastes de hipótesis.....	184
	Previsión	187
	Funciones de respuesta a impulsos	191
	Descomposición de las varianzas de los errores de previsión	193
4	Modelos Multivariantes No Estacionarios	201
4.1	Introducción	201
4.2	Raíces unitarias	211
	Paseo aleatorio	212
	Regresión con variables explicativas no estacionarias	213
	Regresión espuria	214
4.3	Contrastes de raíces unitarias.....	219
	Contrastes DF (Dickey-Fuller) – ADF	222
	Alternativas a los contrastes ADF	230
4.4	Modelos VAR con raíces unitarias	232

4.5	Estimación de relaciones de cointegración	245
	Estimación directa.....	245
	Estimación a través de un modelo de corrección de error	246
	Inferencia en regresiones con series $I(1)$	249
	Estimación a través de un modelo VAR – VEC	253
4.6	Contrastes de cointegración	261
	Contraste de Engle-Granger.....	261
	Contrastes basados en un modelo VAR – VEC	264
	Contrastes de hipótesis sobre vectores de cointegración	269
4.7	Elaboración de modelos multivariantes en la práctica	273

BIBLIOGRAFÍA

LIBROS DE TEXTO

- [01] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994), *Time Series Analysis – Forecasting and Control* (3rd edition), Prentice Hall.
- [02] Peña, D. (2005), *Análisis de Series Temporales*, Alianza.
- [03] Quantitative Micro Software (2002), *EViews 4 User's Guide* (revised for EViews 4.1), QMS-LLC.
- [04] Reinsel, G.C. (1997), *Elements of Multivariate Time Series Analysis* (2nd edition), Springer.

MANUALES COMPLEMENTARIOS

- [05] Brockwell, P.J., Davis, R.A. (2002), *Introduction to Time Series and Forecasting* (2nd edition), Springer.

- [06] Brooks, C. (2002), *Introductory Econometrics for Finance*, Cambridge.
- [07] Davidson, R., MacKinnon, J.G. (2004), *Econometric Theory and Methods*, Oxford.
- [08] Enders, W. (2004), *Applied Econometric Time Series* (2nd edition), Wiley.
- [09] Gandolfo, G. (1997), *Economic Dynamics* (3rd edition), Springer.
- [10] Johnston, J., DiNardo, J. (1997), *Econometric Methods* (4th edition), McGraw-Hill.
- [11] Kennedy, P. (2003), *A Guide to Econometrics* (5th edition), Blackwell.
- [12] Mills, T.C. (1990), *Time Series Techniques for Economists*, Cambridge.
- [13] Mills, T.C. (1999), *The Econometric Modelling of Financial Time Series* (2nd edition), Cambridge.
- [14] Patterson, K. (2000), *An Introduction to Applied Econometrics – A Time Series Approach*, MacMillan.
- [15] Verbeek, M. (2004), *A Guide to Modern Econometrics* (2nd edition), Wiley.

1

INTRODUCCIÓN

1.1 SERIES TEMPORALES

1.1.1 Definición

Una **serie temporal** (o simplemente una **serie**) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie **univariante** o **escalar**) o sobre varias características (serie **multivariante** o **vectorial**) de una unidad observable en diferentes momentos.

Representaciones matemáticas frecuentes de series temporales univariantes:

y_1, y_2, \dots, y_N ; $(y_t)_{t=1}^N$; $(y_t : t = 1, \dots, N)$, donde y_t es la observación n° t ($1 \leq t \leq N$) de la

serie y N es el número de observaciones de que consta la serie completa (el **tamaño** o la **longitud** de la serie). Las N observaciones y_1, y_2, \dots, y_N pueden recogerse en un vector columna $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ de orden $N \times 1$.

Representaciones matemáticas frecuentes de series temporales multivariantes:

$\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$; $(\mathbf{y}_t)_{t=1}^N$; $(\mathbf{y}_t : t = 1, \dots, N)$, donde $\mathbf{y}_t \equiv [y_{t1}, y_{t2}, \dots, y_{tM}]'$ ($M \geq 2$) es la observación n.º t ($1 \leq t \leq N$) de la serie y N es el número de observaciones de que consta la serie completa. Las N observaciones $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ pueden recogerse en una matriz \mathbf{Y} de orden $N \times M$:

$$\mathbf{Y} \equiv \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_N \end{bmatrix} \equiv \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & y_{22} & \cdots & y_{2M} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NM} \end{bmatrix}, \quad (1.1)$$

donde y_{tj} es la observación n.º t ($1 \leq t \leq N$) sobre la característica o variable n.º j ($1 \leq j \leq M$), que es la misma en todo momento t .

1.1.2 Ejemplo – Series temporales

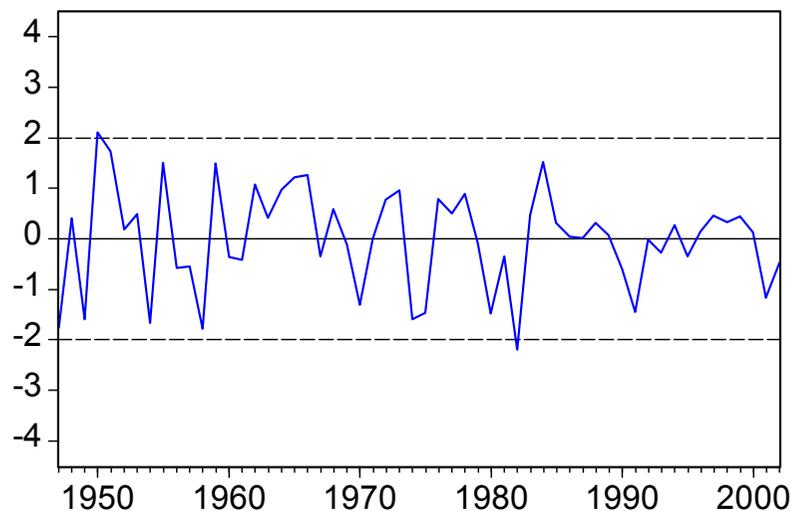


Figura 1.1

Tasa logarítmica de variación anual del PIB real en EEUU
1947 – 2002
(consumption.wf1)

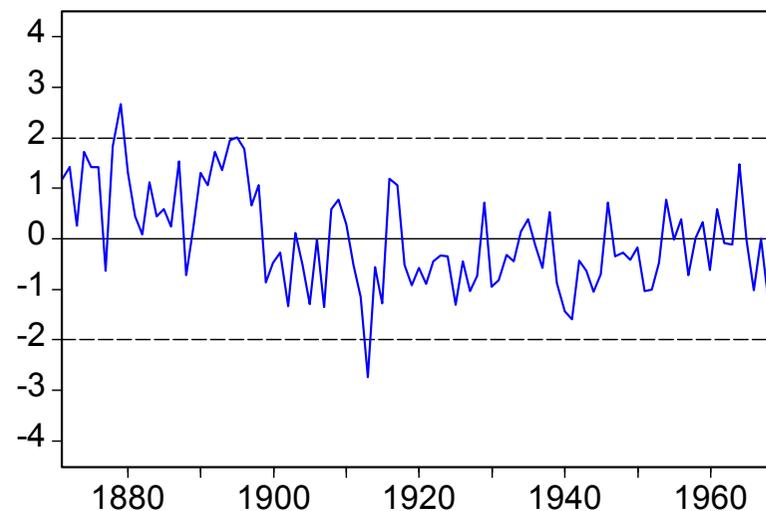
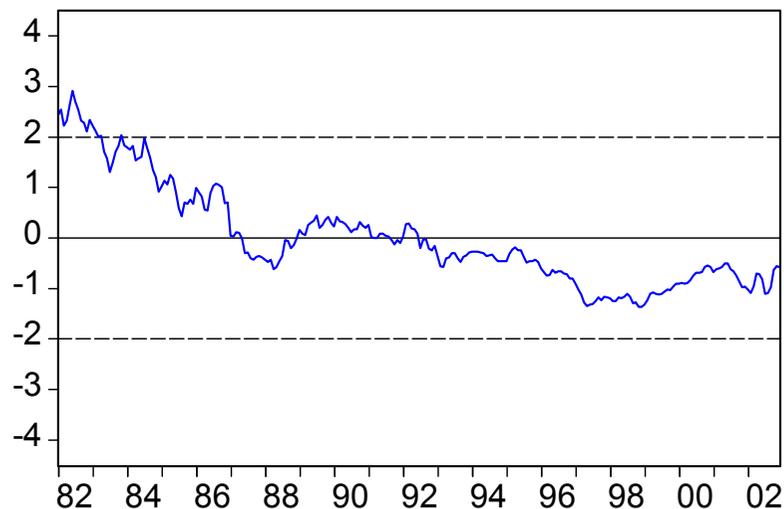


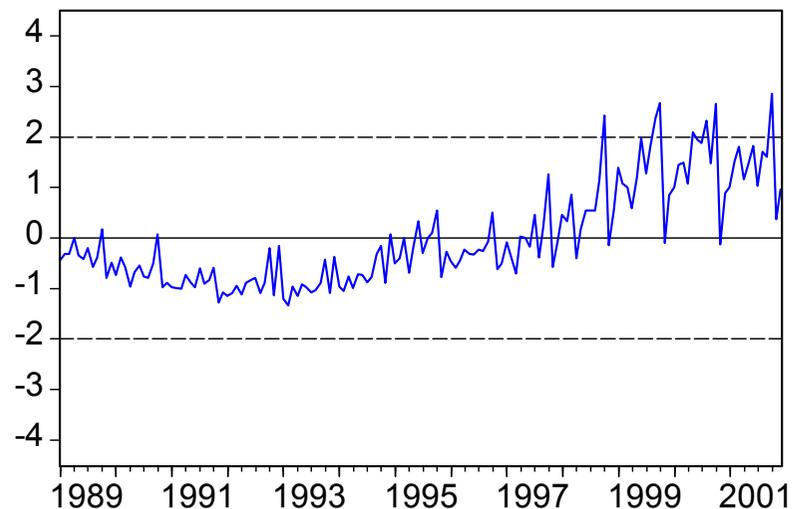
Figura 1.2

Volumen del flujo anual del río Nilo en Asuán
1871 – 1970
(nilo.wf1)

Observación 1: La serie de la Figura 1.1 evoluciona alrededor de un **nivel** constante y presenta una **variabilidad** constante alrededor de dicho nivel. Por el contrario, el nivel de la serie de la Figura 1.2 no parece constante; en concreto, se aprecia que dicho nivel podría haber cambiado de manera permanente a partir del año 1899.

**Figura 1.3**

Inflación interanual en España
Enero 1982 – Diciembre 2002
(exchange-rates.wf1)

**Figura 1.4**

Viviendas iniciadas en España
Enero 1989 – Diciembre 2001
(seasonal.wf1)

Observación 2: La serie de la Figura 1.3 evoluciona alrededor de un nivel que cambia sin seguir aparentemente un patrón concreto; por este motivo, suele decirse que dicha serie presenta una **tendencia estocástica** (aleatoria) en su evolución temporal. El caso de la serie de la Figura 1.4 es semejante; además, la variabilidad local de dicha serie es tanto mayor cuanto más elevado es su nivel local.

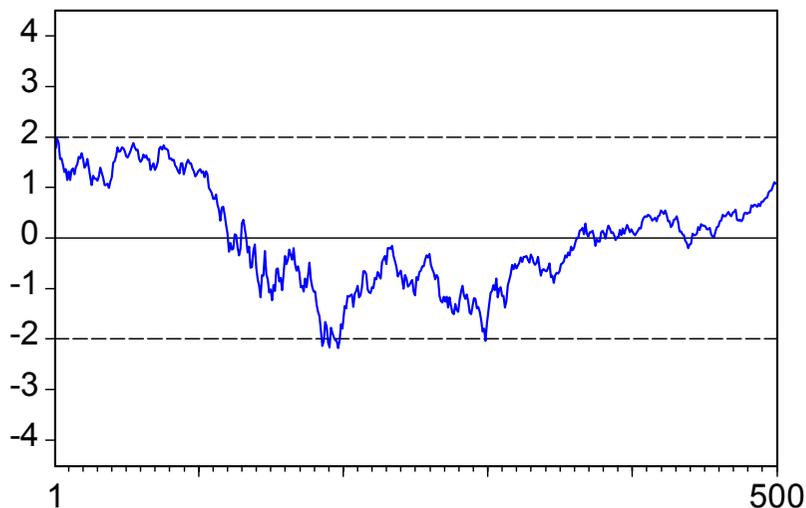


Figura 1.5

IBEX 35 (log) al cierre en la Bolsa de Madrid
2 de enero de 2002 – 30 de diciembre de 2003
(bolsa-madrid.wf1)

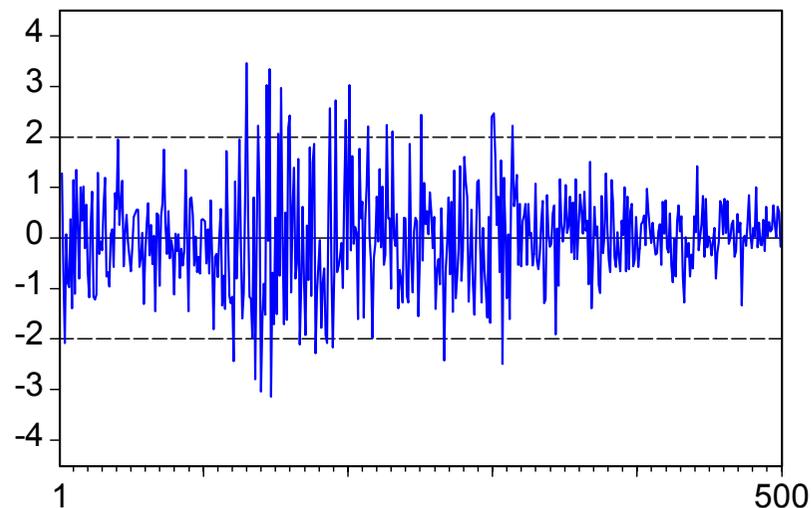


Figura 1.6

Tasa logarítmica de variación diaria del IBEX 35
3 de enero de 2002 – 30 de diciembre de 2003
(bolsa-madrid.wf1)

Observación 3: La serie de la Figura 1.5 evoluciona alrededor de un nivel que cambia sin seguir un patrón concreto, es decir, dicha serie presenta una tendencia estocástica. La serie de la Figura 1.6 presenta un nivel constante; sin embargo, las desviaciones grandes (pequeñas) de dicha serie con respecto a su nivel van seguidas de variaciones grandes (pequeñas), por lo que su **volatilidad condicionada** por su historia reciente presenta algún tipo de inercia.

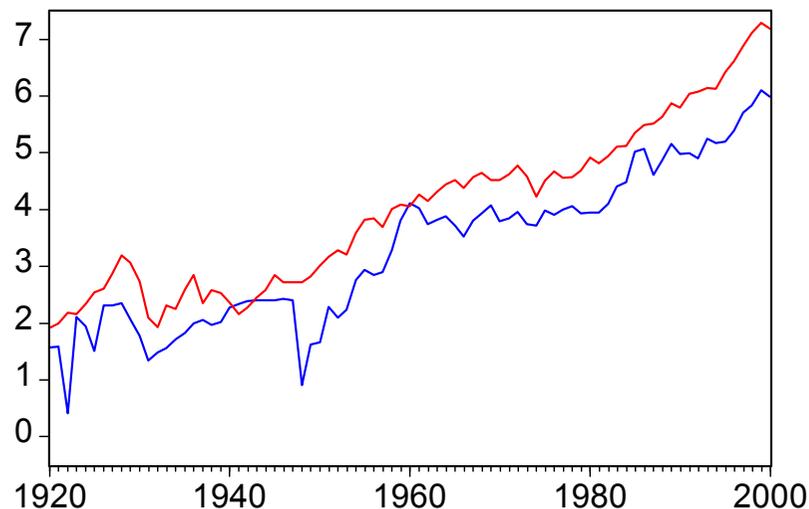


Figura 1.7

Índices bursátiles anuales en Alemania y EEUU
1920 – 2000
(stock-indices.wf1)

■ log CDAX Composite Price Index (Alemania)
■ log S&P 500 Composite Index (EEUU)

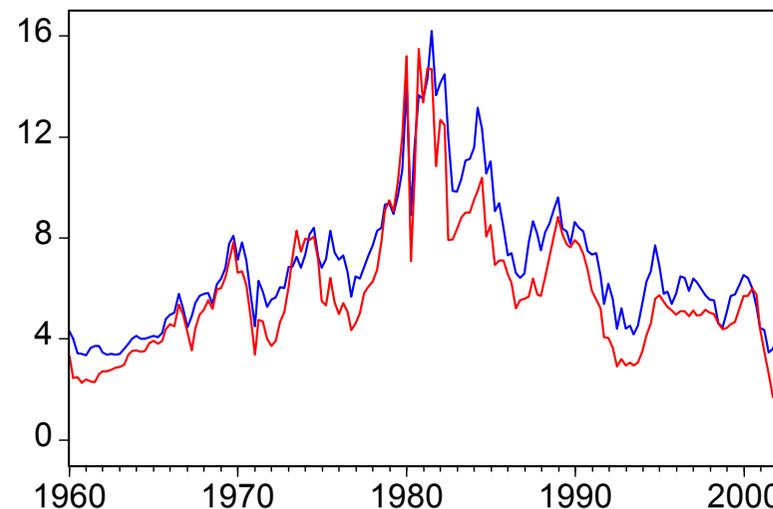
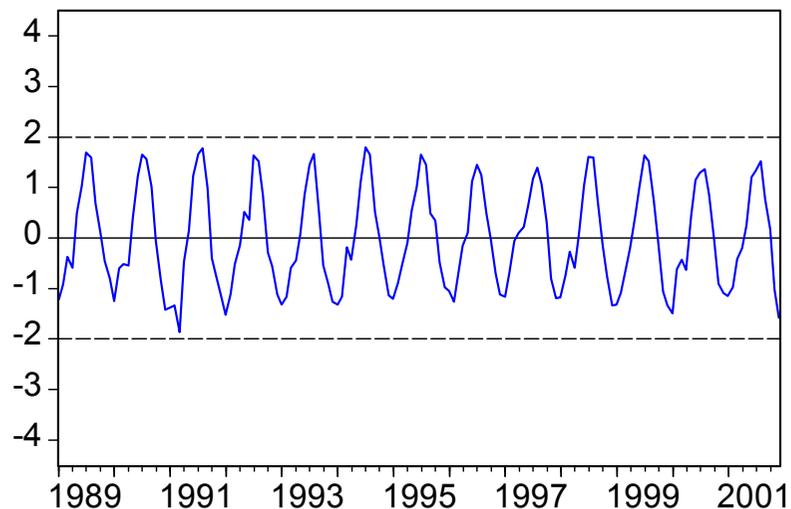


Figura 1.8

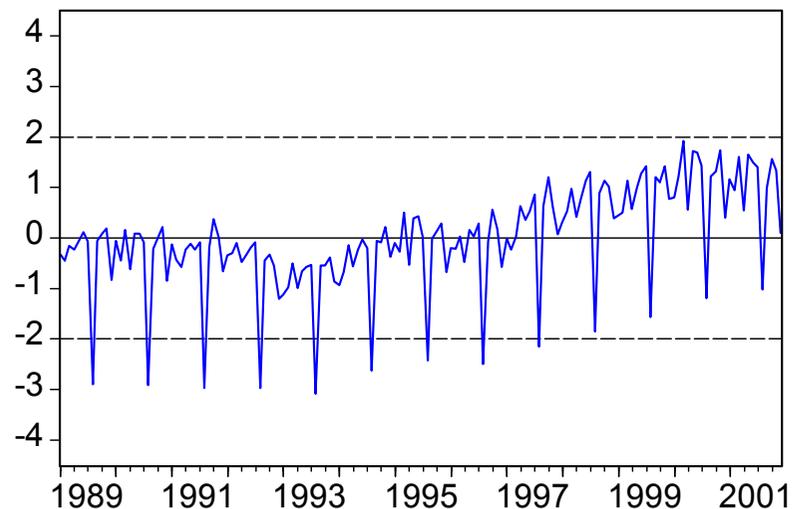
Tipos de interés de la deuda pública en EEUU
1960:I – 2002:IV
(interest-rates-q.wf1)

■ Bonos a 3 años – Vencimiento fijo
■ Letras a 3 meses – Mercado secundario

Observación 4: Las dos series de cada figura presentan **tendencias semejantes**, que parecen más previsibles (porque el patrón en el cambio de nivel es más claro) en las series de la Figura 1.7. Dichas series podrían tener una **tendencia común** con un componente **determinista**, mientras que la **tendencia común** de las series de la Figura 1.8 es **estocástica**.

**Figura 1.9**

Temperatura media registrada en Madrid-Retiro
Enero 1989 – Diciembre 2001
(seasonal.wf1)

**Figura 1.10**

Índice de producción industrial en España
Enero 1989 – Diciembre 2001
(seasonal.wf1)

Observación 5: La serie de la Figura 1.9 presenta reducciones (en invierno) y aumentos (en verano) sistemáticos en su nivel cada 12 meses, por lo que suele decirse que dicha serie es **estacional**. El caso de la serie de la Figura 1.10 es semejante (con caídas sistemáticas en su nivel en agosto de cada año); además, la serie de la Figura 1.10 también presenta (a diferencia de la serie de la Figura 1.9) una tendencia general estocástica.

1.1.3 Objetivos

El primer objetivo del análisis econométrico de una serie temporal consiste en elaborar un **modelo estadístico** que describa adecuadamente la procedencia de dicha serie, de manera que las **implicaciones teóricas** del modelo resulten compatibles con las **pautas muestrales** observadas en la serie temporal. Después, el modelo elaborado a partir de la serie temporal considerada puede utilizarse para:

- ⇒ **Describir** la evolución observada de dicha serie, así como las relaciones contemporáneas y dinámicas entre sus componentes (en el caso de series multivariantes).
- ⇒ **Prever** la evolución futura de dicha serie.
- ⇒ **Contrastar** (presentar evidencia empírica en favor o en contra de) alguna teoría sobre las características o variables a las que se refieren los componentes de dicha serie.

El punto de partida para elaborar un modelo a partir de una serie temporal consiste en considerar dicha serie como una **realización particular finita** de un **proceso estocástico**.

1.2 PROCESOS ESTOCÁSTICOS

1.2.1 Definición

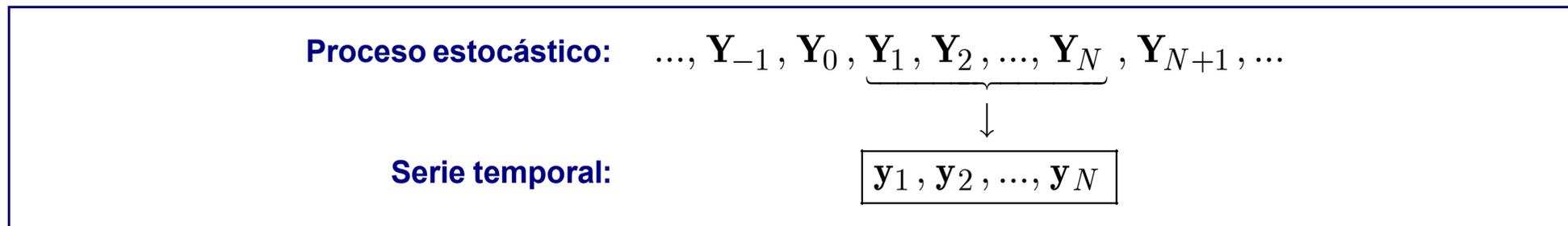
Un **proceso estocástico** es una **secuencia de variables aleatorias**, ordenadas y equidistantes cronológicamente, referidas a una (proceso univariante o escalar) o a varias (proceso multivariante o vectorial) características de una unidad observable en diferentes momentos.

Representaciones matemáticas frecuentes de procesos estocásticos univariantes:

$\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots$; $(Y_t : t = 0, \pm 1, \pm 2, \dots)$; (Y_t) , donde Y_t es una variable aleatoria escalar referida a la unidad observable considerada en el momento t .

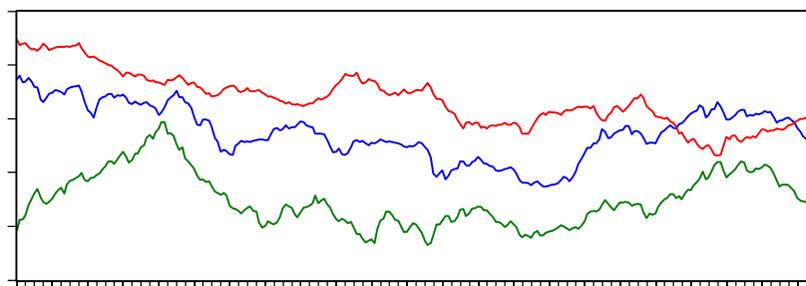
Representaciones matemáticas frecuentes de procesos estocásticos multivariantes:

$\dots, \mathbf{Y}_{-1}, \mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \dots$; $(\mathbf{Y}_t : t = 0, \pm 1, \pm 2, \dots)$; (\mathbf{Y}_t) , donde $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$ ($M \geq 2$) es una variable aleatoria vectorial referida a la unidad observable considerada en el momento t . El componente número j ($1 \leq j \leq M$) de cada \mathbf{Y}_t hace referencia a una característica o variable genérica dada, que es la misma en todo momento t .

**Figura 1.11**

Procedencia de una serie temporal observada

Observación 1: Aunque es posible concebir ciertos experimentos controlados (repetibles varias veces bajo idénticas condiciones de partida) que permitan obtener diferentes realizaciones particulares finitas de un proceso estocástico (como se ilustra en la figura bajo este párrafo), en general es *imposible, prohibitivamente costoso, o moralmente inaceptable* ejecutar dichos experimentos. La imposibilidad de controlar las condiciones a partir de las que se desarrollan las actividades en las que están implicadas las unidades observables a las que se refieren, es lo que hace que muchos procesos estocásticos sólo puedan observarse una única vez.



Tres realizaciones posibles de un mismo proceso estocástico

Observación 2: En general, una serie temporal se refiere a un **período muestral** que tan sólo es una parte de la historia del proceso estocástico del que procede dicha serie; ver Figura 1.11. No obstante, si las circunstancias sociales o naturales del período muestral al que se refiere la serie considerada se mantienen relativamente estables después de dicho período, entonces se espera que las conclusiones obtenidas del análisis de dicha serie sean aplicables también a momentos posteriores, al menos a corto plazo. Esta idea justifica el empleo de un modelo elaborado con una muestra dada para describir la evolución temporal de un proceso estocástico después del período muestral considerado.

Procesos estacionarios

1.2.2 Definición

Un proceso estocástico (Y_t) es **estacionario** cuando las propiedades estadísticas de cualquier secuencia finita $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ ($n \geq 1$) de componentes de (Y_t) son **semejantes** a las de la secuencia $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$ para cualquier número entero $h = \pm 1, \pm 2, \dots$

Observación 1: Aunque en el Tema 2 se dan algunas definiciones más concretas del concepto de estacionariedad, la definición anterior pone de manifiesto que cuando un proceso estocástico es estacionario, sus propiedades estadísticas se simplifican notablemente con respecto a las de un proceso que no sea estacionario (ver Definición 1.2.3), lo cual facilita la descripción de su estructura probabilística completa a partir de una única realización finita del mismo.

Observación 2: En general, una condición necesaria para que un proceso sea estacionario es que la **esperanza incondicional** de cada uno de sus componentes exista y sea la misma a lo largo de toda la historia del proceso. De acuerdo con esto, tan sólo las series de las figuras 1.1 y 1.6 anteriores podrían considerarse series **estacionarias**, es

decir, series compatibles con la hipótesis de haber sido generadas por procesos estacionarios. Las series restantes de las figuras 1.1-1.10 anteriores parecen series claramente **no estacionarias**.

Procesos no estacionarios

1.2.3 Definición

Un proceso estocástico (Y_t) es **no estacionario** cuando las propiedades estadísticas de al menos una secuencia finita $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ ($n \geq 1$) de componentes de (Y_t) , son **diferentes** de las de la secuencia $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$ para al menos un número entero $h > 0$.

Observación 1: En general, una condición suficiente para que un proceso sea no estacionario es que la esperanza incondicional de algunos de sus componentes sea distinta de la de otros (ver las figuras 1.2-1.5 y 1.7-1.10 anteriores). Al menos por este motivo, las propiedades estadísticas de un proceso no estacionario son más complicadas que las de un proceso estacionario. No obstante, si la no estacionariedad de un proceso puede modelizarse de alguna forma sencilla (por ejemplo, transformando adecuadamente dicho proceso para convertirlo en uno estacionario), entonces es posible describir su estructura probabilística completa a partir de una única realización finita del mismo.

Observación 2: Dado que las propiedades estadísticas de un proceso no estacionario son muy diferentes de las de un proceso estacionario, uno de los primeros pasos en el análisis de series temporales consiste en determinar si cada una de las series consideradas resulta o no compatible con la hipótesis de estacionariedad.

1.3 MODELOS

1.3.1 Definición

Un **modelo** para un proceso estocástico es cualquier conjunto de hipótesis bien definidas sobre las propiedades estadísticas de dicho proceso.

Observación: En muchas ocasiones, las propiedades estadísticas sobre las que se plantea un modelo son la esperanza (incondicional o condicional) de cada componente del proceso considerado y las covarianzas (incondicionales o condicionales) entre cada par de componentes del mismo. Además, este planteamiento suele hacerse de forma indirecta, a través de alguna expresión matemática (explícita o implícita) para el componente genérico \mathbf{Y}_t del proceso estocástico (\mathbf{Y}_t) considerado, en lugar de a través de alguna especificación directa de la forma de las esperanzas y de las covarianzas mencionadas.

1.3.2 Ejemplo – Modelos univariantes

Ruido blanco – ARIMA(0,0,0)

Un proceso de **ruido blanco** univariante es una secuencia (A_t) de variables aleatorias escalares idéntica e independientemente distribuidas con media 0 y varianza σ_A^2 , lo cual suele representarse como $(A_t) \sim \text{IID}(0, \sigma_A^2)$. Cuando cada A_t sigue una distribución

Normal, (A_t) se denomina un proceso de ruido blanco **Normal** o **Gaussiano**, lo cual suele representarse como $(A_t) \sim \text{NIID}(0, \sigma_A^2)$.

Modelo AR(1) – ARIMA(1,0,0)

Un proceso estocástico univariante estacionario (Y_t) sigue un modelo **AR(1)** (autorregresivo de orden 1) cuando

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (1.2)$$

donde μ y ϕ_1 son parámetros, $|\phi_1| < 1$ (**condición de estacionariedad**), y $(A_t) \sim \text{IID}(0, \sigma_A^2)$.

Modelo MA(1) – ARIMA(0,0,1)

Un proceso estocástico univariante estacionario (Y_t) sigue un modelo **MA(1)** (media móvil de orden 1) cuando

$$Y_t = \mu + A_t - \theta_1 A_{t-1} \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (1.3)$$

donde μ y θ_1 son parámetros, $|\theta_1| < 1$ (**condición de invertibilidad**), y $(A_t) \sim \text{IID}(0, \sigma_A^2)$.

Paseo aleatorio – ARIMA(0,1,0)

Un proceso estocástico univariante **no** estacionario (Y_t) es un **paseo aleatorio** cuando

$$Y_t = \mu + Y_{t-1} + A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (1.4)$$

donde μ es un parámetro (que en muchas ocasiones vale cero) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$.

Observación 1: Un paseo aleatorio es un modelo AR(1) con $\phi_1 = 1$, es decir, un modelo AR(1) **no** estacionario. [Nótese que cuando $\phi_1 = 1$, la ecuación $1 - \phi_1 x = 0$ asociada con un modelo AR(1) tiene una **raíz unitaria** ($x^* = 1$).] Por otro lado, un paseo aleatorio puede escribirse como $\nabla Y_t = \mu + A_t$, donde $\nabla Y_t \equiv Y_t - Y_{t-1}$, de manera que un proceso estocástico univariante no estacionario (Y_t) es un paseo aleatorio cuando su **diferencia regular de orden 1** ($\nabla Y_t \equiv (Y_t - Y_{t-1})$) es un proceso estacionario $\text{IID}(\mu, \sigma_A^2)$ [donde el parámetro μ puede valer cero, en cuyo caso el proceso (∇Y_t) sería ruido blanco].

Observación 2: Un paseo aleatorio constituye el caso más sencillo de lo que suele denominarse un **proceso integrado de orden 1** ó **I(1)** (es decir, un proceso no estacionario cuya diferencia regular de orden 1 es un proceso estacionario); por extensión de esta terminología, un proceso estacionario suele denominarse un proceso **I(0)**. La notación $(Y_t) \sim \text{I}(1)$ significa que (Y_t) es un proceso I(1) (integrado de orden 1), mientras que la notación $(Y_t) \sim \text{I}(0)$ significa que (Y_t) es un proceso estacionario (integrado de orden 0).

Modelo ARCH(1)

Un proceso estocástico univariante estacionario (U_t) sigue un modelo **ARCH(1)** (del inglés

AutoRegressive Conditional Heteroskedasticity) cuando

$$U_t = \sigma_t A_t, \text{ con } \sigma_t^2 = \mathbb{E}_{t-1}[U_t^2] = \alpha_0 + \alpha_1 U_{t-1}^2, \quad (1.5)$$

donde $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_1 < 1$, y $(A_t) \sim \text{IID}(0, 1)$ es independiente de (σ_t) .

Observación 3: Cuando (U_t) sigue un modelo ARCH(1) como (1.5), los componentes de (U_t) no presentan autocorrelación aunque **no** son estadísticamente independientes (porque la **varianza condicional** de U_t es una función de U_{t-1}). Por otro lado, la **varianza incondicional** de U_t es **constante** e igual a $\alpha_0 / (1 - \alpha_1)$, que es positiva si se cumplen las condiciones anteriores sobre α_0 y α_1 .

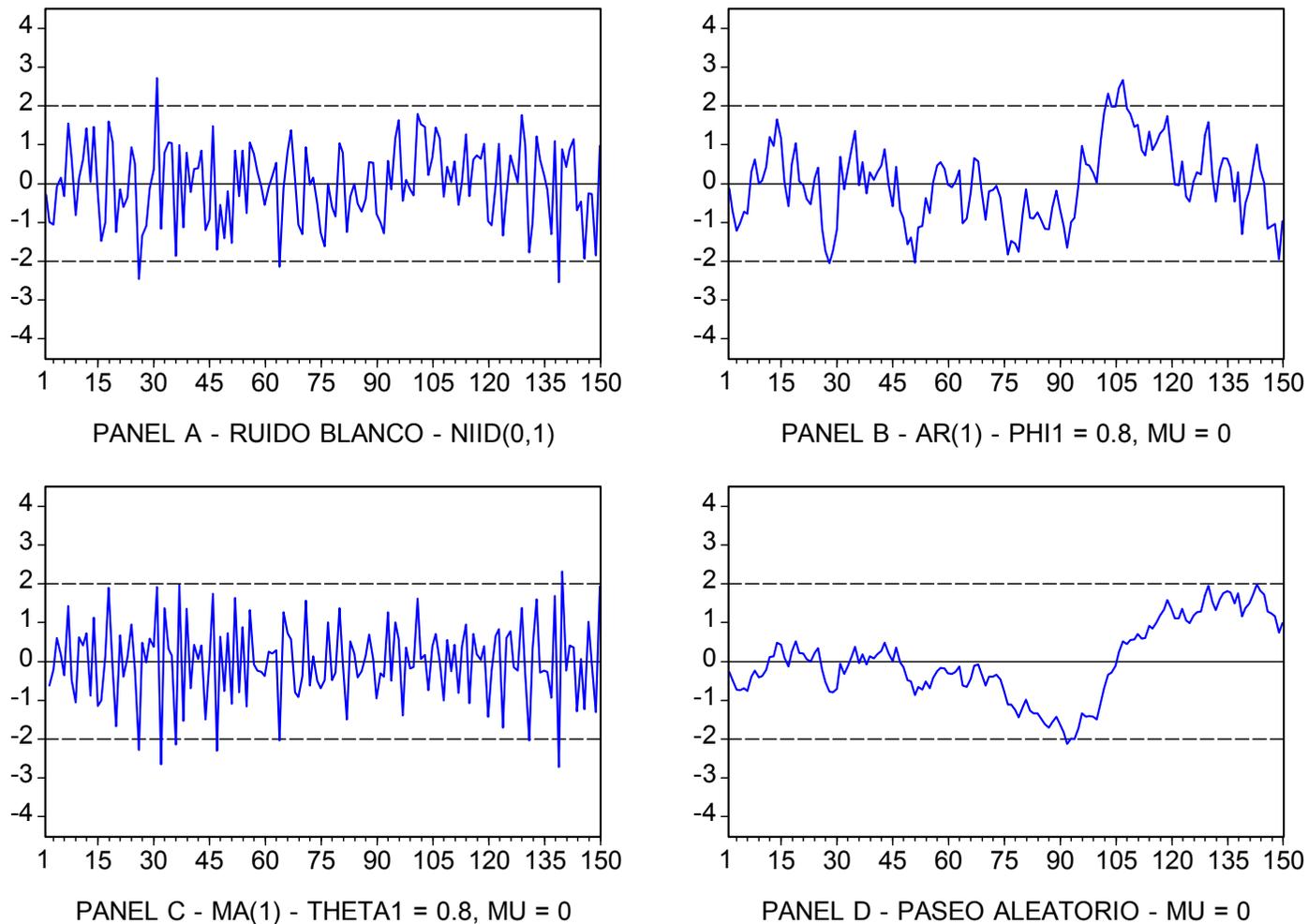
Modelo GARCH(1,1)

Un proceso estocástico univariante estacionario (U_t) sigue un modelo **GARCH(1,1)** (del inglés *Generalized AutoRegressive Conditional Heteroskedasticity*) cuando

$$U_t = \sigma_t A_t, \text{ con } \sigma_t^2 = \mathbb{E}_{t-1}[U_t^2] = \alpha_0 + \alpha_1 U_{t-1}^2 + \delta_1 \sigma_{t-1}^2,$$

donde $\alpha_0 > 0$, $\alpha_1 > 0$, $\delta_1 > 0$, $\alpha_1 + \delta_1 < 1$ y $(A_t) \sim \text{IID}(0, 1)$ es independiente de (σ_t) .

Observación 4: Cuando (U_t) sigue un modelo GARCH(1,1), la **varianza incondicional** de U_t es **constante** e igual a $\alpha_0 / (1 - \alpha_1 - \delta_1)$, que es positiva si se cumplen las condiciones anteriores sobre α_0 , α_1 y δ_1 .

**Figura 1.12**

Series temporales simuladas a partir de varios modelos ARIMA
(sim-series.wf1)

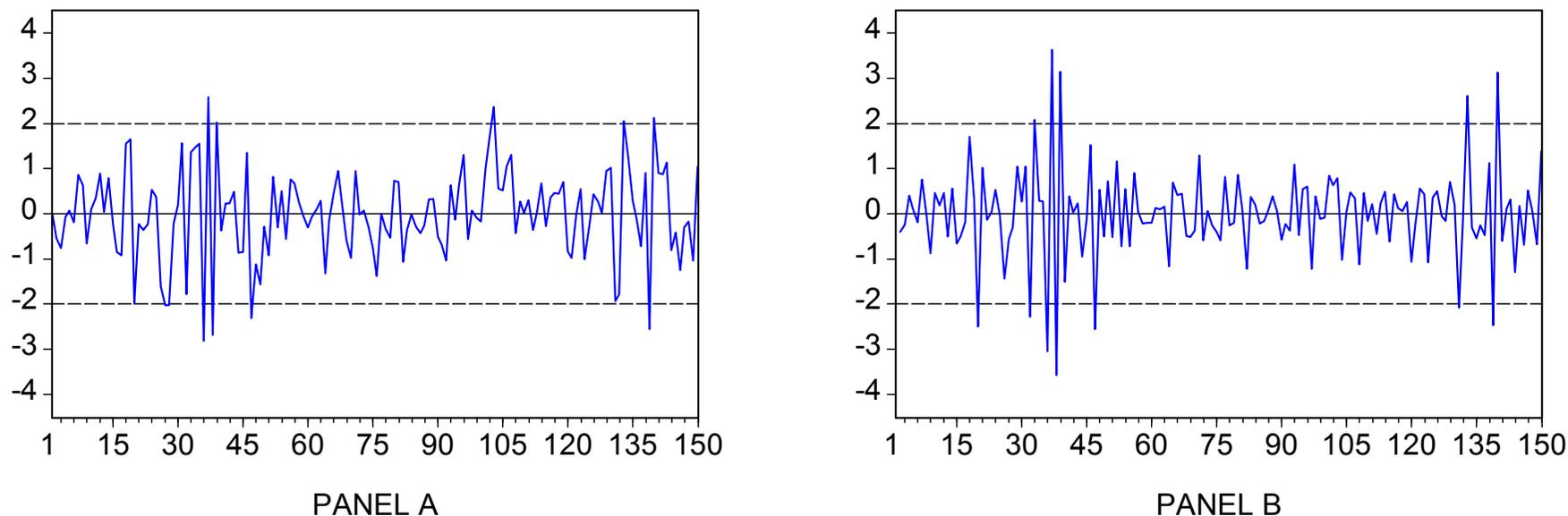


Figura 1.13

Series temporales simuladas a partir del modelo ARCH(1) de la ecuación (1.5)
(sim-series.wf1)

Observación 5: La serie $(u_t)_{t=1}^{150}$ del Panel A se ha obtenido a partir del modelo ARCH(1) de la ecuación (1.5) con $\alpha_0 = 1$, $\alpha_1 = 0.8$ y $(A_t) \sim \text{NIID}(0, 1)$ (comparar con la serie del Panel A de la Figura 1.12). La serie $(y_t)_{t=2}^{150}$ del Panel B se ha obtenido a partir de la ecuación $y_t = u_t - 0.8u_{t-1}$ [un modelo MA(1) invertible; ver (1.3)], donde $(u_t)_{t=1}^{150}$ es la serie del Panel A (comparar con la serie del Panel C de la Figura 1.12).

1.3.3 Ejemplo – Modelos multivariantes estacionarios

Modelo ADL(1,1)

Un proceso bivalente estacionario (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_t, X_t]'$, sigue un modelo **ADL(1,1)** cuando

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + V_t, \quad (1.6)$$

donde $\beta_0, \beta_1, \gamma_0$ y γ_1 son parámetros, $|\beta_1| < 1$, y $(V_t) \sim \text{IID}(0, \sigma_V^2)$.

Observación 1: Cuando $\beta_1 = \gamma_1 = 0$, el modelo ADL(1,1) anterior se reduce a un modelo de regresión estático con perturbaciones IID. Cuando $\gamma_0 = \gamma_1 = 0$, el modelo ADL(1,1) anterior se reduce a un modelo AR(1). Cuando $\gamma_1 = -\beta_1 \gamma_0$, el modelo ADL(1,1) anterior se convierte en un modelo de regresión estático con perturbaciones AR(1). Por último, cuando $\beta_1 = 1$ y $\gamma_1 = -\gamma_0$, el modelo ADL(1,1) anterior se convierte en un modelo de regresión estático entre $\nabla Y_t \equiv Y_t - Y_{t-1}$ y $\nabla X_t \equiv X_t - X_{t-1}$ con perturbaciones IID: $\nabla Y_t = \beta_0 + \gamma_0 \nabla X_t + V_t$.

Observación 2: Manipulando la expresión (1.6) para un modelo ADL(1,1), puede comprobarse que dicho modelo implica que $\nabla Y_t = \beta_0 - (1 - \beta_1)(Y_{t-1} - \lambda X_{t-1}) + \gamma_0 \nabla X_t + V_t$, donde $\lambda \equiv (\gamma_0 + \gamma_1)/(1 - \beta_1)$, que se denomina un **modelo de corrección de error (ECM)**. En el ECM anterior, $(Y_{t-1} - \lambda X_{t-1})$ representa en qué medida Y_{t-1} y X_{t-1} se encuentran fuera de la **relación de equilibrio a largo plazo** entre (Y_t) y (X_t) implícita en (1.6); dado que $-(1 - \beta_1) < 0$, el **término de corrección de error** $-(1 - \beta_1)(Y_{t-1} - \lambda X_{t-1})$ en el ECM anterior representa la proporción del desequilibrio (error) entre Y_{t-1} y X_{t-1} que se corrige a través de ∇Y_t ; ver Figura 1.14.

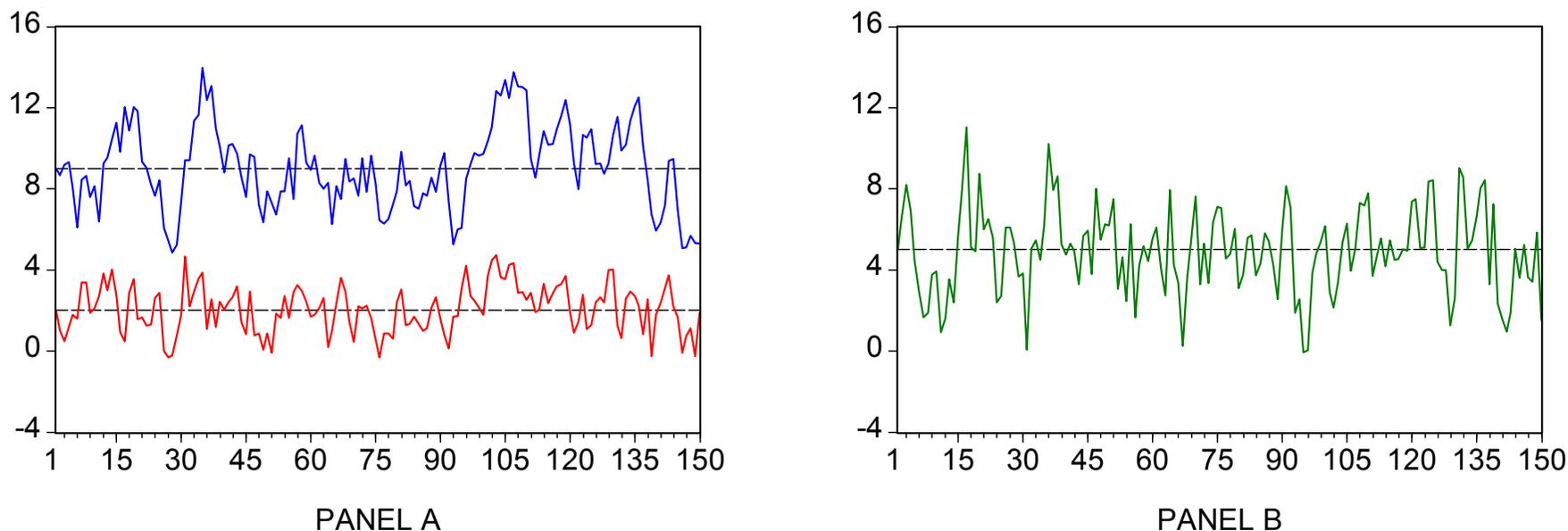


Figura 1.14

Series temporales simuladas a partir del modelo ADL(1,1) de la ecuación (1.6)
(sim-series.wf1)

Panel A - ■ Serie Y ■ Serie X. Panel B - ■ Serie Y - 2 X

Observación 3: La serie X se ha obtenido a partir del modelo AR(1) de la ecuación (1.2) con $\mu = 1$, $\phi_1 = 0.5$ y $(A_t) \sim \text{NIID}(0, 1)$. La serie Y se ha obtenido a partir del modelo ADL(1,1) de la ecuación (1.6) con $\beta_0 = 2$, $\beta_1 = 0.6$, $\gamma_0 = 0.5$, $\gamma_1 = 0.3$ y $(V_t) \sim \text{NIID}(0, 1)$, lo que implica que $\lambda = 2$ en el **modelo de corrección de error (ECM)** asociado; ver Observación 2 en 1.3.3.

Modelo VAR(1) bivalente

Un proceso bivalente estacionario (\mathbf{Y}_t), donde $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$, sigue un modelo **VAR(1)** (autorregresivo vectorial de orden 1) cuando

$$\begin{aligned} Y_{t1} &= \mu_1 + \phi_{11}Y_{t-1,1} + \phi_{12}Y_{t-1,2} + A_{t1}, \\ Y_{t2} &= \mu_2 + \phi_{21}Y_{t-1,1} + \phi_{22}Y_{t-1,2} + A_{t2}, \end{aligned} \quad (1.7)$$

o bien

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \mathbf{A}_t, \quad (1.8)$$

donde $\mathbf{A}_t \equiv [A_{t1}, A_{t2}]'$, $(\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma})$ (ruido blanco vectorial),

$$\boldsymbol{\mu} \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Phi}_1 \equiv \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \boldsymbol{\Sigma} \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \quad (1.9)$$

y las raíces de la ecuación $|\mathbf{I} - \boldsymbol{\Phi}_1 x| = 0$ están **fuera** del círculo unitario (lo cual equivale a que los autovalores de la matriz $\boldsymbol{\Phi}_1$ de (1.8) estén **dentro** del círculo unitario).

Observación 4: A diferencia de lo que ocurre en un modelo ADL(1,1) como (1.6), donde Y_t es una variable **endógena** y X_t es una variable **exógena**, en un modelo VAR(1) como (1.7) tanto Y_{t1} como Y_{t2} son variables endógenas, cuya

evolución temporal se determina de forma conjunta.

Observación 5: El modelo VAR(1) que figura en (1.7)-(1.9) puede interpretarse como la **forma reducida** de un modelo bivalente cuya **forma estructural** es del tipo

$$\begin{aligned} Y_{t1} &= \beta_{01} + \beta_{11}Y_{t-1,1} + \gamma_{01}Y_{t2} + \gamma_{11}Y_{t-1,2} + V_{t1}, \\ Y_{t2} &= \beta_{02} + \beta_{12}Y_{t-1,2} + \gamma_{02}Y_{t1} + \gamma_{12}Y_{t-1,1} + V_{t2}, \end{aligned} \quad (\text{I})$$

o bien

$$\begin{bmatrix} 1 & -\gamma_{01} \\ -\gamma_{02} & 1 \end{bmatrix} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \gamma_{11} \\ \gamma_{12} & \beta_{12} \end{bmatrix} \begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix} + \begin{bmatrix} V_{t1} \\ V_{t2} \end{bmatrix},$$

que puede escribirse de manera más compacta como

$$\mathbf{B}_0 \mathbf{Y}_t = \boldsymbol{\beta}_0 + \mathbf{B}_1 \mathbf{Y}_{t-1} + \mathbf{V}_t, \quad (\text{II})$$

donde $\mathbf{V}_t \equiv [V_{t1}, V_{t2}]'$, $(\mathbf{V}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Upsilon})$,

$$\mathbf{B}_0 \equiv \begin{bmatrix} 1 & -\gamma_{01} \\ -\gamma_{02} & 1 \end{bmatrix}, \boldsymbol{\beta}_0 \equiv \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix}, \mathbf{B}_1 \equiv \begin{bmatrix} \beta_{11} & \gamma_{11} \\ \gamma_{12} & \beta_{12} \end{bmatrix}, \boldsymbol{\Upsilon} \equiv \begin{bmatrix} v_1^2 & 0 \\ 0 & v_2^2 \end{bmatrix}. \quad (\text{III})$$

La expresión (II) suele denominarse un **modelo VAR(1) estructural** para el proceso (\mathbf{Y}_t) , mientras que la expresión (1.8) se denomina un **modelo VAR(1) estándar** para (\mathbf{Y}_t) . Comparando (II) con (1.8), la forma estándar (reducida) de un modelo VAR(1) puede obtenerse a partir de su forma estructural mediante las relaciones $\boldsymbol{\mu} \equiv \mathbf{B}_0^{-1}\boldsymbol{\beta}_0$, $\boldsymbol{\Phi}_1 \equiv \mathbf{B}_0^{-1}\mathbf{B}_1$, $\mathbf{A}_t \equiv \mathbf{B}_0^{-1}\mathbf{V}_t$ y $\boldsymbol{\Sigma} \equiv \mathbf{B}_0^{-1}\boldsymbol{\Upsilon}\mathbf{B}_0^{-1'}$ (que coincide con $\boldsymbol{\Upsilon}$ cuando $\gamma_{01} = \gamma_{02} = 0$, es decir, cuando $\mathbf{B}_0 = \mathbf{I}$). Sin embargo, dado que en (III) figuran 10 parámetros y en (1.9) sólo figuran 9, en general, no es posible **identificar** la forma estructural asociada con un modelo VAR(1) estándar.

Observación 6: Cuando $\gamma_{02} = \gamma_{12} = 0$, el modelo VAR(1) estructural (I) se reduce a

$$\begin{aligned} Y_{t1} &= \beta_{01} + \beta_{11}Y_{t-1,1} + \gamma_{01}Y_{t2} + \gamma_{11}Y_{t-1,2} + V_{t1}, \\ Y_{t2} &= \beta_{02} + \beta_{12}Y_{t-1,2} + V_{t2}, \end{aligned}$$

que es un modelo ADL(1,1) como (1.6) en el que (X_t) sigue un modelo AR(1) estacionario; ver Figura 1.14.

Observación 7: El modelo VAR(1) que figura en (1.8) puede escribirse siempre como $\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi} \mathbf{Y}_{t-1} + \mathbf{A}_t$, con $\nabla \mathbf{Y}_t \equiv \mathbf{Y}_t - \mathbf{Y}_{t-1}$ y $\boldsymbol{\Pi} \equiv \mathbf{I} - \boldsymbol{\Phi}_1$ (que es una matriz no singular cuando los autovalores de $\boldsymbol{\Phi}_1$ están dentro del círculo unitario), lo que permite escribir el valor esperado del proceso estacionario (\mathbf{Y}_t) como $E[\mathbf{Y}_t] = \boldsymbol{\Pi}^{-1} \boldsymbol{\mu}$ (que representa el valor de **equilibrio** alrededor del cual evoluciona dicho proceso).

Observación 8: Un modelo VAR(1) estándar no es más que un caso particular de un modelo VARMA(p, q) del tipo

$$\mathbf{Y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{Y}_{t-2} + \dots + \boldsymbol{\Phi}_p \mathbf{Y}_{t-p} + \mathbf{A}_t - \boldsymbol{\Theta}_1 \mathbf{A}_{t-1} - \boldsymbol{\Theta}_2 \mathbf{A}_{t-2} - \dots - \boldsymbol{\Theta}_q \mathbf{A}_{t-q},$$

que es una extensión para procesos multivariantes del modelo ARMA(p, q) para procesos univariantes. Actualmente, la popularidad de los modelos VAR es mucho mayor entre los economistas que la de los modelos VARMA; de hecho, es muy difícil encontrar manuales de Econometría o de Análisis de Series Temporales aplicado a la Economía en los que se mencione siquiera la posibilidad de modelizar un proceso multivariante a través de un modelo VARMA. Esta posibilidad tiene ventajas muy claras sobre la modelización basada exclusivamente en modelos VAR (por ejemplo, una **parametrización** más **escueta**), aunque sus requisitos computacionales son generalmente más elevados. En todo caso, es una posibilidad que no se contempla en la mayoría de los programas informáticos más populares disponibles actualmente (como EViews) para el análisis de series temporales.

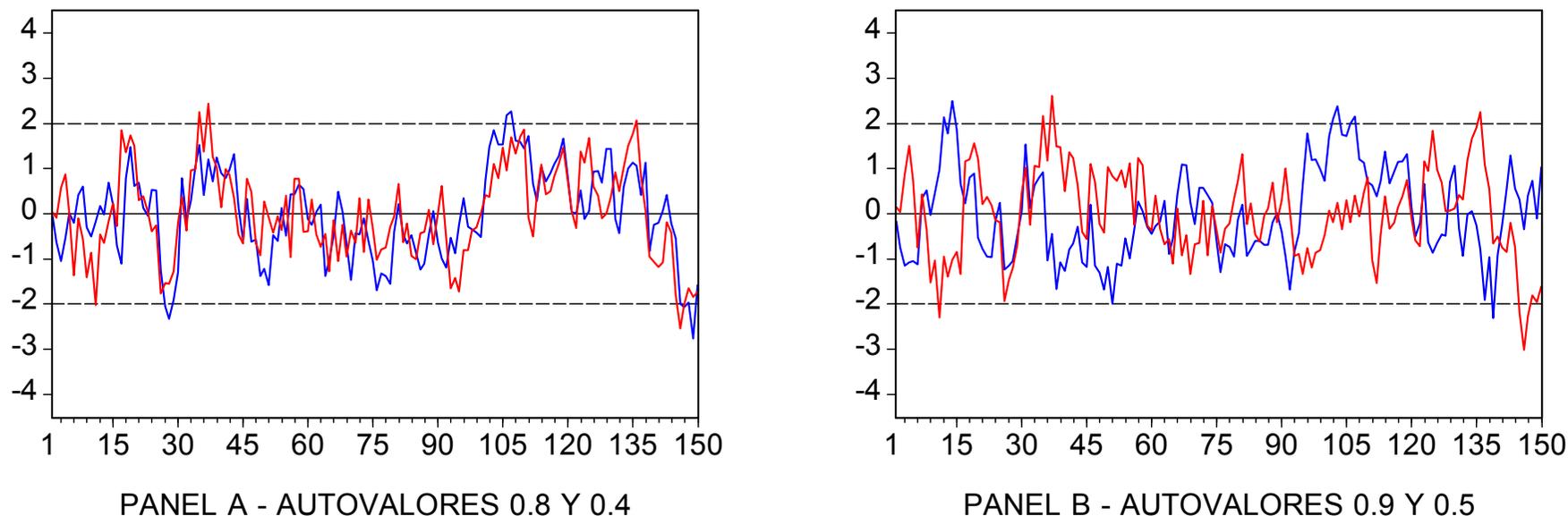


Figura 1.15

Series temporales simuladas a partir del modelo VAR(1) de las ecuaciones (1.7)-(1.9)
(sim-series.wf1)

■ Serie Y1 ■ Serie Y2

Observación 9: Se han utilizado los valores siguientes para los parámetros que figuran en (1.9):

$$\text{Panel A: } \Phi_1 = \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}, \text{ Panel B: } \Phi_1 = \begin{bmatrix} 0.7 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}, \text{ Ambos paneles: } \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}.$$

1.3.4 Ejemplo – Modelos multivariantes no estacionarios

Cointegración

Si un proceso bivalente $(\mathbf{Z}_t) \sim I(1)$ ($\mathbf{Z}_t \equiv [Z_{t1}, Z_{t2}]'$) es tal que cualquier combinación lineal de sus dos componentes también es un proceso $I(1)$, entonces **no** existe entre (Z_{t1}) y (Z_{t2}) relación lineal alguna del tipo $Z_{t1} = \gamma_0 + \gamma_2 Z_{t2} + V_t$ que sea **estable** en el tiempo, ya que en este caso $(V_t) \sim I(1)$ (un proceso **no** estacionario). No obstante, puede que sí exista algún tipo de relación entre (∇Z_{t1}) y (∇Z_{t2}) (dos procesos estacionarios; ver 1.3.3).

Por el contrario, si un proceso bivalente $(\mathbf{Y}_t) \sim I(1)$ ($\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$) es tal que existe una combinación lineal de sus dos componentes que es un proceso $I(0)$, entonces **sí** existe entre (Y_{t1}) e (Y_{t2}) al menos una relación lineal del tipo $Y_{t1} = \beta_0 + \beta_2 Y_{t2} + U_t$ que es **estable** en el tiempo, ya que en este caso $(U_t) \sim I(0)$ (un proceso estacionario).

En el segundo caso, se dice que (Y_{t1}) e (Y_{t2}) presentan **cointegración** de orden $(1,1)$ [porque el orden de integración de los dos procesos es **1** y existe una combinación lineal de ambos que es un proceso $I(\mathbf{1}-\mathbf{1}) = I(0)$], lo cual se representa como $(\mathbf{Y}_t) \sim CI(1,1)$, con **vector de cointegración** $\beta = [1, -\beta_2]'$ [porque $(\beta' \mathbf{Y}_t) \sim I(0)$].

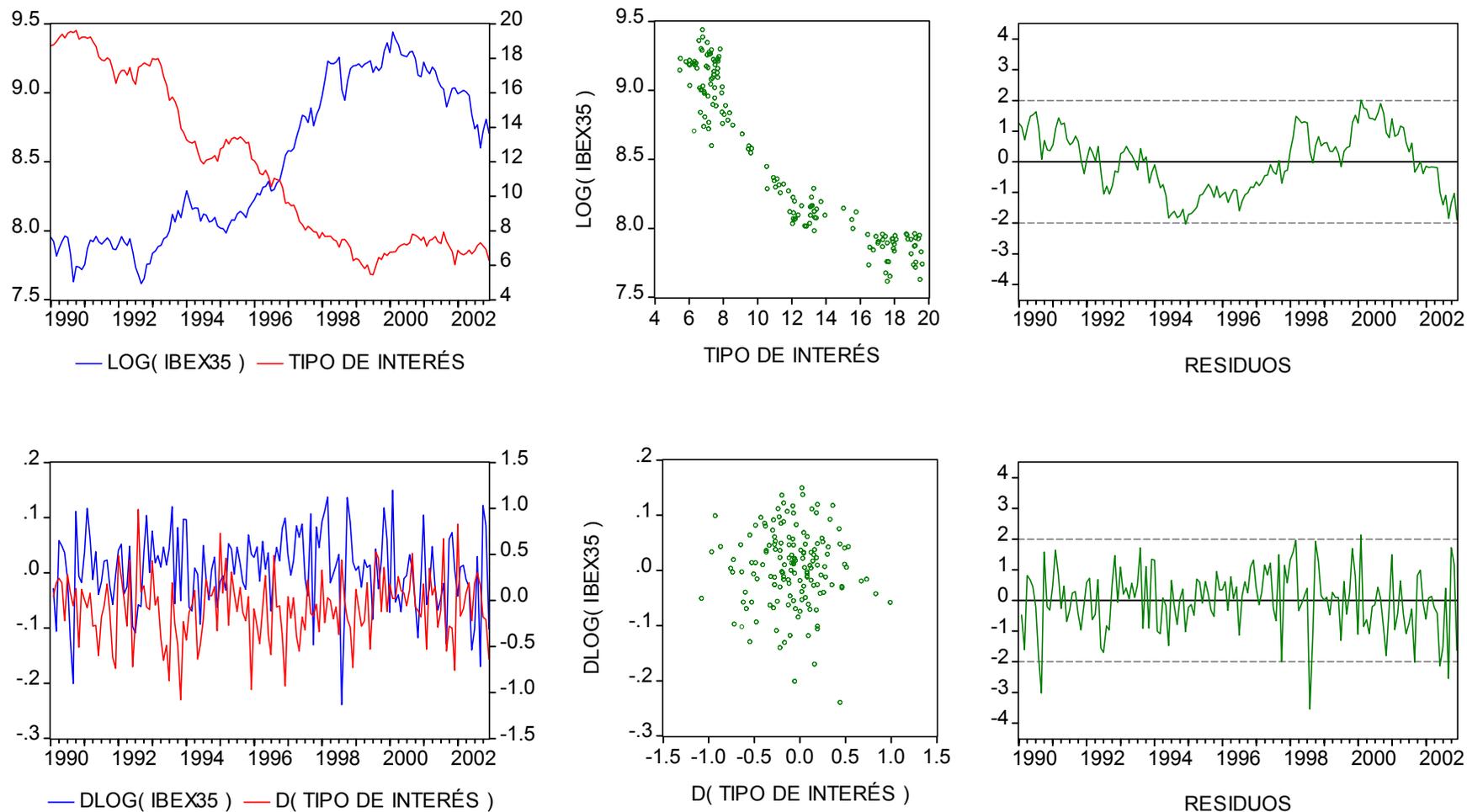


Figura 1.16

Series temporales I(1) no cointegradas - no relacionadas en diferencias (bolsa-int.wf1)

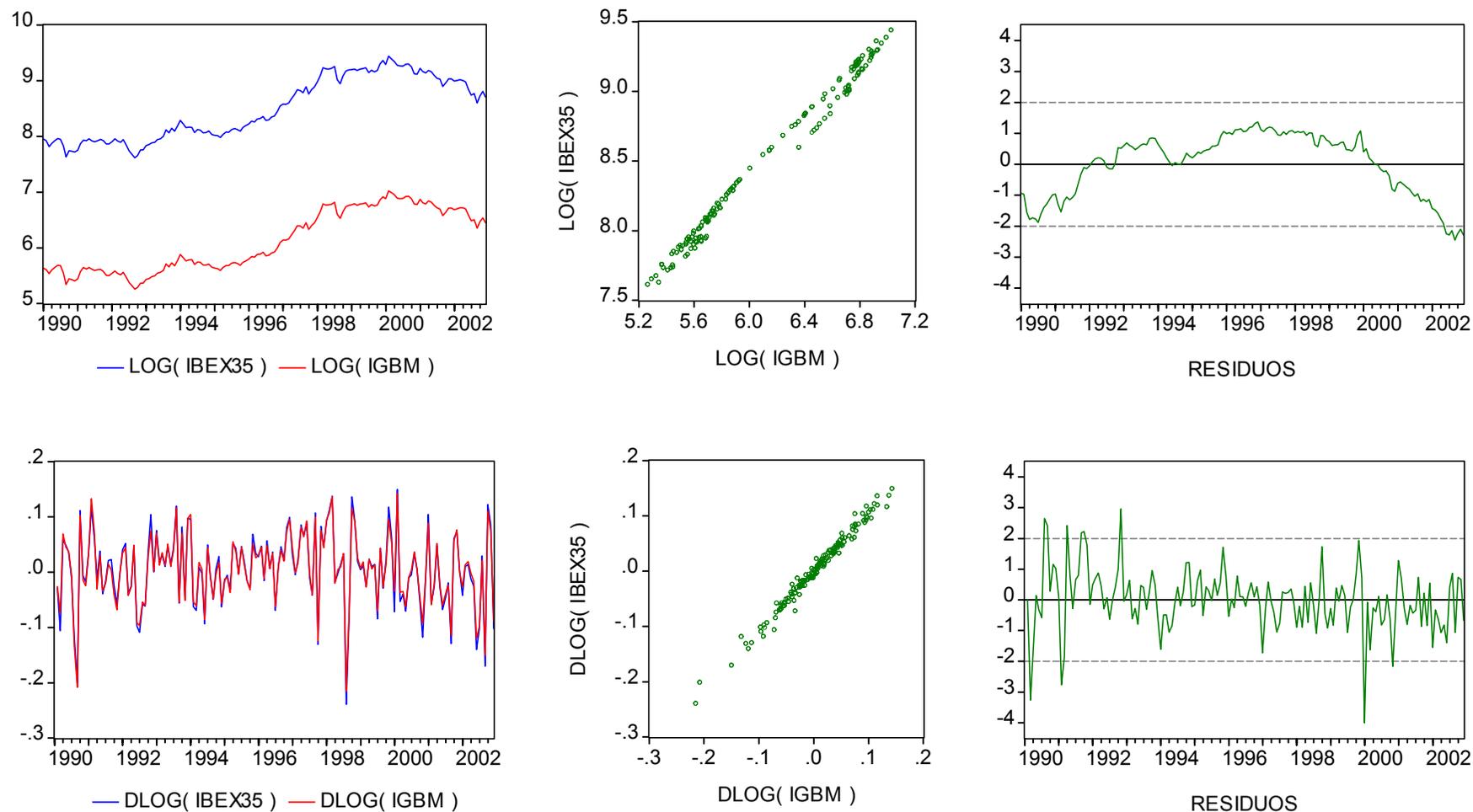


Figura 1.17

Series temporales I(1) no cointegradas - relacionadas en diferencias (bolsa-int.wf1)

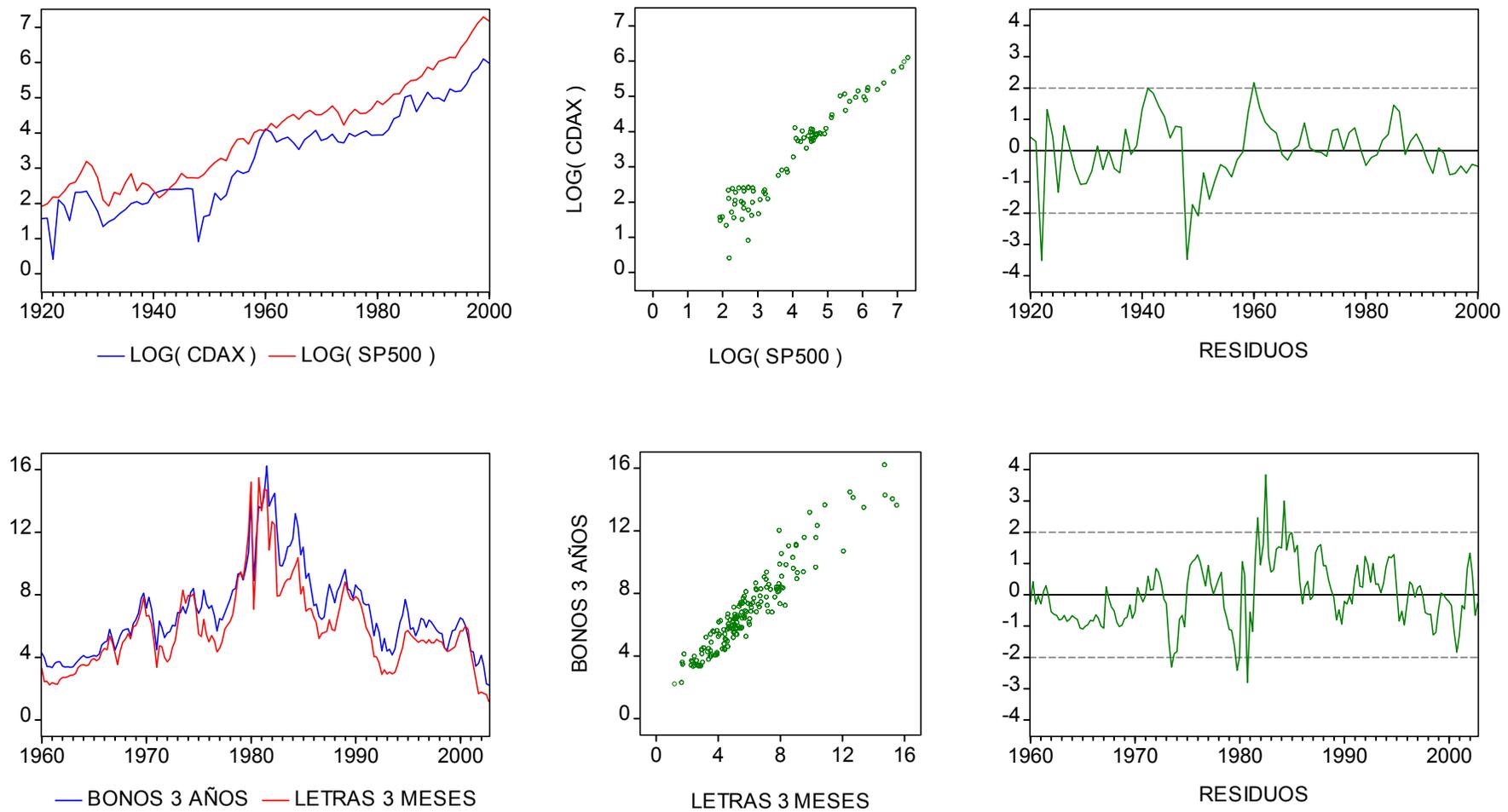


Figura 1.18

Series temporales I(1) posiblemente cointegradas
 (Panel superior: stock-indices.wf1 – Panel inferior: interest-rates-q.wf1)

Modelos para procesos no estacionarios - no cointegrados

Cuando $(\mathbf{Z}_t) \sim I(1)$ ($\mathbf{Z}_t \equiv [Z_{t1}, Z_{t2}]'$) y sus componentes **no** están cointegrados, sólo tiene sentido plantear algún modelo de relación lineal entre (Z_{t1}) y (Z_{t2}) en términos de sus diferencias regulares (∇Z_{t1}) y (∇Z_{t2}) (que son procesos estacionarios; ver 1.3.3).

Plantear en este caso un modelo de relación lineal directamente entre (Z_{t1}) y (Z_{t2}) suele dar lugar a la obtención de **relaciones espurias** que carecen de sentido; ver Figura 1.16.

Modelos para procesos no estacionarios - cointegrados

Cuando $(\mathbf{Y}_t) \sim CI(1, 1)$ ($\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$), sí tiene sentido plantear algún modelo de relación lineal directamente entre (Y_{t1}) e (Y_{t2}) [por ejemplo, un modelo de regresión lineal simple, un modelo ADL(1,1) como el que figura en (1.6), o un ECM como el que figura en la Observación 2 de 1.3.3, donde $(Y_{t1} - \lambda Y_{t2}) \sim I(0)$ para algún $\lambda \neq 0$; ver Figura 1.18].

Modelos VAR(1) con raíces unitarias

Si (\mathbf{Y}_t) sigue un modelo VAR(1) del tipo (1.7)-(1.9) tal que al menos uno de los dos autovalores de Φ_1 es igual a 1 (lo que implica que la ecuación $|\mathbf{I} - \Phi_1 x| = 0$ tiene al menos

una **raíz unitaria**), entonces (\mathbf{Y}_t) es un proceso no estacionario, cuyas características están determinadas por los dos autovalores λ_1 y λ_2 de la matriz Φ_1 :

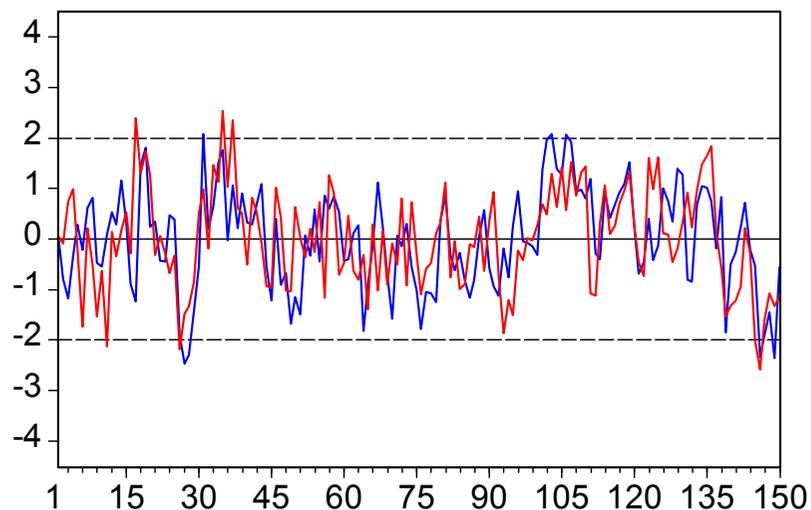
⇨ $\lambda_1 = 1, |\lambda_2| < 1 \Rightarrow (\mathbf{Y}_t) \sim \text{CI}(1,1)$. Nótese que en este caso el rango de la matriz $\Pi \equiv \mathbf{I} - \Phi_1$ que figura en la Observación 7 de 1.3.3 es igual a 1, lo que indica que existe una relación de cointegración entre los dos componentes I(1) de (\mathbf{Y}_t) (Figura 1.19 B).

⇨ $\lambda_1 = 1, \lambda_2 = 1, \Phi_1 \neq \mathbf{I} \Rightarrow (\mathbf{Y}_t) \sim \text{CI}(2,1)$. En este caso, el rango de $\Pi \equiv \mathbf{I} - \Phi_1$ también es igual a 1, lo que indica que existe una relación de cointegración entre los dos componentes I(2) de (\mathbf{Y}_t) (Figura 1.19 C) que proporciona un proceso I(1).

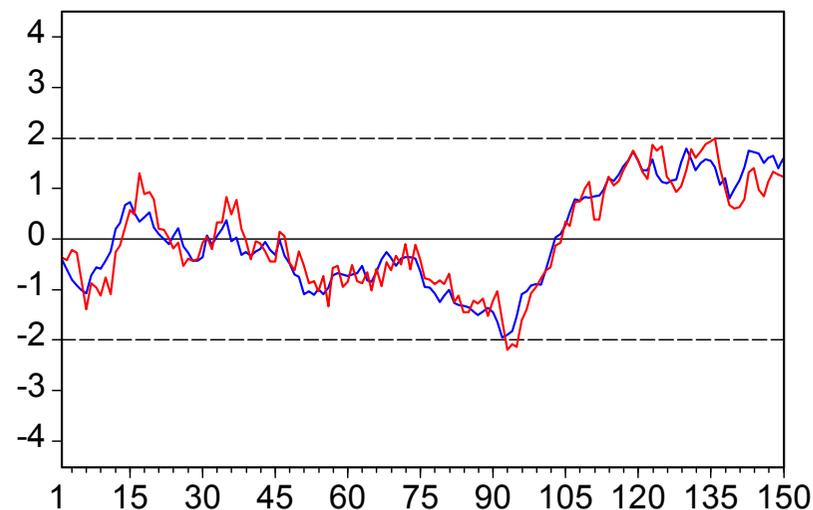
⇨ $\lambda_1 = 1, \lambda_2 = 1, \Phi_1 = \mathbf{I} \Rightarrow (\mathbf{Y}_t) \sim \text{I}(1)$, pero (Y_{t1}) e (Y_{t2}) **no** están cointegrados. En este caso el rango de la matriz $\Pi \equiv \mathbf{I} - \Phi_1$ es igual a 0, lo que indica que no existe ninguna relación de cointegración entre los dos componentes de (\mathbf{Y}_t) (Figura 1.19 D).

Observación 1: El rango de la matriz $\Pi \equiv \mathbf{I} - \Phi_1$ está relacionado con los autovalores λ_1 y λ_2 de la matriz Φ_1 porque los autovalores de $\Pi \equiv \mathbf{I} - \Phi_1$ son $(1 - \lambda_1)$ y $(1 - \lambda_2)$.

Observación 2: Cuando los dos autovalores de Φ_1 están dentro del círculo unitario ($|\lambda_1| < 1, |\lambda_2| < 1$), (\mathbf{Y}_t) es un proceso estacionario (Figura 1.19 A y Figura 1.15). En este caso, el rango de la matriz $\Pi \equiv \mathbf{I} - \Phi_1$ es igual a 2.



PANEL A - AUTOVALORES 0.6 Y 0.2 - RANGO = 2



PANEL B - AUTOVALORES 1.0 Y 0.6 - RANGO = 1

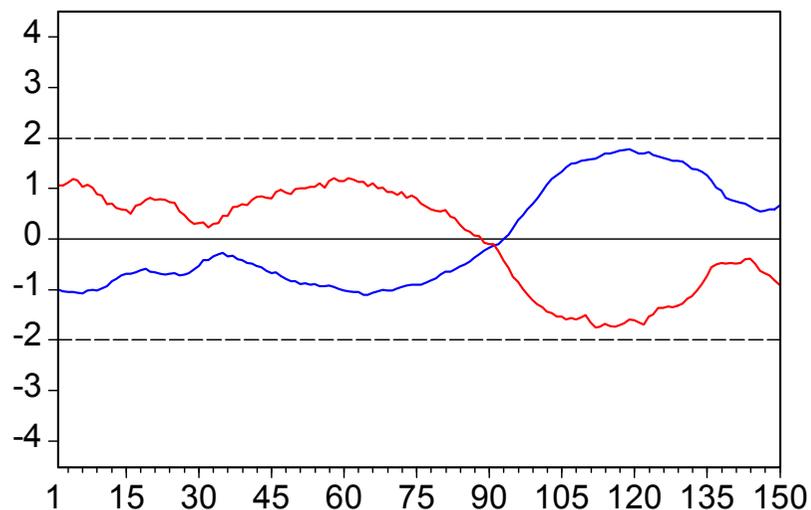
Figura 1.19 (I)

Series temporales simuladas a partir del modelo VAR(1) de las ecuaciones (1.7)-(1.9)
(sim-series.wf1)

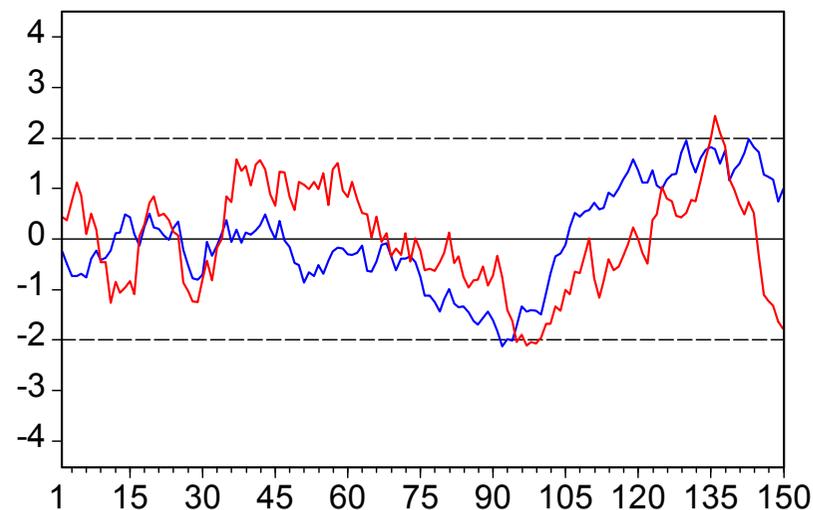
■ Serie Y1 ■ Serie Y2

Observación 3 (I): Se han utilizado los valores siguientes para los parámetros que figuran en (1.9):

$$\text{Panel A: } \Phi_1 = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}, \text{ Panel B: } \Phi_1 = \begin{bmatrix} 1.2 & -0.2 \\ 0.6 & 0.4 \end{bmatrix}, \text{ Ambos paneles: } \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}.$$



PANEL C - AUTOVALORES 1.0 Y 1.0 - RANGO = 1



PANEL D - AUTOVALORES 1.0 Y 1.0 - RANGO = 0

Figura 1.19 (II)

Series temporales simuladas a partir del modelo VAR(1) de las ecuaciones (1.7)-(1.9)
(sim-series.wf1)

■ Serie Y1 ■ Serie Y2

Observación 3 (II): Se han utilizado los valores siguientes para los parámetros que figuran en (1.9):

$$\text{Panel C: } \Phi_1 = \begin{bmatrix} 0.8 & -0.4 \\ 0.1 & 1.2 \end{bmatrix}. \text{ Panel D: } \Phi_1 = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \text{ Ambos paneles: } \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}.$$

1.4 APLICACIONES

1.4.1 Ejemplo – Previsión con modelos ARIMA univariantes

Las **funciones de previsión** en **origen** N a **horizonte** $l \geq 1$ asociadas con los modelos ARIMA considerados en 1.3.2 son las siguientes:

Ruido blanco: $E_N[A_{N+l}] = E[A_{N+l}] = 0$ para todo $l \geq 1$.

AR(1):
$$E_N[Y_{N+l}] = \mu + \phi_1 E_N[Y_{N+l-1}]$$

$$= \left(\sum_{i=0}^{l-1} \phi_1^i \right) \mu + \phi_1^l Y_N$$
 para todo $l \geq 1$.

MA(1):
$$E_N[Y_{N+l}] = \begin{cases} \mu - \theta_1 E_N[A_N], & l = 1, \\ \mu & , l > 1. \end{cases}$$

Paseo aleatorio:
$$E_N[Y_{N+l}] = \mu + E_N[Y_{N+l-1}]$$

$$= l\mu + Y_N$$
 para todo $l \geq 1$.

1.4.2 Ejemplo – Relaciones dinámicas implícitas en un modelo ADL(1,1)

Si un proceso bivalente estacionario (\mathbf{Y}_t) ($\mathbf{Y}_t \equiv [Y_t, X_t]'$) sigue un modelo ADL(1,1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + V_t,$$

donde $|\beta_1| < 1$, $(V_t) \sim \text{IID}(0, \sigma_V^2)$,

$$X_t = \mu + \phi_1 X_{t-1} + A_t,$$

$|\phi_1| < 1$ y $(A_t) \sim \text{IID}(0, \sigma_A^2)$ [un modelo AR(1) estacionario], entonces

$$(\mathbf{Y}_t) \equiv \left(\begin{bmatrix} Y_t \\ X_t \end{bmatrix} \right) \rightarrow \begin{bmatrix} y^* \\ x^* \end{bmatrix} = \begin{bmatrix} \frac{\beta_0}{1 - \beta_1} + \lambda x^* \\ \frac{\mu}{1 - \phi_1} \end{bmatrix}, \quad (1.10)$$

donde

$$\lambda \equiv \frac{\gamma_0 + \gamma_1}{1 - \beta_1}. \quad (1.11)$$

El modelo ADL(1,1) anterior puede escribirse como un ECM (ver Observación 2 en 1.3.3)

$$\nabla Y_t = \beta_0 - (1 - \beta_1)(Y_{t-1} - \lambda X_{t-1}) + \gamma_0 \nabla X_t + V_t, \quad (1.12)$$

donde el término $(Y_{t-1} - \lambda X_{t-1})$ representa en qué medida no se satisface (salvo quizás por una constante) en el momento $t - 1$ la **relación de equilibrio a largo plazo** entre los dos componentes de (\mathbf{Y}_t) implícita en (1.10); ver Figura 1.14.

El ECM (1.12) implica que cualquier desequilibrio (error) entre Y_{t-1} y X_{t-1} es **transitorio** porque tiende a corregirse a través de variaciones adecuadas en Y_t asociadas con el **término de corrección de error** $-(1 - \beta_1)(Y_{t-1} - \lambda X_{t-1})$, donde $-(1 - \beta_1) > 0$ representa la proporción del desequilibrio entre Y_{t-1} y X_{t-1} que se refleja en $\nabla Y_t \equiv Y_t - Y_{t-1}$.

1.4.3 Ejemplo – El tipo de cambio y la paridad del poder adquisitivo

Sea E_t el **tipo de cambio nominal** de la moneda de un país A con respecto a la moneda de un país B (por ejemplo, E_t podría ser el precio en euros de un franco suizo, medido en euros/franco). Si PA_t y PB_t son dos **índices de precios** referidos a los países A y B, respectivamente, la teoría de la **paridad del poder adquisitivo (PPA)** (en inglés *purchasing power parity*) establece la siguiente relación de equilibrio entre E_t y PA_t / PB_t :

$$E^* = \gamma \frac{PA^*}{PB^*}, \quad (1.13)$$

donde $\gamma \neq 0$ puede ser igual a 1 (PPA **absoluta**) o distinto de 1 (PPA **relativa**, que permite ciertas incompatibilidades entre los índices P_t y P_t^*). Nótese que (1.13) implica que

$$\ln E^* = \beta_0 + \ln \frac{PA^*}{PB^*} \quad (\beta_0 \equiv \ln \gamma), \quad (1.14)$$

de manera que, de acuerdo con la PPA, la **elasticidad a largo plazo** del tipo de cambio nominal con respecto a los precios relativos debe ser unitaria. Alternativamente, (1.14) puede escribirse de otras dos formas:

$$\ln RE^* \equiv \ln \left(E^* \frac{PB^*}{PA^*} \right) = \beta_0, \quad \text{o bien} \quad \ln(E^* PB^*) = \beta_0 + \ln PA^*, \quad (1.15)$$

donde $RE^* \equiv E^* (PB^* / PA^*)$ es el **tipo de cambio real** a largo plazo y $E^* PB^*$ representa el nivel de precios a largo plazo en el país B expresado en la moneda del país A. La PPA enunciada en (1.14) puede contrastarse empíricamente de dos formas basadas en (1.15):

⇨ Contrastar si $(\ln RE_t) \sim I(0)$, donde $RE_t \equiv (E_t \times PB_t) / PA_t$; ver figuras 1.20 y 1.21.

⇒ Si $(\ln E_t PB_t) \sim I(1)$ y $(\ln PA_t) \sim I(1)$, contrastar en el modelo de regresión

$$\ln(E_t PB_t) = \beta_0 + \beta_1 \ln PA_t + V_t$$

si $\beta_1 = 1$ y $(V_t) \sim I(0)$, de manera que $(\ln E_t PB_t)$ y $(\ln PA_t)$ constituyan un proceso bivariente $CI(1,1)$ con vector de cointegración igual a $[1, -1]'$.

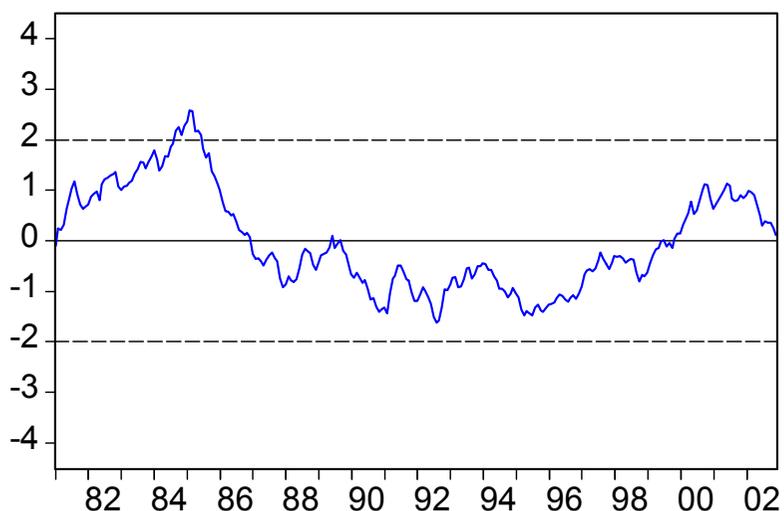


Figura 1.20

Tipo de cambio real ECU-Euro / Dólar EEUU (log)
(exchange-rates.wf1)

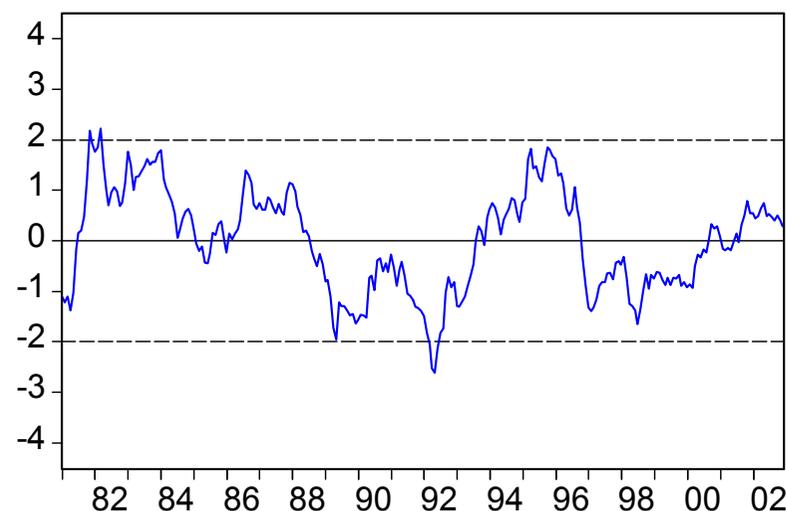


Figura 1.21

Tipo de cambio real ECU-Euro / Franco Suizo (log)
(exchange-rates.wf1)

1.4.4 Ejemplo – La estructura temporal de los tipos de interés

La **estructura temporal de los tipos de interés (ETTI)** (en inglés *term structure of interest rates*) es la relación que existe entre la **rentabilidad** (*yield, interest rate*) y el **plazo hasta el vencimiento** (*term to maturity*) de un conjunto de valores financieros comparables entre sí (por ejemplo, los distintos valores emitidos por el estado **al descuento y sin cupón**).

Si $R_t^{[n]}$ y $R_t^{[m]}$ ($n = km, k > 1$ entero) representan las rentabilidades en el período t de dos valores que vencen dentro de n y m períodos, respectivamente, la **hipótesis de las expectativas (HE)** (*expectations hypothesis*) sobre la ETTI establece que

$$R_t^{[n]} = \frac{1}{k} \sum_{i=0}^{k-1} E_t \left[R_{t+im}^{[m]} \right] + \beta, \quad (1.17)$$

es decir, que la rentabilidad a mayor plazo (n) puede expresarse como una media ponderada de las rentabilidades futuras esperadas a menor plazo (m), más una **prima por plazo** (*term premium*) (β) que es constante. Por ejemplo, si $n = 2$ y $m = 1$ ($k = 2$), (1.17) queda

$$R_t^{[2]} = \frac{1}{2} E_t \left[R_t^{[1]} \right] + \frac{1}{2} E_t \left[R_{t+1}^{[1]} \right] + \beta, \quad (1.18)$$

donde $E_t \left[R_t^{[1]} \right] = R_t^{[1]}$. Por lo tanto, (1.18) también puede escribirse como

$$S_t^{[2,1]} \equiv R_t^{[2]} - R_t^{[1]} = \beta + \frac{1}{2} E_t \left[R_{t+1}^{[1]} - R_t^{[1]} \right] = \beta + \frac{1}{2} E_t \left[\nabla R_{t+1}^{[1]} \right], \quad (1.19)$$

donde $S_t^{[2,1]}$ es el **diferencial** (*spread*) en el período t entre los tipos de interés a plazo 2 (mayor) y a plazo 1 (menor). El diferencial entre rentabilidades para diferentes plazos de vencimiento puede interpretarse como la pendiente de la ETTI.

A **largo plazo**, (1.19) implica la siguiente **relación de equilibrio** entre $R_t^{[2]}$ y $R_t^{[1]}$:

$$S_*^{[2,1]} = \beta, \text{ o bien } R_*^{[2]} = \beta + R_*^{[1]}. \quad (1.20)$$

Por lo tanto, la HE sobre la ETTI enunciada en (1.18)-(1.19) para el caso $n = 2$, $m = 1$, puede contrastarse empíricamente de dos formas basadas en (1.20):

⇒ Contrastar si $(S_t^{[2,1]}) \sim I(0)$; ver figuras 1.22 y 1.23.

⇒ Si $(R_t^{[2]}) \sim I(1)$ y $(R_t^{[1]}) \sim I(1)$, contrastar en el modelo de regresión

$$R_t^{[2]} = \beta_0 + \beta_1 R_t^{[1]} + V_t$$

si $\beta_1 = 1$ y $(V_t) \sim I(0)$, de manera que $(R_t^{[2]})$ y $(R_t^{[1]})$ constituyan un proceso bivalente CI(1,1) con vector de cointegración igual a $[1, -1]'$.

Observación: Las implicaciones anteriores de la HE sobre la formación de la ETTI pueden extenderse teniendo en cuenta que si $(R_t^{[2]})$ está cointegrado con $(R_t^{[1]})$, también deben estarlo $(R_t^{[3]})$, ..., $(R_t^{[n]})$, lo que significa en última instancia que todos los tipos de interés considerados deben compartir la misma tendencia estocástica.

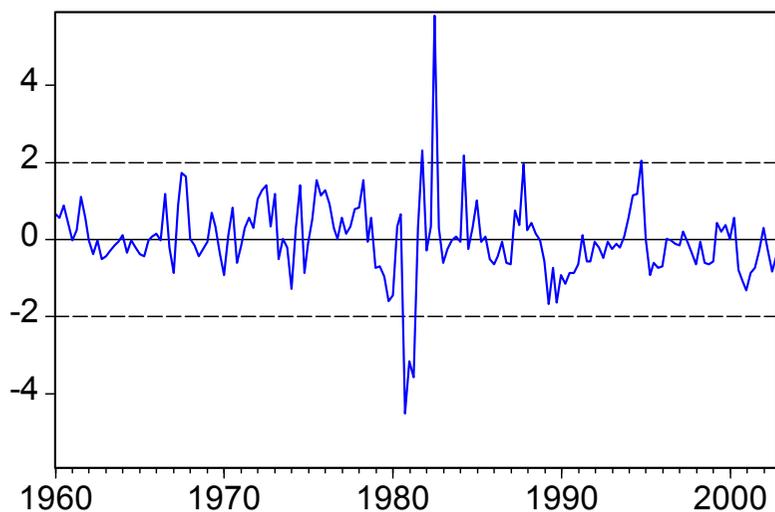


Figura 1.22

Diferencial de la deuda pública (EEUU) a 6 y a 3 meses
(interest-rates-q.wf1)

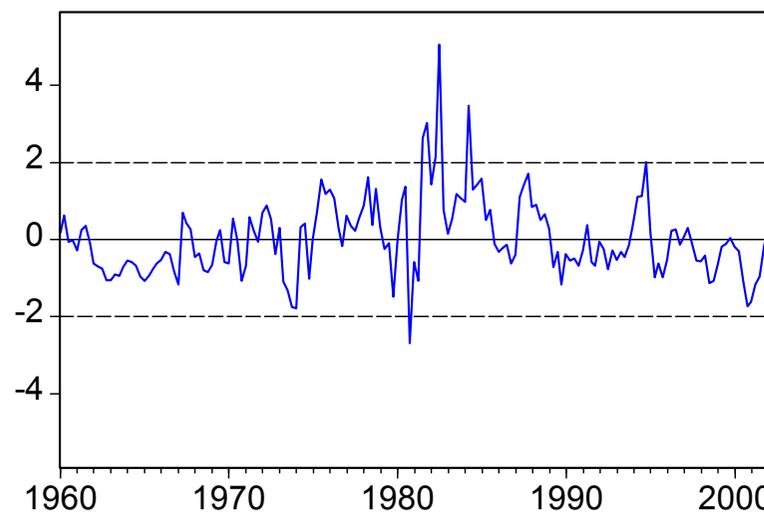
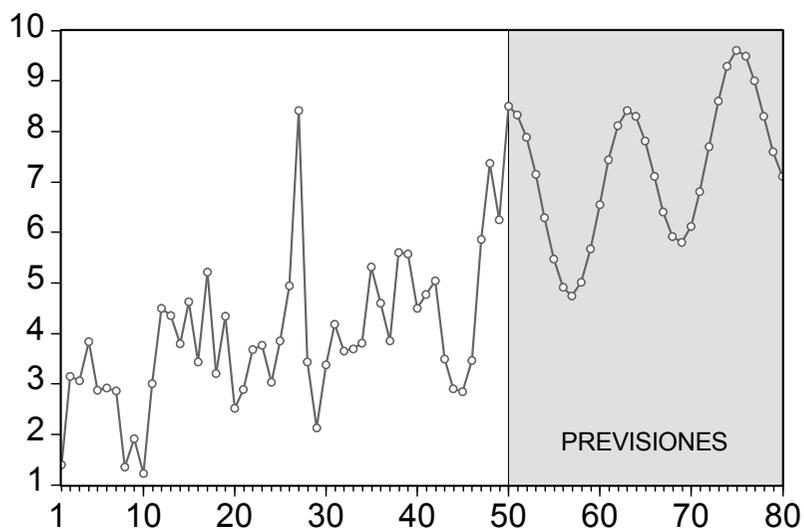


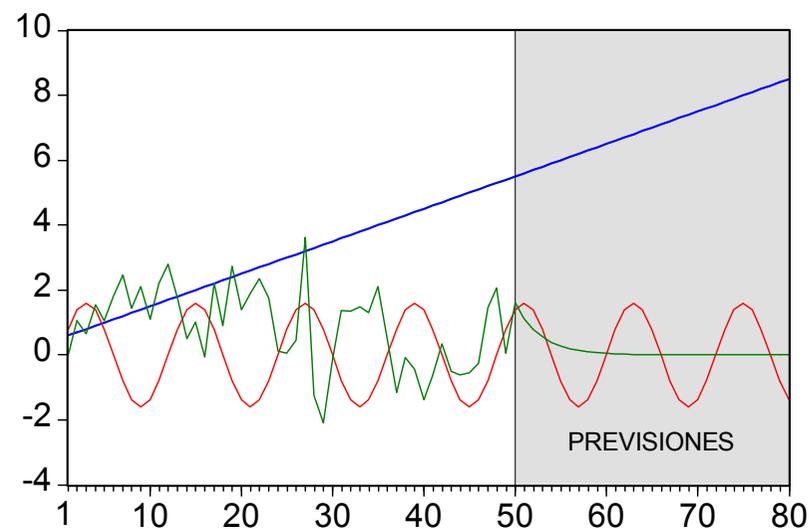
Figura 1.23

Diferencial de la deuda pública (EEUU) a 12 y a 3 meses
(interest-rates-q.wf1)

1.4.5 Ejemplo – Descomposición de series temporales



PANEL A - SERIE TEMPORAL



PANEL B - COMPONENTES

Figura 1.24

Componentes de una serie temporal simulada
(ch1-sim.wf1)

Componentes del Panel B: ■ Tendencia ■ Componente estacional ■ Componente irregular

Observación: La serie del Panel A es la suma de los tres componentes del Panel B, que se han obtenido de la forma siguiente: $T_t = 0.5 + 0.1t$, $S_t = 1.6\text{sen}(\pi t / 6)$, $I_t = 0.7I_{t-1} + A_t$, $A_t \sim \text{NIID}(0, 1)$ ($t = 1, 2, \dots, 50$).

2

MODELOS UNIVARIANTES

2.1 INTRODUCCIÓN

2.1.1 Definición

Un **modelo univariante** para un proceso estocástico univariante o escalar (Y_t) es cualquier conjunto de hipótesis bien definidas sobre ciertas propiedades teóricas de las distribuciones de probabilidad (conjuntas, marginales o condicionales) de los componentes del proceso (Y_t) del que se supone procede una serie temporal observada $(y_t)_{t=1}^N$.

Observación 1: Un modelo de la estructura probabilística completa de (Y_t) consiste en una especificación de las distribuciones conjuntas de todos los vectores del tipo $[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}]'$ ($t_1 < t_2 < \dots < t_n; n = 1, 2, \dots$) que pueden considerarse en relación con (Y_t) . Elaborar un modelo de este tipo requiere al menos estimar el **vector de medias** y la

matriz de varianzas-covarianzas de la muestra $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$ procedente de (Y_t) ,

$$\boldsymbol{\mu} \equiv E[\mathbf{Y}] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix}, \quad \boldsymbol{\Sigma} \equiv \text{Var}[\mathbf{Y}] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{bmatrix},$$

asociada con una serie temporal observada $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$. En conjunto, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ contienen $N + N(N + 1) / 2$ parámetros distintos, que **no** pueden estimarse con precisión utilizando una única serie temporal de N observaciones. Por lo tanto, elaborar un modelo completo para (Y_t) exige suponer alguna simplificación sobre su estructura probabilística que, al menos, reduzca el número de parámetros distintos que contienen $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$.

Observación 2: Un modelo para (Y_t) suele especificarse mediante alguna expresión matemática (por ejemplo, una **ecuación en diferencias**) para Y_t (el componente genérico del proceso considerado), que implique unas **propiedades teóricas** sobre los momentos de primer y segundo orden (medias, varianzas y covarianzas) de las distribuciones conjuntas de los componentes de (Y_t) que sean compatibles con las **propiedades muestrales** correspondientes observadas en una serie temporal $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$. Cuando dichas distribuciones conjuntas son Normales, sus propiedades de primer y segundo orden caracterizan completamente la estructura probabilística de (Y_t) .

Observación 3: Un modelo univariante se utiliza en la práctica: (i) simplemente para proporcionar una **descripción compacta** de la procedencia de los datos que conforman una serie temporal escalar, (ii) para calcular **previsiones a corto plazo** de los valores futuros desconocidos de dicha serie, o (iii) como punto de partida para analizar posibles **relaciones** entre dicha serie y otras series, en el contexto de algún modelo multivariante o vectorial.

2.1.2 Ejemplo

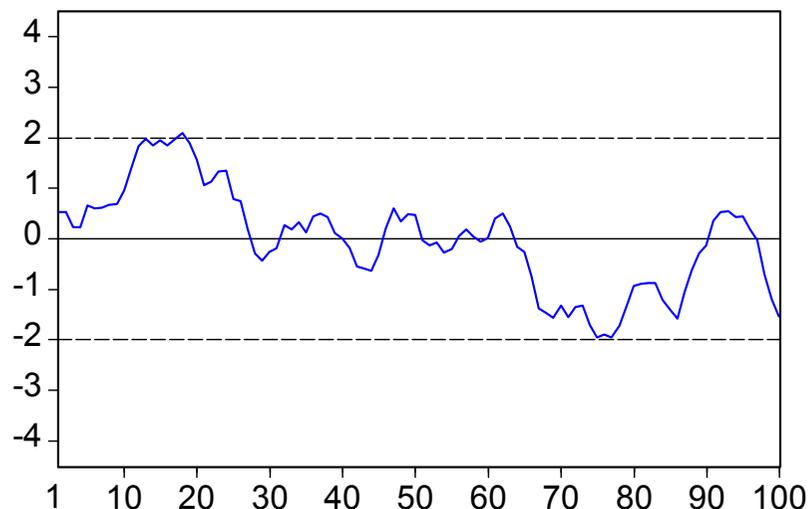


Figura 2.1

Serie no estacionaria simulada
(ch2-sim.wf1)

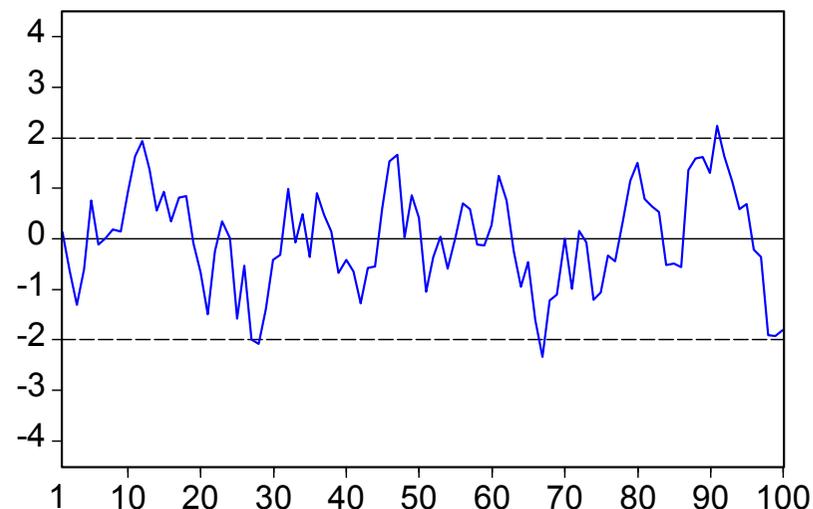
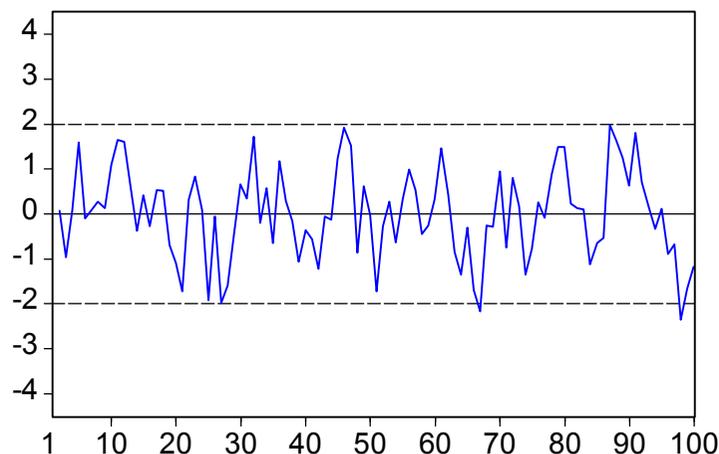


Figura 2.2

Serie estacionaria simulada
(ch2-sim.wf1)

Observación 1: La serie de la Figura 2.1 ha sido generada por un proceso estocástico (Z_t) **no estacionario** tal que $Z_t = 1.5Z_{t-1} - 0.5Z_{t-2} + A_t$ (una ecuación en diferencias de orden 2). La serie de la Figura 2.2 ha sido generada por un proceso estocástico (Y_t) **estacionario** tal que $Y_t = 0.8Y_{t-1} + A_t$ (una ecuación en diferencias de orden 1). En ambos casos, $(A_t) \sim \text{NIID}(0, 1)$.

Observación 2: El proceso estocástico (Z_t) que ha generado la serie de la Figura 2.1 no parece tener una media que sea constante en el tiempo; al menos por este motivo, la estructura probabilística de (Z_t) es más complicada que si su media fuera constante (como parece ocurrir con el proceso estocástico que ha generado la serie de la Figura 2.2). No obstante, la ecuación en diferencias $Z_t = 1.5Z_{t-1} - 0.5Z_{t-2} + A_t$ es equivalente a $\nabla Z_t = 0.5\nabla Z_{t-1} + A_t$, donde $\nabla Z_t \equiv Z_t - Z_{t-1}$ es la **diferencia regular de orden 1** de Z_t . La figura siguiente contiene la diferencia regular de orden 1 de la serie de la Figura 2.1:



Esta figura sugiere que la ecuación en diferencias $Z_t = 1.5Z_{t-1} - 0.5Z_{t-2} + A_t$ describe un proceso estocástico no estacionario cuya diferencia regular de orden uno (∇Z_t) es un proceso estacionario, similar al proceso (Y_t) que ha generado la serie de la Figura 2.2. En ambos casos, todas las propiedades relevantes de primer y segundo orden de (∇Z_t) y de (Y_t) están contenidas en las dos ecuaciones en diferencias de orden 1 $\nabla Z_t = 0.5\nabla Z_{t-1} + A_t$ e $Y_t = 0.8Y_{t-1} + A_t$; en particular, dichas ecuaciones implican que (i) $E[\nabla Z_t] = E[Y_t] = 0$, (ii) $\text{Var}[\nabla Z_t] \approx 1.33$, $\text{Var}[Y_t] \approx 2.78$, y (iii) $\text{Corr}[\nabla Z_t, \nabla Z_{t+k}] = 0.5^k$, $\text{Corr}[Y_t, Y_{t+k}] = 0.8^k$ ($k = 1, 2, \dots$).

2.1.3 Etapas en la construcción de un modelo univariante

- ⇒ **Identificación:** Selección de un modelo que implique ciertas propiedades teóricas para (Y_t) compatibles con las propiedades muestrales observadas en $(y_t)_{t=1}^N$.
- ⇒ **Estimación:** Asignación de valores numéricos a los parámetros del modelo.
- ⇒ **Diagnosis:** Comprobación del ajuste del modelo a los datos utilizados.

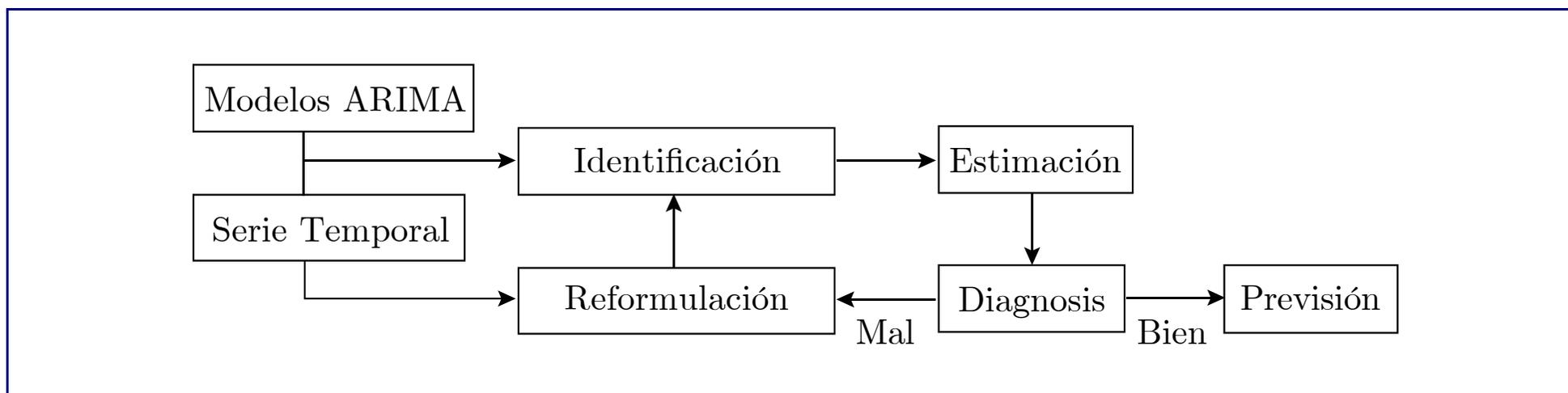


Figura 2.3 – Proceso de construcción de un modelo ARIMA univariante

2.2 PROCESOS ESTOCÁSTICOS ESTACIONARIOS

2.2.1 Definición

Un proceso estocástico (Y_t) es **estrictamente estacionario** si y sólo si para cualesquiera $n \geq 1$ momentos $t_1 < t_2 < \dots < t_n$ de su historia, la distribución de probabilidad conjunta de $[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}]'$ coincide con la de $[Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}]'$ para cualquier número entero $h = \pm 1, \pm 2, \dots$ (distinto de cero).

2.2.2 Definición

Un proceso estocástico (Y_t) con $E[Y_t] < \infty$ para todo $t = 0, \pm 1, \pm 2, \dots$ es **estacionario en media** o **débilmente estacionario de primer orden** si y sólo si $E[Y_t]$ es constante (no depende de t) para todo $t = 0, \pm 1, \pm 2, \dots$

2.2.3 Definición

Un proceso estocástico (Y_t) con $E[Y_t^2] < \infty$ para todo $t = 0, \pm 1, \pm 2, \dots$ es **estacionario en autocovarianza** o **débilmente estacionario de segundo orden** si y sólo si:

- $E[Y_t]$ y $\text{Var}[Y_t]$ son constantes (no dependen de t) para todo $t = 0, \pm 1, \pm 2, \dots$,
- $\text{Cov}[Y_t, Y_{t+k}]$ depende a lo sumo de k (entero) pero no de t para todo $t = 0, \pm 1, \pm 2, \dots$

2.2.4 Definición

Un proceso estocástico (Y_t) es **Normal** o **Gaussiano** cuando para cualesquiera $n \geq 1$ momentos $t_1 < t_2 < \dots < t_n$ de su historia, la distribución de probabilidad conjunta de $[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}]'$ es una distribución Normal n -variante.

Estacionariedad en autocovarianza + Normalidad \Rightarrow Estacionariedad estricta

Figura 2.4 – Estacionariedad

Observación: En general, el tipo de estacionariedad definido en 2.2.3 es suficiente para todos los planteamientos que se abordan en esta asignatura. En adelante, cuando se hable de estacionariedad sin calificativos se entenderá que se está hablando de estacionariedad en autocovarianza. Cuando un proceso estocástico (Y_t) es estacionario, suele decirse que (Y_t) es un proceso **integrado de orden 0**, o **I(0)**, lo que se representa como $(Y_t) \sim I(0)$ (aunque esta terminología no es del todo correcta; ver Definición 2.4.5).

ACF – PACF teóricas

2.2.5 Definición

La **autocovarianza de orden k** ($k > 0$) de un proceso (Y_t) estacionario se representa con el símbolo γ_k y se define como $\gamma_k \equiv \text{Cov}[Y_t, Y_{t+k}]$.

Media:

$$\mu_Y \equiv E[Y_t].$$

Varianza:

$$\sigma_Y^2 \equiv \text{Var}[Y_t] \equiv E[(Y_t - \mu_Y)^2].$$

Autocovarianza de orden k : $\gamma_k \equiv \text{Cov}[Y_t, Y_{t+k}] \equiv E[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]$ ($k = 1, 2, \dots$).

Figura 2.5 – Media, varianza y autocovarianzas de un proceso estacionario

Observación 1: La autocovarianza de orden k de (Y_t) es la covarianza entre cualquier par de componentes de (Y_t) separados entre sí por un intervalo temporal o **retardo** $k > 0$ dado. Por 2.2.3, γ_k puede depender de k , pero no depende de los momentos concretos a los que se refieran los componentes de (Y_t) considerados.

Observación 2: Aunque γ_k se ha definido para cualquier $k > 0$, también puede definirse la autocovarianza de orden cero de un proceso (Y_t) estrictamente estacionario como $\gamma_0 \equiv \text{Cov}[Y_t, Y_t] \equiv \text{Var}[Y_t]$, que es la varianza del proceso

(Y_t) . Por otro lado, la autocovarianza de orden k de (Y_t) también puede definirse como $\gamma_k \equiv \text{Cov}[Y_{t-k}, Y_t]$, ya que para cualquier valor de $k \geq 0$ dado, Y_{t-k} e Y_t están separados entre sí por el mismo intervalo temporal (retardo) que Y_t e Y_{t+k} . Considerada como una función del retardo k , la secuencia $(\gamma_k : k = 0, 1, 2, \dots)$ se denomina **la función de autocovarianza** del proceso estacionario (Y_t) .

2.2.6 Definición

La **autocorrelación simple de orden k** ($k > 0$) de un proceso (Y_t) estacionario se representa con el símbolo ρ_k y se define como

$$\rho_k \equiv \frac{\text{Cov}[Y_t, Y_{t+k}]}{\text{Var}[Y_t]^{1/2} \text{Var}[Y_{t+k}]^{1/2}} \equiv \frac{\gamma_k}{\gamma_0}. \quad (2.1)$$

Autocorrelación simple de orden k : $\rho_k \equiv \frac{\gamma_k}{\gamma_0} \quad (k = 1, 2, \dots); \quad \rho_0 = 1.$

Figura 2.6 – Autocorrelaciones simples de un proceso estacionario

Observación 1: La autocorrelación simple de orden k de (Y_t) es el coeficiente de correlación lineal simple entre cualquier par de componentes de (Y_t) separados entre sí por un retardo $k > 0$ dado. Por 2.2.3, ρ_k puede depender de k , pero no depende de los momentos concretos a los que se refieran los componentes de (Y_t) considerados.

Observación 2: Considerada como una función del retardo k , la secuencia $(\rho_k : k = 1, 2, \dots)$ se denomina **la función de autocorrelación simple (ACF, del inglés *AutoCorrelation Function*)** del proceso estacionario (Y_t) . Dado que cada ρ_k es un coeficiente de correlación, suele decirse que la ACF de (Y_t) representa la **duración** y la **intensidad** de la **memoria** del proceso (Y_t) .

Observación 3: Elaborar un modelo de la estructura probabilística completa de (Y_t) requiere al menos estimar el **vector de medias** y la **matriz de varianzas-covarianzas** de la **muestra** $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$,

$$\boldsymbol{\mu} \equiv E[\mathbf{Y}] = \begin{bmatrix} \mu_Y \\ \mu_Y \\ \vdots \\ \mu_Y \end{bmatrix}, \quad \boldsymbol{\Sigma} \equiv \text{Var}[\mathbf{Y}] = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \cdots & \gamma_0 \end{bmatrix} = \sigma_Y^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \cdots & \rho_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \cdots & 1 \end{bmatrix},$$

asociada con una serie temporal observada $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$. En conjunto, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ contienen $1 + N$ parámetros distintos: $\mu_Y, \gamma_0, \gamma_1, \dots, \gamma_{N-1}$, o bien $\mu_Y, \sigma_Y^2, \rho_1, \dots, \rho_{N-1}$. Las dos matrices que figuran en la expresión anterior suelen representarse como $\boldsymbol{\Gamma}_N$ (**matriz de autocovarianzas** de \mathbf{Y}) y \mathbf{P}_N (**matriz de autocorrelaciones** de \mathbf{Y}), de manera que la matriz de varianzas-covarianzas de la muestra \mathbf{Y} procedente de (Y_t) es $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_N = \sigma_Y^2 \mathbf{P}_N$.

Observación 4: A pesar de que la hipótesis de estacionariedad simplifica notablemente la distribución de probabilidad de $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$, aún no pueden estimarse con precisión $1 + N$ parámetros utilizando tan sólo una única realización particular de \mathbf{Y} (una única serie temporal $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ de N observaciones). La solución a este problema consiste en expresar la media y la función de autocovarianza (o la media, la varianza y la ACF) de (Y_t) en términos de un número reducido de parámetros a través de algún **modelo ARMA** para (Y_t) ; ver Sección 2.3.

2.2.7 Definición

La **autocorrelación parcial de orden k** ($k > 0$) de un proceso (Y_t) estacionario se representa con el símbolo ϕ_{kk} y se define como el parámetro ϕ_{kk} en la regresión

$$\tilde{Y}_t = \phi_{k1}\tilde{Y}_{t-1} + \phi_{k2}\tilde{Y}_{t-2} + \dots + \phi_{kk}\tilde{Y}_{t-k} + U_t, \quad (2.2)$$

donde $\tilde{Y}_{t-i} \equiv Y_{t-i} - \mu_Y$ ($i = 0, 1, \dots, k$) y U_t es independiente de Y_{t-i} para todo $i \geq 1$.

Observación 1: La regresión (2.2) anterior puede escribirse como $Y_t = \phi_{k0} + \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \dots + \phi_{kk}Y_{t-k} + U_t$, donde $\phi_{k0} \equiv (1 - \phi_{k1} - \phi_{k2} - \dots - \phi_{kk})\mu_Y$, lo que implica que $\Delta Y_t = \phi_{kk} \times \Delta Y_{t-k}$ cuando $\Delta Y_{t-1} = \Delta Y_{t-2} = \dots = \Delta Y_{t-k+1} = \Delta U_t = 0$ (es decir, ϕ_{kk} representa el **efecto parcial** o **ceteris paribus** de Y_{t-k} sobre Y_t). Por lo tanto, ϕ_{kk} es una medida del grado de asociación lineal entre dos componentes cualesquiera de (Y_t) separados entre sí por un retardo $k \geq 1$ dado (como Y_t e Y_{t-k}) que **no** es debida a la posible correlación entre cada uno de ellos y todos los componentes de (Y_t) que se encuentran entre ambos ($Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$).

Observación 2: A partir de la regresión anterior, puede comprobarse que $\phi_{kk} = |\mathbf{A}_k| / |\mathbf{B}_k|$, donde

$$\mathbf{A}_k \equiv \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-4} & \rho_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{bmatrix}, \quad \mathbf{B}_k \equiv \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-4} & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{bmatrix}.$$

La matriz \mathbf{B}_k es una matriz de autocorrelaciones del mismo tipo que \mathbf{P}_N en la Observación 3 de 2.2.6. La matriz \mathbf{A}_k es idéntica a \mathbf{B}_k excepto porque la última columna de \mathbf{B}_k está reemplazada por el vector columna $[\rho_1, \rho_2, \dots, \rho_k]'$. En particular, para $k = 1, 2, 3$ resulta que

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}},$$

de manera que los coeficientes de autocorrelación parcial son funciones de los coeficientes de autocorrelación simple de un proceso estrictamente estacionario.

Observación 3: Considerada como una función del retardo k , la secuencia $(\phi_{kk} : k = 1, 2, \dots)$ se denomina **la función de autocorrelación parcial (PACF)**, del inglés *Partial AutoCorrelation Function*) del proceso (Y_t) .

Autocorrelación parcial de orden k : $\phi_{kk} = \frac{|\mathbf{A}_k|}{|\mathbf{B}_k|} \quad (k = 1, 2, \dots); \quad \phi_{11} = \rho_1.$

Figura 2.7 – Autocorrelaciones parciales de un proceso estacionario

ACF – PACF muestrales

- ⇒ En la práctica, es imposible estimar a partir de una serie temporal $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ de N observaciones, la media, la varianza, la ACF y la PACF del proceso estacionario (Y_t) del que supuestamente procede dicha serie. En particular, sólo en el vector de medias y en la matriz de varianzas-covarianzas de la muestra $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$ asociada con la serie $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ figuran $1 + N$ parámetros distintos.
- ⇒ No obstante, en la práctica **sí** pueden estimarse con cierta precisión las K primeras autocorrelaciones simples y parciales de (Y_t) a partir de una serie temporal de N observaciones, siempre que K sea un número bastante más reducido que N .
- ⇒ Dichas estimaciones resultan muy útiles para identificar un modelo inicial para (Y_t) dentro de la clase modelos para procesos estacionarios que se describe en la Sección 2.3.

2.2.8 Definición

La **media muestral** y la **varianza muestral** de una muestra $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$ asociada con una serie temporal $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ son

$$\hat{\mu}_Y \equiv \frac{1}{N} \sum_{t=1}^N Y_t \quad \text{y} \quad \hat{\sigma}_Y^2 \equiv \frac{1}{N} \sum_{t=1}^N (Y_t - \hat{\mu}_Y)^2, \quad (2.3)$$

respectivamente, que suelen emplearse como **estimadores** de la media μ_Y y la varianza σ_Y^2 del proceso estacionario (Y_t) del que procede \mathbf{Y} .

Observación 1: Los estimadores $\hat{\mu}_Y$ y $\hat{\sigma}_Y^2$ proporcionan **estimaciones** numéricas \bar{y} y s_y^2 , respectivamente,

$$\bar{y} \equiv \frac{1}{N} \sum_{t=1}^N y_t, \quad s_y^2 \equiv \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2,$$

cuando se reemplaza en (2.3) cada variable aleatoria Y_t por su valor observado y_t ($t = 1, \dots, N$).

Observación 2: Para contrastar hipótesis sobre μ_Y (el verdadero valor de la media de un proceso estacionario) del estilo de $\mu_Y = m$ frente a alternativas del tipo $\mu_Y \neq m$, $\mu_Y > m$ ó $\mu_Y < m$ (donde m es un número dado, como 0), puede utilizarse un estadístico t basado en que

$$E[\hat{\mu}_Y] = \mu_Y, \quad \text{Var}[\hat{\mu}_Y] \approx \frac{\sigma_Y^2}{N} \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right).$$

Por lo tanto, el estadístico $t_{\mathbf{Y}} = (\hat{\mu}_Y - m) / (\hat{\sigma}_Y^2 / N)^{1/2}$, cuyo valor calculado es $t = (\bar{y} - m) / (s_y^2 / N)^{1/2}$, sigue aproximadamente una distribución $t(N - 1)$.

2.2.9 Definición

La **correlación simple muestral de orden k** ($k > 0$) de una muestra $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$

asociada con una serie temporal $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$ es

$$\hat{\rho}_k \equiv \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (k = 1, 2, \dots) \quad (2.4)$$

(ver Figura 2.6), donde

$$\begin{aligned} \hat{\gamma}_k &\equiv \frac{1}{N} \sum_{t=1}^{N-k} (Y_t - \hat{\mu}_Y)(Y_{t+k} - \hat{\mu}_Y) \\ &\equiv \frac{1}{N} \sum_{t=k+1}^N (Y_{t-k} - \hat{\mu}_Y)(Y_t - \hat{\mu}_Y) \quad (k = 0, 1, 2, \dots). \end{aligned} \quad (2.5)$$

Observación 1: El estimador $\hat{\gamma}_0$ en (2.5) es el mismo estimador que $\hat{\sigma}_Y^2$ en (2.3). Los **estimadores** $\hat{\gamma}_k$ y $\hat{\rho}_k$ proporcionan **estimaciones** numéricas c_k y r_k , respectivamente,

$$\begin{aligned} c_k &\equiv \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \equiv \frac{1}{N} \sum_{t=k+1}^N (y_{t-k} - \bar{y})(y_t - \bar{y}) \quad (k = 0, 1, 2, \dots), \\ r_k &= \frac{c_k}{c_0} \quad (k = 1, 2, \dots), \end{aligned}$$

cuando se reemplaza en (2.5) cada variable aleatoria Y_t por su valor observado y_t ($t = 1, \dots, N$).

Observación 2: La secuencia de valores numéricos ($r_k : k = 1, 2, \dots$) se denomina la **ACF muestral** de la serie temporal y_1, y_2, \dots, y_N . La representación gráfica de la ACF muestral de una serie se denomina el **correlograma** de dicha serie.

Observación 3: Bajo ciertas condiciones, $(\hat{\rho}_k) \sim \text{NIID}(0, 1/N)$, de manera que cualquier autocorrelación simple ρ_k ($k \geq 1$) de (Y_t) puede considerarse *individualmente* significativa al 5% cuando $r_k > 1.96/\sqrt{N}$. Por otro lado, para determinar si las K primeras autocorrelaciones simples de (Y_t) son *conjuntamente* significativas (donde K es un entero bastante más reducido que N), suele emplearse el valor calculado del **estadístico de Ljung-Box**

$$Q_{LB} = N(N-2) \sum_{k=1}^K \frac{r_k^2}{N-k},$$

que bajo la hipótesis nula de que $\rho_1 = \rho_2 = \dots = \rho_K = 0$, sigue aproximadamente una distribución $\chi^2(K)$.

2.2.10 Definición

La **correlación parcial muestral de orden k** ($k > 0$) de una muestra $\mathbf{Y} \equiv [Y_1, Y_2, \dots, Y_N]'$ es el estimador MCO o MV $\hat{\phi}_{kk}$ del parámetro ϕ_{kk} en el modelo de regresión (2.2) considerado para $t = 1, 2, \dots, N$.

Observación 1: Otro estimador razonable para ϕ_{kk} es el que se obtiene al evaluar el cociente de la Figura 2.7, reemplazando en las matrices \mathbf{A}_k y \mathbf{B}_k cada autocorrelación simple teórica ρ_k por su estimador (2.4). La secuencia de valores numéricos $(r_{kk} : k = 1, 2, \dots)$ que se obtiene al aplicar los estimadores $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots$ a una serie y_1, y_2, \dots, y_N , se denomina la **PACF muestral** de dicha serie.

Observación 2: Bajo ciertas condiciones, $(\hat{\phi}_{kk}) \sim \text{NIID}(0, 1/N)$, de manera que cualquier autocorrelación parcial ϕ_{kk} ($k \geq 1$) de (Y_t) puede considerarse *individualmente* significativa al 5% cuando $r_{kk} > 1.96/\sqrt{N}$.

2.2.11 Ejemplo

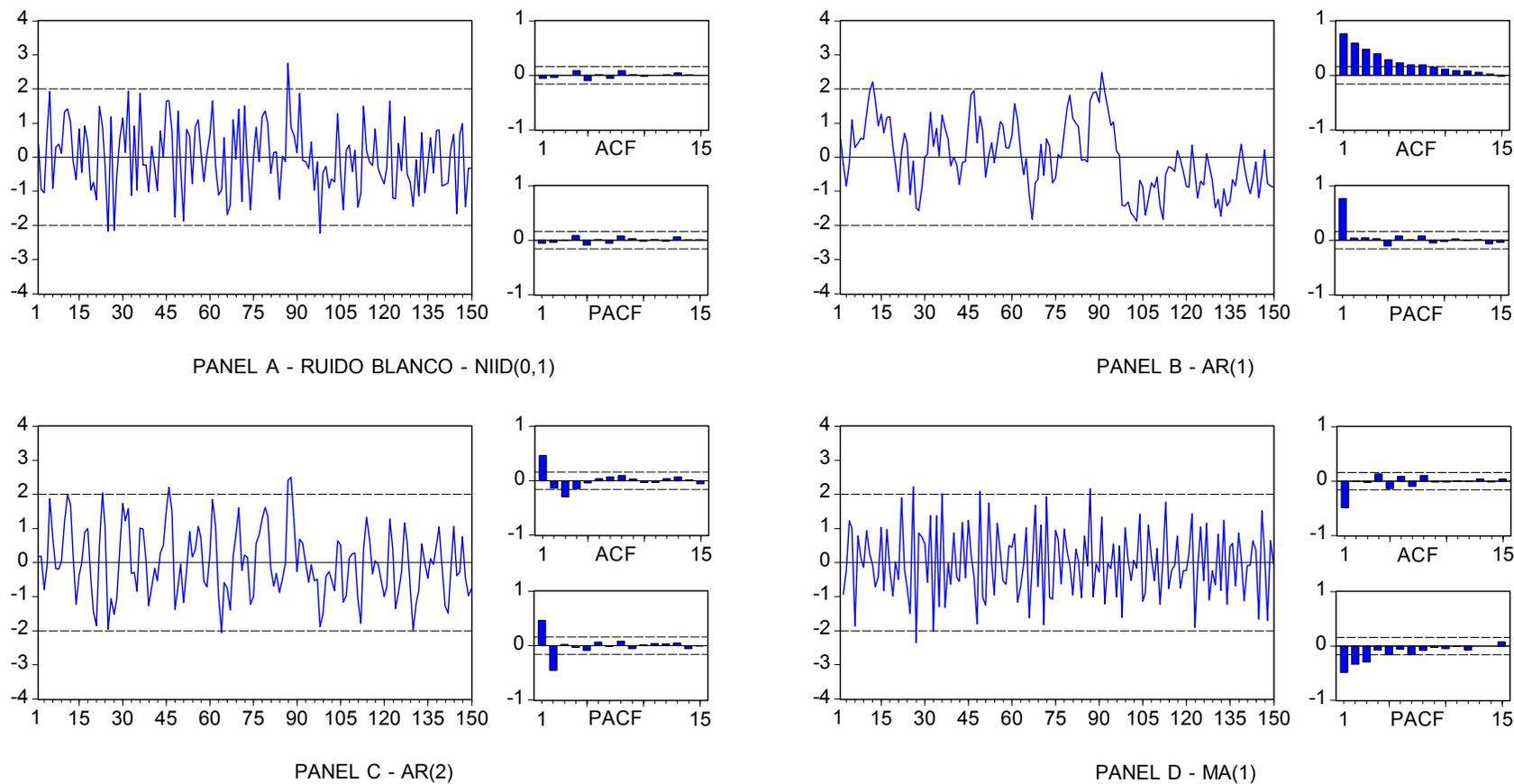
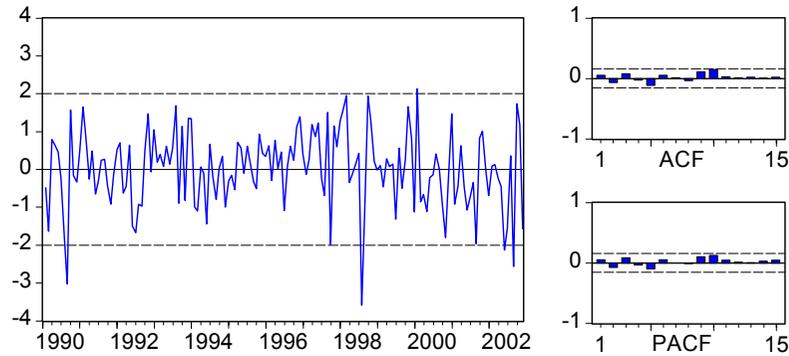


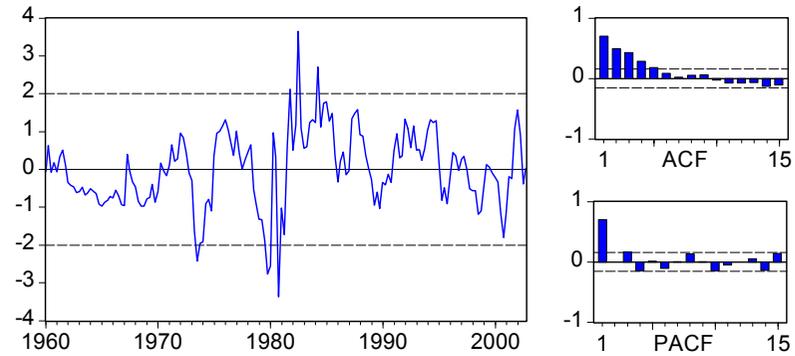
Figura 2.8

Funciones de autocorrelación muestrales de varias series estacionarias simuladas (sim-ARIMA.wf1)

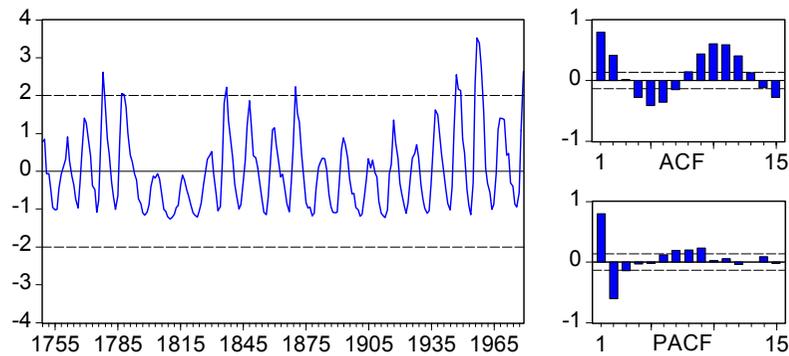
2.2.12 Ejemplo



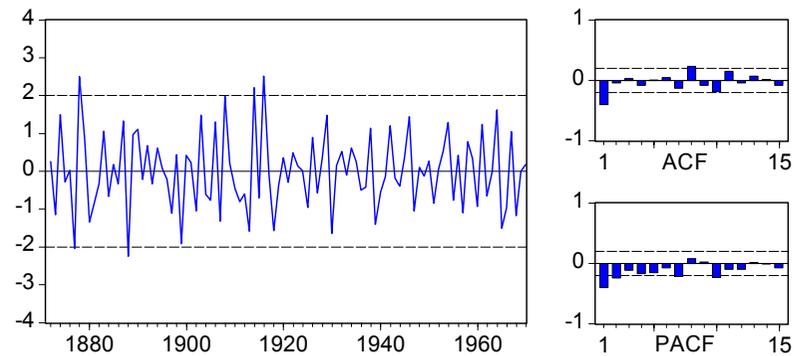
PANEL A - TASA LOGARÍTMICA DE VARIACIÓN MENSUAL DEL IBEX35



PANEL B - DIFERENCIAL DE LA DEUDA (EEUU) A 36 Y A 3 MESES



PANEL C - MANCHAS SOLARES

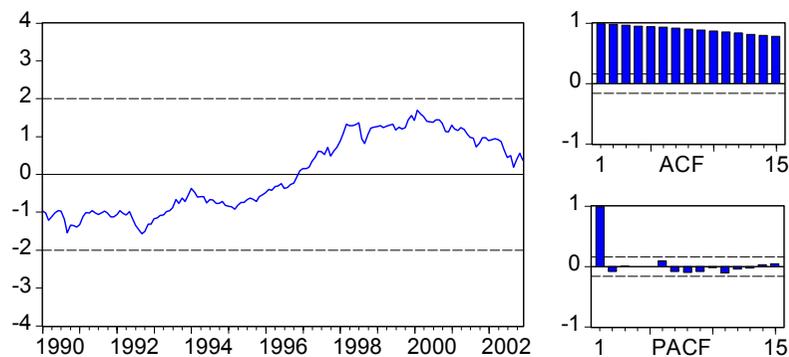


PANEL D - VARIACIÓN ANUAL DEL FLUJO DEL RÍO NILO

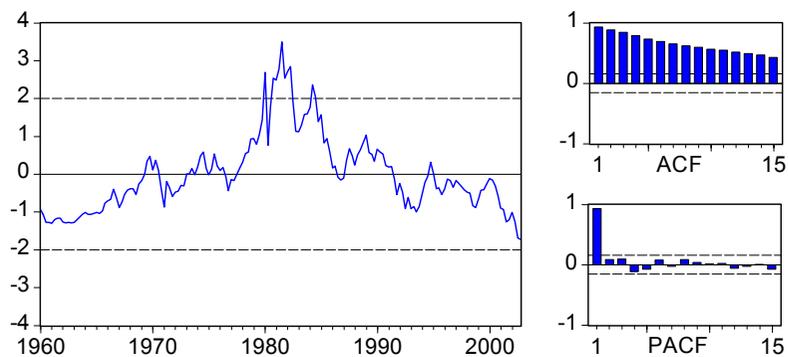
Figura 2.9

Funciones de autocorrelación muestrales de varias series estacionarias reales
(Panel A: bolsa-int.wf1. Panel B: interest-rates-q.wf1. Panel C: sunspots.wf1. Panel D: nilo.wf1)

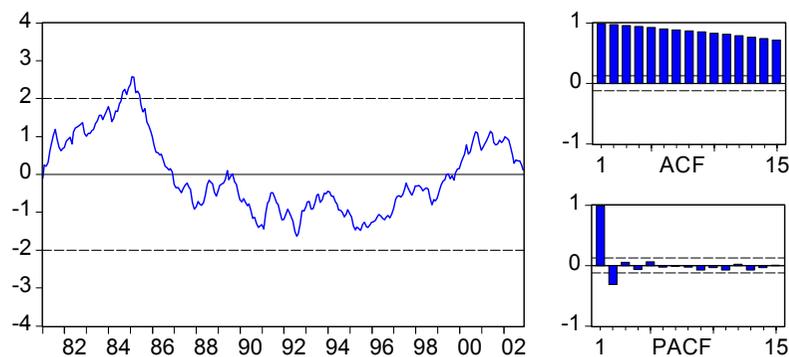
2.2.13 Ejemplo



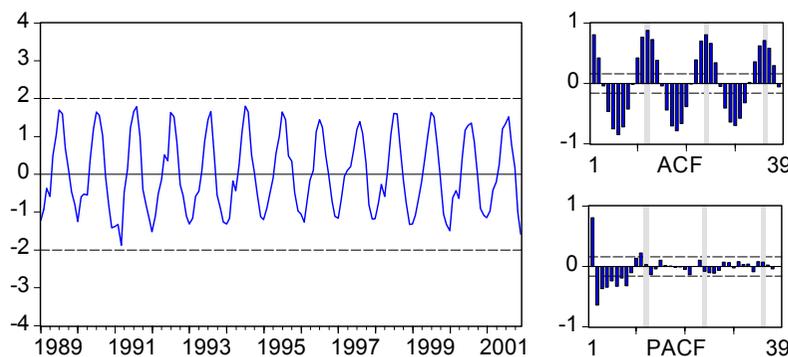
PANEL A - IBEX 35 (LOG)



PANEL B - RENTABILIDAD DE LA DEUDA (EEUU) A 36 MESES



PANEL C - TIPO DE CAMBIO REAL EURO / DÓLAR (LOG)



PANEL D - TEMPERATURA EN MADRID-RETIRO

Figura 2.10

Funciones de autocorrelación muestrales de varias series no estacionarias reales
 (Panel A: bolsa-int.wf1. Panel B: interest-rates-q.wf1. Panel C: exchange-rates.wf1. Panel D: seasonal.wf1)

Observación 1: Aunque la ACF y la PACF muestrales se han definido en 2.2.8-2.2.10 en relación con una serie temporal estacionaria, también pueden calcularse a partir de cualquier secuencia de datos. En particular, la ACF muestral de una **serie no estacionaria** suele decrecer muy lentamente, mientras que la ACF muestral de una **serie estacional** suele presentar el mismo tipo de comportamiento periódico que dicha serie. Por lo tanto, el correlograma de una serie temporal es un instrumento visual muy útil para decidir sobre la estacionariedad de dicha serie.

Observación 2: La ACF y la PACF muestrales de una serie temporal estacionaria son dos instrumentos fundamentales para la **identificación** de un modelo ARMA inicial para el proceso estocástico (Y_t) del que supuestamente procede dicha serie. Por otro lado, la ACF y la PACF muestrales de la **serie de residuos** asociada con un modelo estimado para (Y_t) , también son dos instrumentos fundamentales en la **diagnos**is de dicho modelo.

2.2.14 Procesos estacionarios y series no estacionarias

- ⇒ Muchas series temporales (como las del Ejemplo 2.2.13) no pueden considerarse generadas por procesos estocásticos estacionarios (es decir, son series **no estacionarias**), porque presentan ciertas **tendencias** claras en su evolución temporal (de manera que no presentan afinidad hacia algún valor constante en el tiempo), porque su variabilidad no es constante, porque son **estacionales**, o por varias combinaciones de estos motivos.
- ⇒ No obstante, muchas series temporales no estacionarias se pueden **transformar** de forma adecuada para obtener series de aspecto estacionario (como las de los ejemplos 2.2.11 y

2.2.12), que pueden ser utilizadas para elaborar modelos estadísticos compatibles con la hipótesis de estacionariedad. Estos modelos se presentan en la Sección 2.3.

⇒ En la secciones 2.4 y 2.5 se presentan algunas técnicas sencillas para detectar series no estacionarias y para transformar adecuadamente dichas series (dos operaciones que se llevan a cabo al principio de cualquier análisis empírico), así como los modelos estadísticos que resultan al combinar dichas transformaciones con los modelos estacionarios de la sección 2.3.

2.3 MODELOS ARMA

2.3.1 Definición

Un proceso estocástico **estacionario** (Y_t) sigue un **modelo autorregresivo-media móvil** de orden (p,q) , o ARMA(p,q) (del inglés *AutoRegressive-Moving Average*), si y sólo si

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \dots - \theta_q A_{t-q} \quad (2.6)$$

para todo $t = 0, \pm 1, \pm 2, \dots$, donde $(A_t) \sim \text{IID}(0, \sigma_A^2)$ y $\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ son parámetros tales que todas las raíces de la ecuación polinomial

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \quad (2.7)$$

están fuera del círculo unitario (**condición de estacionariedad**). Un modelo ARMA(p,q) descrito por la ecuación (2.6) es **invertible** si todas las raíces de la ecuación polinomial

$$1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0 \quad (2.8)$$

están fuera del círculo unitario (**condición de invertibilidad**).

2.3.2 Definición

El **operador de retardo** se representa con el símbolo B (a veces L , del inglés *Backshift* o *Lag operator*) y se define de acuerdo con que

$$BX_t \equiv X_{t-1}, B^d X_t \equiv X_{t-d} \quad (d \geq 1 \text{ entero}), \quad (2.9)$$

donde X_t es una variable (real o aleatoria) referida a un momento t determinado.

2.3.3 Definición

La ecuación (2.6) puede escribirse alternativamente como

$$\phi(B)Y_t = \mu + \theta(B)A_t, \quad (2.10)$$

donde

$$\phi(B) \equiv 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.11)$$

es el **operador** o **polinomio autorregresivo** (AR) del modelo [comparar con (2.7)] y

$$\theta(B) \equiv 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.12)$$

es el **operador** o **polinomio media móvil** (MA) [comparar con (2.8)].

Estacionariedad – Invertibilidad

⇒ Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(p, q) descrito por (2.6) ó (2.10), la esperanza incondicional de (Y_t) es

$$\mu_Y \equiv E[Y_t] = \frac{\mu}{1 - \phi_1 - \phi_2 - \dots - \phi_p} = \frac{\mu}{\phi(1)}, \quad (2.13)$$

donde $\phi(1)$ es el valor del operador AR (2.11) cuando $B = 1$. Por lo tanto, (2.10) puede escribirse alternativamente como

$$\phi(B)(Y_t - \mu_Y) = \theta(B)A_t, \quad \text{o bien} \quad \phi(B)\tilde{Y}_t = \theta(B)A_t, \quad (2.14)$$

donde $\tilde{Y}_t \equiv Y_t - E[Y_t] \equiv Y_t - \mu_Y$ para todo $t = 0, \pm 1, \pm 2, \dots$

⇒ La condición de estacionariedad de un modelo ARMA enunciada en 2.3.1 garantiza que los parámetros o **pesos** $\psi_0, \psi_1, \psi_2, \dots$ del polinomio de orden infinito

$$\psi(B) \equiv \frac{\theta(B)}{\phi(B)} \equiv 1 + \psi_1 B + \psi_2 B^2 + \dots \equiv \sum_{i=0}^{\infty} \psi_i B^i \quad (\psi_0 \equiv 1) \quad (2.15)$$

satisfacen la condición $\sum_{i=0}^{\infty} |\psi_i| < \infty$, que es suficiente para que un proceso (Y_t) tal que $\tilde{Y}_t = \psi(B)A_t$ sea estacionario (**Teorema de Wold**).

⇒ La condición de invertibilidad de un modelo ARMA enunciada en 2.3.1 garantiza que los parámetros o **pesos** $\pi_0, \pi_1, \pi_2, \dots$ del polinomio de orden infinito

$$\pi(B) \equiv \frac{\phi(B)}{\theta(B)} \equiv 1 - \pi_1 B - \pi_2 B^2 - \dots \equiv - \sum_{i=0}^{\infty} \pi_i B^i \quad (\pi_0 \equiv -1) \quad (2.16)$$

satisfacen la condición $\sum_{i=0}^{\infty} |\pi_i| < \infty$, lo que implica que cuando (2.14) se escribe como $\pi(B)\tilde{Y}_t = A_t$, (Y_t) es un proceso estacionario tal que la correlación **parcial** entre Y_t e Y_{t-k} ($k > 0$) tiende a cero a medida que el retardo k aumenta.

Observación: Dados unos valores para los parámetros $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ de un modelo ARMA, los pesos ψ y los pesos π de (2.15) y (2.16) pueden calcularse recursivamente teniendo en cuenta que $\phi(B)\psi(B) = \theta(B)$ y que $\theta(B)\psi(B) = \phi(B)$, de manera que, por un lado,

$$\psi_i = \phi_1 \psi_{i-1} + \phi_2 \psi_{i-2} + \dots + \phi_p \psi_{i-p} - \theta_i$$

para todo $i > 0$ (donde $\psi_0 = 1$, $\psi_i = 0$ si $i < 0$ y $\theta_i = 0$ si $i > q$), y, por otro lado,

$$\pi_i = \theta_1 \pi_{i-1} + \theta_2 \pi_{i-2} + \dots + \theta_q \pi_{i-q} + \phi_i$$

para todo $i > 0$ (donde $\pi_0 = -1$, $\pi_i = 0$ si $i < 0$ y $\phi_i = 0$ si $i > p$). Por último, de la relación $\psi(B)\pi(B) = 1$ implícita en (2.15) y (2.16) se deduce que

$$\pi_i = -\sum_{j=0}^{i-1} \pi_j \psi_{i-j} \quad (\pi_0 = -1), \quad \psi_i = \sum_{j=0}^{i-1} \psi_j \pi_{i-j} \quad (\psi_0 = 1) \quad \text{para todo } i \geq 1.$$

Las secuencias $(\psi_i : i = 1, 2, \dots)$ y $(\pi_i : i = 1, 2, \dots)$ calculadas a partir de un modelo ARMA(p, q) para un proceso estacionario (Y_t) , siguen exactamente las mismas pautas que la ACF y la PACF teóricas de (Y_t) , respectivamente, implicadas por dicho modelo.

ACF – PACF teóricas en modelos ARMA

⇒ Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(p, q) descrito por (2.6), (2.10) ó (2.14), la esperanza, la varianza, la ACF y la PACF de (Y_t) dependen únicamente de los $p + q + 2$ parámetros $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \mu$ y $\sigma_A^2)$ que figuran en dicho modelo.

2.3.4 Ejemplo – Modelo autorregresivo de primer orden o AR(1)

Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(1,0) o AR(1)

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t \quad (2.17.1)$$

con $|\phi_1| < 1$ (ver 2.3.1) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$, puede comprobarse lo siguiente:

Media:
$$\mu_Y = \frac{\mu}{1 - \phi_1}. \quad (2.17.2)$$

ACF:
$$\rho_k = \phi_1 \rho_{k-1} \quad (k \geq 1). \quad (2.17.3)$$

Observación: La expresión anterior es una ecuación en diferencias de orden 1 que puede resolverse recursivamente a partir de la condición inicial $\rho_0 = 1$ (ver Figura 2.6), lo que proporciona $\rho_k = \phi_1^k$ para todo $k \geq 0$ (una **exponencial amortiguada**).

PACF:
$$\phi_{kk} = \begin{cases} \phi_1 & \text{si } k = 1, \\ 0 & \text{para todo } k > 1. \end{cases} \quad (2.17.4)$$

Varianza:
$$\sigma_Y^2 = \frac{\sigma_A^2}{1 - \phi_1 \rho_1} = \frac{\sigma_A^2}{1 - \phi_1^2}. \quad (2.17.5)$$

2.3.5 Ejemplo – Modelo autorregresivo de segundo orden o AR(2)

Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(2,0) o AR(2)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + A_t \quad (2.18.1)$$

con $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$ (ver 2.3.1) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$, puede comprobarse lo siguiente:

Media:
$$\mu_Y = \frac{\mu}{1 - \phi_1 - \phi_2}. \quad (2.18.2)$$

ACF:
$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad (k \geq 1). \quad (2.18.3)$$

Observación: La expresión anterior es una ecuación en diferencias que puede resolverse recursivamente a partir de las condiciones iniciales $\rho_0 = 1$ (ver Figura 2.6) y $\rho_1 = \phi_1 / (1 - \phi_2)$ [que se obtiene de (2.18.3) con $k = 1$; por otro lado, las dos ecuaciones que resultan de (2.18.3) con $k = 1$ y $k = 2$ se denominan **ecuaciones de Yule-Walker**]. En particular, si las dos raíces de (2.7) con $p = 2$ son complejas, la solución de (2.18.3) es

$$\rho_k = \frac{D^k \text{sen}(2\pi fk + F)}{\text{sen}(F)} \quad \text{para todo } k \geq 0,$$

cuyo numerador representa una **oscilación armónica amortiguada** de **amplitud** D^k , **período** $1/f$, **frecuencia** f , **fase inicial** $-F / (2\pi f)$ y **factor de amortiguamiento** D ; estas características de ρ_k son funciones de los parámetros ϕ_1 y ϕ_2 de acuerdo con que $D = \sqrt{-\phi_2}$, $\cos(2\pi f) = \phi_1 / (2\sqrt{-\phi_2})$ y $\tan(F) = [(1 - \phi_2) / (1 + \phi_2)] \tan(2\pi f)$.

PACF:
$$\phi_{kk} = \begin{cases} \phi_1 / (1 - \phi_2) & \text{si } k = 1, \\ \phi_2 & \text{si } k = 2, \\ 0 & \text{para todo } k > 2. \end{cases} \quad (2.18.4)$$

$$\text{Varianza: } \sigma_Y^2 = \frac{\sigma_A^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2} = \left[\frac{1 - \phi_2}{1 + \phi_2} \right] \left[\frac{\sigma_A^2}{(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)} \right]. \quad (2.18.5)$$

2.3.6 Ejemplo – Modelo media móvil de primer orden o MA(1)

Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(0,1) o MA(1)

$$Y_t = \mu + A_t - \theta_1 A_{t-1} \quad (2.19.1)$$

con $|\theta_1| < 1$ (ver 2.3.1) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$, puede comprobarse lo siguiente:

$$\text{Media: } \mu_Y = \mu. \quad (2.19.2)$$

$$\text{ACF: } \rho_k = \begin{cases} -\theta_1 / (1 + \theta_1^2) & \text{si } k = 1, \\ 0 & \text{para todo } k > 1. \end{cases} \quad (2.19.3)$$

$$\text{PACF: } \phi_{kk} = - \left[\frac{1}{\sum_{i=0}^k \theta_1^{2i}} \right] \theta_1^k = - \left[\frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}} \right] \theta_1^k \quad \text{para todo } k \geq 1. \quad (2.19.4)$$

$$\text{Varianza: } \sigma_Y^2 = (1 + \theta_1^2) \sigma_A^2. \quad (2.19.5)$$

2.3.7 Ejemplo – Modelo media móvil de segundo orden o MA(2)

Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(0,2) o MA(2)

$$Y_t = \mu + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} \quad (2.20.1)$$

con $\theta_2 + \theta_1 < 1$, $\theta_2 - \theta_1 < 1$, $|\theta_2| < 1$ (ver 2.3.1) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$, puede comprobarse lo siguiente:

Media:
$$\mu_Y = \mu. \quad (2.20.2)$$

ACF:
$$\rho_k = \begin{cases} -[\theta_1(1 - \theta_2)] / (1 + \theta_1^2 + \theta_2^2) & \text{si } k = 1, \\ -\theta_2 / (1 + \theta_1^2 + \theta_2^2) & \text{si } k = 2, \\ 0 & \text{para todo } k > 2. \end{cases} \quad (2.20.3)$$

Varianza:
$$\sigma_Y^2 = (1 + \theta_1^2 + \theta_2^2) \sigma_A^2. \quad (2.20.4)$$

Observación: La expresión analítica de la PACF en un modelo MA(2) depende de la naturaleza de las dos raíces de (2.8) con $q = 2$; en general, dicha expresión es muy complicada. No obstante, la PACF sigue exactamente las mismas pautas que la serie $(\pi_i : i = 1, 2, \dots)$ asociada con (2.16) cuando $\phi(B) = 1$ y $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$. En otros términos, la PACF en un modelo MA(2) es semejante a la ACF en un modelo AR(2); ver Observación en 2.3.5.

2.3.8 Ejemplo – Modelo autorregresivo-media móvil de orden (1,1) o ARMA(1,1)

Cuando un proceso estacionario (Y_t) sigue un modelo ARMA(1,1)

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t - \theta_1 A_{t-1} \quad (2.21.1)$$

con $|\phi_1| < 1$, $|\theta_1| < 1$ (ver 2.3.1) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$, puede comprobarse lo siguiente:

Media:
$$\mu_Y = \frac{\mu}{1 - \phi_1}. \quad (2.21.2)$$

ACF:
$$\rho_k = \begin{cases} [(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)] / (1 - 2\phi_1 \theta_1 + \theta_1^2) & \text{si } k = 1, \\ \rho_1 \phi_1^{k-1} & \text{para todo } k > 1. \end{cases} \quad (2.21.3)$$

Varianza:
$$\sigma_Y^2 = \frac{1 - 2\phi_1 \theta_1 + \theta_1^2}{1 - \phi_1^2} \sigma_A^2 = \left[1 + \frac{(\phi_1 - \theta_1)^2}{1 - \phi_1^2} \right] \sigma_A^2. \quad (2.21.4)$$

Observación: La PACF en un modelo ARMA(1,1) consiste en un valor inicial $\phi_{11} = \rho_1$ seguido de una secuencia de valores que siguen exactamente la pauta de la PACF en un modelo MA(1); ver Ejemplo 2.3.6. Esto es lo mismo que ocurre con la ACF (2.21.3), que consiste en un valor inicial ρ_1 seguido de una secuencia de valores que siguen exactamente la pauta de la ACF en un modelo AR(1); ver Ejemplo 2.3.4.

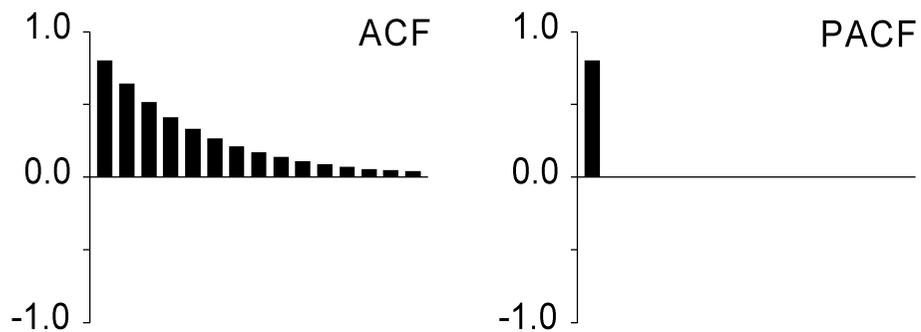


Figura 2.11

AR(1): $Y_t = 0.8Y_{t-1} + A_t$.
Parámetro positivo.

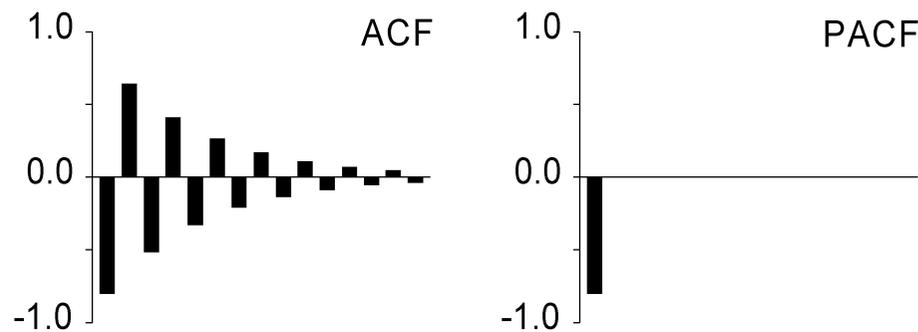


Figura 2.12

AR(1): $Y_t = -0.8Y_{t-1} + A_t$.
Parámetro negativo.

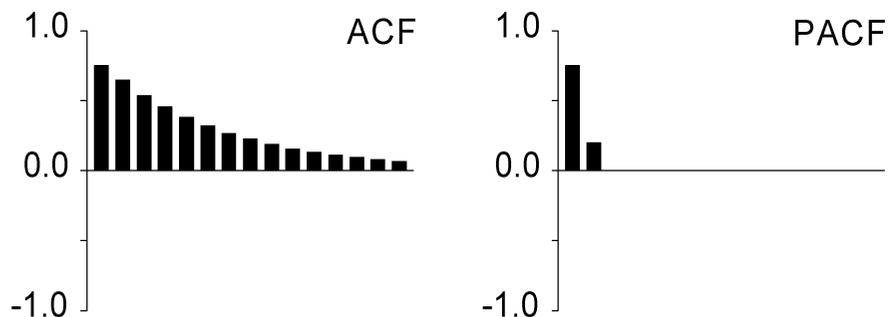


Figura 2.13

AR(2): $Y_t = 0.6Y_{t-1} + 0.2Y_{t-2} + A_t$.
Raíces de (2.7) reales: -4.19, +1.19 (dominante).

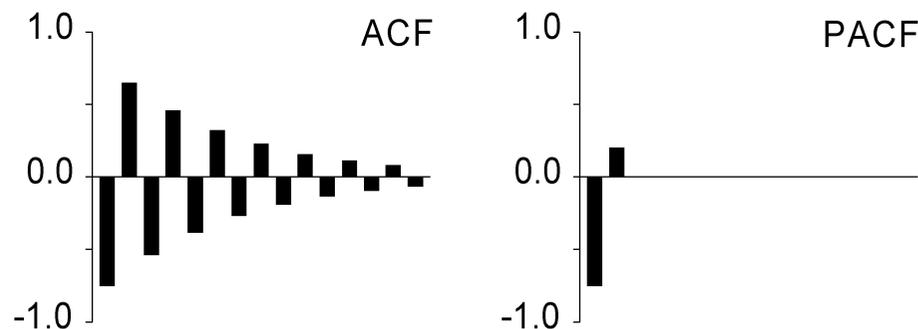


Figura 2.14

AR(2): $Y_t = -0.6Y_{t-1} + 0.2Y_{t-2} + A_t$.
Raíces de (2.7) reales: +4.19, -1.19 (dominante).

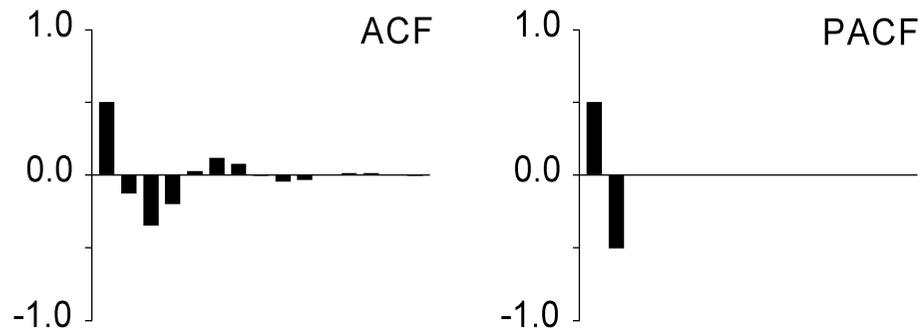


Figura 2.15

AR(2): $Y_t = 0.75Y_{t-1} - 0.5Y_{t-2} + A_t$.
 Raíces de (2.7) complejas: $+0.75 \pm 1.20i$.

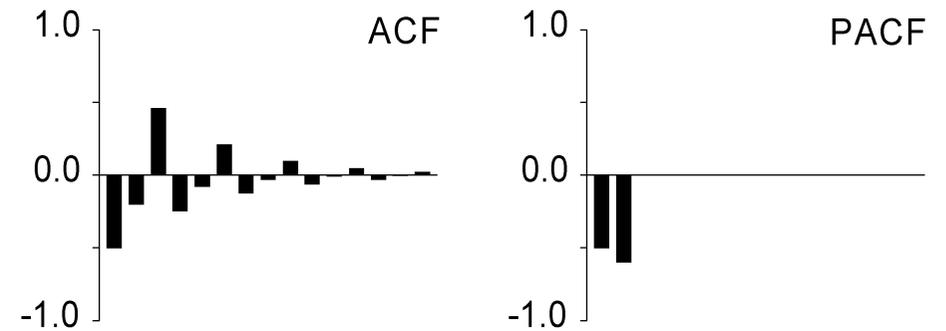


Figura 2.16

AR(2): $Y_t = -0.8Y_{t-1} - 0.6Y_{t-2} + A_t$.
 Raíces de (2.7) complejas: $-0.67 \pm 1.11i$.

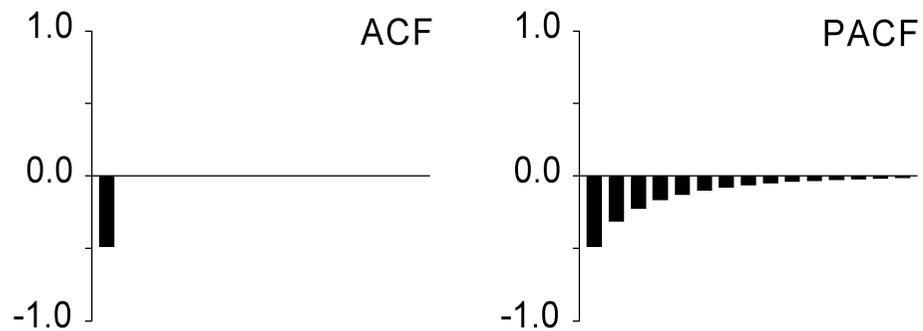


Figura 2.17

MA(1): $Y_t = A_t - 0.8A_{t-1}$.
 Parámetro positivo.

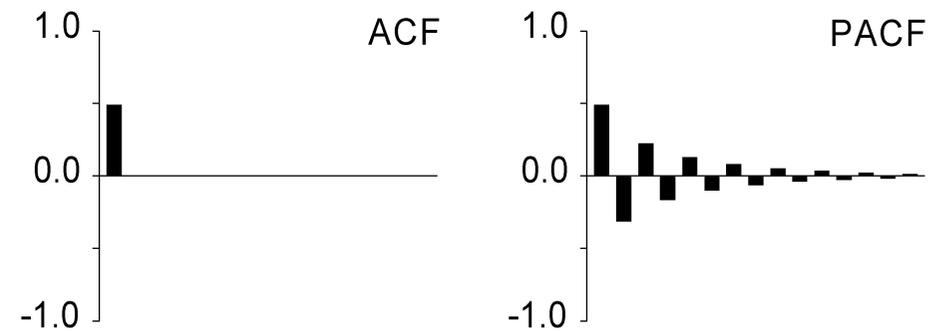


Figura 2.18

MA(1): $Y_t = A_t + 0.8A_{t-1}$.
 Parámetro negativo.

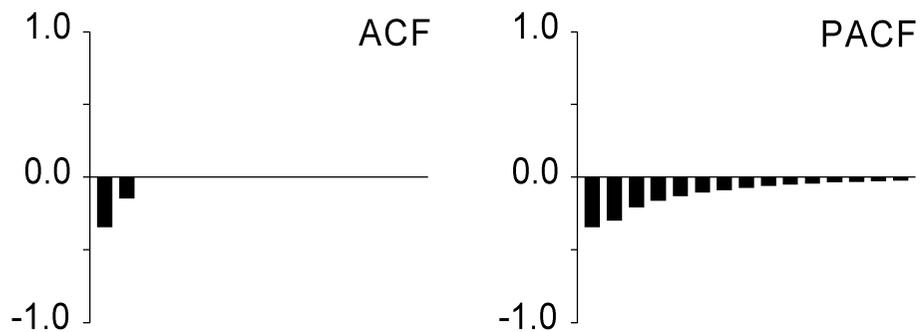


Figura 2.19

MA(2): $Y_t = A_t - 0.6A_{t-1} - 0.2A_{t-2}$.
 Raíces de (2.7) reales: $-4.19, +1.19$ (dominante).

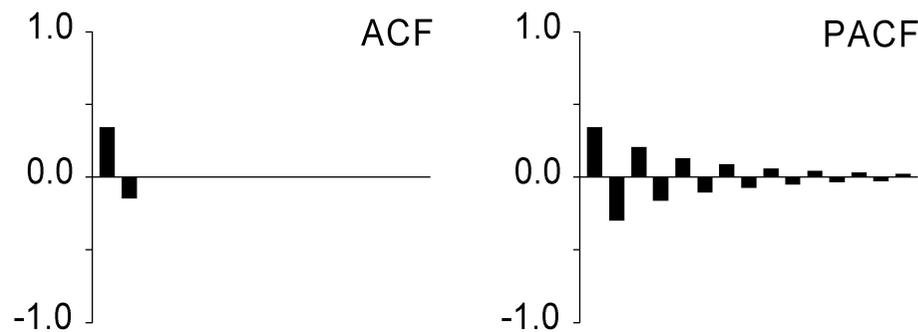


Figura 2.20

MA(2): $Y_t = A_t + 0.6A_{t-1} - 0.2A_{t-2}$.
 Raíces de (2.7) reales: $+4.19, -1.19$ (dominante).

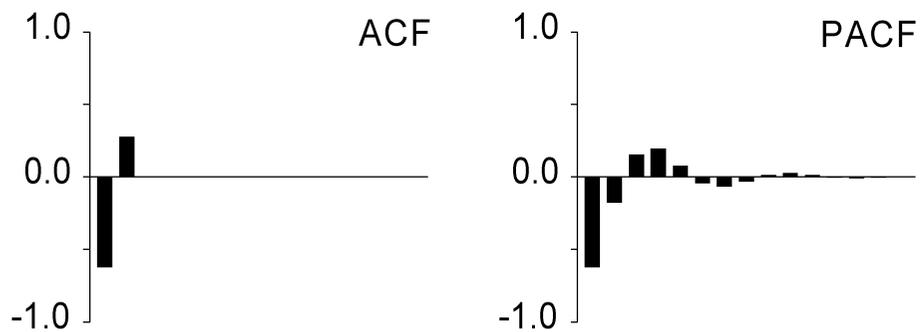


Figura 2.21

MA(2): $Y_t = A_t - 0.75A_{t-1} + 0.5A_{t-2}$.
 Raíces de (2.7) complejas: $+0.75 \pm 1.20i$.

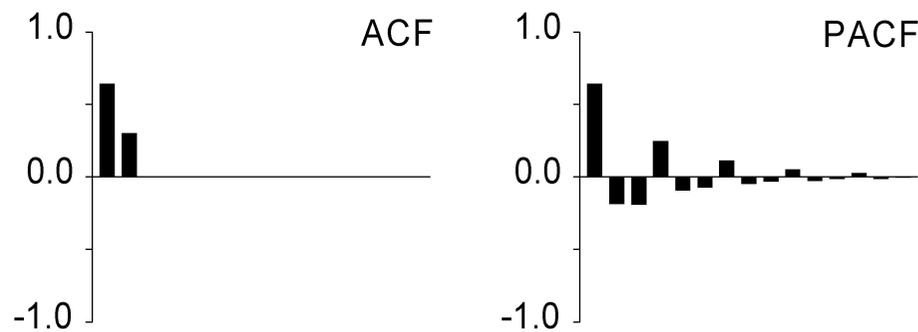


Figura 2.22

MA(2): $Y_t = A_t + 0.8A_{t-1} + 0.6A_{t-2}$.
 Raíces de (2.7) complejas: $-0.67 \pm 1.11i$.

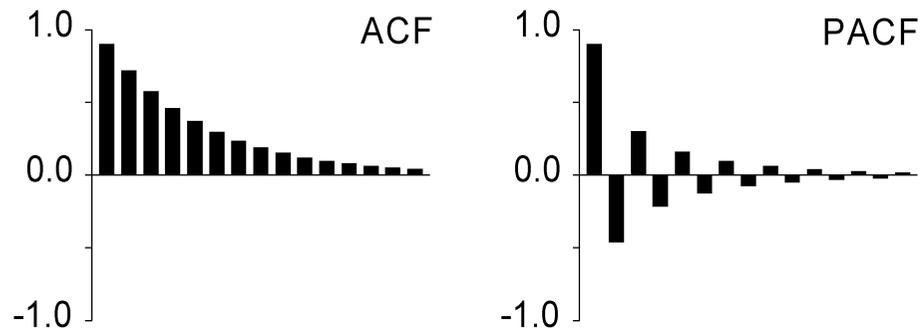


Figura 2.23

$$\text{ARMA}(1,1): Y_t = 0.8Y_{t-1} + A_t + 0.8A_{t-1}.$$

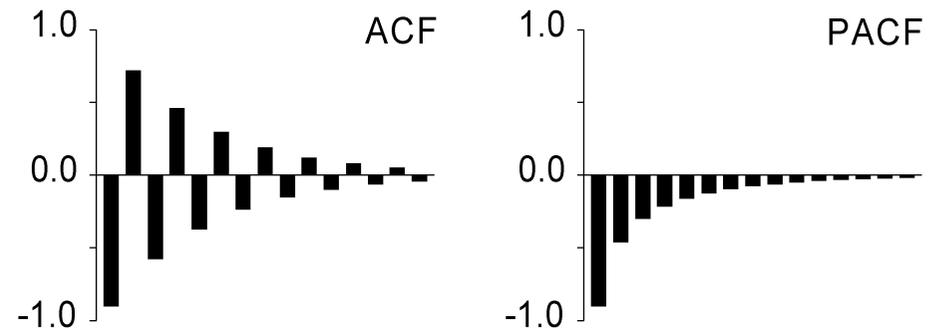


Figura 2.24

$$\text{ARMA}(1,1): Y_t = -0.8Y_{t-1} + A_t - 0.8A_{t-1}.$$

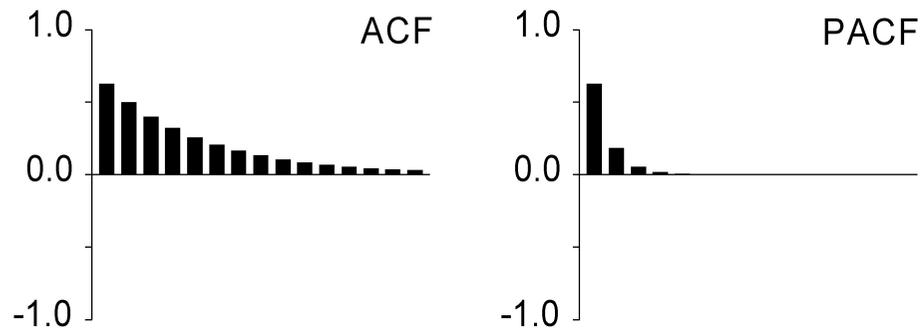


Figura 2.25

$$\text{ARMA}(1,1): Y_t = 0.8Y_{t-1} + A_t - 0.3A_{t-1}.$$

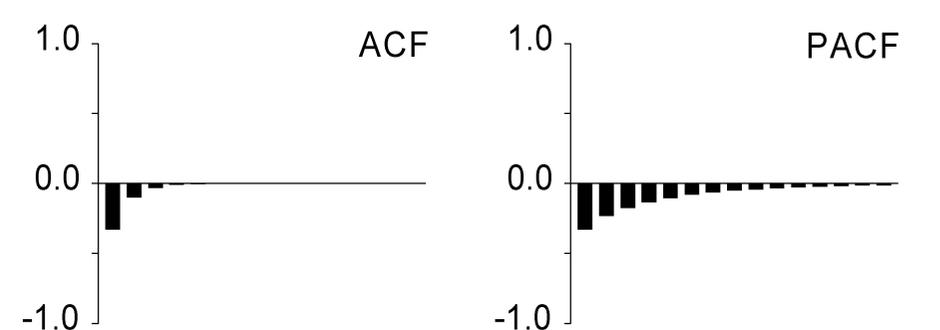


Figura 2.26

$$\text{ARMA}(1,1): Y_t = 0.3Y_{t-1} + A_t - 0.8A_{t-1}.$$

	Modelo AR(p)	Modelo MA(q)	Modelo ARMA(p, q)
Representación de \tilde{Y}_t en términos de \tilde{Y}_{t-i}	$\phi(B)\tilde{Y}_t = A_t$	$\theta^{-1}(B)\tilde{Y}_t = A_t$	$\theta^{-1}(B)\phi(B)\tilde{Y}_t = A_t$
Representación de \tilde{Y}_t en términos de A_{t-i}	$\tilde{Y}_t = \phi^{-1}(B)A_t$	$\tilde{Y}_t = \theta(B)A_t$	$\tilde{Y}_t = \phi^{-1}(B)\theta(B)A_t$
Operador $\pi(B)$	$\pi(B) = \phi(B)$ (finito)	$\pi(B) = \theta^{-1}(B)$ (infinito)	$\pi(B) = \theta^{-1}(B)\phi(B)$ (infinito)
Operador $\psi(B)$	$\psi(B) = \phi^{-1}(B)$ (infinito)	$\psi(B) = \theta(B)$ (finito)	$\psi(B) = \phi^{-1}(B)\theta(B)$ (infinito)
Condición de estacionariedad	Raíces de $\phi(x) = 0$ fuera del círculo unitario	Siempre estacionario	Raíces de $\phi(x) = 0$ fuera del círculo unitario
Condición de invertibilidad	Siempre invertible	Raíces de $\theta(x) = 0$ fuera del círculo unitario	Raíces de $\theta(x) = 0$ fuera del círculo unitario
ACF teórica	Infinita: Mezcla de exponenciales y oscilaciones amortiguadas	Finita: Igual a 0 después de $k = q$	Infinita: Mezcla de exponenciales y oscilaciones amortiguadas después de $k = q - p$
PACF teórica	Finita: Igual a 0 después de $k = p$	Infinita: Dominada por mezcla de exponenciales y oscilaciones amortiguadas	Infinita: Dominada por mezcla de exponenciales y oscilaciones amortiguadas después de $k = p - q$

Observación 1: El archivo `sim-ARIMA.wf1` contiene (entre otras) varias series temporales simuladas (generadas) con las instrucciones del archivo `sim-ARIMA.prg` para EViews 3.1 a partir de todos los modelos ARMA cuyas ACF y PACF teóricas aparecen representadas en las figuras 2.11-2.23 anteriores. Todas las series se han obtenido a partir de la serie RB, que es una realización particular de tamaño 150 de un proceso de ruido blanco Gaussiano con varianza igual a uno. El modelo utilizado para generar cada serie aparece en la vista por defecto en EViews (`SpreadSheet`) de cada serie y en el comentario que precede a cada grupo de instrucciones en el programa `sim-ARIMA.prg`. Las series del archivo `sim-ARIMA.wf1` pueden utilizarse para observar su aspecto en un gráfico temporal estandarizado, para examinar el aspecto de las ACF y PACF muestrales correspondientes (como en la Figura 2.8), y para comparar dichas funciones de autocorrelación muestrales con las funciones de autocorrelación teóricas de las figuras 2.11-2.23.

Observación 2: Cuando una serie temporal estacionaria (y_t) se observa $S > 1$ veces al año, como por ejemplo en el caso de series mensuales ($S = 12$) o trimestrales ($S = 4$), es muy frecuente que un **modelo ARMA multiplicativo**, o $\text{ARMA}(p, q) \times \text{ARMA}(P, Q)_S$, del tipo

$$\phi_p(B)\Phi_P(B^S)Y_t = \mu + \theta_q(B)\Theta_Q(B^S)A_t$$

resulte útil para describir la evolución temporal del proceso estacionario (Y_t) que ha generado dicha serie. En este modelo, $\phi_p(B)$ y $\theta_q(B)$ son los operadores AR (estacionario) y MA (invertible) de (2.11) y (2.12), que en este contexto suelen denominarse **operadores regulares**; $\Phi_P(B^S) \equiv 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}$ (con las raíces de la ecuación $1 - \sum_{i=1}^P \Phi_i x^i = 0$ fuera del círculo unitario) y $\Theta_Q(B^S) \equiv 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$ (con las raíces de la ecuación $1 - \sum_{i=1}^Q \Theta_i x^i = 0$ fuera del círculo unitario), que suelen denominarse **operadores anuales**; y $(A_t) \sim \text{IID}(0, \sigma_A^2)$. En un modelo de este tipo, los operadores regulares se emplean para modelizar la correlación entre pares de componentes de (Y_t) separados entre sí por $k = 1, 2, 3, \dots$ períodos básicos (por ejemplo, meses o trimestres), mientras que los operadores anuales describen la correlación entre pares de componentes de (Y_t) separados entre sí por $k = S, 2S, 3S, \dots$ períodos básicos (o bien 1, 2, 3, ... años); ver figuras 2.27-2.28.

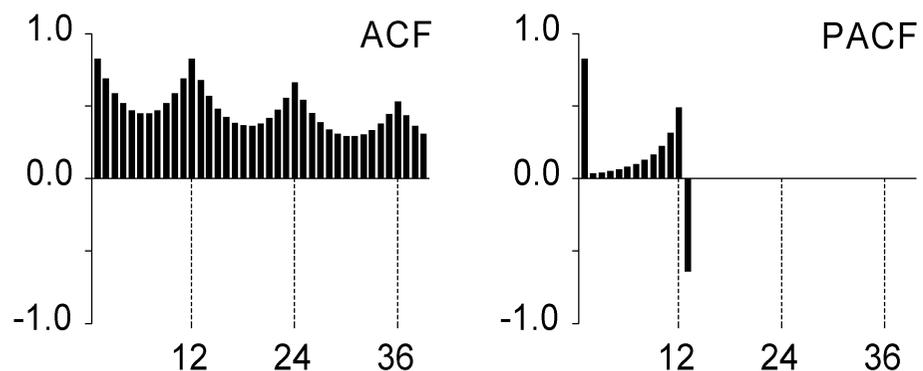


Figura 2.27

$AR(1) \times AR(1)_{12}$, con $\phi_1 = 0.8$, $\Phi_1 = 0.8$.

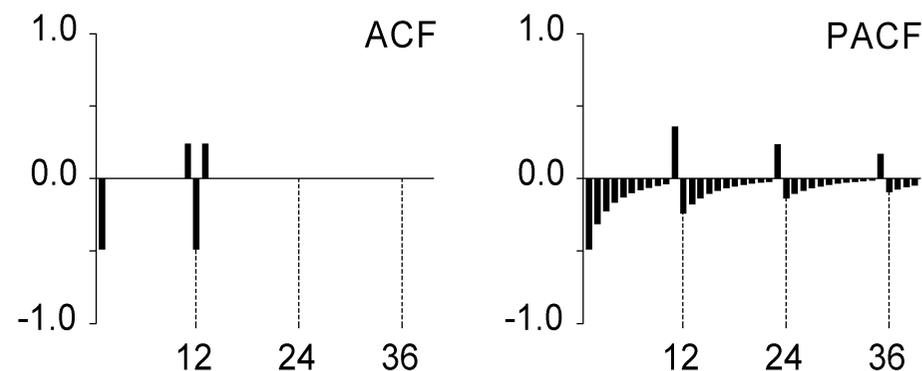


Figura 2.28

$MA(1) \times MA(1)_{12}$, con $\theta_1 = 0.8$, $\Theta_1 = 0.8$.

Observación 3: En general, la ACF teórica de un modelo $ARMA(p, q) \times ARMA(P, Q)_S$ presenta las pautas siguientes: (i) en los primeros retardos (1, 2, 3, ...) aparece la ACF de la parte regular $ARMA(p, q)$, (ii) en los retardos anuales ($S, 2S, 3S, \dots$) aparece la ACF de la parte anual $ARMA(P, Q)_S$, (iii) la pauta de la ACF de la parte regular se repite a la izquierda (en el sentido \leftarrow) y a la derecha (en el sentido \rightarrow) de cada coeficiente anual, con signo igual al producto de los signos del coeficiente anual y de los coeficientes regulares correspondientes. En general, la PACF teórica de un modelo $ARMA(p, q) \times ARMA(P, Q)_S$ presenta las pautas siguientes: (i) en los primeros retardos (1, 2, 3, ...) aparece la PACF de la parte regular $ARMA(p, q)$, (ii) en los retardos anuales ($S, 2S, 3S, \dots$) aparece la PACF de la parte anual $ARMA(P, Q)_S$, (iii) a la derecha de cada coeficiente anual se repite (en el sentido \rightarrow) la pauta de la PACF de la parte regular, con el signo cambiado si el coeficiente anual es positivo, o con su signo si dicho coeficiente es negativo, (iv) a la izquierda de cada coeficiente anual se repite (en el sentido \leftarrow) la pauta de la ACF de la parte regular, con signo igual al producto de los signos del coeficiente anual y de los coeficientes regulares correspondientes.

2.4 PROCESOS ESTOCÁSTICOS NO ESTACIONARIOS

- ⇒ Muchas series temporales (como las del Ejemplo 2.2.13) no pueden considerarse generadas por procesos estocásticos estacionarios (es decir, son series **no estacionarias**), porque presentan ciertas **tendencias** claras en su evolución temporal (de manera que no presentan afinidad hacia algún valor constante en el tiempo), porque su dispersión no es constante, porque son **estacionales**, o por varias combinaciones de estos motivos.
- ⇒ No obstante, muchas series temporales no estacionarias se pueden **transformar** de forma adecuada para obtener series de aspecto estacionario (como las de los ejemplos 2.2.11 y 2.2.12), que pueden ser utilizadas como punto de partida para elaborar en la práctica modelos $ARMA(p,q)$ como los descritos en la Sección 2.3 anterior.
- ⇒ Para convertir una serie no estacionaria en otra estacionaria, suelen emplearse en la práctica dos tipos de transformaciones: un primer tipo para estabilizar su dispersión (es decir, para inducir **estacionariedad en varianza**) y un segundo tipo para estabilizar su nivel (es decir, para eliminar su tendencia o para inducir **estacionariedad en media**).

No estacionariedad en varianza – Transformación de Box-Cox

2.4.1 Definición

Sea (Y_t) un proceso estocástico **no** estacionario tal que

$$\mu_t \equiv E[Y_t] \quad \text{y} \quad \sigma_t^2 \equiv \text{Var}[Y_t]$$

existen y **dependen** de t (por lo que **no** son constantes; ver 2.2.1-2.2.4). Si

$$\sigma_t^2 = \sigma^2 \times h(\mu_t)^2, \quad (2.22)$$

donde $\sigma^2 > 0$ es una constante y $h(\cdot)$ es una función real tal que $h(\cdot) \neq 0$ para cualquier valor de μ_t , entonces una **transformación estabilizadora de la varianza** de (Y_t) es cualquier función real $g(\cdot)$ tal que

$$g'(\cdot) = \frac{1}{h(\cdot)} \quad \text{para cualquier valor de } \mu_t. \quad (2.23)$$

Observación: Una aproximación de Taylor de primer orden (lineal) a $Y_t' \equiv g(Y_t)$ alrededor de μ_t indica que $Y_t' \approx g(\mu_t) + g'(\mu_t)(Y_t - \mu_t)$; por lo tanto, $\text{Var}[Y_t'] \approx g'(\mu_t)^2 \sigma_t^2 = \sigma^2 g'(\mu_t)^2 h(\mu_t)^2$, que es igual σ^2 (constante) cuando $g(\cdot)$ satisface (2.23). En consecuencia, el propósito de una transformación que satisface (2.23) es estabilizar la varianza de un proceso no estacionario (Y_t) cuya **dispersión local** depende de su **nivel local** de acuerdo con (2.22).

2.4.2 Ejemplo

Sea (Y_t) un proceso estocástico no estacionario tal que $\mu_t \equiv E[Y_t]$ y $\sigma_t^2 \equiv \text{Var}[Y_t]$ no son constantes. Si σ_t es proporcional a μ_t , de manera que $\sigma_t^2 = \sigma^2 \mu_t^2$, entonces $h(\mu_t) = \mu_t$ en (2.22) y $g'(\mu_t) = \mu_t^{-1}$ en (2.23), es decir, $g(\cdot) = \ln(\cdot)$; en este caso, la varianza del proceso (Y_t') es (aproximadamente) constante. Si σ_t^2 es proporcional a μ_t , de manera que $\sigma_t^2 = \sigma^2 \mu_t$, entonces $h(\mu_t) = \mu_t^{1/2}$ en (2.22) y $g'(\mu_t) = \mu_t^{-1/2}$ en (2.23), es decir, $g(\cdot) = 2(\cdot)^{1/2}$; en este caso, la varianza de $(Y_t') = (2Y_t^{1/2})$ es (aproximadamente) constante.

Observación 1: Las dos transformaciones del Ejemplo 2.4.2 pueden escribirse como

$$Y_t' \equiv g(Y_t) = \frac{(Y_t + m)^\lambda - 1}{\lambda},$$

que se denomina **transformación de Box-Cox** de parámetros λ y m [donde $|\lambda| \leq 2$ y el -1 del numerador se emplea tan sólo porque $\lim_{\lambda \rightarrow 0} \{[(Y_t + m)^\lambda - 1] / \lambda\} = \ln(Y_t + m)$]. En particular, $\lambda = 0$ y $m = 0$ proporcionan la transformación $Y_t' = \ln Y_t$, mientras que $\lambda = 1/2$ y $m = 0$ proporcionan la transformación $Y_t' = 2Y_t^{1/2}$.

Observación 2: El hecho de multiplicar un proceso estocástico por una constante no afecta a que su varianza sea o no sea constante; por lo tanto, la presencia de λ en el denominador de la transformación de Box-Cox anterior puede ignorarse en general (lo mismo que el -1 del numerador). La Figura 2.29 representa las dos posibilidades del Ejemplo 2.4.2, junto con otras también contempladas (aunque menos frecuentes) en la transformación de Box-Cox.

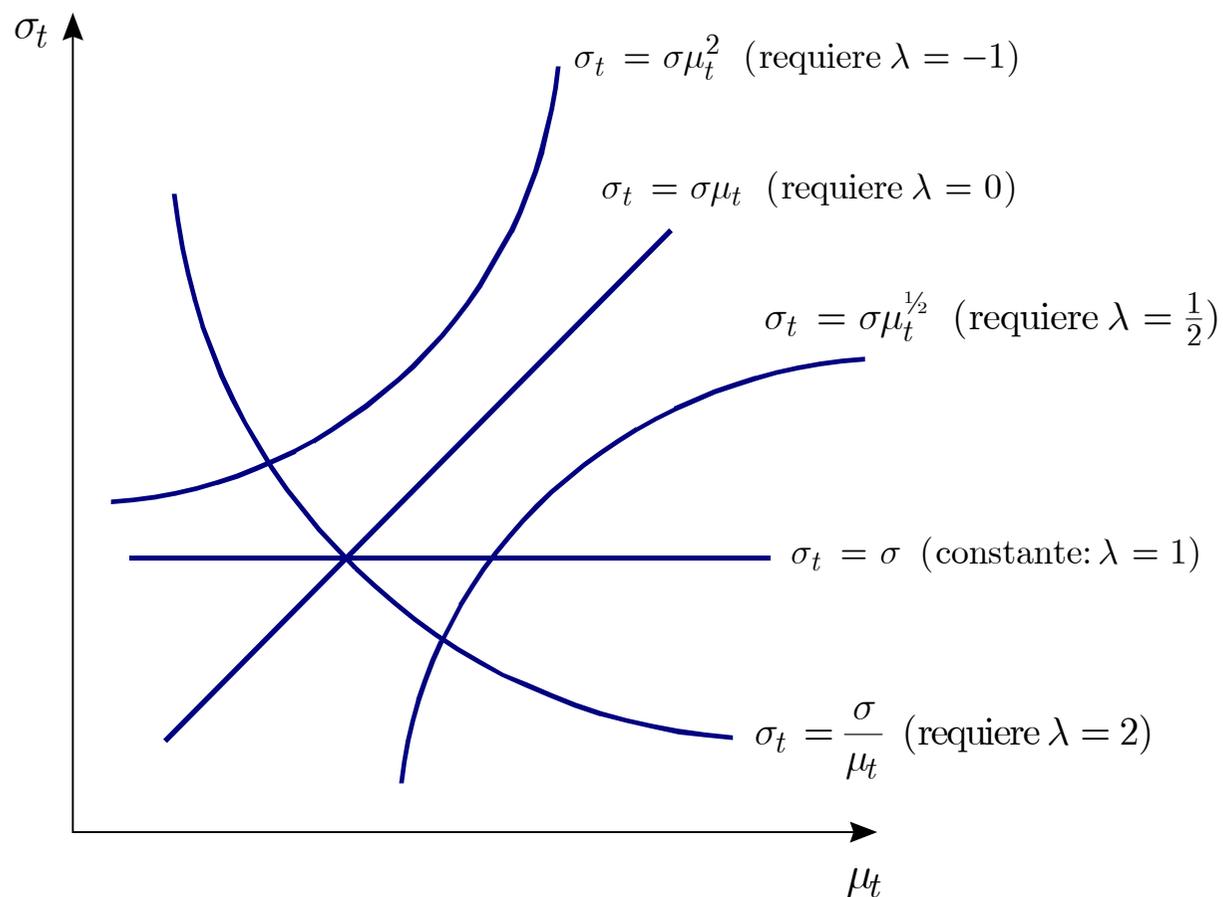


Figura 2.29

Posibles relaciones entre la desviación típica y la media locales de procesos no estacionarios (entre paréntesis figura la transformación de Box-Cox requerida para estabilizar la varianza en cada caso)

2.4.3 Ejemplo

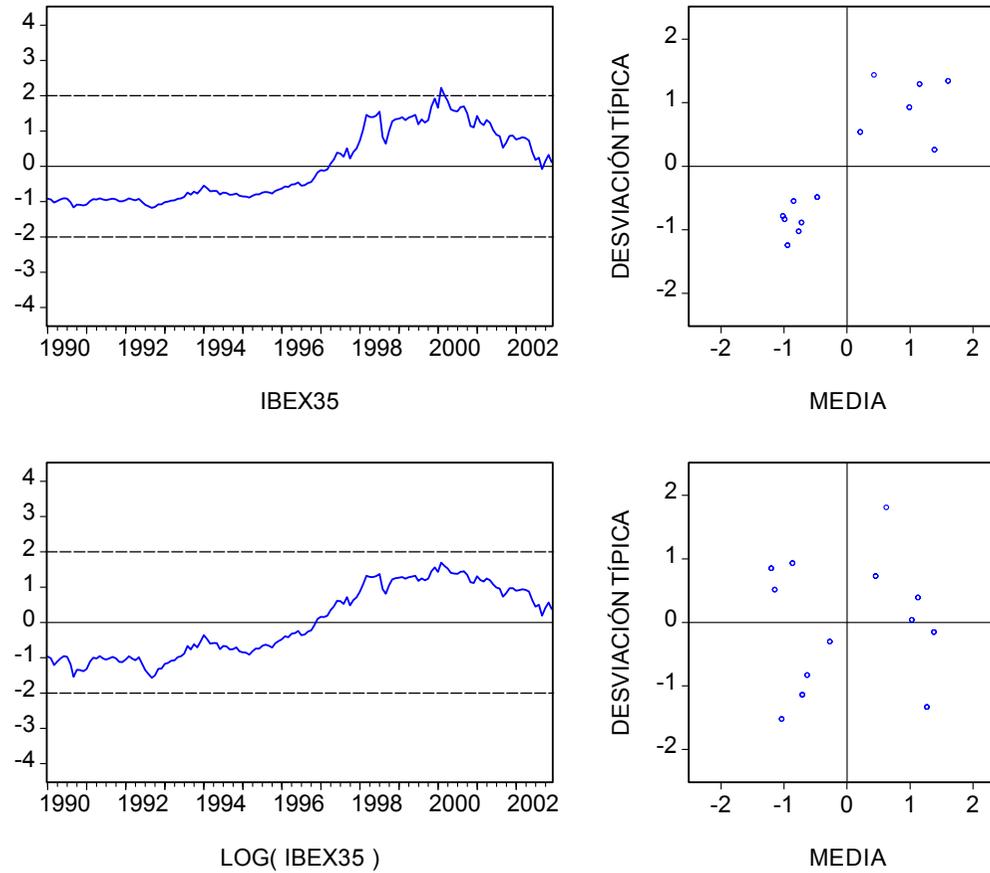


Figura 2.30 (I)

Gráficos desviación típica-media muestrales estandarizados
(Datos: bolsa-int.wf1. Programa: DT-Media.prg)

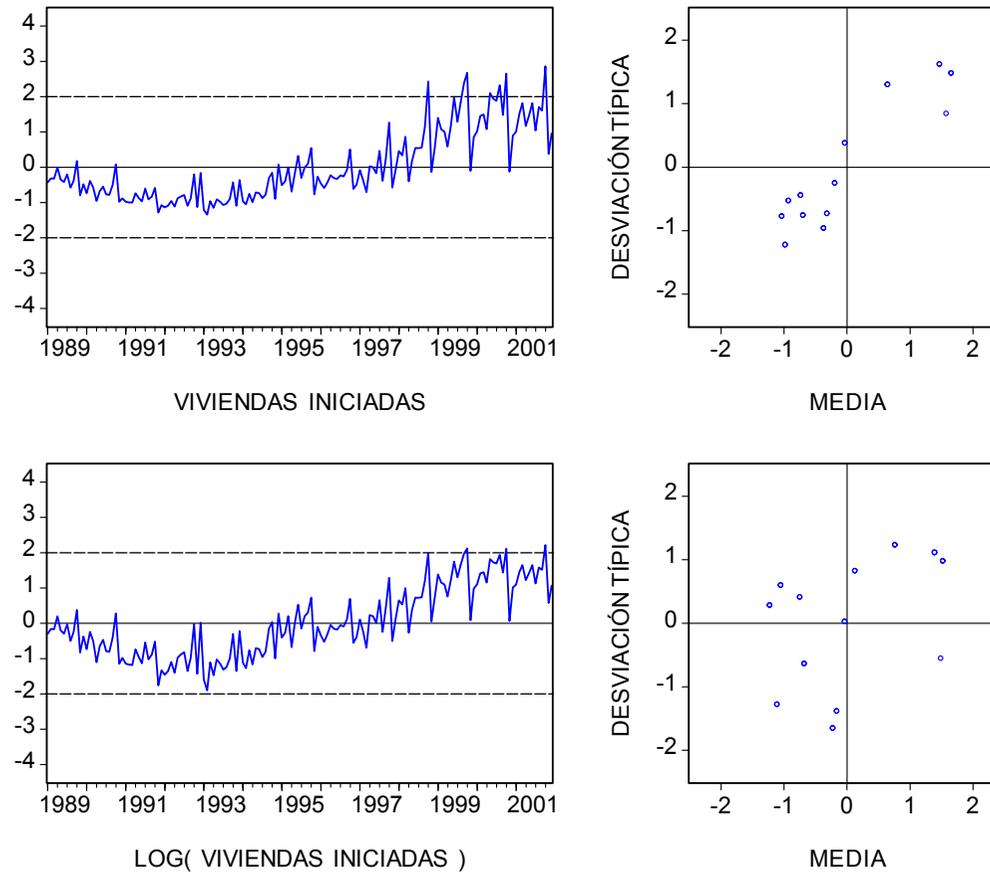


Figura 2.30 (II)

Gráficos desviación típica-media muestrales estandarizados
(Datos: seasonal.wf1. Programa: DT-Media.prg)

No estacionariedad en media – Tendencias

2.4.4 Definición

El operador **diferencia regular de orden d** ($d \geq 1$ entero) se representa con el símbolo ∇^d (a veces Δ^d) y se define como $\nabla^d \equiv (1 - B)^d$ (donde B es el operador de retardo de 2.3.2), de manera que

$$\nabla X_t \equiv X_t - X_{t-1}, \quad \nabla^2 X_t \equiv X_t - 2X_{t-1} + X_{t-2}, \dots, \quad (2.24)$$

donde X_t es una variable (real o aleatoria) referida a un momento t determinado.

2.4.5 Definición

Un proceso estocástico (Y_t) es **integrado de orden d** ($d \geq 0$ entero), o **I(d)**, si y sólo si el proceso $(\nabla^d Y_t)$ sigue un modelo ARMA(p, q) **estacionario** e **invertible** (ver 2.3.1). En tal caso, suele escribirse $(Y_t) \sim I(d)$.

Observación: En particular: (i) un proceso estocástico (Y_t) es integrado de orden 0, o I(0), si y sólo si (Y_t) sigue un modelo ARMA(p, q) estacionario e invertible (para lo cual es necesario que (Y_t) sea un proceso estacionario; ver 2.2.1-2.2.4), (ii) $(Y_t) \sim I(1)$ si y sólo si $(\nabla Y_t) \sim I(0)$, (iii) $(Y_t) \sim I(2)$ si y sólo si $(\nabla^2 Y_t) \sim I(0)$, ... Como se ilustra en los ejemplos 2.4.6-2.4.8, si $(Y_t) \sim I(d)$ con $d \geq 1$, entonces (Y_t) es un proceso **no** estacionario.

2.4.6 Ejemplo

Un **paseo aleatorio** (en inglés *random walk*) es un proceso estocástico (Y_t) tal que

$$Y_t = \mu + Y_{t-1} + A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.25)$$

donde μ es un parámetro (que puede valer 0) y $(A_t) \sim \text{IID}(0, \sigma_A^2)$. (2.25) implica que

$$\nabla Y_t \equiv Y_t - Y_{t-1} = \mu + A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.26)$$

de manera que (∇Y_t) sigue un modelo ARMA(0,0) (con término constante si $\mu \neq 0$). Por lo tanto, $(\nabla Y_t) \sim \text{I}(0)$, lo que implica que un paseo aleatorio es un proceso $\text{I}(1)$.

Por otro lado, (2.25) es una **ecuación en diferencias**, cuya **solución general** a partir de un **origen temporal** $k < t$ dado es

$$Y_t = \left[(y_k - \mu k) + A_t + \sum_{i=1}^{t-k-1} A_{t-i} \right] + \mu t \quad \text{para } t > k. \quad (2.27)$$

Si, por ejemplo, $k = 0$ e $y_0 = 0$, (2.27) implica que $E[Y_t] = \mu t$ y $\text{Var}[Y_t] = \sigma_A^2 t$ dependen de t , por lo que (Y_t) es un proceso no estacionario (incluso cuando $\mu = 0$).

Observación 1: A partir de (2.27) con $k = 0$ e $y_0 = 0$, puede comprobarse que $\text{Corr}[Y_t, Y_{t+h}] = [t/(t+h)]^{1/2}$ para todo $h > 0$, lo que indica nuevamente que un paseo aleatorio es un proceso no estacionario. Además, para cualquier t dado, la correlación anterior tiende a cero con h muy lentamente; de hecho, cuanto mayor es t , más lentamente tiende $\text{Corr}[Y_t, Y_{t+h}]$ a cero (por ejemplo, con $t = 500$, $\text{Corr}[Y_t, Y_{t+h}] > 0.95$ incluso para $h = 50$). Análogamente, puede comprobarse que $\text{Corr}[Y_{t-h}, Y_t] = [(t-h)/t]^{1/2}$ para todo $h > 0$, que no coincide en general con $\text{Corr}[Y_t, Y_{t+h}]$. En cualquier caso, el decrecimiento lento de la ACF de un paseo aleatorio es característico de cualquier proceso integrado de orden $d \geq 1$, lo que contrasta claramente con el decrecimiento exponencial de la ACF de un proceso estacionario $I(0)$ que sigue cualquiera de los modelos ARMA estacionarios e invertibles de la Sección 2.3.

Observación 2: Un proceso integrado de orden $d \geq 1$ también se denomina un proceso **homogéneamente no estacionario** de orden $d \geq 1$, en el sentido que cuando $(Y_t) \sim I(d)$, (Y_t) presenta una **tendencia** en su evolución temporal, pero sus fluctuaciones alrededor de dicha tendencia constituyen un proceso estocástico estacionario con media cero, por lo que dichas fluctuaciones son homogéneas. En otros términos, cuando $(Y_t) \sim I(d)$, su evolución en cada intervalo temporal de su historia (**componente irregular** o **transitorio**) es esencialmente independiente de su tendencia (**componente tendencial** o **permanente**). (El tipo de no estacionariedad que se caracteriza por una clara dependencia de la evolución temporal local de un proceso con respecto a su tendencia, denominado **no estacionariedad explosiva**, no se considera aquí por no ser representativo de la mayoría de las series temporales económicas).

Observación 3: De acuerdo con lo anterior, cuando $(Y_t) \sim I(d)$, Y_t debe admitir una representación del tipo $Y_t = T_t[Y_t] + I_t[Y_t]$, donde $T_t[Y_t]$ es la tendencia de (Y_t) , e $I_t[Y_t]$ es la parte irregular de (Y_t) (que debe ser un proceso estacionario con media cero). La tendencia de un proceso estocástico (Y_t) puede definirse como el nivel al que tiende Y_t desde un **origen temporal** suficientemente anterior al momento t . Por lo tanto, si $T_t[Y_t]$ no es constante, entonces (Y_t) es necesariamente un proceso no estacionario (porque **no** converge a largo plazo a algún nivel de equilibrio).

2.4.7 Ejemplo

Cuando (Y_t) es un paseo aleatorio descrito por (2.25), la forma funcional de (2.27) no varía aunque el origen temporal $k < t$ tienda a $-\infty$. En este caso, puede considerarse que

$$T_t[Y_t] = \left[(y_k - \mu k) + \sum_{i=1}^{t-k-1} A_{t-i} \right] + \mu t, \quad I_t[Y_t] = A_t, \quad (2.28)$$

o bien

$$T_t[Y_t] = \mu + Y_{t-1}, \quad I_t[Y_t] = A_t. \quad (2.29)$$

$T_t[Y_t]$ en (2.28) es la ecuación de una recta de nivel $\beta_0 \equiv \left[(y_k - \mu k) + \sum_{i=1}^{t-k-1} A_{t-i} \right]$ (una variable aleatoria) y pendiente μ (una constante). Por lo tanto, la tendencia de un paseo aleatorio tiene dos componentes: un **componente determinista** asociado con el término μt (que no está presente si $\mu = 0$) y un **componente estocástico** o **aleatorio** asociado con el término β_0 , cada uno de cuyos sumandos $A_{t-1}, A_{t-2}, \dots, A_{k+1}$ representa una contribución permanente aunque aleatoria al nivel de (Y_t) .

Por otro lado, (2.26) implica que $T_t[\nabla Y_t] = \mu$ (constante), $I_t[\nabla Y_t] = A_t$, de manera que (∇Y_t) es en este caso un proceso estocástico sin tendencia.

2.4.8 Ejemplo

Sea (Y_t) un proceso estocástico tal que

$$Y_t = Y_{t-1} + A_t - \theta_1 A_{t-1} \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.30)$$

donde θ_1 es un parámetro tal que $|\theta_1| < 1$ y $(A_t) \sim \text{IID}(0, \sigma_A^2)$. (2.30) implica que

$$\nabla Y_t \equiv Y_t - Y_{t-1} = A_t - \theta_1 A_{t-1} \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.31)$$

de manera que (∇Y_t) sigue un modelo MA(1) invertible. Por lo tanto, $(\nabla Y_t) \sim \text{I}(0)$, lo que implica que un proceso estocástico descrito por (2.30) es un proceso I(1).

Por otro lado, la solución general de (2.30) a partir de un origen temporal $k < t$ dado es

$$Y_t = (y_k - \theta_1 a_k) + A_t + \sum_{i=1}^{t-k-1} (1 - \theta_1) A_{t-i} \quad \text{para } t > k. \quad (2.32)$$

Si $k = 0$ e $y_0 = a_0 = 0$, (2.32) implica que $\text{Var}[Y_t] = \sigma_A^2 + (1 - \theta_1)^2 \sigma_A^2 (t - 1)$ depende de t , por lo que (Y_t) es un proceso no estacionario.

Dado que la forma funcional de la solución (2.32) no varía aunque el origen temporal $k < t$

tienda a $-\infty$, a partir de (2.30) y (2.32) puede considerarse en este caso que

$$T_t[Y_t] = (y_k - \theta_1 a_k) + A_{t-1} + \sum_{i=2}^{t-k-1} (1 - \theta_1) A_{t-i}, \quad I_t[Y_t] = A_t - \theta_1 A_{t-1}, \quad (2.33)$$

o bien

$$T_t[Y_t] = Y_{t-1}, \quad I_t[Y_t] = A_t - \theta_1 A_{t-1}. \quad (2.34)$$

$T_t[Y_t]$ en (2.33) es puramente estocástica, cada uno de cuyos sumandos $A_{t-1}, (1 - \theta_1)A_{t-2}, \dots, (1 - \theta_1)A_{k+1}$ representa una contribución permanente aunque aleatoria al nivel de (Y_t) .

Por otro lado, (2.32) implica que $T_t[\nabla Y_t] = 0$, $I_t[\nabla Y_t] = A_t - \theta_1 A_{t-1}$, de manera que (∇Y_t) es en este caso un proceso estocástico sin tendencia.

2.4.9 Ejemplo

Sea $(Y_t) \sim I(0)$ un proceso estocástico que sigue un modelo AR(1) estacionario del tipo

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.35)$$

donde μ y ϕ_1 son parámetros, $|\phi_1| < 1$, y $(A_t) \sim \text{IID}(0, \sigma_A^2)$. La solución de (2.35) a partir

de un origen temporal $k < t$ dado es:

$$Y_t = \phi_1^{t-k} y_k + \sum_{i=0}^{t-k-1} \phi_1^i A_{t-i} + \left(\sum_{i=0}^{t-k-1} \phi_1^i \right) \mu \quad \text{para } t > k, \quad (2.36)$$

que tiende a

$$Y_t = \frac{\mu}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i A_{t-i} \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots \quad (2.37)$$

cuando $k \rightarrow -\infty$. Por lo tanto, en este caso puede considerarse que

$$T_t[Y_t] = \frac{\mu}{1 - \phi_1}, \quad I_t[Y_t] = \sum_{i=0}^{\infty} \phi_1^i A_{t-i} = \phi_1 I_{t-1}[Y_t] + A_t, \quad (2.38)$$

de manera que (Y_t) es un proceso estocástico sin tendencia.

Observación 1: Los ejemplos 2.4.6-2.4.9 sugieren las siguientes conclusiones generales en relación con un proceso estocástico $(Y_t) \sim I(d)$ ($d \geq 1$) (que se denomina, a veces, en inglés *difference-stationary*):

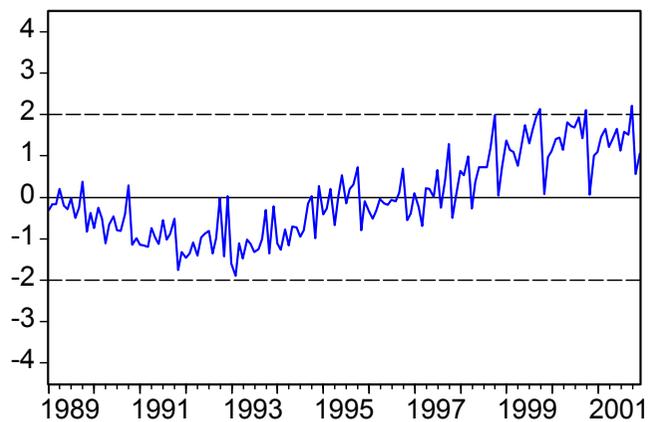
- ⇒ (Y_t) es un proceso **no** estacionario que presenta en su evolución temporal una **tendencia estocástica**, acompañada a veces por alguna tendencia determinista.
- ⇒ $(W_t) \equiv (\nabla^d Y_t)$ es un proceso $I(0)$ que, por lo tanto, no presenta tendencia alguna en su evolución temporal.
- ⇒ En general, los componentes de $(Y_t) \sim I(d)$ en $Y_t = T_t[Y_t] + I_t[Y_t]$ no son únicos. Una forma de identificarlos consiste en imponer que el componente irregular o transitorio de (Y_t) coincida con el de $(W_t) \equiv (\nabla^d Y_t)$.

Observación 2: A veces, la tendencia de un proceso (Y_t) es puramente determinista, como ocurre por ejemplo cuando $Y_t = \beta_0 + \beta_1 t + A_t$ con $\beta_1 \neq 0$ y $(A_t) \sim \text{IID}(0, \sigma_A^2)$; en este caso, $\nabla Y_t = \beta_1 + \nabla A_t$ es un proceso estacionario que sigue un modelo MA(1) **no invertible**, por lo que (Y_t) **no** es un proceso I(1) (un proceso de este tipo se denomina en inglés *trend-stationary*). Por otro lado, si $(Y_t) \sim \text{I}(0)$, como ocurre por ejemplo si $Y_t = \beta_0 + A_t$, (Y_t) no requiere ninguna diferencia regular para transformarlo en un proceso I(0), lo que se observa, por ejemplo, en que $\nabla Y_t = \nabla A_t$ también sigue un modelo MA(1) no invertible. En general, la no invertibilidad indica un uso incorrecto por exceso del operador ∇ (**sobrediferenciación**), lo que puede llevar a caracterizar erróneamente la tendencia de un proceso.

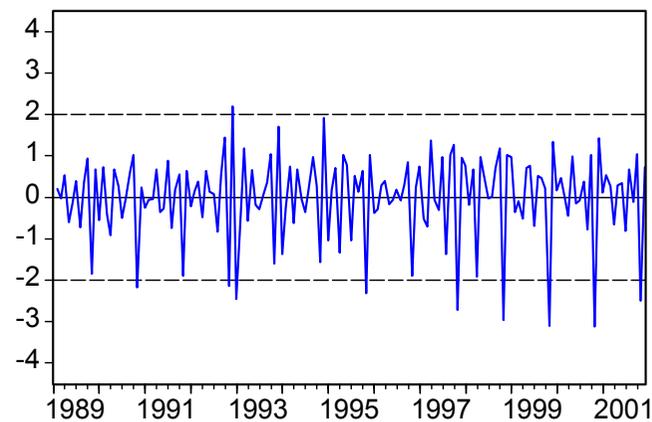
Observación 3: En la práctica, muchas series temporales no estacionarias que presentan tendencias claras en su evolución temporal, se pueden transformar en series estacionarias aplicándoles tan sólo 1 ó 2 diferencias regulares, lo que significa que dichas series han podido ser generadas por procesos integrados de orden 1 ó 2, respectivamente. Por otro lado, la estimación de los componentes $T_t[Y_t]$ e $I_t[Y_t]$ de un proceso no estacionario a partir de una serie temporal, requiere la elaboración previa de un **modelo ARIMA** para dicha serie como se describe en la Sección 2.6.

Observación 4: La **estacionalidad** es una forma de no estacionariedad homogénea que ocurre con mucha frecuencia en la práctica. Por ejemplo, una serie que se observa $S > 1$ veces al año es estacional de **período** S cuando la serie presenta ciertas pautas en su evolución que se repiten sistemáticamente cada S observaciones [ver figuras 1.9, 1.10 y 2.10 (Panel D)]. Para recoger la estacionalidad de un proceso (Y_t) del que supuestamente procede una serie estacional, suele emplearse el operador **diferencia estacional** de **período** S y **orden** $D \geq 1$, $\nabla_S^D \equiv (1 - B^S)^D$, de manera que si (Y_t) es un proceso estacional de período S , entonces $(\nabla_S^D Y_t)$ no presenta estacionalidad (aunque puede ser no estacionario por otros motivos). (Nótese que $\nabla_S \equiv 1 - B^S \neq (1 - B)^S \equiv \nabla^S$ para cualquier valor de $S > 1$.)

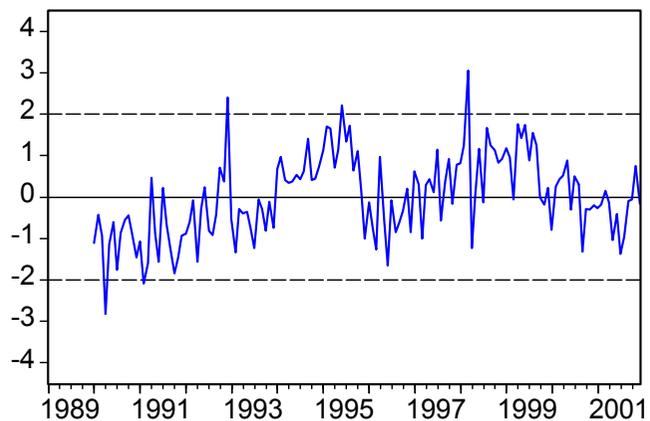
En resumen, si una serie temporal (y_t) es no estacionaria por diversos motivos, $(\nabla^d \nabla_S^D y_t')$ es una serie estacionaria si la transformación de Box-Cox (ver 2.4.1-2.4.3), el número d de diferencias regulares (en muchos casos 0, 1 ó 2), y el número D de diferencias estacionales (en muchos casos 0 ó 1), se escogen adecuadamente; ver Figura 2.31.



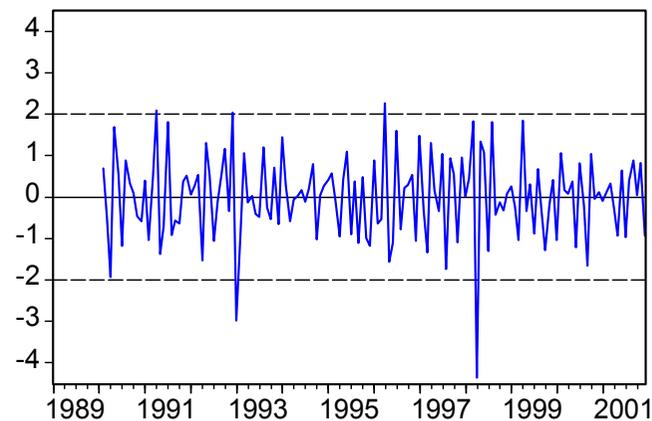
PANEL A - LOG(VIVIN)



PANEL B - DLOG(VIVIN, 1, 0)



PANEL C - DLOG(VIVIN, 0, 12)



PANEL D - DLOG(VIVIN, 1, 12)

Figura 2.31

Diferentes transformaciones de la serie mensual Viviendas iniciadas en España (seasonal.wf1)

2.5 MODELOS ARIMA

2.5.1 Definición

Un proceso estocástico (Y_t) es integrado de orden d ($d \geq 0$ entero) si y sólo si (Y_t) sigue un **modelo autorregresivo-integrado-media móvil** de orden (p, d, q) , o $\text{ARIMA}(p, d, q)$ (del inglés *AutoRegressive-Integrated-Moving Average*), del tipo

$$\phi(B)\nabla^d Y_t = \mu + \theta(B)A_t \quad \text{para todo } t = 0, \pm 1, \pm 2, \dots, \quad (2.39)$$

donde las raíces de las ecuaciones $\phi(x) = 0$ y $\theta(x) = 0$ están fuera del círculo unitario.

Observación 1: (2.39) puede interpretarse como un modelo $\text{ARMA}(p, q)$ estacionario para la diferencia regular de orden d de (Y_t) , en el sentido que (2.39) es equivalente a $\phi(B)W_t = \mu + \theta(B)A_t$ con $(W_t) \equiv (\nabla^d Y_t) \sim I(0)$; ver también (2.10)-(2.14). Por otro lado, (2.39) también puede interpretarse como un modelo $\text{ARMA}(p+d, q)$, **no** estacionario para (Y_t) del tipo $\phi'(B)Y_t = \mu + \theta(B)A_t$, cuyo operador autorregresivo $\phi'(B) \equiv \phi(B)\nabla^d \equiv 1 - \sum_{i=1}^{p'} \phi'_i B^i$ ($p' \equiv p + d$) es tal que la ecuación $\phi'(x) = 0$ tiene exactamente d **raíces unitarias**.

Observación 2: Como se ilustra con dos ejemplos en la Observación 2 del Ejemplo 2.4.9, la invertibilidad de (2.39) garantiza que el modelo ARIMA que sigue $(Y_t) \sim I(d)$ describe adecuadamente el tipo de tendencia que presenta (Y_t) en su evolución temporal (puramente estocástica, o bien con un componente estocástico y otro determinista); para más detalles, ver los dos apartados que siguen al Ejemplo 2.5.2.

2.5.2 Ejemplo

El paseo aleatorio de los ejemplos 2.4.6 y 2.4.7 es un modelo ARIMA(0,1,0) para un proceso I(1). El modelo del Ejemplo 2.4.8 es un modelo ARIMA(0,1,1), o IMA(1,1), para un proceso I(1). Un modelo ARIMA(1,1,0), o ARI(1,1), para un proceso I(1) es

$$(1 - \phi_1 B)\nabla Y_t = \mu + A_t, \text{ o bien } Y_t = \mu + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + A_t,$$

donde $|\phi_1| < 1$, mientras que un modelo ARIMA(1,1,1) para un proceso I(1) es

$$(1 - \phi_1 B)\nabla Y_t = \mu + (1 - \theta_1 B)A_t, \text{ o bien } Y_t = \mu + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + A_t - \theta_1 A_{t-1},$$

donde $|\phi_1| < 1$, $|\theta_1| < 1$ y $\phi_1 \neq \theta_1$. Por su parte, todos los modelos de los ejemplos 2.3.4-2.3.8 y 2.4.9 son modelos ARIMA($p,0,q$), o ARMA(p,q), para procesos I(0).

Tendencias en modelos invertibles

⇨ El modelo ARIMA(p,d,q) para $(Y_t) \sim I(d)$ de la Definición 2.5.1 puede escribirse como

$$\phi(B)[\nabla^d Y_t - \mu_W] = \theta(B)A_t, \quad (2.40)$$

donde $\mu_W \equiv E[\nabla^d Y_t] = \mu / \phi(1)$, o bien como

$$\nabla^d Y_t = \mu_W + \frac{\theta(B)}{\phi(B)} A_t = \mu_W + \psi(B) A_t, \quad (2.41)$$

con $\psi(B) \equiv \theta(B) / \phi(B)$; ver la explicación alrededor de (2.15). Por lo tanto,

$$Y_t = \mu_t + \psi(B) A_t^*, \quad \text{o bien} \quad Y_t = \mu_t + \psi^*(B) A_t, \quad (2.42)$$

donde μ_t , A_t^* y $\psi^*(B) \equiv \sum_{i=0}^{\infty} \psi_i^* B^i$ ($\psi_0^* \equiv 1$) se definen de acuerdo con que

$$\nabla^d \mu_t = \mu_W, \quad \nabla^d A_t^* = A_t, \quad \nabla^d \psi^*(B) = \psi(B). \quad (2.43)$$

⇒ En consecuencia, un modelo ARIMA(p, d, q) para un proceso $(Y_t) \sim I(d)$ implica que

$$Y_t = \underbrace{\mu_t + \psi(B)(A_t^* - A_t)}_{T_t[Y_t]} + \underbrace{\psi(B)A_t}_{I_t[Y_t]} = \underbrace{\mu_t + [\psi^*(B) - \psi(B)]A_t}_{T_t[Y_t]} + \underbrace{\psi(B)A_t}_{I_t[Y_t]}, \quad (2.44)$$

donde la tendencia $T_t[Y_t]$ es la suma de un **componente determinista** y un **componente estocástico**, y la parte irregular $I_t[Y_t]$ es un proceso ARMA estacionario e invertible.

Observación: La expresión $\nabla^d \mu_t = \mu_W$ en (2.43) es una ecuación en diferencias cuya solución general es del tipo $\mu_t = g(t) + [\mu_W / d!] t^d$, donde $g(t)$ es un polinomio de grado igual o inferior a $d - 1$, incluyendo la posibilidad de

que $g(t) = 0$. Por lo tanto, μ_t en (2.44), μ_W en (2.40), ó μ en (2.39), representan un componente tendencial determinista de grado d en (Y_t) , con respecto a cuyos coeficientes un modelo ARIMA tan sólo especifica el número d de diferencias regulares y (explícita o implícitamente) el valor del parámetro μ_W (que puede ser 0, en cuyo caso la tendencia determinista de (Y_t) sería a lo sumo de grado $d - 1$). En particular, puede comprobarse que (2.43)-(2.44) proporcionan exactamente los resultados de los ejemplos 2.4.7-2.4.9.

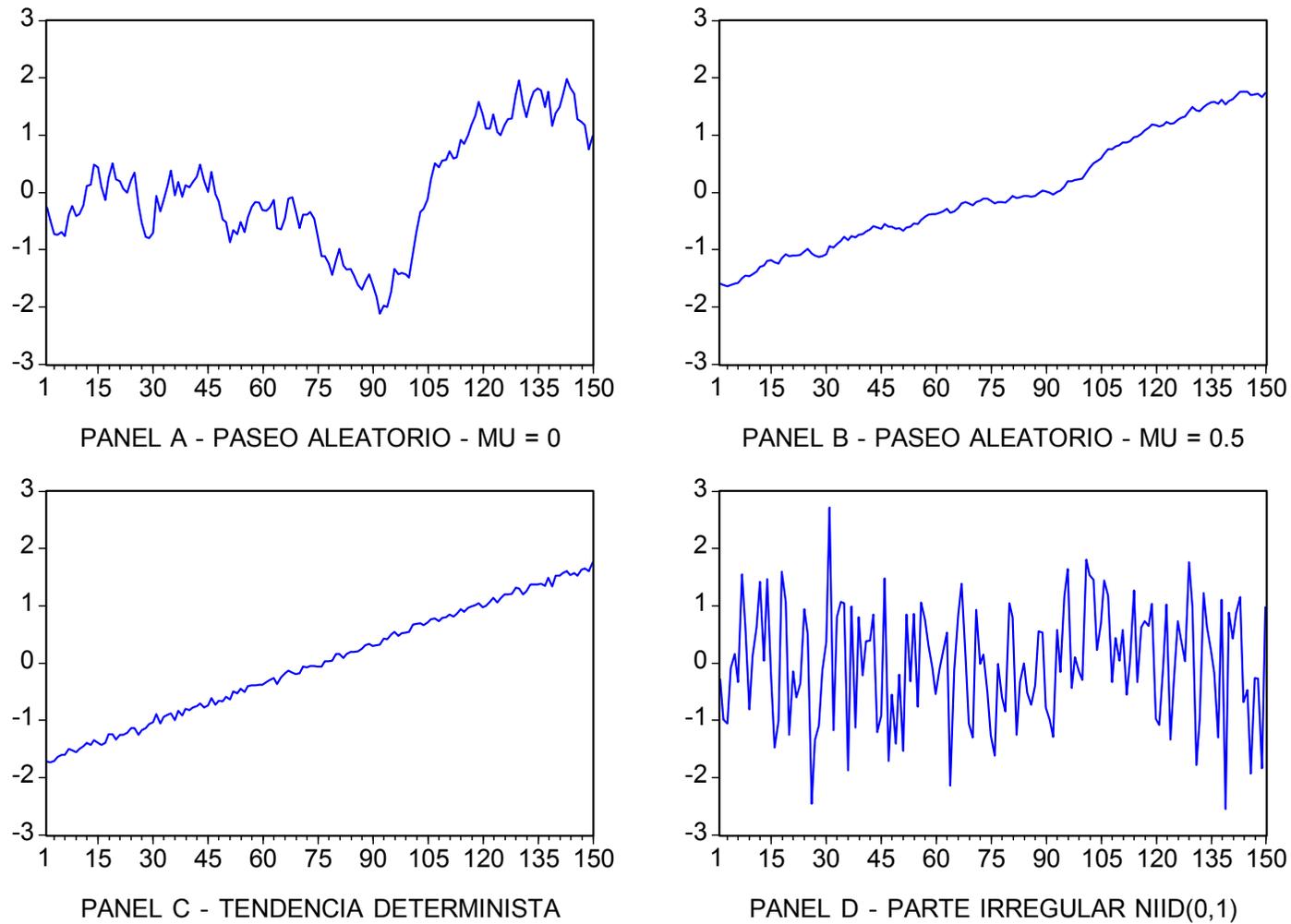
Tendencias en modelos no invertibles

⇒ Cuando el operador MA en (2.40) es tal que $\theta(x) = 0$ tiene exactamente d raíces unitarias, $\theta(B)$ puede factorizarse como $\theta(B) = \vartheta(B)\nabla^d$, donde $\vartheta(B)$ es un operador invertible. En este caso, (Y_t) **no** es un proceso $I(d)$ (ver 2.4.5), y (2.44) daría la falsa impresión de que (Y_t) tiene un componente tendencial estocástico.

⇒ Cuando $\theta(B) = \vartheta(B)\nabla^d$, (2.41)-(2.43) implican que

$$Y_t = \mu_t + \frac{\vartheta(B)}{\phi(B)} A_t,$$

por lo que o bien la tendencia de (Y_t) es puramente determinista (si $d \geq 1$ y $\mu \neq 0$), o bien $(Y_t) \sim I(0)$ y no tiene tendencia (si $d = 1$ y $\mu = 0$); ver Observación 2 en 2.4.9.

**Figura 2.32**

Series simuladas con tendencias diferentes y con la misma parte irregular (sim-series.wf1)

2.6 ELABORACIÓN DE MODELOS ARIMA

El modelo ARIMA más general que se obtiene al considerar conjuntamente todas las posibilidades descritas en las secciones 2.3, 2.4 y 2.5 anteriores, puede escribirse como

$$\phi_p(B)\Phi_P(B^S)[\nabla^d\nabla_S^D Y_t' - \mu_W] = \theta_q(B)\Theta_Q(B^S)A_t. \quad (2.45)$$

Identificación

Entrada: $(y_t)_{t=1}^N$ (serie temporal original); S (período estacional).

- 1.1 Elección de λ y m en la transformación de Box-Cox – Gráfico desviación típica-media. Se obtiene una serie (y_t') , con $y_t' = \ln(y_t + m)$ si $\lambda = 0$, $y_t' = (y_t + m)^\lambda$ si $\lambda \neq 0$.
- 1.2 Elección de d y D – Gráfico temporal (tendencias) y ACF muestral (amortiguamiento) de (y_t') . Se obtiene una serie $(w_t) \equiv (\nabla^d\nabla_S^D y_t')$, supuestamente estacionaria.
- 1.3 Inclusión de μ_W – Contraste de significación de $\mu_W \equiv E[W_t] \equiv E[\nabla^d\nabla_S^D Y_t']$ a partir de la media muestral de la serie (w_t) .
- 1.4 Elección de p , P , q y Q – ACF y PACF muestrales de (w_t) comparadas con ACF y

PACF teóricas de modelos $\text{ARMA}(p, q) \times \text{ARMA}(P, Q)_S$. El criterio central aquí es el reconocimiento de configuraciones:

- Un valor grande en la ACF o en la PACF muestrales en un retardo k dado, puede **no** ser relevante si se encuentra aislado y k es grande.
- Una configuración reconocible de valores pequeños en la ACF o en la PACF muestrales, **sí** puede ser relevante.

Salida: $(w_t)_{t=1}^n \equiv (\nabla^d \nabla_S^D y'_t)_{t=1}^n$ (serie supuestamente estacionaria, con $n \equiv N - d - DS$); órdenes p, P, q y Q – Modelo (2.45) identificado para (Y_t) .

Observaciones sobre la etapa de identificación: La **tasa convencional de variación (neta)** de una serie (y_t) es

$$tcv_t(k) \equiv \frac{y_t}{y_{t-k}} - 1 = \frac{y_t - y_{t-k}}{y_{t-k}} = \frac{(1 - B^k)y_t}{y_{t-k}},$$

mientras que la **tasa logarítmica de variación** de una serie (y_t) se define como

$$tlv_t(k) \equiv \ln[1 + tcv_t(k)] = \ln \frac{y_t}{y_{t-k}} = \ln y_t - \ln y_{t-k} = (1 - B^k) \ln y_t,$$

que es aproximadamente igual a $tcv_t(k)$ para valores de $tcv_t(k)$ pequeños. La tasa convencional es **multiplicativa**, en el sentido que

$$[1 + tcv_t(k)] = \frac{y_t}{y_{t-k}} = \frac{y_t}{y_{t-1}} \frac{y_{t-1}}{y_{t-2}} \dots \frac{y_{t-k+1}}{y_{t-k}} = \prod_{i=0}^{k-1} [1 + tcv_{t-i}(1)],$$

por lo que la tasa logarítmica es **aditiva**:

$$tlv_t(k) = \ln \prod_{i=0}^{k-1} [1 + tcv_{t-i}(1)] = \sum_{i=0}^{k-1} \ln tlv_{t-i}(1).$$

Dado que muchas series temporales requieren una transformación logarítmica y una diferencia regular para darles aspecto estacionario, las series resultantes pueden interpretarse como tasas de variación. Por ejemplo, si (y_t) es una serie mensual, entonces:

- $(\nabla \ln y_t)$ es la tasa logarítmica de variación mensual de (y_t) .
- $(\nabla_{12} \ln y_t)$ es la tasa logarítmica de variación interanual de (y_t) .
- $(\nabla^2 \ln y_t)$ es la variación (absoluta) mensual de $(\nabla \ln y_t)$.
- $(\nabla \nabla_{12} \ln y_t)$ es la variación (absoluta) mensual de $(\nabla_{12} \ln y_t)$, o bien la variación (absoluta) interanual de $(\nabla \ln y_t)$; análogamente, $(\nabla^2 \nabla_{12} \ln y_t)$ puede interpretarse de varias formas alternativas.

Estimación

Entrada: $(w_t)_{t=1}^n \equiv (\nabla^d \nabla_S^D y_t')_{t=1}^n$ (serie de aspecto estacionario); modelo identificado para el proceso supuestamente estacionario $(W_t) \equiv (\nabla^d \nabla_S^D Y_t')$; estimaciones iniciales de los parámetros del modelo; variables de control para el método de estimación escogido.

- 2.1 $MV \Leftrightarrow \text{Min}(\tilde{\mathbf{w}}' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{w}}) | \boldsymbol{\Sigma} |^{\frac{1}{n}}$, con $\boldsymbol{\Sigma} \equiv \sigma_A^{-2} \text{Var}[\mathbf{W}]$ (teórica, implicada por el modelo).
- 2.2 Problemas: evaluar la forma cuadrática $\tilde{\mathbf{w}}' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{w}}$ y el determinante $|\boldsymbol{\Sigma}|$; estimación no lineal (no existen soluciones analíticas).
- 2.3 Alternativa a MV: MC $\Leftrightarrow \text{Min} \tilde{\mathbf{w}}' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{w}}$ (aproximaciones: *retroprevisión*, EViews, ...).

Salida: $\hat{\mu}_W$, $\hat{\boldsymbol{\phi}} \equiv [\hat{\phi}_1, \dots, \hat{\phi}_p]'$, $\hat{\boldsymbol{\Phi}} \equiv [\hat{\Phi}_1, \dots, \hat{\Phi}_P]'$, $\hat{\boldsymbol{\theta}} \equiv [\hat{\theta}_1, \dots, \hat{\theta}_q]'$, $\hat{\boldsymbol{\Theta}} \equiv [\hat{\Theta}_1, \dots, \hat{\Theta}_Q]'$, $\hat{\sigma}_A^2$ (modelo estimado); \mathbf{V} (matriz de covarianzas estimadas); serie $(\hat{a}_t)_{t=1}^n$ de **residuos**.

Diagnosis

Entrada: Modelo estimado; residuos.

- 3.1 Parámetros: significación (individual y conjunta); correlaciones estimadas entre estimadores; estacionariedad e invertibilidad.
- 3.2 Residuos (reemplazan a un proceso de ruido blanco en el modelo estimado): gráfico temporal; ACF y PACF residuales (Ljung-Box); media muestral; Normalidad.
- 3.3 Posible reformulación según los resultados de 3.1 y 3.2.

Salida: Si modelo reformulado \Rightarrow Volver a estimación. Si modelo adecuado \Rightarrow Utilizar para prever (otros fines: análisis de intervención, análisis multivariante, ...).

Observaciones sobre la etapa de diagnosis: Las operaciones siguientes referidas a los **parámetros** del modelo identificado y estimado, son especialmente útiles en la etapa de diagnosis:

- Contrastar la **significación individual** de cada parámetro, o la **significación conjunta** de varios parámetros, según interese. A veces, estos contrastes indican que pueden eliminarse algunos parámetros no significativos.
- Examinar la matriz **R** de **correlaciones** estimadas entre los estimadores de los parámetros, que puede calcularse a partir de **V** como $\mathbf{R} = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}$, donde **D** es una matriz diagonal cuyos elementos son las raíces cuadradas de los elementos en la diagonal principal de **V**. La presencia de correlaciones elevadas suele indicar que el modelo está **sobrep parametrizado**.
- Factorizar los operadores AR y MA estimados y comprobar que no tienen **factores comunes**. Si dichos operadores tienen factores aproximadamente comunes, en general deben cancelarse.
- Comprobar que los operadores AR y MA estimados satisfacen las condiciones de estacionariedad y de invertibilidad, respectivamente. La presencia de una raíz de $\hat{\phi}(x) = 0$ o de $\hat{\Phi}(x) = 0$ próxima o igual a 1 indica **subdiferenciación**, por lo que es conveniente aplicar una diferencia regular o estacional, respectivamente, a la serie (w_t) . Por su parte, la presencia de una raíz de $\hat{\theta}(x) = 0$ o de $\hat{\Theta}(x) = 0$ próxima o igual a 1 indica **sobrediferenciación**, por lo que es conveniente eliminar una diferencia regular o estacional, respectivamente, de la serie (w_t) (además, quizás, de reconsiderar la naturaleza de la tendencia implícita en el modelo reformulado).
- Cuando en la etapa de identificación no está claro si d (el número de diferencias regulares) debe ser igual a uno o igual a dos (también cuando se duda entre $d = 1$ y $d = 0$), una buena práctica consiste en escoger $d = 2$ y añadir

un factor MA(1) regular (**testigo de sobrediferenciación**) al modelo identificado con $d = 2$; si el modelo estimado resulta tal que $\hat{\theta}(x) = 0$ tiene una raíz unitaria, se reformula el modelo con $d = 1$ y sin el factor MA(1) no invertible (reconsiderando, además, la naturaleza de la tendencia implícita en el modelo reformulado).

- Cuando en la etapa de identificación no están claros los órdenes p , P , q y Q , se puede llevar a cabo un ejercicio de **sobreajuste**, que consiste en (i) incluir (sin un respaldo muestral claro) de uno en uno factores AR y MA de orden 1 ó 2, (ii) estimar el modelo resultante, y (iii) comprobar si cada factor añadido es significativo.

Las operaciones siguientes referidas a los **residuos** del modelo identificado y estimado, son especialmente útiles en la etapa de diagnóstico:

- Examinar el gráfico temporal de la serie de residuos estandarizada. La presencia de cualquier tipo de tendencia suele indicar la conveniencia de aplicar una diferencia regular o estacional (según sea la tendencia) a la serie (w_t) . Si se observa que la dispersión de los residuos no es constante, probablemente la transformación de Box-Cox incluida en (w_t) no es adecuada (o quizás las perturbaciones del modelo presentan **heteroscedasticidad condicional** del tipo que se describe en la Sección 2.8).
- Examinar el histograma de la serie de residuos y utilizar el estadístico de Jarque-Bera para contrastar la hipótesis de Normalidad de las perturbaciones.
- Si el modelo no contiene el parámetro μ_W , contrastar la significación de la media de las perturbaciones utilizando la media residual; si se rechaza la hipótesis nula del contraste, añadir el parámetro μ_W al modelo.
- Examinar la ACF y la PACF residuales. Un amortiguamiento muy lento de la ACF indica que la serie de residuos no es estacionaria, por lo que conviene aplicar una diferencia (regular o estacional, según convenga) adicional a la serie (w_t) . Si la ACF y la PACF presentan alguna configuración reconocible, añadir al modelo los operadores AR y MA correspondientes. Para determinar si las autocorrelaciones simples entre las perturbaciones del modelo son

todas iguales a cero hasta cierto retardo g , puede utilizarse el valor calculado del estadístico de Ljung-Box,

$$Q_{LB}(\hat{a}_t) = n(n-2) \sum_{k=1}^g \frac{r_k(\hat{a}_t)^2}{n-k},$$

que bajo la hipótesis nula de que $\rho_1(A_t) = \rho_2(A_t) = \dots = \rho_g(A_t) = 0$, sigue aproximadamente una distribución $\chi^2(g-m)$, donde m es el número total de parámetros estimados (excluyendo la varianza de las perturbaciones).

Criterios de información

2.6.1 Definición

Para un modelo que contiene m parámetros y ha sido estimado con n observaciones, el **criterio de información de Akaike** es la cantidad

$$\text{AIC} \equiv -\frac{2L^*}{n} + \frac{2m}{n}, \quad (2.46)$$

donde L^* es el logaritmo neperiano de la función de verosimilitud asociada con el modelo estimado, evaluada en las estimaciones obtenidas de los parámetros de dicho modelo.

2.6.2 Definición

Para un modelo que contiene m parámetros y ha sido estimado con n observaciones, el

criterio de información bayesiano de Schwarz es la cantidad

$$\text{BIC} \equiv -\frac{2L^*}{n} + \frac{m \ln(n)}{n}, \quad (2.47)$$

donde L^* es el logaritmo neperiano de la función de verosimilitud asociada con el modelo estimado, evaluada en las estimaciones obtenidas de los parámetros de dicho modelo.

Observación 1: En el contexto de la elaboración de un modelo ARIMA a partir de una serie temporal (y_t) , los criterios de información AIC y BIC (que a veces se denomina SC) suelen emplearse para decidir sobre los órdenes p , P , q y Q asociados con una transformación (estacionaria) dada de (y_t) . Dichos criterios **no** deben utilizarse para discriminar ni entre diferentes transformaciones alternativas de (y_t) , ni entre modelos estimados con diferentes tamaños muestrales efectivos (que pueden variar en función del método de estimación empleado).

Observación 2: Teniendo en cuenta lo anterior, un modelo es tanto más adecuado cuanto más **reducidos** son los valores de sus AIC y BIC. Ambos criterios expresan un compromiso entre la obtención de un buen ajuste (a través del primer sumando de cada uno de ellos) y la obtención de un modelo con un número reducido de parámetros (a través del segundo sumando). En este sentido, puede comprobarse que el BIC siempre selecciona un modelo más **escueto** que el AIC. De hecho, el BIC es siempre preferible al AIC con muestras grandes, aunque el AIC puede ser preferible con algunas muestras cortas. En todo caso, la selección de un modelo no debería basarse únicamente en el empleo de estos criterios (aunque pueden incorporarse al procedimiento de identificación descrito al principio de esta sección).

Contraste de no estacionariedad de Shin-Fuller

2.6.3 Definición

Cuando el operador autorregresivo de orden p en un modelo ARMA(p, q) puede factorizarse como $\phi_p(B) = (1 - \phi B)\phi'_{p-1}(B)$, el **contraste de no estacionariedad de Shin-Fuller (1998)** consiste en **rechazar** la hipótesis nula $\phi = 1$ en favor de la alternativa $\phi < 1$ (con un nivel de significación dado) cuando

$$SF \equiv \begin{cases} L^* - \bar{L} & , \text{ si } \hat{\phi} \leq 1 - \frac{4}{n-1} \\ 0 & , \text{ si } \hat{\phi} > 1 - \frac{4}{n-1} \end{cases} \quad (2.48)$$

es **mayor** que el valor crítico correspondiente de la Tabla 2.1. En (2.48), L^* es el logaritmo neperiano de la función de verosimilitud en la estimación por máxima verosimilitud exacta (MVE) no restringida del modelo, $\hat{\phi}$ es la estimación MVE no restringida del parámetro ϕ , y \bar{L} es el logaritmo neperiano de la función de verosimilitud en la estimación por MVE restringida a que $\phi = 1 - 4/(n - 1)$.

Observación: Ver Shin, D.W., y Fuller, W.A. (1998), Unit root tests based on unconditional maximum likelihood

estimation for the autoregressive moving average, *Journal of Time Series Analysis*, 19, 591-599.

Nivel de significación			
$n - 1$	10%	5%	1%
25	1.02	1.68	3.33
50	1.06	1.75	3.41
100	1.07	1.75	3.41
250	1.07	1.76	3.44
500	1.08	1.77	3.46

Tabla 2.1

Valores críticos para el contraste de no estacionariedad de Shin-Fuller (1998)

Contrastes de no invertibilidad de Davis-Chen-Dunsmuir

2.6.4 Definición

En un modelo MA(1) del tipo $Y_t = (1 - \theta_1 B)A_t$, el **contraste MLE de no invertibilidad de Davis-Chen-Dunsmuir (1995)** consiste en **rechazar** la hipótesis nula $\theta_1 = 1$ en favor de la alternativa $\theta_1 < 1$ (con un nivel de significación dado) cuando $n(1 - \hat{\theta}_1)$ es **mayor** que el

valor crítico b_{MLE} correspondiente de la Tabla 2.2 (o bien cuando $\hat{\theta}_1 < 1 - \frac{b_{MLE}}{n}$), donde $\hat{\theta}_1$ es la estimación MVE no restringida de θ_1 .

	Nivel de significación			
	10%	5%	2.5%	1%
b_{MLE}	4.90	6.80	8.97	11.93
b_{GLR}	1.00	1.94	2.95	4.41

Tabla 2.2

Valores críticos asintóticos para los contrastes de no invertibilidad de Davis-Chen-Dunsmuir (1995)

2.6.5 Definición

En un modelo MA(1) del tipo $Y_t = (1 - \theta_1 B)A_t$, el **contraste GLR de no invertibilidad de Davis-Chen-Dunsmuir (1995)** consiste en **rechazar** la hipótesis nula $\theta_1 = 1$ en favor de la alternativa $\theta_1 < 1$ (con un nivel de significación α dado) cuando

$$DCD \equiv 2(L^* - \bar{L}) \quad (2.49)$$

es **mayor** que el valor crítico b_{GLR} correspondiente de la Tabla 2.2 anterior. En (2.49), L^*

es el logaritmo neperiano de la función de verosimilitud en la estimación por MVE no restringida del modelo MA(1), y \bar{L} es el logaritmo neperiano de la función de verosimilitud evaluada en $\theta_1 = 1$.

Observación: Ver Davis, R.A., Chen, M., y Dunsmuir, W.T.M. (1995), Inference for MA(1) processes with a root on or near the unit circle, *Probability and Mathematical Statistics*, 15, 227-242.

2.6.6 Ejemplo

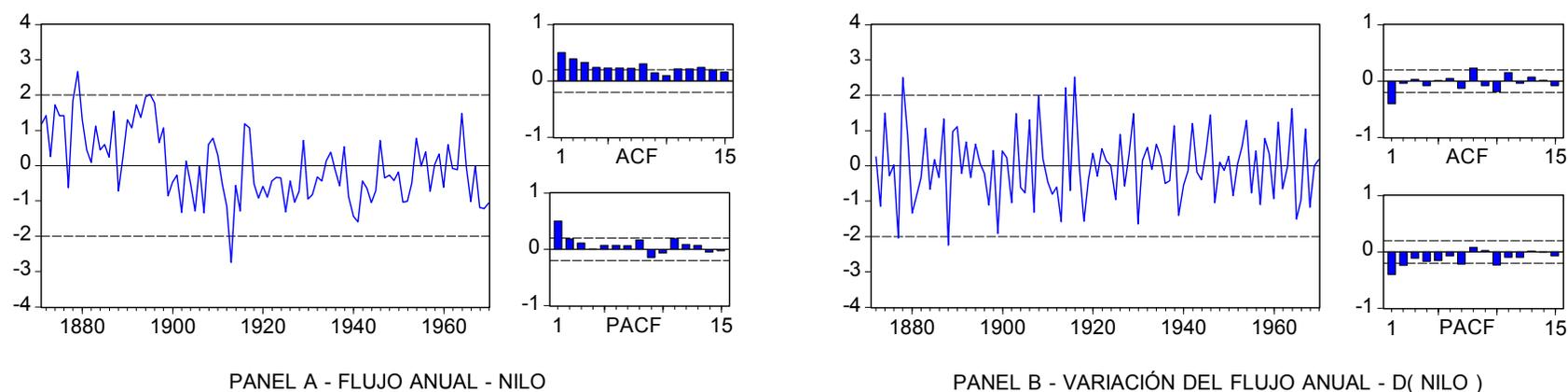


Figura 2.33

Volumen del flujo anual del río Nilo en Asuán (1871 – 1970)
(nilo.wf1)

Si (Y_t) representa el proceso estocástico Volumen del flujo anual del río Nilo en Asuán, la Figura 2.33 sugiere que o bien $(Y_t) \sim I(0) \sim AR(1)$, o bien $(Y_t) \sim I(1) \sim IMA(1,1)$. Para decidir sobre esta cuestión:

- 1 Contrastar $\phi_1 = 1$ frente a $\phi_1 < 1$ en el modelo

$$(1 - \phi_1 B)Y_t = \mu + (1 - \theta_1 B)A_t. \quad (2.50)$$

Si $\phi_1 = 1$, (2.50) indica que $(Y_t) \sim I(1) \sim IMA(1,1)$ (probablemente con $\mu = 0$). Si $\phi_1 < 1$, (2.50) indica que $(Y_t) \sim I(0) \sim ARMA(1,1)$ (quizás con $\theta_1 = 0$, de acuerdo con el Panel A de la Figura 2.33).

- 2 Contrastar $\theta_1 = 1$ frente a $\theta_1 < 1$ en el modelo

$$(1 - \phi_1 B)\nabla Y_t = (1 - \theta_1 B)A_t \quad (2.51)$$

Si $\theta_1 = 1$, (2.51) es (2.50) con $\theta_1 = 0$, lo que indica que $(Y_t) \sim I(0) \sim AR(1)$. Si $\theta_1 < 1$, (2.51) implica que $(Y_t) \sim I(1) \sim ARIMA(1,1,1)$ (probablemente con $\phi_1 = 0$, de acuerdo con el Panel B de la Figura 2.33).

1 Contraste de no estacionariedad de Shin-Fuller

La estimación por MVE de (2.50) proporciona $\hat{\phi}_1 = 0.8610 < 1 - 4 / (100 - 1) \approx 0.9596$, $L^* = 53.7367$. Por otro lado, la estimación por MVE de (2.50) bajo la restricción de que $\phi_1 = 0.9596$ proporciona $\bar{L} = 53.0453$. Por lo tanto, $SF = L^* - \bar{L} = 0.6914$, que es **menor** que cualquier valor crítico de la Tabla 2.1. En consecuencia, la hipótesis de que $\phi_1 = 1$ en (2.50) **no se rechaza** ni siquiera al 10%.

2 Contrastes de no invertibilidad de Davis-Chen-Dunsmuir

La estimación por MVE de (2.51) proporciona $\hat{\theta}_1 = 0.8741$, $L^* = 53.2404$; por lo tanto, $n(1 - \hat{\theta}_1) = 12.46$, que es **mayor** que cualquier valor crítico b_{MLE} de la Tabla 2.2. Por otro lado, la estimación por MVE de (2.51) bajo la restricción de que $\theta_1 = 1$ proporciona $\bar{L} = 48.2245$; por lo tanto, $DCD = 2(L^* - \bar{L}) = 10.0318$, que es **mayor** que cualquier valor crítico b_{GLR} de la Tabla 2.2. En ambos casos, la hipótesis de que $\theta_1 = 1$ en (2.51) **se rechaza** incluso al 1%.

Los resultados de los contrastes anteriores sugieren claramente que $(Y_t) \sim I(1)$.

2.6.7 Ejemplo

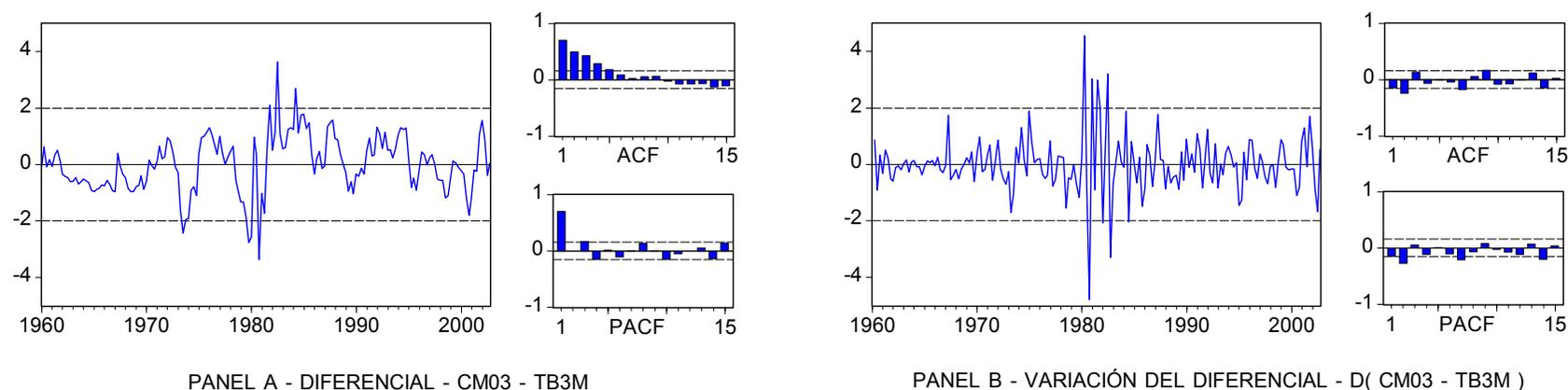


Figura 2.34

Diferencial de la rentabilidad de la deuda pública en EEUU a 3 años y a 3 meses (1960:I – 2002:IV)
(interest-rates-q.wf1)

Si (Y_t) representa el proceso estocástico Diferencial 3 años – 3 meses, la Figura 2.34 sugiere que o bien $(Y_t) \sim I(0) \sim AR(1)$, o bien $(Y_t) \sim I(1)$ (probablemente un paseo aleatorio).

1 Contrastar $\phi_1 = 1$ frente a $\phi_1 < 1$ en el modelo

$$(1 - \phi_1 B)Y_t = \mu + A_t. \quad (2.52)$$

Si $\phi_1 = 1$, (2.52) indica que $(Y_t) \sim I(1)$ (un paseo aleatorio, quizás con $\mu = 0$). Si $\phi_1 < 1$, (2.52) indica que $(Y_t) \sim I(0) \sim AR(1)$.

2 Contrastar $\theta_1 = 1$ frente a $\theta_1 < 1$ en el modelo

$$(1 - \phi_1 B)\nabla Y_t = (1 - \theta_1 B)A_t \quad (2.53)$$

Si $\theta_1 = 1$, (2.53) es (2.52), lo que indica que $(Y_t) \sim I(0) \sim AR(1)$. Si $\theta_1 < 1$, (2.53) implica que $(Y_t) \sim I(1) \sim ARIMA(1, 1, 1)$ (probablemente con $\phi_1 = \theta_1 = 0$, de acuerdo con el Panel B de la Figura 2.34).

1 Contraste de no estacionariedad de Shin-Fuller

La estimación por MVE de (2.52) proporciona $\hat{\phi}_1 = 0.6990 < 1 - 4/(172 - 1) \approx 0.9766$, $L^* = -157.0281$. Por otro lado, la estimación por MVE de (2.52) bajo la restricción de que $\phi_1 = 0.9766$ proporciona $\bar{L} = -170.0937$. Por lo tanto, $SF = L^* - \bar{L} = 13.0656$, que es **mayor** que cualquier valor crítico de la Tabla 2.1. En consecuencia, la hipótesis de que $\phi_1 = 1$ en (2.52) **se rechaza** incluso al 1%.

2 Contrastes de no invertibilidad de Davis-Chen-Dunsmuir

La estimación por MVE de (2.53) proporciona $\hat{\theta}_1 = 0.9895$, $L^* = -157.9638$; por lo tanto, $n(1 - \hat{\theta}_1) = 1.7955$, que es **menor** que cualquier valor crítico b_{MLE} de la Tabla 2.2. Por otro lado, la estimación por MVE de (2.53) bajo la restricción de que $\theta_1 = 1$ proporciona $\bar{L} = -157.9860$; por lo tanto, $DCD = 2(L^* - \bar{L}) = 0.0444$, que es **menor** que cualquier valor crítico b_{GLR} de la Tabla 2.2. En ambos casos, la hipótesis de que $\theta_1 = 1$ en (2.53) **no se rechaza** ni siquiera al 10%.

Los resultados de los contrastes anteriores sugieren claramente que $(Y_t) \sim I(0)$.

2.6.8 Ejemplo

Si (Y_t) representa el proceso estocástico PIB real anual en España, la Figura 2.35 sugiere que o bien $(\ln Y_t) \sim I(1) \sim \text{ARI}(1, 1)$ [lo que equivale a que $(\nabla \ln Y_t) \sim I(0) \sim \text{AR}(1)$], o bien $(\ln Y_t) \sim I(2) \sim \text{ARIMA}(0, 2, 0)$ [lo que equivale a que $(\nabla \ln Y_t) \sim I(1)$ es un paseo aleatorio]. Para decidir sobre esta cuestión:

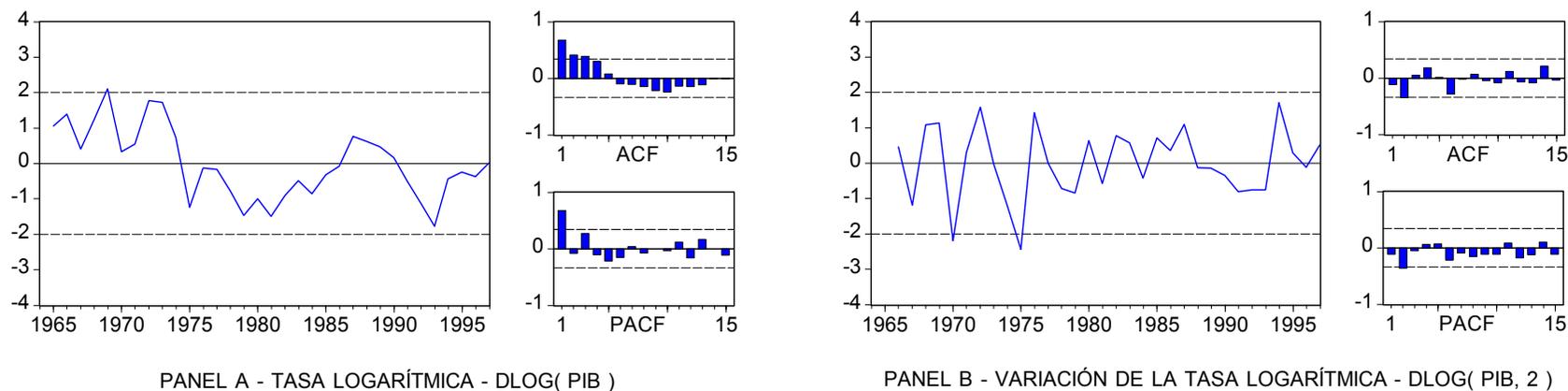


Figura 2.35

Tasa logarítmica de variación anual del PIB real en España (1965 – 1997)
(pib-cpi.wf1)

- 1 Contrastar $\phi_1 = 1$ frente a $\phi_1 < 1$ en el modelo

$$(1 - \phi_1 B) \nabla \ln Y_t = \mu + A_t. \quad (2.54)$$

Si $\phi_1 = 1$, (2.54) indica que $(\ln Y_t) \sim I(2)$ (un paseo aleatorio, quizás con $\mu = 0$). Si $\phi_1 < 1$, (2.54) indica que $(\ln Y_t) \sim I(1) \sim \text{ARI}(1, 1)$.

- 2 Contrastar $\theta_1 = 1$ frente a $\theta_1 < 1$ en el modelo

$$(1 - \phi_1 B)\nabla^2 \ln Y_t = (1 - \theta_1 B)A_t \quad (2.55)$$

Si $\theta_1 = 1$, (2.55) es (2.54), lo que indica que $(\ln Y_t) \sim I(1) \sim \text{ARI}(1, 1)$. Si $\theta_1 < 1$, (2.55) implica que $(Y_t) \sim I(2) \sim \text{ARIMA}(1, 2, 1)$ (probablemente con $\phi_1 = \theta_1 = 0$, de acuerdo con el Panel B de la Figura 2.35).

1 Contraste de no estacionariedad de Shin-Fuller

La estimación por MVE de (2.54) proporciona $\hat{\phi}_1 = 0.6790 < 1 - 4 / (33 - 1) = 0.875$, $L^* = 87.0613$. Por otro lado, la estimación por MVE de (2.54) bajo la restricción de que $\phi_1 = 0.875$ proporciona $\bar{L} = 85.7555$. Por lo tanto, $SF = L^* - \bar{L} = 1.3058$; de acuerdo con la Tabla 2.1, la hipótesis de que $\phi_1 = 1$ en (2.54) **se rechaza** al 10%, aunque **no se puede rechazar** al 5%.

2 Contrastes de no invertibilidad de Davis-Chen-Dunsmuir

La estimación por MVE de (2.55) proporciona $\hat{\theta}_1 = 0.9038$, $L^* = 83.3913$; por lo tanto, $n(1 - \hat{\theta}_1) = 3.1746$, que es **menor** que cualquier valor crítico b_{MLE} de la Tabla 2.2. Por otro lado, la estimación por MVE de (2.55) bajo la restricción de que $\theta_1 = 1$

proporciona $\bar{L} = 83.3211$; por lo tanto, $DCD = 2(L^* - \bar{L}) = 0.1404$, que es **menor** que cualquier valor crítico b_{GLR} de la Tabla 2.2. En ambos casos, la hipótesis de que $\theta_1 = 1$ en (2.55) **no se rechaza** ni siquiera al 10%.

En este ejemplo, el contraste de Shin-Fuller podría sugerir que $(\ln Y_t) \sim I(2)$, aunque los contrastes de Davis-Chen-Dunsmuir parecen indicar que $(\ln Y_t) \sim I(1)$.

Observación 1: Como indican los tres ejemplos anteriores, los contrastes de **raíces unitarias** de Shin-Fuller (SF) y de Davis-Chen-Dunsmuir (DCD) son útiles para decidir sobre el orden de integración de un proceso estocástico. En general, cuando se duda entre $d = 0$ y $d = 1$: **[1]** escribir un modelo plausible para (∇Y_t) , reemplazar ∇Y_t en dicho modelo por $(1 - \phi B)$, y contrastar con SF $\phi = 1$ frente a $\phi < 1$; alternativamente, **[2]** escribir un modelo plausible para (Y_t) , añadir en dicho modelo ∇ en el lado izquierdo y $(1 - \theta B)$ en el lado derecho (testigo de posible sobrediferenciación), y contrastar con DCD $\theta = 1$ frente a $\theta < 1$. Cuando se duda entre $d = 1$ y $d = 2$: **[1]** escribir un modelo plausible para $(\nabla^2 Y_t)$, reemplazar $\nabla^2 Y_t$ en dicho modelo por $(1 - \phi B)\nabla$, y contrastar con SF $\phi = 1$ frente a $\phi < 1$; o bien **[2]** escribir un modelo plausible para (∇Y_t) , añadir en dicho modelo ∇ en el lado izquierdo y $(1 - \theta B)$ en el lado derecho, y contrastar con DCD $\theta = 1$ frente a $\theta < 1$.

Observación 2: Los contrastes SF y DCD son los contrastes de raíces unitarias más recomendables entre los disponibles actualmente. Sin embargo, dado que su empleo requiere la estimación por MVE (no condicionada) de modelos ARIMA, no son ni remotamente tan populares como otros contrastes que requieren métodos de estimación más sencillos (aunque, igualmente, menos recomendables que MVE). En particular, los contrastes SF y DCD no están implementados en EViews, aunque sí lo están varios contrastes menos recomendables en general; ver Sección 4.3.

2.7 PREVISIÓN CON MODELOS ARIMA

2.7.1 Definición

La **previsión puntual** en **origen** k (dado) a **horizonte** $l \geq 1$ (variable) de un proceso estocástico (Y_t) , se representa con el símbolo $Y_k(l)$ y se define como el **valor esperado** de Y_{k+l} **condicionado** por toda la información disponible hasta un momento k dado de la historia de (Y_t) ; es decir,

$$Y_k(l) \equiv E_k[Y_{k+l}] \text{ para } l = 1, 2, \dots \quad (2.56)$$

Observación: Considerada como una función del horizonte de previsión $l \geq 1$, $Y_k(l)$ se denomina la **función de previsión** en origen k (dado) del proceso (Y_t) . En la práctica, el origen de previsión k suele ser igual al momento N de la última observación de una serie temporal y_1, y_2, \dots, y_N empleada en la elaboración de un modelo para (Y_t) .

2.7.2 Definición

El **error de previsión** en origen k (dado) a horizonte $l \geq 1$ (variable) de un proceso estocástico (Y_t) , es la variable aleatoria

$$E_k(l) \equiv Y_{k+l} - Y_k(l) \text{ para } l = 1, 2, \dots, \quad (2.57)$$

es decir, la diferencia entre el **valor observado** Y_{k+l} de (Y_t) en un momento $k + l$ de su historia y el **valor previsto** $Y_k(l)$ en origen k a horizonte l de (Y_t) .

2.7.3 Definición

La **función de previsión** en origen k a horizonte $l \geq 1$ de un proceso $(Y_t) \sim I(d)$ que sigue un **modelo ARIMA** del tipo $\phi'(B)Y_t = \mu + \theta(B)A_t$ (ver Observación 1 de 2.5.1), donde $\phi'(B) \equiv \phi(B)\nabla^d \equiv 1 - \sum_{i=1}^{p'} \phi'_i B^i$ ($p' \equiv p + d$), es

$$Y_k(l) \equiv \mathbf{E}_k[Y_{k+l}] = \mu + \sum_{i=1}^{p'} \phi'_i \mathbf{E}_k[Y_{k+l-i}] + \mathbf{E}_k[A_{k+l}] - \sum_{i=1}^q \theta_i \mathbf{E}_k[A_{k+l-i}], \quad (2.58)$$

o bien [ver (2.40)-(2.43)]

$$Y_k(l) \equiv \mathbf{E}_k[Y_{k+l}] = \mu_{k+l} + \mathbf{E}_k[\psi^*(B)A_{k+l}], \quad (2.59)$$

donde $\psi^*(B) \equiv \sum_{i=0}^{\infty} \psi_i^* B^i$ ($\psi_0^* \equiv 1$) se define de acuerdo con que $\nabla^d \psi^*(B) = \psi(B)$.

Observación 1: La función de previsión (2.58) de un proceso $(Y_t) \sim \text{ARIMA}(p, d, q)$ se utiliza en la práctica para calcular previsiones puntuales $\hat{y}_k(1), \hat{y}_k(2), \dots$ a partir de un modelo estimado. Para ello, en (2.58) se reemplazan los parámetros $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ por sus valores estimados, se reemplaza cada $\mathbf{E}_k[Y_j]$ por y_j (un dato) si $j \leq k$,

o bien por $\hat{y}_k(j-k)$ (una previsión) si $j > k$, y se reemplaza cada $E_k[A_j]$ por \hat{a}_j (un residuo) si $j \leq k$, o bien por 0 (la esperanza incondicional de A_j) si $j > k$. Los errores de previsión calculados según (2.57), asociados con las previsiones $\hat{y}_k(1), \hat{y}_k(2), \dots$, suelen representarse como $\hat{e}_k(l) \equiv y_{k+l} - \hat{y}_k(l)$ ($l = 1, 2, \dots$) (que sólo pueden calcularse cuando y_{k+1}, y_{k+2}, \dots son datos conocidos); por su parte, la función de previsión (2.59) se utiliza para describir las propiedades estadísticas de los errores de previsión (ver Definición 2.7.4).

Observación 2: Un modelo ARIMA(p, d, q) estimado puede escribirse como $\hat{\phi}(B)\nabla^d y_t = \hat{\mu} + \hat{\theta}(B)\hat{a}_t$, o bien como $\hat{\phi}'(B)y_t = \hat{\mu} + \hat{\theta}(B)\hat{a}_t$, $\hat{\phi}'(B) \equiv \hat{\phi}(B)\nabla^d \equiv (1 - \sum_{i=1}^p \hat{\phi}_i B^i)(1 - B)^d \equiv (1 - \sum_{i=1}^{p'} \hat{\phi}'_i B^i)$, $\hat{\theta}(B) \equiv 1 - \sum_{i=1}^q \hat{\theta}_i B^i$; en consecuencia, cada residuo \hat{a}_t de un modelo estimado puede escribirse como

$$\hat{a}_t = y_t - \left[\hat{\mu} + \sum_{i=1}^{p'} \hat{\phi}'_i y_{t-i} - \sum_{i=1}^q \hat{\theta}_i \hat{a}_{t-i} \right],$$

donde el término entre corchetes es un **valor ajustado**, que puede interpretarse como una previsión en origen $t-1$ a horizonte 1. Por lo tanto, \hat{a}_t puede interpretarse como un error de previsión en origen $t-1$ a horizonte 1.

2.7.4 Definición

El **error de previsión** en origen k a horizonte $l \geq 1$ de un proceso $(Y_t) \sim I(d)$ que sigue un **modelo ARIMA** del tipo $\phi'(B)Y_t = \mu + \theta(B)A_t$ (ver 2.7.3), puede escribirse, de acuerdo con (2.42) y (2.59), como

$$E_k(l) \equiv Y_{k+l} - Y_k(l) = \psi^*(B)A_{k+l} - E_k[\psi^*(B)A_{k+l}],$$

donde $\psi^*(B)A_{k+l} = \sum_{i=0}^{\infty} \psi_i^* A_{k+l-i} = A_{k+l} + \psi_1^* A_{k+l-1} + \psi_2^* A_{k+l-2} + \dots$ Por lo tanto,

$$E_k(l) = A_{k+l} + \psi_1^* A_{k+l-1} + \dots + \psi_{l-1}^* A_{k+1} = \sum_{i=0}^{l-1} \psi_i^* A_{k+l-i}, \quad (2.60)$$

donde los parámetros ψ_i^* del polinomio $\psi^*(B) \equiv \sum_{i=0}^{\infty} \psi_i^* B^i$ ($\psi_0^* \equiv 1$) pueden calcularse teniendo en cuenta que $\nabla^d \psi^*(B) = \psi(B)$, con $\psi(B) \equiv \theta(B) / \phi(B)$; ver la explicación alrededor de (2.15).

Observación: (2.60) implica que $E[E_k(l)] = 0$ y que $\text{Var}[E_k(l)] = \sigma_A^2 \sum_{i=0}^{l-1} (\psi_i^*)^2$ en cualquier modelo ARIMA con cualquier valor para $d \geq 0$; en particular, si $d = 0$, $\psi^*(B) = \psi(B)$, por lo que $\text{Var}[E_k(l)] = \sigma_A^2 \sum_{i=0}^{l-1} \psi_i^2$ en este caso. Como $\text{Var}[E_k(l)]$ depende en cualquier caso del horizonte $l \geq 1$ pero **no** del origen k de previsión, $\text{Var}[E_k(l)]$ suele escribirse simplemente como $V(l)$, $l = 1, 2, \dots$

2.7.5 Ejemplo – Previsión con un modelo ARMA(1,1)

El modelo ARMA(1,1), o ARIMA(1,0,1), para un proceso $(Y_t) \sim I(0)$ es

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t - \theta_1 A_{t-1}.$$

Escribiendo esta expresión en $t = N + l$ ($l \geq 1$) y aplicando al resultado el operador $E_N[\cdot]$, se obtiene que la función de previsión en origen N a horizonte $l \geq 1$ es

$$Y_N(l) = \mu + \phi_1 E_N[Y_{N+l-1}] + E_N[A_{N+l}] - \theta_1 E_N[A_{N+l-1}]. \quad (2.61)$$

Para $l = 1$, (2.61) queda

$$Y_N(1) = \mu + \phi_1 Y_N - \theta_1 A_N = \mu + \phi_1 \left(Y_N - \frac{\theta_1}{\phi_1} A_N \right), \quad (2.62)$$

mientras que para $l \geq 2$, (2.61) queda

$$Y_N(l) = \mu + \phi_1 Y_N(l-1). \quad (2.63)$$

(2.63) es una ecuación en diferencias, cuya solución a partir de la condición inicial (2.62) es

$$Y_N(l) = \left(\sum_{i=0}^{l-1} \phi_1^i \right) \mu + \phi_1^l \left(Y_N - \frac{\theta_1}{\phi_1} A_N \right) \text{ para todo } l \geq 1. \quad (2.64)$$

(2.64) se denomina la **función final de previsión** de $(Y_t) \sim \text{ARMA}(1,1)$, que converge al valor esperado de (Y_t) , $E[Y_t] = \mu / (1 - \phi_1)$, cuando $|\phi_1| < 1$ y $l \rightarrow \infty$.

Observación: Cada error de previsión puede calcularse (junto con su varianza) a partir de la diferencia entre cada $Y_{N+l} = \mu + \phi_1 Y_{N+l-1} + A_{N+l} - \theta_1 A_{N+l-1}$ y la previsión (2.62)-(2.63) correspondiente, para $l = 1, 2, 3, \dots$:

$$\begin{aligned} E_N(1) &= A_{N+1}, \\ E_N(2) &= \phi_1 E_N(1) + A_{N+2} - \theta_1 A_{N+1} = A_{N+2} + (\phi_1 - \theta_1) A_{N+1}, \\ E_N(3) &= \phi_1 E_N(2) + A_{N+3} - \theta_1 A_{N+2} = A_{N+3} + (\phi_1 - \theta_1) A_{N+2} + \phi_1 (\phi_1 - \theta_1) A_{N+1}, \\ &\dots, \end{aligned}$$

$$E_N(l) = \sum_{i=0}^{l-1} \psi_i A_{N+l-i}, \text{ donde } \psi_0 \equiv 1, \psi_i = \phi_1^{i-1} (\phi_1 - \theta_1) \text{ (} i \geq 1 \text{)}.$$

Por lo tanto,

$$V(l) = \sigma_A^2 \sum_{i=0}^{l-1} \psi_i^2 = \left[1 + (\phi_1 - \theta_1)^2 \sum_{i=1}^{l-1} \phi_1^{2(i-1)} \right] \sigma_A^2,$$

que converge a la varianza de (Y_t) , $\text{Var}(Y_t) = [1 + (\phi_1 - \theta_1)^2 / (1 - \phi_1^2)] \sigma_A^2$, cuando $|\phi_1| < 1$ y $l \rightarrow \infty$.

2.7.6 Ejemplo – Previsión con un modelo AR(1)

El modelo AR(1), o ARIMA(1,0,0), para un proceso $(Y_t) \sim I(0)$ es

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t.$$

Escribiendo esta expresión en $t = N + l$ ($l \geq 1$), y aplicando al resultado el operador $E_N[\cdot]$, se obtiene que la función de previsión en origen N a horizonte $l \geq 1$ es

$$Y_N(l) = \mu + \phi_1 E_N[Y_{N+l-1}] + E_N[A_{N+l}]. \quad (2.65)$$

Para $l = 1$, (2.65) queda

$$Y_N(1) = \mu + \phi_1 Y_N, \quad (2.66)$$

mientras que para $l \geq 2$, (2.65) queda

$$Y_N(l) = \mu + \phi_1 Y_N(l-1). \quad (2.67)$$

(2.67) es una ecuación en diferencias, cuya solución a partir de la condición inicial (2.66) es

$$Y_N(l) = \left(\sum_{i=0}^{l-1} \phi_1^i \right) \mu + \phi_1^l Y_N \quad \text{para todo } l \geq 1. \quad (2.68)$$

(2.68) es la **función final de previsión** de $(Y_t) \sim \text{AR}(1)$, que converge al valor esperado de (Y_t) , $E[Y_t] = \mu / (1 - \phi_1)$, cuando $|\phi_1| < 1$ y $l \rightarrow \infty$.

Observación: Cada error de previsión puede calcularse (junto con su varianza) a partir de la diferencia entre cada $Y_{N+l} = \mu + \phi_1 Y_{N+l-1} + A_{N+l}$ y la previsión (2.66)-(2.67) correspondiente, para $l = 1, 2, 3, \dots$:

$$\begin{aligned} E_N(1) &= A_{N+1}, \\ E_N(2) &= \phi_1 E_N(1) + A_{N+2} = A_{N+2} + \phi_1 A_{N+1}, \\ E_N(3) &= \phi_1 E_N(2) + A_{N+3} = A_{N+3} + \phi_1 A_{N+2} + \phi_1^2 A_{N+1}, \\ &\dots, \\ E_N(l) &= \sum_{i=0}^{l-1} \psi_i A_{N+l-i}, \quad \text{donde } \psi_i = \phi_1^i \quad (i \geq 0). \end{aligned}$$

Por lo tanto,

$$V(l) = \sigma_A^2 \sum_{i=0}^{l-1} \psi_i^2 = \left(\sum_{i=0}^{l-1} \phi_1^{2i} \right) \sigma_A^2,$$

que converge a la varianza de (Y_t) , $\text{Var}(Y_t) = [1 / (1 - \phi_1^2)] \sigma_A^2$, cuando $|\phi_1| < 1$ y $l \rightarrow \infty$.

2.7.7 Ejemplo – Previsión con un modelo MA(1)

El modelo MA(1), o ARIMA(0,0,1), para un proceso $(Y_t) \sim I(0)$ es

$$Y_t = \mu + A_t - \theta_1 A_{t-1}.$$

Escribiendo esta expresión en $t = N + l$ ($l \geq 1$), y aplicando al resultado el operador $E_N[\cdot]$, se obtiene que la función de previsión en origen N a horizonte $l \geq 1$ es

$$Y_N(l) = \mu + E_N[A_{N+l}] - \theta_1 E_N[A_{N+l-1}]. \quad (2.69)$$

Para $l = 1$, (2.69) queda

$$Y_N(1) = \mu - \theta_1 A_N, \quad (2.70)$$

mientras que para $l \geq 2$, (2.69) queda

$$Y_N(l) = \mu. \quad (2.71)$$

(2.71) es la **función final de previsión** de $(Y_t) \sim \text{MA}(1)$, que coincide con $E[Y_t] = \mu$.

Observación 1: Cada error de previsión puede calcularse (junto con su varianza) a partir de la diferencia entre cada $Y_{N+l} = \mu + A_{N+l} - \theta_1 A_{N+l-1}$ y la previsión (2.70)-(2.71) correspondiente, para $l = 1, 2, 3, \dots$:

$$\begin{aligned}
 E_N(1) &= A_{N+1}, \\
 E_N(2) &= A_{N+2} - \theta_1 A_{N+1}, \\
 E_N(3) &= A_{N+3} - \theta_1 A_{N+2}, \\
 &\dots,
 \end{aligned}$$

$$E_N(l) = \sum_{i=0}^{l-1} \psi_i A_{N+l-i}, \text{ donde } \psi_0 \equiv 1, \psi_1 = -\theta_1, \psi_i = 0 \ (i \geq 2).$$

Por lo tanto,

$$V(1) = \sigma_A^2, \quad V(l) = (1 + \theta_1^2) \sigma_A^2 \quad (l \geq 2),$$

que a partir de $l = 2$ es constante y coincide con $\text{Var}(Y_t) = (1 + \theta_1^2) \sigma_A^2$.

Observación 2: En general, puede comprobarse si (Y_t) es un proceso $I(0)$, de manera que $(Y_t) \sim \text{ARMA}(p, q)$ estacionario e invertible, $Y_N(l) \rightarrow E(Y_t)$ y $V(l) \rightarrow \text{Var}(Y_t)$ cuando el horizonte de previsión l tiende a infinito.

2.7.8 Ejemplo – Previsión con un modelo IMA(1,1)

El modelo IMA(1,1), o ARIMA(0,1,1), para un proceso $(Y_t) \sim I(1)$, puede escribirse como

$$Y_t = \mu + Y_{t-1} + A_t - \theta_1 A_{t-1}.$$

Escribiendo esta expresión en $t = N + l$ ($l \geq 1$) y aplicando al resultado el operador $E_N[\cdot]$, se obtiene que la función de previsión en origen N a horizonte $l \geq 1$ es

$$Y_N(l) = \mu + E_N[Y_{N+l-1}] + E_N[A_{N+l}] - \theta_1 E_N[A_{N+l-1}], \quad (2.72)$$

Para $l = 1$, (2.72) queda

$$Y_N(1) = \mu + Y_N - \theta_1 A_N, \quad (2.73)$$

mientras que para $l \geq 2$, (2.72) queda

$$Y_N(l) = \mu + Y_N(l-1). \quad (2.74)$$

(2.75) es una ecuación en diferencias, cuya solución a partir de la condición inicial (2.73) es

$$Y_N(l) = l\mu + Y_N - \theta_1 A_N \quad \text{para todo } l \geq 1. \quad (2.75)$$

(2.75) es la **función final de previsión** de $(Y_t) \sim \text{IMA}(1, 1)$, que no depende de l si $\mu = 0$, pero diverge cuando $l \rightarrow \infty$ si $\mu \neq 0$.

Observación 1: Cada error de previsión puede calcularse (junto con su varianza) a partir de la diferencia entre cada $Y_{N+l} = \mu + Y_{N+l-1} + A_{N+l} - \theta_1 A_{N+l-1}$ y la previsión (2.73)-(2.74) correspondiente, para $l = 1, 2, 3, \dots$:

$$E_N(1) = A_{N+1},$$

$$E_N(2) = E_N(1) + A_{N+2} - \theta_1 A_{N+1} = A_{N+2} + (1 - \theta_1)A_{N+1},$$

$$E_N(3) = E_N(2) + A_{N+3} - \theta_1 A_{N+2} = A_{N+3} + (1 - \theta_1)A_{N+2} + (1 - \theta_1)A_{N+1},$$

$$\dots,$$

$$E_N(l) = \sum_{i=0}^{l-1} \psi_i^* A_{N+l-i}, \text{ donde } \psi_0^* \equiv 1, \psi_i^* = (1 - \theta_1) \text{ (} i \geq 1 \text{)}.$$

Por lo tanto,

$$V(l) = \sigma_A^2 \sum_{i=0}^{l-1} (\psi_i^*)^2 = [1 + (1 - \theta_1)^2 (l - 1)] \sigma_A^2,$$

de manera que $V(l) \rightarrow \infty$ cuando $l \rightarrow \infty$.

Observación 2: Cuando (Y_t) es un paseo aleatorio, $Y_t = \mu + Y_{t-1} + A_t$, los resultados del Ejemplo (2.7.8) con $\theta_1 = 0$ indican que $Y_N(l) = l\mu + Y_N$ para todo $l \geq 1$, con $V(l) = l\sigma_A^2$. En general, puede comprobarse que cuando (Y_t) es un proceso $I(d)$ con $d \geq 1$, de manera que $(Y_t) \sim \text{ARIMA}(p, d, q)$ con $d \geq 1$, tanto $Y_N(l)$ como $V(l)$ son funciones que divergen cuando el horizonte de previsión l tiende a infinito; en particular, $V(l) \rightarrow \infty$ cuando $l \rightarrow \infty$, lo cual contrasta claramente con los resultados de los ejemplos 2.7.5-2.7.7 para procesos $I(0)$.

Previsiones puntuales – Intervalos de confianza

El modelo ARIMA más general considerado en la Sección 2.6,

$$\phi_p(B)\Phi_P(B^S)[\nabla^d \nabla_S^D Y_t' - \mu_W] = \theta_q(B)\Theta_Q(B^S)A_t,$$

puede escribirse como

$$\Phi^*(B)Y_t' = \mu + \Theta^*(B)A_t, \quad (2.76)$$

donde

$$\Phi^*(B) \equiv \phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D \equiv 1 - \sum_{i=1}^{p^*} \Phi_i^* B^i,$$

$$\Theta^*(B) \equiv \theta_q(B)\Theta_Q(B^S) \equiv 1 - \sum_{i=1}^{q^*} \Theta_i^* B^i,$$

$\mu \equiv [\phi_p(1)\Phi_P(1)]\mu_W$, $p^* \equiv (p+d) + (P+D)S$, $q^* \equiv q + QS$. Por lo tanto, (2.76) queda

$$Y_t' = \mu + \sum_{i=1}^{p^*} \Phi_i Y_{t-i}' + A_t - \sum_{i=1}^{q^*} \Theta_i^* A_{t-i}. \quad (2.77)$$

Cálculo de previsiones puntuales para la serie transformada:

$$\hat{y}'_N(l) = \hat{\mu} + \sum_{i=1}^{p^*} \hat{\Phi}_i [y'_{N+l-i}] - \sum_{i=1}^{q^*} \hat{\Theta}_i^* [a_{N+l-i}] \quad (l \geq 1), \quad (2.78)$$

donde

$$[y'_j] \equiv \begin{cases} y'_j & , \text{ si } j \leq N, \\ \hat{y}'_N(j-N) & , \text{ si } j > N. \end{cases} \quad [a_j] \equiv \begin{cases} \hat{a}_j \equiv y'_j - \hat{y}'_{j-1}(1) & , \text{ si } j \leq N, \\ 0 & , \text{ si } j > N. \end{cases}$$

Cálculo de intervalos de confianza para la serie transformada:

$$\text{IC}'_{1-\alpha}(l) = [\hat{y}'_{N+l}(\mp)] = [\hat{y}'_N(l) \mp u \times \sqrt{\hat{v}(l)}], \quad (2.79)$$

donde $u = 1.65, 1.96$ ó 2.58 , según si $\alpha = 0.10, 0.05$ ó 0.01 , $\hat{v}(l) \equiv \hat{\sigma}_A^2 \sum_{i=0}^{l-1} (\hat{\psi}_i^*)^2$, y

$$\hat{\psi}_i^* = -\hat{\Theta}_i^* + \sum_{j=1}^{p^*} \hat{\Phi}_j^* \hat{\psi}_{i-j}^* \quad (i \geq 1), \quad (2.80)$$

con $\hat{\psi}_0^* \equiv 1$, $\hat{\psi}_i^* = 0$ si $i < 0$, y $\hat{\Theta}_i^* = 0$ si $i > q^*$.

Cálculo de previsiones puntuales e intervalos de confianza para la serie original:

$$\hat{y}_N(l) \equiv \begin{cases} \exp [\hat{y}'_N(l)] , & \text{si } \lambda = 0, \\ [\hat{y}'_N(l)]^{\frac{1}{\lambda}} , & \text{si } \lambda \neq 0, \end{cases} \quad \hat{y}_{N+l}(\mp) \equiv \begin{cases} \exp [\hat{y}'_{N+l}(\mp)] , & \text{si } \lambda = 0, \\ [\hat{y}'_{N+l}(\mp)]^{\frac{1}{\lambda}} , & \text{si } \lambda \neq 0, \end{cases} \quad (2.81)$$

donde λ es el parámetro de la transformación Box-Cox empleada (si el parámetro m en dicha transformación es distinto de cero, se resta m en las cuatro expresiones anteriores).

Criterios de evaluación de previsiones

2.7.9 Definición

Sea $\hat{y}_N(1), \hat{y}_N(2), \dots, \hat{y}_N(L)$ una secuencia de previsiones calculadas en origen N a horizontes $1, 2, \dots, L$. La **raíz del error cuadrático medio (RMSE, del inglés *Root Mean Squared Error*)** asociada con dicha secuencia es

$$RMSE \equiv \sqrt{\frac{1}{L} \sum_{l=1}^L [\hat{y}_N(l) - y_{N+l}]^2} \equiv \sqrt{\frac{1}{L} \sum_{l=1}^L \hat{e}_N(l)^2}, \quad (2.82)$$

es decir, la desviación típica muestral de los errores de previsión $\hat{e}_N(1)$, $\hat{e}_N(2)$, ..., $\hat{e}_N(L)$.

2.7.10 Definición

Sea $\hat{y}_N(1)$, $\hat{y}_N(2)$, ..., $\hat{y}_N(L)$ una secuencia de previsiones calculadas en origen N a horizontes $1, 2, \dots, L$. El **error absoluto medio** (**MAE**, del inglés *Mean Absolute Error*) asociado con dicha secuencia es

$$MAE \equiv \frac{1}{L} \sum_{l=1}^L |\hat{y}_N(l) - y_{N+l}| \equiv \frac{1}{L} \sum_{l=1}^L |\hat{e}_N(l)|, \quad (2.83)$$

es decir, la media de los valores absolutos de los errores de previsión.

2.7.11 Definición

Sea $\hat{y}_N(1)$, $\hat{y}_N(2)$, ..., $\hat{y}_N(L)$ una secuencia de previsiones calculadas en origen N a horizontes $1, 2, \dots, L$. El **error porcentual absoluto medio** (**MAPE**, del inglés *Mean Absolute Percentage Error*) asociado con dicha secuencia es

$$MAPE \equiv \frac{100}{L} \sum_{l=1}^L \left| \frac{\hat{y}_N(l) - y_{N+l}}{y_{N+l}} \right| \equiv \frac{100}{L} \sum_{l=1}^L \left| \frac{\hat{e}_N(l)}{y_{N+l}} \right|, \quad (2.84)$$

es decir, la media de los valores absolutos de los errores de previsión porcentuales.

2.7.12 Ejemplo

Modelo estimado con EViews 4.1 (May 23 2003 build) para el proceso Viviendas iniciadas en España, (Y_t) , con datos mensuales desde 1989:01 hasta 1999:12 (132 observaciones):

$$\nabla \nabla_{12} \ln y_t = (1 - 0.675B) (1 - 0.892B^{12}) + \hat{a}_t, \quad (2.85)$$

(0.070) (0.024)

$$n = 119, \quad \hat{\sigma}_A = 0.137.$$

Teniendo en cuenta que $\nabla \nabla_{12} \equiv (1 - B)(1 - B^{12}) = 1 - B - B^{12} + B^{13}$, (2.85) queda

$$y'_t = y'_{t-1} - y'_{t-12} + y'_{t-13} + \hat{a}_t - 0.675\hat{a}_{t-1} - 0.892\hat{a}_{t-12} + 0.602\hat{a}_{t-13}, \quad (2.86)$$

donde $y'_t \equiv \ln y_t$. A partir de (2.86), las previsiones puntuales del proceso original (Y_t) , calculadas en origen $n = 119$ (1999:12) a horizontes 1, 2, ..., 24, son:

$$\hat{y}_{119}(1) = \exp [y'_{119} - y'_{108} + y'_{107} - 0.675\hat{a}_{119} - 0.892\hat{a}_{108} + 0.602\hat{a}_{107}] = 38.455,$$

$$\hat{y}_{119}(2) = \exp [\hat{y}'_{119}(1) - y'_{109} + y'_{108} - 0.892\hat{a}_{109} + 0.602\hat{a}_{108}] = 36.995,$$

...

$$\hat{y}_{119}(24) = \exp [\hat{y}'_{119}(23) - \hat{y}'_{119}(12) + \hat{y}'_{119}(11)] = 43.975.$$

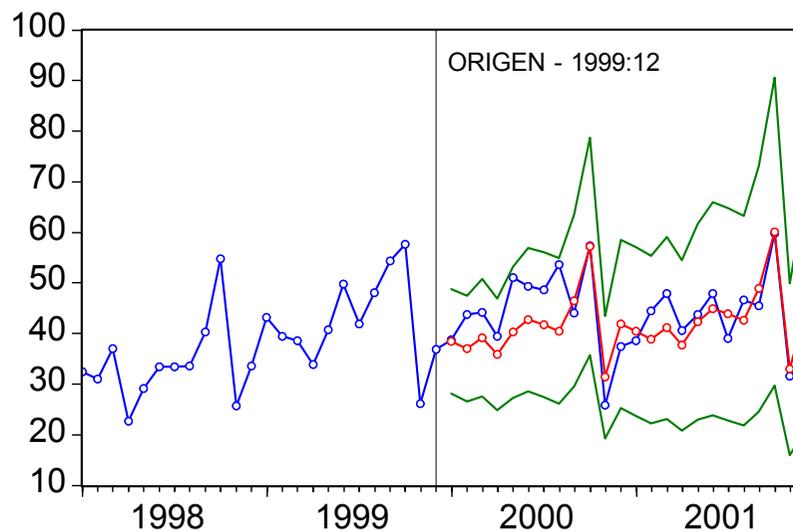


Figura 2.36

Previsiones del proceso Viviendas iniciadas en España calculadas a partir de (2.85)-(2.86)
(seasonal.wf1)

En este caso, $RMSE = 5.43$, $MAE = 4.47$, y $MAPE = 10.26\%$, que podrían compararse con las mismas medidas calculadas a partir de otros modelos para (Y_t) alternativos a (2.85).

2.7.13 Ejemplo

Modelo estimado con EViews 4.1 (May 23 2003 build) para el proceso PIB real en España, (Y_t) , con datos anuales desde 1964 hasta 1997 (34 observaciones):

$$\begin{aligned} (1 - 0.6747B) [\nabla \ln y_t - 0.0311] &= \hat{a}_t, \\ (0.1302) \qquad \qquad \qquad (0.0096) & \\ n = 32, \hat{\sigma}_A &= 0.0176. \end{aligned} \tag{2.87}$$

(2.87) también puede escribirse como

$$w_t = 0.0101 + 0.6747w_{t-1} + \hat{a}_t, \tag{2.88}$$

donde $w_t \equiv \nabla \ln y_t$, o bien como

$$y'_t = 0.0101 + 1.6747y'_{t-1} - 0.6747y'_{t-2} + \hat{a}_t, \tag{2.89}$$

donde $y'_t \equiv \ln y_t$, dependiendo de si se quiere prever la tasa logarítmica de variación anual

del PIB real, $(W_t) \equiv (\nabla \ln Y_t)$, o bien el PIB real directamente, (Y_t) .

(2.88) es un modelo AR(1) estimado para $(W_t) \equiv (\nabla \ln Y_t)$; por lo tanto (ver Ejemplo 2.7.6), $\hat{w}_{32}(1) = 0.0101 + 0.6747w_{32}$, $\hat{w}_{32}(2) = 0.0101 + 0.6747\hat{w}_{32}(1)$, ..., que converge a la media estimada de $(W_t) \equiv (\nabla \ln Y_t)$, $\hat{\mu}_W = 0.0311$.

Por su parte, (2.89) es un modelo ARI(1,1) estimado para $(Y'_t) \equiv (\ln Y_t)$; por lo tanto:

$$\hat{y}'_{32}(1) = \exp [0.0101 + 1.6747 y'_{32} - 0.6747 y'_{31}],$$

$$\hat{y}'_{32}(2) = \exp [0.0101 + 1.6747 \hat{y}'_{32}(1) - 0.6747 y'_{32}],$$

$$\hat{y}'_{32}(3) = \exp [0.0101 + 1.6747 \hat{y}'_{32}(2) - 0.6747 \hat{y}'_{32}(1)],$$

...

que es una secuencia creciente (no convergente) con el horizonte de previsión.

2.8 HETEROSCEDASTICIDAD CONDICIONAL AUTORREGRESIVA

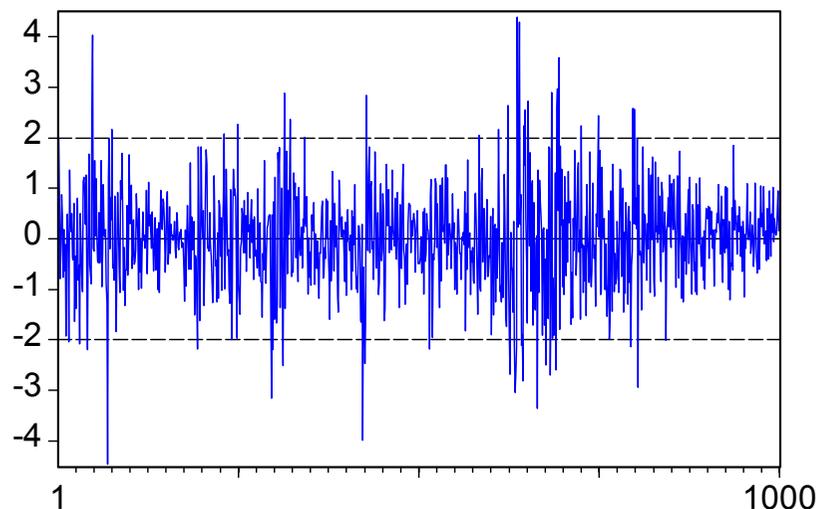


Figura 2.37

TLV diaria del NYSE Composite Index
7 de enero de 2000 – 31 de diciembre de 2003
(nyse-composite.wf1)

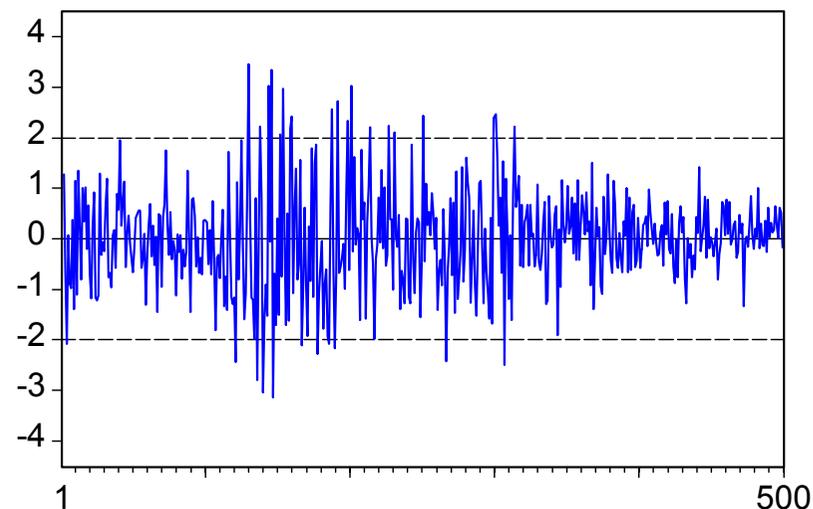


Figura 2.38

TLV diaria del IBEX 35
3 de enero de 2002 – 30 de diciembre de 2003
(bolsa-madrid.wf1)

Observación: Las dos series anteriores presentan un nivel constante. Sin embargo, las desviaciones grandes (pequeñas) de cada serie con respecto a su nivel van seguidas de variaciones grandes (pequeñas). Por lo tanto, la **volatilidad** de cada serie, **condicionada** por su historia reciente, presenta algún tipo de inercia.

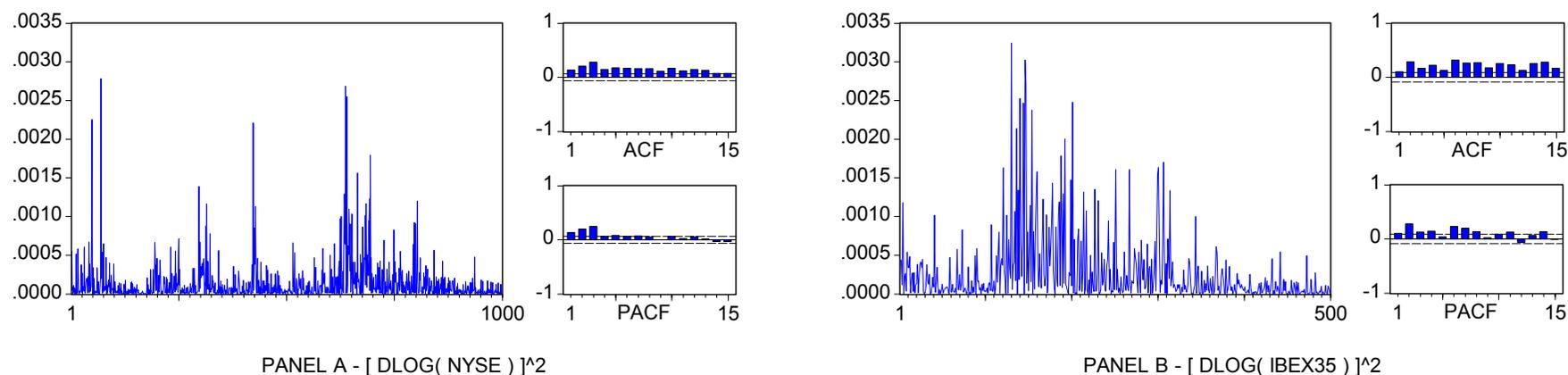


Figura 2.39

Rendimientos diarios al cuadrado del NYSE Composite Index (Panel A) y del IBEX 35 (Panel B)
(Panel A: nyse-composite.wf1. Panel B: bolsa-madrid.wf1)

Modelos ARCH – GARCH

2.8.1 Definición

Un proceso estocástico univariante estacionario (U_t) sigue un modelo **ARCH(1)** (del inglés *AutoRegressive Conditional Heteroskedasticity*) si y sólo si

$$U_t = \sigma_t A_t, \text{ con } \sigma_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2, \quad (2.90)$$

donde $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_1 < 1$, y $(A_t) \sim \text{IID}(0, 1)$ es independiente de (σ_t) .

Observación 1: Cuando $(U_t) \sim \text{ARCH}(1)$, puede comprobarse que $E[U_t] = E_{t-1}[U_t] = 0$, $E[U_t^2] = \alpha_0 / (1 - \alpha_1)$ (que está bien definida si se cumplen las condiciones anteriores sobre α_0 y α_1), $E[U_t U_{t+k}] = 0$ para todo $k \neq 0$, y $E_{t-1}[U_t^2] = \sigma_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2$. En particular, un modelo $\text{ARCH}(1)$ como (2.90) implica que los componentes del proceso (U_t) **no** están autocorrelacionados aunque **no** son estadísticamente independientes, ya que la **varianza condicional** de U_t es una función de U_{t-1} . Por otro lado, si se define $V_t \equiv U_t^2 - \sigma_t^2 \equiv \sigma_t^2 (A_t^2 - 1)$, (2.90) implica que $U_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2 + V_t$, que es un modelo $\text{AR}(1)$ para (U_t^2) con $E[V_t] = 0$ pero con $\text{Var}[V_t]$ no constante.

Observación 2: En general, un proceso univariante estacionario (U_t) sigue un modelo $\text{ARCH}(q)$ cuando $U_t = \sigma_t A_t$, con $\sigma_t^2 = E_{t-1}[U_t^2] = \alpha_0 + \sum_{i=1}^q \alpha_i U_{t-i}^2$, $\alpha_i > 0$ ($i = 0, 1, \dots, q$), $\sum_{i=1}^q \alpha_i < 1$, y $(A_t) \sim \text{IID}(0, 1)$ es independiente de (σ_t) . En este caso, la varianza incondicional de U_t es constante e igual a $\alpha_0 / (1 - \sum_{i=1}^q \alpha_i)$, que está bien definida si se cumplen las condiciones anteriores sobre los parámetros $\alpha_0, \alpha_1, \dots, \alpha_q$.

Observación 3: En la práctica, un modelo $\text{ARCH}(q)$ puede resultar poco satisfactorio cuando la volatilidad condicional de una serie (u_t) no es constante pero las autocorrelaciones muestrales de (u_t^2) no son muy elevadas (dos características frecuentes en muchas series financieras; ver Figura 2.39). Para acomodarse a estas pautas, el orden q de un modelo ARCH debe ser elevado, lo que dificulta el cálculo de estimaciones $\hat{\alpha}_i$ ($i = 0, 1, \dots, q$) que satisfagan las condiciones de la Observación 2. El modelo $\text{GARCH}(1,1)$ de la Definición 2.8.2 resuelve este inconveniente en muchas situaciones prácticas.

2.8.2 Definición

Un proceso estocástico univariante estacionario (U_t) sigue un modelo **GARCH(1,1)** (del

inglés *Generalized AutoRegressive Conditional Heteroskedasticity*) si y sólo si

$$U_t = \sigma_t A_t, \text{ con } \sigma_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2 + \delta_1 \sigma_{t-1}^2, \quad (2.91)$$

donde $\alpha_0 > 0$, $\alpha_1 > 0$, $\delta_1 > 0$, $\alpha_1 + \delta_1 < 1$ y $(A_t) \sim \text{IID}(0, 1)$ es independiente de (σ_t) .

Observación 1: Cuando $(U_t) \sim \text{GARCH}(1, 1)$, $E[U_t] = E_{t-1}[U_t] = 0$, $E[U_t^2] = \alpha_0 / (1 - \alpha_1 - \delta_1)$ (que está bien definida si se cumplen las condiciones anteriores sobre α_0 , α_1 y δ_1), $E[U_t U_{t+k}] = 0$ para todo $k \neq 0$, y $E_{t-1}[U_t^2] = \sigma_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2 + \delta_1 \sigma_{t-1}^2$. Un modelo $\text{GARCH}(1, 1)$ como (2.91) implica que los componentes del proceso (U_t) **no** están autocorrelacionados aunque **no** son estadísticamente independientes, ya que la **varianza condicional** de U_t es una función de U_{t-1} . Por otro lado, si se define $V_t \equiv U_t^2 - \sigma_t^2 \equiv \sigma_t^2 (A_t^2 - 1)$, (2.91) implica que $U_t^2 = \alpha_0 + (\alpha_1 + \delta_1) U_{t-1}^2 + V_t - \delta_1 V_{t-1}$, que es un modelo $\text{ARMA}(1, 1)$ para (U_t^2) con $E[V_t] = 0$ pero con $\text{Var}[V_t]$ no constante.

Observación 2: Aunque es fácil extender la definición anterior al caso de modelos $\text{GARCH}(p, q)$ del tipo $U_t = \sigma_t A_t$, con $\sigma_t^2 = E_{t-1}[U_t^2] = \alpha_0 + \sum_{i=1}^q \alpha_i U_{t-i}^2 + \sum_{i=1}^p \delta_i \sigma_{t-i}^2$, en la práctica el modelo $\text{GARCH}(1, 1)$ parece funcionar bastante bien en muchas ocasiones. En este sentido, se observa con frecuencia que $\hat{\alpha}_1$ es un número pequeño y positivo, que $\hat{\delta}_1$ es bastante mayor que $\hat{\alpha}_1$, y que $0.9 < \hat{\alpha}_1 + \hat{\delta}_1 < 1$, lo que sugiere (ver Observación 1) que (U_t^2) es un proceso altamente persistente; ver Figura 2.39.

Observación 3: Cuando (W_t) es un proceso estacionario, a veces los residuos de un modelo ARMA estimado para (W_t) presentan grupos (*clusters*) con alta dispersión y otros con baja dispersión; esto puede sugerir que las perturbaciones del modelo considerado no son ruido blanco, sino que siguen algún modelo $\text{GARCH}(p, q)$. En tal caso, un modelo razonable para (W_t) podría ser $\phi(B)W_t = \mu + \theta(B)U_t$, donde $(U_t) \sim \text{GARCH}(p, q)$. Como un modelo

GARCH(p, q) para la varianza condicional σ_t^2 de (U_t) puede interpretarse como un modelo ARMA para el proceso (U_t^2) (ver Observación 1 en 2.8.1 y en 2.8.2), la serie de los residuos al cuadrado $\hat{u}_1^2, \hat{u}_2^2, \dots, \hat{u}_N^2$ del modelo estimado $\hat{\phi}(B)w_t = \hat{\mu} + \hat{\theta}(B)\hat{u}_t$, debería presentar algún tipo de autocorrelación si $(U_t) \sim \text{GARCH}(p, q)$.

Contrastes de heteroscedasticidad condicional

Para decidir si un proceso estocástico estacionario (U_t) con $E[U_t] = 0$ presenta algún tipo de estructura ARCH o GARCH, puede utilizarse alguno de los procedimientos siguientes que hacen uso de una serie temporal u_1, u_2, \dots, u_N generada por (U_t) :

- ⇒ Examinar la ACF y la PACF muestrales de la serie $u_1^2, u_2^2, \dots, u_N^2$ (ver Figura 2.39) para identificar alguna pauta de autocorrelación reconocible.
- ⇒ Estimar una regresión auxiliar del tipo $u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2 + v_t$ y calcular el *p-value* $\Pr[\chi^2(q) > NR_*^2]$, donde R_*^2 es el coeficiente de determinación de la regresión anterior; alternativamente, calcular el *p-value* $\Pr[F(q, N - q - 1) > F_*]$, donde F_* es el valor del estadístico F para el contraste de significación global en la regresión anterior. En cualquier caso, si el *p-value* es suficientemente pequeño, rechazar la hipótesis nula de ausencia de estructura ARCH(q).

Observación: El primer procedimiento es más informativo que el segundo (que, además, requiere decidir de antemano un valor adecuado para q). Siempre es recomendable examinar la ACF y la PACF muestrales de (u_t^2) , además quizás de utilizar el segundo procedimiento con $q = 1$, ya que si la hipótesis nula de ausencia de estructura ARCH(q) se rechaza con $q = 1$, entonces existe seguramente algún tipo estructura ARCH en (U_t) , quizás de tipo GARCH(1,1).

2.8.3 Ejemplo

Modelo estimado con EViews 4.1 (May 23 2003 build) para el proceso NYSE Composite Index, (Y_t) , con datos diarios desde el 7 de enero de 2000 hasta el 31 de diciembre de 2003 (1000 observaciones):

$$\nabla \ln y_t = \hat{u}_t, \text{ donde } \hat{u}_t = \hat{\sigma}_t \hat{a}_t, \quad (2.92)$$

$$\hat{\sigma}_t^2 = 2.99\text{E-}06 + 0.0989\hat{u}_{t-1}^2 + 0.8802\hat{\sigma}_{t-1}^2, \quad (2.93)$$

(0.96E-06) (0.0154) (0.0168)

$$n = 1000, \hat{\sigma}_U = 0.0118.$$

Observación 1: La serie de residuos (Ordinary Residual Series) calculada por EViews es la serie (\hat{u}_t) de (2.92) [o, en general, la serie (\hat{u}_t) de un modelo ARIMA estimado para (Y_t) del tipo $\hat{\phi}(B)\nabla^d y_t' = \hat{\mu} + \hat{\theta}(B)\hat{u}_t$], mientras que la serie de residuos estandarizados (Standardized Residual Series) calculada por EViews es la serie $(\hat{a}_t) = (\hat{\sigma}_t^{-1}\hat{u}_t)$ de (2.92), donde $\hat{\sigma}_t$ es la desviación típica condicional estimada de las perturbaciones (U_t) , que se calcula como la raíz

cuadrada de (2.93); ver Figura (2.40). Por lo tanto, si la varianza condicional de (U_t) está bien modelizada, entonces la serie (\hat{u}_t) debe presentar estructura de tipo GARCH(1,1) [ver figuras 2.37 y 2.39 (Panel A)], mientras que la serie $(\hat{a}_t) = (\hat{\sigma}_t^{-1}\hat{u}_t)$ debe resultar compatible con la hipótesis de que $(A_t) \sim \text{IID}(0, 1)$; ver Figura 2.41.

Observación 2: El modelo (2.92)-(2.93) estimado para la rentabilidad diaria del NYSE Composite Index describe una característica habitual de muchas series de rentabilidades financieras: la **esperanza condicional** de la rentabilidad **no** depende de rentabilidades pasadas, pero la **varianza condicional** de la rentabilidad **sí**.

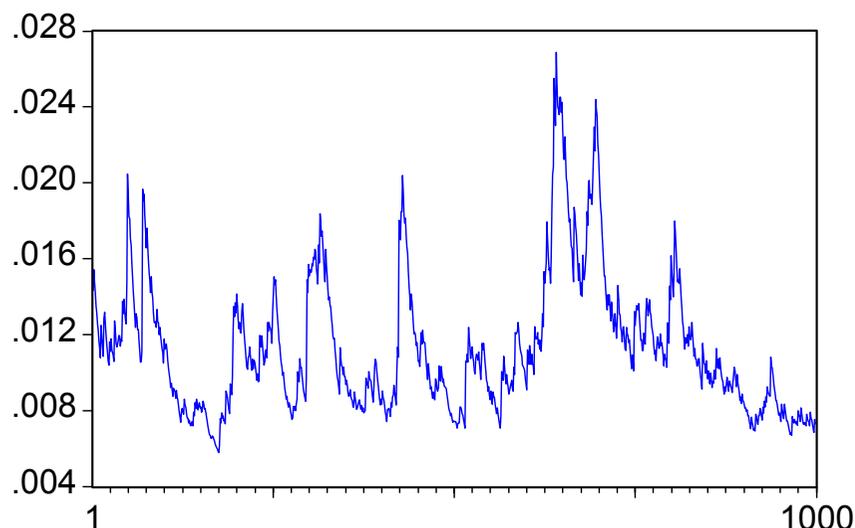


Figura 2.40

Desviación típica condicional estimada en el modelo
(2.92) – (2.93)

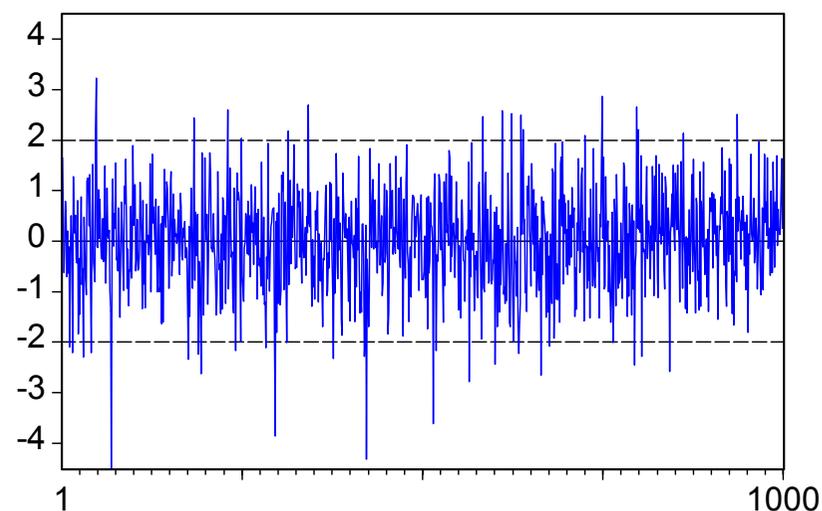


Figura 2.41

Residuos estandarizados del modelo
(2.92) – (2.93)

Previsión de la varianza condicional con un modelo GARCH(1,1)

(2.91) implica que $U_{t+l} = \sigma_{t+l}A_{t+l}$, con $\sigma_{t+l}^2 = \alpha_0 + \alpha_1 U_{t+l-1}^2 + \delta_1 \sigma_{t+l-1}^2$, de manera que cuando $l = 1$,

$$\sigma_t^2(1) \equiv \mathbb{E}_t[\sigma_{t+1}^2] = \alpha_0 + \alpha_1 U_t^2 + \delta_1 \sigma_t^2, \quad (2.94)$$

y cuando $l \geq 2$,

$$\sigma_t^2(l) \equiv \mathbb{E}_t[\sigma_{t+l}^2] = \alpha_0 + \alpha_1 \mathbb{E}_t[U_{t+l-1}^2] + \delta_1 \mathbb{E}_t[\sigma_{t+l-1}^2]. \quad (2.95)$$

Como $\mathbb{E}_t[U_{t+l-1}^2] = \mathbb{E}_t[\sigma_{t+l-1}^2 A_{t+l-1}^2] = \mathbb{E}_t[\sigma_{t+l-1}^2] \mathbb{E}_t[A_{t+l-1}^2] = \sigma_t^2(l-1)$, (2.95) queda

$$\sigma_t^2(l) = \alpha_0 + (\alpha_1 + \delta_1) \sigma_t^2(l-1),$$

que puede resolverse recursivamente a partir de la condición inicial $\sigma_t^2(1)$ en (2.94):

$$\sigma_t^2(l) = \left[\sum_{i=1}^{l-1} (\alpha_1 + \delta_1)^{i-1} \right] \alpha_0 + (\alpha_1 + \delta_1)^{l-1} \sigma_t^2(1) \quad (2.96)$$

para cualquier horizonte de previsión $l \geq 2$. Si $\alpha_1 + \delta_1 < 1$, $\sigma_t^2(l) \rightarrow \alpha_0 / (1 - \alpha_1 - \delta_1)$ cuando $l \rightarrow \infty$, que es la varianza incondicional del proceso (U_t) .

3

MODELOS MULTIVARIANTES ESTACIONARIOS

3.1 INTRODUCCIÓN

3.1.1 Definición

Un proceso estocástico multivariante (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, cuyos momentos de primer y segundo orden son finitos, es **estacionario en autocovarianza** si y sólo si

- $E[\mathbf{Y}_t]$ y $\text{Var}[\mathbf{Y}_t]$ son constantes (no dependen de t) para todo $t = 0, \pm 1, \pm 2, \dots$,
- $\text{Cov}[\mathbf{Y}_t, \mathbf{Y}_{t+k}]$ depende a lo sumo de k (entero) pero no de t para todo $t = 0, \pm 1, \pm 2, \dots$

Observación: El hecho de considerar los componentes $(Y_{t1}), (Y_{t2}) \dots, (Y_{tM})$ de (\mathbf{Y}_t) conjuntamente, en lugar de considerar cada uno de ellos por separado (como en el Tema 2), permite elaborar modelos estadísticos para (\mathbf{Y}_t) que describan no sólo la evolución temporal individual de cada (Y_{ti}) ($i = 1, 2, \dots, M$), sino también las posibles relaciones

contemporáneas y dinámicas entre ellos. Dos clases generales de modelos que resultan útiles en este contexto son los **modelos ARMA multivariantes** o **vectoriales** y los **modelos de función de transferencia-ruido**.

3.1.2 Ejemplo

Un proceso bivalente estacionario (\mathbf{Y}_t), donde $\mathbf{Y}_t = [Y_t, X_t]'$, sigue un modelo **VAR(1)** (autorregresivo vectorial de orden 1) bivalente cuando

$$\begin{aligned} Y_t &= \mu_1 + \phi_{11}Y_{t-1} + \phi_{12}X_{t-1} + A_{t1}, \\ X_t &= \mu_2 + \phi_{21}Y_{t-1} + \phi_{22}X_{t-1} + A_{t2}, \end{aligned} \quad (3.1)$$

o bien

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \mathbf{A}_t, \quad (3.2)$$

donde $\mathbf{A}_t \equiv [A_{t1}, A_{t2}]'$, $(\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_A)$ (ruido blanco vectorial),

$$\boldsymbol{\mu} \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Phi}_1 \equiv \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \boldsymbol{\Sigma}_A \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \quad (3.3)$$

y los ceros de $|\mathbf{I} - \boldsymbol{\Phi}_1 x|$ (los autovalores de $\boldsymbol{\Phi}_1$) están fuera (dentro) del círculo unitario.

Observación: Cuando $(\mathbf{Y}_t) \sim \text{VAR}(1)$, para describir las relaciones contemporáneas y dinámicas entre (Y_t) y (X_t) se puede considerar en primer lugar la función de autocorrelación (ACF) teórica de (\mathbf{Y}_t) implicada por (3.2)-(3.3). Por extensión del caso univariante (ver Tema 2), si $\boldsymbol{\mu}_Y \equiv E[\mathbf{Y}_t]$ y $\boldsymbol{\Gamma}(k) \equiv E[\tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}'_{t+k}] \equiv E[\tilde{\mathbf{Y}}_{t-k} \tilde{\mathbf{Y}}'_t]$ ($k = 0, 1, 2, \dots$) [la media y la función de autocovarianza teóricas de (\mathbf{Y}_t)], entonces (3.2) implica que

$$\boldsymbol{\mu}_Y = (\mathbf{I} - \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\mu}, \quad \tilde{\mathbf{Y}}_t = \boldsymbol{\Phi}_1 \tilde{\mathbf{Y}}_{t-1} + \mathbf{A}_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \mathbf{A}_{t-i} \quad (\text{con } \boldsymbol{\Psi}_i \equiv \boldsymbol{\Phi}_1^i),$$

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}(0) \boldsymbol{\Phi}_1' + \boldsymbol{\Sigma}_A, \quad \boldsymbol{\Gamma}(k) = \boldsymbol{\Gamma}(k-1) \boldsymbol{\Phi}_1' = \boldsymbol{\Gamma}(0) [\boldsymbol{\Phi}_1']^k \quad \text{para todo } k \geq 1,$$

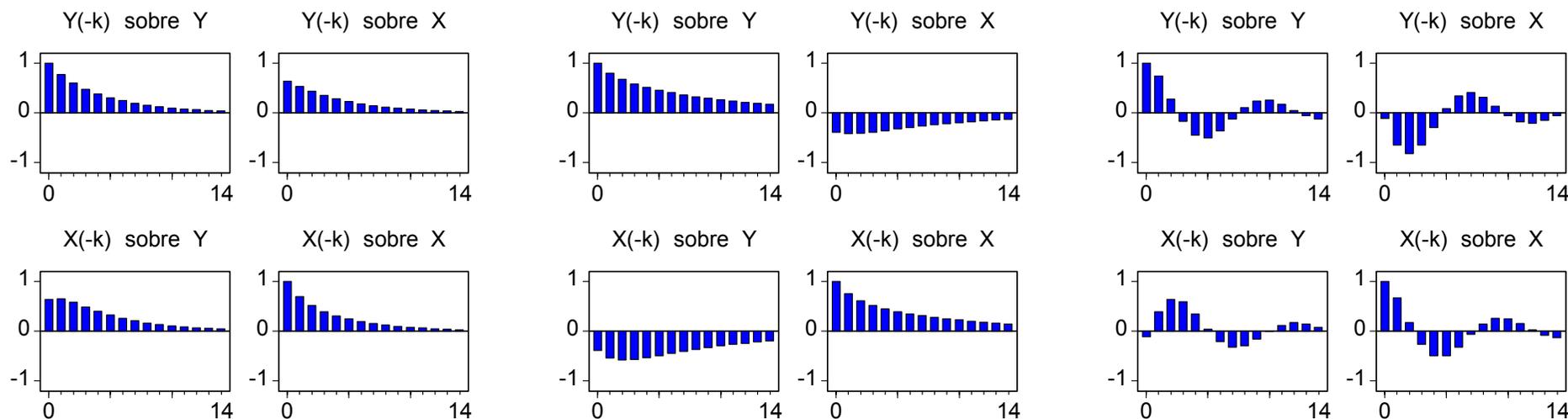
donde

$$\boldsymbol{\mu}_Y \equiv \begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix} \equiv \begin{bmatrix} E[Y_t] \\ E[X_t] \end{bmatrix}, \quad \boldsymbol{\Gamma}(k) \equiv \begin{bmatrix} \gamma_{YY}(k) & \gamma_{YX}(k) \\ \gamma_{XY}(k) & \gamma_{XX}(k) \end{bmatrix} \equiv \begin{bmatrix} \text{Cov}[Y_{t-k}, Y_t] & \text{Cov}[Y_{t-k}, X_t] \\ \text{Cov}[X_{t-k}, Y_t] & \text{Cov}[X_{t-k}, X_t] \end{bmatrix} \quad (k = 0, 1, 2, \dots).$$

[Nótese que $\boldsymbol{\Gamma}(k) = \boldsymbol{\Gamma}(-k)'$.] Por lo tanto, la ACF de (\mathbf{Y}_t) , $\mathbf{P}(k) \equiv \mathbf{D}^{-1/2} \boldsymbol{\Gamma}(k) \mathbf{D}^{-1/2}$ ($k = 0, 1, 2, \dots$), es en este caso $\mathbf{P}(k) = \mathbf{P}(0) (\mathbf{D}^{1/2} [\boldsymbol{\Phi}_1']^k \mathbf{D}^{-1/2})$, donde

$$\mathbf{D} \equiv \begin{bmatrix} \gamma_{YY}(0) & 0 \\ 0 & \gamma_{XX}(0) \end{bmatrix}, \quad \mathbf{P}(k) \equiv \begin{bmatrix} \rho_{YY}(k) & \rho_{YX}(k) \\ \rho_{XY}(k) & \rho_{XX}(k) \end{bmatrix} \equiv \begin{bmatrix} \text{Corr}[Y_{t-k}, Y_t] & \text{Corr}[Y_{t-k}, X_t] \\ \text{Corr}[X_{t-k}, Y_t] & \text{Corr}[X_{t-k}, X_t] \end{bmatrix} \quad (k = 0, 1, 2, \dots).$$

En $\mathbf{P}(k)$, $\rho_{YY}(k)$ y $\rho_{XX}(k)$ describen las **autocorrelaciones** de (Y_t) y de (X_t) , respectivamente, mientras que $\rho_{YX}(k)$ y $\rho_{XY}(k)$ describen **correlaciones cruzadas** del tipo $(Y_{t-k}) \rightarrow (X_t)$ y $(X_{t-k}) \rightarrow (Y_t)$, respectivamente. Así, los dos componentes de (\mathbf{Y}_t) reciben un tratamiento simétrico en un modelo VAR(1), lo que permite considerar la existencia de una relación de **causalidad bidireccional** entre ellos. [Nótese que la información contenida en $\rho_{YX}(k)$ y en $\rho_{XY}(k)$ para $k = 0, 1, 2, \dots$, también está contenida en $\rho_{YX}(k)$ o bien en $\rho_{XY}(k)$ para $k = 0, \pm 1, \pm 2, \dots$, cada una de las cuales se denomina una **función de correlación cruzada** (CCF, del inglés *Cross Correlation Function*).]



Los tres diagramas anteriores representan la ACF de $(\mathbf{Y}_t) \sim \text{VAR}(1)$ para diferentes valores de Φ_1 en (3.2)-(3.3):

$$\Phi_1 = \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} 0.7 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{bmatrix}; \quad \Sigma_A = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix} \text{ en los tres casos.}$$

3.1.3 Definición

Un proceso multivariante estacionario (\mathbf{Y}_t) sigue un **modelo ARMA multivariante** o **vectorial** de orden (p,q) , o $\text{VARMA}(p,q)$ (autorregresivo-media móvil vectorial, del inglés *Vector Autoregressive-Moving Average*), si y sólo si

$$\begin{aligned} \mathbf{Y}_t = & \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{Y}_{t-2} + \dots + \boldsymbol{\Phi}_p \mathbf{Y}_{t-p} + \mathbf{A}_t \\ & - \boldsymbol{\Theta}_1 \mathbf{A}_{t-1} - \boldsymbol{\Theta}_2 \mathbf{A}_{t-2} - \dots - \boldsymbol{\Theta}_q \mathbf{A}_{t-q} \end{aligned} \quad (3.4)$$

para todo $t = 0, \pm 1, \pm 2, \dots$, donde $(\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_A)$ y $\boldsymbol{\mu}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_p, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \dots, \boldsymbol{\Theta}_q$ son matrices de parámetros tales que todas las raíces de la ecuación

$$\left| \mathbf{I} - \boldsymbol{\Phi}_1 x - \boldsymbol{\Phi}_2 x^2 - \dots - \boldsymbol{\Phi}_p x^p \right| = 0$$

están fuera del círculo unitario (**condición de estacionariedad**). Un modelo VARMA(p, q) descrito por (3.4) es **invertible** si todas las raíces de la ecuación polinomial

$$\left| \mathbf{I} - \boldsymbol{\Theta}_1 x - \boldsymbol{\Theta}_2 x^2 - \dots - \boldsymbol{\Theta}_q x^q \right| = 0$$

están fuera del círculo unitario (**condición de invertibilidad**).

Observación 1: (3.4) puede escribirse alternativamente como $\boldsymbol{\Phi}(B)\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Theta}(B)\mathbf{A}_t$, donde $\boldsymbol{\Phi}(B) \equiv \mathbf{I} - \sum_{i=1}^p \boldsymbol{\Phi}_i B^i$ es el operador autorregresivo (AR) del modelo, y $\boldsymbol{\Theta}(B) \equiv \mathbf{I} - \sum_{i=1}^q \boldsymbol{\Theta}_i B^i$ es el operador media móvil (MA) del modelo. (3.4) se denomina un modelo VARMA **estándar** o normalizado, en el sentido que $\boldsymbol{\Phi}(0) = \boldsymbol{\Theta}(0) = \mathbf{I}$.

Observación 2: Dado que $\boldsymbol{\Sigma}_A \equiv \text{Var}[\mathbf{A}_t]$ es una matriz definida positiva, $\boldsymbol{\Sigma}_A$ puede factorizarse como $\boldsymbol{\Sigma}_A = \mathbf{T}\boldsymbol{\Sigma}^*\mathbf{T}'$, donde $\boldsymbol{\Sigma}^*$ es una matriz diagonal definida positiva y \mathbf{T} es una matriz triangular inferior con unos en su diagonal

principal (**factorización de Cholesky**). La matriz \mathbf{T} puede emplearse para transformar $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \Theta(B)\mathbf{A}_t$ en un modelo con **perturbaciones ortogonales** del tipo $\Phi^*(B)\mathbf{Y}_t = \boldsymbol{\mu}^* + \Theta^*(B)\mathbf{A}_t^*$, con $\Phi^*(B) \equiv \mathbf{T}^{-1}\Phi(B)$, $\boldsymbol{\mu}^* \equiv \mathbf{T}^{-1}\boldsymbol{\mu}$, $\Theta^*(B) \equiv \mathbf{T}^{-1}\Theta(B)\mathbf{T}$, $\mathbf{A}_t^* \equiv \mathbf{T}^{-1}\mathbf{A}_t$, $\Phi^*(0) = \mathbf{T}^{-1}$ (triangular inferior), $\text{Var}[\mathbf{A}_t^*] = \mathbf{T}^{-1}\boldsymbol{\Sigma}_A\mathbf{T}^{-1'} = \boldsymbol{\Sigma}^*$ (diagonal), y $\Theta^*(0) = \mathbf{I}$. Por lo tanto, $\Phi^*(0)$ muestra explícitamente las relaciones contemporáneas entre los componentes del proceso (\mathbf{Y}_t), mientras que en la representación estándar (3.4) la información sobre dichas relaciones está recogida implícitamente en $\boldsymbol{\Sigma}_A$ (una matriz no diagonal en general). La representación no estándar anterior de un modelo VARMA para (\mathbf{Y}_t) puede depender crucialmente del **orden** en el que figuran los componentes de \mathbf{Y}_t .

Observación 3: En el modelo VAR(1) bivalente (ver Ejemplo 3.1.2)

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} - \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}, \quad \boldsymbol{\Sigma}_A = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}, \quad (\text{I})$$

la factorización de Cholesky de $\boldsymbol{\Sigma}_A$ es

$$\boldsymbol{\Sigma}_A = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}, \text{ de manera que } \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix}. \quad (\text{II})$$

Premultiplicando (I) por la matriz \mathbf{T}^{-1} de (II),

$$\begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} - \begin{bmatrix} 0.6 & 0.2 \\ -0.1 & 0.5 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} A_{t1}^* \\ A_{t2}^* \end{bmatrix}, \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}, \quad (\text{II})$$

es decir, $Y_t = 0.6Y_{t-1} + 0.2X_{t-1} + A_{t1}^*$, $X_t = 0.5X_{t-1} + 0.5Y_t - 0.1Y_{t-1} + A_{t2}^*$, donde (A_{t1}^*) y (A_{t2}^*) son dos procesos de ruido blanco **ortogonales** o **incorrelacionados** (independientes si son Normales o Gaussianos). (II) indica

que $(Y_t) \rightarrow (X_t)$ contemporáneamente, pero $(X_t) \rightarrow (Y_t)$ no. Por otro lado, el modelo estándar (I) es idéntico, excepto por el orden en el que figuran Y_t y X_t , a

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} - \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} A_{t2} \\ A_{t1} \end{bmatrix}, \quad \Sigma_{\mathbf{A}} = \begin{bmatrix} 10 & 2 \\ 2 & 4 \end{bmatrix}. \quad (\text{IV})$$

Ahora, la factorización de Cholesky de $\Sigma_{\mathbf{A}}$ en (IV) es

$$\Sigma_{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 3.6 \end{bmatrix} \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}, \quad \text{de manera que ahora } \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 \\ -0.2 & 1 \end{bmatrix}. \quad (\text{V})$$

Premultiplicando (IV) por la matriz \mathbf{T}^{-1} de (V),

$$\begin{bmatrix} 1 & 0 \\ -0.2 & 1 \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} - \begin{bmatrix} 0.6 & 0.2 \\ 0.08 & 0.56 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} A_{t2}^* \\ A_{t1}^* \end{bmatrix}, \quad \Sigma^* = \begin{bmatrix} 10 & 0 \\ 0 & 3.6 \end{bmatrix},$$

o bien

$$\begin{bmatrix} 1 & -0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} - \begin{bmatrix} 0.56 & 0.08 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} A_{t1}^* \\ A_{t2}^* \end{bmatrix}, \quad \Sigma^* = \begin{bmatrix} 3.6 & 0 \\ 0 & 10 \end{bmatrix}, \quad (\text{VI})$$

es decir, $Y_t = 0.56Y_{t-1} + 0.2X_t + 0.08X_{t-1} + A_{t1}^*$, $X_t = 0.6X_{t-1} + 0.2Y_{t-1} + A_{t2}^*$, donde (A_{t1}^*) y (A_{t2}^*) son dos procesos de ruido blanco ortogonales distintos de los que figuran en (II). (VI) indica que ahora $(X_t) \rightarrow (Y_t)$ contemporáneamente, pero $(Y_t) \rightarrow (X_t)$ no, exactamente lo contrario que ocurre en (II).

Observación 4: Dado que $\Theta^{-1}(B) = \text{Adj}[\Theta(B)] / |\Theta(B)|$, $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \Theta(B)\mathbf{A}_t$ también puede escribirse como $\text{Adj}[\Theta(B)]\Phi(B)\mathbf{Y}_t = \text{Adj}[\Theta(1)]\boldsymbol{\mu} + |\Theta(B)|\mathbf{A}_t$. A partir de esta expresión, puede obtenerse una representación de un

modelo ARMA M -variante en términos de M **funciones de transferencia**; ver Ejemplo 3.1.4 y Definición 3.1.5. En (II) de la Observación 3 anterior, $(1 - 0.6B)Y_t = 0.2X_{t-1} + A_{t1}^*$, $(1 - 0.5B)X_t = (0.5 - 0.1B)Y_t + A_{t2}^*$, que **no** coinciden con $(1 - 0.56B)Y_t = (0.2 + 0.08B)X_t + A_{t1}^*$, $(1 - 0.6B)X_t = 0.2Y_{t-1} + A_{t2}^*$ en (VI).

3.1.4 Ejemplo

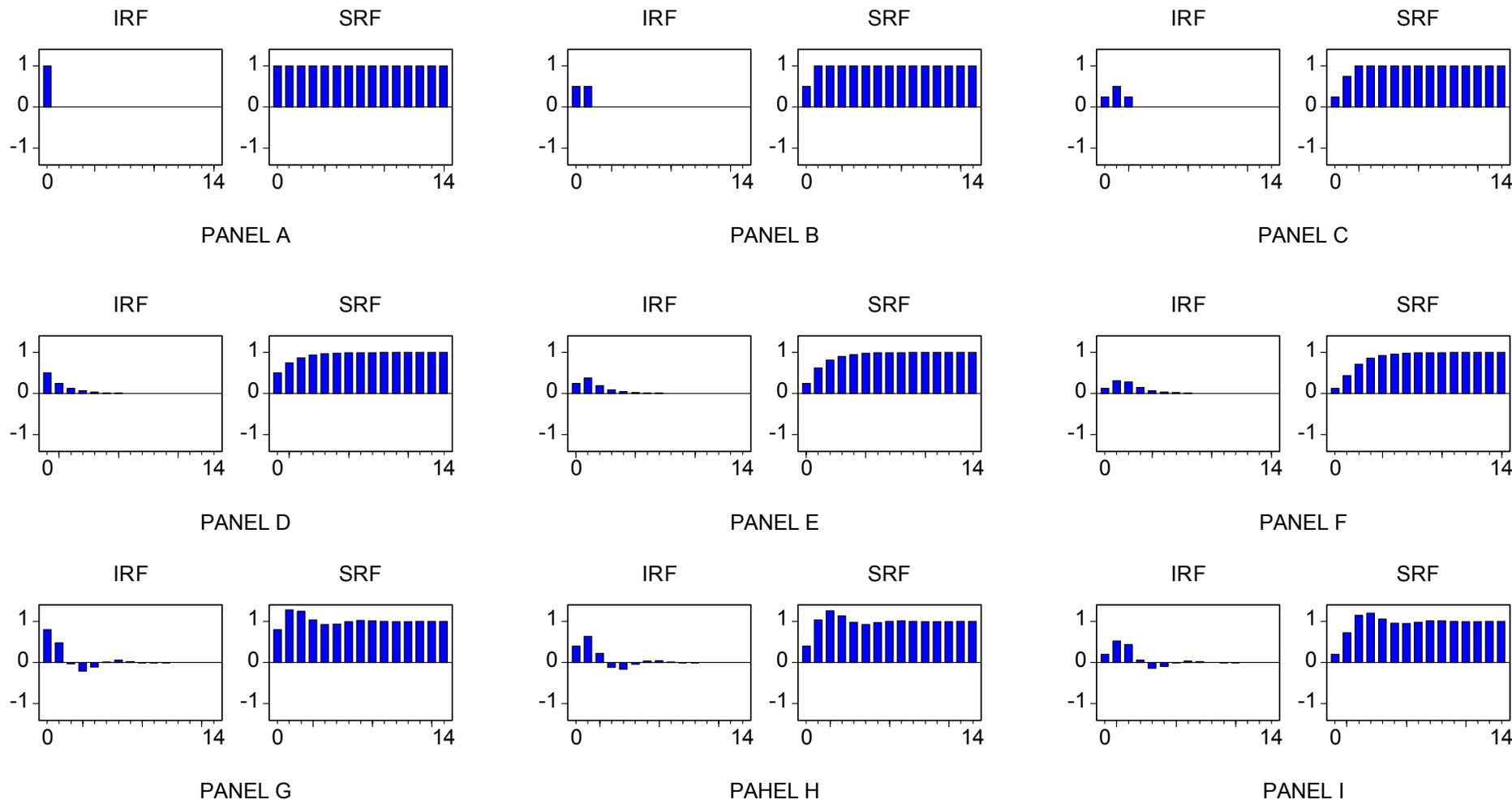
Un proceso bivalente estacionario (\mathbf{Y}_t) , donde $\mathbf{Y}_t = [Y_t, X_t]'$, sigue un **modelo de función de transferencia-ruido** de orden (s, b, r) cuando

$$Y_t = \alpha_0 + \frac{\omega(B)}{\delta(B)} B^b X_t + N_t, \quad (3.5)$$

con $\omega(B) \equiv \omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_s B^s$, $\delta(B) \equiv 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r$, las raíces de la ecuación $\delta(x) = 0$ están **fuera** del círculo unitario, y $(N_t) \sim I(0)$ es un **proceso de ruido** (**no** necesariamente ruido blanco) con media cero, **independiente** de (X_t) .

Observación 1: (3.5) representa una relación de **causalidad unidireccional** del tipo $(X_{t-k}) \rightarrow (Y_t)$ ($k \geq 0$), a través de la **función de transferencia** $v(B) \equiv [\omega(B) / \delta(B)] B^b \equiv \sum_{k=0}^{\infty} v_k B^k$ del **input** (X_t) sobre el **output** (Y_t) . La secuencia (v_0, v_1, v_2, \dots) de coeficientes de $v(B) \equiv v_0 + v_1 B + v_2 B^2 + \dots$ se denomina la **función de respuesta al impulso** (IRF, del inglés *Impulse Response Function*) del output con respecto al input del modelo, en el sentido que $v_k = \partial Y_t / \partial X_{t-k} = \partial Y_{t+k} / \partial X_t$ ($k \geq 0$). (3.5) puede resultar útil, en particular, para prever (Y_t) cuando (X_t) es un **indicador adelantado** (en inglés *leading indicator*) de (Y_t) , es decir, cuando el **tiempo muerto** $b \geq 1$.

Observación 2: La figura siguiente contiene la IRF asociada con varias funciones de transferencia, así como la **función de respuesta al escalón** (SRF, del inglés *Step Response Function*), $V_k \equiv \sum_{i=0}^k v_i$ ($k \geq 0$), correspondiente. En todos los casos, la **ganancia a largo plazo** (en inglés *steady-state gain*), $g \equiv \sum_{k=0}^{\infty} v_k$, es igual a 1.



En la figura anterior se han utilizado las funciones de transferencia siguientes:

$$\text{Panel A : } v(B) = 1. \quad \text{Panel B : } v(B) = 0.5 + 0.5B. \quad \text{Panel C : } v(B) = 0.25 + 0.5B + 0.25B^2.$$

$$\text{Panel D : } v(B) = \frac{0.5}{1 - 0.5B}. \quad \text{Panel E : } v(B) = \frac{0.25 + 0.25B}{1 - 0.5B}. \quad \text{Panel F : } v(B) = \frac{0.125 + 0.25B + 0.125B^2}{1 - 0.5B}.$$

$$\text{Panel G : } v(B) = \frac{0.8}{1 - 0.6B + 0.4B^2}. \quad \text{Panel H : } v(B) = \frac{0.4 + 0.4B}{1 - 0.6B + 0.4B^2}. \quad \text{Panel I : } v(B) = \frac{0.2 + 0.4B + 0.2B^2}{1 - 0.6B + 0.4B^2}.$$

En todos los casos, $b = 0$. Para valores del tiempo muerto $b \geq 1$, todas las funciones de la figura anterior estarían desplazadas b retardos hacia la derecha.

Observación 3: Si $\phi(B)N_t = \theta(B)A_{t1}^*$ y $\phi_X(B)X_t = \mu_0 + \theta_X(B)A_{t2}^*$, donde (A_{t1}^*) y (A_{t2}^*) son dos procesos de ruido blanco **independientes**, entonces (3.5) puede escribirse como

$$\begin{bmatrix} \delta(B)\phi(B) & -\omega(B)\phi(B)B^b \\ 0 & \phi_X(B) \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix} + \begin{bmatrix} \delta(B)\theta(B) & 0 \\ 0 & \theta_X(B) \end{bmatrix} \begin{bmatrix} A_{t1}^* \\ A_{t2}^* \end{bmatrix}, \quad \Sigma^* = \begin{bmatrix} \sigma_{1*}^2 & 0 \\ 0 & \sigma_{2*}^2 \end{bmatrix}, \quad (\text{I})$$

donde $\mu_1^* \equiv \delta(1)\phi(1)\alpha_0$, $\mu_2^* \equiv \mu_0$, o bien como

$$\begin{bmatrix} \Phi_{11}^*(B) & \Phi_{12}^*(B) \\ 0 & \Phi_{22}^*(B) \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix} + \begin{bmatrix} \Theta_{11}^*(B) & 0 \\ 0 & \Theta_{22}^*(B) \end{bmatrix} \begin{bmatrix} A_{t1}^* \\ A_{t2}^* \end{bmatrix}, \quad \Sigma^* = \begin{bmatrix} \sigma_{1*}^2 & 0 \\ 0 & \sigma_{2*}^2 \end{bmatrix}.$$

Si el tiempo muerto $b \geq 1$ en (3.5), entonces la expresión anterior es un modelo VARMA estándar para el proceso bivalente (\mathbf{Y}_t) . Por el contrario, si $b = 0$, $\Phi_{12}^*(0) = -\omega_0$, de manera que

$$\Phi^*(0) \equiv \begin{bmatrix} \Phi_{11}^*(0) & \Phi_{12}^*(0) \\ 0 & \Phi_{22}^*(0) \end{bmatrix} = \begin{bmatrix} 1 & -\omega_0 \\ 0 & 1 \end{bmatrix} \Rightarrow \mathbf{T} \equiv [\Phi^*(0)]^{-1} = \begin{bmatrix} 1 & \omega_0 \\ 0 & 1 \end{bmatrix}. \quad (\text{II})$$

Si (I) se escribe como $\Phi^*(B)\mathbf{Y}_t = \boldsymbol{\mu}^* + \Theta^*(B)\mathbf{A}_t^*$, entonces la representación VARMA estándar o normalizada de (I) es $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \Theta(B)\mathbf{A}_t$, donde $\Phi(B) \equiv \mathbf{T}\Phi^*(B)$, $\boldsymbol{\mu} \equiv \mathbf{T}\boldsymbol{\mu}^*$, $\Theta(B) \equiv \mathbf{T}\Theta^*(B)\mathbf{T}^{-1}$, $\mathbf{A}_t \equiv \mathbf{T}\mathbf{A}_t^*$; por lo tanto, $\Phi(0) = \Theta(0) = \mathbf{I}$, y $\text{Var}[\mathbf{A}_t] = \mathbf{T}\boldsymbol{\Sigma}^*\mathbf{T}' = \boldsymbol{\Sigma}_\mathbf{A}$ (una matriz **no** diagonal). Explícitamente:

$$\Phi(B) = \begin{bmatrix} \delta(B)\phi(B) & \omega_0\phi_X(B) - \omega(B)\phi(B) \\ 0 & \phi_X(B) \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1^* + \omega_0\mu_2^* \\ \mu_2^* \end{bmatrix},$$

$$\Theta(B) = \begin{bmatrix} \delta(B)\theta(B) & \omega_0[\theta_X(B) - \delta(B)\theta(B)] \\ 0 & \theta_X(B) \end{bmatrix}, \quad \boldsymbol{\Sigma}_\mathbf{A} = \begin{bmatrix} \sigma_{1*}^2 + \omega_0^2\sigma_{2*}^2 & \omega_0\sigma_{2*}^2 \\ \omega_0\sigma_{2*}^2 & \sigma_{2*}^2 \end{bmatrix}.$$

Por ejemplo, un modelo del tipo

$$Y_t = \alpha_0 + \frac{\omega_0}{1 - \delta_1 B} B^b X_t + N_t, \quad \text{con } N_t = (1 - \theta_1 B)A_{t1}^* \text{ y } (1 - \phi_1 B)X_t = \mu_0 + A_{t2}^*,$$

puede escribirse como (I),

$$\begin{bmatrix} 1 - \delta_1 B & -\omega_0 B^b \\ 0 & 1 - \phi_1 B \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} (1 - \delta_1)\alpha_0 \\ \mu_0 \end{bmatrix} + \begin{bmatrix} (1 - \delta_1 B)(1 - \theta_1 B) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A_{t1}^* \\ A_{t2}^* \end{bmatrix}, \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} \sigma_{1*}^2 & 0 \\ 0 & \sigma_{2*}^2 \end{bmatrix},$$

que es un modelo VARMA($b,2$) estándar si $b \geq 1$, pero es un modelo VARMA(1,2) no estándar si $b = 0$. En este caso, la representación estándar o normalizada correspondiente es $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \Theta(B)\mathbf{A}_t$, donde

$$\Phi(B) = \begin{bmatrix} 1 - \delta_1 B & -\omega_0 \phi_1 B \\ 0 & 1 - \phi_1 B \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} (1 - \delta_1)\alpha_0 + \omega_0 \mu_0 \\ \mu_0 \end{bmatrix},$$

$$\Theta(B) = \begin{bmatrix} 1 - (\delta_1 + \theta_1)B + \delta_1 \theta_1 B^2 & \omega_0 [(\delta_1 + \theta_1)B - \delta_1 \theta_1 B^2] \\ 0 & 1 \end{bmatrix}, \quad \Sigma_{\mathbf{A}} = \begin{bmatrix} \sigma_{1*}^2 + \omega_0^2 \sigma_{2*}^2 & \omega_0 \sigma_{2*}^2 \\ \omega_0 \sigma_{2*}^2 & \sigma_{2*}^2 \end{bmatrix}.$$

3.1.5 Definición

Un proceso multivariante estacionario (\mathbf{Y}_t), donde $\mathbf{Y}_t = [Y_t, X_{t1}, X_{t2}, \dots, X_{tM}]'$, sigue un **modelo de función de transferencia-ruido**, o TFN (del inglés *Transfer Function-Noise*), con **inputs múltiples** si y sólo si

$$Y_t = \alpha_0 + \frac{\omega_1(B)}{\delta_1(B)} B^{b_1} X_{t1} + \frac{\omega_2(B)}{\delta_2(B)} B^{b_2} X_{t2} + \dots + \frac{\omega_M(B)}{\delta_M(B)} B^{b_M} X_{tM} + N_t, \quad (3.6)$$

con $\omega_i(B) \equiv \omega_{i0} - \omega_{i1}B - \omega_{i2}B^2 - \dots - \omega_{is_i}B^{s_i}$, $\delta_i(B) \equiv 1 - \delta_{i1}B - \delta_{i2}B^2 - \dots - \delta_{ir_i}B^{r_i}$, las raíces de $\delta_i(x) = 0$ están **fuera** del círculo unitario ($i = 1, 2, \dots, M$), y $(N_t) \sim I(0)$ es un **proceso de ruido** (no necesariamente ruido blanco) con media cero, **independiente** de (\mathbf{X}_t) , $\mathbf{X}_t = [X_{t1}, X_{t2}, \dots, X_{tM}]'$, donde (\mathbf{X}_t) sigue un modelo VARMA estacionario e invertible de dimensión M (ver Definición 3.1.3).

Observación 1: En general, (3.6) implica que $E[Y_t] = \alpha_0 + \sum_{i=1}^M g_i E[X_{ti}]$, donde $g_i \equiv \omega_i(1) / \delta_i(1)$ ($i = 1, 2, \dots, M$) es la ganancia a largo plazo del output con respecto a cada input. Por ejemplo, en el modelo de la última parte de la Observación 3 de 3.1.4, $E[Y_t] = \alpha_0 + [\omega_0 / (1 - \delta_1)] E[X_t]$, con $E[X_t] = \mu_0 / (1 - \phi_1)$. Esta cuestión es importante para facilitar la interpretación de los modelos de las secciones 3.2 y 3.3, todos los cuales son casos particulares del modelo TFN (3.6).

Observación 2: Prácticamente todos los llamados **modelos de regresión** para procesos estacionarios que figuran en los manuales de Econometría, son casos particulares (generalmente arbitrarios y restringidos a priori) del modelo (3.6). En la actualidad, la popularidad de dichos modelos es mucho mayor entre los economistas que la de los modelos TFN; de hecho, es muy difícil encontrar manuales de Econometría o de Análisis de Series Temporales aplicado a la Economía en los que se mencione siquiera la posibilidad de modelizar relaciones de causalidad unidireccionales entre procesos estacionarios a través de un modelo TFN. Esta posibilidad tiene ventajas muy claras sobre la modelización basada exclusivamente en modelos de regresión, aunque sus requisitos computacionales son generalmente más elevados. En todo caso, es una posibilidad que no se contempla en la mayoría de los programas informáticos más populares disponibles actualmente (como EViews) para el análisis de series temporales; ver Box, Jenkins y Reinsel (1994) capítulos 10-12, Brockwell y Davis (2002) secciones 10.1-10.2, y Mills (1990) capítulos 12-13. La situación actual de los modelos VARMA en el Análisis de Series Temporales aplicado a la Economía es muy semejante; ver Brockwell y Davis (2002) capítulo 7, Mills (1990) capítulo 14, y Reinsel (1997) capítulos 1-2 y 4-6.

3.2 MODELOS DE REGRESIÓN CON PERTURBACIONES ARMA

3.2.1 Definición

Un modelo de regresión con perturbaciones ARMA para un proceso estacionario (\mathbf{Y}_t) , donde $\mathbf{Y}_t = [Y_t, X_{t1}, \dots, X_{tM}]'$, es un caso particular de (3.6) con $s_i = b_i = r_i = 0$ para todo $i = 1, 2, \dots, M$. En este caso, (3.6) queda

$$Y_t = \alpha_0 + \omega_{10}X_{t1} + \dots + \omega_{M0}X_{tM} + N_t, \quad (3.7)$$

donde $(N_t) \sim I(0)$ tiene media cero, es **independiente** de (\mathbf{X}_t) , $\mathbf{X}_t = [X_{t1}, \dots, X_{tM}]'$, y sigue un modelo ARMA(p, q) estacionario e invertible del tipo $\phi(B)N_t = \theta(B)A_t$.

Observación 1: El modelo (3.7) puede escribirse como $Y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{X}_t + N_t$, $\boldsymbol{\beta}' \equiv [\beta_1, \dots, \beta_M]$, que representa una relación de causalidad unidireccional **contemporánea** del tipo $(\mathbf{X}_t) \rightarrow (Y_t)$ con $E[Y_t] = \beta_0 + \boldsymbol{\beta}'E[\mathbf{X}_t]$ (constante). (3.7) se puede interpretar como un modelo ARMA para el proceso $(N_t) \equiv (Y_t - \beta_0 - \boldsymbol{\beta}'\mathbf{X}_t) \sim I(0)$, donde (Y_t) , (X_{t1}) , ..., (X_{tM}) son estacionarios. En Econometría, un modelo como (3.7) se denomina a veces un modelo **estático** de regresión con perturbaciones autocorrelacionadas.

Observación 2: Aunque (3.7) se ha planteado en 3.2.1 como un modelo para un proceso (\mathbf{Y}_t) multivariante estacionario, (3.7) también puede ser un modelo adecuado para $(\mathbf{Y}_t) \sim I(1)$ cuando (Y_t) , (X_{t1}) , ..., (X_{tM}) están **cointegrados**; ver Tema 4.

Análisis de intervención

Sea (Y_t) un proceso estocástico tal que $Y_t = \eta_t + N_t$, donde (η_t) representa el efecto sobre el nivel de (Y_t) derivado de un **suceso especial** o de **intervención** (quizás fácilmente identificable) ocurrido en un momento t^* dado de la historia de (Y_t) .

Si $(N_t) = (Y_t - \eta_t) \sim I(0)$, un modelo plausible para (Y_t) podría ser

$$Y_t = \alpha_0 + \omega_0 \xi_t + N_t, \quad (3.8)$$

donde $(N_t) \sim \text{ARMA}(p, q)$ y ξ_t es una **variable determinista** o de **intervención** del tipo

Impulso:

$$\xi_t^{I, t^*} \equiv \begin{cases} 1, & \text{si } t = t^*; \\ 0, & \text{si } t \neq t^*. \end{cases} \quad (3.9.1)$$

Escalón:

$$\xi_t^{E, t^*} \equiv \begin{cases} 1, & \text{si } t \geq t^*; \\ 0, & \text{si } t < t^*. \end{cases} \quad (3.9.2)$$

Observación 1: Un impulso como (3.9.1) representa un cambio **puntual** o **transitorio** en el nivel de (Y_t) , mientras que un escalón como (3.9.2) representa un cambio **permanente** en el nivel de (Y_t) .

Observación 2: El modelo (3.8) puede generalizarse para contemplar: (i) los posibles efectos de varios sucesos de intervención en la historia de (Y_t) , (ii) respuestas de (Y_t) ante dichos sucesos más complejas que una respuesta instantánea, y (iii) la posibilidad de que $(Y_t - \eta_t)$ sea un proceso no estacionario. Un **modelo de intervención** en el que caben todas estas posibilidades es el siguiente (ver 3.1.5):

$$Y_t = \frac{\omega_1(B)}{\delta_1(B)} B^{b_1} \xi_{t1} + \frac{\omega_2(B)}{\delta_2(B)} B^{b_2} \xi_{t2} + \dots + \frac{\omega_M(B)}{\delta_M(B)} B^{b_M} \xi_{tM} + U_t,$$

donde (U_t) es un proceso no estacionario tal que $\phi(B)\Phi(B^S)\nabla^d\nabla_S^D U_t = \theta(B)\Theta(B^S)A_t$, o bien

$$\nabla^d\nabla_S^D Y_t = \frac{\omega_1(B)B^{b_1}}{\delta_1(B)} \nabla^d\nabla_S^D \xi_{t1} + \frac{\omega_2(B)B^{b_2}}{\delta_2(B)} \nabla^d\nabla_S^D \xi_{t2} + \dots + \frac{\omega_M(B)B^{b_M}}{\delta_M(B)} \nabla^d\nabla_S^D \xi_{tM} + N_t,$$

donde $(N_t) \equiv (\nabla^d\nabla_S^D U_t) \sim I(0)$. Si se incluye un término constante α_0 en el modelo anterior, de manera que

$$\nabla^d\nabla_S^D Y_t = \alpha_0 + \frac{\omega_1(B)B^{b_1}}{\delta_1(B)} \nabla^d\nabla_S^D \xi_{t1} + \frac{\omega_2(B)B^{b_2}}{\delta_2(B)} \nabla^d\nabla_S^D \xi_{t2} + \dots + \frac{\omega_M(B)B^{b_M}}{\delta_M(B)} \nabla^d\nabla_S^D \xi_{tM} + N_t,$$

entonces (Y_t) tiene un componente determinista en su tendencia (lo cual **no** resulta apropiado en muchos casos).

Observación 3: En muchas ocasiones, los valores **anómalos** o **atípicos** presentes en una serie temporal pueden corregirse mediante términos de intervención semejantes a los considerados anteriormente. No obstante, en general, dichas correcciones sólo deben figurar explícitamente en un modelo cuando se dispone de **información extramuestral** que justifica su presencia, o bien cuando los valores atípicos correspondientes son **influyentes** (aunque no se disponga en este caso de información extramuestral al respecto).

3.2.2 Ejemplo

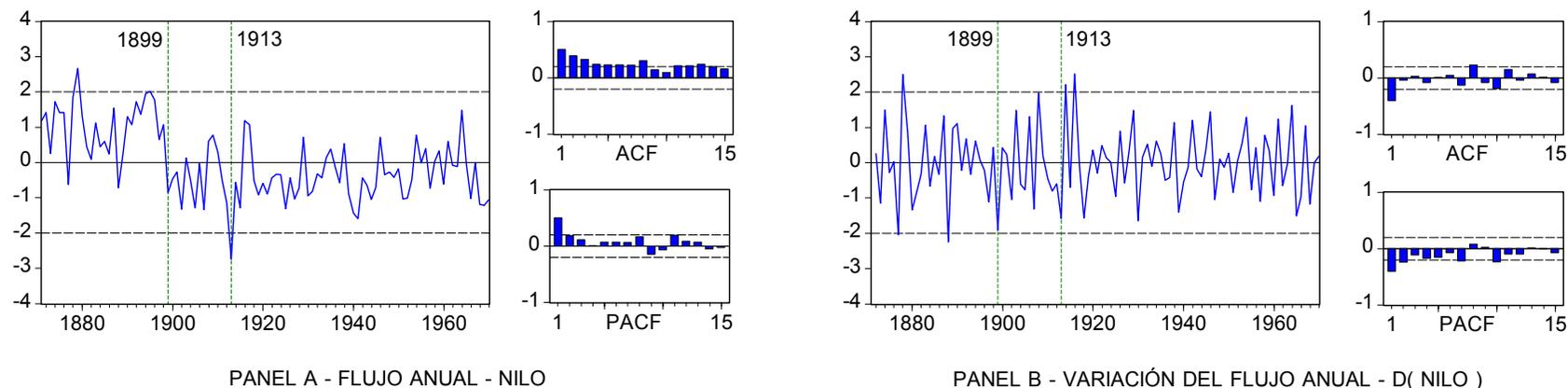


Figura 3.1

Volumen del flujo anual del río Nilo en Asuán (1871 – 1970)
(nilo.wf1)

Si (Y_t) representa el proceso estocástico Volumen del flujo anual del río Nilo en Asuán, los resultados de los contrastes del Ejemplo 2.6.6 sugieren que $(Y_t) \sim I(1) \sim IMA(1,1)$. La Figura 3.2 contiene un modelo IMA(1,1) estimado con EViews 4.1 (May 23 2003 build) para (Y_t) , junto con los residuos y la ACF y la PACF residuales correspondientes.

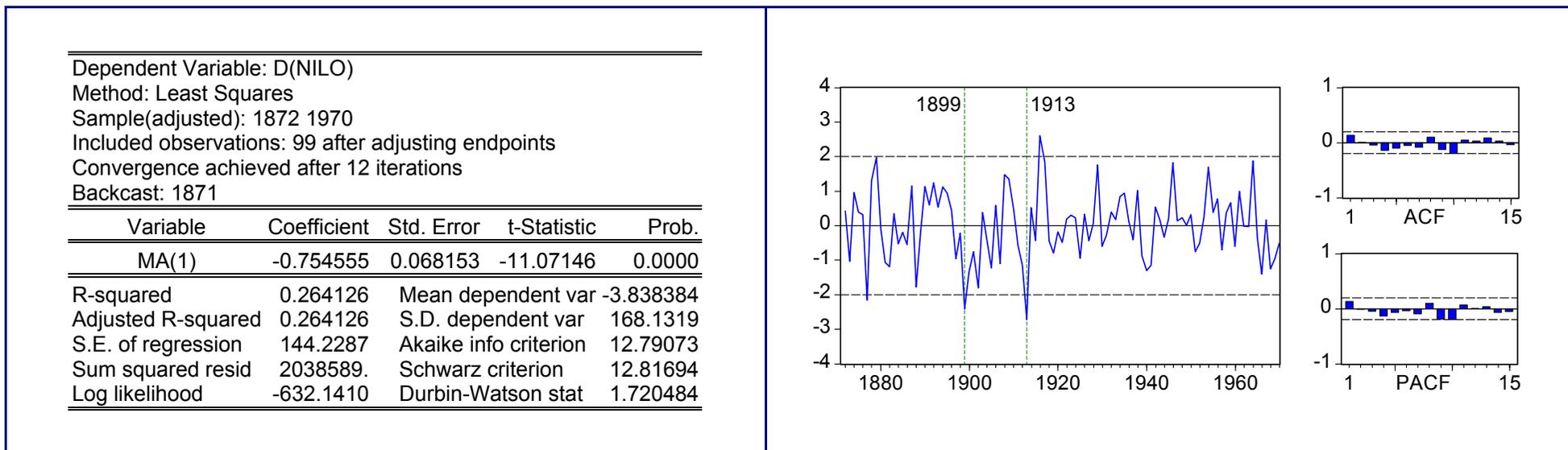


Figura 3.2

Estimación de un modelo MA(1) para la serie D(NILO)

La Figura 3.1 (Panel A) sugiere que el nivel medio de (Y_t) cambió de forma permanente a partir de 1899 (debido probablemente al inicio de la construcción de una presa nueva en Asuán), y que en 1913 el nivel de (Y_t) fue especialmente bajo. Estos dos sucesos pueden analizarse incluyendo en el modelo de la Figura 3.2 un escalón en 1899 y un impulso en 1913. La Figura 3.3 contiene el modelo resultante estimado con EViews 4.1.

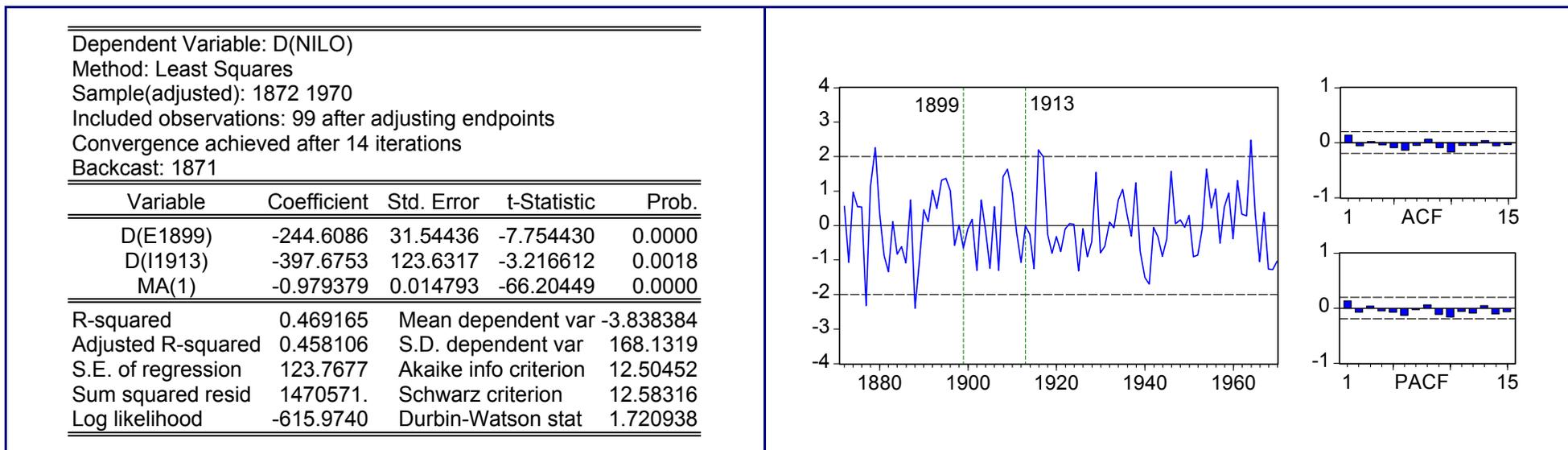


Figura 3.3

Estimación de un modelo MA(1) con dos intervenciones para la serie D(NILO)

La estimación de θ_1 está muy próxima a la unidad, lo que sugiere que el término MA(1) es esencialmente igual a ∇ (la hipótesis $\theta_1 = 1$ puede contrastarse formalmente como en el Ejemplo 2.6.6). Si $\theta_1 = 1$, $Y_t = \omega_{10}\xi_t^{E,1899} + \omega_{20}\xi_t^{I,1913} + U_t$ con $\nabla U_t = (1 - \theta_1 B)A_t$ es equivalente a $Y_t = \alpha_0 + \omega_{10}\xi_t^{E,1899} + \omega_{20}\xi_t^{I,1913} + N_t$, donde (N_t) es ruido blanco. La Figura 3.4 contiene la estimación de este último modelo; la Figura 3.5 contiene el

componente de intervención $\hat{\eta}_t = 1097.75 - 242.23\xi_t^{E,1899} - 399.52\xi_t^{I,1913}$ estimado para (Y_t) , junto con la serie (y_t) de la Figura 3.1 (Panel A).

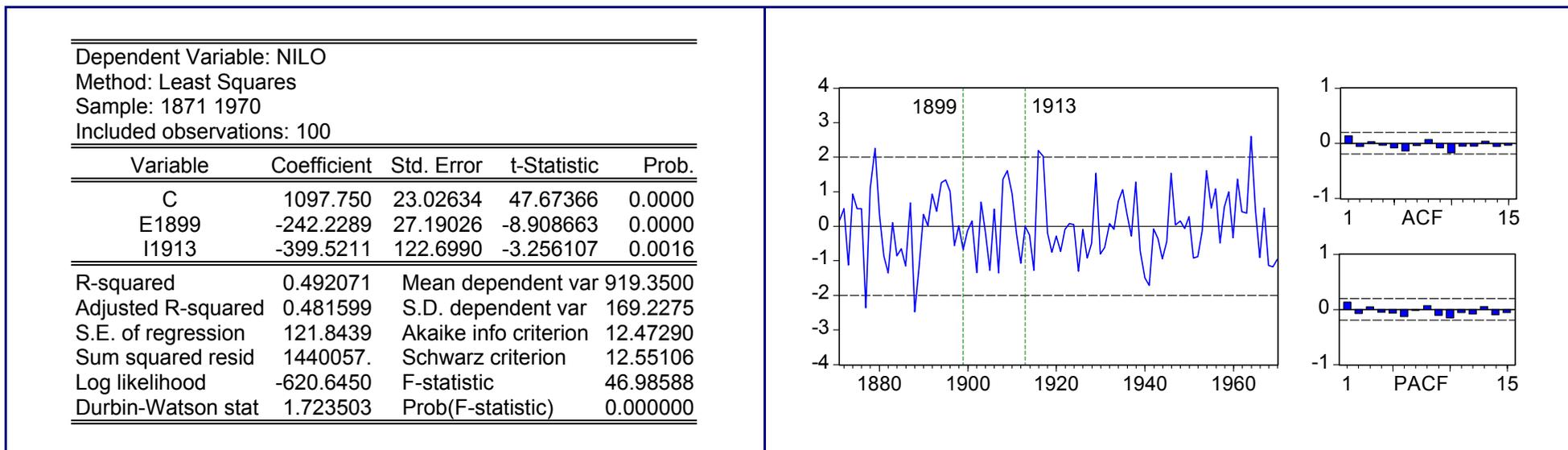


Figura 3.4

Estimación de un modelo con dos intervenciones para la serie NILO

Observaciones: La ocurrencia de sucesos especiales en la historia de un proceso estocástico puede distorsionar el aspecto de una serie temporal, provocando apariencia de no estacionariedad y estructuras de autocorrelación espurias. Un análisis de intervención sistemático, basado en el empleo de información extramuestral y en la detección de observaciones influyentes, permite modelizar adecuadamente dichos sucesos. Las intervenciones de tipo impulso

afectan sólo a una observación y pueden incluirse en un modelo con cierta discrecionalidad; por el contrario, las intervenciones de tipo escalón son más comprometidas y conviene justificarlas con información detallada.

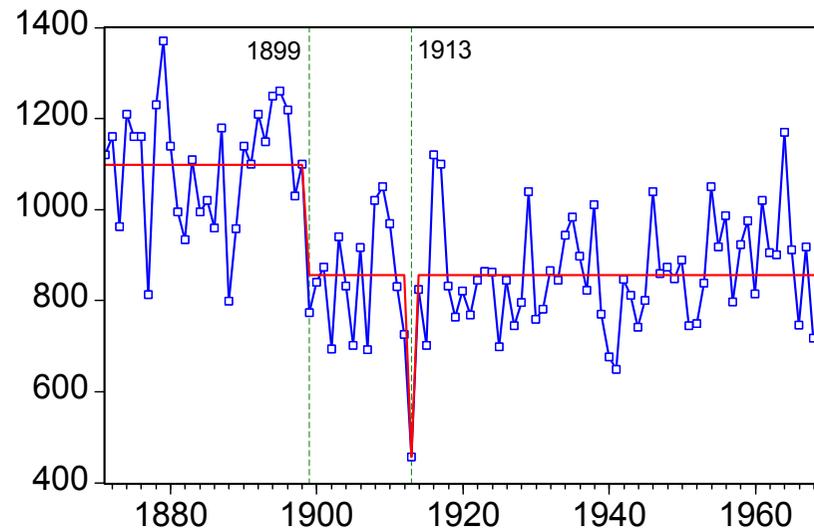


Figura 3.5

Componente de intervención estimado (■) del volumen del flujo anual del río Nilo en Asuán (■)
(nilo.wf1)

3.2.3 Ejemplo

Un ejemplo muy popular del análisis de intervención trata del efecto de legislaciones acerca del uso del cinturón de seguridad sobre el número de víctimas en accidentes de tráfico.

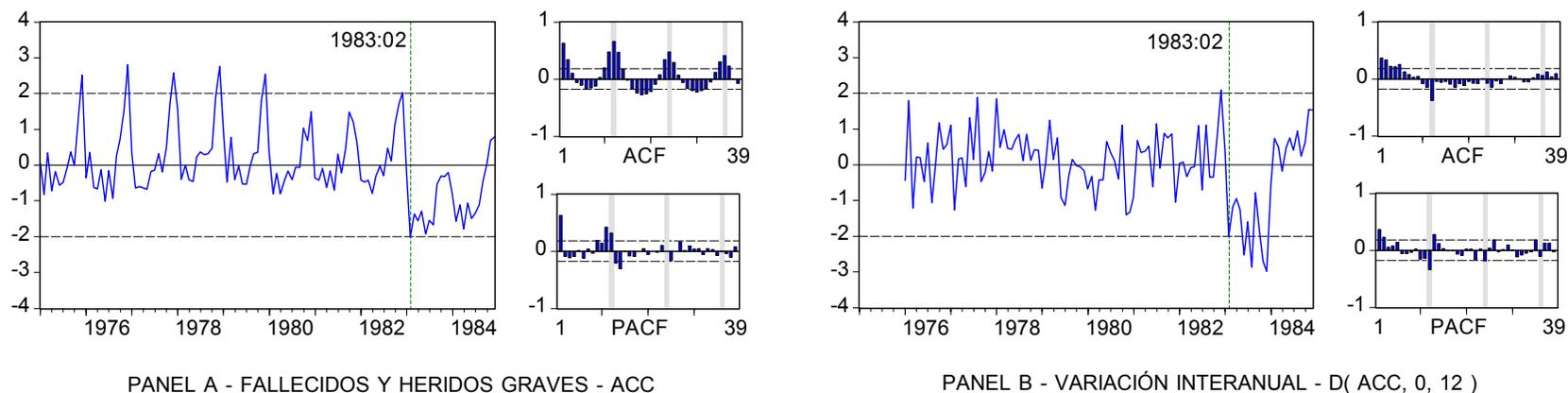


Figura 3.6

Fallecidos y heridos graves en carreteras del Reino Unido (Enero 1975 – Diciembre 1984)
(seat-belt-uk.wf1)

La Figura 3.6 sugiere estimar para (Y_t) un modelo del tipo $Y_t = \omega_0 \xi_t^{E,1983:02} + U_t$, donde $\xi_t^{E,1983:02}$ es un escalón que representa la entrada en vigor de una nueva legislación sobre el uso del cinturón de seguridad en febrero de 1983, y $\nabla_{12}U_t = (1 - \Theta_1 B^{12})A_t$. La Figura 3.7 contiene el modelo estimado con EViews 4.1 (May 23 2003 build) para (Y_t) , junto con los residuos y la ACF y la PACF residuales correspondientes; la Figura 3.8 contiene el componente de intervención $\nabla_{12}\hat{\eta}_t = -287.41 \nabla_{12}\xi_t^{E,1983:02}$ estimado para $(\nabla_{12}Y_t)$, junto

con la serie $(\nabla_{12}y_t)$ de la Figura 3.6 (Panel B).

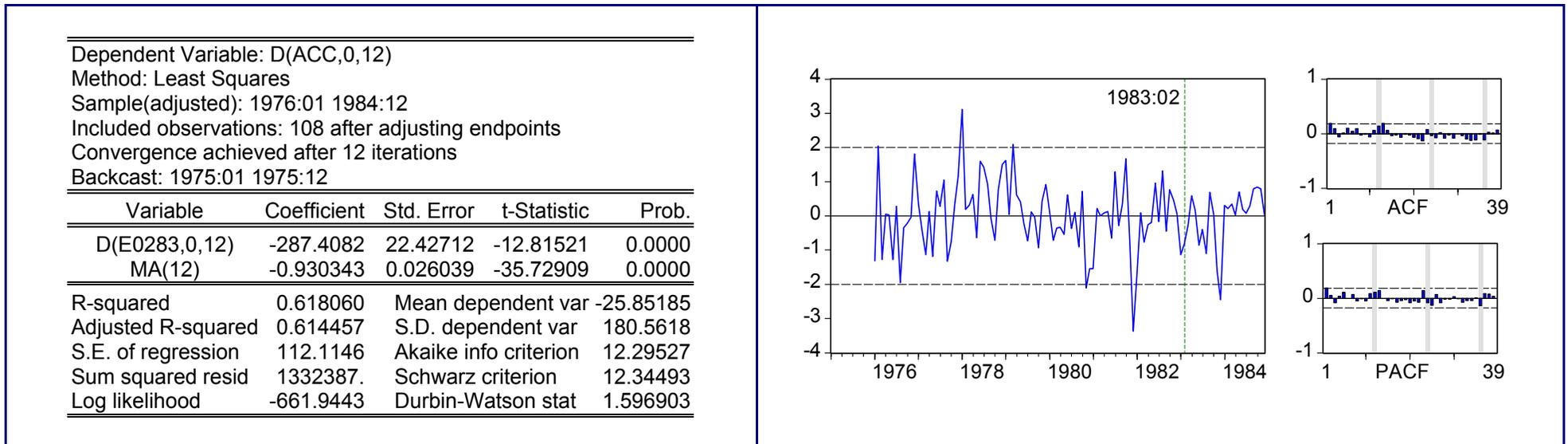


Figura 3.7

Estimación de un modelo $IMA(1,1)_{12}$ con una intervención para la serie ACC

Observaciones: El efecto estimado de la entrada en vigor de la nueva legislación es una **reducción permanente** en el número de fallecidos y heridos graves aproximadamente igual a 287 personas cada mes a partir de febrero de 1983 (un efecto notable, teniendo en cuenta que la media de la serie ACC hasta enero de 1983 es igual a 1621). Por otro lado, la modelización de este suceso de intervención permite descartar una estructura de tipo autorregresivo en $(\nabla_{12}Y_t)$, que parece clara en la Figura 3.6 (Panel B), pero que no está presente en los residuos de la Figura 3.7.

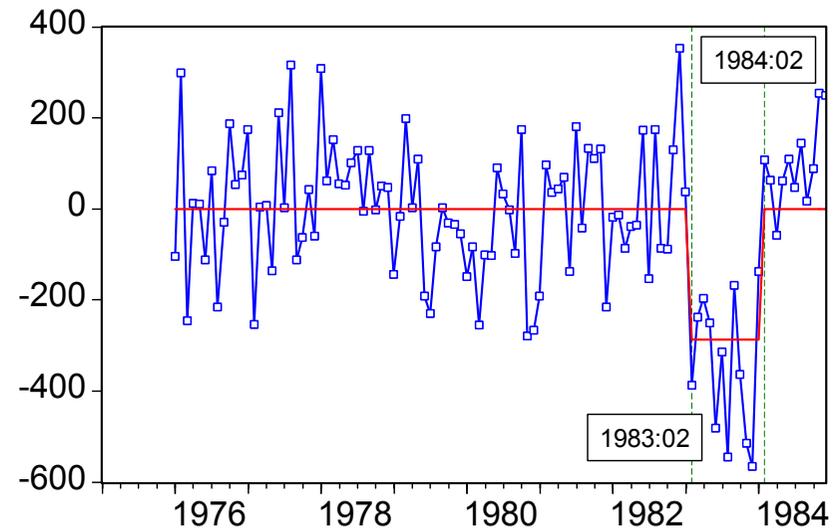


Figura 3.8

Componente de intervención estimado (■) de la variación interanual de la serie ACC (■)
(seat-belt-uk.wf1)

3.3 MODELOS DE RETARDOS DISTRIBUIDOS EN ECONOMETRÍA

3.3.1 Ejemplo

Un modelo de **retardos distribuidos finitos** para un proceso bivalente estacionario (\mathbf{Y}_t) , donde $\mathbf{Y}_t = [Y_t, X_t]'$, es un caso particular de (3.5) con $\delta(B) = 1$ y $b = 0$:

$$Y_t = \alpha_0 + \omega_0 X_t + \omega_1 X_{t-1} + \dots + \omega_s X_{t-s} + N_t,$$

donde $(N_t) \sim I(0)$ es independiente de (X_t) . [En Econometría, suele suponerse, además, que (N_t) es un proceso de ruido blanco.]

Observación 1: En el modelo anterior, ω_0 suele denominarse un **multiplicador de impacto** (en inglés *impact propensity* o *impact multiplier*), en el sentido que $\omega_0 = \partial Y_t / \partial X_t$. La FRI $v_k = \partial Y_t / \partial X_{t-k} = \partial Y_{t+k} / \partial X_t$ ($k \geq 0$) es en este caso $(\omega_0, \omega_1, \dots, \omega_s, 0, \dots)$, mientras que la ganancia a largo plazo (en inglés *steady state gain*, *long-run propensity* o *long-run multiplier*) es simplemente $g = \sum_{k=0}^s \omega_k$, que representa una relación de equilibrio a largo plazo entre los dos componentes de (\mathbf{Y}_t) del tipo $y_* = \alpha_0 + g x_*$.

Observación 2: En Econometría, un modelo TFN como (3.5) [o, en general, como (3.6)] con $b = 0$ y (N_t) ruido blanco, suele denominarse un modelo de **retardos distribuidos infinitos** (o, en inglés, *rational distributed lags*), en el sentido que la FRI es infinita cuando $r \geq 1$ en (3.5); ver Ejemplo 3.1.4. No obstante, la manera de elaborar modelos de este tipo en Econometría suele ser (en el mejor de los casos) arbitraria y muy diferente de la empleada en el Análisis de Series Temporales; ver Observación 2 de 3.1.5.

Modelos ADL

3.3.2 Ejemplo

Un modelo ADL (del inglés *Autoregressive Distributed Lag*) para un proceso bivariante estacionario (\mathbf{Y}_t), donde $\mathbf{Y}_t = [Y_t, X_t]'$, es un caso particular de (3.5) con $\phi(B) = \delta(B)$ y $b = 0$ [ver (I) en Observación 3 de 3.1.4]:

$$\delta(B)Y_t = \beta_0 + \omega(B)X_t + \theta(B)A_t, \quad (3.10)$$

donde $\beta_0 \equiv \delta(1)\alpha_0$, $(A_t) \sim \text{IID}(0, \sigma_A^2)$, y las raíces de $\delta(x) = 0$ están fuera del círculo unitario. En Econometría, (3.10) se escribe a veces como

$$Y_t = \beta_0 + \sum_{i=1}^r \beta_i Y_{t-i} + \sum_{i=0}^s \gamma_i X_{t-i} + V_t, \quad (3.11)$$

donde $\beta_i \equiv \delta_i$ ($i = 1, 2, \dots, r$), $\gamma_0 \equiv \omega_0$, $\gamma_i \equiv -\omega_i$ ($i = 0, 1, \dots, s$), $V_t \equiv \theta(B)A_t$, y suele suponerse que $\theta(B) = 1$, de manera que (V_t) en (3.11) sea un proceso de ruido blanco.

Observación: En Econometría, el modelo de causalidad unidireccional más general que suele considerarse es el modelo ADL de la Definición 3.3.2 (quizás con más de una variable explicativa), que incluye la **restricción fundamental** de que $\phi(B) = \delta(B)$ en el modelo TFN (3.5). En este sentido, nótese que para un proceso estacionario bivariante (\mathbf{Y}_t), con $\mathbf{Y}_t = [Y_t, X_t]'$, (i) un modelo estático de regresión con perturbaciones esféricas es (3.10) con $\omega(B) = \omega_0$ y

$\delta(B) = \theta(B) = 1$, (ii) un modelo estático con perturbaciones autocorrelacionadas es (3.10) con $\omega(B) / \delta(B) = \omega_0$, (iii) un modelo de retardos distribuidos finitos es (3.10) con $\delta(B) = \theta(B) = 1$, y (iii) un modelo de retardos distribuidos infinitos es (3.10) con $\theta(B) = \delta(B)$.

3.3.3 Ejemplo

En Econometría, un modelo ADL(1,1) es (3.11) con $r = s = 1$:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + V_t, \quad (3.12)$$

que describe una relación de equilibrio a largo plazo del tipo $y_* = \beta_0 / (1 - \beta_1) + \lambda x_*$, con $\lambda \equiv (\gamma_0 + \gamma_1) / (1 - \beta_1)$.

Observación: Si se supone que $(V_t) \sim \text{IID}(0, \sigma_V^2)$ en (3.12): (i) cuando $\beta_1 = \gamma_1 = 0$, (3.12) se reduce a un modelo estático de regresión estático con perturbaciones IID, (ii) cuando $\gamma_0 = \gamma_1 = 0$, (3.12) se reduce a un modelo AR(1), (iii) cuando $\gamma_1 = -\beta_1 \gamma_0$, (3.12) se convierte en un modelo de regresión estático con perturbaciones AR(1), y (iv) cuando $\beta_1 = 1$ y $\gamma_1 = -\gamma_0$, (3.12) se convierte en un modelo de regresión estático entre ∇Y_t y ∇X_t con perturbaciones IID, $\nabla Y_t = \beta_0 + \gamma_0 \nabla X_t + V_t$.

Modelos de corrección de error

3.3.4 Ejemplo

El modelo ADL(1,1) de (3.12) puede escribirse como

$$\nabla Y_t = \beta_0 - (1 - \beta_1)(Y_{t-1} - \lambda X_{t-1}) + \gamma_0 \nabla X_t + V_t, \quad (3.13)$$

donde $\lambda \equiv (\gamma_0 + \gamma_1)/(1 - \beta_1)$. El término $(Y_{t-1} - \lambda X_{t-1})$ representa en qué medida no se satisface (salvo quizás por una constante) en el momento $t - 1$ la relación de equilibrio a largo plazo entre los dos componentes de (\mathbf{Y}_t) implícita en (3.12).

(3.13) se denomina un **modelo de corrección de error** (ECM, del inglés *Error Correction Model*). (3.13) implica que cualquier desequilibrio (error) entre Y_{t-1} y X_{t-1} es **transitorio** porque tiende a corregirse a través de variaciones adecuadas en Y_t asociadas con el **término de corrección de error** $-(1 - \beta_1)(Y_{t-1} - \lambda X_{t-1})$, donde $-(1 - \beta_1) > 0$ representa la proporción del desequilibrio entre Y_{t-1} y X_{t-1} que se refleja en $\nabla Y_t \equiv Y_t - Y_{t-1}$.

Observación: Aunque los modelos de esta sección se han planteado para un proceso (\mathbf{Y}_t) bivalente estacionario, todos ellos [en particular el modelo ADL(1,1) de (3.12) en su versión ECM de (3.13)] también pueden ser modelos adecuados para $(\mathbf{Y}_t) \sim I(1)$ cuando (Y_t) y (X_t) están **cointegrados**; ver Tema 4.

3.3.5 Ejemplo

La Figura 3.9 contiene un gráfico de las series $(y_t) \equiv (100 \nabla \ln m1_t^*)$ y $(x_t) \equiv (\nabla tb3m_t)$, así como la ACF muestral bivalente de (\mathbf{y}_t) , $\mathbf{y}_t \equiv [y_t, x_t]'$ (Ver Observación en Ejemplo

3.1.2), donde $(m1_t^*)$ y $(tb3m_t)$ son series anuales sobre M1 (real) y tipos de interés de las Letras del Tesoro a 3 meses en Estados Unidos.

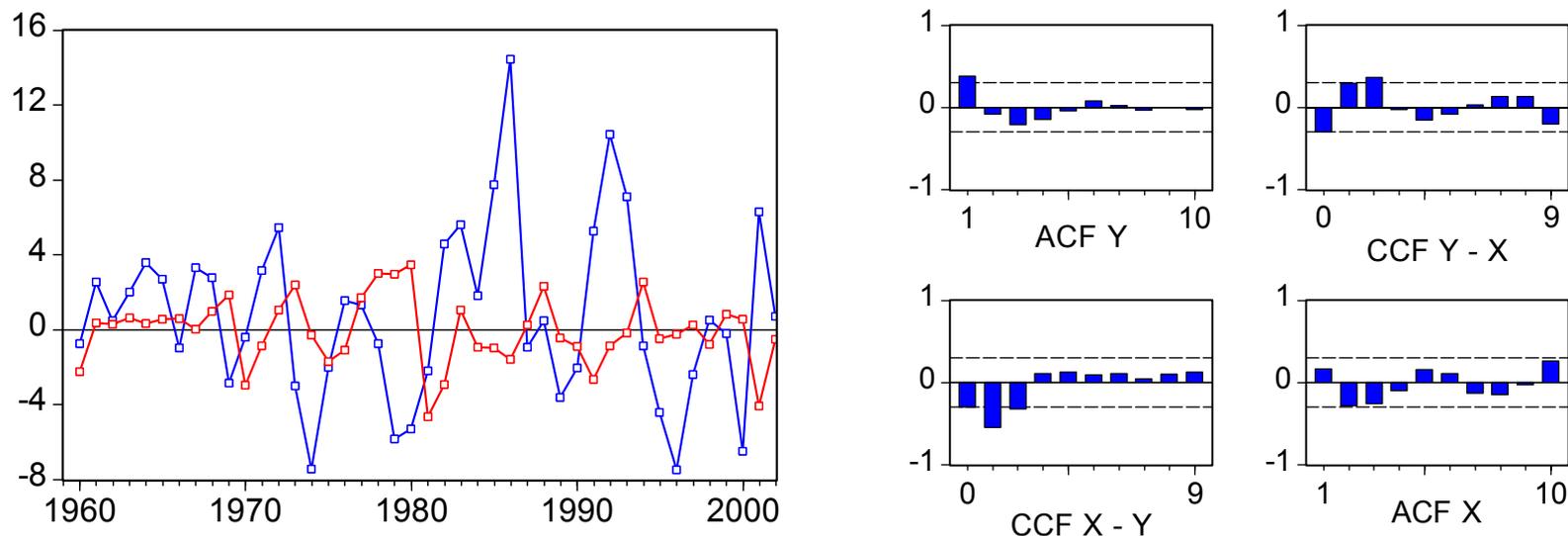


Figura 3.9

Series Y (■) y X (■) del Ejemplo 3.3.5 (1960 – 2002)
(money-demand.wf1)

La Figura 3.10 contiene un modelo ADL(1,1) como (3.12) estimado con EViews 4.1 (May 23 2003 build), junto con los residuos y la ACF y la PACF residuales correspondientes.

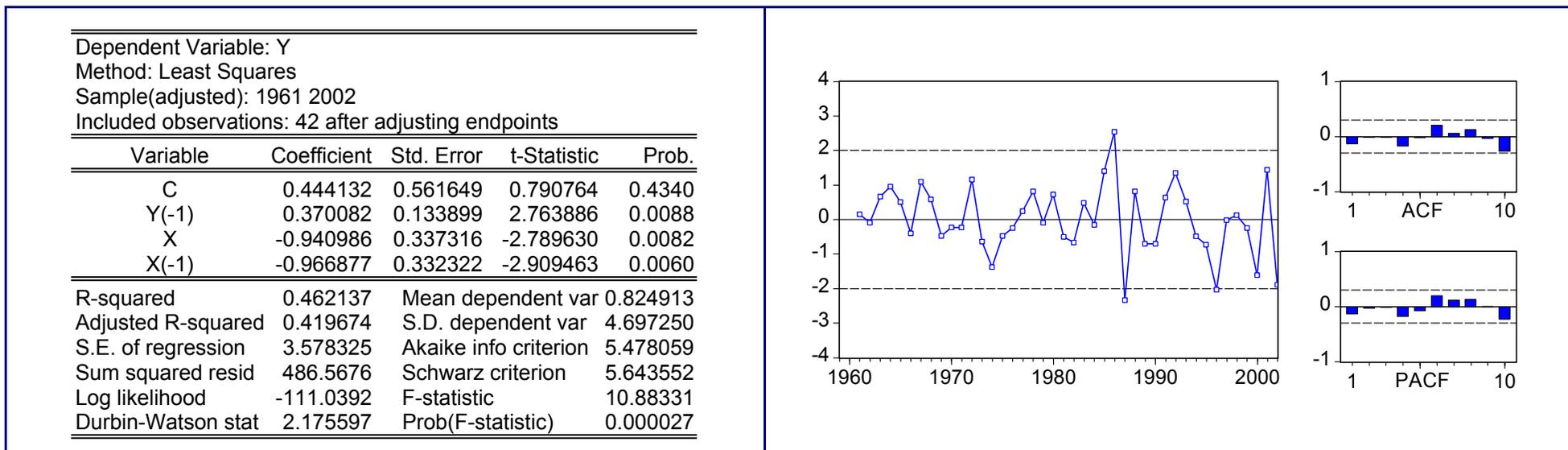


Figura 3.10

Estimación de un modelo ADL(1,1) con las series del Ejemplo 3.3.5

Observación 1: El modelo estimado de la Figura 3.10 puede escribirse como

$$y_t = 0.705 + \frac{-0.941 - 0.967B}{1 - 0.370B} x_t + \frac{1}{1 - 0.370B} \hat{a}_t,$$

que es un modelo TFN cuyo ruido sigue un modelo AR(1) restringido a que $\hat{\phi}(B) = \hat{\delta}(B)$; ver 3.3.2. La IRF y la SRF (ver observaciones 1 y 2 en 3.1.4) implicadas por el modelo TFN anterior están representadas en la Figura 3.11 (comparar con el panel CCF X - Y de la Figura 3.9); la ganancia estimada es $\hat{g} = \hat{\omega}(1) / \hat{\delta}(1) = -3.029$, que implica una relación de equilibrio a largo plazo entre (Y_t) y (X_t) del tipo $y_* = 0.705 - 3.029x_*$.

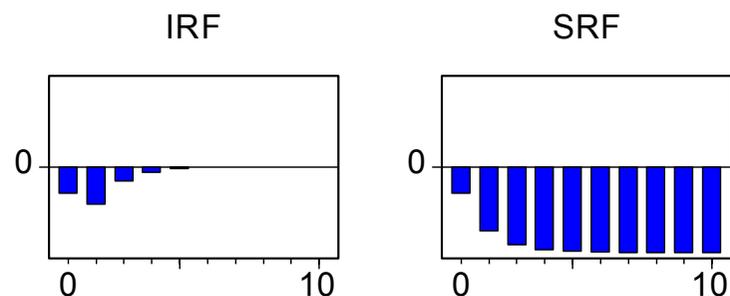


Figura 3.11

IRF y SRF implicadas por el modelo ADL(1,1) estimado de la Figura 3.10

Observación 2: El modelo estimado de la Figura 3.10 también puede escribirse como

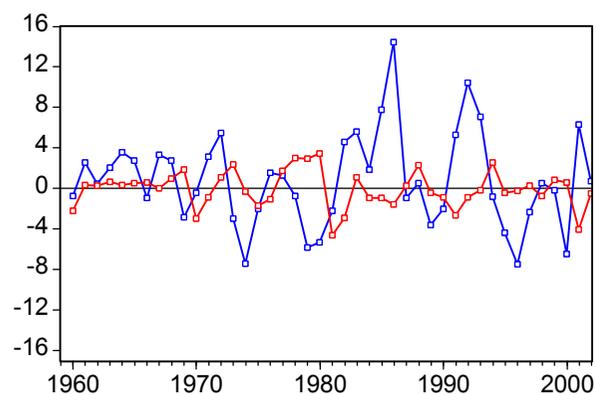
$$\nabla y_t = 0.444 - 0.630(y_{t-1} + 3.029x_{t-1}) - 0.941\nabla x_t + \hat{a}_t,$$

que es un ECM (ver Ejemplo 3.3.4) cuyo término de corrección de error, $-0.630(y_{t-1} + 3.029x_{t-1})$, indica que un 63% del desequilibrio (error) anual observado entre y_{t-1} y x_{t-1} se corrige en el año siguiente. Nótese que la estimación de λ (-3.029) [ver (3.13)] coincide con la ganancia estimada a partir del modelo TFN de la Observación 1. En este sentido, una manera de obtener directamente el error estándar del estimador de λ consiste en estimar el ECM (3.13) directamente por **mínimos cuadrados no lineales**, lo que proporciona (comparar con la Figura 3.10):

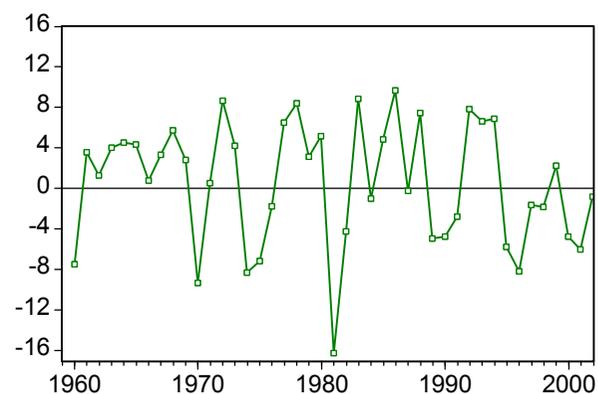
Dependent Variable: D(Y)
 Method: Least Squares
 Sample(adjusted): 1961 2002
 Included observations: 42 after adjusting endpoints
 Convergence achieved after 17 iterations
 D(Y) = C(1) + C(2) * (Y(-1) - C(3) * X(-1)) + C(4) * D(X)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.444132	0.561649	0.790764	0.4340
C(2)	-0.629918	0.133899	-4.704418	0.0000
C(3)	-3.028749	0.906482	-3.341212	0.0019
C(4)	-0.940986	0.337316	-2.789630	0.0082
R-squared	0.568401	Mean dependent var	0.034347	
Adjusted R-squared	0.534327	S.D. dependent var	5.243720	
S.E. of regression	3.578325	Akaike info criterion	5.478059	
Sum squared resid	486.5676	Schwarz criterion	5.643552	
Log likelihood	-111.0392	Durbin-Watson stat	2.175597	

Observación 3: La figura siguiente contiene la representación de las series (y_t) (■) y (x_t) (■) (Panel A) junto con la serie de desequilibrios transitorios estimados $(y_t + 3.029x_t)$ (■) (Panel B):



PANEL A



PANEL B

3.4 MODELOS VAR ESTACIONARIOS

3.4.1 Definición

Un proceso multivariante estacionario (\mathbf{Y}_t) sigue un **modelo AR multivariante** o **vectorial** de orden p , o VAR(p) (autorregresivo vectorial, del inglés *Vector AutoRegressive*), si y sólo si

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{Y}_{t-2} + \dots + \boldsymbol{\Phi}_p \mathbf{Y}_{t-p} + \mathbf{A}_t \quad (3.14)$$

para todo $t = 0, \pm 1, \pm 2, \dots$, donde $(\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_A)$ y $\boldsymbol{\mu}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_p$ son matrices de parámetros tales que todas las raíces de la ecuación

$$\left| \mathbf{I} - \boldsymbol{\Phi}_1 x - \boldsymbol{\Phi}_2 x^2 - \dots - \boldsymbol{\Phi}_p x^p \right| = 0$$

están fuera del círculo unitario (**condición de estacionariedad**).

Observación 1: (3.14) puede escribirse alternativamente como $\boldsymbol{\Phi}(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t$, donde $\boldsymbol{\Phi}(B) \equiv \mathbf{I} - \sum_{i=1}^p \boldsymbol{\Phi}_i B^i$ es el operador autorregresivo (AR) del modelo. (3.14) se denomina un modelo VAR **estándar** o normalizado, en el sentido que $\boldsymbol{\Phi}(0) = \mathbf{I}$.

Observación 2: Dado que $\boldsymbol{\Sigma}_A \equiv \text{Var}[\mathbf{A}_t]$ es una matriz definida positiva, $\boldsymbol{\Sigma}_A$ puede factorizarse como $\boldsymbol{\Sigma}_A = \mathbf{T}\boldsymbol{\Sigma}^*\mathbf{T}'$,

donde Σ^* es una matriz diagonal definida positiva y \mathbf{T} es una matriz triangular inferior con unos en su diagonal principal (**factorización de Cholesky**). La matriz \mathbf{T} puede emplearse para transformar $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t$ en un modelo con **perturbaciones ortogonales** del tipo $\Phi^*(B)\mathbf{Y}_t = \boldsymbol{\mu}^* + \mathbf{A}_t^*$, con $\Phi^*(B) \equiv \mathbf{T}^{-1}\Phi(B)$, $\boldsymbol{\mu}^* \equiv \mathbf{T}^{-1}\boldsymbol{\mu}$, $\mathbf{A}_t^* \equiv \mathbf{T}^{-1}\mathbf{A}_t$; en este caso, $\Phi^*(0) = \mathbf{T}^{-1}$ (triangular inferior) y $\text{Var}[\mathbf{A}_t^*] = \mathbf{T}^{-1}\Sigma_{\mathbf{A}}\mathbf{T}^{-1'} = \Sigma^*$ (diagonal); por lo tanto, $\Phi^*(0)$ muestra explícitamente las relaciones contemporáneas entre los componentes del proceso (\mathbf{Y}_t), mientras que en la representación estándar (3.14) la información sobre dichas relaciones está recogida implícitamente en $\Sigma_{\mathbf{A}}$ (una matriz no diagonal en general). La representación no estándar anterior de un modelo VAR para (\mathbf{Y}_t) puede depender crucialmente del **orden** en el que figuran los componentes de \mathbf{Y}_t ; ver observaciones 1 a 4 de 3.1.3.

3.4.2 Ejemplo

La Tabla 3.1 contiene la estimación con EViews 4.1 (May 23 2003 build) de un modelo VAR(1) bivalente con la serie (\mathbf{y}_t), $\mathbf{y}_t = [y_t, x_t]'$, del Ejemplo 3.3.5; el modelo estimado puede resumirse como

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0.254 & -1.222 \\ 0.141 & 0.275 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{a}_{t1} \\ \hat{a}_{t2} \end{bmatrix}, \hat{\Sigma}_{\mathbf{A}} = \begin{bmatrix} 14.974 & -2.717 \\ -2.717 & 2.829 \end{bmatrix}, \quad (3.15)$$

$$N = 42, \quad |\hat{\Sigma}_{\mathbf{A}}| = 34.97, \quad L^* = -193.84, \quad \text{AIC} = 9.42, \quad \text{BIC} = 9.59.$$

La Figura 3.12 contiene la serie de residuos ($\hat{\mathbf{a}}_t$), $\hat{\mathbf{a}}_t = [\hat{a}_{t1}, \hat{a}_{t2}]'$, del modelo (3.15).

Vector Autoregression Estimates		
Sample(adjusted): 1961 2002		
Included observations: 42 after adjusting endpoints		
Standard errors in () & t-statistics in []		
	Y	X
Y(-1)	0.253606 (0.13289) [1.90841]	0.141040 (0.05776) [2.44199]
X(-1)	-1.222001 (0.34493) [-3.54278]	0.275128 (0.14991) [1.83525]
R-squared	0.337901	0.153942
Adj. R-squared	0.321349	0.132790
Sum sq. resids	598.9547	113.1407
S.E. equation	3.869608	1.681820
F-statistic	20.41396	7.278070
Log likelihood	-115.4033	-80.40563
Akaike AIC	5.590631	3.924078
Schwarz SC	5.673378	4.006824
Mean dependent	0.824913	-0.025238
S.D. dependent	4.697250	1.805998
Determinant Residual Covariance		34.97280
Log Likelihood (d.f. adjusted)		-193.8368
Akaike Information Criteria		9.420801
Schwarz Criteria		9.586293

Tabla 3.1

Estimación de un modelo VAR(1) con las series del Ejemplo 3.3.5

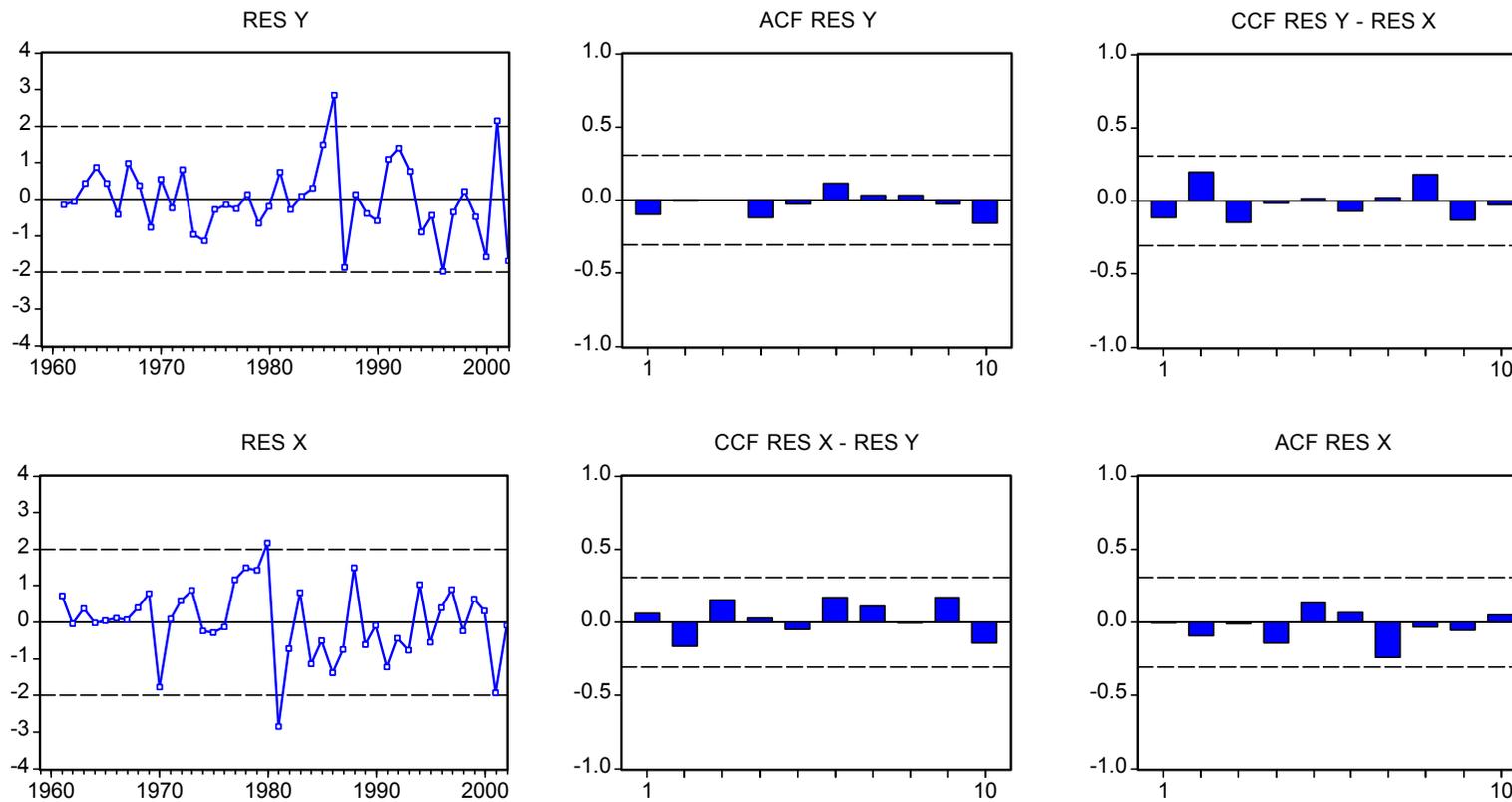


Figura 3.12

Residuos del modelo VAR(1) estimado de la Tabla 3.1

Observación 1: En las dos CCFs de la Figura 3.12 no está representada la correlación muestral contemporánea entre las dos series de residuos (el valor de ambas CCFs para el retardo 0) porque dicha correlación figura implícitamente

en la matriz de varianzas-covarianzas $\hat{\Sigma}_{\mathbf{A}}$ de (2.15); la matriz de correlaciones correspondiente es

$$\hat{\mathbf{P}}_{\mathbf{A}} = \begin{bmatrix} 1 & -0.417 \\ -0.417 & 1 \end{bmatrix},$$

donde -0.417 es la correlación muestral contemporánea entre las dos series de residuos.

Observación 2: La factorización de Cholesky $\hat{\Sigma}_{\mathbf{A}} = \hat{\mathbf{T}}\hat{\Sigma}^*\hat{\mathbf{T}}'$ de la matriz $\hat{\Sigma}_{\mathbf{A}}$ de (3.15) es

$$\hat{\Sigma}_{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ -0.181 & 1 \end{bmatrix} \begin{bmatrix} 14.974 & 0 \\ 0 & 2.336 \end{bmatrix} \begin{bmatrix} 1 & -0.181 \\ 0 & 1 \end{bmatrix}, \text{ de manera que } \hat{\mathbf{T}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0.181 & 1 \end{bmatrix}.$$

Premultiplicando el modelo estimado (3.15) por $\hat{\mathbf{T}}^{-1}$, se obtiene que:

$$\begin{bmatrix} 1 & 0 \\ 0.181 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} - \begin{bmatrix} 0.254 & -1.222 \\ 0.187 & 0.053 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} \hat{a}_{t1}^* \\ \hat{a}_{t2}^* \end{bmatrix}, \quad \hat{\Sigma}^* = \begin{bmatrix} 14.974 & 0 \\ 0 & 2.336 \end{bmatrix},$$

es decir,

$$\begin{aligned} y_t &= 0.254y_{t-1} - 1.222x_{t-1} + \hat{a}_{t1}^*, \\ x_t &= 0.053x_{t-1} - 0.181y_t + 0.187y_{t-1} + \hat{a}_{t2}^*, \end{aligned} \tag{I}$$

donde las series de residuos (\hat{a}_{t1}^*) y (\hat{a}_{t2}^*) son ortogonales. Por otro lado, (3.15) es idéntico, excepto por el orden en el que figuran y_t y x_t , a

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} - \begin{bmatrix} 0.275 & 0.141 \\ -1.222 & 0.254 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \hat{a}_{t2} \\ \hat{a}_{t1} \end{bmatrix}, \quad \hat{\Sigma}_{\mathbf{A}} = \begin{bmatrix} 2.829 & -2.717 \\ -2.717 & 14.974 \end{bmatrix}. \tag{II}$$

Ahora, la factorización de Cholesky $\hat{\Sigma}_{\mathbf{A}} = \hat{\mathbf{T}}\hat{\Sigma}^*\hat{\mathbf{T}}'$ de la matriz $\hat{\Sigma}_{\mathbf{A}}$ de (II) es

$$\hat{\Sigma}_{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ -0.961 & 1 \end{bmatrix} \begin{bmatrix} 2.829 & 0 \\ 0 & 12.364 \end{bmatrix} \begin{bmatrix} 1 & -0.961 \\ 0 & 1 \end{bmatrix}, \text{ de manera que ahora } \hat{\mathbf{T}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0.961 & 1 \end{bmatrix}.$$

Premultiplicando (II) por $\hat{\mathbf{T}}^{-1}$, se obtiene que

$$\begin{bmatrix} 1 & 0 \\ 0.961 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} - \begin{bmatrix} 0.275 & 0.141 \\ -0.958 & 0.389 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \hat{a}_{t2}^* \\ \hat{a}_{t1}^* \end{bmatrix}, \quad \hat{\Sigma}^* = \begin{bmatrix} 2.829 & 0 \\ 0 & 12.364 \end{bmatrix},$$

es decir

$$\begin{aligned} y_t &= 0.389y_{t-1} - 0.961x_t - 0.958x_{t-1} + \hat{a}_{t1}^*, \\ x_t &= 0.275x_{t-1} + 0.141y_{t-1} + \hat{a}_{t2}^*. \end{aligned} \tag{III}$$

(III) implica que $(X_t) \rightarrow (Y_t)$ contemporáneamente pero $(Y_t) \rightarrow (X_t)$ no, exactamente al contrario que en (I); por lo tanto, la manera de ordenar y_t y x_t en un modelo VAR(1) estándar, influye notablemente sobre las implicaciones del modelo VAR(1) estructural resultante. En general, dichas implicaciones deberían ser consistentes con algún tipo de información extramuestral o teórica. Por ejemplo, si se cree que $(X_t) \rightarrow (Y_t)$ contemporáneamente pero $(Y_t) \rightarrow (X_t)$ no, entonces (III) es una forma estructural adecuada mientras que (I) no lo es; en este sentido, nótese muy cuidadosamente que la primera ecuación de (III) es perfectamente compatible con el modelo ADL(1,1) estimado en el Ejemplo 3.3.5 bajo la hipótesis de causalidad unidireccional $(X_t) \rightarrow (Y_t)$.

Observación 3: La ACF teórica bivariante (ver Observación de 3.1.2) implicada por el modelo estimado (3.15), está representada en la Figura 3.13. La ACF teórica resulta compatible con las pautas observadas en la ACF muestral bivariante de la Figura 3.9, por lo que el modelo VAR(1) estimado de la Tabla 3.1 parece adecuado en este caso.

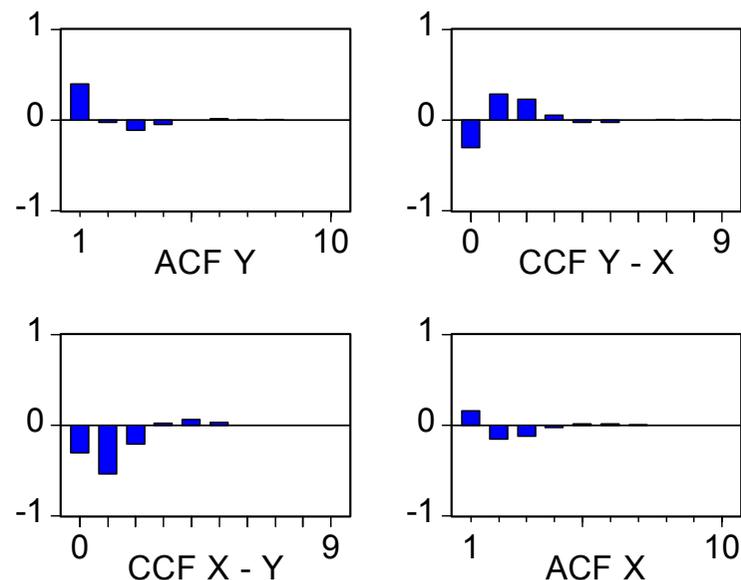


Figura 3.13

ACF teórica bivalente implicada por el modelo VAR(1) estimado de la Tabla 3.1

Identificación – Estimación – Contrastes de hipótesis

Para identificar un orden inicial p adecuado para un modelo VAR a partir de una serie temporal M -variante $(\mathbf{y}_t)_{t=1}^N$, suelen emplearse en la práctica los instrumentos siguientes:

⇒ Los **modelos univariantes** elaborados a partir de cada componente de (\mathbf{y}_t) .

- ⇒ La **ACF muestral multivariante** de (\mathbf{y}_t) (como en la Figura 3.9), que consta de M ACFs individuales más $M(M - 1)$ CCFs entre cada par de componentes de (\mathbf{y}_t) .
- ⇒ Los **criterios de información** AIC y BIC (ver definiciones 2.6.1 y 2.6.2) para varios modelos VAR estimados con órdenes distintos y con la misma muestra efectiva.
- ⇒ El valor calculado del **estadístico de razón de verosimilitudes** (LR, del inglés *Likelihood Ratio*), $LR \approx N \left(\ln \left| \hat{\Sigma}_{\mathbf{A}}^0 \right| - \ln \left| \hat{\Sigma}_{\mathbf{A}}^1 \right| \right)$, donde $\hat{\Sigma}_{\mathbf{A}}^0$ y $\hat{\Sigma}_{\mathbf{A}}^1$ son las matrices de varianzas-covarianzas estimadas de (\mathbf{A}_t) bajo H_0 (por ejemplo $p = p_0$) y bajo H_1 (por ejemplo $p = p_1$, con $p_1 > p_0$), respectivamente, a partir de la misma muestra efectiva. Bajo H_0 , el estadístico LR sigue aproximadamente una distribución $\chi^2(l)$, donde l es el número total de restricciones incluidas en H_0 [en general, $l = M^2(p_1 - p_0)$].

Observación 1: El estadístico LR anterior puede emplearse para contrastar una gran variedad de hipótesis a partir de un modelo VAR estimado. Un uso bastante popular del estadístico LR (además de la identificación inicial de un orden p adecuado) tiene que ver con el llamado **contraste de causalidad de Granger**. Si un proceso multivariante (\mathbf{Y}_t) de dimensión M se divide en dos bloques de componentes (\mathbf{Y}_{t1}) e (\mathbf{Y}_{t2}) de dimensiones M_1 y M_2 , respectivamente, entonces se dice que (\mathbf{Y}_{t2}) no causa a (\mathbf{Y}_{t1}) según Granger [en inglés, (\mathbf{Y}_{t2}) *does not Granger cause* (\mathbf{Y}_{t1})] cuando la esperanza de (\mathbf{Y}_{t1}) condicionada por todo el pasado de (\mathbf{Y}_{t1}) y de (\mathbf{Y}_{t2}) coincide con la esperanza de (\mathbf{Y}_{t1}) condicionada sólo por su propio pasado. Si (\mathbf{Y}_t) sigue un modelo VAR(p) del tipo

$$\begin{bmatrix} \mathbf{Y}_{t1} \\ \mathbf{Y}_{t2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \sum_{j=1}^p \begin{bmatrix} \boldsymbol{\Phi}_j^{11} & \boldsymbol{\Phi}_j^{12} \\ \boldsymbol{\Phi}_j^{21} & \boldsymbol{\Phi}_j^{22} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-j,1} \\ \mathbf{Y}_{t-j,2} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{t1} \\ \mathbf{A}_{t2} \end{bmatrix}, \quad \boldsymbol{\Sigma}_A = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

entonces cuando (\mathbf{Y}_{t2}) no causa a (\mathbf{Y}_{t1}) debe ocurrir que $\boldsymbol{\Phi}_j^{12} = \mathbf{0}$ para todo $j = 1, \dots, p$. [Análogamente, cuando (\mathbf{Y}_{t1}) no causa a (\mathbf{Y}_{t2}) debe ocurrir que $\boldsymbol{\Phi}_j^{21} = \mathbf{0}$ para todo $j = 1, \dots, p$.] Dado que las p matrices $\boldsymbol{\Phi}_j^{12}$ sólo aparecen en la primera fila de la expresión anterior, para contrastar la hipótesis nula (H_0) de que $\boldsymbol{\Phi}_j^{12} = \mathbf{0}$ para todo $j = 1, \dots, p$, se puede (i) estimar los dos modelos siguientes:

$$\text{Bajo } H_0 : \mathbf{Y}_{t1} = \boldsymbol{\mu}_1 + \sum_{j=1}^p \boldsymbol{\Phi}_j^{11} \mathbf{Y}_{t-j,1} + \mathbf{A}_{t1},$$

$$\text{Bajo } H_1 : \mathbf{Y}_{t1} = \boldsymbol{\mu}_1 + \sum_{j=1}^p \left[\boldsymbol{\Phi}_j^{11} \mathbf{Y}_{t-j,1} + \boldsymbol{\Phi}_j^{12} \mathbf{Y}_{t-j,2} \right] + \mathbf{A}_{t1},$$

y (ii) calcular el valor del estadístico $LR \approx N \left(\ln \left| \hat{\boldsymbol{\Sigma}}_{11}^0 \right| - \ln \left| \hat{\boldsymbol{\Sigma}}_{11}^1 \right| \right)$, que bajo H_0 sigue aproximadamente una distribución $\chi^2(l)$, donde l es el número de restricciones incluidas en H_0 (en general, $l = pM_1M_2$).

La estimación de un modelo VAR suele llevarse a cabo aplicando MCO a cada ecuación (fila) del modelo por separado (lo que implica que los p primeros residuos **no** pueden calcularse después de estimar el modelo), aunque en general es recomendable estimar cualquier modelo VAR (lo mismo que cualquier modelo VARMA) por máxima verosimilitud exacta (MVE); ver Observación 8 en 1.3.3 y Observación 2 en 3.1.5.

Observación 2: En general, un modelo VAR(p) puede estar **sobreparametrizado** debido a que para recoger la estructura MA presente en muchas series estacionarias es necesario que el orden p sea elevado. Esto implica que las estimaciones de los parámetros de un VAR suelen ser poco precisas, lo cual puede no ser un problema si sólo se pretende usar el modelo estimado para prever. En general, cualquier análisis estructural debería estar basado en modelos que recojan explícitamente tanto la estructura AR como la estructura MA de una serie multivariante, aunque esto **no** es lo habitual en la práctica econométrica actual; ver Observación 8 en 1.3.3.

Previsión

3.4.3 Definición

El modelo VAR(p) estacionario de la Definición 3.4.1 puede escribirse como

$$\mathbf{Y}_t = \boldsymbol{\mu}_Y + \boldsymbol{\Psi}(B)\mathbf{A}_t, \quad (3.16)$$

donde $\boldsymbol{\mu}_Y \equiv E[\mathbf{Y}_t] = \boldsymbol{\Phi}(1)^{-1}\boldsymbol{\mu}$, $\boldsymbol{\Psi}(B) \equiv \boldsymbol{\Phi}(B)^{-1} = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i B^i$, $\boldsymbol{\Psi}_i = \sum_{j=1}^p \boldsymbol{\Phi}_j \boldsymbol{\Psi}_{i-j}$ (con $\boldsymbol{\Psi}_0 \equiv \mathbf{I}$, $\boldsymbol{\Psi}_i \equiv \mathbf{0}$ si $i < 0$), y $\sum_{i=0}^{\infty} \|\boldsymbol{\Psi}_i\| < \infty$. (3.16) se denomina la **representación de Wold** de un modelo AR(p) estacionario.

3.4.4 Definición

La **función de previsión** en origen k a horizonte $l \geq 1$ de un proceso $(\mathbf{Y}_t) \sim I(0)$ que sigue

un modelo VAR(p) del tipo $\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t$, donde $\Phi(B) \equiv \mathbf{I} - \sum_{i=1}^p \Phi_i B^i$, es

$$\mathbf{Y}_k(l) \equiv E_k[\mathbf{Y}_{k+l}] = \boldsymbol{\mu} + \sum_{i=1}^p \Phi_i E_k[\mathbf{Y}_{k+l-i}], \quad (3.17)$$

o bien

$$\mathbf{Y}_k(l) \equiv E_k[\mathbf{Y}_{k+l}] = \boldsymbol{\mu}_Y + E_k[\Psi(B)\mathbf{A}_{k+l}] = \boldsymbol{\mu}_Y + \sum_{i=l}^{\infty} \Psi_i \mathbf{A}_{k+l-i}, \quad (3.18)$$

donde $\Psi(B)$ es el polinomio de la representación de Wold (3.16).

Observación 1: La función de previsión (3.17) de un proceso $(\mathbf{Y}_t) \sim \text{VAR}(p)$ se utiliza en la práctica para calcular previsiones puntuales $\hat{\mathbf{y}}_k(1), \hat{\mathbf{y}}_k(2), \dots$ a partir de un modelo estimado. Para ello, en (3.17) se reemplazan los parámetros $\boldsymbol{\mu}, \Phi_1, \dots, \Phi_p$ por sus valores estimados, y se reemplaza cada $E_k[\mathbf{Y}_j]$ por \mathbf{y}_j (un dato) si $j \leq k$, o bien por $\hat{\mathbf{y}}_k(j-k)$ (una previsión) si $j > k$. Los **errores de previsión** asociados con las previsiones $\hat{\mathbf{y}}_k(1), \hat{\mathbf{y}}_k(2), \dots$, suelen representarse como $\hat{\mathbf{e}}_k(l) \equiv \mathbf{y}_{k+l} - \hat{\mathbf{y}}_k(l)$ ($l = 1, 2, \dots$) (que sólo pueden calcularse cuando $\mathbf{y}_{k+1}, \mathbf{y}_{k+2}, \dots$ son datos conocidos); por su parte, la función de previsión (3.18) se utiliza para describir las propiedades estadísticas de los errores de previsión (ver Definición 3.4.5).

Observación 2: Un modelo VAR(p) estimado puede escribirse como $\hat{\Phi}(B)\mathbf{y}_t = \hat{\boldsymbol{\mu}} + \hat{\mathbf{a}}_t$, con $\hat{\Phi}(B) \equiv \mathbf{I} - \sum_{i=1}^p \hat{\Phi}_i B^i$; en consecuencia, cada residuo $\hat{\mathbf{a}}_t$ de un modelo estimado puede escribirse como

$$\hat{\mathbf{a}}_t = \mathbf{y}_t - \left[\hat{\boldsymbol{\mu}} + \sum_{i=1}^p \hat{\Phi}_i \mathbf{y}_{t-i} \right],$$

donde el término entre corchetes es un **valor ajustado**, que puede interpretarse como una previsión en origen $t-1$ a horizonte 1. Por lo tanto, $\hat{\mathbf{a}}_t$ puede interpretarse como un error de previsión en origen $t-1$ a horizonte 1.

3.4.5 Definición

El **error de previsión** en origen k a horizonte $l \geq 1$ de un proceso $(\mathbf{Y}_t) \sim \mathbf{I}(0)$ que sigue un modelo VAR(p) puede escribirse, de acuerdo con (3.16) y (3.18), como

$$\mathbf{E}_k(l) \equiv \mathbf{Y}_{k+l} - \mathbf{Y}_k(l) = \boldsymbol{\Psi}(B)\mathbf{A}_{k+l} - \mathbf{E}_k[\boldsymbol{\Psi}(B)\mathbf{A}_{k+l}] = \sum_{i=0}^{l-1} \boldsymbol{\Psi}_i \mathbf{A}_{k+l-i}. \quad (3.19)$$

Observación 1: (3.19) implica que $\mathbf{E}[\mathbf{E}_k(l)] = \mathbf{0}$ y que $\text{Var}[\mathbf{E}_k(l)] = \sum_{i=0}^{l-1} \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_A \boldsymbol{\Psi}_i'$. Como $\text{Var}[\mathbf{E}_k(l)]$ depende del horizonte $l \geq 1$ pero **no** del origen k de previsión, $\text{Var}[\mathbf{E}_k(l)]$ suele escribirse simplemente como $\boldsymbol{\Sigma}(l)$, $l = 1, 2, \dots$. Por otro lado, teniendo en cuenta que $\sum_{i=0}^{\infty} \|\boldsymbol{\Psi}_i\| < \infty$, $\mathbf{Y}_k(l) \rightarrow \boldsymbol{\mu}_Y$ [ver (3.18)] y $\boldsymbol{\Sigma}(l) \rightarrow \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_A \boldsymbol{\Psi}_i' = \boldsymbol{\Sigma}_Y$ cuando el horizonte de previsión l tiende a infinito.

Observación 2: Si (\mathbf{A}_t) es un proceso Gaussiano (Normal), entonces $\mathbf{E}_k \sim \mathbf{N}[\mathbf{0}, \boldsymbol{\Sigma}(l)]$. Por lo tanto, los elementos $\sigma_{ii}(l)$ en la diagonal principal de $\boldsymbol{\Sigma}(l)$ pueden utilizarse, junto con las previsiones puntuales $Y_{ki}(l)$, para construir de la forma habitual intervalos de confianza a horizonte l para $Y_{k+l,i}$ ($1 \leq i \leq M; l = 1, 2, \dots$); ver, por ejemplo, (2.79) y (2.81) para el caso univariante.

3.4.6 Ejemplo

El modelo VAR(1) para un proceso $(\mathbf{Y}_t) \sim \mathbf{I}(0)$ es

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \mathbf{A}_t.$$

Escribiendo esta expresión en $t = N + l$ ($l \geq 1$), y aplicando al resultado el operador $E_N[\cdot]$, se obtiene que la función de previsión en origen N a horizonte $l \geq 1$ es

$$\mathbf{Y}_N(l) = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 E_N[\mathbf{Y}_{N+l-1}] + E_N[\mathbf{A}_{N+l}]. \quad (3.20)$$

Para $l = 1$, (3.20) queda

$$\mathbf{Y}_N(1) = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_N, \quad (3.21)$$

mientras que para $l \geq 2$, (3.20) queda

$$\mathbf{Y}_N(l) = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_N(l-1). \quad (3.22)$$

(3.22) es una ecuación en diferencias, cuya solución a partir de la condición inicial (3.21) es

$$\mathbf{Y}_N(l) = \left(\sum_{i=0}^{l-1} \boldsymbol{\Phi}_1^i \right) \boldsymbol{\mu} + \boldsymbol{\Phi}_1^l \mathbf{Y}_N \quad \text{para todo } l \geq 1. \quad (3.23)$$

(3.23) es la **función final de previsión** de $(\mathbf{Y}_t) \sim \text{VAR}(1)$, que converge al valor esperado de (\mathbf{Y}_t) , $E[\mathbf{Y}_t] = (\mathbf{I} - \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\mu}$, cuando $l \rightarrow \infty$.

Observación: El polinomio $\boldsymbol{\Psi}(B)$ en la expresión (3.16) para un modelo VAR(1) es $\boldsymbol{\Psi}(B) = \boldsymbol{\Phi}(B)^{-1} = \sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i B^i$, por lo que $\boldsymbol{\Sigma}(l) = \sum_{i=0}^{l-1} \boldsymbol{\Phi}_1^i \boldsymbol{\Sigma}_A \boldsymbol{\Phi}_1^{i'}$ en este caso.

Funciones de respuesta a impulsos

La representación de Wold (3.16) de un proceso $(\mathbf{Y}_t) \sim \text{VAR}(p)$ estacionario puede escribirse de manera detallada como

$$\underbrace{\begin{bmatrix} Y_{t1} \\ Y_{t2} \\ \vdots \\ Y_{tM} \end{bmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \\ \vdots \\ \mu_{Y_M} \end{bmatrix}}_{\boldsymbol{\mu}_Y} + \underbrace{\sum_{k=0}^{\infty} \begin{bmatrix} \psi_k(11) & \psi_k(12) & \cdots & \psi_k(1M) \\ \psi_k(21) & \psi_k(22) & \cdots & \psi_k(2M) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_k(M1) & \psi_k(M2) & \cdots & \psi_k(MM) \end{bmatrix}}_{\boldsymbol{\Psi}(B)\mathbf{A}_t} \begin{bmatrix} A_{t-k,1} \\ A_{t-k,2} \\ \vdots \\ A_{t-k,M} \end{bmatrix}, \quad (3.24)$$

lo que indica que $\psi_k(ij) = \partial Y_{ti} / \partial A_{t-k,j} = \partial Y_{t+k,i} / \partial A_{tj}$ ($k = 0, 1, 2, \dots$).

3.4.7 Definición

La **IRF** de Y_{ti} ($1 \leq i \leq M$) ante un **impulso unitario** en A_{tj} ($1 \leq j \leq M$) es la secuencia $(\partial Y_{t+k,i} / \partial A_{tj}) = (\psi_k(ij))$ para $k = 0, 1, 2, \dots$. De forma compacta, la IRF de \mathbf{Y}_t ante impulsos unitarios en cada elemento de \mathbf{A}_t es $(\partial \mathbf{Y}_{t+k} / \partial \mathbf{A}_t') = (\boldsymbol{\Psi}_k)$ para $k = 0, 1, 2, \dots$

3.4.8 Definición

La **IRF** de Y_{ti} ($1 \leq i \leq M$) ante un **impulso de cuantía** $\sigma_j \equiv \text{DT}[A_{tj}]$ en A_{tj} ($1 \leq j \leq M$) es la secuencia $(\psi_{k(ij)}\sigma_j)$ para $k = 0, 1, 2, \dots$. De forma compacta, la IRF de \mathbf{Y}_t ante impulsos de cuantía $\sigma_1, \sigma_2, \dots, \sigma_M$ en cada elemento de \mathbf{A}_t es $(\Psi_k \mathbf{\Lambda})$ para $k = 0, 1, 2, \dots$, donde $\mathbf{\Lambda} \equiv \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_M]$.

Observación: En 3.4.7, $\psi_{k(ij)}$ representa la consecuencia de un aumento igual a 1 en A_{tj} (la perturbación asociada con la variable n° j del modelo VAR en el momento t) sobre $Y_{t+k,i}$ (la variable n° i del modelo VAR en el momento $t+k$, $k \geq 0$), suponiendo que todos los elementos de (\mathbf{A}_t) que no son A_{tj} no varían. En 3.4.8, $\psi_{k(ij)}\sigma_j$ representa esencialmente lo mismo (salvo porque en 3.4.8 el aumento en A_{tj} es igual a σ_j en lugar de 1). En ambos casos, el hecho de suponer que todos los elementos de (\mathbf{A}_t) que no son A_{tj} no varían cuando A_{tj} sí varía, resulta poco creíble cuando al menos una de las covarianzas contemporáneas entre A_{tj} y los otros componentes de \mathbf{A}_t es distinta de cero.

3.4.9 Definición

La matriz de varianzas-covarianzas de (\mathbf{A}_t) en un modelo VAR(p) puede factorizarse como $\Sigma_{\mathbf{A}} = \mathbf{L}\mathbf{L}'$, donde \mathbf{L} es una matriz triangular inferior (**factorización de Cholesky**), de manera que $\mathbf{L}^{-1}\Sigma_{\mathbf{A}}\mathbf{L}^{-1'} = \mathbf{I}$. Entonces, la representación (3.16) puede escribirse como $\mathbf{Y}_t = \boldsymbol{\mu}_{\mathbf{Y}} + \sum_{k=0}^{\infty} \Psi_k \mathbf{L}\mathbf{L}^{-1} \mathbf{A}_{t-k} = \boldsymbol{\mu}_{\mathbf{Y}} + \sum_{k=0}^{\infty} \Psi_k^* \mathbf{A}_{k-i}^*$, con $(\mathbf{A}_t^*) \equiv (\mathbf{L}^{-1} \mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \mathbf{I})$,

de manera que los M componentes de \mathbf{A}_t^* son **ortogonales** entre sí. En este contexto, la **IRF** de Y_{ti} ($1 \leq i \leq M$) ante un **impulso unitario** en A_{tj}^* ($1 \leq j \leq M$) es la secuencia $(\partial Y_{t+k,i} / \partial A_{tj}^*)$ para $k = 0, 1, 2, \dots$. De forma compacta, la IRF de \mathbf{Y}_t ante impulsos unitarios en cada elemento de \mathbf{A}_t^* es $(\partial \mathbf{Y}_{t+k} / \partial \mathbf{A}_t^{*'}) = (\Psi_k^*) \equiv (\Psi_k \mathbf{L})$ para $k = 0, 1, 2, \dots$

Observación 1: En 3.4.9, el elemento en la posición (i, j) de Ψ_k^* representa la consecuencia de un aumento unitario en A_{tj}^* (una perturbación ortogonal asociada con la variable n° j del modelo VAR en el momento t) sobre $Y_{t+k,i}$ (la variable n° i del modelo VAR en el momento $t+k$, $k \geq 0$), suponiendo que todos los elementos de (\mathbf{A}_t^*) que no son A_{tj}^* no varían; ahora este supuesto sí es razonable, ya que las covarianzas contemporáneas entre A_{tj}^* y los demás componentes de \mathbf{A}_t^* son todas iguales a cero.

Observación 2: La ortogonalización de las perturbaciones de un modelo VAR a través de la factorización de Cholesky de $\Sigma_{\mathbf{A}}$, tiene el inconveniente de que dicha factorización **no** es única, ya que depende crucialmente del **orden** en el que figuran los componentes de (\mathbf{Y}_t) ; ver Observación 3 en 3.1.3 y Observación 2 en 3.4.2.

Descomposición de las varianzas de los errores de previsión

De acuerdo con la factorización $\Sigma_{\mathbf{A}} = \mathbf{L}\mathbf{L}'$ de 3.4.9, la matriz de varianzas-covarianzas del error de previsión a horizonte $l \geq 1$ (ver Observación 1 en 3.4.5) puede escribirse como

$$\Sigma(l) = \sum_{k=0}^{l-1} \Psi_k \Sigma_{\mathbf{A}} \Psi_k' = \sum_{k=0}^{l-1} \Psi_k^* \Psi_k^{*'} \quad (3.25)$$

de manera que cada elemento en la diagonal principal de $\Sigma(l)$ es

$$\sigma_{ii}(l) = \sum_{k=0}^{l-1} \left(\sum_{j=1}^M \psi_{k(ij)}^{*2} \right) = \sum_{j=1}^M \left(\sum_{k=0}^{l-1} \psi_{k(ij)}^{*2} \right), \quad (3.26)$$

donde

$$\sum_{k=0}^{l-1} \psi_{k(ij)}^{*2} = \psi_{0(ij)}^{*2} + \psi_{1(ij)}^{*2} + \dots + \psi_{l-1(ij)}^{*2} \quad (3.27)$$

representa la contribución de la perturbación ortogonal A_{tj}^* ($1 \leq j \leq M$) a la varianza del error de previsión a horizonte $l \geq 1$ del proceso (Y_{ti}) ($1 \leq i \leq M$).

Dado que $\Sigma(l) \rightarrow \sum_{i=0}^{\infty} \Psi_i \Sigma_A \Psi_i' = \Sigma_Y$ (ver Observación 1 de 3.4.5) cuando l tiende a infinito, (3.27) permite evaluar la contribución de A_{tj}^* ($1 \leq j \leq M$) a la varianza del proceso (Y_{ti}) ($1 \leq i \leq M$) cuando l es suficientemente grande en (3.27).

3.4.10 Ejemplo

Continuando con el Ejemplo 3.4.2, la Tabla 3.2 contiene los resultados de los contrastes de causalidad de Granger para las hipótesis nulas: (i) (X_t) no causa a (Y_t) , (ii) (Y_t) no causa a (X_t) . La hipótesis (i) se rechaza claramente; la hipótesis (ii) **no** puede rechazarse al 1%.

VAR Pairwise Granger Causality/Block Exogeneity Wald Tests			
Sample: 1959 2002			
Included observations: 42			
Dependent variable: Y			
Exclude	Chi-sq	df	Prob.
X	12.55127	1	0.0004
All	12.55127	1	0.0004
Dependent variable: X			
Exclude	Chi-sq	df	Prob.
Y	5.963297	1	0.0146
All	5.963297	1	0.0146

Tabla 3.2

Contrastes de causalidad de Granger a partir del modelo VAR(1) del Ejemplo 3.4.2

A partir del modelo VAR(1) estimado en (3.15) (ver también Tabla 1), resulta que:

$$\begin{aligned} \hat{\Psi}_0 &\equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \hat{\Psi}_1 = \begin{bmatrix} 0.2536 & -1.2220 \\ 0.1410 & 0.2751 \end{bmatrix}, \hat{\Psi}_2 = \begin{bmatrix} -0.1080 & -0.6461 \\ 0.0746 & -0.0967 \end{bmatrix}, \\ \hat{\Psi}_3 &= \begin{bmatrix} -0.1185 & -0.0457 \\ 0.0053 & -0.1177 \end{bmatrix}, \hat{\Psi}_4 = \begin{bmatrix} -0.0365 & 0.1323 \\ -0.0153 & -0.0388 \end{bmatrix}, \dots \end{aligned} \quad (3.28)$$

donde $\hat{\Psi}_k = \hat{\Phi}_1^k$ para todo $k \geq 0$; ver Observación en Ejemplo 3.4.6. Por otro lado,

$$Y \ X: \quad \hat{\Sigma}_A = \begin{bmatrix} 14.9739 & -2.7168 \\ -2.7168 & 2.8285 \end{bmatrix} = \hat{\mathbf{L}}\hat{\mathbf{L}}' = \begin{bmatrix} 3.8696 & 0 \\ -0.7021 & 1.5283 \end{bmatrix} \begin{bmatrix} 3.8696 & -0.7021 \\ 0 & 1.5283 \end{bmatrix} \quad (3.29)$$

Si se cambia el orden en el que figuran los dos componentes de (\mathbf{Y}_t) , entonces

$$X \ Y: \quad \hat{\Sigma}_A = \begin{bmatrix} 2.8285 & -2.7168 \\ -2.7168 & 14.9739 \end{bmatrix} = \hat{\mathbf{L}}\hat{\mathbf{L}}' = \begin{bmatrix} 1.6818 & 0 \\ -1.6154 & 3.5163 \end{bmatrix} \begin{bmatrix} 1.6818 & -1.6154 \\ 0 & 3.5163 \end{bmatrix} \quad (3.30)$$

La Tabla 3.3 contiene las funciones de respuesta a impulsos (estimadas) consideradas en 3.4.7, 3.4.8 (con $\hat{\sigma}_1 = 14.9739^{1/2} = 3.8696$, $\hat{\sigma}_2 = 2.8285^{1/2} = 1.6818$) y 3.4.9 [con los dos órdenes posibles representados en (3.29) y (3.30)], respectivamente.

La Tabla 3.4 contiene las proporciones (estimadas) de las varianzas de los errores de previsión de (Y_t) y de (X_t) asociadas con cada perturbación ortogonal en cada uno de los dos órdenes posibles representados en (3.29) y (3.30).

Observación 1: Por ejemplo, la varianza (estimada) del error de previsión a horizonte $l = 3$ de (Y_t) es

$$\hat{\sigma}_Y^2(3) = \left[\hat{\psi}_{0(11)}^{*2} + \hat{\psi}_{1(11)}^{*2} + \hat{\psi}_{2(11)}^{*2} \right] + \left[\hat{\psi}_{0(12)}^{*2} + \hat{\psi}_{1(12)}^{*2} + \hat{\psi}_{2(12)}^{*2} \right] \quad (I)$$

[ver (3.2.6)-(3.2.7) con $i = 1$, $l = 3$], donde $\hat{\psi}_{k(11)}^*$ y $\hat{\psi}_{k(12)}^*$ son los elementos (1,1) y (1,2) de $\hat{\Psi}_k^* \equiv \hat{\Psi}_k \hat{\mathbf{L}}$ ($k = 0, 1, 2$),

respectivamente, $\hat{\Psi}_k$ ($k = 0, 1, 2$) figuran en (3.28), y $\hat{\mathbf{L}}$ figura en (3.29):

$$Y \ X: \quad \hat{\Psi}_0^* = \begin{bmatrix} 3.8696 & 0 \\ -0.7021 & 1.5283 \end{bmatrix}, \hat{\Psi}_1^* = \begin{bmatrix} 1.8393 & -1.8675 \\ 0.3526 & 0.4205 \end{bmatrix}, \hat{\Psi}_2^* = \begin{bmatrix} 0.0356 & -0.9874 \\ 0.3564 & -0.1477 \end{bmatrix}, \quad (\text{II})$$

En este caso, (I) queda

$$\hat{\sigma}_Y^2(3) = \underbrace{\left[3.8696^2 + 1.8393^2 + 0.0356^2 \right]}_{Y: 18.3581 \ (80.44\%)} + \underbrace{\left[0^2 + (-1.8675)^2 + (-0.9874)^2 \right]}_{X: 4.4625 \ (19.56\%)} = 4.7771^2.$$

Alternativamente, cambiando el orden en el que figuran Y_t y X_t en el modelo, la varianza (estimada) del error de previsión a horizonte $l = 3$ de (Y_t) también puede calcularse como

$$\hat{\sigma}_Y^2(3) = \left[\hat{\psi}_{0(21)}^{*2} + \hat{\psi}_{1(21)}^{*2} + \hat{\psi}_{2(21)}^{*2} \right] + \left[\hat{\psi}_{0(22)}^{*2} + \hat{\psi}_{1(22)}^{*2} + \hat{\psi}_{2(22)}^{*2} \right] \quad (\text{III})$$

[ver (3.2.6)-(3.27) con $i = 2$, $l = 3$], donde ahora $\hat{\Psi}_k$ ($k = 0, 1, 2$) son las matrices de (3.28) con sus elementos cambiados de orden adecuadamente y $\hat{\mathbf{L}}$ figura en (3.30):

$$X \ Y: \quad \hat{\Psi}_0^* = \begin{bmatrix} 1.6818 & 0 \\ -1.6154 & 3.5163 \end{bmatrix}, \hat{\Psi}_1^* = \begin{bmatrix} 0.2349 & 0.4959 \\ -2.4649 & 0.8918 \end{bmatrix}, \hat{\Psi}_2^* = \begin{bmatrix} -0.2830 & 0.2622 \\ -0.9121 & -0.3799 \end{bmatrix}, \quad (\text{IV})$$

En este caso, (III) queda

$$\hat{\sigma}_Y^2(3) = \underbrace{\left[(-1.6154)^2 + (-2.4649)^2 + (-0.9121)^2 \right]}_{X: 9.5172 \ (41.70\%)} + \underbrace{\left[3.5163^2 + 0.8918^2 + (-0.3799)^2 \right]}_{Y: 13.3040 \ (58.30\%)} = 4.7771^2.$$

Response of Y			Response of Y			Response of Y			Response of Y		
LAG	Y	X	LAG	Y	X	LAG	Y	X	LAG	Y	X
0	1.000000	0.000000	0	3.869608	0.000000	0	3.869608	0.000000	0	3.516299	-1.615398
1	0.253606	-1.222001	1	0.981356	-2.055185	1	1.839308	-1.867539	1	0.891754	-2.464859
2	-0.108035	-0.646113	2	-0.418054	-1.086646	2	0.035575	-0.987432	2	-0.379884	-0.912126
3	-0.118526	-0.045745	3	-0.458650	-0.076935	3	-0.426533	-0.069910	3	-0.416774	0.114533
4	-0.036511	0.132254	4	-0.141283	0.222427	4	-0.234136	0.202118	4	-0.128383	0.281406
5	0.009394	0.081003	5	0.036350	0.136232	5	-0.020521	0.123794	5	0.033031	0.121058
6	0.013807	0.010807	6	0.053428	0.018176	6	0.045840	0.016516	6	0.048549	-0.004128
7	0.005026	-0.013899	7	0.019448	-0.023375	7	0.029206	-0.021241	7	0.017672	-0.031494
8	-0.000686	-0.009965	8	-0.002653	-0.016760	8	0.004343	-0.015230	8	-0.002411	-0.015652
9	-0.001579	-0.001904	9	-0.006112	-0.003202	9	-0.004775	-0.002910	9	-0.005554	-0.000650
Response of X			Response of X			Response of X			Response of X		
LAG	Y	X	LAG	Y	X	LAG	Y	X	LAG	Y	X
0	0.000000	1.000000	0	0.000000	1.681820	0	-0.702089	1.528263	0	0.000000	1.681820
1	0.141040	0.275128	1	0.545770	0.462716	1	0.352606	0.420468	1	0.495939	0.234880
2	0.074573	-0.096656	2	0.288567	-0.162558	2	0.356428	-0.147715	2	0.262220	-0.283022
3	0.005280	-0.117721	3	0.020431	-0.197985	3	0.103081	-0.179908	3	0.018565	-0.206514
4	-0.015264	-0.038840	4	-0.059067	-0.065322	4	-0.031798	-0.059358	4	-0.053674	-0.040664
5	-0.009349	0.007967	5	-0.036178	0.013399	5	-0.041771	0.012176	5	-0.032874	0.028502
6	-0.001247	0.013617	6	-0.004827	0.022901	6	-0.014387	0.020810	6	-0.004386	0.024916
7	0.001604	0.005271	7	0.006207	0.008864	7	0.002507	0.008055	7	0.005641	0.006273
8	0.001150	-0.000510	8	0.004451	-0.000858	8	0.004809	-0.000780	8	0.004044	-0.002716
9	0.000220	-0.001546	9	0.000850	-0.002600	9	0.001936	-0.002363	9	0.000773	-0.002955
Nonfactorized One Unit			Nonfactorized One Std.Dev.			Cholesky – Order Y X			Cholesky – Order X Y		

Tabla 3.3

Funciones de respuesta a impulsos estimadas a partir del modelo VAR(1) del Ejemplo 3.4.2

Variance Decomposition of Y				Variance Decomposition of Y			
HORIZON	S.E.	Y	X	HORIZON	S.E.	Y	X
1	3.869608	100.0000	0.000000	1	3.869608	82.57291	17.42709
2	4.673823	84.03405	15.96595	2	4.673823	60.24175	39.75825
3	4.777124	80.44459	19.55541	3	4.777124	58.29695	41.70305
4	4.796637	80.58213	19.41787	4	4.796637	58.57856	41.42144
5	4.806600	80.48572	19.51428	5	4.806600	58.40732	41.59268
6	4.808237	80.43273	19.56727	6	4.808237	58.37226	41.62774
7	4.808484	80.43356	19.56644	7	4.808484	58.37646	41.62354
8	4.808620	80.43271	19.56729	8	4.808620	58.37452	41.62548
9	4.808646	80.43192	19.56808	9	4.808646	58.37391	41.62609
10	4.808649	80.43191	19.56809	10	4.808649	58.37397	41.62603

Variance Decomposition of X				Variance Decomposition of X:			
HORIZON	S.E.	Y	X	HORIZON	S.E.	Y	X
1	1.681820	17.42709	82.57291	1	1.681820	0.000000	100.0000
2	1.769079	19.72300	80.27700	2	1.769079	7.858915	92.14108
3	1.810664	22.70245	77.29755	3	1.810664	9.599360	90.40064
4	1.822497	22.72850	77.27150	4	1.822497	9.485485	90.51452
5	1.823741	22.72791	77.27209	5	1.823741	9.559170	90.44083
6	1.824260	22.76741	77.23259	6	1.824260	9.586207	90.41379
7	1.824435	22.76925	77.23075	7	1.824435	9.584941	90.41506
8	1.824455	22.76896	77.23104	8	1.824455	9.585692	90.41431
9	1.824461	22.76949	77.23051	9	1.824461	9.586115	90.41388
10	1.824464	22.76954	77.23046	10	1.824464	9.586106	90.41389

Cholesky – Order Y X

Cholesky – Order X Y

Tabla 3.4

Proporciones de las varianzas de los errores de previsión estimadas a partir del modelo VAR(1) del Ejemplo 3.4.2

Observación 2: En general, en cualquier modelo VAR bivalente, la descomposición de la varianza de los errores de previsión describe la proporción de la dinámica de cada componente del modelo debida a sus propias perturbaciones

frente a las perturbaciones del otro componente. Si, por ejemplo, las perturbaciones de (X_t) explican una proporción insignificante de la varianza de los errores de previsión de (Y_t) , puede decirse que (Y_t) es una variable esencialmente **exógena** en el modelo considerado; en este caso, (Y_t) evoluciona independientemente de (X_t) . Por el contrario, si las perturbaciones de (X_t) explican una proporción elevada de la varianza de los errores de previsión de (Y_t) , (Y_t) es una variable esencialmente **endógena**.

4

MODELOS MULTIVARIANTES NO ESTACIONARIOS

4.1 INTRODUCCIÓN

4.1.1 Definición

Un proceso estocástico multivariante (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, es **integrado de orden d** ($d \geq 0$ entero), o **$\mathbf{I}(d)$** , si y sólo si $(Y_{ti}) \sim I(d)$ (ver 2.4.5) para todo $i = 1, 2, \dots, M$. En tal caso, suele escribirse $(\mathbf{Y}_t) \sim \mathbf{I}(d)$.

Observación 1: El único tipo de no estacionariedad que se contempla en este Tema 4 es la no estacionariedad homogénea implicada por el hecho de que cada componente $(Y_{t1}), (Y_{t2}) \dots, (Y_{tM})$ de (\mathbf{Y}_t) sea integrado de orden $d \geq 1$; ver 2.4.6-2.4.9. En particular, se supone cada componente de (\mathbf{Y}_t) puede representar una transformación de Box-Cox (ver 2.4.1-2.4.3) de alguna característica o variable original de interés, de manera que $(Y_{t1}), (Y_{t2}) \dots, (Y_{tM})$ son procesos univariantes que únicamente requieren $d \geq 1$ diferencias regulares para hacerlos estacionarios.

Observación 2: Una posibilidad conservadora a la hora de elaborar un modelo de relación entre los componentes de un proceso $(\mathbf{Y}_t) \sim I(d)$, consiste en plantear un modelo estacionario (ver Tema 3) para $(\nabla^d \mathbf{Y}_t) \sim I(0)$. Sin embargo, en algunos casos especiales este planteamiento puede llevar a ignorar la existencia de ciertas relaciones importantes entre los componentes no estacionarios de (\mathbf{Y}_t) ; esto ocurre cuando dichos componentes están **cointegrados**.

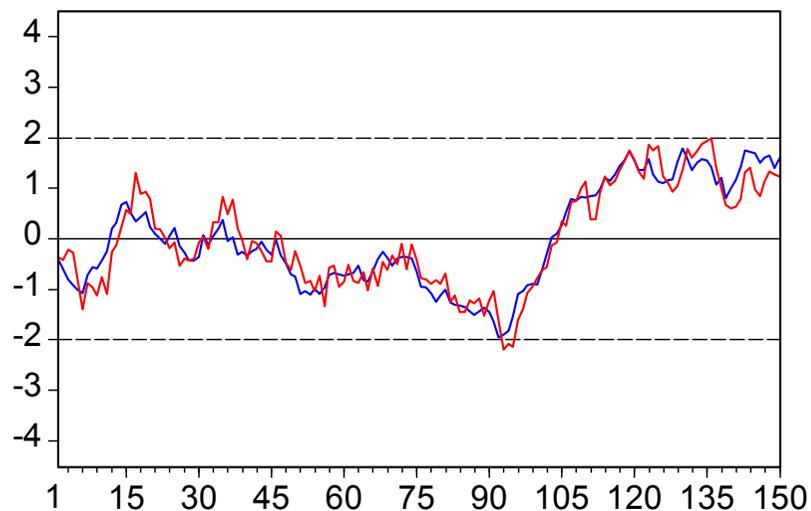
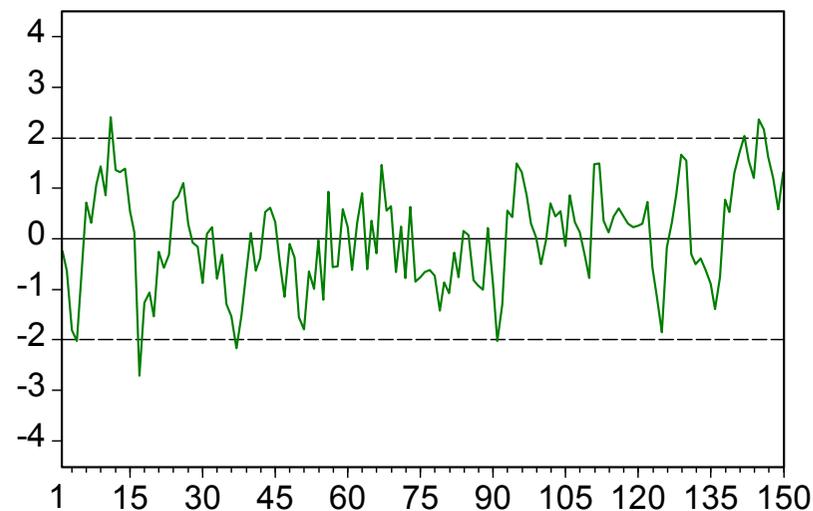
4.1.2 Definición

Un proceso multivariante $(\mathbf{Y}_t) \sim I(d)$, $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, presenta **cointegración** de orden (d, b) (con $1 \leq b \leq d$) si y sólo si existe al menos un **vector de cointegración** $\boldsymbol{\beta} \equiv [\beta_1, \beta_2, \dots, \beta_M]'$ con al menos dos de sus componentes distintos de cero, tal que el **proceso de error** $(U_t) \equiv (\beta_0 + \boldsymbol{\beta}'\mathbf{Y}_t) \sim I(d - b)$; en tal caso, se escribe $(\mathbf{Y}_t) \sim CI(d, b)$.

Observación 1: La cointegración se refiere a $M \geq 2$ procesos univariantes integrados del **mismo orden**. En particular, si $(Y_{t1}) \sim I(d_1)$ e $(Y_{t2}) \sim I(d_2)$, con $d_1 \neq d_2$, entonces (Y_{t1}) e (Y_{t2}) **no** pueden estar cointegrados; en este caso puede comprobarse que cualquier combinación lineal de (Y_{t1}) e (Y_{t2}) es un proceso integrado de orden $\max\{d_1, d_2\}$.

Observación 2: La cointegración se refiere a una combinación **lineal** de $M \geq 2$ procesos $I(d)$ (en la que el término constante β_0 puede perfectamente ser igual a cero), aunque podrían concebirse situaciones en las que existieran funciones **no lineales** de dichos procesos que fuesen $I(d - b)$. Por otro lado, como el orden de integración de un proceso (U_t) no varía si se multiplica (U_t) por cualquier constante distinta de cero, un vector de cointegración **no** es único en el sentido que si $\boldsymbol{\beta}$ es un vector de cointegración, también lo es $c\boldsymbol{\beta} = [c\beta_1, c\beta_2, \dots, c\beta_M]'$ para cualquier $c \neq 0$; para identificar de forma única un vector de cointegración, suele imponerse la condición $\beta_1 \equiv 1$ (porque, por ejemplo, el primer componente de (\mathbf{Y}_t) es especialmente relevante por algún motivo).

4.1.3 Ejemplo – Series $I(1)$ cointegradas y no cointegradas

PANEL A - SERIES $I(1)$ COINTEGRADAS

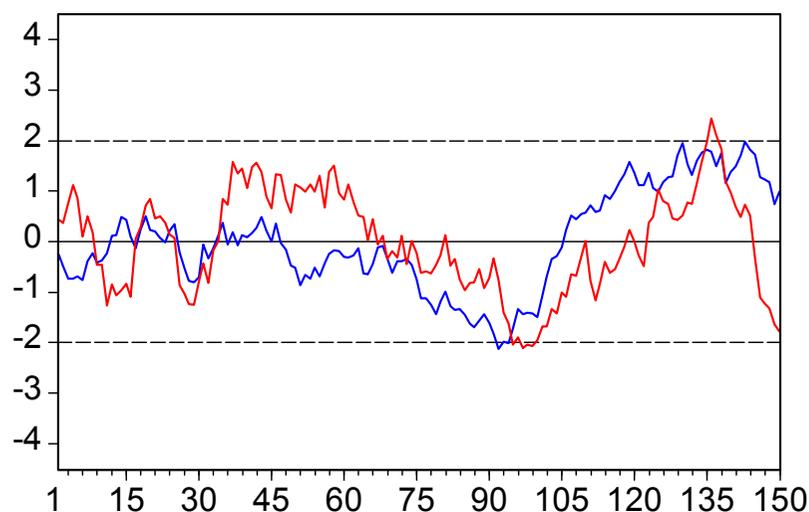
PANEL B - DESEQUILIBRIO ESTACIONARIO

Figura 4.1

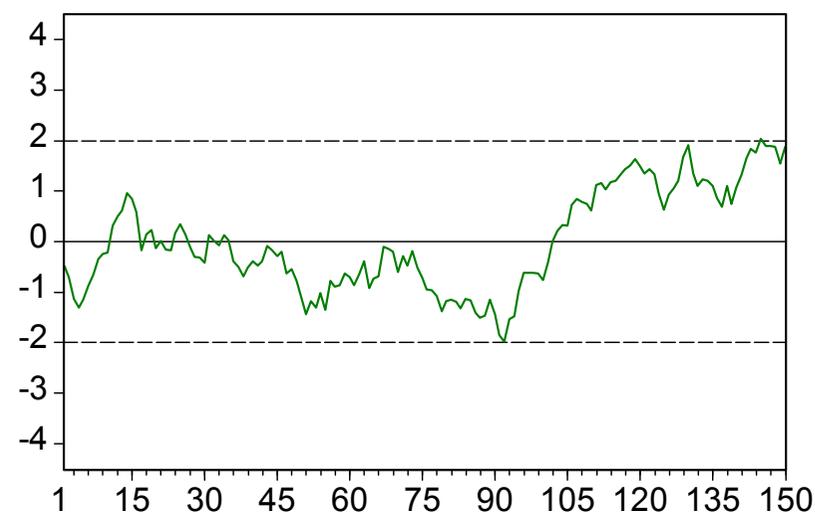
Dos series simuladas $I(1)$ cointegradas
(sim-series.wf1)

Observación 1: Las dos series del Panel A de la Figura 4.1 son no estacionarias, pero tienden a moverse juntas porque comparten una **tendencia estocástica común**; en otros términos, dichas series están ligadas por una **relación estable** o de **equilibrio a largo plazo**. En este caso, existe al menos una combinación lineal de ambas series que es una serie $I(0)$

(estacionaria); en particular, la serie del Panel B es $(\hat{u}_t) = (-\hat{\gamma}_1 + [1, -\hat{\gamma}_2]\mathbf{y}_t)$, donde $\mathbf{y}_t \equiv [y_{t1}, y_{t2}]'$ contiene las dos series del Panel A y $\hat{\gamma}_1$, $\hat{\gamma}_2$ son las estimaciones MCO del término constante y de la pendiente en la regresión lineal de y_{t1} sobre y_{t2} . Como $(\hat{u}_t) \sim I(0)$, $(\mathbf{y}_t) \sim CI(1, 1)$ con vector de cointegración $\boldsymbol{\beta} = [\beta_1, \beta_2]' = [1, -\hat{\gamma}_2]'$. La serie (\hat{u}_t) puede interpretarse en este caso como una secuencia de **desequilibrios transitorios** entre (y_{t1}) e (y_{t2}) con respecto a una relación (aproximadamente) estable o de equilibrio a largo plazo del tipo $y_{t1} - \hat{\gamma}_2 y_{t2} \approx \hat{\gamma}_1$.



PANEL A - SERIES I(1) NO COINTEGRADAS

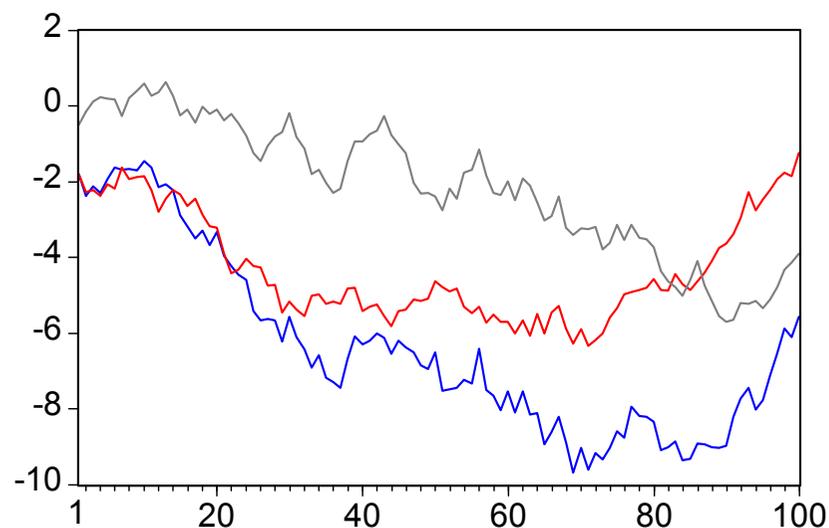
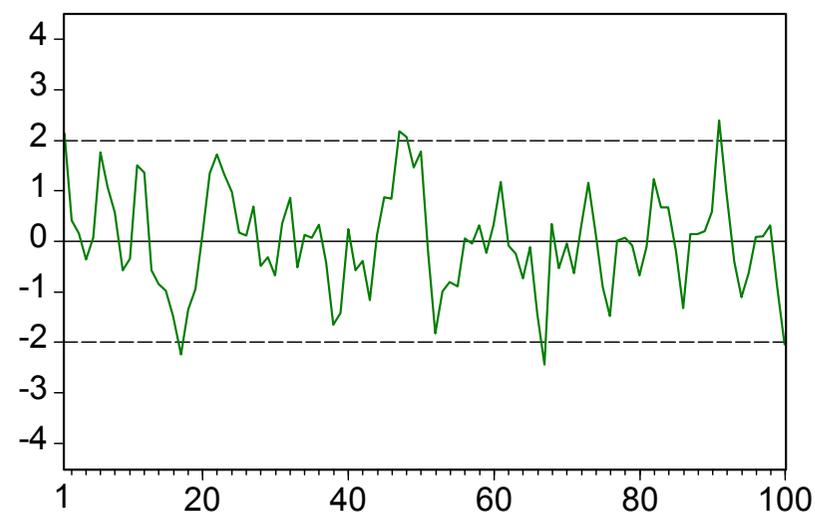


PANEL B - DESEQUILIBRIO NO ESTACIONARIO

Figura 4.2

Dos series simuladas I(1) no cointegradas
(sim-series.wf1)

Observación 2: Las dos series del Panel A de la Figura 4.2 son no estacionarias y deambulan sin presentar tendencia alguna a moverse juntas, por lo que cualquier combinación lineal de las mismas es una serie $I(1)$ (no estacionaria). En particular, la serie de residuos del Panel B de la Figura 4.2 es no estacionaria, por lo que en este caso no existe una relación estable (una relación de cointegración) entre las dos series consideradas.

PANEL A - SERIES $I(1)$ COINTEGRADAS

PANEL B - DESEQUILIBRIO ESTACIONARIO

Figura 4.3

Tres series simuladas $I(1)$ cointegradas
(coint6.wf1)

Observación 3: Las tres series del Panel A de la Figura 4.3 son no estacionarias; además, ningún par de los tres

posibles parece presentar una tendencia común. Sin embargo, las tres series están cointegradas. La serie del Panel B de la Figura 4.3 es la serie de residuos $(\hat{u}_t) = (-\hat{\gamma}_1 + [1, -\hat{\gamma}_2, -\hat{\gamma}_3]\mathbf{y}_t)$, donde $\mathbf{y}_t \equiv [y_{t1}, y_{t2}, y_{t3}]'$ contiene las tres series del Panel A, y $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ son las estimaciones MCO del término constante y de las pendientes, respectivamente, en la regresión lineal de y_{t1} sobre y_{t2} e y_{t3} ; como $(\hat{u}_t) \sim I(0)$, $(\mathbf{y}_t) \sim CI(1,1)$ con vector de cointegración $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]' = [1, -\hat{\gamma}_2, -\hat{\gamma}_3]'$. En consecuencia, la dinámica de al menos una de las tres series consideradas está determinada por la dinámica de las otras dos. Como en la Observación 1, la serie (\hat{u}_t) puede interpretarse en este caso como una secuencia de desequilibrios transitorios entre (y_{t1}) , (y_{t2}) e (y_{t3}) con respecto a una relación (aproximadamente) estable o de equilibrio a largo plazo del tipo $y_{t1} - \hat{\gamma}_2 y_{t2} - \hat{\gamma}_3 y_{t3} \approx \hat{\gamma}_1$.

Observación 4: En Econometría, es bastante frecuente plantear modelos de relación entre procesos no estacionarios del mismo tipo que los modelos para procesos estacionarios considerados en el Tema 3 (especialmente en las secciones 3.2 y 3.3). En general, para que un modelo de alguna teoría sobre relaciones estables entre procesos no estacionarios tenga sentido, es necesario que exista alguna relación de cointegración entre dichos procesos; en caso contrario, el modelo carece de interés y sólo puede dar lugar a la estimación de **relaciones espurias**. La Sección 4.2, las figuras 1.16-1.18, y los ejemplos 1.4.3-1.4.4 del Tema 1, contienen algunas ilustraciones de esta cuestión.

4.1.4 Ejemplo – Un modelo VAR(1) para un proceso bivalente cointegrado

Sea (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$, un proceso bivalente que sigue el modelo VAR(1)

$$\begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \begin{bmatrix} 1.2 & -0.2 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}. \quad (4.1)$$

(4.1) es un modelo para $(\mathbf{Y}_t) \sim I(1)$, **restringido** a que $(\mathbf{Y}_t) \sim CI(1,1)$. Para verlo, nótese en primer lugar que los dos autovalores de Φ_1 son $\lambda_1 = 1$ y $\lambda_2 = 0.6 < 1$; por lo tanto,

$$\Phi_1 = \begin{bmatrix} 1.2 & -0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} -0.5 & -0.5 \\ -0.5 & -1.5 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.6 \end{bmatrix} \begin{bmatrix} -3.0 & 1.0 \\ 1.0 & -1.0 \end{bmatrix} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}. \quad (4.2)$$

[Los dos autovectores en las columnas de \mathbf{P} no son únicos.] (4.1) puede escribirse como

$$\underbrace{\mathbf{P}^{-1}\mathbf{Y}_t}_{\mathbf{Z}_t} = \underbrace{\mathbf{P}^{-1}\Phi_1\mathbf{P}}_{\mathbf{\Lambda}} \underbrace{\mathbf{P}^{-1}\mathbf{Y}_{t-1}}_{\mathbf{Z}_{t-1}} + \underbrace{\mathbf{P}^{-1}\mathbf{A}_t}_{\mathbf{U}_t}, \quad (4.3)$$

es decir,

$$\underbrace{\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix}}_{\mathbf{Z}_t} = \underbrace{\begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.6 \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{bmatrix}}_{\mathbf{Z}_{t-1}} + \underbrace{\begin{bmatrix} U_{t1} \\ U_{t2} \end{bmatrix}}_{\mathbf{U}_t}, \quad (4.4)$$

de manera que $(1 - B)Z_{t1} = U_{t1}$, $(1 - 0.6B)Z_{t2} = U_{t2}$. En consecuencia, $(Z_{t1}) \sim I(1)$ (un paseo aleatorio) y $(Z_{t2}) \sim I(0)$ [que sigue un modelo AR(1) estacionario]. Dado que (4.3) implica que $\mathbf{Y}_t = \mathbf{P}\mathbf{Z}_t$, o bien

$$\begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \begin{bmatrix} -0.5 & -0.5 \\ -0.5 & -1.5 \end{bmatrix} \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} = -0.5 \begin{bmatrix} Z_{t1} + Z_{t2} \\ Z_{t1} + 3Z_{t2} \end{bmatrix}, \quad (4.5)$$

entonces $(\mathbf{Y}_t) \sim I(1)$, ya que sus dos componentes son combinaciones lineales de un proceso $I(1)$ (un paseo aleatorio) y de un proceso $I(0)$ [que sigue un modelo $AR(1)$ estacionario].

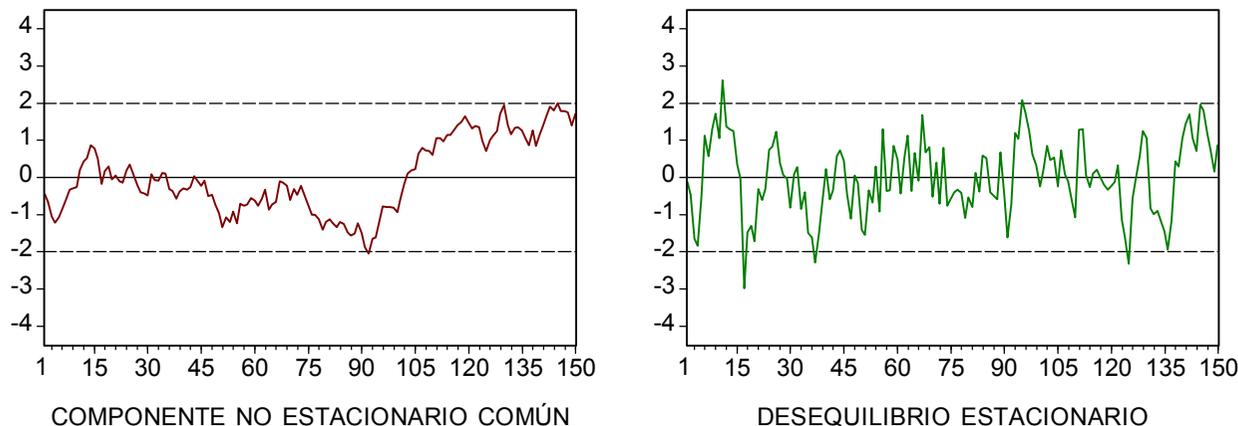
Recíprocamente, (4.3) también implica que $\mathbf{Z}_t = \mathbf{P}^{-1}\mathbf{Y}_t$, o bien

$$\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} = \begin{bmatrix} -3.0 & 1.0 \\ 1.0 & -1.0 \end{bmatrix} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \begin{bmatrix} -3Y_{t1} + Y_{t2} \\ Y_{t1} - Y_{t2} \end{bmatrix}, \quad (4.6)$$

cuya segunda fila y la segunda fila de (4.4) implican que $(Z_{t2}) = (Y_{t1} - Y_{t2}) \sim I(0)$, es decir, $(\mathbf{Y}_t) \sim CI(1,1)$ con vector de cointegración $\boldsymbol{\beta} = [1, -1]'$.

Observación 1: De acuerdo con la primera fila de (4.4)-(4.6), $(-0.5Z_{t1}) = (1.5Y_{t1} - 0.5Y_{t2}) \sim I(1)$ se puede interpretar como el **componente no estacionario común** en (Y_{t1}) e (Y_{t2}) ; por su parte, $(Z_{t2}) = (Y_{t1} - Y_{t2}) \sim I(0)$ puede interpretarse como el **proceso de error** en una relación estable entre (Y_{t1}) e (Y_{t2}) del tipo $Y_{t1} - Y_{t2} \approx 0$. Las dos series simuladas del Panel A de la Figura 4.1 proceden de un modelo como (4.1) con $(\mathbf{A}_t) \sim NIID(\mathbf{0}, \boldsymbol{\Sigma})$, donde los elementos de $\boldsymbol{\Sigma}$ son $\sigma_1^2 = 4$, $\sigma_2^2 = 10$, $\sigma_{12} = \sigma_{21} = 2$; ver también Figura 1.19 (I) (Panel B). La figura siguiente contiene las series $(-0.5z_{t1}) = (1.5y_{t1} - 0.5y_{t2})$ (izquierda) y $(z_{t2}) = (y_{t1} - y_{t2})$ (derecha) (que es muy semejante a

la serie del Panel B de la Figura 4.1):



Observación 2: El modelo VAR(1) (4.1) puede escribirse como $\nabla \mathbf{Y}_t = -(\mathbf{I} - \Phi_1)\mathbf{Y}_{t-1} + \mathbf{A}_t = -\mathbf{\Pi}\mathbf{Y}_{t-1} + \mathbf{A}_t$, que se denomina un **modelo de corrección de error vectorial** (VEC, del inglés *Vector Error Correction*). Dado que los autovalores de $\mathbf{\Pi} \equiv \mathbf{I} - \Phi_1$ son $1 - \lambda_1 = 0$, $1 - \lambda_2 = 0.4$, y los autovectores de $\mathbf{\Pi} \equiv \mathbf{I} - \Phi_1$ coinciden con los de Φ_1 , el modelo VEC asociado con el modelo VAR(1) (4.1) puede escribirse como $\nabla \mathbf{Y}_t = -\mathbf{P}(\mathbf{I} - \Lambda)\mathbf{P}^{-1}\mathbf{Y}_{t-1} + \mathbf{A}_t$, o bien

$$\begin{bmatrix} \nabla Y_{t1} \\ \nabla Y_{t2} \end{bmatrix} = - \begin{bmatrix} -0.5 & -0.5 \\ -0.5 & -1.5 \end{bmatrix} \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.4 \end{bmatrix} \begin{bmatrix} -3.0 & 1.0 \\ 1.0 & -1.0 \end{bmatrix} \begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix} = - \begin{bmatrix} -0.2 \\ -0.6 \end{bmatrix} [1, -1] \begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}.$$

Es decir, $\nabla \mathbf{Y}_t = -\alpha\beta'\mathbf{Y}_{t-1} + \mathbf{A}_t$; por lo tanto, $\mathbf{\Pi} \equiv \mathbf{I} - \Phi_1$ puede factorizarse como $\mathbf{\Pi} = \alpha\beta'$, donde β es el vector de cointegración, $\beta = [1, -1]'$, y $\alpha = [-0.2, -0.6]'$ contiene los **factores de corrección** o de ajuste de (Y_{t1}) e (Y_{t2}) en cada momento t ante un **desequilibrio** o **error** transitorio, $\beta'\mathbf{Y}_{t-1} = Y_{t-1,1} - Y_{t-1,2}$, que tiene lugar en el momento anterior en la relación de cointegración entre (Y_{t1}) e (Y_{t2}) . Nótese que el **rango** de la matriz $\mathbf{\Pi} = \alpha\beta'$ es igual a **uno**

mientras que su dimensión es igual a 2, lo que indica que existe **una** relación de cointegración entre (Y_{t1}) e (Y_{t2}) .

4.1.5 Definición

El **modelo de corrección de error vectorial** (VEC) asociado con un modelo VAR(p) del tipo

$$\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t, \text{ o bien } \mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i=1}^p \Phi_i \mathbf{Y}_{t-i} + \mathbf{A}_t, \quad (4.7)$$

para un proceso multivariante $(\mathbf{Y}_t) \sim I(d)$, $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, es

$$\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \mathbf{\Pi} \mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{F}_i \nabla \mathbf{Y}_{t-i} + \mathbf{A}_t, \quad (4.8)$$

donde

$$\mathbf{\Pi} \equiv \Phi(1) \equiv \mathbf{I} - \sum_{i=1}^p \Phi_i, \quad \mathbf{F}_i \equiv -\sum_{j=i+1}^p \Phi_j \quad (i = 1, \dots, p-1). \quad (4.9)$$

Cuando $d \geq 1$, el rango de la matriz $\mathbf{\Pi}$ determina el número de relaciones de cointegración entre los componentes de $(\mathbf{Y}_t) \sim I(d)$. [Si el rango de $\mathbf{\Pi}$ es M , entonces $(\mathbf{Y}_t) \sim I(0)$.]

Observación: En general, cuando $(\mathbf{Y}_t) \sim I(1) \sim \text{VAR}(p)$, con $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$, si $(\mathbf{Y}_t) \sim \text{CI}(1, 1)$ entonces el modelo adecuado para (\mathbf{Y}_t) es (4.8) con $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ [$\Rightarrow \text{rango}(\mathbf{\Pi}) = 1$]; en este caso, un modelo VAR estacionario para $(\nabla \mathbf{Y}_t)$ **no** sería adecuado, ya que en dicho modelo se habría omitido el **término de corrección de error** $-\mathbf{\Pi}\mathbf{Y}_{t-1}$.

4.2 RAÍCES UNITARIAS

- ⇒ Cuando un proceso estocástico (\mathbf{Y}_t) es integrado de orden $d \geq 1$ (ver 4.1.1), suele decirse que (\mathbf{Y}_t) tiene d **raíces unitarias** (ver Observación 1 en 2.5.1). Por ejemplo, un proceso univariante $(Y_t) \sim I(1)$ tiene una raíz unitaria porque (Y_t) sigue en este caso un modelo ARIMA($p,1,q$) del tipo $\phi'(B)Y_t = \mu + \theta(B)A_t$, cuyo operador autorregresivo $\phi'(B) \equiv \phi(B)\nabla$ es tal que la ecuación $\phi'(x) = 0$ tiene una raíz igual a 1; en la práctica, muchas series temporales no estacionarias son compatibles con la hipótesis de haber sido generadas por procesos $I(1)$ (por ejemplo, paseos aleatorios; ver 2.4.6-2.4.7).
- ⇒ La problemática fundamental asociada con la elaboración de modelos para procesos con raíces unitarias puede resumirse en dos puntos:
- ⇒ En primer lugar, los estimadores (MCO, MVE, ...) y los estadísticos de contraste (t , F , LR , ...) habituales, **no** siguen distribuciones estándar **ni siquiera asintóticamente**. En general, en modelos para procesos con raíces unitarias, cualquier inferencia realizada sobre la base de dichas distribuciones estándar (t , F , Chi Cuadrado, Normal, ...) es incorrecta y puede llevar a conclusiones erróneas.

- ⇒ En segundo lugar, la diferenciación regular conjunta de todos los componentes de un proceso multivariante $(\mathbf{Y}_t) \sim I(d)$, aunque proporciona un proceso $(\nabla^d \mathbf{Y}_t) \sim I(0)$, **no** es adecuada cuando los componentes de (\mathbf{Y}_t) están cointegrados; ver Ejemplo 4.1.4.
- ⇒ Por lo tanto, es importante disponer en la práctica de procedimientos fiables para: (i) decidir sobre el orden de integración de un proceso univariante [especialmente para distinguir entre procesos $I(0)$ e $I(1)$], (ii) decidir si los componentes de un proceso multivariante están cointegrados, y (iii) hacer inferencias correctas en presencia de procesos con raíces unitarias.

Paseo aleatorio

Uno de los pilares fundamentales en la teoría estadística necesaria para resolver las tres cuestiones planteadas en el punto anterior es el **paseo aleatorio**. Aunque un paseo aleatorio es un proceso $I(1)$ muy sencillo, sus características básicas son esencialmente iguales que las de cualquier proceso integrado de orden $d \geq 1$; ver 2.4.6-2.4.7 y la Observación siguiente.

Observación: Un paseo aleatorio es un proceso estocástico (Y_t) tal que $Y_t = \mu + Y_{t-1} + A_t$, o bien $\nabla Y_t = \mu + A_t$, donde $(A_t) \sim \text{IID}(0, \sigma_A^2)$ (ruido blanco). Un paseo aleatorio puede escribirse a partir de un origen temporal $k < t$

dado como $Y_t = T_t + I_t$, donde $T_t \equiv [(y_k - \mu k) + \sum_{i=1}^{t-k-1} A_{t-i}] + \mu t$ es la **tendencia** de (Y_t) (su **componente permanente**, que tiene una parte estocástica y otra determinista) e $I_t \equiv A_t$ es la **parte irregular** de (Y_t) (su **componente transitorio**). Un paseo aleatorio es un proceso $I(1)$ porque $(\nabla Y_t) \sim \text{ARMA}(0, 0)$; en otros términos, un paseo aleatorio tiene una raíz unitaria porque en el modelo $\phi'(B)Y_t = \mu + A_t$, el operador autorregresivo $\phi'(B) \equiv \nabla \equiv 1 - B$ es tal que $\phi'(x) = 0$ tiene una raíz igual a 1. La forma más sencilla de comprobar que un paseo aleatorio es un proceso no estacionario, consiste en comprobar que su varianza (calculada con respecto a un origen temporal $k < t$ dado) no es constante (de hecho, tiende a infinito cuando $t \rightarrow \infty$; ver 2.4.6).

Por otro lado, cuando (Y_t) es un paseo aleatorio, $E_{t-1}[Y_t] = Y_{t-1}$ (característica que se denomina en inglés *martingale property*). Esta propiedad tiene cierto sentido económico en muchas ocasiones, especialmente en el estudio de diversos mercados financieros; ver, por ejemplo, las figuras 2.30 y 2.32. Un paseo aleatorio es el proceso en tiempo discreto correspondiente a un **proceso de Wiener** (o, en inglés, *Brownian motion*) en tiempo continuo, que juega un papel fundamental en el desarrollo de la teoría asintótica relevante en presencia de procesos con raíces unitarias.

Regresión con variables explicativas no estacionarias

- ⇒ En cualquiera de los modelos considerados en las secciones 3.2 y 3.3, las condiciones de regularidad requeridas para poder hacer inferencia estándar **no** se cumplen cuando alguna de las variables explicativas es un proceso no estacionario.
- ⇒ En concreto, la condición de que la matriz de covarianzas muestrales entre las variables explicativas converja en probabilidad a una matriz finita definida positiva (condición

requerida para que los estimadores MCO y MV sean consistentes y asintóticamente Normales), **no** se cumple cuando alguna de las variables explicativas es un proceso con una raíz unitaria.

Observación: Por ejemplo, si en un modelo como $Y_t = \beta_1 + \beta_2 X_t + U_t$ la variable explicativa es un paseo aleatorio del tipo $X_t = X_{t-1} + A_t$, donde $(A_t) \sim \text{IID}(0, 1)$, entonces $E \left[\sum_{t=1}^N X_t^2 \right] = \frac{1}{2} N(N + 1)$, lo que indica que $\frac{1}{N} \sum_{t=1}^N X_t^2$ no converge en probabilidad a una cantidad finita.

⇒ No obstante, algunos estadísticos habituales (t y F , fundamentalmente) **sí** siguen distribuciones estándar en ciertos modelos con variables explicativas no estacionarias; ver el apartado sobre inferencia en regresiones con series $I(1)$ en la Sección 4.5.

Regresión espuria

⇒ Si (Y_t) , (X_t) son procesos independientes, cabría esperar al estimar un modelo del tipo

$$Y_t = \beta_1 + \beta_2 X_t + U_t, \text{ donde } (U_t) \sim \text{IID}(0, \sigma^2), \quad (4.10)$$

que β_2 resultase no significativo y que el valor calculado del R^2 fuese próximo a cero. No obstante, esto sólo ocurre bajo condiciones muy restrictivas sobre (Y_t) y (X_t) .

- ⇒ En particular, si (Y_t) , (X_t) son paseos aleatorios independientes, el estadístico t para $\beta_2 = 0$ **no** sigue ni una distribución t de Student ni una distribución $N(0,1)$, **incluso asintóticamente**. En su lugar, el valor absoluto del estadístico t tiende a ser cada vez mayor a medida que el tamaño muestral N aumenta; en última instancia (cuando N tiende a infinito), el contraste habitual basado en el estadístico t para $\beta_2 = 0$ llevaría con toda probabilidad a rechazar esta hipótesis nula. Por otro lado, el R^2 no converge en este caso a cero, sino a un número aleatorio positivo distinto para cada muestra.
- ⇒ Cuando una regresión como (4.10) parece indicar la existencia de una relación que en realidad **no** existe, se denomina una **regresión espuria** (en inglés, *spurious regression*); ver las figuras 1.16-1.17 (panel superior).
- ⇒ Aunque existe una completa teoría estadística que justifica rigurosamente la posibilidad de que una regresión como (4.10) sea espuria, resulta más ilustrativo exponer dicha posibilidad mediante algunos **ejercicios de simulación (experimentos de Monte Carlo)**. La Figura 4.4 contiene un resumen de varios ejercicios de este tipo, que están comentados en las observaciones siguientes.

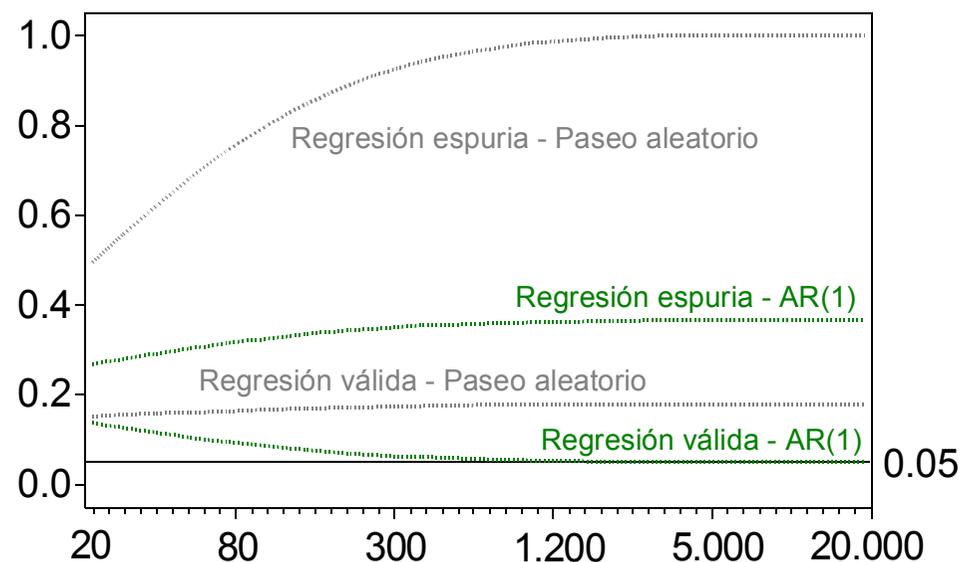


Figura 4.4

Frecuencias de rechazo de ausencia de relación en regresiones espurias y válidas
Adaptado de la Figura 14.1 en Davidson y MacKinnon (2004)

Observación 1: El primer ejercicio de simulación resumido en la Figura 4.4 consiste en llevar a cabo las operaciones de los pasos (1)-(3) siguientes para tamaños muestrales comprendidos entre $N = 20$ y $N = 20.000$: (1) generar un millón de series bivariantes diferentes (\mathbf{y}_t), $\mathbf{y}_t \equiv [y_t, x_t]'$, de tamaño N a partir de un paseo aleatorio bivariante (\mathbf{Y}_t), $\mathbf{Y}_t \equiv (Y_t, X_t)'$, con sus dos componentes **independientes**, (2) estimar (4.10) por MCO un millón de veces [una vez con cada una de las series generadas en (1)] y contrastar cada vez con el estadístico t habitual la hipótesis nula $\beta_2 = 0$ al 5%, (3) registrar el número de rechazos y dividirlo por un millón. La frecuencia (proporción) obtenida en el paso (3) está representada en la curva **Regresión espuria - Paseo aleatorio** de la Figura 4.4 como una función de N . Dicha

frecuencia es elevada incluso para tamaños muestrales pequeños, y tiende claramente a uno a medida que N aumenta. En cierto sentido, esta conclusión no es sorprendente: si $\beta_2 = 0$, (4.10) implica que $(Y_t) \sim \text{IID}(\beta_1, \sigma^2)$; por lo tanto, $\beta_2 = 0$ es una hipótesis falsa en (4.10), y es lógico esperar que un contraste rechace una hipótesis nula falsa incluso si la hipótesis alternativa también es falsa (como ocurre en este caso).

Sin embargo, el fenómeno anterior no se debe sólo a que la hipótesis $\beta_2 = 0$ sea falsa en (4.10). Para verlo, puede plantearse el contraste de $\beta_2 = 0$ en la regresión

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + U_t, \quad (4.11)$$

que, bajo $\beta_2 = 0$, proporciona un modelo válido para (Y_t) (un paseo aleatorio). La frecuencia (proporción) de rechazo de $\beta_2 = 0$ (obtenida mediante una simulación análoga a la descrita anteriormente) está representada en la curva **Regresión válida - Paseo aleatorio** de la Figura 4.4 como una función de N . Aunque dicha frecuencia ahora no tiende a uno, tiende a un valor claramente **mayor** que 0.05. Este **sobrerrechazo** de $\beta_2 = 0$ [una hipótesis nula que es cierta en (4.11)] se debe a la presencia de variables $I(1)$ en (4.11), lo que implica, en general, que el estadístico t no sigue ni una distribución t de Student ni una distribución $N(0,1)$, incluso asintóticamente.

Observación 2: El segundo ejercicio de simulación resumido en la Figura 4.4 consiste en llevar a cabo las operaciones de los pasos (1)-(3) siguientes para tamaños muestrales comprendidos entre $N = 20$ y $N = 20.000$: (1) generar un millón de series bivariantes diferentes (\mathbf{y}_t) , $\mathbf{y}_t \equiv [y_t, x_t]'$, de tamaño N a partir de un modelo VAR(1) bivalente del tipo $(1 - 0.8B)Y_{t1} \equiv A_{t1}$, $(1 - 0.8B)Y_{t2} \equiv A_{t2}$, con (A_{t1}) , (A_{t2}) **independientes**, (2) estimar (4.10) y (4.11) por MCO un millón de veces [una vez con cada una de las series generadas en (1)] y contrastar cada vez en cada modelo con el estadístico t habitual la hipótesis nula $\beta_2 = 0$ al 5%, (3) registrar el número de rechazos en cada caso y dividirlo por un millón. Las frecuencias (proporciones) obtenidas en el paso (3) están representadas en las curvas **Regresión espuria - AR(1)** [para el modelo (4.10)] y **Regresión válida - AR(1)** [para el modelo (4.11)] de la Figura 4.4 como

funciones de N . En este segundo ejercicio, todas las variables que figuran en (4.10) y (4.11) son $I(0)$. Sin embargo, la frecuencia de rechazo de $\beta_2 = 0$ en (4.10) tiende a un valor claramente superior a 0.05, aunque (a diferencia de lo que ocurre en el ejercicio de la Observación 1) dicha frecuencia no tiende a crecer constantemente hacia 1 con N . De hecho, la frecuencia de rechazo de $\beta_2 = 0$ en (4.11) es esencialmente correcta (un 5%) al menos para tamaños muestrales grandes.

Lo que provoca que (4.10) sea una regresión espuria es que $\beta_2 = 0$ es también en este caso una hipótesis falsa en (4.10): si $\beta_2 = 0$, (4.10) implica que $(Y_t) \sim \text{IID}(\beta_1, \sigma^2)$, cuando en realidad $(Y_t) \sim \text{AR}(1)$.

Observación 3: En resumen, los dos ejercicios anteriores ilustran la posibilidad de encontrar relaciones aparentes entre variables que son en realidad independientes. Esta posibilidad no es remota cuando las variables consideradas están autocorrelacionadas, y es particularmente probable con variables no estacionarias. Ver Granger, C.W.J., y Newbold, P. (1974), Spurious regressions in econometrics, *Journal of Econometrics*, 2, 111-120; Phillips, P.C.B., (1986), Understanding spurious regressions in econometrics, *Journal of Econometrics*, 33, 311-340; Durlauf, S.N., y Phillips, P.C.B., (1988), Trends versus random walks in time series analysis, *Econometrica*, 56, 1333-1354; Granger, C.W.J., (2001), Spurious regressions in econometrics, Cap. 26 en *A Companion to Econometric Theory*, ed. B. Baltagi, Blackwell, 557-561.

4.3 CONTRASTES DE RAÍCES UNITARIAS

- ⇒ En muchas ocasiones, el orden de integración de un proceso estocástico se puede determinar de manera bastante fiable a la vista del gráfico temporal y del correlograma de una serie generada por el proceso considerado.
- ⇒ En los casos en los que el orden de integración de un proceso no está claro a la vista de dichos instrumentos gráficos, un contraste formal de raíces unitarias puede ayudar a decidir si, por ejemplo, un proceso es $I(0)$ ó $I(1)$, o si es $I(1)$ ó $I(2)$.
- ⇒ En cualquier caso, la determinación fiable del orden de integración de un proceso es importante en relación con: (i) la forma correcta de hacer inferencias en cada caso, (ii) la posibilidad de estimar relaciones espurias, (iii) la posibilidad de estimar relaciones de cointegración, y (iv) las implicaciones económicas correspondientes.

4.3.1 Ejemplo

Si un proceso estocástico univariante (Y_t) sigue un modelo AR(1) del tipo

$$Y_t = \phi_1 Y_{t-1} + A_t, \quad (4.12)$$

entonces $(Y_t) \sim I(0)$ si $|\phi_1| < 1$, mientras que $(Y_t) \sim I(1)$ (un paseo aleatorio) si $\phi_1 = 1$. Una manera de contrastar formalmente $H_0 : (Y_t) \sim I(1)$ frente a $H_1 : (Y_t) \sim I(0)$ en este caso, consiste en contrastar $H_0 : \phi_1 = 1$ frente a $H_1 : \phi_1 < 1$ en (4.12).

Observación 1: En la práctica, H_1 significa que $0 < \phi_1 < 1$, ya que $\phi_1 < 0$ en (4.12) es una posibilidad muy poco frecuente. Por otro lado, la alternativa $H_1 : \phi_1 > 1$ no se considera habitualmente porque implicaría que (Y_t) es un proceso no estacionario explosivo.

El contraste de $H_0 : \phi_1 = 1$ frente a $H_1 : \phi_1 < 1$ en (4.12) es equivalente [restando Y_{t-1} en ambos lados de (4.12)] al contraste de

$$H_0 : \gamma = 0 \quad \text{frente a} \quad H_1 : \gamma < 0 \quad (4.13)$$

en el modelo

$$\nabla Y_t = \gamma Y_{t-1} + A_t, \quad (4.14)$$

donde $\gamma \equiv \phi_1 - 1$. El estadístico t habitual asociado con γ para contrastar (4.13) en (4.14), **no** sigue bajo H_0 en (4.13) ni una distribución t de Student ni una distribución $N(0,1)$, ni siquiera asintóticamente. No obstante, la distribución del estadístico t bajo H_0 en (4.13) es

conocida y suele denominarse una **distribución de Dickey-Fuller** (DF). El estadístico t asociado con γ en (4.14) se representa habitualmente como τ (**tau**).

El contraste DF de (4.13) en (4.14) consiste en **rechazar** H_0 en favor de H_1 cuando el valor calculado del estadístico τ es **menor** que el valor crítico de la distribución DF correspondiente para un nivel de significación dado. La primera fila de la Tabla 4.1 contiene los valores críticos asintóticos para distintos niveles de significación.

Distribución	Nivel de significación		
	1%	5%	10%
DF - Estadístico τ	-2.56	-1.94	-1.62
DF - Estadístico τ_μ	-3.43	-2.86	-2.57
DF - Estadístico τ_τ	-3.96	-3.41	-3.13
N(0,1) - Estadístico t	-2.33	-1.65	-1.28

Tabla 4.1

Valores críticos asintóticos para los contrastes DF – ADF de raíces unitarias

Observación 2: Los valores críticos de la distribución DF de τ son **más negativos** que los de la distribución N(0,1) (que figuran en la última fila de la Tabla 4.1); esto significa que para rechazar H_0 , el valor del estadístico τ debe ser más

negativo (mayor en valor absoluto) que si se comparase con un valor crítico de la distribución $N(0,1)$. En otros términos, si el valor del estadístico τ se comparase (incorrectamente) con un valor crítico de la distribución $N(0,1)$, H_0 se rechazaría más frecuentemente; ver Figura 4.4.

Contrastes DF (Dickey-Fuller) – ADF

4.3.2 Ejemplo

Una primera extensión del contraste básico de 4.3.1 consiste en permitir la presencia de un término constante en (4.12) o, equivalentemente, en (4.14), de manera que

$$\nabla Y_t = \mu + \gamma Y_{t-1} + A_t. \quad (4.15)$$

En este caso, el contraste DF de (4.13) en (4.15) consiste en **rechazar** H_0 en favor de H_1 cuando el valor calculado del estadístico t asociado con γ en (4.15), que suele representarse como τ_μ , es **menor** que el valor crítico de la distribución DF correspondiente para un nivel de significación dado. La segunda fila de la Tabla 4.1 contiene los valores críticos asintóticos para niveles de significación del 1%, del 5% y del 10%.

Observación: El estadístico t asociado con μ en (4.15) **sí** sigue una distribución $N(0,1)$ asintóticamente (que puede aproximarse por la distribución t de Student habitual en muestras pequeñas), lo que permite contrastar de forma

estándar la significación individual de μ en (4.15). Este resultado es un caso particular del caso más general que se describe en el apartado sobre inferencia en regresiones con series I(1) en la Sección 4.5.

4.3.3 Ejemplo

Una segunda extensión del contraste básico de 4.3.1 consiste en permitir que (Y_t) siga un modelo AR(p) con término constante y $p \geq 2$, de manera que

$$Y_t = \mu + \sum_{i=1}^p \phi_i Y_{t-i} + A_t,$$

o bien (ver 4.1.5)

$$\nabla Y_t = \mu + \gamma Y_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \nabla Y_{t-i} + A_t, \quad (4.16)$$

donde $\gamma \equiv -\phi(1) \equiv -1 + \sum_{i=1}^p \phi_i$, $\phi_i^* \equiv -\sum_{j=i+1}^p \phi_j$ ($i = 1, \dots, p-1$). En este caso, si el operador $\phi(B) \equiv 1 - \sum_{i=1}^p \phi_i B$ tiene una raíz unitaria, entonces $\phi(1) = -\gamma = 0$; por lo tanto, contrastar $H_0 : (Y_t) \sim I(1)$ frente a $H_1 : (Y_t) \sim I(0)$ en este caso es equivalente (de nuevo) a contrastar (4.13) en (4.16). Este contraste (que se denomina **ADF**, del inglés *Augmented Dickey-Fuller*) se lleva a cabo exactamente igual que en 4.3.1 (cuando $\mu = 0$) o en 4.3.2 (si $\mu \neq 0$).

Observación 1: Los estadísticos t asociados con μ , ϕ_1^* , ..., ϕ_{p-1}^* en (4.16) siguen distribuciones $N(0,1)$ asintóticamente

(que pueden aproximarse por la distribución t de Student habitual en muestras pequeñas), lo que permite contrastar de forma estándar la significación individual o conjunta de $\phi_1^*, \dots, \phi_{p-1}^*$ en (4.16).

Observación 2: En la práctica, un contraste ADF suele plantearse a partir de una regresión del tipo

$$\nabla Y_t = \mu + \gamma Y_{t-1} + \sum_{i=1}^m \rho_i \nabla Y_{t-i} + V_t, \quad (4.17)$$

donde m se escoge de tal manera que el término $\sum_{i=1}^m \rho_i \nabla Y_{t-i}$ recoja esencialmente toda la autocorrelación presente en (∇Y_t) [para que (V_t) sea aproximadamente ruido blanco]. Nótese que si (∇Y_t) presenta algún tipo de estructura MA, o bien es estacional, entonces m debe ser un número elevado; en tal caso, la potencia de los contrastes ADF puede ser reducida. Recíprocamente, si m es pequeño, el nivel de significación efectivo del contraste puede ser mayor que el nivel de significación escogido, ya que la validez de los valores críticos de la Tabla 4.1 depende crucialmente de que la dinámica presente en (∇Y_t) esté recogida adecuadamente en (4.17); ver Figura 4.4.

Observación 3: En ocasiones, un contraste ADF se plantea a partir de una regresión del tipo

$$\nabla Y_t = \mu + \beta t + \gamma Y_{t-1} + \sum_{i=1}^m \rho_i \nabla Y_{t-i} + V_t, \quad (4.18)$$

donde la presencia de βt suele justificarse para no confundir un proceso cuya tendencia es puramente determinista con un proceso I(1) (cuyo componente de tendencia esencial es estocástico). Nótese, no obstante, que la hipótesis nula en (4.13) representaría en este caso la posibilidad de que (∇Y_t) tuviera una tendencia lineal determinista (algo muy poco usual en la práctica). En todo caso, el estadístico t para γ en (4.18) suele representarse como τ_τ , y los valores críticos para el contraste ADF correspondiente figuran en la tercera fila de la Tabla 4.1.

Observación 4: En general, cualquier contraste ADF para decidir si un proceso es I(1) ó I(0), también puede utilizarse para decidir si un proceso es I(2) ó I(1), definiendo adecuadamente Y_t en (4.14)-(4.19) [por ejemplo, $Y_t \equiv \nabla Z_t$ si se desea contrastar $(Z_t) \sim I(2)$ frente a $(Z_t) \sim I(1)$].

4.3.4 Ejemplo – Contrastes ADF de raíces unitarias

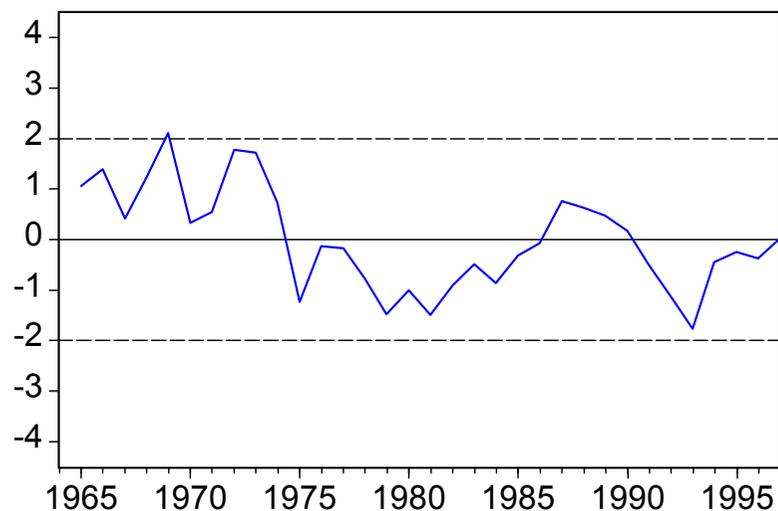


Figura 4.5

Tasa logarítmica de variación anual del PIB en España
(pib-cpi.wf1)

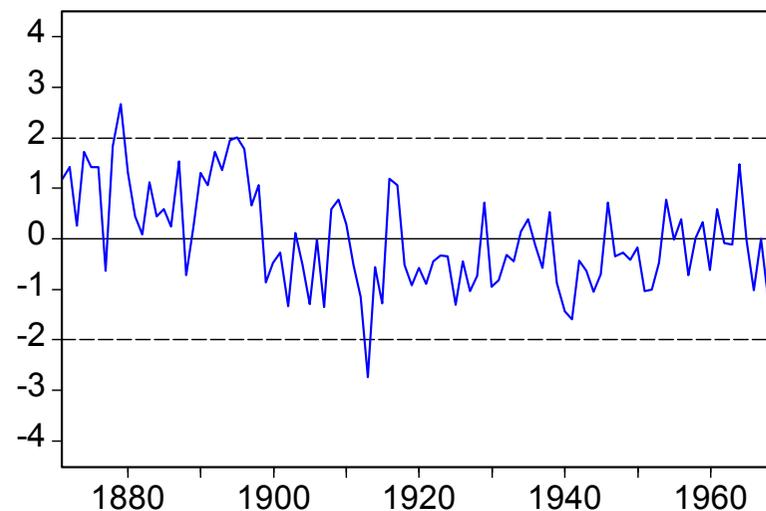


Figura 4.6

Volumen del flujo anual del río Nilo en Asuán
(nilo.wf1)

Observación 1: La Tabla 4.2 contiene el resultado del contraste DF basado en (4.17) con $m = 0$ para la serie (z_t) de la Figura 4.5 (obtenido con EViews 4.1 – May 23 2003 build); en este caso, no puede rechazarse que $(Z_t) \sim I(1)$ ni siquiera al 10% ($\tau_\mu = -2.498$, $p\text{-value} = 12.53\%$). Nótese que: (i) la conclusión habría sido distinta si se hubiera utilizado erróneamente una distribución t para τ_μ ($p\text{-value} = 1.82\%$), (ii) el contraste de Shin-Fuller del Ejemplo

2.6.8 presenta cierta evidencia en favor de que $(Z_t) \sim I(1)$, y (iii) los contrastes de no invertibilidad de Davis-Chen-Dunsmuir del Ejemplo 2.6.8 presentan bastante evidencia en favor de que $(Z_t) \sim I(0)$.

Null Hypothesis: Z has a unit root Exogenous: Constant Lag Length: 0 (Automatic based on AIC, MAXLAG=4)			
	t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic	-2.498350	0.1253	
Test critical values: 1% level	-3.653730		
5% level	-2.957110		
10% level	-2.617434		

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation Dependent Variable: D(Z) Method: Least Squares Sample(adjusted): 1966 1997 Included observations: 32 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.325272	0.130195	-2.498350	0.0182
C	0.010127	0.005366	1.887046	0.0689

Tabla 4.2

Contraste ADF basado en (4.17) con $m = 0$ para la serie de la Figura 4.5

Null Hypothesis: NILO has a unit root Exogenous: Constant Lag Length: 1 (Automatic based on AIC, MAXLAG=4)			
	t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic	-4.048705	0.0018	
Test critical values: 1% level	-3.498439		
5% level	-2.891234		
10% level	-2.582678		

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation Dependent Variable: D(NILO) Method: Least Squares Sample(adjusted): 1873 1970 Included observations: 98 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
NILO(-1)	-0.406281	0.100348	-4.048705	0.0001
D(NILO(-1))	-0.198787	0.100227	-1.983361	0.0502
C	368.3168	93.58492	3.935643	0.0002

Tabla 4.3

Contraste ADF basado en (4.17) con $m = 1$ para la serie de la Figura 4.6

Observación 2: La Tabla 4.3 contiene el resultado del contraste DF basado en (4.17) con $m = 1$ para la serie (z_t) de la Figura 4.6 (obtenido con EViews 4.1 – May 23 2003 build); en este caso, se rechaza claramente que $(Z_t) \sim I(1)$

incluso al 1% ($\tau_\mu = -4.049$, $p\text{-value} = 0.18\%$). Nótese que: (i) la conclusión habría sido la misma si se hubiera utilizado erróneamente una distribución t para τ_μ ($p\text{-value} = 0.01\%$), (ii) el contraste de Shin-Fuller del Ejemplo 2.6.6 presenta bastante evidencia en favor de que $(Z_t) \sim I(1)$, y (iii) los contrastes de no invertibilidad de Davis-Chen-Dunsmuir del Ejemplo 2.6.8 presentan bastante evidencia en contra de que $(Z_t) \sim I(0)$; ver también Ejemplo 3.2.2.

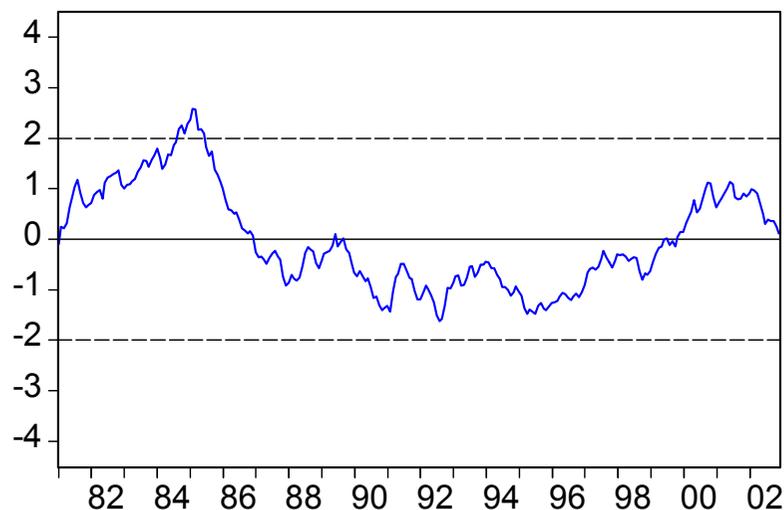


Figura 4.7

Tipo de cambio real ECU-Euro / Dólar EEUU (log)
(exchange-rates.wf1)

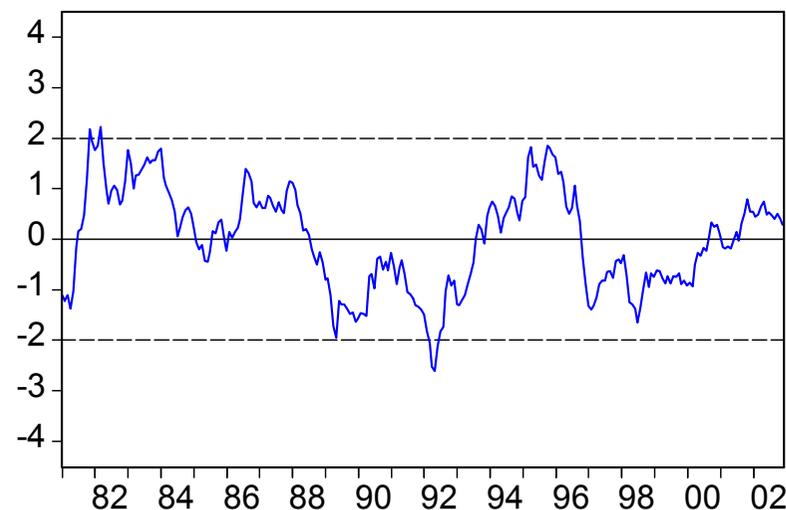


Figura 4.8

Tipo de cambio real ECU-Euro / Franco Suizo (log)
(exchange-rates.wf1)

Observación 3: La Tabla 4.4 contiene el resultado del contraste DF basado en (4.17) con $m = 1$ para la serie (z_t) de la Figura 4.7 (obtenido con EViews 4.1 – May 23 2003 build); en este caso, claramente no se rechaza que $(Z_t) \sim I(1)$

($\tau_\mu = -1.598$, $p\text{-value} = 48.22\%$). Por su parte, la Tabla 4.5 contiene el resultado del contraste DF basado en (4.17) con $m = 1$ para la serie (z_t) de la Figura 4.8; en este caso, (Z_t) $\sim I(1)$ puede rechazarse al 5% aunque no al 1% ($\tau_\mu = -3.026$, $p\text{-value} = 3.38\%$). Estos resultados contienen cierta evidencia en favor de la teoría de la paridad del poder adquisitivo (PPA) entre el Euro y el Franco Suizo, pero bastante evidencia en contra de dicha teoría entre el Euro y el Dólar EEUU; ver Ejemplo 1.4.3.

Null Hypothesis: Z has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic based on SIC, MAXLAG=24)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.597903	0.4822
Test critical values: 1% level			-3.455193	
5% level			-2.872370	
10% level			-2.572615	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Sample(adjusted): 1981:03 2002:12				
Included observations: 262 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.013200	0.008261	-1.597903	0.1113
D(Z(-1))	0.320106	0.058226	5.497629	0.0000
C	-0.000460	0.001533	-0.300207	0.7643

Tabla 4.4

Contraste ADF basado en (4.17) con $m = 1$ para la serie de la Figura 4.7

Null Hypothesis: Z has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic based on SIC, MAXLAG=24)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-3.026169	0.0338
Test critical values: 1% level			-3.455193	
5% level			-2.872370	
10% level			-2.572615	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Sample(adjusted): 1981:03 2002:12				
Included observations: 262 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.050210	0.016592	-3.026169	0.0027
D(Z(-1))	0.253849	0.059947	4.234591	0.0000
C	-0.024599	0.008215	-2.994403	0.0030

Tabla 4.5

Contraste ADF basado en (4.17) con $m = 1$ para la serie de la Figura 4.8

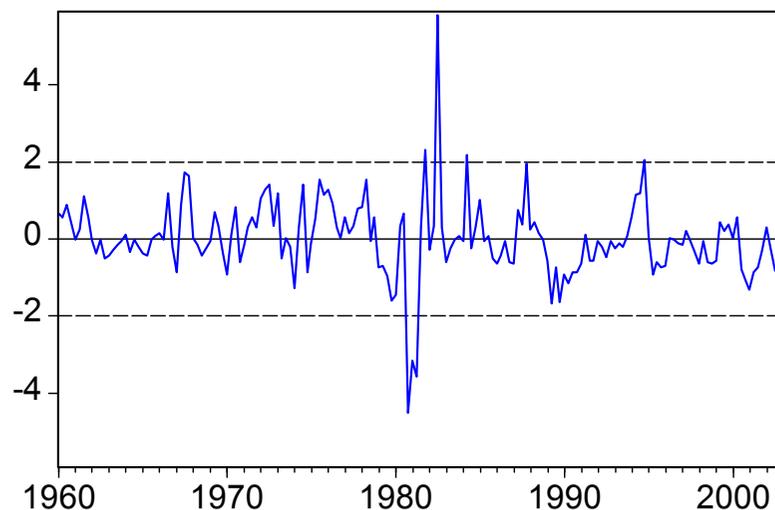


Figura 4.9

Diferencial de la deuda pública (EEUU) a 6 y a 3 meses
(interest-rates-q.wf1)

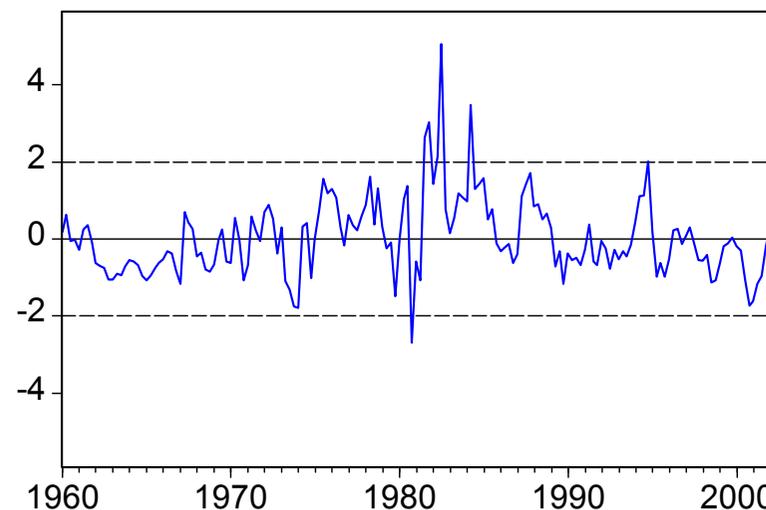


Figura 4.10

Diferencial de la deuda pública (EEUU) a 12 y a 3 meses
(interest-rates-q.wf1)

Observación 4: La Tabla 4.6 contiene el resultado del contraste DF basado en (4.17) con $m = 0$ para la serie (z_t) de la Figura 4.9 (obtenido con EViews 4.1 – May 23 2003 build); en este caso, se rechaza claramente que $(Z_t) \sim I(1)$ en favor de $(Z_t) \sim I(0)$ ($\tau_\mu = -8.498$, $p\text{-value} = 0.00\%$). Por su parte, la Tabla 4.7 contiene el resultado del contraste DF basado en (4.17) con $m = 0$ para la serie (z_t) de la Figura 4.10; también en este caso se rechaza claramente que $(Z_t) \sim I(1)$ en favor de $(Z_t) \sim I(0)$ ($\tau_\mu = -6.504$, $p\text{-value} = 0.00\%$). Estos resultados contienen bastante evidencia en favor de la hipótesis de las expectativas sobre la estructura temporal de los tipos de interés; ver Ejemplo 1.4.4.

Null Hypothesis: Z has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=12)				
		t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic				
Test critical values:	1% level	-3.468749		
	5% level	-2.878311		
	10% level	-2.575791		
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Sample(adjusted): 1960:2 2002:4				
Included observations: 171 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.598199	0.070389	-8.498442	0.0000
C	0.091311	0.019087	4.783835	0.0000

Tabla 4.6

Contraste ADF basado en (4.17) con $m = 0$ para la serie de la Figura 4.9

Null Hypothesis: Z has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=12)				
		t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic				
Test critical values:	1% level	-3.468749		
	5% level	-2.878311		
	10% level	-2.575791		
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Sample(adjusted): 1960:2 2002:4				
Included observations: 171 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.402058	0.061822	-6.503527	0.0000
C	0.247491	0.047686	5.190045	0.0000

Tabla 4.7

Contraste ADF basado en (4.17) con $m = 0$ para la serie de la Figura 4.10

Alternativas a los contrastes ADF

⇒ A pesar de que los contrastes ADF basados en (4.17)-(4.18) son probablemente los más populares en la práctica actual, sus inconvenientes (ver Observación 2 en 4.3.3) han

provocado la aparición de decenas de contrastes alternativos (ninguno de los cuales, por otro lado, parece haber alcanzado la popularidad de los contrastes ADF).

- ⇒ Entre ellos, en algunos se plantea como hipótesis nula el que un proceso sea $I(0)$ (al contrario que en los contrastes ADF); ver, por ejemplo, Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., y Shin, Y. (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics*, 54, 159-178.
- ⇒ En general, los contrastes de raíces unitarias más fiables disponibles actualmente son el contraste de no estacionariedad de Shin-Fuller y los contrastes de no invertibilidad de Davis-Chen-Dunsmuir, descritos en los dos últimos apartados de la Sección 2.6.

4.4 MODELOS VAR CON RAÍCES UNITARIAS

4.4.1 Ejemplo

Sea (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}]'$, un proceso bivalente que sigue un modelo VAR(1) del tipo

$$\underbrace{\begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{12} & \phi_{22} \end{bmatrix}}_{\boldsymbol{\Phi}_1} \underbrace{\begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix}}_{\mathbf{Y}_{t-1}} + \underbrace{\begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}}_{\mathbf{A}_t}, \quad (\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_A), \quad (4.19)$$

tal que los dos autovalores de $\boldsymbol{\Phi}_1$, λ_1 y λ_2 , son **distintos** entre sí. En este caso,

$$\boldsymbol{\Phi}_1 = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} p(1)1 & p(1)2 \\ p(2)1 & p(2)2 \end{bmatrix} = [\mathbf{p}_1, \mathbf{p}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{p}'(1) \\ \mathbf{p}'(2) \end{bmatrix}, \quad (4.20)$$

donde \mathbf{p}_1 y \mathbf{p}_2 son dos autovectores de $\boldsymbol{\Phi}_1$ linealmente independientes, asociados con λ_1 y λ_2 , respectivamente. El modelo VAR(1) en (4.19) puede escribirse como

$$\underbrace{\mathbf{P}^{-1}\mathbf{Y}_t}_{\mathbf{Z}_t} = \underbrace{\mathbf{P}^{-1}\boldsymbol{\mu}}_{\boldsymbol{\mu}_0} + \underbrace{\mathbf{P}^{-1}\boldsymbol{\Phi}_1\mathbf{P}}_{\boldsymbol{\Lambda}} \underbrace{\mathbf{P}^{-1}\mathbf{Y}_{t-1}}_{\mathbf{Z}_{t-1}} + \underbrace{\mathbf{P}^{-1}\mathbf{A}_t}_{\mathbf{U}_t}, \quad (4.21)$$

donde

$$\mathbf{Z}_t = \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} = \begin{bmatrix} \mathbf{p}'_{(1)} \mathbf{Y}_t \\ \mathbf{p}'_{(2)} \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} p_{(1)1}Y_{t1} + p_{(1)2}Y_{t2} \\ p_{(2)1}Y_{t1} + p_{(2)2}Y_{t2} \end{bmatrix}, \quad (4.22)$$

es decir,

$$\underbrace{\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix}}_{\mathbf{Z}_t} = \underbrace{\begin{bmatrix} \mu_{01} \\ \mu_{02} \end{bmatrix}}_{\boldsymbol{\mu}_0} + \underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{bmatrix}}_{\mathbf{Z}_{t-1}} + \underbrace{\begin{bmatrix} U_{t1} \\ U_{t2} \end{bmatrix}}_{\mathbf{U}_t}, \quad (\mathbf{U}_t) \sim \text{IID}(\mathbf{0}, \mathbf{P}^{-1} \boldsymbol{\Sigma}_A \mathbf{P}^{-1'}), \quad (4.23)$$

de manera que $(1 - \lambda_1 B)Z_{t1} = \mu_{01} + U_{t1}$, $(1 - \lambda_2 B)Z_{t2} = \mu_{02} + U_{t2}$. En consecuencia, $(Z_{ti}) \sim \text{I}(0)$ o bien $(Z_{ti}) \sim \text{I}(1)$ en función de si $|\lambda_i| < 1$ o bien $\lambda_i = 1$, respectivamente. Por otro lado, (4.21) también implica que $\mathbf{Y}_t = \mathbf{P}\mathbf{Z}_t$, es decir,

$$\mathbf{Y}_t = \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \mathbf{p}_1 Z_{t1} + \mathbf{p}_2 Z_{t2} = \begin{bmatrix} p_{11}Z_{t1} + p_{12}Z_{t2} \\ p_{21}Z_{t1} + p_{22}Z_{t2} \end{bmatrix}. \quad (4.24)$$

Caso 1: $|\lambda_1| < 1$, $|\lambda_2| < 1$ ($\lambda_1 \neq \lambda_2$).

(4.23) implica que $(Z_{t1}) \sim \text{I}(0)$ y $(Z_{t2}) \sim \text{I}(0)$, por lo que (4.24) implica que $(\mathbf{Y}_t) \sim \text{I}(0)$.

Caso 2: $\lambda_1 = 1, |\lambda_2| < 1$.

(4.23) implica que $(Z_{t1}) \sim I(1)$ y $(Z_{t2}) \sim I(0)$, por lo que (4.24) implica que $(\mathbf{Y}_t) \sim I(1)$. Juntando este resultado con la segunda fila de (4.22), $Z_{t2} = \mathbf{p}'_{(2)} \mathbf{Y}_t = p_{(2)1} Y_{t1} + p_{(2)2} Y_{t2}$, resulta que $(\mathbf{p}'_{(2)} \mathbf{Y}_t) \sim I(0)$, por lo que $(\mathbf{Y}_t) \sim CI(1,1)$ con vector de cointegración $\boldsymbol{\beta} \equiv [\beta_1, \beta_2]' = \mathbf{p}_{(2)} \equiv [p_{(2)1}, p_{(2)2}]'$ [la segunda fila de \mathbf{P}^{-1} en (4.20)]. En este caso, el modelo VEC (ver 4.1.5) asociado con (4.19) es $\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi} \mathbf{Y}_{t-1} + \mathbf{A}_t$, donde $\boldsymbol{\Pi} \equiv \mathbf{I} - \boldsymbol{\Phi}_1$ puede escribirse [ver (4.20)] como

$$\boldsymbol{\Pi} = \mathbf{P}(\mathbf{I} - \boldsymbol{\Lambda})\mathbf{P}^{-1} = [\mathbf{p}_1, \mathbf{p}_2] \begin{bmatrix} 0 & 0 \\ 0 & 1 - \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{p}'_{(1)} \\ \mathbf{p}'_{(2)} \end{bmatrix} = [\mathbf{p}_2(1 - \lambda_2)][\mathbf{p}'_{(2)}] = \boldsymbol{\alpha}\boldsymbol{\beta}'. \quad (4.25)$$

Es decir, $\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \mathbf{A}_t$, o bien

$$\begin{bmatrix} \nabla Y_{t1} \\ \nabla Y_{t2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} - \begin{bmatrix} \alpha_1(\beta_1 Y_{t-1,1} + \beta_2 Y_{t-1,2}) \\ \alpha_2(\beta_1 Y_{t-1,1} + \beta_2 Y_{t-1,2}) \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}, \quad (4.26)$$

donde $\alpha_1 \equiv p_{12}(1 - \lambda_2)$, $\alpha_2 \equiv p_{22}(1 - \lambda_2)$, $|\lambda_2| < 1$, $\beta_1 \equiv p_{(2)1}$, $\beta_2 \equiv p_{(2)2}$. Nótese que: (i) el rango de $\boldsymbol{\Pi}$ en (4.25) es 1, y (ii) (4.26) **no** es un modelo VAR para $(\nabla \mathbf{Y}_t) \sim I(0)$.

Caso 3: $\lambda_1 = \lambda_2 = \lambda = 1$.

Si Φ_1 en (4.19) **no** es una matriz diagonal, pero sus dos autovalores son iguales, entonces

$$\Phi_1 = \mathbf{Q}\mathbf{J}\mathbf{Q}^{-1} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} q_{(1)1} & q_{(1)2} \\ q_{(2)1} & q_{(2)2} \end{bmatrix} = [\mathbf{q}_1, \mathbf{q}_2] \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \mathbf{q}'_{(1)} \\ \mathbf{q}'_{(2)} \end{bmatrix}, \quad (4.27)$$

que se denomina la **forma canónica de Jordan** de Φ_1 . Ahora, (4.19) puede escribirse como

$$\underbrace{\mathbf{Q}^{-1}\mathbf{Y}_t}_{\mathbf{Z}_t} = \underbrace{\mathbf{Q}^{-1}\boldsymbol{\mu}}_{\boldsymbol{\mu}_0} + \underbrace{\mathbf{Q}^{-1}\Phi_1\mathbf{Q}}_{\mathbf{J}} \underbrace{\mathbf{Q}^{-1}\mathbf{Y}_{t-1}}_{\mathbf{Z}_{t-1}} + \underbrace{\mathbf{Q}^{-1}\mathbf{A}_t}_{\mathbf{U}_t}, \quad (4.28)$$

donde

$$\mathbf{Z}_t = \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} = \begin{bmatrix} \mathbf{q}'_{(1)}\mathbf{Y}_t \\ \mathbf{q}'_{(2)}\mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} q_{(1)1}Y_{t1} + q_{(1)2}Y_{t2} \\ q_{(2)1}Y_{t1} + q_{(2)2}Y_{t2} \end{bmatrix}, \quad (4.29)$$

es decir,

$$\underbrace{\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix}}_{\mathbf{Z}_t} = \underbrace{\begin{bmatrix} \mu_{01} \\ \mu_{02} \end{bmatrix}}_{\boldsymbol{\mu}_0} + \underbrace{\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}}_{\mathbf{J}} \underbrace{\begin{bmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{bmatrix}}_{\mathbf{Z}_{t-1}} + \underbrace{\begin{bmatrix} U_{t1} \\ U_{t2} \end{bmatrix}}_{\mathbf{U}_t}, \quad (\mathbf{U}_t) \sim \text{IID}(\mathbf{0}, \mathbf{Q}^{-1}\boldsymbol{\Sigma}_A\mathbf{Q}^{-1'}), \quad (4.30)$$

de manera que $(1 - \lambda B)Z_{t1} = \mu_{01} + Z_{t-1,2} + U_{t1}$, $(1 - \lambda B)Z_{t2} = \mu_{02} + U_{t2}$. Por lo tanto, cuando $\lambda = 1$, $(Z_{t2}) \sim I(1)$ y $(Z_{t1}) \sim I(2)$. Por otro lado, (4.28) también implica que $\mathbf{Y}_t = \mathbf{Q}\mathbf{Z}_t$, es decir,

$$\mathbf{Y}_t = \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} = \mathbf{q}_1 Z_{t1} + \mathbf{q}_2 Z_{t2} = \begin{bmatrix} q_{11}Z_{t1} + q_{12}Z_{t2} \\ q_{21}Z_{t1} + q_{22}Z_{t2} \end{bmatrix}, \quad (4.31)$$

de manera que $(\mathbf{Y}_t) \sim I(2)$ cuando $\lambda = 1$. Juntando este resultado con la segunda fila de (4.29), $Z_{t2} = \mathbf{q}'_{(2)}\mathbf{Y}_t = q_{(2)1}Y_{t1} + q_{(2)2}Y_{t2}$, resulta que $(\mathbf{q}'_{(2)}\mathbf{Y}_t) \sim I(1)$, por lo que $(\mathbf{Y}_t) \sim CI(2, 1)$ con vector de cointegración $\boldsymbol{\beta} \equiv [\beta_1, \beta_2]' = \mathbf{q}_{(2)} \equiv [q_{(2)1}, q_{(2)2}]'$ [la segunda fila de \mathbf{Q}^{-1} en (4.27)]. En este caso, el modelo VEC (ver 4.1.5) asociado con (4.19) es $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi}\mathbf{Y}_{t-1} + \mathbf{A}_t$, donde $\boldsymbol{\Pi} \equiv \mathbf{I} - \boldsymbol{\Phi}_1$ puede escribirse [ver (4.27)] como

$$\boldsymbol{\Pi} = \mathbf{Q}(\mathbf{I} - \mathbf{J})\mathbf{Q}^{-1} = [\mathbf{q}_1, \mathbf{q}_2] \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q}'_{(1)} \\ \mathbf{q}'_{(2)} \end{bmatrix} = -[\mathbf{q}_1][\mathbf{q}'_{(2)}] = \boldsymbol{\alpha}\boldsymbol{\beta}'. \quad (4.32)$$

Es decir, $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \mathbf{A}_t$, o bien

$$\begin{bmatrix} \nabla Y_{t1} \\ \nabla Y_{t2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} - \begin{bmatrix} \alpha_1 (\beta_1 Y_{t-1,1} + \beta_2 Y_{t-1,2}) \\ \alpha_2 (\beta_1 Y_{t-1,1} + \beta_2 Y_{t-1,2}) \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}, \quad (4.33)$$

donde $\alpha_1 \equiv -q_{12}$, $\alpha_2 \equiv -q_{22}$, $\beta_1 \equiv p_{(2)1}$, $\beta_2 \equiv p_{(2)2}$. Nótese que: (i) el rango de $\mathbf{\Pi}$ en (4.32) es 1, y (ii) (4.33) **no** es un modelo VAR para $(\nabla \mathbf{Y}_t) \sim I(1)$. La situación en este caso es completamente análoga a la del Caso 2 anterior, aunque ahora $(\mathbf{Y}_t) \sim CI(2, 1)$.

Caso 4: $\lambda_1 = \lambda_2 = \lambda = 1$ porque $\mathbf{\Phi}_1 = \mathbf{I}$.

En este caso especial, (4.19) queda $\nabla \mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t$, de manera que $(\mathbf{Y}_t) \sim I(1)$. Además, en la expresión $\nabla \mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t$ no está presente ningún término de corrección de error, lo que indica que no existe ninguna combinación lineal de los dos componentes de (\mathbf{Y}_t) que sea $I(0)$; en otros términos, $\mathbf{\Pi} = \mathbf{I} - \mathbf{\Phi}_1 = \mathbf{0}$, cuyo rango es igual a cero. En este caso, los dos componentes de $(\mathbf{Y}_t) \sim I(1)$ **no** están cointegrados.

Observación: En resumen, cuando (\mathbf{Y}_t) es un proceso bivalente que sigue un modelo VAR(1) como (4.19), ocurre que: (i) $(\mathbf{Y}_t) \sim I(0)$ cuando $\text{rango}(\mathbf{\Pi}) = 2$, (ii) $(\mathbf{Y}_t) \sim CI(1, 1)$ o bien $(\mathbf{Y}_t) \sim CI(2, 1)$ cuando $\text{rango}(\mathbf{\Pi}) = 1$, y (iii) $(\mathbf{Y}_t) \sim I(1)$ con sus dos componentes no cointegrados cuando $\text{rango}(\mathbf{\Pi}) = 0$, donde $\mathbf{\Pi} \equiv \mathbf{I} - \mathbf{\Phi}_1$. Además, cuando $(\mathbf{Y}_t) \sim CI(1, 1)$, $(\nabla \mathbf{Y}_t)$ **no** sigue un modelo VAR estándar, sino un modelo VEC como (4.26).

4.4.2 Ejemplo

Sea (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, Y_{t3}]'$, un proceso trivariante que sigue un modelo VAR(1),

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \mathbf{A}_t, \quad (\mathbf{A}_t) \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma}_A), \quad (4.34)$$

tal que los tres autovalores de $\boldsymbol{\Phi}_1$, λ_1 , λ_2 y λ_3 , son **distintos** entre sí. En este caso,

$$\boldsymbol{\Phi}_1 = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \mathbf{p}'_{(1)} \\ \mathbf{p}'_{(2)} \\ \mathbf{p}'_{(3)} \end{bmatrix}, \quad (4.35)$$

donde \mathbf{p}_1 , \mathbf{p}_2 y \mathbf{p}_3 son tres autovectores de $\boldsymbol{\Phi}_1$ linealmente independientes, asociados con λ_1 , λ_2 y λ_3 , respectivamente. El modelo VAR(1) en (4.34) puede escribirse como

$$\underbrace{\mathbf{P}^{-1}\mathbf{Y}_t}_{\mathbf{Z}_t} = \underbrace{\mathbf{P}^{-1}\boldsymbol{\mu}}_{\boldsymbol{\mu}_0} + \underbrace{\mathbf{P}^{-1}\boldsymbol{\Phi}_1\mathbf{P}}_{\boldsymbol{\Lambda}} \underbrace{\mathbf{P}^{-1}\mathbf{Y}_{t-1}}_{\mathbf{Z}_{t-1}} + \underbrace{\mathbf{P}^{-1}\mathbf{A}_t}_{\mathbf{U}_t}, \quad (4.36)$$

lo que implica que $\mathbf{Z}_t \equiv [Z_{t1}, Z_{t2}, Z_{t3}]' \equiv [\mathbf{p}'_{(1)}\mathbf{Y}_t, \mathbf{p}'_{(2)}\mathbf{Y}_t, \mathbf{p}'_{(3)}\mathbf{Y}_t]'$ [combinaciones lineales de los componentes de (\mathbf{Y}_t)]; recíprocamente, $\mathbf{Y}_t \equiv \mathbf{P}\mathbf{Z}_t$.

Caso 1: $|\lambda_i| < 1$ ($i = 1, 2, 3$)

(4.36) implica que $(Z_{ti}) \sim I(0)$ ($i = 1, 2, 3$), por lo que $(\mathbf{Y}_t) \equiv (\mathbf{PZ}_t) \sim I(0)$.

Caso 2: $\lambda_1 = 1$, $|\lambda_2| < 1$, $|\lambda_3| < 1$ ($\lambda_2 \neq \lambda_3$).

(4.36) implica que $(Z_{t1}) \sim I(1)$, $(Z_{t2}) \sim I(0)$ y $(Z_{t3}) \sim I(0)$, por lo que $(\mathbf{Y}_t) \sim I(1)$. Juntando este resultado con que $(Z_{t2}) \equiv (\mathbf{p}'_{(2)}\mathbf{Y}_t)$ y $(Z_{t3}) \equiv (\mathbf{p}'_{(3)}\mathbf{Y}_t)$, resulta en este caso que existen **dos** relaciones de cointegración entre los tres componentes de (\mathbf{Y}_t) , cuyos vectores de cointegración son $\beta_1 = \mathbf{p}_{(2)}$ y $\beta_2 = \mathbf{p}_{(3)}$, respectivamente.

En este caso, el modelo VEC (ver 4.1.5) asociado con (4.34) es $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi}\mathbf{Y}_{t-1} + \mathbf{A}_t$, donde $\boldsymbol{\Pi} \equiv \mathbf{I} - \boldsymbol{\Phi}_1$ puede escribirse [ver (4.35) y (4.25)] como

$$\boldsymbol{\Pi} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 - \lambda_2 & 0 \\ 0 & 0 & 1 - \lambda_3 \end{bmatrix} \begin{bmatrix} \mathbf{p}'_{(1)} \\ \mathbf{p}'_{(2)} \\ \mathbf{p}'_{(3)} \end{bmatrix} = [\mathbf{p}_2(1 - \lambda_2), \mathbf{p}_3(1 - \lambda_3)] \begin{bmatrix} \mathbf{p}'_{(2)} \\ \mathbf{p}'_{(3)} \end{bmatrix} = \mathbf{AB}',$$

donde \mathbf{A} y \mathbf{B} son matrices 3×2 con rango igual a 2. Es decir, $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \mathbf{AB}'\mathbf{Y}_{t-1} + \mathbf{A}_t$, o

bien $\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \mathbf{A}\mathbf{Z}_{t-1}^* + \mathbf{A}_t$, donde $\mathbf{Z}_{t-1}^* \equiv [Z_{t-1,2}, Z_{t-1,3}]' \equiv \mathbf{B}'\mathbf{Y}_{t-1}$. Nótese que en este caso el rango de $\boldsymbol{\Pi} = \mathbf{A}\mathbf{B}'$ es 2 (porque existen 2 relaciones de cointegración).

Caso 3: $\lambda_1 = \lambda_2 = \lambda = 1$, $|\lambda_3| < 1$.

Cuando dos de los tres autovalores de $\boldsymbol{\Phi}_1$ en (4.34) son iguales, en ciertas ocasiones

$$\boldsymbol{\Phi}_1 = \mathbf{Q}\mathbf{J}\mathbf{Q}^{-1} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \mathbf{q}'_{(1)} \\ \mathbf{q}'_{(2)} \\ \mathbf{q}'_{(3)} \end{bmatrix}, \quad (4.37)$$

que se denomina la **forma canónica de Jordan** de $\boldsymbol{\Phi}_1$. Ahora, (4.34) puede escribirse como

$$\underbrace{\mathbf{Q}^{-1}\mathbf{Y}_t}_{\mathbf{Z}_t} = \underbrace{\mathbf{Q}^{-1}\boldsymbol{\mu}}_{\boldsymbol{\mu}_0} + \underbrace{\mathbf{Q}^{-1}\boldsymbol{\Phi}_1\mathbf{Q}}_{\mathbf{J}} \underbrace{\mathbf{Q}^{-1}\mathbf{Y}_{t-1}}_{\mathbf{Z}_{t-1}} + \underbrace{\mathbf{Q}^{-1}\mathbf{A}_t}_{\mathbf{U}_t}, \quad (4.38)$$

lo que implica cuando $\lambda = 1$ que $(Z_{t3}) \sim I(0)$, $(Z_{t2}) \sim I(1)$ y $(Z_{t1}) \sim I(2)$; por otro lado, (4.38) también implica que $\mathbf{Y}_t = \mathbf{Q}\mathbf{Z}_t$, por lo que $(\mathbf{Y}_t) \sim I(2)$ cuando $\lambda = 1$. Juntando estos dos resultados, se obtiene que $(\mathbf{q}'_{(2)}\mathbf{Y}_t) \sim I(1)$, por lo que $(\mathbf{Y}_t) \sim CI(2,1)$, y, al

mismo tiempo, $(\mathbf{q}'_{(3)}\mathbf{Y}_t) \sim I(0)$, por lo que $(\mathbf{Y}_t) \sim CI(2, 2)$. El modelo VEC asociado con (4.34) es $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi}\mathbf{Y}_{t-1} + \mathbf{A}_t$, donde $\boldsymbol{\Pi} = \mathbf{I} - \boldsymbol{\Phi}_1$ puede escribirse [ver (4.37)] como

$$\boldsymbol{\Pi} = \mathbf{Q}(\mathbf{I} - \mathbf{J})\mathbf{Q}^{-1} = [-\mathbf{q}_1, \mathbf{q}_3(1 - \lambda_3)] \begin{bmatrix} \mathbf{q}'_{(2)} \\ \mathbf{q}'_{(3)} \end{bmatrix} = \mathbf{A}\mathbf{B}', \quad (4.39)$$

donde \mathbf{A} y \mathbf{B} son matrices 3×2 con rango igual a 2. Es decir, $\nabla\mathbf{Y}_t = \boldsymbol{\mu} - \mathbf{A}\mathbf{B}'\mathbf{Y}_{t-1} + \mathbf{A}_t$, donde ahora $(\nabla\mathbf{Y}_t) \sim I(1)$ y $(\mathbf{B}'\mathbf{Y}_t)$ tiene un componente $I(1)$ y otro $I(0)$. También en este caso el rango de $\boldsymbol{\Pi} = \mathbf{A}\mathbf{B}'$ es 2 (porque existen 2 relaciones de cointegración).

Observación 1: En resumen, cuando (\mathbf{Y}_t) es un proceso trivariante que sigue un modelo VAR(1) como (4.34), puede ocurrir que: (i) $(\mathbf{Y}_t) \sim I(0)$ cuando $\text{rango}(\boldsymbol{\Pi}) = 3$, (ii) (\mathbf{Y}_t) es un proceso no estacionario cuyos componentes presentan dos relaciones de cointegración cuando $\text{rango}(\boldsymbol{\Pi}) = 2$. En general, cuando $(\mathbf{Y}_t) \sim I(d)$ es un proceso de orden $M \geq 2$, pueden existir hasta $M - 1$ relaciones de cointegración entre sus componentes; además, si $d = 2$, no todas esas relaciones proporcionan necesariamente un proceso $I(0)$; estas posibilidades son, no obstante, difíciles de interpretar en la práctica en muchas ocasiones.

Observación 2: Una posibilidad fácilmente interpretable en la práctica, consiste en que los componentes de $(\mathbf{Y}_t) \sim I(1)$ presenten una única relación de cointegración, de manera que $(\mathbf{Y}_t) \sim CI(1, 1)$ con un único vector de cointegración. Esta posibilidad existe incluso cuando $M \geq 3$. Por ejemplo, considérese (4.34) con

$$\Phi_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & \phi_{33} \end{bmatrix},$$

donde $|\phi_{33}| < 1$. En este caso, los autovalores de Φ_1 son como en el Caso 3 anterior ($\lambda_1 = \lambda_2 = 1, \lambda_3 = \phi_{33}$); sin embargo, en este caso sí existen dos autovectores linealmente independientes asociados con el autovalor igual a 1. Por ejemplo, si $\phi_{33} = 0.6$, entonces

$$\Phi_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0.6 \end{bmatrix} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \begin{bmatrix} 0.4 & 0 & 0 \\ 0 & 0.4 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} 2.5 & 0 & 0 \\ 0 & 2.5 & 0 \\ -2.5 & -2.5 & 1 \end{bmatrix},$$

$$\mathbf{\Pi} \equiv \mathbf{I} - \Phi_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.4 \end{bmatrix} [-2.5, -2.5, 1].$$

La expresión anterior para Φ_1 indica [ver (4.36)] que $\mathbf{Z}_t \equiv [Z_{t1}, Z_{t2}, Z_{t3}]' \equiv \mathbf{P}^{-1}\mathbf{Y}_t$ es tal que (Z_{t1}) y (Z_{t2}) son $I(1)$, pero (Z_{t3}) es $I(0)$, lo que implica que $(\mathbf{Y}_t) \sim I(1)$. Estos resultados junto con la expresión anterior para $\mathbf{\Pi}$ implican que $(\mathbf{Y}_t) \sim CI(1,1)$ con vector de cointegración $\beta = [-2.5, -2.5, 1]'$; dado que $\text{rango}(\mathbf{\Pi}) = 1$, en este caso existe una única relación de cointegración entre los tres componentes $I(1)$ de (\mathbf{Y}_t) .

4.4.3 Definición

El **rango de cointegración** de un proceso (\mathbf{Y}_t) , $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, integrado de orden $d \geq 1$, que sigue un modelo VAR(p) del tipo

$$\Phi(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}_t, \text{ o bien } \mathbf{Y}_t = \boldsymbol{\mu} + \sum_{i=1}^p \Phi_i \mathbf{Y}_{t-i} + \mathbf{A}_t, \quad (4.40)$$

es el rango de la matriz $\boldsymbol{\Pi} \equiv \Phi(1) \equiv \mathbf{I} - \sum_{i=1}^p \Phi_i$ que figura en la representación VEC (ver 4.1.5) asociada con (4.40):

$$\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi} \mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{F}_i \nabla \mathbf{Y}_{t-i} + \mathbf{A}_t. \quad (4.41)$$

Si $P \equiv \text{rango}(\boldsymbol{\Pi})$, entonces P es el número de relaciones de cointegración linealmente independientes que existen entre los componentes de $(\mathbf{Y}_t) \sim \text{I}(d)$. Nótese que:

- (1) Si $P = M$, entonces $|\boldsymbol{\Pi}| \equiv |\Phi(1)|$ es distinto de cero, lo que indica que $|\Phi(x)| = 0$ no tiene raíces unitarias; por lo tanto, en este caso $(\mathbf{Y}_t) \sim \text{I}(0)$. En consecuencia, si $(\mathbf{Y}_t) \sim \text{I}(d)$ con $d \geq 1$, entonces $0 \leq P \leq M - 1$.
- (2) Si $P = 0$, entonces $\boldsymbol{\Pi} = \mathbf{0}$, por lo que $(\nabla \mathbf{Y}_t) \sim \text{VAR}(p - 1)$ (estacionario si $d = 1$).

- (3) La situación intermedia entre las dos anteriores ($1 \leq P \leq M - 1$) indica que existen P relaciones de cointegración entre los componentes de (\mathbf{Y}_t) , que es un proceso integrado de orden uno o superior. En este caso, $\mathbf{\Pi}$ puede factorizarse como $\mathbf{\Pi} = \mathbf{A}\mathbf{B}'$, donde \mathbf{A} y \mathbf{B} son matrices de orden $M \times P$, con $\text{rango}(\mathbf{A}) = \text{rango}(\mathbf{B}) = P$; las P columnas de \mathbf{B} (o las P filas de \mathbf{B}') contienen los P vectores de cointegración (linealmente independientes) entre los componentes de (\mathbf{Y}_t) .

4.5 ESTIMACIÓN DE RELACIONES DE COINTEGRACIÓN

- ⇒ Cuando $(\mathbf{Y}_t) \sim \text{CI}(1, 1)$, es importante poder hacer inferencias sobre al menos uno de los vectores de cointegración $\boldsymbol{\beta} \equiv [\beta_1, \beta_2, \dots, \beta_M]'$ asociado con una de las $M - 1$ posibles relaciones de cointegración entre los componentes de (\mathbf{Y}_t) .
- ⇒ Aunque es lógico contrastar la presencia de relaciones de cointegración antes de estimarlas, también es cierto que resulta más sencillo discutir cómo estimar relaciones de cointegración antes de discutir cómo contrastar su presencia; ver Sección 4.6.

Estimación directa

La manera más sencilla y directa de estimar una posible relación de cointegración del tipo $(U_t) \equiv (\beta_0 + \boldsymbol{\beta}'\mathbf{Y}_t) \sim \text{I}(0)$ (ver 4.1.2) entre los componentes de $(\mathbf{Y}_t) \sim \text{CI}(1, 1)$, consiste en estimar por MCO la regresión lineal de un componente dado de (\mathbf{Y}_t) sobre todos los demás. Si $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, se trata de estimar por MCO una regresión del tipo

$$Y_{t1} = -\beta_0 - \boldsymbol{\beta}'_2 \mathbf{Y}_{t2} + U_t, \quad (4.42)$$

donde $\mathbf{Y}_{t2} \equiv [Y_{t2}, \dots, Y_{tM}]'$, $\boldsymbol{\beta}_2 \equiv [\beta_2, \dots, \beta_M]'$, de manera que el vector de cointegración en la (posible) relación considerada es $\boldsymbol{\beta} \equiv [1, \boldsymbol{\beta}'_2]'$ (ver Observación 2 en 4.1.2).

Observación 1: El estimador MCO de $\boldsymbol{\beta}_2$ en (4.42) suele representarse como $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$. A simple vista, (4.42) parece un modelo en el que se violan todas las hipótesis que garantizan que $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$ sea un estimador útil en la práctica (como mínimo, consistente). Por ejemplo, cuando $(\mathbf{Y}_t) \sim \text{VAR}(p)$, todas las variables en (4.42) son **endógenas**, de manera que $\text{Cov}[\mathbf{Y}_{t2}, U_t] \neq \mathbf{0}$ [a menos que las perturbaciones del modelo VAR que sigue (\mathbf{Y}_t) sean ortogonales]; además, (U_t) en (4.42) suele presentar algún tipo de autocorrelación. No obstante, se puede demostrar que $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$ converge en probabilidad a $\boldsymbol{\beta}_2$ a la misma velocidad que N^{-1} (es decir, más rápido que $N^{-1/2}$); por este motivo, suele decirse que $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$ es un estimador **superconsistente** de $\boldsymbol{\beta}_2$ en (4.42). Este resultado ilustra el hecho de que la teoría asintótica asociada con procesos I(1) es muy diferente de la teoría asintótica estándar para procesos estacionarios.

Observación 2: La distribución asintótica de $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$ **no** es una distribución estándar, por lo que (a pesar de lo mencionado en la Observación 1) cualquier contraste habitual sobre $\boldsymbol{\beta}_2$ basado en $\hat{\boldsymbol{\beta}}_2^{\text{MCO}}$ es (en general) incorrecto.

Estimación a través de un modelo de corrección de error

Cuando $(\mathbf{Y}_t) \sim \text{CI}(1, 1)$, (\mathbf{Y}_t) sigue un modelo VEC como (4.41). Si se considera una única relación de cointegración, entonces $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ en (4.41), donde $\boldsymbol{\alpha} \equiv [\alpha_1, \dots, \alpha_M]'$ contiene al menos un elemento distinto de cero, y $\boldsymbol{\beta} \equiv [1, \boldsymbol{\beta}'_2]'$ (como en el apartado anterior) es el vector de cointegración (normalizado) que se pretende estimar.

La primera fila de (4.41) puede escribirse en este caso como

$$\nabla Y_{t1} = \mu_1 - \alpha_1 (Y_{t-1,1} + \beta_2' \mathbf{Y}_{t-1,2}) + \sum_{i=1}^{p-1} \phi_{i1}' \nabla \mathbf{Y}_{t-i} + A_{t1}, \quad (4.43)$$

donde ϕ_{i1}' representa la primera fila de \mathbf{F}_i en (4.41). Si, por ejemplo, $M = 2$, $p = 1$, la expresión anterior queda

$$\nabla Y_{t1} = \mu_1 - \alpha_1 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) + A_{t1}, \quad (4.44)$$

que coincide con la primera fila de (4.26) en el Caso 2 del Ejemplo 4.4.1 (imponiendo la normalización $\beta_1 = 1$). La expresión anterior es un modelo de corrección de error (EC) del tipo presentado en la Sección 3.3 para procesos estacionarios [ver, en particular, (3.13) en el Ejemplo 3.3.4], lo que sugiere plantear para estimar β_2 un modelo del tipo

$$\nabla Y_{t1} = \mu_1 - \alpha_1 (Y_{t-1,1} + \beta_2' \mathbf{Y}_{t-1,2}) + \phi_0' \nabla \mathbf{Y}_{t2} + \sum_{i=1}^m \phi_i' \nabla \mathbf{Y}_{t-i} + U_t \quad (4.45)$$

en lugar de (4.43), donde $\phi_0 \equiv [\phi_{02}, \dots, \phi_{0M}]'$ recoge los posibles efectos contemporáneos de $\nabla \mathbf{Y}_{t2} \equiv [\nabla Y_{t2}, \dots, \nabla Y_{tM}]'$ sobre ∇Y_{t1} , y $\phi_i \equiv [\phi_{i1}, \phi_{i2}, \dots, \phi_{iM}]'$ recoge los posibles efectos de $\nabla \mathbf{Y}_{t-i} \equiv [\nabla Y_{t-i,1}, \nabla Y_{t-i,2}, \dots, \nabla Y_{t-i,M}]'$ sobre ∇Y_{t1} para $i = 1, \dots, m$.

Si, por ejemplo, $M = 2$, $m = 0$, (4.45) queda

$$\nabla Y_{t1} = \mu_1 - \alpha_1 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) + \phi_{02} \nabla Y_{t2} + U_t, \quad (4.46)$$

que es formalmente idéntico al modelo EC (3.13) en el Ejemplo 3.3.4.

Observación 3: En principio, (4.45)-(4.46) deben estimarse por **mínimos cuadrados no lineales** (MCNL) (como en la Observación 2 del Ejemplo 3.3.5). El estimador MCNL de β_2 en (4.45) suele representarse como $\hat{\beta}_2^{\text{EC}}$, que se denomina el **estimador EC** de β_2 . Al igual que $\hat{\beta}_2^{\text{MCO}}$ (ver Observación 1), $\hat{\beta}_2^{\text{EC}}$ también es superconsistente, aunque suele tener un sesgo más reducido que el estimador MCO. Por otro lado, la distribución asintótica de $\hat{\beta}_2^{\text{EC}}$ **no** es una distribución estándar, por lo que (como en el caso del estimador MCO) cualquier contraste estándar sobre β_2 basado en $\hat{\beta}_2^{\text{EC}}$ es (en general) incorrecto.

En el apartado siguiente se discuten algunas posibilidades para hacer inferencia estándar sobre ciertas funciones de α_1 y β_2 en (4.45), o de α_1 y β_2 en (4.46), basadas en las representaciones alternativas de (4.45)-(4.46) siguientes:

$$(4.45) \Rightarrow \quad \nabla Y_{t1} = \mu_1 - \alpha_1 Y_{t-1,1} - \delta_2' Y_{t-1,2} + \phi_0' \nabla Y_{t2} + \sum_{i=1}^m \phi_i' \nabla Y_{t-i} + U_t, \quad (4.47)$$

$$(4.46) \Rightarrow \quad \nabla Y_{t1} = \mu_1 - \alpha_1 Y_{t-1,1} - \delta_2 Y_{t-1,2} + \phi_{02} \nabla Y_{t2} + U_t, \quad (4.48)$$

donde $\delta_2 \equiv \alpha_1 \beta_2$ en (4.47) y $\delta_2 \equiv \alpha_1 \beta_2$ en (4.48). Nótese que (4.47)-(4.48) pueden estimarse por MCO porque son regresiones lineales, al contrario que (4.45)-(4.46), de manera que $\hat{\beta}_2^{\text{EC}} \equiv \hat{\delta}_2 / \hat{\alpha}_1$ en (4.45), o $\hat{\beta}_2^{\text{EC}} \equiv \hat{\delta}_2 / \hat{\alpha}_1$ en (4.46).

Observación 4: Un enfoque alternativo a los considerados en general en (4.42), (4.45) y (4.47), que es muy sencillo de utilizar en la práctica, consiste en estimar β_2 por MCO en la regresión

$$Y_{t1} = -\beta_0 - \beta_2' Y_{t2} + \sum_{i=-m}^m \theta_i' \nabla Y_{t+i,2} + A_t, \quad (4.49)$$

que es (4.42) con ∇Y_{t2} y m **retardos** y **adelantos** de ∇Y_{t2} añadidos. Esta modificación de (4.42) se espera que sirva para eliminar la endogeneidad de Y_{t2} [así como la posible autocorrelación de (U_t)] en (4.42). En tal caso, el estimador MCO de β_2 en (4.49) es consistente, asintóticamente Normal y asintóticamente eficiente, lo que permite hacer inferencias sobre el vector de cointegración β_2 utilizando dicho estimador de forma estándar; ver Saikkonen, P. (1991), Asymptotically efficient estimation of cointegration regressions, *Econometric Theory*, 7, 1-21.

Inferencia en regresiones con series I(1)

- ⇒ Por lo visto hasta ahora, parece que la teoría asintótica estándar nunca es aplicable cuando una regresión contiene variables explicativas I(1). Esto es cierto en regresiones espurias como (4.10), en las regresiones (4.17)-(4.18) asociadas con los contrastes ADF, y en algunas regresiones asociadas con la estimación de una relación de cointegración, como (4.42) y los modelos EC (4.45)-(4.46).
- ⇒ Sin embargo, **no** es cierto que el estadístico t de cada parámetro en cualquier regresión con regresores I(1) siga una distribución no estándar asintóticamente. De hecho, ni siquiera es cierto que el estadístico t de cada parámetro asociado con un regresor I(1) siga una distribución de ese tipo.

- ⇒ Por el contrario, puede demostrarse que el estadístico t de cualquier parámetro asociado **únicamente** con un regresor $I(0)$ (quizás después de reorganizar la regresión correspondiente), sigue una distribución $N(0,1)$ asintóticamente.
- ⇒ De manera análoga, el estadístico F para un contraste de significación conjunta, sigue la distribución asintótica habitual cuando todos los parámetros implicados en dicho contraste pueden figurar **al mismo tiempo** multiplicando a regresores $I(0)$.
- ⇒ Por el contrario, cualquier estadístico t o F que no satisfaga las condiciones anteriores, sigue en general una distribución asintótica no estándar.

Observación 5: Ver Sims, C.A., Stock, J.H., y Watson, M.W. (1990), Inference in linear time series models with some unit roots, *Econometrica*, 58, 113-144; Phillips, P.C.B. (1991), Optimal inference in cointegrated systems, *Econometrica*, 59, 283-306.

4.5.1 Ejemplo

En la representación (4.48) del modelo EC (4.46),

$$\nabla Y_{t1} = \mu_1 - \alpha_1 Y_{t-1,1} - \delta_2 Y_{t-1,2} + \phi_{02} \nabla Y_{t2} + U_t, \quad (4.50)$$

si $(Y_{t1}) \sim I(1)$ e $(Y_{t2}) \sim I(1)$ **no** están cointegrados, entonces no es posible reorganizar (4.50) para que α_1 sólo multiplique a una variable $I(0)$; por lo tanto, el estadístico t para α_1 en (4.50) sigue en este caso una distribución no estándar.

Dado que (4.50) es equivalente a (4.46), cuando (Y_{t1}) e (Y_{t2}) **sí** están cointegrados con vector de cointegración $[1, \beta_2]'$, α_1 en (4.46) multiplica a $(Y_{t-1,1} + \beta_2 Y_{t-1,2})$, que es $I(0)$, por lo que en este caso el estadístico t para α_1 en (4.50) sigue una distribución $N(0,1)$ asintóticamente.

Por otro lado, (4.50) también puede escribirse como

$$\nabla Y_{t1} = \mu_1 - \delta_2 (\beta_1 Y_{t-1,1} + Y_{t-1,2}) + \phi_{02} \nabla Y_{t2} + U_t, \quad (4.51)$$

donde $\beta_1 \equiv 1/\beta_2 \equiv \alpha_1/\delta_2$. En (4.51), el término al que multiplica δ_2 es $I(0)$ cuando (Y_{t1}) e (Y_{t2}) están cointegrados con vector de cointegración $[1, \beta_2]'$ ó $[\beta_1, 1]'$, por lo que en este caso el estadístico t para δ_2 en (4.50) también es $N(0,1)$ asintóticamente.

En todos los casos, los estadísticos t asociados con μ_1 y ϕ_{02} también siguen distribuciones $N(0,1)$ asintóticamente. En resumen, si (Y_{t1}) e (Y_{t2}) en (4.50) están cointegrados con vector de cointegración $[1, \beta_2]'$, entonces el estadístico t para cada uno de los parámetros

de (4.50) sigue una distribución $N(0,1)$ asintóticamente.

Observación 6: A pesar de lo anterior, cualquier estadístico F asociado con α_1 y δ_2 conjuntamente en (4.50), no sigue la distribución asintótica habitual porque no es posible escribir (4.50) de manera que α_1 y δ_2 multipliquen sólo a variables $I(0)$ al mismo tiempo. De hecho, si los estimadores $\hat{\alpha}_1$ y $\hat{\delta}_2$ siguieran asintóticamente una distribución conjunta Normal, entonces el estimador $\hat{\beta}_2^{\text{EC}} \equiv \hat{\delta}_2 / \hat{\alpha}_1$ de β_2 en (4.46) también sería asintóticamente Normal (con una tasa de convergencia asociada de orden $N^{-1/2}$), lo que contradiría el hecho de que $\hat{\beta}_2^{\text{EC}}$ es superconsistente; ver las observaciones 1 y 3 anteriores.

4.5.2 Ejemplo

La regresión (4.17) con $m = 1$ asociada con un contraste ADF para (Y_t) ,

$$\nabla Y_t = \mu + \gamma Y_{t-1} + \rho_1 \nabla Y_{t-1} + V_t, \quad (4.52)$$

puede escribirse como

$$\begin{aligned} \nabla Y_t &= \mu + \gamma Y_{t-1} + \rho_1 Y_{t-1} - \rho_1 Y_{t-2} + V_t \\ &= \mu + \gamma(Y_{t-1} - Y_{t-2}) + \rho_1 Y_{t-1} + (\gamma - \rho_1) Y_{t-2} + V_t, \end{aligned}$$

lo que indica que cuando $(Y_t) \sim I(1)$, (4.52) no puede escribirse de tal forma que γ sólo multiplique a una variable $I(0)$. Por lo tanto, el estadístico t para γ en (4.52) no sigue una

distribución estándar; ver Sección 4.3.

Por otro lado, $(Y_t) \sim I(1) \Leftrightarrow (\nabla Y_t) \sim I(0)$, por lo que el estadístico t para ρ_1 en (4.52) sí sigue una distribución $N(0,1)$ asintóticamente.

En general, cualquier estadístico F para el contraste de significación conjunta de varios retardos de ∇Y_t en (4.17), sigue la distribución asintótica habitual, lo que permite decidir de forma estándar cuántos retardos de ∇Y_t deben incluirse en (4.17).

Estimación a través de un modelo VAR – VEC

De acuerdo con los resultados de la Sección 4.4, cuando $(\mathbf{Y}_t) \sim CI(1,1)$, donde $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$, si (\mathbf{Y}_t) sigue un modelo VAR(p), entonces (ver 4.4.3)

$$\nabla \mathbf{Y}_t = \boldsymbol{\mu} - \mathbf{A}\mathbf{B}'\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{F}_i \nabla \mathbf{Y}_{t-i} + \mathbf{A}_t, \quad (4.53)$$

donde \mathbf{A} y \mathbf{B} son dos matrices $M \times P$ con $\text{rango}(\mathbf{A}) = \text{rango}(\mathbf{B}) = P$ ($1 \leq P \leq M - 1$). Las P columnas de \mathbf{B} contienen P vectores de cointegración (linealmente independientes) entre los componentes de (\mathbf{Y}_t) , de manera que $(\mathbf{B}'\mathbf{Y}_t) \sim I(0)$.

La estimación del modelo VEC (4.53) proporciona una estimación de cada una de las P relaciones de cointegración entre los M componentes de $(\mathbf{Y}_t) \sim I(1)$.

4.5.3 Ejemplo

Cuando $M = 2$, $P = 1$, $p = 1$ (ver Caso 2 en Ejemplo 4.4.1), (4.53) queda

$$\begin{bmatrix} \nabla Y_{t1} \\ \nabla Y_{t2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} - \begin{bmatrix} \alpha_1 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) \\ \alpha_2 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix}, \quad (4.54)$$

que es (4.26) imponiendo la normalización $\beta_1 = 1$. (4.54) es un modelo bivariante no lineal en los parámetros, que puede estimarse por MVE (utilizando procedimientos numéricos de optimización no lineal) suponiendo que $(\mathbf{A}_t) \sim \text{NIID}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{A}})$, con $\mathbf{A}_t \equiv [A_{t1}, A_{t2}]'$.

Alternativamente, (4.54) puede estimarse como se describe en la Observación 7 siguiente para el modelo VEC general (4.53); ver, por ejemplo, Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Observación 7: El procedimiento que se describe a continuación para estimar (4.53) es probablemente el más empleado en la actualidad; sin embargo, **no** es un procedimiento de estimación por MVE, sino (a lo sumo) por MVC (Máxima

Verosimilitud Condicionada), que consiste básicamente en estimar primero \mathbf{B} en (4.53) para estimar después $\boldsymbol{\mu}$, \mathbf{A} , \mathbf{F}_i ($i = 1, \dots, p - 1$) y $\boldsymbol{\Sigma}_A$ por MCO (ecuación por ecuación, como en un modelo VAR no restringido).

Paso 1: Estimar por MCO la regresión multivariante con término constante de $\nabla \mathbf{y}_t$ sobre $\nabla \mathbf{y}_{t-i}$ ($i = 1, \dots, p - 1$) y guardar los vectores ($M \times 1$) de residuos $\hat{\mathbf{v}}_{t1} \equiv \nabla \mathbf{y}_t - \nabla \hat{\mathbf{y}}_t$.

Paso 2: Estimar por MCO la regresión multivariante con término constante de \mathbf{y}_{t-1} sobre $\nabla \mathbf{y}_{t-i}$ ($i = 1, \dots, p - 1$) y guardar los vectores ($M \times 1$) de residuos $\hat{\mathbf{v}}_{t2} \equiv \mathbf{y}_{t-1} - \hat{\mathbf{y}}_{t-1}$.

Paso 3: Calcular las matrices de covarianzas muestrales entre los dos conjuntos de residuos anteriores:

$$\hat{\boldsymbol{\Sigma}}_{ij} \equiv \frac{1}{N} \sum_{t=1}^N \hat{\mathbf{v}}_{ti} \hat{\mathbf{v}}'_{tj} \quad (i = 1, 2; j = 1, 2),$$

donde N representa el número de residuos calculados en los pasos 1 y 2.

Paso 4: Calcular los escalares λ_i y los vectores \mathbf{z}_i ($i = 1, \dots, M$) que satisfacen las ecuaciones

$$(\lambda_i \hat{\boldsymbol{\Sigma}}_{22} - \hat{\boldsymbol{\Sigma}}_{21} \hat{\boldsymbol{\Sigma}}_{11}^{-1} \hat{\boldsymbol{\Sigma}}_{12}) \mathbf{z}_i = \mathbf{0} \quad (i = 1, \dots, M). \quad (\text{I})$$

Para ello: (1) Calcular los autovalores de la matriz simétrica y definida positiva

$$\boldsymbol{\Psi} \equiv \hat{\boldsymbol{\Psi}}'_{22} \hat{\boldsymbol{\Sigma}}_{21} \hat{\boldsymbol{\Sigma}}_{11}^{-1} \hat{\boldsymbol{\Sigma}}_{12} \hat{\boldsymbol{\Psi}}_{22}, \quad (\text{II})$$

donde $\hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\Psi}}'_{22} = \hat{\boldsymbol{\Sigma}}_{22}^{-1}$; los autovalores de $\boldsymbol{\Psi}$ son los escalares λ_i de (I). (2) Ordenar los autovalores de $\boldsymbol{\Psi}$ de mayor a menor ($\lambda_j > \lambda_i$ si $j < i$), calcular los autovectores correspondientes, y recogerlos en las columnas de \mathbf{P} tal que $\boldsymbol{\Psi} \equiv \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}'$, $\mathbf{P}' \mathbf{P} = \mathbf{I}$, donde $\boldsymbol{\Lambda} \equiv \text{diag}[\lambda_1, \dots, \lambda_M]$; las columnas de $\mathbf{Z} \equiv \hat{\boldsymbol{\Psi}}_{22} \mathbf{P}$ son los vectores \mathbf{z}_i de (I). La matriz \mathbf{Z} satisface además la relación

$$\mathbf{Z}' \hat{\boldsymbol{\Sigma}}_{22} \mathbf{Z} = \mathbf{I}. \quad (\text{III})$$

Paso 5: Las primeras P columnas de \mathbf{Z} constituyen la estimación $\hat{\mathbf{B}}$ por MV de \mathbf{B} en (4.53); $\hat{\mathbf{B}}$ está identificada de forma única por (III). Por último, estimar $\boldsymbol{\mu}$, \mathbf{A} , \mathbf{F}_i ($i = 1, \dots, p - 1$) y $\boldsymbol{\Sigma}_A$ en (4.53) por MCO, reemplazando \mathbf{B} por su estimación $\hat{\mathbf{B}}$. Los autovalores de $\hat{\boldsymbol{\Pi}} \equiv \hat{\mathbf{A}}\hat{\mathbf{B}}'$ son esencialmente los escalares $\lambda_1, \dots, \lambda_P$ del Paso 4 anterior.

Observación 8: Como en los casos de los estimadores MCO y EC discutidos en los apartados anteriores, el **estimador VEC** de \mathbf{B} en (4.53) asociado con el procedimiento de la Observación 7, también es superconsistente. En general (excepto quizás en muestras especialmente cortas), el sesgo del estimador VEC es menor que el de los estimadores MCO y EC. En muestras cortas, la varianza del estimador VEC puede ser ligeramente mayor que la de los otros dos estimadores, aunque este fenómeno tiende a desaparecer a medida que aumenta el tamaño muestral.

4.5.4 Ejemplo

En este ejemplo se ilustran los procedimientos de estimación descritos en esta sección con la serie bivalente simulada de la Figura 4.1 (Panel A); ver también Ejemplo 4.1.4.

Estimación directa – MCO en (4.42)

$$Y_{t1} = \gamma_{01} + \gamma_{21}Y_{t2} + U_{t1}, \text{ o bien } Y_{t2} = \gamma_{02} + \gamma_{22}Y_{t1} + U_{t2}. \quad (4.55)$$

Observación 1: Las tablas 4.8 y 4.9 contienen los resultados de las estimaciones por MCO (obtenidas con EViews 4.1 – May 23 2003 build) de las dos regresiones en (4.55). De la Tabla 4.8 se deduce que $\hat{\boldsymbol{\beta}}^{\text{MCO}} = [1, -0.8984]'$, mientras que de la Tabla 4.9 se deduce que $\hat{\boldsymbol{\beta}}^{\text{MCO}} = [-0.9935, 1]'$. En general, al colocar en el lado izquierdo de la regresión (4.42) diferentes componentes de (\mathbf{Y}_t) , se obtienen **diferentes** estimaciones del vector de cointegración.

Dependent Variable: VAR31				
Method: Least Squares				
Sample: 1 150				
Included observations: 150				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.917682	0.333624	2.750648	0.0067
VAR32	0.898398	0.025622	35.06348	0.0000
R-squared	0.892555	Mean dependent var	5.023577	
Adjusted R-squared	0.891829	S.D. dependent var	11.63319	
S.E. of regression	3.826084	Akaike info criterion	5.534804	
Sum squared resid	2166.560	Schwarz criterion	5.574946	
Log likelihood	-413.1103	F-statistic	1229.448	
Durbin-Watson stat	0.711044	Prob(F-statistic)	0.000000	

Tabla 4.8

Estimación por MCO de la primera regresión en (4.55)

Dependent Variable: VAR32				
Method: Least Squares				
Sample: 1 150				
Included observations: 150				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.420663	0.358027	-1.174947	0.2419
VAR31	0.993496	0.028334	35.06348	0.0000
R-squared	0.892555	Mean dependent var	4.570239	
Adjusted R-squared	0.891829	S.D. dependent var	12.23341	
S.E. of regression	4.023490	Akaike info criterion	5.635420	
Sum squared resid	2395.894	Schwarz criterion	5.675562	
Log likelihood	-420.6565	F-statistic	1229.448	
Durbin-Watson stat	0.783306	Prob(F-statistic)	0.000000	

Tabla 4.9

Estimación por MCO de la segunda regresión en (4.55)

Estimación EC – MCNL en (4.45)

$$\begin{aligned}\nabla Y_{t1} &= \mu_1 - \alpha_1 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) + A_{t1}, \\ \nabla Y_{t2} &= \mu_2 - \alpha_2 (Y_{t-1,1} + \beta_2 Y_{t-1,2}) + A_{t1}.\end{aligned}\tag{4.56}$$

Observación 2: Las tablas 4.10 y 4.11 contienen los resultados de las estimaciones por MCNL (obtenidas con EViews 4.1 – May 23 2003 build) de las dos regresiones en (4.56). De la Tabla 4.10 se deduce que $\hat{\beta}^{EC} = [1, -1.0545]'$, $\hat{\alpha}_1 = -0.1854$, mientras que de la Tabla 4.11 se deduce que $\hat{\beta}^{EC} = [1, -0.9771]'$, $\hat{\alpha}_1 = -0.5998$. Al igual que en la Observación 1, en este caso también se obtienen dos estimaciones diferentes del vector de cointegración.

Dependent Variable: D(VAR31)
 Method: Least Squares
 Sample(adjusted): 2 150
 Included observations: 149 after adjusting endpoints
 Convergence achieved after 4 iterations
 D(VAR31) = C(1) - C(2) * (VAR31(-1) + C(3) * VAR32(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.123727	0.181733	0.680817	0.4971
C(2)	-0.185393	0.043928	-4.220396	0.0000
C(3)	-1.054513	0.082920	-12.71721	0.0000

R-squared	0.133462	Mean dependent var	0.158273
Adjusted R-squared	0.121592	S.D. dependent var	2.168691
S.E. of regression	2.032572	Akaike info criterion	4.276410
Sum squared resid	603.1771	Schwarz criterion	4.336892
Log likelihood	-315.5925	Durbin-Watson stat	1.899665

Dependent Variable: D(VAR32)
 Method: Least Squares
 Sample(adjusted): 2 150
 Included observations: 149 after adjusting endpoints
 Convergence achieved after 7 iterations
 D(VAR32) = C(1) - C(2) * (VAR31(-1) + C(3) * VAR32(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.187440	0.292152	-0.641584	0.5221
C(2)	-0.599812	0.070618	-8.493790	0.0000
C(3)	-0.977096	0.037923	-25.76506	0.0000

R-squared	0.345770	Mean dependent var	0.131851
Adjusted R-squared	0.336807	S.D. dependent var	4.012361
S.E. of regression	3.267532	Akaike info criterion	5.225876
Sum squared resid	1558.808	Schwarz criterion	5.286358
Log likelihood	-386.3277	Durbin-Watson stat	2.112185

Tabla 4.10

Estimación por MCNL de la primera regresión en (4.56)

Tabla 4.11

Estimación por MCNL de la segunda regresión en (4.56)

Estimación VEC – MV en (4.53)

$$\begin{bmatrix} \nabla Y_{t1} \\ \nabla Y_{t2} \end{bmatrix} = - \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [1, \beta_2] \begin{bmatrix} Y_{t-1,1} \\ Y_{t-2,1} \end{bmatrix} + \begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix} \tag{4.57}$$

Observación 3: Estimar (4.57) por MV consiste en estimar las dos regresiones de (4.56) conjuntamente (con la restricción $\mu_1 = \mu_2 = 0$, a la vista de los resultados de las tablas 4.10-4.11). La Tabla 4.12 contiene el resultado de la estimación (obtenida con EViews 4.1 – May 23 2003 build) de (4.57), de donde se deduce que $\hat{\beta}^{VEC} = [1, -0.9919]'$,

$\hat{\alpha}_1 = -0.1979$, $\hat{\alpha}_2 = -0.5875$. Ahora se obtiene una **única** estimación del vector de cointegración.

Vector Error Correction Estimates		
Sample (adjusted): 2 150		
Included observations: 149 after adjusting endpoints		
Standard errors in () & t-statistics in []		
Cointegrating Eq:	CointEq1	
VAR31(-1)	1.000000	
VAR32(-1)	-0.991874	
	(0.03415)	
	[-29.0421]	
Error Correction:	D(VAR31)	D(VAR32)
CointEq1	0.197924	0.587534
	(0.04149)	(0.06661)
	[4.77033]	[8.82015]
R-squared	0.128619	0.343826
Adj. R-squared	0.128619	0.343826
Sum sq. resids	606.5483	1563.439
S.E. equation	2.024426	3.250196
Log likelihood	-316.0078	-386.5487
Akaike AIC	4.255138	5.201996
Schwarz SC	4.275299	5.222157
Mean dependent	0.158273	0.131851
S.D. dependent	2.168691	4.012361
Determinant Residual Covariance	38.42034	
Log Likelihood	-693.6601	
Log Likelihood (d.f. adjusted)	-694.6634	
Akaike Information Criteria	9.378033	
Schwarz Criteria	9.458675	

Tabla 4.12

Estimación por MV del modelo VEC (4.57)

Observación 4: La estimación de (4.57) por MVE proporciona unos resultados esencialmente idénticos a los que figuran en la Tabla 4.12. En la práctica, el método más recomendable para estimar relaciones de cointegración es el método basado en la estimación conjunta de todos los parámetros de un modelo VEC como (4.53); este método permite, además, contrastar hipótesis sobre los vectores de cointegración y sobre los factores de ajuste de una manera muy sencilla (ver Sección 4.6). No obstante, los estimadores MCO y EC pueden resultar útiles en los primeros pasos de un análisis aplicado, para obtener una primera impresión de las relaciones de cointegración consideradas, y para obtener estimaciones iniciales de los parámetros implicados en dichas relaciones.

4.6 CONTRASTES DE COINTEGRACIÓN

- ⇒ Los tres métodos para estimar relaciones de cointegración de la sección anterior, pueden extenderse para contrastar la existencia de dichas relaciones entre un conjunto de procesos univariantes $I(1)$.
- ⇒ El método basado en un modelo VEC también puede emplearse para determinar el número de relaciones de cointegración entre dichos procesos.
- ⇒ En esta sección sólo se discuten contrastes basados en la estimación de regresiones lineales como (4.42) y en la estimación de modelos VEC como (4.53).

Contraste de Engle-Granger

La manera más sencilla y probablemente más popular de contrastar la existencia de relaciones de cointegración entre los componentes de un proceso $(\mathbf{Y}_t) \sim I(1)$ consiste en:

- ⇒ Estimar (4.42) por MCO y guardar los residuos $\hat{u}_t = \hat{\beta}_0 + y_{t1} + \hat{\beta}'_2 \mathbf{y}_{t2}$.
- ⇒ Contrastar $H_0 : (U_t) \sim I(1)$ frente a $H_1 : (U_t) \sim I(0)$ mediante el contraste de

$H_0 : \gamma = 0$ frente a $H_1 : \gamma < 0$ en la regresión

$$\nabla U_t = \gamma U_{t-1} + \sum_{i=1}^m \rho_i \nabla U_{t-i} + A_t, \quad (4.58)$$

utilizando la serie de residuos (\hat{u}_t) del paso anterior en lugar de (U_t).

Observación 1: La hipótesis nula en el contraste anterior es que los componentes de $(\mathbf{Y}_t) \sim I(1)$ **no** están cointegrados; la hipótesis alternativa es que $(\mathbf{Y}_t) \sim CI(1,1)$. El contraste basado en la regresión (4.58) es, de hecho, un contraste de tipo ADF (ver Sección 4.3) sobre el proceso no observable $(U_t) \equiv (\beta_0 + Y_{t1} + \beta_2' \mathbf{Y}_{t2})$. Nótese, por otro lado, que (4.58) no incluye un término constante porque la media aritmética de la serie de residuos (\hat{u}_t) es exactamente igual a cero [por la presencia de β_0 en (4.42)].

Observación 2: La serie de residuos (\hat{u}_t) **no** es una realización particular del proceso $(U_t) \equiv (\beta_0 + Y_{t1} + \beta_2' \mathbf{Y}_{t2})$ (que no es observable), sino (en cierto sentido) una estimación superconsistente de dicho proceso. Por este motivo, la distribución asintótica del estadístico t para γ en (4.58) no es la distribución DF del estadístico τ en el contraste ADF correspondiente; en particular, dicha distribución depende ahora explícitamente de M [la dimensión de (\mathbf{Y}_t)], por lo que el estadístico t para γ en (4.58) suele representarse como $\tau(M)$. Los valores críticos de la distribución DF de τ (ver Tabla 4.1) **no** son aplicables; algunos valores críticos adecuados en este caso figuran en la Tabla 4.13.

Observación 3: Un problema importante asociado con el estadístico $\tau(M)$ es que su valor calculado (aunque no su distribución asintótica) depende de cuál sea el componente de (\mathbf{Y}_t) que figura en el lado izquierdo de la regresión (4.42). En consecuencia, el contraste de cointegración de Engle-Granger puede llevar (y de hecho lleva con frecuencia) a distintas conclusiones en función de cuál sea dicho componente.

N	Nivel de significación – $M = 2$			Nivel de significación – $M = 3$		
	1%	5%	10%	1%	5%	10%
50	-4.123	-3.461	-3.130	-4.592	-3.915	-3.578
100	-4.008	-3.398	-3.087	-4.441	-3.828	-3.514
200	-3.954	-3.368	-3.067	-4.368	-3.785	-3.483
500	-3.921	-3.350	-3.054	-4.326	-3.760	-3.464

Tabla 4.13

Valores críticos para el contraste de cointegración de Engle-Granger

4.6.1 Ejemplo

En este ejemplo se emplea la misma serie bivalente que en el Ejemplo 4.5.4. Los residuos de las dos regresiones en (4.55) estimadas por MCO (ver tablas 4.8-4.9) están representados en la Figura 4.11. Utilizando dichos residuos para estimar (4.58) con $m = 0$, los valores de los estadísticos $\tau(2)$ son -5.5878 (con la primera serie de residuos) y -5.9767 (con la segunda serie de residuos). En ambos casos, se rechaza claramente que $(U_t) \sim I(1)$ en favor de que $(U_t) \sim I(0)$; por lo tanto, $(\mathbf{Y}_t) \sim CI(1,1)$.

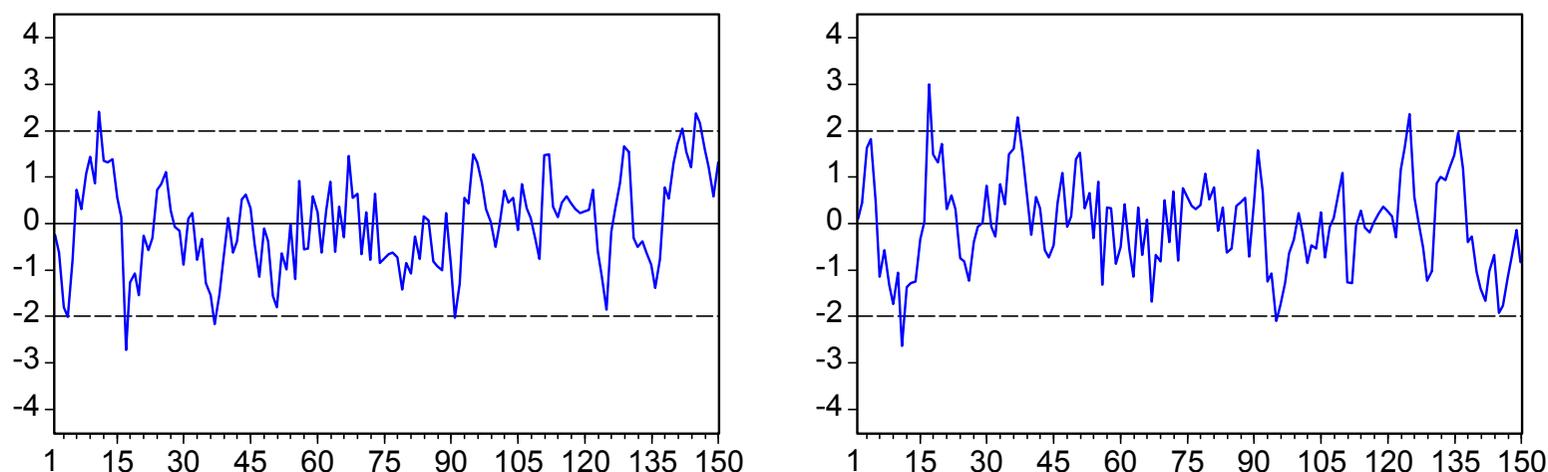


Figura 4.11

Residuos de las dos regresiones en (4.55) estimadas por MCO

Contrastes basados en un modelo VAR – VEC

- ⇒ En la práctica, el método más recomendable para contrastar la existencia de relaciones de cointegración entre los M componentes de $(\mathbf{Y}_t) \sim I(1)$, es algún método basado en la estimación conjunta de todos los parámetros de un modelo VEC como (4.53).
- ⇒ Un método de este tipo permite, además, determinar el número de relaciones de

cointegración [el rango P ($0 \leq P \leq M - 1$) de la matriz $\mathbf{\Pi} \equiv \mathbf{\Phi}(1) \equiv \mathbf{I} - \sum_{i=1}^p \mathbf{\Phi}_i$; ver 4.4.3], así como contrastar hipótesis sobre los vectores de cointegración y sobre los factores de ajuste de una manera muy sencilla.

⇒ Los procedimientos que se describen a continuación están basados en la estimación de (4.53) siguiendo los pasos de la Observación 7 de la Sección 4.5.

4.6.2 Definición

El logaritmo neperiano de la función de verosimilitud asociada con el modelo VEC (4.53) [que es un modelo VAR como (4.40)-(4.41) para $(\mathbf{Y}_t) \sim I(1)$, restringido a que existan P relaciones de cointegración ($0 \leq P \leq M - 1$) entre los M componentes de (\mathbf{Y}_t)], evaluada en las estimaciones de $\boldsymbol{\mu}$, \mathbf{A} , \mathbf{B} , \mathbf{F}_i ($i = 1, \dots, p - 1$) y $\boldsymbol{\Sigma}_A$ obtenidas siguiendo los pasos de la Observación 7 de la Sección 4.5, es igual a

$$L(P) \equiv -\frac{N}{2} \left[M(1 + \ln 2\pi) + \ln \left| \hat{\boldsymbol{\Sigma}}_{11} \right| \right] - \frac{N}{2} \sum_{i=1}^P \ln(1 - \lambda_i), \quad (4.59)$$

donde $\hat{\boldsymbol{\Sigma}}_{11}$ y λ_i ($i = 1, \dots, P$) están definidos en dicha Observación [en concreto, $\lambda_1, \dots, \lambda_P$ son los P autovalores más grandes de la matriz $\boldsymbol{\Psi}$ que figura en la expresión (II)].

Observación 1: (4.59) puede calcularse para cualquier valor de P (el número de relaciones de cointegración, que coincide con el rango de $\mathbf{\Pi} \equiv \mathbf{\Phi}(1) \equiv \mathbf{I} - \sum_{i=1}^P \mathbf{\Phi}_i$; ver 4.4.3) comprendido entre 0 y M , una vez que se han calculado y ordenado de mayor a menor los M autovalores λ_i ($i = 1, \dots, M$) de la matriz $\mathbf{\Psi}$ que figura en la expresión (II) de la Observación 7 de la Sección 4.5. Nótese que $0 < \lambda_i < 1$ ($i = 1, \dots, M$), ya que $\mathbf{\Psi}$ es simétrica y definida positiva ($\Rightarrow \lambda_i > 0$) y (4.59) no existiría si $\lambda_i \geq 1$ ($\Rightarrow \lambda_i < 1$).

4.6.3 Definición

El **estadístico de la traza** (en inglés, *trace statistic*) para contrastar $H_0: P \leq P_0$ (o bien $H_0: P = P_0$ cuando $P_0 = 0$) frente a $H_1: P \geq P_0 + 1$ (o bien $H_1: P = M$ cuando $P_0 = M - 1$), con $0 \leq P_0 \leq M - 1$, es el estadístico LR basado en (4.59)

$$\lambda_{\text{TR}}(P_0) \equiv 2[L(M) - L(P_0)] \equiv -N \sum_{i=P_0+1}^M \ln(1 - \lambda_i). \quad (4.60)$$

Observación 2: El estadístico (4.60) puede interpretarse como la suma de los últimos $M - P_0$ elementos en la diagonal principal de la matriz $-N \ln(\mathbf{I} - \mathbf{\Lambda})$, donde $\mathbf{\Lambda} \equiv \text{diag}[\lambda_1, \dots, \lambda_M]$. (4.60) puede utilizarse para llevar a cabo la secuencia (con $P_0 = 0, 1, \dots, M - 1$) siguiente de contrastes:

$$H_0: P = 0, H_1: P \geq 1;$$

$$H_0: P \leq 1, H_1: P \geq 2;$$

$$\vdots$$

$$H_0: P \leq M - 1, H_1: P = M.$$

Esta secuencia suele concluir cuando **no** se rechaza H_0 por primera vez. Por otro lado, el estadístico (4.60) **no** sigue una distribución estándar bajo H_0 porque la matriz $\mathbf{\Pi} \equiv \mathbf{\Phi}(1) \equiv \mathbf{I} - \sum_{i=1}^p \mathbf{\Phi}_i$ en (4.41) no se puede escribir como una matriz que sólo multiplique a variables $I(0)$. La Tabla 4.14 contiene algunos valores críticos adecuados para los contrastes basados en 4.6.3 y 4.6.4; ver MacKinnon, J.G., Haug, A.A., y Michelis, L. (1999), Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration, *Journal of Applied Econometrics*, 14, 563-577.

4.6.4 Definición

El **estadístico del máximo** (en inglés, *max statistic*) para contrastar $H_0 : P = P_0$ frente a $H_1 : P = P_0 + 1$, con $0 \leq P_0 \leq M - 1$, es

$$\lambda_{\text{MAX}}(P_0) \equiv -N \ln(1 - \lambda_{P_0+1}) \equiv -N \ln(1 - \lambda_{\text{MAX}}). \quad (4.61)$$

Observación 3: En (4.61), $\lambda_{\text{MAX}} \equiv \lambda_{P_0+1}$ es el mayor autovalor entre los asociados con los autovectores **no** incluidos en la estimación de \mathbf{B} bajo H_0 ; ver el Paso 4 en la Observación 7 de la Sección 4.5. (4.61) puede utilizarse para llevar a cabo la secuencia (con $P_0 = 0, 1, \dots, M - 1$) siguiente de contrastes:

$$\begin{aligned} H_0 : P = 0, H_1 : P = 1; \\ H_0 : P = 1, H_1 : P = 2; \\ \vdots \\ H_0 : P = M - 1, H_1 : P = M. \end{aligned}$$

Esta secuencia puede concluir cuando **no** se rechaza H_0 por primera vez. Como en el caso de (4.60), (4.61) **no** sigue una distribución estándar bajo H_0 ; ver Tabla 4.14.

$M - P_0$	Nivel de significación – $\lambda_{TR}(P_0)$			Nivel de significación – $\lambda_{MAX}(P_0)$		
	1%	5%	10%	1%	5%	10%
Modelo VAR sin término constante – Relación de cointegración sin término constante						
1	6.9406	4.1299	2.9762	6.9406	4.1299	2.9762
2	16.3619	12.3209	10.4746	15.0913	11.2248	9.4748
3	29.5135	24.2760	21.7772	22.2517	17.7973	15.7174
4	46.5716	40.1749	37.0354	29.0603	24.1592	21.8367
Modelo VAR sin término constante – Relación de cointegración con término constante						
1	12.7608	9.1645	7.5567	12.7608	9.1645	7.5567
2	25.0781	20.2618	17.9804	20.1612	15.8921	13.9059
3	41.1950	35.1928	32.2684	27.0678	22.2996	20.0501
4	61.2669	54.0790	50.5253	33.7329	28.5881	26.1212
Modelo VAR con término constante – Relación de cointegración con término constante						
1	6.6349	3.8415	2.7055	6.6349	3.8415	2.7055
2	19.9371	15.4947	13.4288	18.5200	14.2646	12.2965
3	35.4582	29.7971	27.0669	25.8612	21.1316	18.8928
4	54.6815	47.8561	44.4936	32.7153	27.5843	25.1241

Tabla 4.14

Valores críticos asintóticos (MacKinnon-Haug-Michelis) para los contrastes basados en (4.60)-(4.61)

Observación 4: En algunas ocasiones, los resultados de los contrastes basados en los estadísticos (4.60) y (4.61) pueden no coincidir. Dado que la hipótesis alternativa del contraste basado en (4.61) es más concreta, dicho contraste suele ser preferible para determinar el número de relaciones de cointegración en caso de conflicto.

4.6.5 Ejemplo

En este ejemplo se emplea la misma serie bivalente que en el Ejemplo 4.5.4 (ver también Ejemplo 4.6.1). La Tabla 4.12 contiene la estimación del modelo VEC (4.57) [que incluye la restricción $(\mathbf{Y}_t) \sim CI(1,1)$]. La Tabla 4.15 contiene los resultados de los contrastes de cointegración basados en 4.6.3 y en 4.6.4 (obtenidos con EViews 4.1 – May 23 2003 build).

Contrastes de hipótesis sobre vectores de cointegración

- ⇒ Los resultados de la Sección 4.5 sobre las propiedades de los estimadores MCO y EC de vectores de cointegración, indican que (en general) no es posible hacer inferencias estándar sobre dichos vectores utilizando los estimadores MCO y EC.
- ⇒ Por ejemplo, los estadísticos t asociados con γ_{21} , γ_{22} en (4.55) (ver tablas 4.8-4.9) y con β_2 en (4.56) (ver tablas 4.10-4.11) **no** siguen distribuciones $N(0,1)$ asintóticamente [al contrario, por ejemplo, que los estadísticos t asociados con μ_1 , α_1 , μ_2 , α_2 en (4.56)].

Sample(adjusted): 2 150 - Included observations: 149 after adjusting endpoints
 Series: VAR31 VAR32 - Lags interval (in first differences): No lags
 Trend assumption: No deterministic trend

Trace Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.355096	65.85655	12.53	16.31
At most 1	0.003331	0.497188	3.84	6.51

*(**) denotes rejection of the hypothesis at the 5%(1%) level
 Trace test indicates 1 cointegrating equation(s) at both 5% and 1% levels

Max-Eigenvalue Test

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.355096	65.35936	11.44	15.69
1	0.003331	0.497188	3.84	6.51

*(**) denotes rejection of the hypothesis at the 5%(1%) level
 Max-eigenvalue test indicates 1 cointegrating equation(s) at both 5% and 1% levels

Tabla 4.15

Contrastes de cointegración en el Ejemplo 4.6.5

⇒ La utilización del estimador VEC resuelve este problema. En general, para contrastar l restricciones sobre los P vectores de cointegración contenidos en la matriz \mathbf{B} del modelo VEC (4.53), puede utilizarse el estadístico LR basado en (4.59)

$$\lambda^*(l) \equiv -N \sum_{i=1}^P \left[\ln(1 - \lambda_i) - \ln(1 - \lambda_i^*) \right], \quad (4.62)$$

donde $\lambda_1^*, \dots, \lambda_P^*$ son los autovalores del modelo VEC restringido por la hipótesis nula correspondiente. (4.62) sigue una distribución $\chi^2(l)$ (asintóticamente) cuando las l restricciones contrastadas son ciertas; ver, por ejemplo, Harris, R. (1995), *Cointegration Analysis in Econometric Modelling*, Prentice-Hall.

⇒ (4.62) también puede utilizarse para contrastar hipótesis sobre los factores de ajuste contenidos en la matriz \mathbf{A} del modelo VEC (4.53). Si $P = 1$ y sólo se desea contrastar una hipótesis sobre un factor de ajuste α_i ($1 \leq i \leq M$), también puede emplearse el estadístico t asociado con α_i en un modelo EC como (4.45)-(4.48) (en este caso, si no se rechaza que $\alpha_i = 0$, suele decirse que (Y_{ti}) es una variable **débilmente exógena**).

4.6.6 Ejemplo

En este ejemplo se emplea la misma serie bivalente que en el Ejemplo 4.5.4 (ver también ejemplos 4.6.1 y 4.6.5). La Tabla 4.12 contiene la estimación del modelo VEC (4.57) [que incluye la restricción $(\mathbf{Y}_t) \sim CI(1,1)$]. La Tabla 4.16 contiene el resultado del contraste de la hipótesis nula $\beta_2 = -1$ en (4.57) (obtenido con EViews 4.1 – May 23 2003 build), que indica claramente que no puede rechazarse dicha hipótesis nula.

Vector Error Correction Estimates		
Sample (adjusted): 2 150		
Included observations: 149 after adjusting endpoints		
Standard errors in () & t-statistics in []		
Cointegration Restrictions: $B(1,1) = 1, B(1,2) = -1$		
Restrictions identify all cointegrating vectors		
LR test for binding restrictions (rank = 1):		
Chi-square(1)	0.052593	
Probability	0.818611	
Cointegrating Eq:	CointEq1	
VAR31(-1)	1.000000	
VAR32(-1)	-1.000000	
Error Correction:	D(VAR31)	D(VAR32)
CointEq1	0.197194 (0.04120) [4.78605]	0.583260 (0.06620) [8.81000]
R-squared	0.129385	0.343305
Adj. R-squared	0.129385	0.343305
Sum sq. resids	606.0150	1564.679
S.E. equation	2.023535	3.251485
Log likelihood	-315.9422	-386.6078
Akaike AIC	4.254258	5.202790
Schwarz SC	4.274419	5.222950
Mean dependent	0.158273	0.131851
S.D. dependent	2.168691	4.012361
Determinant Residual Covariance		38.43391
Log Likelihood		-693.6864
Log Likelihood (d.f. adjusted)		-694.6897
Akaike Information Criteria		9.378386
Schwarz Criteria		9.459028

Tabla 4.16

Estimación restringida del modelo VEC (4.57)

4.7 ELABORACIÓN DE MODELOS MULTIVARIANTES EN LA PRÁCTICA

En esta última sección se ilustran algunos aspectos importantes de la elaboración de modelos multivariantes en la práctica, usando la serie trivariante (\mathbf{y}_t) , $\mathbf{y}_t \equiv [w_t, y_t, z_t]'$, simulada de la Figura 4.3 (Panel A) [azul $\equiv (w_t)$, rojo $\equiv (y_t)$, gris $\equiv (z_t)$].

Paso 1

- ⇒ Elaborar un **modelo univariante** para cada serie (Sección 2.6), prestando especial atención a: (i) el **orden de integración** adecuado en cada caso (Sección 4.3), y (ii) la presencia de **sucesos de intervención** y de **observaciones influyentes** (Sección 3.2).
- ⇒ Los **instrumentos gráficos** son tan importantes en las operaciones anteriores como cualesquiera contrastes formales (que son, en general, bastante menos informativos).
- ⇒ Si todas las series consideradas son $I(0)$, elaborar un **modelo estacionario** como alguno de los descritos en el Tema 3, prestando especial atención a si las posibles **relaciones de causalidad** van en una sola dirección (Sección 3.3) o en varias direcciones (Sección 3.4).
- ⇒ Si todas las series consideradas son $I(d)$ (con d en muchos casos igual a 1), ir al Paso 2.

⇒ Cada una de las series (w_t) , (y_t) , (z_t) está adecuadamente descrita por un paseo aleatorio sin término constante, de manera que $(\mathbf{y}_t) \sim I(1)$. La Figura 4.12 contiene una representación gráfica de las series (∇w_t) , (∇y_t) , (∇z_t) .

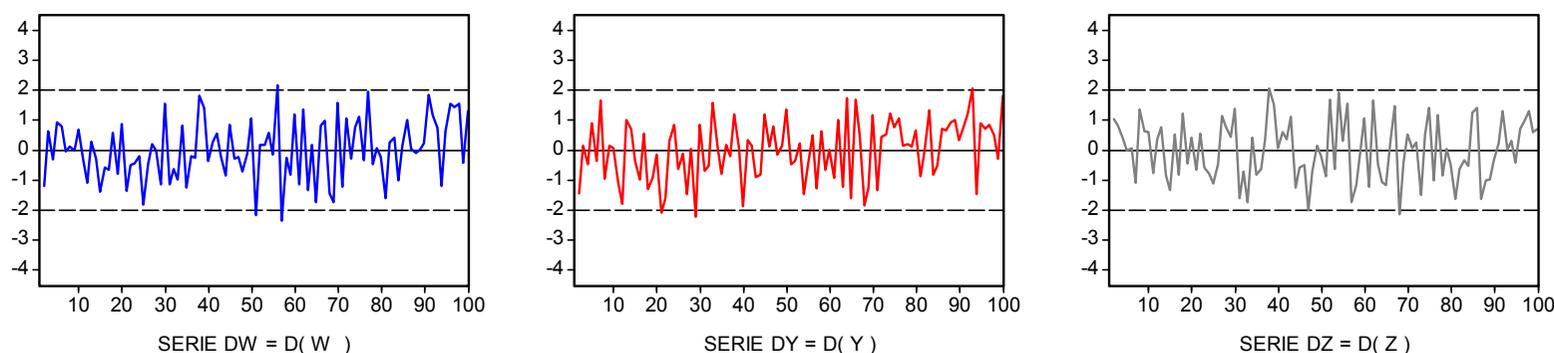


Figura 4.12

Series de la Figura 4.3 (Panel A) diferenciadas

Paso 2

⇒ Contrastar la existencia de **cointegración** entre las series consideradas (Sección 4.6), prestando especial atención a: (i) el número de relaciones de cointegración, y (ii) el aspecto gráfico de cualquier serie potencial de desequilibrios transitorios. Para ello:

- ⇒ Elegir el **orden** adecuado de un **modelo VAR** como (4.40) para las series **no** diferenciadas, con los mismos criterios que si las series fuesen estacionarias (Sección 3.4).
- ⇒ Estimar el modelo VAR anterior y determinar el **rango** de la matriz $\mathbf{\Pi}$ en la representación VEC (4.41) (Sección 4.6).
- ⇒ Si el rango de $\mathbf{\Pi}$ es igual a cero (ausencia de cointegración), elaborar un modelo (estacionario si $d = 1$) para las series diferenciadas. En caso contrario, estimar la representación VEC (4.41) imponiendo la presencia de cointegración, lo que resulta en un modelo VEC como (4.53) (Sección 4.5).

Observación 1: En la práctica, el método más recomendable para estimar relaciones de cointegración es el método basado en la estimación conjunta de todos los parámetros de un modelo VEC como (4.53); este método permite, además, contrastar hipótesis sobre los vectores de cointegración y sobre los factores de ajuste de una manera muy sencilla (Sección 4.6). No obstante, los estimadores MCO y EC de la Sección 4.5 pueden resultar útiles en los primeros pasos de un análisis aplicado, para obtener una primera impresión de las relaciones de cointegración consideradas, y para obtener estimaciones iniciales de los parámetros implicados en dichas relaciones.

Observación 2: En algunas ocasiones, para determinar el rango de la matriz $\mathbf{\Pi}$ en (4.41), así como para estimar conjuntamente todos los parámetros de (4.53), se puede incluir un término constante en cada relación de cointegración. En la práctica, las posibilidades más frecuentes a este respecto son las que se indican en la Tabla 4.14. Los estimadores MCO y EC también pueden utilizarse para decidir sobre esta cuestión; ver Ejemplo 4.5.4.

⇒ La Tabla 4.17 indica que un modelo VAR(1) ó VAR(2) (probablemente sin término constante) con raíces unitarias, podría resultar adecuado para la serie $(\mathbf{y}_t) \sim I(1)$ **no** diferenciada. La Tabla 4.18 contiene las raíces inversas de $|\mathbf{I} - \hat{\Phi}_1 x - \hat{\Phi}_2 x^2| = 0$ asociadas con un modelo VAR(2) (sin término constante) estimado para $(\mathbf{y}_t) \sim I(1)$.

VAR Lag Order Selection Criteria				
Endogenous variables: W Y Z				
Sample: 1 100				
Included observations: 96				
Exogenous variables: C				
Lag	LogL	LR	AIC	SC
0	-360.6905	NA	7.576886	7.657022
1	-37.32356	619.7867	1.027574	1.348118*
2	-26.45854	20.14557*	0.988720*	1.549671
3	-22.29582	7.458204	1.089496	1.890855
4	-18.86455	5.933236	1.205511	2.247278
Exogenous variables: None				
1	-38.10346	NA	0.981322	1.221730*
2	-27.15590	20.52668*	0.940748*	1.421563
3	-23.00314	7.526876	1.041732	1.762955
4	-19.56024	6.025067	1.157505	2.119136

LR: sequential modified LR test statistic (each test at 5% level)
AIC: Akaike information criterion
SC: Schwarz information criterion

Tabla 4.17

Selección del orden de un modelo VAR para las series de la Figura 4.3 (Panel A)

Roots of Characteristic Polynomial	
Endogenous variables: W Y Z	
Exogenous variables: None	
Lag specification: 1 2	
Root	Modulus
0.992508 - 0.023204 i	0.992779
0.992508 + 0.023204 i	0.992779
0.314149 - 0.274306 i	0.417053
0.314149 + 0.274306 i	0.417053
-0.266018	0.266018
-0.018241	0.018241

Tabla 4.18

Raíces de un modelo VAR(2) estimado para las series de la Figura 4.3 (Panel A)

⇒ La Tabla 4.19 (cuyas tres secciones corresponden a las tres posibilidades indicadas en la Tabla 4.14) indica la existencia de una única relación de cointegración para un nivel de significación del 1% (aunque en algunos casos no puede rechazarse la existencia de dos relaciones de cointegración al 5%).

Observación 3: La estimación por MCO de las tres regresiones posibles (4.42) (una con cada una de las tres series en el lado izquierdo), así como la estimación por MCNL de los tres modelos EC posibles (4.45) y la estimación de un modelo VAR(2) para las tres series no diferenciadas, sugieren que el caso más plausible entre los tres considerados en la Tabla 4.19 es el Caso 1. La conclusión de que existe una única relación de cointegración habría sido más clara si el modelo sobre el que se realizan los contrastes de cointegración hubiera sido un VAR(1) en vez de un VAR(2).

Sample (adjusted): 3 100 - Included observations: 98 after adjusting endpoints
Series: W Y Z - Lags interval (in first differences): 1 to 1

Case 1: Trend assumption: No deterministic trend

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325919	54.91357	24.31	29.75
At most 1	0.138488	16.26189	12.53	16.31
At most 2	0.016729	1.653347	3.84	6.51

Trace test indicates 2 cointegrating equation(s) at the 5% level

Trace test indicates 1 cointegrating equation(s) at the 1% level

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325919	38.65168	17.89	22.99
1	0.138488	14.60855	11.44	15.69
2	0.016729	1.653347	3.84	6.51

Max-eigenvalue test indicates 2 cointegrating equation(s) at the 5% level

Max-eigenvalue test indicates 1 cointegrating equation(s) at the 1% level

Case 2: Trend assumption: No deterministic trend (restricted constant)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325996	56.78582	34.91	41.07
At most 1	0.140322	18.12291	19.96	24.60
At most 2	0.033168	3.305551	9.24	12.97

Trace test indicates 1 cointegrating equation(s) at both 5% and 1% levels

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325996	38.66291	22.00	26.81
1	0.140322	14.81736	15.67	20.20
2	0.033168	3.305551	9.24	12.97

Max-eigenvalue test indicates 1 cointegrating equation(s) at both 5% and 1% levels

Case 3: Trend assumption: Linear deterministic trend				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325399	55.10275	29.68	35.65
At most 1	0.133274	16.52668	15.41	20.04
At most 2	0.025282	2.509450	3.76	6.65
Trace test indicates 2 cointegrating equation(s) at the 5% level				
Trace test indicates 1 cointegrating equation(s) at the 1% level				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.325399	38.57607	20.97	25.52
1	0.133274	14.01723	14.07	18.63
2	0.025282	2.509450	3.76	6.65
Max-eigenvalue test indicates 1 cointegrating equation(s) at both 5% and 1% levels				

Tabla 4.19

Contrastes de cointegración entre las series de la Figura 4.3 (Panel A)

⇒ La Tabla 4.20 contiene un resumen de la estimación de un modelo VEC como (4.53) con $p = 2$, que incluye una única relación de cointegración entre las tres series consideradas. La Figura 4.13 contiene las tres series de residuos asociadas con el modelo estimado. Por último, la Figura 4.14 contiene la serie de desequilibrios transitorios [que es claramente una serie $I(0)$] con respecto a la relación de cointegración estimada.

Vector Error Correction Estimates			
Sample(adjusted): 3 100			
Included observations: 98 after adjusting endpoints			
Standard errors in ()			
Cointegrating Eq:	CointEq1		
W(-1)	1.000000		
Y(-1)	-0.981563 (0.01267)		
Z(-1)	-1.017802 (0.02103)		
Error Correction:	D(W)	D(Y)	D(Z)
CointEq1	0.221151 (0.18256)	0.553842 (0.12556)	0.170359 (0.15636)
D(W(-1))	-0.537182 (0.21918)	-0.441359 (0.15075)	-0.344034 (0.18773)
D(Y(-1))	0.236430 (0.21771)	0.181466 (0.14974)	0.155884 (0.18647)
D(Z(-1))	0.467920 (0.21957)	0.414865 (0.15102)	0.323425 (0.18807)
Determinant Residual Covariance			0.000453
Log Likelihood			-33.77638
Log Likelihood (d.f. adjusted)			-39.90227
Akaike Information Criteria			1.120454
Schwarz Criteria			1.516113

Tabla 4.20

Estimación de un modelo VEC con las series de la Figura 4.3 (Panel A)

Observación 4: El modelo VEC de la Tabla 4.20 puede escribirse como:

$$\begin{bmatrix} \nabla w_t \\ \nabla y_t \\ \nabla z_t \end{bmatrix} = - \begin{bmatrix} -0.22 \\ -0.55 \\ -0.17 \end{bmatrix} \underbrace{[1, -0.98, -1.02]}_{\hat{\alpha}'\hat{\mathbf{y}}_{t-1}} \begin{bmatrix} w_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} -0.54 & 0.24 & 0.47 \\ -0.44 & 0.18 & 0.41 \\ -0.34 & 0.16 & 0.32 \end{bmatrix}}_{\hat{\mathbf{F}}_1} \begin{bmatrix} \nabla w_{t-1} \\ \nabla y_{t-1} \\ \nabla z_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{a}_{t1} \\ \hat{a}_{t2} \\ \hat{a}_{t3} \end{bmatrix},$$

de manera que la serie $(\hat{u}_t) \equiv (\hat{\beta}'\mathbf{y}_t) \equiv (w_t - 0.98y_t - 1.02z_t)$ representa la secuencia de desequilibrios transitorios entre los componentes de (\mathbf{y}_t) con respecto a su relación de cointegración (una relación estable o de equilibrio a largo plazo) estimada $w_t - 0.98y_t - 1.02z_t \approx 0$; ver Figura 4.14. Por otro lado, el vector $\hat{\alpha} \equiv [-0.22, -0.55, -0.17]'$ contiene los factores de ajuste estimados de cada componente de (\mathbf{y}_t) frente a (\hat{u}_t) .

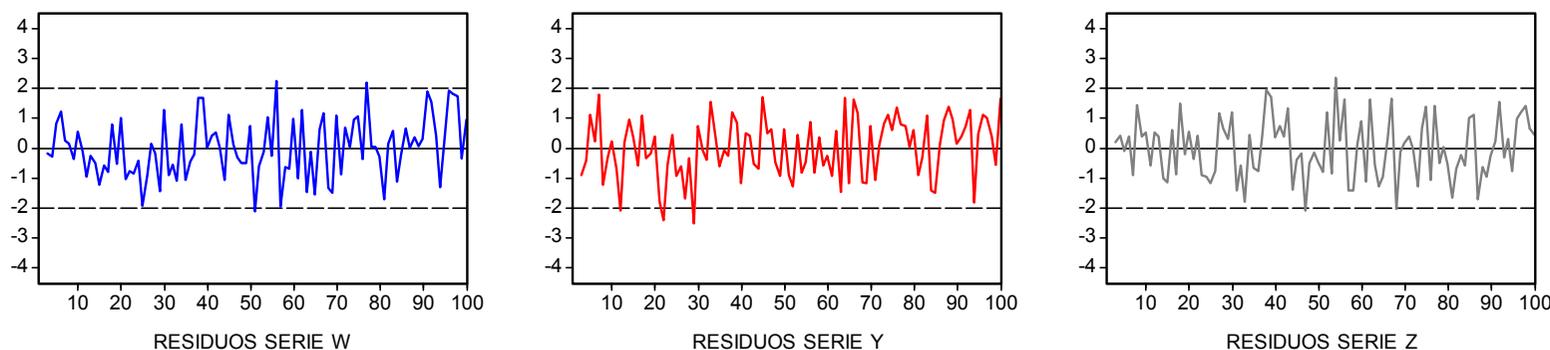


Figura 4.13

Residuos asociados con el modelo VEC de la Tabla 4.20

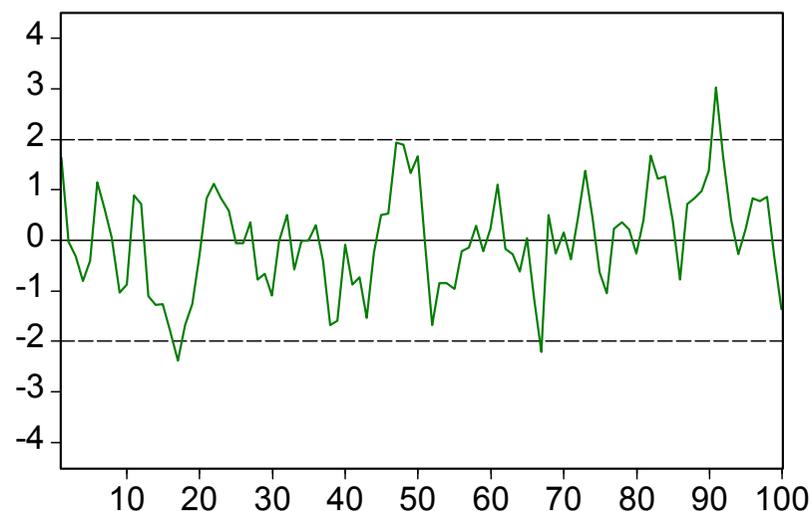


Figura 4.14

Serie de desequilibrios transitorios (estacionarios) asociados con el modelo VEC de la Tabla 4.20

Paso 3

- ⇒ Analizar cada uno de los vectores de cointegración estimados y los factores de ajuste correspondientes. En particular, contrastar las hipótesis sobre **B** y **A** en (4.53) que puedan ser relevantes en cada caso (Sección 4.6).
- ⇒ La Tabla 4.21 contiene un resumen de la estimación de un modelo VEC como (4.53) con $p = 2$, restringido bajo la hipótesis de que $\beta_1 = 1$, $\beta_2 = \beta_3 = -1$, $\alpha_1 = \alpha_3 = 0$.

Restricted Vector Error Correction Estimates			
Sample (adjusted): 3 100			
Included observations: 98 after adjusting endpoints			
Cointegration Restrictions:			
B(1,1) = 1, B(1,2) = -1, B(1,3) = -1			
A(1,1) = 0, A(3,1) = 0			
LR test for binding restrictions (rank = 1):			
Chi-square(4): 2.203389 Probability: 0.698409			
Cointegrating Eq: CointEq1			
W(-1)	1.000000		
Y(-1)	-1.000000		
Z(-1)	-1.000000		
Error Correction:	D(W)	D(Y)	D(Z)
CointEq1	0.000000 (0.00000)	0.496677 (0.07617)	0.000000 (0.00000)
D(W(-1))	-0.483588 (0.21998)	-0.419140 (0.15256)	-0.305272 (0.18816)
D(Y(-1))	0.205808 (0.22049)	0.185601 (0.15291)	0.134115 (0.18860)
D(Z(-1))	0.401630 (0.22199)	0.397586 (0.15395)	0.275711 (0.18988)
Determinant Residual Covariance			0.000462
Log Likelihood			-34.87808
Log Likelihood (d.f. adjusted)			-40.80307
Akaike Information Criteria			1.138838
Schwarz Criteria			1.534497

Tabla 4.21

Estimación de un modelo VEC restringido con las series de la Figura 4.3 (Panel A)

Observación 5: El estadístico LR (4.62) siempre puede calcularse como $\lambda^*(l) = 2[L(P) - L^*(P)]$, donde $L(P)$ es el logaritmo neperiano de la función de verosimilitud (4.59) (no restringida) y $L^*(P)$ es el logaritmo de la función de verosimilitud restringida bajo la hipótesis nula que se desea contrastar; ver, por ejemplo, la fila **Log Likelihood** en las tablas 4.20 y 4.21. Por otro lado, nótese que el número de restricciones que figuran en la Tabla 4.21 es $l = 4$ porque la condición de normalización $\beta_1 = 1$ también está incluida en la Tabla 4.20.

Paso 4

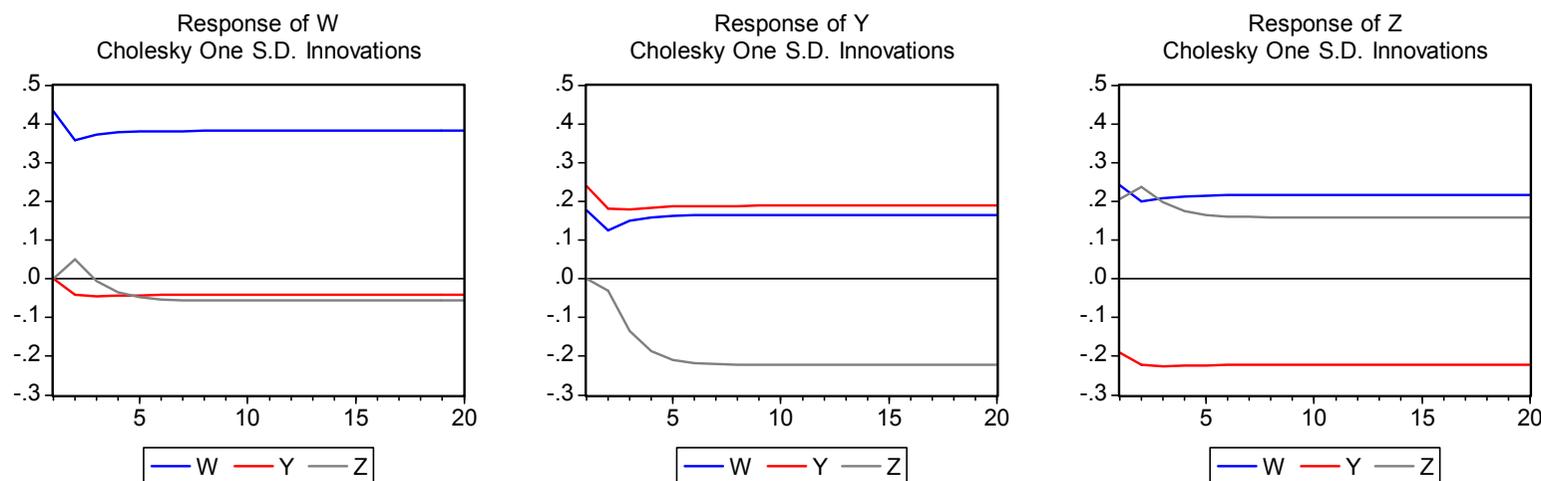
⇒ Analizar las funciones de respuesta a impulsos y las descomposiciones de las varianzas de los errores de previsión (Sección 3.4), a partir de la representación VAR estándar del modelo VEC (4.53). En última instancia, utilizar el modelo estimado para calcular previsiones de las series consideradas (Sección 3.4).

Observación 5: El modelo VEC de la Tabla 4.20 escrito como en la Observación 4, $\nabla \mathbf{y}_t = -\hat{\Pi} \mathbf{y}_{t-1} + \hat{\mathbf{F}}_1 \nabla \mathbf{y}_{t-1} + \hat{\mathbf{a}}_t$, donde $\hat{\Pi} \equiv \hat{\alpha} \hat{\beta}'$, se puede escribir como un modelo VAR(2) estándar, $\mathbf{y}_t = \hat{\Phi}_1 \mathbf{y}_{t-1} + \hat{\Phi}_2 \mathbf{y}_{t-2} + \hat{\mathbf{a}}_t$, donde (ver las definiciones 4.1.5 y 4.1.3) $\hat{\Phi}_2 \equiv -\hat{\mathbf{F}}_1$, $\hat{\Phi}_1 \equiv \mathbf{I} - \hat{\Pi} - \hat{\Phi}_2$, es decir

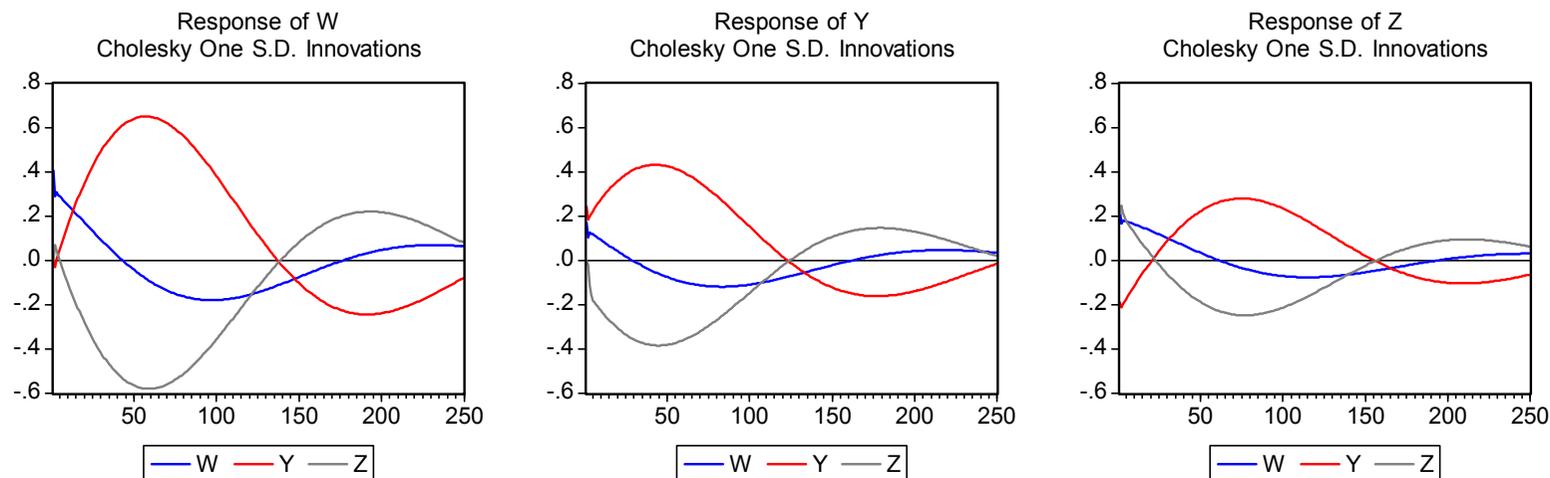
$$\begin{bmatrix} w_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.68 & 0.02 & 0.24 \\ 0.11 & 0.64 & -0.15 \\ -0.17 & -0.01 & 1.15 \end{bmatrix} \begin{bmatrix} w_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 0.54 & -0.24 & -0.47 \\ 0.44 & -0.18 & -0.41 \\ 0.34 & -0.16 & -0.32 \end{bmatrix} \begin{bmatrix} w_{t-2} \\ y_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} \hat{a}_{t1} \\ \hat{a}_{t2} \\ \hat{a}_{t3} \end{bmatrix},$$

que puede utilizarse como cualquier modelo VAR estándar para los fines mencionados anteriormente (Sección 3.4).

Observación 6: En la representación de Wold (ver Definición 3.4.3) de un modelo VAR(p) estimado con raíces unitarias, ocurre que la secuencia de matrices $\hat{\Psi}_i = \sum_{j=1}^p \hat{\Phi}_j \hat{\Psi}_{i-j}$ ($i = 0, 1, 2, \dots$), con $\hat{\Psi}_0 \equiv \mathbf{I}$, $\hat{\Psi}_i \equiv \mathbf{0}$ si $i < 0$, **no** converge a $\mathbf{0}$ (como ocurre en el caso de un modelo VAR estacionario). Esto significa que cualquier función de respuesta a impulsos (ver las definiciones 3.4.7-3.4.9) asociada con un modelo VEC estimado que incluye alguna relación de cointegración, converge a una cantidad **distinta** de cero. Las propiedades asintóticas de los estimadores de Ψ_i ($i = 0, 1, 2, \dots$) en este caso, son muy diferentes de las asociadas con un modelo VAR estacionario; ver, por ejemplo, Phillips, P.C.B. (1998), Impulse response and forecast error asymptotics in nonstationary VARs, *Journal of Econometrics*, 83, 21-56. No obstante, lo anterior pone de manifiesto que la secuencia ($\hat{\Psi}_i: i = 0, 1, 2, \dots$) debe calcularse en el caso de un modelo VEC con relaciones de cointegración, a partir de la representación VAR de dicho modelo estimado (como en la figura siguiente),



no a partir de un modelo VAR no restringido estimado con las series no diferenciadas (como en la figura siguiente).



En las dos figuras anteriores, se ha representado la IRF de la Definición 3.4.9, ordenando las series como se indica en el pie de cada gráfico. La figura de la página anterior se ha obtenido a partir de la representación VAR modelo VEC de la Tabla 4.20; ver Observación 5. La figura de esta página se ha obtenido a partir de un modelo VAR(2) no restringido estimado con las tres series no diferenciadas; ver Tabla 4.8.