

CAPÍTULO 2

MODELOS DE SERIES DE TIEMPO

ESTACIONARIAS UNIVARIADOS

En esta sección desarrollamos las herramientas básicas del análisis de series de tiempo para modelos con una sola variable. Ciertamente, modelar una variable puede parecer una propuesta modesta; la mayoría de los problemas en economía se refieren a la interacción entre variables económicas y, en particular, en la influencia que unas tienen sobre otras.

Existen dos razones para enfocarse en modelos univariados. La primera es de tipo pedagógico y se refiere a que los conceptos que se desarrollan en este capítulo pueden ser extendidos a modelos multivariados directamente, pero resulta más fácil entenderlos en un contexto univariado. Segundo, y más importante, los modelos univariados proveen una importante fuente de conocimiento en dos dimensiones: primero, para identificar las características propias de cada variable de tiempo (tamaño de sus innovaciones, persistencia, dinámica hacia el equilibrio) y, segundo, para predecir el comportamiento futuro de una variable económica con confianza y economía de medios.

2.01 Primera Representación Fundamental de un Proceso Estocástico

Gran parte del análisis de modelos univariados de series de tiempo estacionarias lo haremos sobre la base de la función de autocovarianzas (γ_{jt}) del proceso $\{y_t\}$, que se define como:

$$\gamma_{jt} = E[y_t - \mu_t][y_{t-j} - \mu_{t-j}] \quad (2.1)$$

donde μ_t es la media de la variable en el instante t y E es la esperanza matemática condicional en la información disponible hasta el instante $t-1$. Naturalmente, γ_{0t} es la varianza de la variable en el instante t .

La función de autocovarianza es la primera *representación fundamental* de un proceso estocástico. Existen tres de estas representaciones fundamentales; estudiaremos las otras en los capítulos siguientes. El concepto de “fundamental” se refiere a que dicha función identifica y caracteriza el proceso estocástico completamente. Si dos variables tienen la misma función de autocovarianzas, entonces son el mismo proceso estocástico.

2.02 Estacionariedad y Ergodicidad

Se define la estacionariedad como aquella propiedad del proceso generador de los datos (PGD) que da origen a la muestra, tal que:

- *Estacionariedad débil: cuando la media y covarianzas no dependen del tiempo (sólo de j).*
- *Estacionariedad fuerte: cuando la función de probabilidad conjunta de las realizaciones del proceso estocástico no depende del tiempo y sólo depende de " j ".*

Note que la estacionariedad fuerte compromete todos los momentos de la distribución, en tanto que la estacionariedad débil se limita a dos de ellos (media y covarianzas). En realidad, la denominación de estacionariedad débil es simplemente coloquial. Uno podría definir estacionariedad para cualquier momento de la función de distribución; así, el proceso sería estacionario de primer orden si la media es constante, estacionario de segundo orden si la covarianza es constante, etc.

Por otro lado, el que un proceso cumpla con estacionariedad fuerte no implica que cumpla con estacionariedad débil. Por ejemplo, datos generados según la distribución de Cauchy cumplirán con la propiedad de ser estacionarios en el sentido fuerte, pero no con ser estacionarios en el sentido débil. Por otro lado, estacionariedad débil más normalidad implican estacionariedad fuerte, por la simple razón que la distribución normal queda completamente descrita por sus primeros dos momentos.

Cuando una variable aleatoria es débilmente estacionaria, su función de autocovarianzas puede ser escrita, y eventualmente estimada, como:

$$\gamma_j = E[y_t - \mu][y_{t-j} - \mu] \quad (2.2)$$

La importancia que tiene el que una variable sea estacionaria queda graficada en la reducción en el número de parámetros necesarios para describirla, como se desprende al comparar la ecuación (2.1) con (2.2). Si la distribución fuese normal pero no estacionaria, en cada instante de tiempo habría que parametrizar su media, varianza y covarianzas. En una muestra de tamaño n , ello equivale a estimar $n + n(n+1)/2$ parámetros. Si la variable, por otro lado, es estacionaria, basta con parametrizar la media y la función de autocovarianzas, con sólo $n+1$ parámetros.

¿Cómo sabemos si la variable es estacionaria? Esta es una pregunta compleja, a la cual le dedicaremos bastante análisis en el Capítulo 4. Por el momento baste decir que el problema que ella plantea es el siguiente: ¿es posible usar la muestra $\{y_t\}$ y calcular los momentos $E[y]$ y $Cov[y]$? Con la excepción de las series estacionarias, ello sería incorrecto porque los estimadores de esos momentos serían obtenidos en el tiempo y no entre muestras.

Un concepto relacionado, pero no igual, es el de ergodicidad.¹ En series de tiempo se usa para designar aquellos procesos en los cuales una secuencia o muestra de éste es igualmente representativa del proceso completo. Una colección de secuencias forma un conjunto ergódico si la estructura de comportamiento del proceso observado en distintos intervalos de tiempo reproduce la que se observa en cualquier otro instante de tiempo escogido al azar. Una muestra permite, entonces, la estimación de sus momentos.

Así, una variable estacionaria en covarianza es *ergódica en la media* si

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T y_t = E(y_t) \quad (2.3)$$

En la práctica, y como demostraremos formalmente más adelante, lo que se requiere es que las autocovarianzas tiendan a cero cuando $j \rightarrow \infty$.² La intuición es que para que la variable sea estacionaria el pasado remoto ($j \rightarrow \infty$) no debe contener información útil para hacer predicciones sobre las realizaciones contemporáneas o futuras del proceso estocástico. Esto exige:

$$\sum_{j=0}^{\infty} |y_j| < \infty \quad (2.4)$$

Igualmente, una variable estacionaria en covarianza es *ergódica en los segundos momentos* si

$$\lim_{T \rightarrow \infty} \frac{1}{T-j} \sum_{t=j+1}^T (y_t - \mu)(y_{t-j} - \mu) \rightarrow y_j \quad (2.5)$$

Usualmente ergodicidad y estacionariedad son equivalentes. No obstante, considere el proceso $y_t^i = \mu^i + \epsilon_t$, donde μ^i es la media de cada secuencia y se extrae de una $N(0, \eta^2)$. El proceso es estacionario porque $E[y_t^i] = E[\mu^i] = 0$ y $V[y_t^i] = \eta^2 + \sigma_\epsilon^2$, pero no es ergódico porque el promedio de una secuencia dada computado usando (2.3) no es cero sino μ^i .

-
- 1 Ludwig Boltzmann (1844-1906) inventó en 1884 el término ergódico a partir de las palabras griegas *erg* (trabajo) y *ode* (forma) (Über einige Fragen der Kinetische Gastheorie, *Wiener Berichte*, 96: 891-918). Inicialmente el término se usó para designar la probabilidad que algún estado de un sistema mecánico vuelva a suceder y, en particular, a la probabilidad cero que algún estado nunca vuelva a recurrir.
 - 2 "Hay un concepto que es el corruptor y desatinador de los otros. No hablo del Mal cuyo limitado imperio es la ética; hablo del infinito". Jorge Luis Borges, Avatares de la Tortuga, en *Discusión*, 1932.

2.03 Función de Autocorrelación³

En la práctica, la función de autocovarianzas no se utiliza para el análisis de series de tiempo porque depende de unidades de medida que hacen compleja su interpretación. La función de autocorrelación –definida como la función de autocovarianzas normalizada por la varianza– tiene la ventaja de ser adimensional (sin unidades) y estar acotada al intervalo [-1,1].

$$\rho_j = \frac{\gamma_j}{\gamma_0} \quad (2.6)$$

La representación gráfica de las correlaciones para distintos valores de j , llamados rezagos, es llamado *correlograma*. En las secciones siguientes usaremos la función de autocorrelación y su correlograma asociado como la base para identificar el tipo de modelo de serie de tiempo que mejor representa una variable aleatoria.

2.04 PROCESO RUIDO BLANCO⁴

Este proceso, que es la base del análisis de series de tiempo, se describe como aquel que satisface las siguientes propiedades:

$$E[\epsilon_t] = E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0 \quad (2.7)$$

$$V[\epsilon_t] = V(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \sigma^2 \quad (2.8)$$

$$E[\epsilon_t \epsilon_{t-j}] = E(\epsilon_t \epsilon_{t-j} | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0 \quad j \neq 0 \quad (2.9)$$

Las condiciones (2.7) y (2.9) se refieren a la inexistencia de correlación serial o predecibilidad de la variable. Es decir, condicional en la información disponible en el instante t no es posible predecir el valor futuro condicional del proceso $\{\epsilon_t\}$. La condición (2.8) señala que el proceso tiene varianza condicional constante

3 La noción de correlograma analítico fue desarrollada por A. Wold en 1938 (*A Study in the Analysis of Stationary Time Series*). Sin embargo, Yule ya había usado gráficos con correlaciones empíricas en su trabajo de 1926 "Why Do We Sometimes Get Nonsense Correlations between Time-series", *Journal of the Royal Statistical Society*, 89: 1-69, (1926).

4 Originalmente, el término ruido blanco se refería a interferencia radial o eléctrica y fue acuñado por L. D. Carson, W. R. Miles & S. S. Stevens en 1943. Ellos señalan que *Inside the plane ... we hear all frequencies added together at once, producing a noise which is to sound what white light is to light*. La luz blanca es aquella que combina todas las otras luces que existen en distintas frecuencias ("Vision, Hearing and Aeronautical Design," *Scientific Monthly*, 56:446-451, (1943).

(homocedasticidad condicional). Si, además, se tiene que las innovaciones son $\epsilon_t \sim N(0, \sigma^2)$, entonces este proceso se llama Ruido Blanco Gaussiano.⁵

Evidentemente, este es un proceso estacionario en media y covarianza. Como éstas son constantes, no dependen del tiempo.

El correlograma teórico del ruido blanco es bastante simple:

$$\begin{aligned} \rho_j &= 1 & j &= 0 \\ \rho_j &= 0 & j &\neq 0 \end{aligned} \quad (2.10)$$

lo que indica que el proceso no tiene “memoria” del pasado.

Como representación de variables económicas este proceso es limitado porque no da cuenta de una de sus principales características: la persistencia. En el ruido blanco, una realización –positiva o negativa– no tiende a ser seguida por una realización de tamaño o dirección similar. Aunque el proceso parece poco atractivo, está en la base de muchos modelos económicos modernos (considere el tipo de perturbación de un modelo de expectativas racionales) y de todos los modelos econométricos (considere el papel del residuo en el modelo clásico de regresión $y_t = x_t\beta + \epsilon_t$). La falta de memoria característica del ruido blanco es la que lleva a que usualmente se le denomine “innovación”, queriendo indicar que la información contenida en $\{\epsilon_t\}$ no está contenida en ninguna realización previa del proceso y es, por tanto, nueva. Más aún, algunos modelos económicos predicen que el comportamiento de variables claves de la economía debe ser precisamente el de un ruido blanco (p.e., las ganancias extraordinarias de las acciones, según el modelo de Fama, 1970⁶).

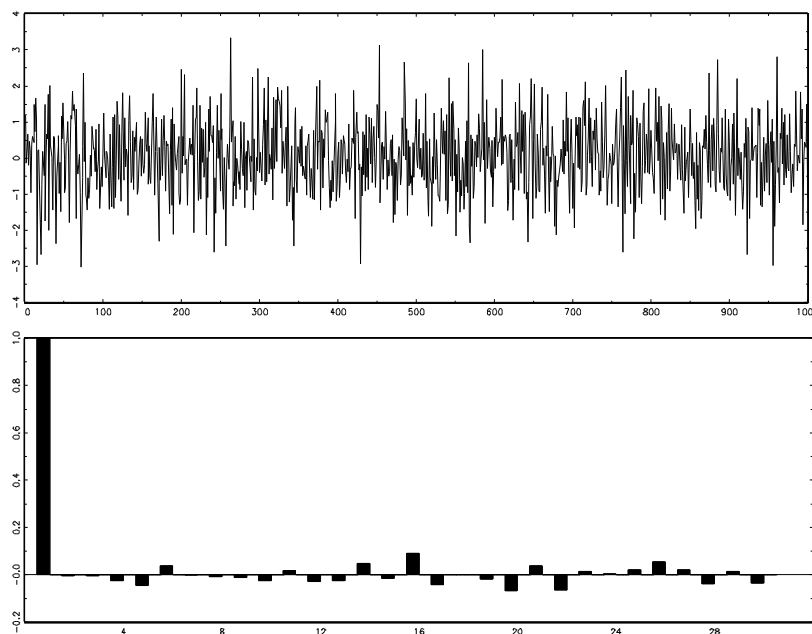
La Figura 1 presenta 1.000 realizaciones de un proceso tipo ruido blanco Gaussiano estándar (es decir, con $\sigma^2=1$) hechas usando Gauss y su correlograma empírico.⁷ Note que, aunque el proceso por construcción es ruido blanco, las correlaciones de orden superior a cero estimadas sobre la base de una muestra no son exactamente cero. Esto se debe a que las correlaciones muestrales están estimadas con información limitada y, por ello, son susceptibles de depender de la conformación de la muestra. Si realiza el experimento con 10,000 observaciones verá como las correlaciones muestrales son más parecidas a las teóricas.

5 Es posible generalizar el proceso ruido blanco. Por ejemplo, puede asumirse que las innovaciones no son independientes. Si solo se cumple la primera condición, entonces ϵ_t corresponde a una “martingale difference sequence”.

6 E. Fama, “Efficient capital market: a review of theory and empirical work”, *Journal of Finance* 25:2 (May), 383-417, (1970).

7 El correlograma empírico depende, naturalmente, de la muestra y su tamaño, por lo que no resulta extraño que tenga correlaciones “medidas” distintas de cero cuando j es grande. El programa de Gauss está en el Apéndice A. Úselo para estudiar cómo el tamaño de muestra y la varianza de las innovaciones afectan el correlograma muestral.

Figura 2.1
Proceso Ruido Blanco
Serie y Correlograma



2.05 Procesos Media Móvil⁸

Este proceso se contruye usando una combinación de ruidos blancos, pero a diferencia de éstos tiene memoria temporal. El proceso media móvil de orden finito q , $MA(q)$ de y_t , se define como:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (2.11)$$

donde μ y el vector θ son parámetros constantes y distintos de infinito. La media del proceso (2.11) es $E[y_t] = \mu$, obviamente, y su varianza es $\gamma_0 = \sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$. La función de autocovarianzas es:

⁸ El primer economista en estudiar seriamente procesos de media móvil fue G.U. Yule en un trabajo de 1909 que utilizaba “promedios instantáneos” (*Journal of the Royal Statistical Society*, 72: 721-730). El término se popularizó con el texto de W. I. King's *Elements of Statistical Methods* (1912). No obstante, el uso de promedios móviles para suavizar datos se había popularizado varias décadas antes de los trabajos de Yule.

$$\gamma_j = \begin{cases} \sigma^2 (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j}) & \forall j=0,1,\dots,q \\ 0 & \forall j>q \end{cases} \quad (2.12)$$

La función de autocorrelación correspondiente es muy similar:

$$\rho_j = \begin{cases} \frac{\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j}}{1 + \theta_1^2 + \dots + \theta_q^2} & \forall j=0,1,\dots,q \\ 0 & \forall j>q \end{cases} \quad (2.13)$$

con $\theta_0=1$. Como se deduce de la ecuación (2.13) los modelos de media móvil tienen q autocorrelaciones distintas de cero. Es decir, la memoria se les acaba en el rezago q y la información del pasado distante –mayor a q – no influye en la predicción de y_t ni en la de su varianza. En la jerga de series de tiempo se dice que el correlograma “muere súbito en q ”.⁹ Nótese que aunque la función de autocorrelación se construye usando los coeficientes del proceso MA, no es posible identificar de manera única un proceso MA(q) a partir del correlograma. Por ejemplo, los siguientes procesos tienen el mismo correlograma:

$$(a) \ y_t = \theta \epsilon_{t-1} + \epsilon_t \quad (b) \ y_t = \frac{1}{\theta} \epsilon_{t-1} + \epsilon_t \quad (2.14)$$

Adicionalmente, resulta obvio que el MA(q) es estacionario porque su función de autocovarianza no depende del tiempo y es finita. Note que dicha función de autocovarianzas es una suma *finita* de parámetros θ_j que en sí son *finitos*. El resultado es más general aún, la suma de los parámetros θ_j es siempre menor que infinito. Esto es lo que vuelve estacionario al proceso estocástico y_t .

¿Qué sucede si el MA es de orden infinito, es decir si $q=\infty$? Esta pregunta es bastante compleja pues nos lleva a pensar en “límites de secuencias”. Por el momento usaremos un resultado que demostraremos más adelante: *Si la secuencia de coeficientes del MA(∞) es sumable absoluta (absolute summable), entonces la secuencia de autocovarianzas es sumable absoluta y el proceso es estacionario.* Consideremos la varianza del proceso (2.11), recordando que las innovaciones son independientes, cuando $q \rightarrow \infty$

$$V[y_t] = \lim_{q \rightarrow \infty} \sum_{j=0}^q \theta_j^2 \sigma^2 = \sigma^2 \lim_{q \rightarrow \infty} \sum_{j=0}^q \theta_j^2 \quad (2.15)$$

⁹ Una disciplina no puede ser enseñada sin su jerga, pero no siempre es necesario enseñar dicha disciplina, Adolfo Bioy Casares, *De Jardines Ajenos*, Bruguera.

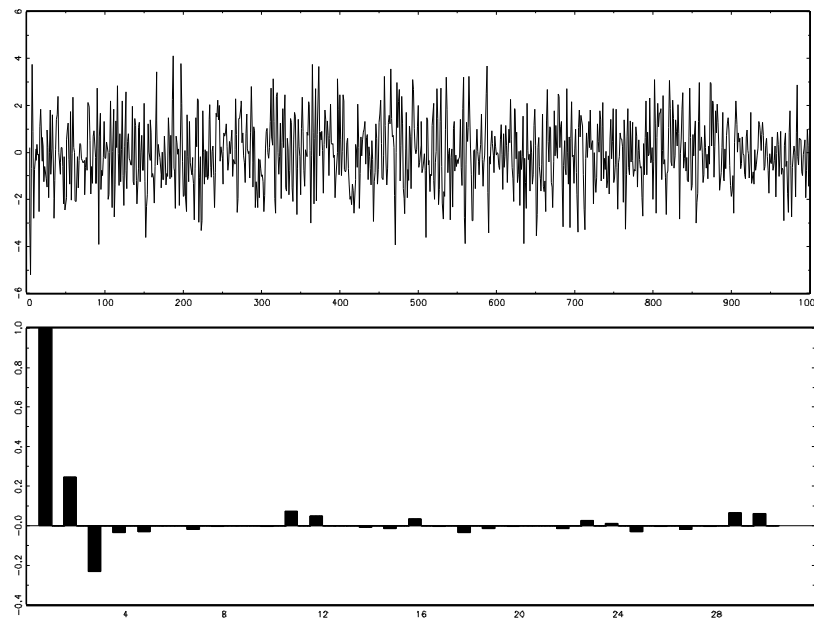
Este límite será finito –es decir, la varianza de la variable será finita– sólo si la secuencia de coeficientes del MA es sumable cuadrática (*square summability*). Las series estacionarias cumplen una condición un poco más estricta llamada sumable en valor absoluto (*absolute summability*)¹⁰:

$$\sum_{j=0}^{\infty} |y_j| < \infty \quad (2.16)$$

Naturalmente, si el MA(q) tiene coeficientes que son sumables en valor absoluto, la variable será ergódica en la media. Si, además, $\{\epsilon_t\}$ es Gaussiano entonces el proceso será ergódico en todos los momentos. Esto lo vamos a usar más adelante para construir funciones de verosimilitud y estimar los parámetros de procesos estocásticos de interés.

En la Figura 2.2 se presentan 1.000 realizaciones del siguiente proceso MA(2) $y_t = 0.2\epsilon_{t-1} - 0.4\epsilon_{t-2} + \epsilon_t$, incluyendo su correlograma empírico. Como se puede ver, la memoria del proceso se acaba en el segundo rezago.

Figura 2.2
Proceso Media Móvil MA(2)
Serie y Correlograma



¹⁰ Ver la demostración en J. Hamilton (1994) *Time Series Analysis*, Princeton University Press, Apéndice 3.A, pp.69-70.

2.06 Procesos Autoregresivos

Los procesos autoregresivos tienen una lógica distinta de los modelos media móvil porque no se construyen a partir de la historia de las innovaciones sino la de la misma variable (más la innovación contemporánea).

Tomemos el más simple de los procesos autoregresivos, un AR(1) del tipo:

$$y_t = \phi y_{t-1} + \epsilon_t \quad (2.17)$$

nótese las diferencias entre este proceso y cualquier MA(q).

Por recursión, podemos escribir este proceso como:

$$y_t = \epsilon_t + \phi [\epsilon_{t-1} + \phi y_{t-2}] = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 y_{t-2} \quad (2.18)$$

continuando la recursión ad-infinitum obtenemos:

$$y_t = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \phi^3 \epsilon_{t-3} \dots \quad (2.19)$$

es decir, es un MA(∞). Para que el proceso sea estacionario la secuencia tiene que ser sumable en valor absoluto, lo que a su vez requiere que $|\phi| < 1$.

La media del proceso (2.17) es cero, pero si el proceso fuese $y_t = \mu + \phi y_{t-1} + \epsilon_t$, su media no-condicional sería:

$$E[y_t] = \frac{\mu}{1 - \phi} \quad (2.20)$$

la implicación de este resultado es que si en cualquier modelo autoregresivo incluimos una constante, ésta dará cuenta de la media de la serie.

La varianza del proceso AR(1) es $\gamma_0 = \frac{\sigma^2}{1 - \phi^2}$ y su función de autocovarianzas es:

$$\gamma_j = \phi^j \frac{\sigma^2}{1 - \phi^2} \quad (2.21)$$

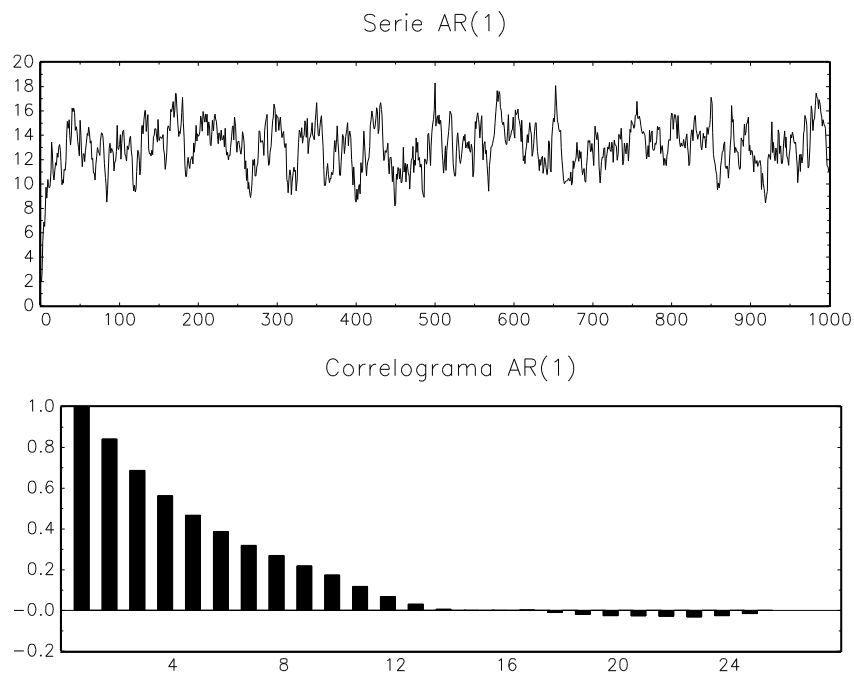
lo que da origen a una función de autocorrelación decreciente suave, del tipo $\rho_j = \phi^j$.

Hay dos elementos que se deben hacer notar:

1. Si $\phi < 0$, el correlograma exhibirá ciclos, i.e., valores positivos alternados con negativos. La magnitud de los ciclos, sin embargo, es decreciente.
2. La propiedad de estacionariedad de la variable implica que $Cov[y_t, y_{t-j}] = Cov[y_t, y_{t+j}]$. Mecánicamente esto sucede porque podemos hacer la recursión hacia atrás o hacia adelante. Conceptualmente, la razón es que siendo estacionaria la variable lo único que determina una covarianza es el largo del rezago j , es decir, la distancia entre y_t e y_{t-j} y no " t ".

En la Figura 2.3 se presentan 1.000 realizaciones del proceso $y_t = 0.85 y_{t-1} + \epsilon_t$ donde las innovaciones son ruido blanco Gaussiano estándar y $y_0 = 0$. Se incluye, además, el correlograma empírico en el que se puede ver claramente el patrón suavemente declinante de las autocorrelaciones.

Figura 2.3
Proceso Autoregresivo AR(1)
Serie y Correlograma



Nótese que, a diferencia de los procesos $MA(q)$, los procesos autoregresivos dependen del punto de partida de los datos. Frecuentemente en los estudios empíricos con datos simulados se eliminan las primeras observaciones de la muestra (p.e., 50) para evitar la dependencia del punto de partida.

El manejo de modelos de serie de tiempo es más fácil si se usa el operador de rezagos, L , que permite operar con la estructura dinámica de una variable cómodamente de manera algebraica.¹¹ Se define el operador como $Ly_t = y_{t-1}$. Así, $L^2 y_t = y_{t-2}$ y $L^{-2} y_t = y_{t+2}$.

El modelo AR(1) se puede re-escribir como:

$$(1 - \phi L) y_t = \epsilon_t \quad (2.22)$$

de donde se desprenden las ecuaciones (2.19) y (2.20) directamente. Note que si por simplicidad se reduce $\mu = 0$, entonces de la ecuación (2.22) se deduce:

$$(1 - \phi L) y_t = \epsilon_t \Rightarrow y_t = \frac{\epsilon_t}{(1 - \phi L)} = \epsilon_t (1 + \phi L + \phi^2 L^2 + \dots) \quad (2.23)$$

es decir, obtenemos nuevamente el resultado anterior –que todo AR(1) tiene una representación MA(∞)– pero además queda en evidencia que, para que el proceso sea invertible, se requiere imponer restricciones sobre ϕ . Note que la ecuación (2.23) se obtiene usando una expansión de Taylor.

Si extendemos el modelo anterior a un caso general que llamaremos AR(p) tenemos:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (2.24)$$

La media y función de autocovarianzas del proceso AR(p) son bastante más difíciles de obtener. Vamos a usar un modelo autoregresivo de segundo orden y el operador de rezagos para estudiar los procesos generales. Re-escribiendo (2.24) como un AR(2) tenemos:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad (2.25)$$

usamos el operador de rezagos

$$y_t (1 - \phi_1 L - \phi_2 L^2) = \epsilon_t \quad (2.26)$$

si factorizamos el término en L como un cuadrado de binomio tenemos:

$$(1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L)(1 - \lambda_2 L) \quad (2.27)$$

Usando (2.26) y (2.27) tenemos las siguientes restricciones

11 Ver el Apéndice B para una descripción más completa de este operador y sus propiedades.

$$\begin{aligned}\lambda_1 \lambda_2 &= -\phi_2 \\ \lambda_1 + \lambda_2 &= \phi_1\end{aligned}\tag{2.28}$$

es decir

$$y_t = (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1} \epsilon_t = \left[\sum_{j=0}^{\infty} \lambda_1^j L^j \right] \left[\sum_{j=0}^{\infty} \lambda_2^j L^j \right] \epsilon_t\tag{2.29}$$

El último término puede ser escrito como

$$\left[\sum_{j=0}^{\infty} \lambda_1^j L^j \right] \left[\sum_{j=0}^{\infty} \lambda_2^j L^j \right] = \sum_{j=0}^{\infty} \left[\sum_{k=0}^j \lambda_1^k \lambda_2^{j-k} \right] L^j\tag{2.30}$$

es decir, nuevamente encontramos que el $AR(p)$ puede ser escrito como un proceso MA de orden infinito.

Resulta evidente que el operador de rezagos puede ser usado también en los procesos media móvil directamente. En particular,

$$y_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \epsilon_t \Rightarrow (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)^{-1} y_t = \epsilon_t\tag{2.31}$$

es decir, todo $MA(q)$ puede ser representado como un $AR(\infty)$, provisto que el proceso $MA(q)$ sea invertible. Este resultado, y su contraparte para procesos $AR(p)$, puede ser interpretado como que “en la práctica todo proceso $MA(AR)$ puede ser aproximado arbitrariamente cerca por un proceso $AR(MA)$ con un número suficientemente largo de rezagos”.

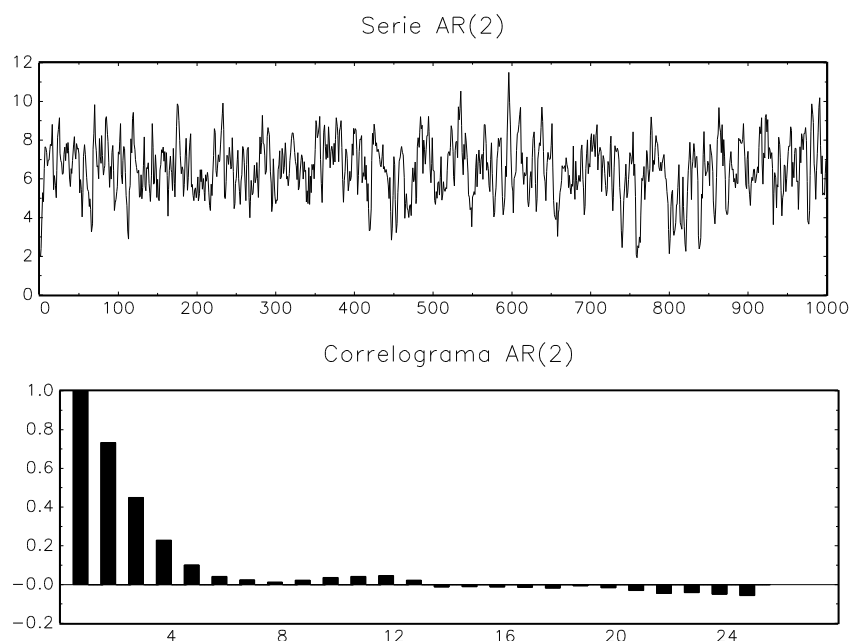
Para el $AR(p)$ es directo –aunque tedioso– demostrar que la función de autocorrelación obedece al siguiente sistema de ecuaciones (llamado ecuaciones de Yule-Walker).

$$\begin{aligned}\rho_0 &= \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_p \rho_p + \sigma^2 \\ \rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \vdots &= \vdots + \vdots + \ddots + \vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \rho_0\end{aligned}\tag{2.32}$$

de donde se desprende que, conociendo las autocorrelaciones, se puede hacer un sistema de ecuaciones para obtener los parámetros del modelo. Ello resulta usualmente demasiado engorroso y más adelante se propone un método más eficiente y asintóticamente equivalente.

La Figura 2.4 se presentan 1.000 realizaciones del proceso $y_t = 0.85y_{t-1} - 0.15y_{t-2} + \epsilon_t$ donde las innovaciones son ruido blanco Gaussiano y $y_0 = 0$. Se incluye, además, el correlograma empírico en el que se puede ver claramente que, al igual que en el caso del AR(1), las autocorrelaciones tienen un patrón suavemente declinante. De hecho, esa es la característica de todo modelo autoregresivo.

Figura 2.4
Proceso Autoregresivo AR(2)
Serie y Correlograma



Por ello, la observación del correlograma no es suficiente para identificar el orden del modelo autoregresivo, por que no sabemos si p es 1, 2, 3 ... Vamos a tener que usar una función adicional para identificar el orden del AR. Esta función es llamada función de autocorrelación parcial. La identificación del orden del AR(p) se hará sobre la base de ambos correlogramas.

La intuición de la función de autocorrelación parcial es simple. Suponga que el verdadero proceso es AR(p). Pero especificamos y parametrizamos los datos usando un modelo AR($p-1$). Sin demasiada formalidad escribimos:

$$y_t = AR(p-1)y_{t-1} + \eta_t \quad (2.33)$$

¿Como sería η_t en este caso? Obviamente, no podría ser ruido blanco porque contiene tanto las verdaderas innovaciones ϵ_t como el término $\phi_p y_{t-p}$ que no hemos

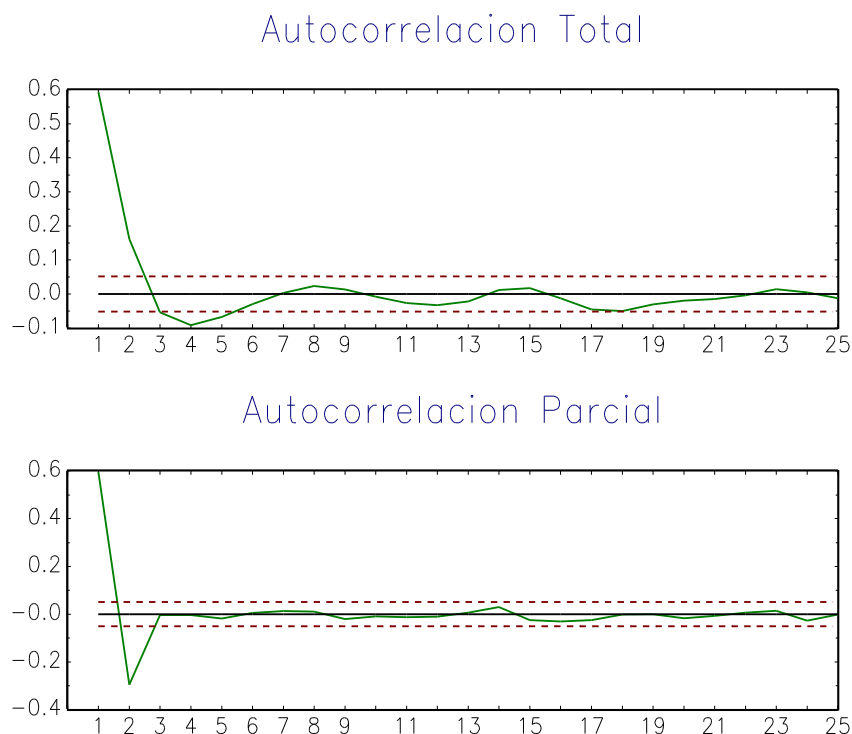
incluido en el modelo. Por lo tanto, η_t va a estar correlacionado con y_{t-p} . Vamos a usar exactamente esta observación para construir la función de correlación parcial y para determinar el orden p del modelo autoregresivo. Estudiemos la siguiente covarianza:

$$\gamma_j^* = \text{cov}[y_t - E[y_t | y_{t-1}, y_{t-2}, \dots, y_{t-j+1}], y_{t-j}] \quad (2.34)$$

donde E es la mejor predicción lineal de y_t hecha con sus propios j rezagos. El primer término de la covarianza es el error de predicción de y_t que se comete al usar un modelo lineal basado en j rezagos. Cuando no hay covarianza, significa que el error de predicción no contiene información útil para predecir y_t porque no quedan términos del tipo y_{t-p} omitidos del modelo. Así, cuando γ_j^* es cero, ese es el orden del modelo AR. Naturalmente, normalizamos la covarianza parcial por la varianza para evitar problemas de unidades de medida.

Un ejemplo del uso de ambos correlogramas se muestra en la Figura 2.5. El correlograma total presenta un patrón de caída suave aunque cíclico. Ello identifica un modelo tipo $\text{AR}(p)$. El correlograma parcial muere súbitamente en la segunda correlación. Ello identifica el modelo como un $\text{AR}(2)$.

Figura 2.5
Correlograma Total y Parcial AR(2)



Ya que hemos derivado la función de autocorrelación parcial, conviene preguntarse por la forma que toma ésta en el caso de un proceso de media móvil. Recordando que todo proceso media móvil puede ser escrito como un $AR(\infty)$ se deduce que el correlograma parcial tendrá infinitos términos distintos de cero. Además, como debe cumplirse la sumabilidad absoluta, el correlograma será decreciente.

2.07 Procesos Autoregresivos y de Media Móvil

Nada le impide a un proceso estocástico tener ambos componentes (AR y MA), reflejando distintas fuerzas que determinan la evolución de la variable, en cuyo caso los llamamos proceso autoregresivo y de media móvil o $ARMA(p,q)$. Naturalmente, es posible interpretar un modelo AR o MA como un caso particular del modelo ARMA. En términos genéricos el modelo $ARMA(p,q)$ es:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (2.35)$$

Calcular la funciones de correlación total y parcial es directo pero tedioso. Apelamos a nuestros conocimientos de modelos AR y MA para intuir cómo son dichas funciones:

- Respecto del correlograma total, éste tiene ambos componentes hasta el rezago q cuando la parte MA deja de operar y sólo queda la parte AR que decae suavemente con o sin oscilaciones.
- Respecto del correlograma parcial, éste tiene ambos componentes hasta el rezago p cuando la parte AR deja de operar y sólo queda la parte MA que decae suavemente con o sin oscilaciones.

La combinación de ambos correlogramas nos permite determinar la naturaleza del proceso ARMA y el largo de rezagos de cada componente.

2.08 Identificación de Procesos Estocásticos Estacionarios

Las reglas de pulgar para identificar procesos estocásticos se presentan en el Cuadro 2.1:

Cuadro 2.1
Reglas de identificación de procesos estacionarios

Proceso	Autocorrelación Total	Autocorrelación Parcial
Ruido Blanco	Muere súbito en 0	Muere súbito en 0
MA(q)	Muere súbito en q	Muere suave (con o sin ciclos)
AR(p)	Muere suave (con o sin ciclos)	Muere súbito en p
ARMA(p, q)	Muere suave desde q (con o sin ciclos)	Muere suave desde p (con o sin ciclos)

2.09 Funciones de Correlación Admisibles y el Teorema de Wold

Antes de avanzar hacia la parte empírica de este tipo de modelos univariados (llamada usualmente técnica de Box-Jenkins), debemos preguntarnos si es legítimo modelar una variable usando esta metodología. Hay tres preguntas que debemos responder:

- (1) ¿qué nos permite asegurar que representar una variable usando modelos ARMA(p, q) es adecuado?,
- (2) ¿qué nos hace pensar que al estudiar una función de autocorrelación podremos encontrar el “verdadero modelo”?
- (3) ¿cómo sabemos que una secuencia de números corresponde efectivamente a una realización de la función de autocorrelación de algún proceso estocástico de interés?

Funciones de autocorrelación admisibles

No toda secuencia de números $\{1, \rho_1, \rho_2, \rho_3, \dots\}$ corresponde a una función de autocorrelación de un proceso estocástico. Obviamente debemos restringir los valores de ρ para que cumplan $|\rho_i| \leq 1$. Pero ello no es necesariamente suficiente. Además se debe exigir que las varianzas ρ_0 sean positivas. Consideremos un caso particular pero ilustrativo:

$$V[\alpha_0 y_t + \alpha_1 y_{t-1}] = y_0 [\alpha_0 \quad \alpha_1] \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \geq 0 \quad (2.36)$$

Dado que y_0 es positiva, entonces $\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}$ tiene que ser positiva semi-definida. Así, la única restricción necesaria es que $|\rho_1| \leq 1$ (nótese que no hay restricciones sobre los coeficientes).

Si el modelo tuviese un segundo rezago, su varianza quedaría descrita de manera análoga a la de la ecuación (2.36), excepto que la matriz de correlaciones sería ahora $\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}$. Para que dicha matriz sea definida positiva, se debe cumplir que el determinante y los subdeterminantes sean positivos. Esto último requiere que $|\rho_1| \leq 1$ y $|\rho_2| \leq 1$, en tanto que el primero exige

$$1 + 2\rho_1^2\rho_2 - 2\rho_1^2 - 2\rho_2^2 \geq 0 \quad \Rightarrow \quad (\rho_2 - (2\rho_1^2 - 1))(\rho_2 - 1) \leq 0 \quad (2.37)$$

pero como $|\rho_2| \leq 1$, entonces se debe cumplir que:

$$\rho_2 - (2\rho_1^2 - 1) \geq 0 \quad \Rightarrow \quad \rho_2 \geq (2\rho_1^2 - 1) \quad \Rightarrow \quad -1 \leq \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \leq 1 \quad (2.38)$$

esto impone restricciones a los valores de ρ_2 dado los valores de ρ_1 . Por ejemplo si este último es 0.9, entonces $0.62 \leq \rho_2 \leq 1$.

Teorema de Wold¹²

Aún si contamos con una secuencia de valores que pueda ser congruente con una función de autocorrelación y, por necesidad, con una función de autocovarianzas de algún proceso estocástico, ¿qué nos permite asegurar que es posible representar variables económicas con modelos ARMA? El siguiente teorema es la base del análisis de series de tiempo:

“Cualquier proceso con media cero y estacionario en covarianza $\{y_t\}$ puede ser representado de manera única por un componente lineal determinístico y un componente lineal no determinístico”.

$$y_t = P^*[y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}] + \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i} \quad (2.39)$$

donde P^* es un predictor lineal y se cumple:

- $\epsilon_t = y_t - P^*[y_t | y_{t-1}, y_{t-2}, \dots]$ es un error de predicción.
- ϵ_t es ruido blanco, es decir tiene media cero, varianza constante, no covaría con $P^*[y_t | y_{t-1}, y_{t-2}, \dots]$ ni con los rezagos de y_t .
- $\theta_0 = 1$ y se cumple $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ (sumabilidad cuadrática)
- Todas las raíces de $\theta(L)$ están fuera del círculo unitario.

¿Cuál es el mejor predictor lineal de y_t ? Podemos aproximar este problema usando un modelo AR con p rezagos.

$$y_t = \sum_{i=1}^p \gamma_i y_{t-i} + \sum_{i=0}^{\infty} \pi_i \epsilon_{t-i} \quad (2.40)$$

Este es un modelo ARMA. El primer término es un AR(p) pero el segundo es una media móvil de orden infinito. Obviamente, éste no puede ser estimado con muestras aleatorias de tamaño finito. Usamos como solución de compromiso buscar el mejor ajuste con rezagos finitos.

$$y_t = \sum_{i=1}^p \gamma_i y_{t-i} + \sum_{i=0}^q \pi_i \epsilon_{t-i} \quad (2.41)$$

¹² Ver la prueba del teorema en Thomas Sargent, *Dynamic Macroeconomics*, 1987. La demostración requiere conocer la mecánica de los espacios de Hilbert, pero es relativamente intuitivo.

Algunas observaciones importantes respecto del Teorema de Wold.

1. No se requiere que los errores ϵ_t se distribuyan normal o sean i.i.d.
2. No se requiere que $E(y_t, \epsilon_t) = 0$ sino que $P^*(y_t, \epsilon_t) = 0$. La diferencia radica en que $P^*(\cdot)$ es una proyección lineal, en tanto que $E(\cdot)$ considera todas las posibles formas de proyección de y_t , tanto lineales como no lineales.
3. Los errores ϵ_t no tienen por qué ser las verdaderas innovaciones del proceso.
4. La representación de la variable como un $MA(\infty)$ es *una* posible representación del PGD que captura el segundo momento: no es la única o la mejor representación de la variable. Otras representaciones pueden capturar, además, momentos de orden superior.
5. La representación de la variable según la ecuación (2.39) es única entre modelos lineales en que los errores de proyección son lineales.

2.10 Metodología de Box-Jenkins

La metodología de Box y Jenkins¹³ para modelar series de tiempo apareció como una alternativa al fracaso de los modelos econométricos de gran escala (basados en gran número de ecuaciones simultáneas tradicionales, ver Klein y Evans¹⁴) durante los años 1970-1975. Esta metodología considera seis pasos, que discutimos a continuación:

1. Dado que se aplica a variables estacionarias, primero debe hacerse estacionaria la variable aleatoria. Para ello usualmente se aplica algún filtro.
2. Se retiran, además, los componentes determinísticos (estacionalidad, constante).
3. Se computan los correlogramas muestrales y se hace una *identificación preliminar* del modelo, es decir se determina el tipo y orden del modelo subyacente (AR, MA, ARMA).
4. Se estiman los parámetros del modelo propuesto.
5. Se hacen tests de especificación y se evalúa si la identificación preliminar es adecuada. Si no lo es, se modifica la especificación.
6. Se hacen predicciones del modelo estimado dentro y fuera de la muestra.

¹³ Box, G. E. P., and Jenkins, G. *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

¹⁴ L. Klein and G. Evans, *The Wharton Econometric Forecasting Model*, Economic Research Unit, Department of Economics, Wharton School of Finance and Economics, University of Pennsylvania, 1967.

2.11 ¿Cómo Volver Estacionaria una Variable?

La metodología Box-Jenkins enfrenta el problema de remover la tendencia de las variables mediante dos técnicas básicas.¹⁵ Si la variable es estacionaria basta con hacer una regresión contra una tendencia lineal y ocupar los residuos de dicha estimación. Considere la ecuación (2.42)

$$y_t = \alpha + \beta t + \psi(L)\epsilon_t \quad (2.42)$$

donde α y β son dos parámetros y $\psi(L)$ es la representación media móvil de un modelo ARMA(p, q) para las innovaciones ϵ_t . La remoción de tendencia correspondería a obtener $y_t^f = y_t - \hat{\alpha} - \hat{\beta}t$. Esto puede ser hecho usando técnicas de mínimos cuadrados para filtrar la serie.

Si la variable no es estacionaria –lo que se verifica cuando el correlograma muestral no cae o lo hace demasiado lento– se debe tomar la primera diferencia de la variable. Es decir,

$$(1-L)y_t = \delta + \psi^*(L)\epsilon_t \quad (2.43)$$

Exactamente cuándo se debe usar cada método es lo que nos preocupará más adelante cuando se estudie el tema de raíces unitarias (ver Capítulo 4). En esta sección supondremos que el econométrista sabe *a-priori* cuándo usar primeras diferencias o diferencias con respecto a una tendencia lineal.

También es necesario remover los componentes determinísticos, tales como los efectos estacionales, usando algún tipo de filtro.¹⁶ Soto (2002) demuestra que los métodos más populares de ajuste estacional (p.e., ARIMA X-11) no pueden ser considerados una simplificación inocente de los datos sin pérdidas sustanciales de información valiosa. El tratamiento estándar para remover la estacionalidad supone que la variable es estacionaria. Si este supuesto es correcto, Sims (1974) demuestra que la remoción de componentes estacionales usando un proceso de filtrado común para todas las variables mantiene las condiciones de congruencia del modelo original. Ghysels y Perron (1993)¹⁷, por otro lado, demuestran que la conclusión anterior no puede ser extendida a modelos dinámicos en los cuales se incluyen variables endógenas rezagadas. Soto (2002) muestra que si el supuesto de estacionariedad de la estacionalidad no se cumple, la remoción de componentes estacionales con métodos tipo ARIMA X-11 es inadecuada porque induce correlación serial aún si ésta no existe en los datos originales.

¹⁵ Ver Apéndice C.

¹⁶ "Ajuste Estacional e Integración en Variables Macroeconómicas", *Cuadernos de Economía*, 39:135-155, 2002.

¹⁷ Sims, C. A. (1974). "Seasonality in regression", *Journal of the American Statistical Association*, 69: 618-627. Ghysels, E. and P. Perron (1993). "The effect of seasonal adjustment filters on tests for a unit root", *Journal of Econometrics*, 55(1-2):57-98.

2.12 Identificación del Tipo de Modelo ARMA

En las secciones anteriores hemos identificado los distintos tipos de modelos (AR, MA, ARMA) mediante el uso de correlogramas teóricos. Si tenemos una muestra finita, sin embargo, tendremos *estimadores muestrales* de las correlaciones. Es decir, tendremos que hacer inferencias basados en la evidencia muestral, para lo cual es necesario determinar la distribución del correlograma muestral.

Si los datos son de un MA(q) con innovaciones Gaussianas, entonces a partir del rezago q las correlaciones muestrales debieran ser cero en valor esperado, pero como éstas son una variable estocástica podrían ser numéricamente distintas de cero. Como dichas estimaciones tienen varianza ello nos permite hacer intervalos de confianza para la hipótesis nula $H_0: \rho_j = 0$. La varianza de las correlaciones es:

$$Var(\hat{\rho}_j) \approx \frac{1}{T} \left[1 + 2 \sum_{i=1}^q \rho_i^2 \right] \quad j = q+1, q+2, \dots \quad (2.44)$$

Por ello, si el MA tiene innovaciones Gaussianas el intervalo de confianza para un test de la hipótesis nula $\rho_j = 0$ de tamaño 95% es $\pm 2/\sqrt{T}$.¹⁸

Si los datos provienen de un AR(p), y considerando que la función de autocorrelación parcial es:

$$y_t = \hat{\mu} + \hat{\alpha}_1^{(m)} y_{t-1} + \hat{\alpha}_2^{(m)} y_{t-2} + \dots + \hat{\alpha}_m^{(m)} y_{t-m} + \epsilon_t^* \quad (2.45)$$

entonces, el último coeficiente α_m debiese ser 0 en valor esperado si $m > p$. Su varianza es:

$$Var(\hat{\alpha}_m^{(m)}) \approx \frac{1}{T} \quad (2.46)$$

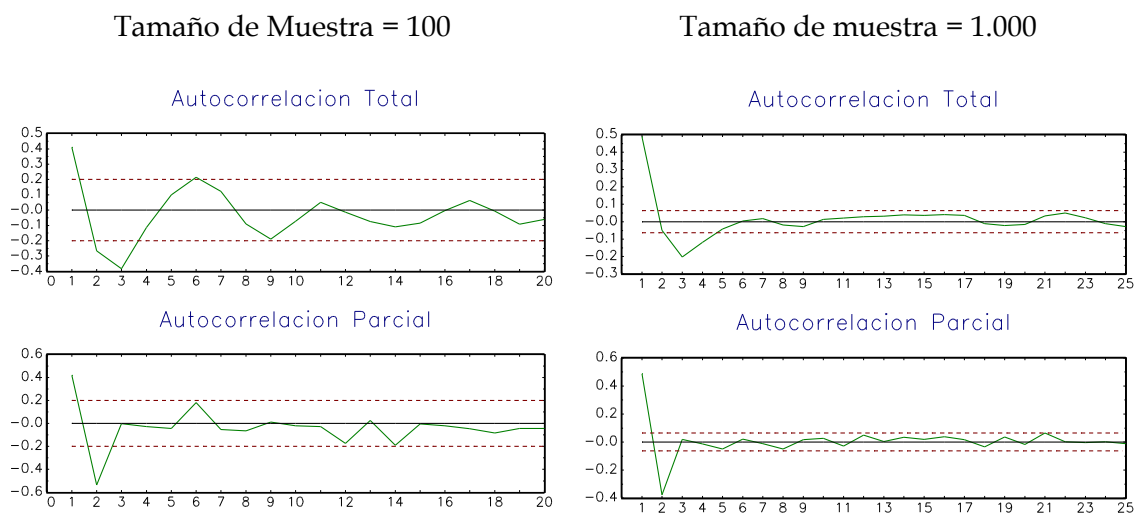
por lo que el intervalo de confianza para el test de la hipótesis nula $\alpha_m = 0$ de tamaño 95% es $\pm 2/\sqrt{T}$. Esto soluciona el problema de inferencia.

La figura 2.6 presenta gráficamente dos simulaciones hechas para un AR(2) con parámetros comunes, pero cambiando el tamaño de muestra. El panel de la izquierda corresponde a una muestra de 100 observaciones, el de la derecha corresponde a 1.000 observaciones. Hay dos elementos destacables. Primero, la reducción de la incertidumbre en el estimador de la correlación. Segundo, la mayor precisión de la estimación de las correlaciones. No obstante, ambos gráficos sugieren que el modelo es un AR(2).

¹⁸ Note que ésta no es la varianza asintótica estimada, la que requiere un estimador de σ^2 .

Algunos autores señalan que se deben usar los correlogramas muestrales “sólo para tener una idea” del tipo de proceso y su largo, pero que la clave consiste en asegurar que las innovaciones sean ruido blanco usando tests de diagnóstico adecuados.

Figura 2.6
Efecto del Tamaño de Muestra en Identificación en un AR(2)



Debe reconocerse que la identificación por correlogramas es, en la práctica, algo difícil porque los correlogramas empíricos son menos nítidos que los correlogramas teóricos, sobre todo en muestras pequeñas. Por ello, la siguiente etapa del método de Box y Jenkins (estimación de modelos ARMA) es crucial. Estimar el modelo ARMA es relativamente directo por máxima verosimilitud, pero lo vamos a ver paso por paso.

2.13 Estimación de Modelos Autoregresivos

El caso del AR(1)

Sea $y_t = c + \phi y_{t-1} + \epsilon_t$. Queremos construir la función de verosimilitud de la muestra que incluye $\{y_1, y_2, \dots, y_n\}$. Como el modelo de y_t depende de la observación anterior, tendremos problemas computando para $t=1$ porque no hay observación en $t=0$. Por eso, necesitamos computar la esperanza y la varianza no condicionales para la primera observación (y_1):

$$E[y_1] = \frac{c}{1-\phi} \quad V[y_1] = \frac{\sigma^2}{1-\phi^2} \quad (2.47)$$

La función de densidad no-condicional de y_1 , si la innovación es Gaussiana, es:

$$f(y_1; \theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2/(1-\phi^2)}} e^{\left[\frac{-(y_1 - c/(1-\phi))^2}{2\sigma^2/(1-\phi^2)} \right]} \quad (2.48)$$

La función de densidad de y_2 condicional en haber observado y_1 es:

$$f(y_2|y_1; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma^2} \right]} \quad (2.49)$$

Note que para la densidad de y_2 condicional en y_1 , el valor observado de y_1 es una constante no una variable aleatoria. Entonces, $E[y_2] = c + \phi y_1$ y la única varianza es de la innovación, por lo que y_2 se distribuye $N(c + \phi y_1, \sigma^2)$. En general, las densidades condicionales para los datos desde 2 a n se pueden escribir como:

$$f(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \theta) = f(y_t|y_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right]} \quad (2.50)$$

Dado que los y_i son independientes, la densidad conjunta de la muestra es el producto de las marginales. Para la primera y segunda observación ésta es:

$$f(y_2, y_1; \theta) = f(y_2|y_1; \theta) f(y_1; \theta) \quad (2.51)$$

Podemos hacer esto para todas las observaciones de la muestra y obtener la función de verosimilitud exacta:

$$\mathcal{L}(\theta) = f(y_1; \theta) \cdot \prod_{t=2}^T f(y_t|y_{t-1}; \theta) \quad (2.52)$$

Tomando logaritmos:

$$\log \mathcal{L}(\theta) = \log f(y_1; \theta) + \sum_{t=2}^T \log f(y_t|y_{t-1}; \theta) \quad (2.53)$$

o, lo que es equivalente:

$$\begin{aligned} \log \mathcal{L}(\theta) = & \frac{-1}{2} \log 2\pi - \frac{1}{2} \log \frac{\sigma^2}{1-\phi^2} - \frac{(y_1 - \phi/(1-\phi))^2}{2\sigma^2(1-\phi^2)} \\ & - \frac{T-1}{2} \log 2\pi - \frac{T-1}{2} \log \sigma^2 - \sum_{t=2}^T \frac{(y_t - \phi - \phi y_{t-1})^2}{2\sigma^2} \end{aligned} \quad (2.54)$$

Esta función se optimiza por cálculo numérico. Hay que reconocer, no obstante, que la estimación puede ser engorrosa, en particular porque los métodos de optimización directa de la función de verosimilitud son sensibles a los puntos de partida, al algoritmo de estimación, e incluso a la conformación de la muestra.

Una alternativa práctica consiste en optimizar la densidad condicional en y_1

$$f(y_t, y_{t-1}, y_{t-2}, \dots | y_1; \theta) = \prod_{t=2}^T f(y_t | y_{t-1}; \theta) \quad (2.55)$$

entonces

$$\log \mathcal{L}(\theta) = \frac{-T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \frac{\sum_{t=2}^T (y_t - \phi - \phi y_{t-1})^2}{2\sigma^2} \quad (2.56)$$

La ecuación (2.56) tiene tres componentes, de los cuales los dos primeros son constantes. Por ello, maximizar $\log \mathcal{L}(\theta)$ es equivalente a minimizar el último término. Pero esto último es equivalente a hacer una regresión de mínimos cuadrados ordinarios. Así, la maximización de la densidad condicional en y_1 de la muestra produce estimadores de los parámetros que son equivalentes a los obtenidos por una regresión mínimos cuadrados ordinarios de y_t en sus rezagos. Este resultado se extiende directamente a un proceso AR(p).

Como es sabido desde los años 1950, los estimadores de mínimos cuadrados de procesos autoregresivos son sesgados en muestra pequeña. El sesgo es creciente en el número de parámetros que se estimen, pero desaparece asintóticamente. Tanizaki et al. (2005) estudian el tamaño del sesgo y demuestran que es posible hacer correcciones que mejoren la calidad de la estimación. Si bien esta corrección –obtenida por simulación– es útil, es importante recordar que en modelos de series de tiempo no estamos interesados en los parámetros *per-se* sino en la capacidad del modelo completo para dar cuenta de manera razonable de las fluctuaciones de la variable de interés.¹⁹

¹⁹ Tanizaki, H; S. Hamori and S. Matsubayashi “On least-squares bias in the AR(p) models”, *Statistical Papers*, 47:109-124 (2005).

2.14 Estimación de Modelos Media Móvil

Estimación de un Modelo MA(1)

Sea $y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}$. Como vimos en el caso del modelo autorregresivo, maximizar la densidad condicional es más simple que hacerlo sobre la densidad exacta. En el caso del MA(1) condicionamos sobre ϵ_{t-1} .

$$y_t | \epsilon_{t-1} \rightarrow N[(\mu + \theta \epsilon_{t-1}), \sigma^2] \quad (2.57)$$

Lamentablemente, no es posible estimar este proceso por mínimos cuadrados ordinarios. De hecho, el problema es más complejo de lo que parece: para estimar θ se necesitan los residuos pero para determinar los residuos se necesita estimar θ .

Por ello necesitamos un algoritmo –que en este caso será de tipo iterativo–:

- Si sabemos que $\epsilon_0 = 0$, entonces se cumple $y_1 | \epsilon_0 = 0 \rightarrow N[\mu, \sigma^2]$.
- Si conocemos y_1 , entonces podemos conocer $\epsilon_1 = y_1 - \mu$.
- Si conocemos y_2 y ϵ_1 , obtenemos $\epsilon_2 = y_2 - \mu - \theta \epsilon_1$.
- Si conocemos y_3 y ϵ_2 , obtenemos $\epsilon_3 = y_3 - \mu - \theta \epsilon_2$, etc.

De esta forma, tenemos toda la secuencia de ϵ_t condicional en $\epsilon_0 = 0$.

Podemos formar entonces la densidad conjunta condicional y la función de verosimilitud condicional:

$$\log \mathcal{L}(\theta) = -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad (2.58)$$

El problema de la ecuación (2.58) es que no podemos estimarla de manera convencional. Pero sí podemos parametrizarla usando el siguiente algoritmo. La solución es darle un conjunto de parámetros arbitrarios θ_0 , computar las innovaciones ϵ_t , y optimizar $\log \mathcal{L}(\theta)$ determinando un nuevo conjunto de parámetros θ_1 . Con ese nuevo conjunto de parámetros se re-computa $\log \mathcal{L}(\theta)$ y se itera hasta que la función de verosimilitud converja. Si $|\theta|$ es pequeño, el efecto de suponer $\epsilon_0 = 0$ es poco importante y desaparece luego. Si $|\theta| > 1$, el algoritmo diverge.

Este mismo algoritmo se utiliza para estimar un MA(q), sólo que deberemos darle q parámetros iniciales.

2.15 Estimación de un Modelo ARMA

La aproximación a la función de verosimilitud en el caso de los AR fue condicionada a los y_p iniciales. La aproximación a la función de verosimilitud en el caso de los MA fue condicionada a los ϵ_q iniciales. La aproximación a la función de verosimilitud de los ARMA, naturalmente, condiciona respecto de los y_p y ϵ_q iniciales.

$$\log \mathcal{L}(\theta) = \frac{-T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad (2.59)$$

$$\epsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (2.60)$$

2.16 Predicciones con Modelos ARMA²⁰

La predicción de una variable económica envuelve dos dimensiones: el valor que se espera que la variable tome en el período $t+j$ y, debido a la aleatoriedad de los modelos econométricos, la imprecisión o varianza de la predicción. Es decir, queremos:

$$\begin{aligned} E[y_{t+j}] &= E[y_{t+j} \mid y_{t+j-1}, y_{t+j-2}, \dots, \epsilon_t, \epsilon_{t-1}, \dots] \\ V[y_{t+j}] &= V[y_{t+j} \mid y_{t+j-1}, y_{t+j-2}, \dots, \epsilon_t, \epsilon_{t-1}, \dots] \end{aligned} \quad (2.61)$$

Predicción de un modelo AR(1)

Consideremos el modelo $y_t = \phi y_{t-1} + \epsilon_t$, entonces

$$\begin{aligned} E_t[y_{t+1}] &= E_t[\phi y_t + \epsilon_{t+1}] &= \phi y_t \\ E_t[y_{t+2}] &= E_t[\phi y_{t+1} + \epsilon_{t+2}] &= \phi^2 y_t \\ E_t[y_{t+3}] &= E_t[\phi y_{t+2} + \epsilon_{t+3}] &= \phi^3 y_t \\ &\vdots \\ E_t[y_{t+j}] &= E_t[\phi y_{t+j-1} + \epsilon_{t+j}] &= \phi^j y_t \end{aligned} \quad (2.62)$$

Nótese que la predicción del AR(1) es su representación de media móvil de orden infinito. Además, si tomamos la predicción en el infinito (es decir, el límite cuando $j \rightarrow \infty$), vemos que es cero, lo que es razonable en un proceso estacionario con media no condicional cero:

²⁰ La predicción es algo muy difícil, sobre todo acerca del futuro, Niels Bohr.

$$\lim_{j \rightarrow \infty} E_t(y_{t+j}) = 0 = E[y_t] \quad (2.63)$$

recuerde que los momentos no-condicionales son los límites de los momentos condicionales.

Por otro lado la varianza de la predicción es:

$$\begin{aligned} V_t[y_{t+1}] &= V_t[\phi y_t + \epsilon_{t+1}] = \sigma_\epsilon^2 \\ V_t[y_{t+2}] &= V_t[\phi y_{t+1} + \epsilon_{t+2}] = (1 + \phi^2) \sigma_\epsilon^2 \\ V_t[y_{t+3}] &= V_t[\phi y_{t+2} + \epsilon_{t+3}] = (1 + \phi^2 + \phi^4) \sigma_\epsilon^2 \\ &\vdots \\ V_t[y_{t+j}] &= V_t[\phi y_{t+j-1} + \epsilon_{t+j}] = (1 + \phi^2 + \dots + \phi^{2(j-1)}) \sigma_\epsilon^2 \end{aligned} \quad (2.64)$$

De la ecuación (2.64) se desprende que la imprecisión de la predicción es creciente en el tiempo. Nuevamente, tomamos el límite de la varianza para descubrir que la incertidumbre en el proceso estacionario crece con j pero está acotada:

$$\lim_{j \rightarrow \infty} V_t(y_{t+j}) = \frac{1}{1 - \phi^2} \sigma_\epsilon^2 = V[y_t] \quad (2.65)$$

Este análisis puede extenderse tediosamente aunque sin dificultades analíticas para el $AR(p)$. Existe una técnica que estudiaremos más adelante que permite escribir un $AR(p)$ como si fuese un $AR(1)$ de modo de computar las predicciones y sus varianzas de acuerdo a las ecuaciones (2.64) y (2.65) fácilmente.

Predicción de un modelo $MA(q)$

Como hemos visto, la predicción de un modelo $AR(p)$ es equivalente a obtener su representación $MA(\infty)$ y tomar esperanza y varianza de la predicción. En el caso del $MA(q)$ esto es directo. Si consideramos el modelo $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$, entonces:

$$\begin{aligned} E[y_{t+j}] &= E[\mu + \epsilon_{t+j} + \theta_1 \epsilon_{t+j-1} + \theta_2 \epsilon_{t+j-2} + \dots + \theta_q \epsilon_{t+j-q}] \\ &= \mu + \theta_j \epsilon_t + \theta_{j+1} \epsilon_{t-1} + \dots + \theta_q \epsilon_{t+j-q} \end{aligned} \quad (2.66)$$

donde nuevamente en el límite, la predicción condicional equivale a la no-condicional. Por otro lado la varianza de la predicción es:

$$V[y_{t+j}] = V[\mu + \epsilon_{t+j} + \theta_1 \epsilon_{t+j-1} + \dots + \theta_q \epsilon_{t+j-q}] = (1 + \theta_j^2 + \theta_{j+1}^2 + \dots + \theta_q^2) \sigma_\epsilon^2 \quad (2.67)$$

donde es evidente que en el límite ($j \rightarrow \infty$) la varianza es una constante.

Las predicciones de modelos ARMA son llamadas a menudo las "funciones impulso-respuesta". La razón es que si pensamos en el efecto de una innovación que toma valor 1 en el instante 0 y cero en adelante, lo que obtenemos es exactamente lo descrito por las ecuaciones de predicción del AR(p) o del MA(q).

Recordemos, sin embargo, que hemos usado modelos sin tendencia, sin estacionalidad, sin media y a veces en diferencias, por lo que las predicciones de estos modelos en realidad corresponden más bien a simulaciones de los modelos que a predicciones "realistas" (se pueden obtener estas últimas añadiendo los componentes que hemos removido). No obstante, podemos entender bastante de la dinámica de las variables económicas estudiando su respuesta frente a innovaciones.

2.17 Simulación Histórica con Modelos ARMA

Una vez estimado el ARMA(p,q), predecir es relativamente simple. Dado el modelo:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (2.68)$$

(A) podremos predecir usando los p rezagos de y_t y los q rezagos de ϵ_t . Este tipo de simulación se llama *predicción estática*. Ver Figura 2.7.

(B) podremos predecir usando los p rezagos de y_t y los q rezagos de ϵ_t ya predichos por el modelo en iteraciones anteriores. Este tipo de simulación se llama *predicción dinámica*. Ver Figura 2.8.

Figura 2.7
Predicción Estática

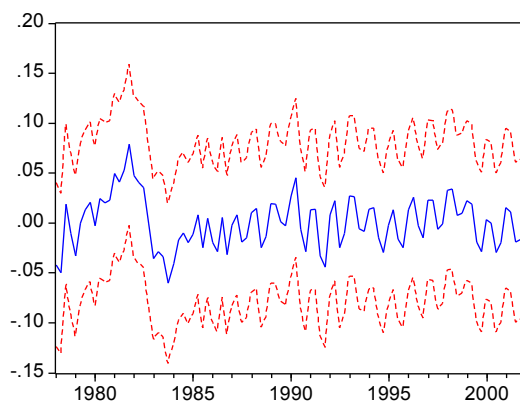
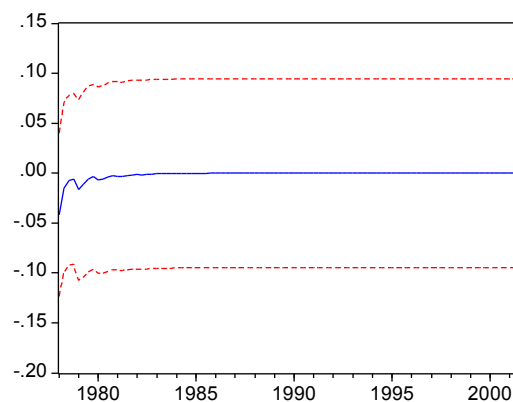


Figura 2.8
Predicción Dinámica



2.18 Validación del Modelo Económico ARMA

Una vez estimado el modelo, tenemos que estar seguros que éste tenga sentido econométrico. Algunas reglas útiles al respecto son:

- Los residuos deben ser ruido blanco (sin estructura). Existen tests de verificación que los residuos son no correlacionados (p.e., Ljung-Box o Box-Pierce).
- El ajuste del modelo debe ser "razonable". Considere que, como un modelo ARMA explota la memoria de las series, la bondad de ajuste debe ser alta.
- La significancia de los parámetros individuales es irrelevante, porque va a haber colinealidad. No obstante, la significancia de los parámetros en bloque es relevante (test con F).
- Hay que tener cuidado con el problema de sobreparametrizar (*overfitting*). Debido a que estamos trabajando con polinomios, mientras más polinomios usemos el R^2 tenderá a crecer. Un criterio más preciso que la identificación visual de los correlogramas son los tests de largo de regazos. Entre los más comunes están aquellos de "contenido informacional", que consisten en hacer un *trade-off* entre ajuste (medido por la suma de residuos al cuadrado) y exceso de parámetros (k) respecto del tamaño de muestra (T):

- | | | |
|------------------------|--------|--|
| • Test de Akaike: | \min | $-2 \frac{\log \mathcal{L}(\theta)}{T} + \frac{2k}{T}$ |
| • Test de Schwartz | \min | $-2 \frac{\log \mathcal{L}(\theta)}{T} + \frac{k \log T}{T}$ |
| • Test de Hannan-Quinn | \min | $-2 \frac{\log \mathcal{L}(\theta)}{T} + \frac{2k \log(\log(T))}{T}$ |

Apéndice A: Código Gauss para simulación de modelos ARMA

```

/* PROGRAMA QUE SIMULA CUALQUIER TIPO DE ARMA(p,q) y COMPUTA AUTOCORRELACIONES TOTALES Y PARCIALES
*/

new;
cls;
library pgraph;
"Largo de la serie (50+)" ;;
len=con(1,1);
p=0;q=0;
"Coefficientes del proceso ARMA(p,q)" ;
"Orden de p:" ;;
p=con(1,1);
if p ne 0;
    "Ingresa coeficientes de la parte AR: ";;
    phi=con(p,1);
endif;
"Orden de q:" ;;
q=con(1,1);
if q ne 0;
    "Ingresa coeficientes de la parte MA: ";;
    theta=con(q,1);
    theta=1|theta;
endif;

data=rndn(len+50,1);

if p==0 and q==0;
    tit="Ruido Blanco";
    serie=trimr(data,50,0);
elseif p==0 and q>0;
    field=1;fmat="%*.*f";lb=ftos(q,fmat,field,0);
    tit="MA(" $+lb$+")";
    i=1;
    do while i<=q;
        data=packr(data~shiftr(data[.,cols(data)]',1,(miss(0,0))''));
        i=i+1;
    endo;
    maq=data*theta;
    serie=trimr(maq,50-q,0);
elseif p>0 and q==0;
    field=1;fmat="%*.*f";lb=ftos(p,fmat,field,0);
    tit="AR(" $+lb$+")";
    ar=ones(len+50,1)*100;
    i=p+1;
    do while i<=len+50;
        j=1;
        aux=zeros(p,1);
        do while j<=p;
            aux[j,1]=phi[j,1]*ar[i-j,1];
            j=j+1;
        endo;
        ar[i,1]=sumc(aux)+data[i,1];
        i=i+1;
    endo;
    serie=trimr(ar,50,0);
else;
    field=1;fmat="%*.*f";
    lbp=ftos(p,fmat,field,0);
    lbq=ftos(q,fmat,field,0);
    tit="ARMA(" $+lbp$+"," $+lbq$+")";
    ar=ones(len+50,1)*0;
    i=maxc(p|q)+1;
    do while i<=len+50;
        j=1;
        auxp=zeros(p,1);
        do while j<=p;

```

```

        auxp[j,1]=phi[j,1]*ar[i-j,1];
        j=j+1;
    endo;
    k=1;
    auxq=zeros(q,1);
    do while k<=q;
        auxq[k,1]=theta[k,1]*data[i-k,1];
        k=k+1;
    endo;
    ar[i,1]=sumc(auxp)+sumc(auxq)+data[i,1];
    i=i+1;
end;
serie=trimr(ar,50,0);

{rhot,rhop,inter}=myauto(serie);
mono(serie,ones(1,1)|rhot,ones(1,1)|rhop,inter[1,1]|inter);

end;

proc(0)=mono(serie,rhot,rhop,inter);
local z,zz,linea;
graphset;

fonts("simplex simgrma");
_pnumht={0.25};
_pnum=2;
_pdate="";
_plwidth = 3;
_pltype={6};
_ptitlt={0.35};
_protate=0;
begwind;
window(3,1,0);
setwind(1);
_protate=0;
z="Serie ";
title(z$+tit);
xy(seqa(0,1,rows(serie)),serie);
nextwind;
z="Correlograma Total de ";
title(z$+tit);
_pltype={6,6,1,1};
xtics(0,rows(rhot)-1,1,1);
_pcolor={1,2,3,4};
linea=zeros(rows(rhot),1);
xy(seqa(0,1,rows(rhot)),rhot~linea~inter~(-1*inter));
nextwind;
z="Correlograma Parcial de ";
title(z$+tit);
xy(seqa(0,1,rows(rhot)),rhop~linea~inter~(-1*inter));
endwind;
endp;

proc(3)=myauto(data);
local z,data,vardata,i,autoc,autop,inter,bhat;
z=minc((trunc(rows(data)/4)|24));
data=data-meanc(data);
vardata=data'data;
autoc=zeros(z,1);
autop=autoc;
i=1;
do while i<=z;
    autoc[i,1]=trimr(data,i,0)'packr(lagn(data,i))/vardata;
    i=i+1;
end;

inter=ones(rows(autoc),1)*2/sqrt(rows(data));

i=1;
do while i<=z;
    data=packr(data~shiftr(data[.,cols(data)]',1,(miss(0,0))''));

```

```
bhat=data[:,1]/(ones(rows(data),1)~data[:,2:cols(data)]); /*autocorrelacion parcial */
autop[i,1]=bhat[rows(bhat),1];
i=i+1;
end;

retp(autoc,autop,inter);
endp;
```


Apéndice B: Operador de Rezagos

1. El operador de rezagos L es definido como: $Ly_t = y_{t-1}$.
2. Si se aplica repetidamente el operador L a la variable y_t se obtiene:
 $L^p y_t = L \circ L \circ L \dots \circ L y_t = y_{t-p}$.
3. El operador tiene función inversa L^{-1} de modo tal que $LL^{-1}y_t = L^{-1}Ly_t = y_t$.
4. El operador es lineal, $L^{-1}\alpha y_t = \alpha L^{-1}Ly_t = \alpha y_{t-1}$.
5. Note que $(1 - \phi L)^{-1} = (1 + \phi L + \phi^2 L^2 \dots)$.
6. Se puede usar el operador para definir un operador similar que se utiliza en modelos multivariados, llamado el operador de rezagos polinomial. Este se define como $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$. Al igual que en el caso anterior, este operador tiene inversa $\phi(L)\phi(L)^{-1}y_t = \phi(L)^{-1}\phi(L)y_t = y_t$.
7. Propiedades del operador polinomial de rezagos:
 - a. Multiplicación: $a(L)b(L) = (a_0 + a_1 L + \dots)(b_0 + b_1 L + \dots)$.
 - b. Propiedad conmutativa: $a(L)b(L) = b(L)a(L)$.
 - c. Potencias enteras positivas: $a(L)^2 = a(L)a(L)$.
 - d. Inversión: $a(L) = (1 - \lambda_1)(1 - \lambda_2) \Rightarrow a(L)^{-1} = (1 - \lambda_1)^{-1}(1 - \lambda_2)^{-1}$

Apéndice C: Metodologías de Remoción de Estacionalidad para Variables Estacionarias

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

En esta sección consideramos las dos principales metodologías de remoción de la estacionalidad que parten de la base que las series son estacionarias: el método de las variables mudas (*dummies*) y la metodología X-11²¹. Estas son las metodologías más antiguas y, pese a ser aún populares, son limitadas e imponen fuertes restricciones a la naturaleza estocástica de las series y de la estacionalidad.

Método de Regresión

Esta es, probablemente, la manera más simple de remover componentes estacionales. No obstante es la que impone mayores restricciones sobre las características que debe cumplir el fenómeno de estacionalidad. En el método de regresión se utiliza una serie de variables mudas para capturar los efectos estacionales:

$$y_t^* = \alpha + \beta t + \sum_{i=1}^q \hat{\beta}_i D_i + ARMA(p, q)\epsilon_t$$

donde cada variable D_i toma valores 1 en el periodo que ésta representa (p.e., enero de cada año) y cero en cualquier otro periodo. El modelo puede ser extendido para incluir otras variables explicativas o un proceso más complejo –aunque estacionario– para el error (p.e., un ARMA).

En la medida que los coeficientes $\hat{\beta}_i$ sean estadísticamente significativos, habrán componentes estacionales. Consecuentemente, la serie desestacionalizada y_t^* se calcula como:

$$y_t^* = y_t - \sum_{i=1}^q \hat{\beta}_i D_i$$

donde $\hat{\beta}_i$ son los parámetros estimados para cada componente estacional que resulten significativos.

21 La metodología de la variación en x-periodos implícitamente supone que la serie es no estacionaria –al menos en el componente estacional.

Como resulta evidente, al modelar la estacionalidad de esta manera se asume que el efecto estacional sea determinístico (o constante en valor esperado), es decir que en cada año el cambio en la variable por razones estacionales sea exactamente el mismo. Ello no es un supuesto adecuado para muchas de las variables económicas debido a que el comportamiento estacional se determina por numerosas y muy heterogéneas fuentes, lo que sugiere que éste sea modelado como una variable aleatoria.

El uso de variables mudas para modelar estacionalidad en un análisis de regresión presenta el problema de inducir correlaciones espúreas entre las variables²², en tanto que su uso para predecir variables en modelos de series de tiempo (p.e., ARIMA, VAR) produce predicciones fuera de muestra que son de peor calidad que las que se obtienen de otros métodos (X-11 o la variación en x -periodos), en especial en modelos multivariados²³.

No obstante, este método satisface algunas características que son deseables en cualquier método de remoción de estacionalidad: que se preserve el promedio de la serie original, que los componentes estacionales sean ortogonales entre sí, y que al aplicar el método a la serie desestacionalizada no se obtengan nuevos factores estacionales (idempotencia).

Pese a su popularidad, las metodologías tradicionales tienen grandes limitaciones. Implícitamente, ellas suponen que la variable que se desea desestacionalizar cumple determinadas propiedades estadísticas, las que usualmente no son verificadas previamente en los datos. Algunas técnicas suponen que los factores estacionales son determinísticos en tanto que otras asumen que éstos son estocásticos y estacionarios. En general, estas metodologías requieren que las variables originales sean estacionarias para estimar adecuadamente los componentes estacionales. Inadvertidamente, algunos métodos imponen estructuras dinámicas que las variables no necesariamente poseen (p.e., la variación en x -periodos). Cuando dichas propiedades no se verifican, las metodologías de remoción del componente estacional entregan series filtradas que presentan fuertes distorsiones y que pueden llevar a mala especificación en modelos o a conclusiones equivocadas cuando éstas se usan para hacer tests de hipótesis.

Más allá de las limitaciones técnicas de los distintos métodos de desestacionalización, existe una razón más profunda para investigar el comportamiento estacional de las principales variables macroeconómicas y que se relaciona con nuestro conocimiento de las fuentes y efectos de las fluctuaciones de la economía. Beaulieu y Miron²⁴ señalan que, en ocasiones, el estudio de los componentes estacionales puede

22 Ver Abeyasinghe, T. (1991). Inappropriate use of seasonal dummies in regression, *Economics Letters*, 36(2):175-179.

23 Ver Abeyasinghe, T. (1994). Deterministic seasonal models and spurious regressions, *Journal of Econometrics*, 61(2):259-272.

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

convertirse en una herramienta útil para comprender la estructura de los ciclos económicos. La razón fundamental para ello es que el estudio de ciclos requiere imponer condiciones de identificación sobre los shocks (transitorios, permanentes) que afectan a la variable de interés, de la misma manera que se hace para distinguir en un mercado cualquiera los shocks de oferta y demanda. Según estos autores, resulta más creíble imponer dichas restricciones sobre los componentes estacionales que sobre los componentes fundamentales.

Método de los Promedios Móviles

En líneas genéricas, la metodología se compone de dos etapas que funcionan de manera independiente. El objetivo de la primera etapa es eliminar de la serie aquellas observaciones "extremas" que, de no ser eliminadas, podrían distorsionar la medición de los componentes estacionales. Para ello, se estima un modelo ARIMA que incluye regresores (dummies) para capturar efectos de eventos especiales (feriados, años bisiestos, etc.), observaciones atípicas (outliers), cambios de niveles de las series (quiebres), tendencia y un conjunto de variables mudas que operan como estimadores preliminares de la estacionalidad. El modelo ARIMA se escoge de modo tal que satisfaga varios requisitos, entre ellos un bajo error de predicción fuera de muestra, ausencia de correlación y estacionariedad de residuos. Una vez estimado el modelo ARIMA se realizan dos transformaciones a la serie original: (1) se reemplazan los valores extremos por valores filtrados eliminándose la posible distorsión, y (2) se predice un cierto número de periodos (típicamente 2 años) fuera de cada extremo de la muestra y se añaden los valores predichos a la serie original. El objetivo de esto último se describe a continuación.

En la segunda etapa se estiman los componentes estacionales por medio de promedios móviles. En una primera parte, se estima el componente de tendencia de la serie transformada en la etapa anterior utilizando los llamados promedios móviles de Henderson. El promedio móvil de Henderson es un promedio ponderado de manera decreciente de 3, 5, 7, ... observaciones centrado en la observación de interés, por lo que considera valores pasados y futuros. El número de observaciones es frecuentemente seleccionado de modo automático por el programa. En la segunda etapa, se estiman los componentes estacionales de la serie ajustada en la primera etapa a la cual se le han eliminado adicionalmente los componentes de tendencia. El papel que cumplen las observaciones añadidas en la primera etapa es mejorar el cálculo de los componentes estacionales usando promedios móviles en los extremos de la muestra donde la información está truncada.

El procedimiento para datos trimestrales es el siguiente (para datos mensuales el procedimiento es análogo):

- Se transforman los datos usando la técnica de Box y Cox (1964).²⁵
- Se filtra la serie usando $x_t = \frac{[0.5 y_{t+2} + y_{t+1} + y_t + y_{t-1} + 0.5 y_{t-2}]}{4}$.
- Se obtiene la serie $r_t = \frac{y_t}{x_t}$.
- Se calculan los índices estacionales, i , como el promedio de los valores de r estimados para cada trimestre (por ejemplo, para el primer trimestre se usan r_1, r_5, r_9, \dots).
- Se normalizan los índices estacionales para preservar la media de la serie original de la siguiente manera: $s_i = \frac{i_q}{\sqrt{i_1 i_2 i_3 i_4}}$.
- Se obtiene la serie desestacionalizada como $y_t^* = \frac{y_t}{s_i}$ y se verifica que la modelación es adecuada haciendo tests de especificación y de estacionalidad residual.

Aunque el método ARIMA X-11 es muy popular no está exento de limitaciones. En primer lugar, el método produce en la mayor parte de las aplicaciones prácticas un filtro no lineal, porque en los subfiltros se utilizan promedios de Henderson con distinto número de rezagos determinados de manera endógena en función de la presencia de valores extremos en los datos (Ghysels y Perron, 1993).²⁶ Ello distorsiona las propiedades dinámicas de las series. Segundo, el filtro induce correlación en datos distantes producto del uso de promedios móviles de Henderson que no necesariamente existe en las series originales. Tercero, para varias clases de modelos el uso de un ARIMA para la predicción en las puntas de la muestra utilizando variables expresadas en logaritmos produce un sesgo sistemático hacia abajo en las series ajustadas por estacionalidad (Franses, 1997)²⁷. La magnitud y sentido del sesgo son desafortunadamente, desconocidos porque dependen de la estructura particular de cada muestra y modelo. Cuarto, el tratamiento de los componentes irregulares (por ejemplo, Semana Santa que cae en distintos días del mes de abril de cada año) no es necesariamente óptimo en esta metodología, favoreciéndose su remoción por métodos indirectos previos al filtrado de los datos con X-11 (Lee, 1998).²⁸

25 Box, G and D. Cox (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B*, pp. 211-264.

26 Ghysels, E. and P. Perron (1993). "The effect of seasonal adjustment filters on tests for a unit root", *Journal of Econometrics* (55)1-2. pp. 57-98

27 Franses Ph.H.B.F. (1997) Are Many Current Seasonally Adjusted Data Downward Biased? Discussion Paper, EUR-FEW-EI-97-17/A, Erasmus University at Rotterdam.

28 Lee, G. (1998). X-12 ARIMA Seasonal Adjustment in Korean Economic Time Series, *The Bank of Korea Economic Papers*, vol 1, No 2, September.

Apéndice D: Ejercicios

1. Explique en qué consiste la estacionariedad de un proceso estocástico. ¿Cuál es la diferencia entre estacionariedad débil y fuerte? ¿Qué es ergodicidad? ¿Por qué el siguiente proceso no es estacionario (demuestre)? $y_t = y_{t-1} + \epsilon_t$ con ϵ_t i.i.d.
2. ¿Para cuáles valores de θ y ϕ son estacionarios los siguientes procesos?
Demuestre. $y_t = (1 + 1.4L + \theta L^2)\epsilon_t$ $\epsilon_t \sim i.i.d.(0,1)$
 $z_t(1.4L - \phi L^2) = \epsilon_t$ $\epsilon_t \sim i.i.d.(0,1)$
3. ¿Por qué la función de autocorrelación parcial de $AR(p)$ muere súbitamente?
4. Demuestre que la función de autocorrelación parcial de un $MA(q)$ cae suavemente.
5. Demuestre que $(1 - \phi L)^{-1} = (1 + \phi L + \phi^2 L^2 + \dots)$ usando una expansión de Taylor.
6. Suponga que $\epsilon \in R^2$ tal que es $i.i.d. N[0, \Omega]$, donde Ω es una matriz de 2×2 positiva y definida. Suponga que $z_t = [x_t, y_t]$. La solución estacionaria del modelo primitivo (estructural) es

$$A_0 z_t = A_1 z_{t-1} + \epsilon_t$$

donde $z_0 = 0$. Los coeficientes de las matrices son:

$$A_0 = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \quad A_1 = \begin{bmatrix} \phi & \beta \\ 0 & \theta \end{bmatrix}$$

- a) encuentre la forma reducida de z_t . Exprese la matriz de varianzas de residuos de la forma reducida –que usted llamará Σ – en función de Ω , A_0 y A_1 . Exprese la matriz de parámetros de la forma reducida –que usted llamará Π – en función de Ω , A_0 y A_1 .
 - b) escriba la función de verosimilitud del problema y encuentre los estimadores de la forma reducida.
 - c) haga el test $\beta=0$.
 - d) Suponiendo que tiene estimados los parámetros de la forma reducida y computados los residuos, ¿cómo puede imponer restricciones para identificar los parámetros de la forma primitiva?
7. Suponga que $x_t = \alpha x_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$ donde $|\theta| < 1$ y $x_0 = 0$. Considere la regresión de x_t en x_{t-1} y el estimador:

- a) Suponga que $|\alpha| < 1$. ¿Es un estimador consistente de α ? Demuestre y encuentre la distribución límite del estimador.
- b) Suponga ahora que $\alpha = 1$. ¿Es un estimador consistente de α ? Encuentre la distribución límite del estimador.
8. Demuestre que la función de autocorrelación de un proceso autoregresivo x_t de orden 2 de la forma $x_t = \frac{1}{3}x_{t-1} + \frac{2}{9}x_{t-2} + \epsilon_t$ -donde ϵ_t un ruido blanco- es:

$$\rho(\tau) = \frac{16}{21} \left(\frac{2}{3} \right)^\tau + \frac{5}{21} \left(\frac{-1}{3} \right)^\tau \quad \text{para } \tau = 0; 1; 2; \dots$$

9. Sobre la base de un correlograma muestral Juan Nosé, econometrista aficionado, ha determinado que un proceso de segundo orden es el apropiado para modelar un fenómeno de interés. Usando una muestra de tamaño 200, obtuvo $\text{corr}(x_t, x_{t-1}) = 0.7$ y $\text{corr}(x_t, x_{t-2}) = 0.5$ y propuso los siguientes modelos:

$$\begin{aligned} (1) \quad x_t &= 0.7x_{t-1} + 0.5x_{t-2} + \epsilon_t \\ (2) \quad x_t &= 0.7\epsilon_{t-1} + 0.5\epsilon_{t-2} + \epsilon_t \end{aligned}$$

pero la simulación de estos modelos resultó ser un fracaso.

- a) ¿Por qué los modelos propuestos son inadecuados?
- b) ¿Para cuáles valores de parámetros ϕ_1 y ϕ_2 es posible hacer un modelo AR(2) para x_t ?
- c) Si se simula la variable x_t sobre la base del modelo 1 propuesto ¿qué espera usted que suceda?
10. Una secuencia de observaciones $\{x_1, x_2, \dots, x_n\}$ tiene varianza muestral $c_0 = 14.5$ y autocovarianza de orden 1 $c_1 = 5$. Demuestre que hay más de un proceso MA(1) que es congruente con esta estructura dinámica para esa misma muestra, pero que sólo uno de ellos es invertible.
11. Considere un ARMA(2,1) cuyos parámetros estimados son $\hat{\phi}_1, \hat{\phi}_2$ y $\hat{\theta}$. Considere que la varianza de las innovaciones es σ^2 . Demuestre que el intervalo de confianza de la predicción de este modelo crece con el horizonte de proyección pero converge asintóticamente.
12. Comente el siguiente razonamiento: *Cada modelo VAR o ARIMA tiene un número infinito de representaciones MA. Pero sólo una de ellas es la representación fundamental de acuerdo al Teorema de Wold. Sin embargo, no existen razones para descartar a-priori representaciones no-fundamentales, ya que éstas podrían corresponder a descripciones más fidedignas de una economía.*

13. Suponga que y es una variable generada por un modelo $AR(p_1)$ estacionario con innovaciones ϵ_t^1 . Suponga que z es generada por un modelo $AR(p_2)$ estacionario con innovaciones ϵ_t^2 . Las innovaciones ϵ_t^1 y ϵ_t^2 son independientes y no correlacionadas. ¿Qué proceso tiene $w=y+z$?
14. Considere el Teorema de Wold. ¿Se sostiene este teorema si las innovaciones del proceso estocástico de interés no son i.i.d? ¿Se sostiene el teorema si el DGP es no lineal? ¿Se sostiene el teorema si hay componentes determinísticos estacionales? Explique cuidadosamente sus respuestas.
15. Suponga que la variable aleatoria y es generada por un modelo $AR(p_1)$ estacionario con innovaciones ϵ_1 . Suponga que la variable aleatoria z es generada por un modelo $ARMA(p_2, q)$ estacionario con innovaciones ϵ_2 . Las innovaciones ϵ_1 y ϵ_2 son independientes y no correlacionadas. ¿Qué proceso tiene $w=y+z$?

Contenidos

Capítulo 2: Modelos de Series de Tiempo Estacionarias Univariados.....	1
2.01 Primera Representación Fundamental de un Proceso Estocástico.....	1
2.02 Estacionariedad y Ergodicidad.....	2
2.03 Función de Autocorrelación.....	4
2.04 Proceso Ruido Blanco.....	4
2.05 Procesos Media Móvil.....	6
2.06 Procesos Autoregresivos.....	9
2.07 Procesos Autoregresivos y de Media Móvil.....	15
2.08 Identificación de Procesos Estocásticos Estacionarios.....	16
2.09 Funciones de Correlación Admisibles y el Teorema de Wold.....	16
2.10 Metodología de Box-Jenkins.....	19
2.11 ¿Cómo Volver Estacionaria una Variable?.....	20
2.12 Identificación del Tipo de Modelo ARMA.....	21
2.13 Estimación de Modelos Autoregresivos.....	22
2.14 Estimación de Modelos Media Móvil.....	25
2.15 Estimación de un Modelo ARMA.....	26
2.16 Predicciones con Modelos ARMA.....	26
2.17 Simulación Histórica con Modelos ARMA.....	28
2.18 Validación del Modelo Econométrico ARMA.....	29
Apéndice A: Código Gauss para simulación de modelos ARMA.....	30
Apéndice B: Operador de Rezagos.....	33
Apéndice C: Metodologías de Remoción de Estacionalidad para Variables Estacionarias.....	34
Apéndice D: Ejercicios.....	38
Contenidos.....	41

Ultima corrección: 06/03/10