

Series de Tiempo Económicas

Modelos estacionarios

Alejandro Mosiño - *Universidad de Guanajuato*

v. 2026.03.09

Introducción

Estacionariedad (1/2)

Recordemos que una serie de tiempo, y_t , es estacionaria si:

1. Su media es constante:

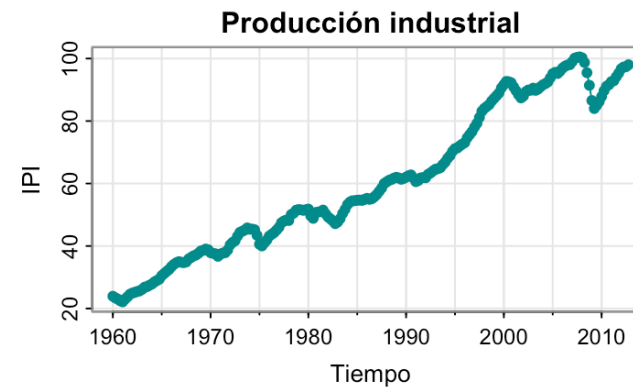
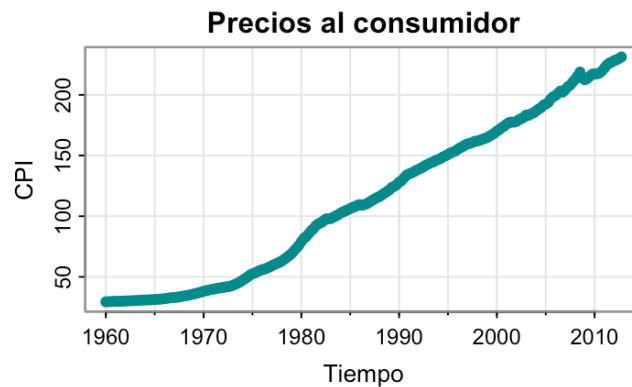
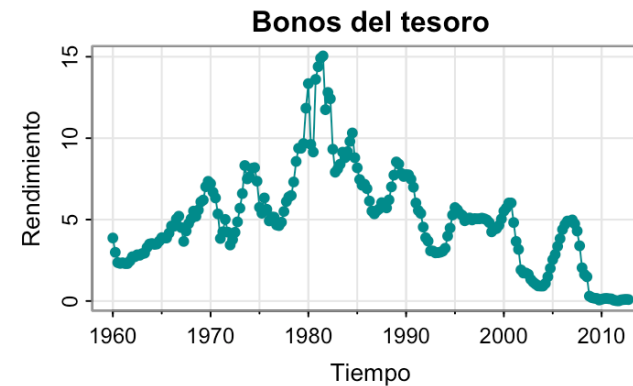
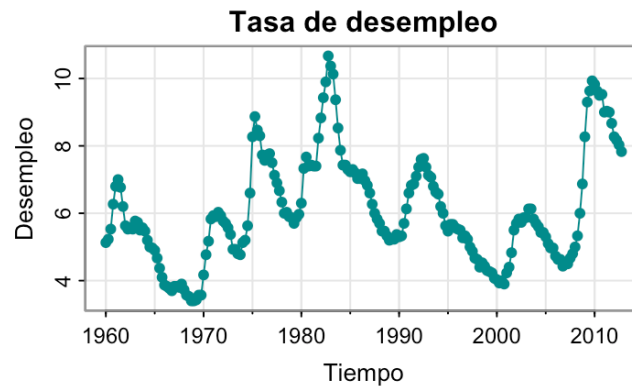
$$\begin{aligned}\mathbb{E}(y_t) &= \mu_{y_t} \\ &= \mu\end{aligned}$$

2. Su función de autocovarianza, $Cov(y_t, y_{t-h})$, es tal que:

$$\begin{aligned}Cov(y_t, y_{t-h}) &= \gamma_y(h) \\ &= \gamma(h); \quad \forall h \in \mathbb{Z}\end{aligned}$$

Estacionariedad (2/2)

¿Cuáles de estas series de tiempo parecen estacionarias?



Modelos estacionarios

Para modelar series de tiempo estacionarias usamos modelos **Autorregresivos con Promedios Móviles (ARMA)**. Estos combinan:

- *Parte autorregresiva, AR*: La serie se modela como una combinación lineal de sus valores pasados.
- *Parte de medias móviles, MA*: La serie se modela en función de sus errores pasados.

Los modelos ARMA nos permiten capturar la **parte cíclica** de una serie de tiempo, es decir, las fluctuaciones que ocurren alrededor de una media constante.

Modelos AR

Formalmente, un modelo **AR(p)** se define como:

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\ &= \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \end{aligned}$$

Donde:

- ϕ_i son los coeficientes autorregresivos, $i = 1, \dots, p$.
- ε_t es un ruido blanco.

Modelos MA

Un modelo **MA(q)** se define como:

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \\ &= \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \end{aligned}$$

Donde:

- θ_j son los coeficientes de promedios móviles, $j = 1, \dots, q$.
- ε_t es un ruido blanco.

Modelos ARMA

Un modelo **ARMA**(p,q) combina los modelos AR(p) y MA(p):

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Por ejemplo, un modelo ARMA(**1**,**1**):

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

El operador de rezagos

El operador de rezagos

El **operador de rezagos**, L o B , es un *operador lineal* tal que, aplicado a una serie de tiempo, y_t :

$$Ly_t = y_{t-1}$$

Naturalmente:

$$\begin{aligned} L^2 y_t &= L(Ly_t) \\ &= Ly_{t-1} \\ &= y_{t-2} \end{aligned}$$

En general:

$$L^i y_t = y_{t-i}$$

Propiedades del operador de rezagos (1/3)

Las siguientes son las propiedades del operador de rezagos:

1. $Lc = c$; para cualquier constante c arbitraria.
2. $(L^i + L^j)y_t = y_{t-i} + y_{t-j}$; *propiedad distributiva*.
3. $L^i L^j y_t = y_{t-i-j}$; *propiedad asociativa*.
4. $L^{-i} y_t = y_{t+i}$; **operador de liderazgo**.

Propiedades del operador de rezagos (2/3)

Las siguientes son las propiedades del operador de rezagos:

5. Si $|a| < 1$, entonces:

$$\left[1 + aL + (aL)^2 + (aL)^3 + \dots\right] y_t = \frac{1}{1 - aL} y_t$$

Esto es:

$$\frac{1}{1 - aL} = 1 + aL + (aL)^2 + (aL)^3 + \dots$$

Propiedades del operador de rezagos (3/3)

Las siguientes son las propiedades del operador de rezagos:

6. Si $|a| > 1$, entonces:

$$\left[1 + (aL)^{-1} + (aL)^{-2} + (aL)^{-3} + \dots\right] y_t = -\frac{aL}{1 - aL} y_t$$

Esto es:

$$\frac{1}{1 - aL} = -\frac{1}{aL} \left(1 + (aL)^{-1} + (aL)^{-2} + (aL)^{-3} + \dots\right)$$

El operador de rezagos y los modelos ARMA (1/2)

Un modelo ARMA(p, q) puede reescribirse en términos del operador de rezagos. En particular:

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

puede reescribirse como:

$$y_t = \phi_1 L y_t + \phi_2 L^2 y_t + \cdots + \phi_p L^p y_t + \varepsilon_t \\ + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \cdots + \theta_q L^q \varepsilon_t$$

El operador de rezagos y los modelos ARMA (2/2)

Un modelo ARMA(p, q) puede reescribirse en términos del operador de rezagos:

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$

Donde:

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \cdots - \phi_p L^p$$

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \cdots + \theta_q L^q$$

Invertibilidad

Invertibilidad

En ocasiones, podría resultar útil transformar un proceso AR en un MA y viceversa, o un modelo ARMA en un AR o un MA. *Si esto es posible*, decimos que el proceso es **invertible**.

Invertibilidad, AR (1/4)

En general, un proceso AR(p):

$$\Phi(L)y_t = \varepsilon_t$$

es invertible si:

$$y_t = \Phi(L)^{-1} \varepsilon_t$$

Para esto, es necesario que el polinomio $\Phi(L)^{-1}$ exista.

Nota. Se puede demostrar que una condición *suficiente* para que el polinomio $\Phi(L)^{-1}$ exista es que *las raíces del polinomio $\Phi(L) = 0$ se encuentren fuera del círculo unitario*. El resultado es un MA(∞).

Invertibilidad, AR (2/4)

Por ejemplo, considera un proceso AR(1):

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

En este caso, es muy fácil demostrar que, si $|\phi_1| < 1$, este proceso se puede invertir en:

$$y_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

el cual es un proceso del tipo $MA(\infty)$. Nota que la única raíz del polinomio $\Phi(L) = 0$ es $1/\phi_1$. Esta está fuera del círculo unitario si:

$$\left| \frac{1}{\phi_1} \right| > 1 \iff |\phi_1| < 1$$

Invertibilidad, AR (3/4)

Ejercicio

Considera un proceso AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

1. ¿Cómo es el polinomio $\Phi(L)$?
2. ¿Cuáles son las raíces del polinomio $\Phi(L) = 0$?
3. ¿Cómo aseguramos que las raíces del polinomio $\Phi(L) = 0$ se encuentren fuera del círculo unitario?

Invertibilidad, AR (4/4)

Ejercicio

Muestra que, si las raíces del polinomio $\Phi(L) = 0$ se encuentran fuera del círculo unitario, un proceso AR(2) puede expresarse como:

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

donde: $\psi_0 = 1, \psi_1 = \phi_1, \psi_2 = \phi_1^2 + \phi_2$ y $\psi_k = \phi_1 \psi_{k-1} + \phi_2 \psi_{k-2}$, para $k > 2$.

Invertibilidad, MA (1/2)

Al igual que los procesos AR, los procesos MA(q):

$$y_t = \Theta(L)\varepsilon_t$$

pueden invertirse si:

$$\Theta(L)^{-1} y_t = \varepsilon_t$$

Para esto, es necesario que el polinomio $\Theta(L)^{-1}$ exista.

Nota. Se puede demostrar que una condición *necesaria y suficiente* para que el polinomio $\Theta(L)^{-1}$ exista es que *las raíces del polinomio $\Theta(L) = 0$ se encuentren fuera del círculo unitario*. El resultado es un $AR(\infty)$.

Invertibilidad, MA (2/2)

Ejercicio

Muestra que, si las raíces del polinomio $\Theta(L) = 0$ se encuentran fuera del círculo unitario, un proceso MA(1) puede expresarse como:

$$y_t = - \sum_{i=1}^{\infty} \psi_i y_{t-1} + \varepsilon_t$$

donde: $\psi_i = (-\theta_1)^i$, para $i \geq 1$.

Invertibilidad, ARMA (1/2)

Finalmente, los procesos ARMA(p, q) pueden invertirse en procesos MA(∞) o AR(∞) dependiendo de las raíces de los polinomios $\Phi(L) = 0$ y $\Theta(L) = 0$.

Invertibilidad, ARMA (2/2)

Ejercicio

Muestra que, si las raíces del polinomio $\Phi(L) = 0$ se encuentran fuera del círculo unitario, un proceso ARMA(1, 1) puede invertirse en un proceso MA(∞):

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

donde: $\psi_0 = 1, \psi_1 = \phi_1 + \theta_1, \psi_k = \phi_1 \psi_{k-1}$, para $k \geq 2$.

Función de Impulso-Respuesta

Función de Impulso-Respuesta (1/2)

La **Función de Impulso-Respuesta (FIR)**:

- Muestra cómo un choque en el término de error, ε_t , afecta a una serie de tiempo, y_t , en periodos futuros.
- Permite analizar la *dinámica temporal* de un proceso y la persistencia de los efectos de un shock.

Función de Impulso-Respuesta (2/2)

Consideremos un proceso ARMA(p, q):

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$

Deseamos saber el efecto que tiene un shock (*normalizado a una unidad*) en el periodo s -ésimo, ε_s , sobre la variable y_t . Este efecto depende de los parámetros p y q .

Función de Impulso-Respuesta: MA(1)

Consideremos un proceso MA(1):

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

En este caso, la FIR:

Periodo	$s-2$	$s-1$	s	$s+1$	$s+2$	$s+3$	...
ε_t	0	0	1	0	0	0	...
y_t	0	0	1	θ_1	0	0	...

Función de Impulso-Respuesta: MA(q)

Consideremos un proceso MA(q):

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

En este caso, la FIR:

Periodo	$s-2$	$s-1$	s	$s+1$	$s+2$	$s+3$	...	$s+q$	$s+q+1$	...
ε_t	0	0	1	0	0	0	...	0	0	...
y_t	0	0	1	θ_1	θ_2	θ_3	...	θ_q	0	...

Función de Impulso-Respuesta: AR(1)

Para calcular la FIR de un proceso AR(1), primero tenemos que invertirlo. Como hemos visto antes, un proceso AR(1) puede invertirse en un proceso MA(∞) en tanto el coeficiente $|\phi_1| < 1$:

$$y_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

Entonces, la FIR es:

Periodo	$s-2$	$s-1$	s	$s+1$	$s+2$	$s+3$	...
ε_t	0	0	1	0	0	0	...
y_t	0	0	1	ϕ_1	ϕ_1^2	ϕ_1^3	...

Función de Impulso-Respuesta: ARMA(p, q)

Para calcular la FIR de un proceso ARMA(p, q), primero lo invertimos en un MA(∞). Si esto es posible:

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

Entonces, la FIR es:

Periodo	$s-2$	$s-1$	s	$s+1$	$s+2$	$s+3$	...
ε_t	0	0	1	0	0	0	...
y_t	0	0	1	ψ_1	ψ_2	ψ_3	...

Función de Impulso-Respuesta: Ejemplo (1/2)

Ejercicio

Considera el proceso AR(2):

$$y_t = 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$

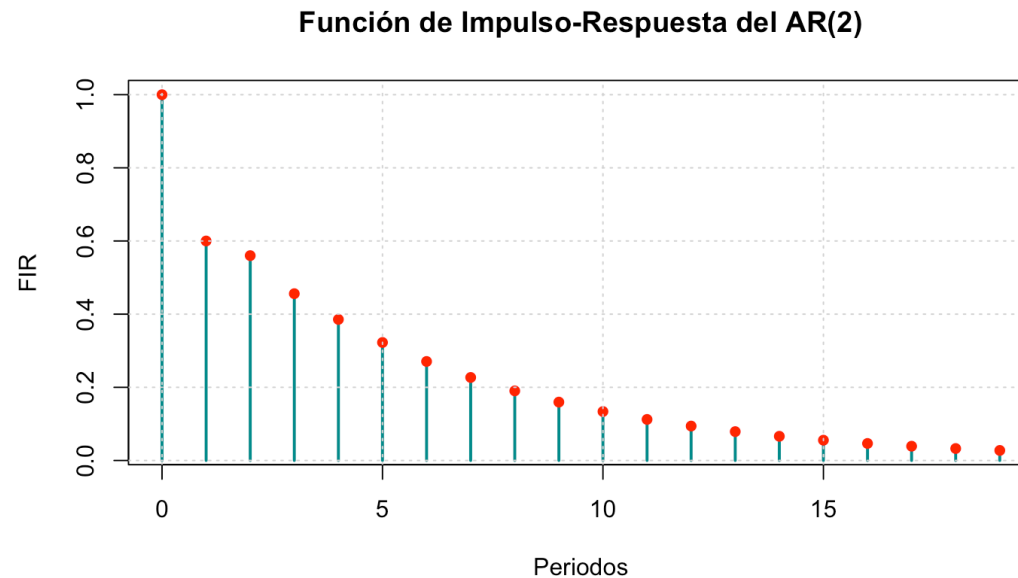
1. Muestra que la FIR es:

Periodo	$s - 2$	$s - 1$	s	$s + 1$	$s + 2$	$s + 3$	...
ε_t	0	0	1	0	0	0	...
y_t	0	0	1	0.6	0.516	0.4579	...

Función de Impulso-Respuesta: Ejemplo (2/2)

Ejercicio

2. Grafica la FIR de este proceso. Muestra que esta es:



Función de autocovarianza y autocorrelación

Función de autocovarianza y autocorrelación

Consideremos un modelo ARMA(p, q):

$$\begin{aligned}\Phi(L)y_t &= \Theta(L)\varepsilon_t \\ \varepsilon_t &\sim WN(0, \sigma^2)\end{aligned}$$

Deseamos calcular la funciones de autocovarianza y de autocorrelación. Para esto, conviene invertir el proceso en su forma MA(∞).

Función de Autocovarianza y Autocorrelación: MA(1)

Para un proceso MA(1):

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

es fácil demostrar que:

$$\gamma(h) = \begin{cases} (1 + \theta_1^2)\sigma^2 & h = 0 \\ \theta_1 \sigma^2 & |h| = 1 \\ 0 & |h| > 1 \end{cases} \quad \rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta_1}{1 + \theta_1^2} & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

Nota que este proceso cumple con las condiciones de estacionariedad y ergodicidad.

Función de Autocovarianza y Autocorrelación: MA(q)

Ahora, para un proceso MA(q):

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

tenemos:

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{i=0}^q \theta_i^2 & h = 0 \\ \sigma^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h} & 1 \leq |h| \leq q \\ 0 & |h| > q \end{cases} \quad \rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{i=0}^{q-h} \theta_i \theta_{i+h}}{\sum_{i=0}^q \theta_i^2} & 1 \leq |h| \leq q \\ 0 & |h| > 1 \end{cases}$$

donde $\theta_0 = 1$. Este proceso es, naturalmente, estacionario y ergódico.

Función de Autocovarianza y Autocorrelación: MA(∞)

(1/4)

¿Será un proceso MA(∞) estacionario? Si lo fuera:

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{i=0}^{\infty} \theta_i^2 & h = 0 \\ \sigma^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+h} & |h| \geq 1 \end{cases} \quad \rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+h}}{\sum_{i=0}^{\infty} \theta_i^2} & |h| \geq 1 \end{cases}$$

donde $\theta_0 = 1$. Nota que la estacionariedad de este proceso depende del término $\sum_{i=0}^{\infty} \theta_i^2$. ¿Es este término finito?

Función de Autocovarianza y Autocorrelación: MA(∞)

(2/4)

Proposición

Si ε_t es una secuencia de ruidos blancos con media cero y varianza finita, $\sigma^2 < \infty$, y si $\sum_{i=0}^{\infty} \theta_i^2 < \infty$, entonces la serie:

$$y_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}$$

converge en media cuadrática.

Función de Autocovarianza y Autocorrelación: MA(∞)

(3/4)

En ocasiones, resulta más fácil utilizar la condición de **convergencia absoluta**. Esta requiere:

$$\sum_{i=0}^{\infty} |\theta_i| < \infty$$

Se puede demostrar que, *si un proceso converge en valor absoluto, también lo hace en media cuadrática*.

Función de Autocovarianza y Autocorrelación: MA(∞)

(4/4)

Ejemplo

Consideremos un proceso AR(1):

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

Si $|\phi_1| < 1$, este proceso es *invertible y estacionario*, puesto que:

$$y_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \iff \sum_{i=0}^{\infty} |\phi_1^i| < \infty$$

Descomposición de Wold (1/2)

Teorema (Wold)

Sea $\{y_t\}$ un proceso débilmente estacionario con media cero y varianza finita. Entonces existe una representación única de la forma:

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \psi_0 = 1, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty,$$

donde $\{\varepsilon_t\}$ es un proceso de ruido blanco con media cero y varianza finita σ^2 .

Teorema de Wold (2/2)

Nota

En los procesos ARMA(p, q):

- *Estacionariedad*: se garantiza cuando las raíces del polinomio $\Phi(L) = 0$ están fuera del círculo unitario.
- *Invertibilidad*: se garantiza cuando las raíces del polinomio $\Theta(L) = 0$ están fuera del círculo unitario.

Ecuaciones de Yule-Walker

Ecuaciones de Yule-Walker (1/6)

Para calcular las funciones de autocovarianza y autocorrelación de un proceso $AR(p)$, también podemos utilizar las **ecuaciones de Yule-Walker**:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

- Multiplicamos ambos lados de la ecuación por y_t y calculamos el valor esperado:

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma^2$$

Ecuaciones de Yule-Walker (2/6)

Para calcular las funciones de autocovarianza y autocorrelación de un proceso $AR(p)$, podemos utilizar las ecuaciones de Yule-Walker:

- Dividimos el resultado por $\gamma(0)$:

$$1 = \phi_1\rho(1) + \phi_2\rho(2) + \cdots + \phi_p\rho(p) + \frac{\sigma^2}{\gamma(0)}$$

- Repetimos los dos pasos previos, pero ahora multiplicando ambos lados del proceso por y_{t-1} :

$$\rho(1) = \phi_1 + \phi_2\rho(1) + \phi_3\rho(2) + \cdots + \phi_p\rho(p-1)$$

Ecuaciones de Yule-Walker (3/6)

Para calcular las funciones de autocovarianza y autocorrelación de un proceso $AR(p)$, podemos utilizar las ecuaciones de Yule-Walker:

- Repetimos los primeros dos pasos, pero ahora multiplicando ambos lados del proceso por y_{t-2} :

$$\rho(2) = \phi_1\rho(1) + \phi_2 + \phi_3\rho(1) + \cdots + \phi_p\rho(p-2)$$

- Repetimos los primeros dos pasos, pero ahora multiplicando ambos lados del proceso por $y_{t-3}, y_{t-4}, \dots$. En general, si multiplicamos por y_{t-k} :

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2) + \cdots + \phi_{p-k} + \phi_{p-k+1}\rho(1) + \cdots + \phi_p\rho(p-k)$$

Ecuaciones de Yule-Walker (4/6)

Para calcular las funciones de autocovarianza y autocorrelación de un proceso $AR(p)$, podemos utilizar las ecuaciones de Yule-Walker:

- Nota que, *si conocemos los valores de los coeficientes*, $\phi_1, \phi_2, \dots, \phi_p$, podemos obtener la función de autocorrelación resolviendo el sistema de ecuaciones de forma recursiva.

Ecuaciones de Yule-Walker (5/6)

Para calcular las funciones de autocovarianza y autocorrelación de un proceso $AR(p)$, podemos utilizar las ecuaciones de Yule-Walker:

- Finalmente, nota que:

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \phi_2\rho(2) - \cdots - \phi_p\rho(p)}$$

- La función de autocovarianza se encuentra usando este resultado y recordando que:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Ecuaciones de Yule-Walker (6/6)

Ejercicio

Muestra que, para un proceso AR(2), la función de autocorrelación es:

$$\rho(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\rho(2) = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

$$\rho(3) = \frac{\phi_1^3}{1 - \phi_2} + \phi_1\phi_2 + \frac{\phi_1\phi_2}{1 - \phi_2}$$

$\vdots$

Estimación de los coeficientes de un AR(p) (1/2)

El procedimiento de Yule-Walker nos permite plantear la **fórmula de Yule-Walker**:

$$\mathbf{R}\Phi = \mathbf{r}$$

Donde:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_{p-2} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \rho_{p-2} & \rho_{p-3} & \rho_{p-4} & \dots & 1 & \rho_1 \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{pmatrix}; \quad \Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix}; \quad \mathbf{r} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{p-1} \\ \rho_p \end{pmatrix}$$

Estimación de los coeficientes de un AR(p) (2/2)

Nota

1. *Si conocemos (o estimamos) la función de autocorrelación*, $\rho_1, \rho_2, \dots, \rho_p$, podemos encontrar estimadores para los coeficientes de un proceso AR(p):

$$\hat{\Phi}^{YW} = \mathbf{R}^{-1} \mathbf{r}$$

2. Dado que la matriz $\mathbf{R}$ es cuadrada, simétrica y de rango completo, la matriz $\mathbf{R}^{-1}$ existe. Además, cada elemento del vector $\hat{\Phi}^{YW}$ puede encontrarse mediante la **regla de Cramer**.

La función de autocorrelación parcial (1/5)

Nota que un proceso $AR(p)$ puede escribirse como:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + 0 \cdot y_{t-p-1} + 0 \cdot y_{t-p-2} + \cdots + 0 \cdot y_{t-p-m} + \varepsilon_t$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

Esto es, *si realmente tenemos un proceso $AR(p)$* todos los coeficientes:

$$\begin{aligned} \phi_{p+1} &= \phi_{p+2} = \cdots = \phi_{p+m} \\ &= 0 \end{aligned}$$

La función de autocorrelación parcial (2/5)

Lo anterior indica que, si tenemos un proceso $AR(p)$, pero usamos la fórmula de Yule-Walker para estimar los coeficientes de un $AR(p + m)$, deberíamos encontrar:

- Los coeficientes estimados $\phi_1^{YW}, \phi_2^{YW}, \dots, \phi_p^{YW}$ son estadísticamente diferentes de cero.
- Los coeficientes estimados $\phi_{p+1}^{YW}, \phi_{p+2}^{YW}, \dots, \phi_{p+m}^{YW}$ son estadísticamente iguales a cero.

La función de autocorrelación parcial (3/5)

La **función de autocorrelación parcial (FACP)**, no es más que la secuencia de coeficientes ϕ_{kk} del sistema:

$$y_t = \phi_{11} y_{t-1} + \varepsilon_{1t}$$

$$y_t = \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \varepsilon_{2t}$$

$$y_t = \phi_{31} y_{t-1} + \phi_{32} y_{t-2} + \phi_{33} y_{t-3} + \varepsilon_{3t}$$

$$\vdots$$

$$y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + \varepsilon_{kt}$$

La función de autocorrelación parcial (4/5)

La FACP puede encontrarse mediante la aplicación sucesiva de la fórmula de Yule-Walker:

1. Se establece un punto de corte, por ejemplo, sea $p = T/4$.
2. Sea $1 \leq k \leq p$ y sean $\mathbf{R}^{(i)}$ y $\mathbf{r}^{(i)}$ las versiones de las matrices $\mathbf{R}$ y $\mathbf{r}$ de la fórmula de Yule-Walker para el caso k -ésimo. Para cada valor de k :

$$\hat{\Phi}^{(k)} = (\mathbf{R}^{(k)})^{-1} \mathbf{r}^{(k)} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_k \end{pmatrix}$$

3. Naturalmente, $\phi_{kk} = \hat{\phi}_k$.

La función de autocorrelación parcial (5/5)

Ejercicio

Muestra que, para un proceso AR(2), la función de autocorrelación parcial es:

$$\begin{aligned}\phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}\end{aligned}$$

Estimación por Máxima verosimilitud

Función de verosimilitud: Preliminares

Pensemos en una serie de tiempo como una secuencia de variables aleatorias $\{Y_t\}_{t=1}^T$, con realizaciones $\{y_t\}_{t=1}^T$. Si esta serie de tiempo sigue un proceso ARMA(p, q) y *todos sus parámetros son conocidos*, podemos plantear su **función de densidad conjunta**.

Para esto, hacemos uso de lo siguiente:

1. Sean X, Y y Z tres variables aleatorias con realizaciones x, y y z , respectivamente. Entonces:

$$f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x)$$

2. Además, la densidad conjunta condicional se descompone como:

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|X,Z}(y|x, z)$$

Función de densidad conjunta (1/5)

Para empezar, sea, $X = Y_1$ y $Y = Y_2, Y_3, \dots, Y_T$. Entonces:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y|x) \\ &= f_{Y_1}(y_1)f_{Y_2,Y_3,\dots,Y_T|Y_1}(y_2,y_3,\dots,y_T|y_1) \end{aligned}$$

Función de densidad conjunta (2/5)

Pero, nota que si ahora $X = Y_2$, $Y = Y_3, Y_4, \dots, Y_T$ y $Z = Y_1$, tenemos:

$$\begin{aligned} f_{X,Y|Z}(x,y|z) &= f_{X|Z}(x|z)f_{Y|X,Z}(y|x,z) \\ &= f_{Y_2|Y_1}(y_2|y_1)f_{Y_3,Y_4,\dots,Y_T|Y_1,Y_2}(y_3,y_4,\dots,y_T|y_1,y_2) \end{aligned}$$

Por lo tanto:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y|x) \\ &= f_{Y_1}(y_1)f_{Y_2|Y_1}(y_2|y_1)f_{Y_3,Y_4,\dots,Y_T|Y_1,Y_2}(y_3,y_4,\dots,y_T|y_1,y_2) \end{aligned}$$

Función de densidad conjunta (3/5)

Podemos encontrar $f_{Y_3, Y_4, \dots, Y_T | Y_1, Y_2}(y_3, y_4, \dots, y_T | y_1, y_2)$ si definimos ahora $X = Y_3$, $Y = Y_4, Y_5, \dots, Y_T$ y $Z = Y_1, Y_2$:

$$\begin{aligned} f_{X,Y|Z}(x,y|z) &= f_{X|Z}(x|z)f_{Y|X,Z}(y|x,z) \\ &= f_{Y_3|Y_1,Y_2}(y_3|y_1,y_2)f_{Y_4,Y_5,\dots,Y_T|Y_1,Y_2,Y_3}(y_4,y_5,\dots,y_T|y_1,y_2,y_3) \end{aligned}$$

Por lo tanto:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y|x) \\ &= f_{Y_1}(y_1)f_{Y_2|Y_1}(y_2|y_1)f_{Y_3|Y_1,Y_2}(y_3|y_1,y_2)f_{Y_4,Y_5,\dots,Y_T|Y_1,Y_2,Y_3}(y_4,y_5,\dots,y_T|y_1,y_2,y_3) \end{aligned}$$

Función de densidad conjunta (4/5)

Si continuamos con el mismo razonamiento, es fácil encontrar que la función de densidad conjunta las variables aleatorias $\{Y_t\}_{t=1}^T$ es:

$$\begin{aligned} f_{Y_1, Y_2, \dots, Y_T}(y_1, y_2, \dots, y_T) &= f_{Y_1}(y_1) \times f_{Y_2|Y_1}(y_2|y_1) \times f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1) \times \dots \\ &\quad \times f_{Y_T|Y_{T-1}, Y_{T-2}, \dots, Y_1}(y_T|y_{T-1}, y_{T-2}, \dots, y_1) \\ &= f_{Y_1}(y_1) \times \prod_{t=2}^T f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1}(y_t|y_{t-1}, y_{t-2}, \dots, y_1) \end{aligned}$$

Función de densidad conjunta (5/5)

Lo anterior puede aplicarse para cualquier distribución que se le asigne a las variables $\mathbf{Y}_t$. Sin embargo, conviene recordar que si $Y_t \sim \square(\mu_Y, \sigma_Y^2)$:

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left\{-\frac{1}{2\sigma_Y^2}(Y_t - \mu_Y)^2\right\}$$

Y si $\mathbf{Y} \sim \square(\mu_Y, \Sigma_Y)$

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mu_Y)' \Sigma^{-1} (\mathbf{Y} - \mu_Y)\right\}$$

Función de densidad conjunta: ejemplo (1/6)

Supongamos que la serie de tiempo, y_t sigue un proceso AR(1):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim iid \square(0, \sigma^2)$$

donde, de momento, *todos los parámetros del modelo son conocidos*. De antes, sabemos que:

$$\mu_y = \frac{\phi_0}{1 - \phi_1}$$
$$= \mu$$
$$\sigma_y^2 = \frac{\sigma^2}{1 - \phi_1^2}$$
$$= \gamma_0$$

Función de densidad conjunta: ejemplo (2/6)

Primero, necesitamos calcular $f_{Y_1}(y_1)$. Dado que hemos supuesto que ε_t sigue una distribución *iid*, podemos concluir que:

$$y_t \sim \mathcal{N}(\mu, \gamma_0)$$

Entonces:

$$\begin{aligned} f_{Y_1}(y_1) &= \frac{1}{\sqrt{2\pi\gamma_0}} \exp\left\{-\frac{1}{2\gamma_0}(y_1 - \mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\gamma_0}} \exp\left\{-\frac{1}{2\gamma_0}\varepsilon_1^2\right\} \end{aligned}$$

Función de densidad conjunta: ejemplo (3/6)

Ahora necesitamos $\mathbb{E}(y_2 | y_1)$. Nota que:

$$\begin{aligned}\mathbb{E}(y_2 | y_1) &= \phi_0 + \phi_1 y_1 \\ \mathbb{V}ar(y_2 | y_1) &= \sigma^2\end{aligned}$$

Por lo tanto:

$$\begin{aligned}f_{Y_2|Y_1}(y_2 | y_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_2 - \phi_0 - \phi_1 y_1)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_2^2\right\}\end{aligned}$$

Función de densidad conjunta: ejemplo (4/6)

Ahora necesitamos $\mathbb{E}(y_3 | y_1, y_2)$. Nota que:

$$\begin{aligned}\mathbb{E}(y_3 | y_1, y_2) &= \phi_0 + \phi_1 y_2 \\ \mathbb{V}ar(y_3 | y_1, y_2) &= \sigma^2\end{aligned}$$

Por lo tanto:

$$\begin{aligned}f_{Y_3|Y_1,Y_2}(y_3 | y_1, y_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_3 - \phi_0 - \phi_1 y_2)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_3^2\right\}\end{aligned}$$

Función de densidad conjunta: ejemplo (5/6)

Dado que tenemos un proceso AR(1), es fácil demostrar que para $k \geq 2$:

$$\begin{aligned}\mathbb{E}(y_k | y_1, y_2, \dots, y_{k-1}) &= \phi_0 + \phi_1 y_{k-1} \\ \mathbb{V}ar(y_k | y_1, y_2, \dots, y_{k-1}) &= \sigma^2\end{aligned}$$

Por lo tanto:

$$\begin{aligned}f_{Y_k | Y_1, Y_2, \dots, Y_{k-1}}(y_k | y_1, y_2, \dots, y_{k-1}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_k - \phi_0 - \phi_1 y_{k-1})^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_k^2\right\}\end{aligned}$$

Función de densidad conjunta: ejemplo (6/6)

Entonces, la función de densidad conjunta de un proceso AR(1) es:

$$f_{Y_1, Y_2, \dots, Y_T}(y_1, y_2, \dots, y_T) = \frac{1}{\sqrt{2\pi\gamma_0}} \exp\left\{-\frac{1}{2\gamma_0} \varepsilon_1^2\right\} \times \prod_{k=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_k^2\right\}$$

Función de verosimilitud

La **función de verosimilitud** no es más que la función de densidad conjunta cuando *los parámetros del modelo son desconocidos*.

En el caso de un proceso AR(1), por ejemplo, los coeficientes ϕ_0 , ϕ_1 , así como σ^2 , son desconocidos, por lo que su función de verosimilitud es:

$$L(\phi_0, \phi_1, \sigma^2) = \frac{1}{\sqrt{2\pi\gamma_0}} \exp\left\{-\frac{1}{2\gamma_0} \left(y_1 - \frac{\phi_0}{1-\phi_1}\right)^2\right\} \\ \times \prod_{k=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (y_k - \phi_0 - \phi_1 y_{k-1})^2\right\}$$

Función de log-verosimilitud

En ocasiones, resulta más conveniente calcular la **función de log-verosimilitud**. Esta no es más que el *logaritmo natural de la función de verosimilitud*.

En el caso de un proceso AR(1), por ejemplo, tenemos que la función de log-verosimilitud es:

$$\begin{aligned} \ln(\phi_0, \phi_1, \sigma^2) = & -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln\left(\frac{\sigma^2}{1 - \phi_1^2}\right) \\ & - \frac{[y_1 - \phi_0/(1 - \phi_1)]^2}{2\sigma^2/(1 - \phi_1)} - \left(\frac{T-1}{2}\right) \ln(\sigma^2) \\ & - \sum_{k=2}^T \left\{ -\frac{1}{2\sigma^2} (y_k - \phi_0 - \phi_1 y_{k-1})^2 \right\} \end{aligned}$$

Función de verosimilitud de un $AR(p)$ (1/6)

Consideremos ahora un proceso $AR(p)$:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$
$$\varepsilon_t \sim iid \square(0, \sigma^2)$$

Para calcular la función de verosimilitud, necesitamos primero calcular la función de densidad conjunta *de las primeras p observaciones*, las cuales agruparemos en el vector $\mathbf{y}_p$:

$$\mathbf{y}_p = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

Función de verosimilitud de un $AR(p)$ (2/6)

De antes, podemos deducir que:

$$\mathbf{y}_p \sim \square(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

donde:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}; \quad \boldsymbol{\Sigma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \gamma_{p-4} & \dots & \gamma_0 \end{pmatrix}$$

Y, naturalmente:

$$\mu = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p}; \quad \gamma_h = \gamma(h)$$

Función de verosimilitud de un AR(p) (3/6)

Entonces, para el vector $\mathbf{y}_p$:

$$f_{Y_1, Y_2, \dots, Y_p}(y_1, y_2, \dots, y_p) = (2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_p - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{y}_p - \boldsymbol{\mu})\right\}$$

Además, siguiendo un procedimiento similar al de un AR(1), es fácil encontrar que para el resto de las observaciones ($k > p$):

$$\begin{aligned} f_{Y_k|Y_1, Y_2, \dots, Y_{k-1}}(y_k|y_1, y_2, \dots, y_{k-1}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_k - \phi_0 - \phi_1 y_{k-1} - \phi_2 y_{k-2} - \dots - \phi_p y_{k-p})^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_k^2\right\} \end{aligned}$$

Función de verosimilitud de un AR(p) (4/6)

Entonces, la función de verosimilitud para un proceso ARMA(p) es:

$$L(\phi_0, \phi_1, \dots, \phi_p, \sigma^2) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_p - \mu)' \Sigma^{-1} (\mathbf{y}_p - \mu)\right\} \\ \times \prod_{k=p+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_k^2\right\}$$

Y la función de log-verosimilitud:

$$\ln L(\phi_0, \phi_1, \dots, \phi_p, \sigma^2) = -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\Sigma|^{-\frac{1}{2}} - \frac{1}{2} (\mathbf{y}_p - \mu)' \Sigma^{-1} (\mathbf{y}_p - \mu) \\ - \frac{1}{2\sigma^2} \sum_{k=p+1}^T (y_t - \phi_0 - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2$$

Función de verosimilitud de un $AR(p)$ (5/6)

Nota

Para encontrar la función de verosimilitud de un $AR(p)$ es necesario encontrar la forma de la matriz $\mathbf{\Sigma}^{-1} = \sigma^2 \mathbf{V}^{-1}$. Sean v_{ij} los elementos de $\mathbf{V}^{-1}$. Galbraith y Galbraith (1974) demostraron que:

$$v_{ij} = \left[\sum_{k=0}^{i-1} \phi_k \phi_{k+j-i} - \sum_{k=p+1-j}^{p+i-j} \phi_k \phi_{k+j-i} \right]$$

para $1 \leq i \leq j \leq p$ y $\phi_0 = -1$. Todos los elementos $i > j$ se encuentran por simetría.

Función de verosimilitud de un $AR(p)$ (6/6)

Ejercicio

Muestra que para un proceso $AR(2)$, la función de log verosimilitud es:

$$\begin{aligned} \ln L(\phi_0, \phi_1, \phi_2, \sigma^2) = & -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) + \frac{1}{2} \ln\{(1 + \phi_2)^2[(1 - \phi_2)^2 - \phi_1^2]\} \\ & - \left(\frac{1 + \phi_2}{2\sigma^2}\right) [(1 - \phi_2)(y_1 - \mu)^2 - 2\phi_1(y_1 - \mu)(y_2 - \mu) + (1 - \phi_2)(y_2 - \mu)^2] \\ & - \frac{1}{2\sigma^2} \sum_{k=3}^T (y_k - \phi_0 - \phi_1 y_{k-1} - \phi_2 y_{k-2})^2 \end{aligned}$$

donde:

$$\mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

Estimación por máxima verosimilitud

Para estimar los coeficientes de un proceso $AR(p)$, tenemos que *maximizar* la función de verosimilitud (o, de forma equivalente, la de log-verosimilitud) con respecto a los coeficientes desconocidos. Esta técnica se conoce como **estimación por máxima verosimilitud**.

Sin embargo, nota que esta es una función no lineal de los parámetros, por lo que *no existe solución cerrada*.

Máxima verosimilitud condicional (1/2)

Nota que, si suponemos que las *primeras p observaciones* de la serie de tiempo, $y_1, y_2, \dots, y_p$, son deterministas, la función de verosimilitud de un proceso AR(p) se simplifica a:

$$L(\phi_0, \phi_1, \dots, \phi_p, \sigma^2) = \prod_{k=p+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_k^2\right\}$$

Y, la función de log-verosimilitud:

$$\begin{aligned} \ln L(\phi_0, \phi_1, \sigma^2 | y_1) = & -\frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln(\sigma^2) \\ & - \frac{1}{2\sigma^2} \sum_{k=p+1}^T (y_k - \phi_0 - \phi_1 y_{k-1} - \dots - \phi_p y_{k-p})^2 \end{aligned}$$

Estas se conocen como funciones de **verosimilitud condicional** y de **log-verosimilitud condicional**, respectivamente. Nota que, a diferencia de la función de verosimilitud exacta, la función de verosimilitud condicional *permite una solución cerrada para los coeficientes estimados*.

Estimación por máxima verosimilitud condicional (2/2)

Ejercicio

Muestra que si las primeras p observaciones de un $AR(p)$ son deterministas, los coeficientes estimados por máxima verosimilitud (condicional) son equivalentes a los estimadores de MCO.

Muestra, además, que el estimador de la varianza del error es:

$$\sigma^2 = \frac{1}{T - p} \sum_{k=p+1}^T \varepsilon_t^2$$

Verosimilitud condicional: Modelos ARMA(p,q)

Función de verosimilitud condicional: MA(1) (1/10)

Para comenzar, consideremos un proceso MA(1):

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$
$$\varepsilon_t \sim iid \mathcal{N}(0, \sigma^2)$$

Primero, nota que, si conocemos ε_{t-1} , siempre es posible calcular:

$$\mathbb{E}(y_t | \varepsilon_{t-1}) = \mu + \theta\varepsilon_{t-1}$$
$$\mathbb{V}ar(y_t | \varepsilon_{t-1}) = \sigma^2$$

Función de verosimilitud condicional: MA(1) (2/10)

Lo anterior implica que:

$$y_t | \varepsilon_{t-1} \sim \square(\mu + \theta \varepsilon_{t-1}, \sigma^2)$$

Y, por lo tanto, para todo t :

$$f_{Y_t | \varepsilon_{t-1}}(y_t | \varepsilon_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_t - \mu - \theta \varepsilon_{t-1})^2 \right\}$$

Función de verosimilitud condicional: MA(1) (3/10)

Nota que, cuando $t = 1$:

$$y_1 = \mu + \varepsilon_1 + \theta_1 \varepsilon_0$$

Dado que no conocemos ε_0 , podemos hacer el supuesto adicional que $\varepsilon_0 = 0$. Entonces, en este caso:

$$\begin{aligned} f_{Y_1|\varepsilon_0}(y_1|\varepsilon_0 = 0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_1 - \mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_1^2\right\} \end{aligned}$$

Función de verosimilitud condicional: MA(1) (4/10)

Ahora, para $t = 2$:

$$y_2 = \mu + \varepsilon_2 + \theta_1 \varepsilon_1$$

Si condicionamos a que $Y_1 = y_1$, entonces ε_1 se vuelve conocido. Entonces:

$$\begin{aligned} f_{Y_2|Y_1, \varepsilon_0}(y_2|y_1, \varepsilon_0 = 0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_2 - \mu - \theta_1 \varepsilon_1)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_2^2\right\} \end{aligned}$$

Función de verosimilitud condicional: MA(1) (5/10)

Ahora, para $t = 3$:

$$y_3 = \mu + \varepsilon_3 + \theta_1 \varepsilon_2$$

Y, si condicionamos a que $Y_1 = y_1$ $Y_2 = y_2$, entonces:

$$\begin{aligned} f_{Y_3|Y_1,Y_2,\varepsilon_0}(y_3|y_1,y_2,\varepsilon_0=0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_3 - \mu - \theta_1\varepsilon_2)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_3^2\right\} \end{aligned}$$

Función de verosimilitud condicional: MA(1) (6/10)

Queda claro que, para todo t :

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

Y:

$$\begin{aligned} f_{Y_t|Y_1,Y_2,\dots,Y_{t-1},\varepsilon_0}(y_t|y_1,y_2,\dots,y_{t-1},\varepsilon_0=0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \mu - \theta_1\varepsilon_{t-1})^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_t^2\right\} \end{aligned}$$

Función de verosimilitud condicional: MA(1) (7/10)

Entonces, la función de densidad conjunta condicionada a que $\varepsilon_0 = 0$ para las T observaciones es:

$$f_{Y_1, Y_2, \dots, Y_T | \varepsilon_0 = 0}(y_1, y_2, \dots, y_T | \varepsilon_0 = 0) = f_{Y_1 | \varepsilon_0}(y_1 | \varepsilon_0 = 0) \times \prod_{k=2}^T f_{Y_k | Y_1, Y_2, \dots, Y_{k-1}, \varepsilon_0}(y_k | y_1, y_2, \dots, y_{k-1}, \varepsilon_0 = 0)$$

Función de verosimilitud condicional: MA(1) (8/10)

Cuando no conocemos los parámetros del proceso, tenemos la función de verosimilitud:

$$L(\mu, \theta_1, \sigma^2 | \varepsilon_0 = 0) = \prod_{k=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \varepsilon_t^2\right\}$$

Y la función de log-verosimilitud:

$$\ln L(\mu, \theta_1, \sigma^2 | \varepsilon_0 = 0) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^T \varepsilon_t^2$$

Con:

$$\varepsilon_t = \begin{cases} y_1 - \mu; & t = 1 \\ y_t - \mu - \theta_1 \varepsilon_{t-1}; & t > 1 \end{cases}$$

Función de verosimilitud condicional: MA(1) (9/10)

Nota

Para realizar el cálculo anterior, hemos supuesto que $\varepsilon_0 = 0$. ¿Qué tan importante es este supuesto?

Nota que, en cada periodo t :

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1}$$

Esta no es más que una *ecuación en diferencias* para ε_t con condición inicial $\varepsilon_0 = 0$.

Función de verosimilitud condicional: MA(1) (10/10)

Nota

Si resolvemos la ecuación en diferencias para ε_t :

$$\varepsilon_t = \sum_{i=0}^{t-1} (-1)^i \theta_1^i (y_{t-i} - \mu) + (-1)^t \theta_1^t \varepsilon_0$$

Entonces, si $|\theta_1| < 1$, el coeficiente θ_1^t converge a cero conforme t crece. Por lo tanto, el supuesto de $\varepsilon_0 = 0$ deja de tener importancia en tanto la serie de tiempo sea lo suficientemente larga.

Función de verosimilitud condicional: MA(q) (1/3)

Podemos seguir el procedimiento anterior para encontrar la función de verosimilitud para un MA(q):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$
$$\varepsilon_t \sim iid \square(0, \sigma^2)$$

Igual que antes, debemos establecer algunos supuestos sobre el comportamiento de ε_t para poder plantear la función de verosimilitud condicional.

Función de Verosimilitud condicional: MA(q) (2/3)

Ahora, es necesario suponer que:

$$\varepsilon_0 = \varepsilon_1 = \cdots = \varepsilon_{q-1} = 0$$

O bien que:

$$\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-q+1} = 0$$

¿Por qué?

Función de Verosimilitud condicional: MA(q) (3/3)

Bajo alguno de los supuestos anteriores, podemos calcular:

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Lo cual nos permite deducir que:

$$\square(\mu, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{\varepsilon}_0 = 0) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^T \varepsilon_t^2$$

donde $\boldsymbol{\theta} = (\theta_1 \ \theta_2 \ \dots \ \theta_q)^T$ y $\boldsymbol{\varepsilon}_0 = (\varepsilon_0 \ \varepsilon_{-1} \ \dots \ \varepsilon_{-q+1})^T$.

Función de Verosimilitud condicional: ARMA(p,q) (1/3)

Finalmente, para encontrar la función de verosimilitud condicional para un proceso ARMA(p,q) establecemos (por ejemplo):

$$y_t = \frac{\mu}{1 - \phi_1 - \phi_2 \dots - \phi_p}; \quad t = 0, -1, \dots, -p + 1$$

Y:

$$\varepsilon_t = 0; \quad t = 0, -1, \dots, -q + 1 = 0$$

Función de Verosimilitud condicional: ARMA(p,q) (2/3)

En este caso:

$$\ln(\mu, \phi, \theta, \sigma^2 | \mathbf{y}_0, \boldsymbol{\varepsilon}_0) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^T \varepsilon_t^2$$

Donde $\mathbf{y}_0$ se refiere a las condiciones iniciales establecidas para $\mathbf{y}_t$ y $\boldsymbol{\varepsilon}_0$ se refiere a las condiciones iniciales establecidas para $\boldsymbol{\varepsilon}_t$.

Función de Verosimilitud condicional: ARMA(p, q) (3/3)

Nota

Para poder calcular la función de verosimilitud, otra posibilidad consiste en establecer:

$$\varepsilon_t = 0; \quad t = 0, -1, \dots, -q + 1 = 0,$$

pero fijar $y_t, t = 1, 2, \dots, p$ en sus valores observados. Esta opción, sin embargo, implica perder las primeras p observaciones de la muestra. En este caso:

$$\ln(\mu, \phi, \theta, \sigma^2 | y_0, \varepsilon_0) = -\frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=p+1}^T \varepsilon_t^2$$

Identificación del orden de un $ARMA(p,q)$

Identificación del orden de un ARMA(p,q) (1/8)

En ausencia de parámetros, y bajo los supuestos de *estacionariedad* y *ergodicidad*, usamos las observaciones $\{y_t\}_{t=1}^T$ para encontrar:

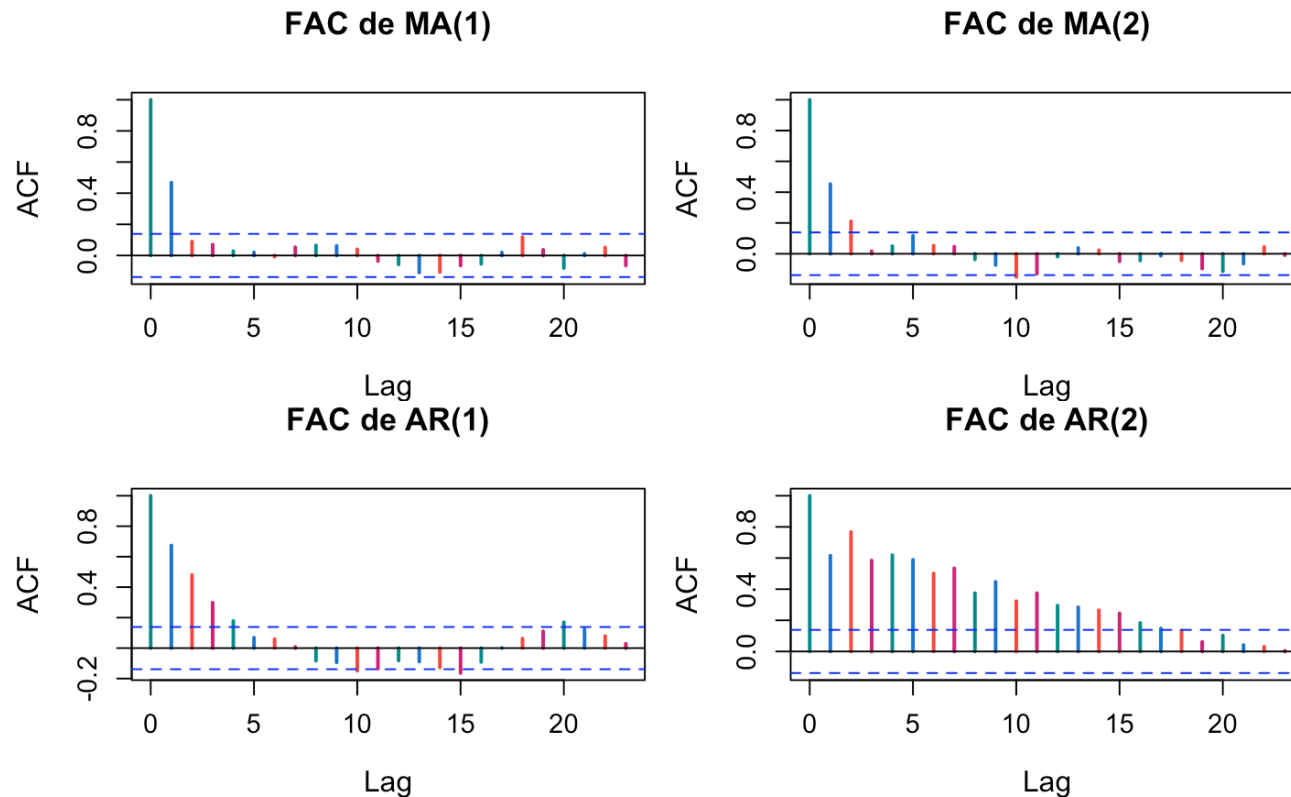
- Promedio muestral
- Autocovarianza muestral
- Autocorrelación muestral
- Autocorrelación parcial

Para identificar el orden de un ARMA(p,q), *comparamos el comportamiento de las medidas de dependencia serial muestrales con el comportamiento teórico* para diferentes valores de p y de q .

Identificación del orden de un $ARMA(p,q)$ (2/8)

Regla 1

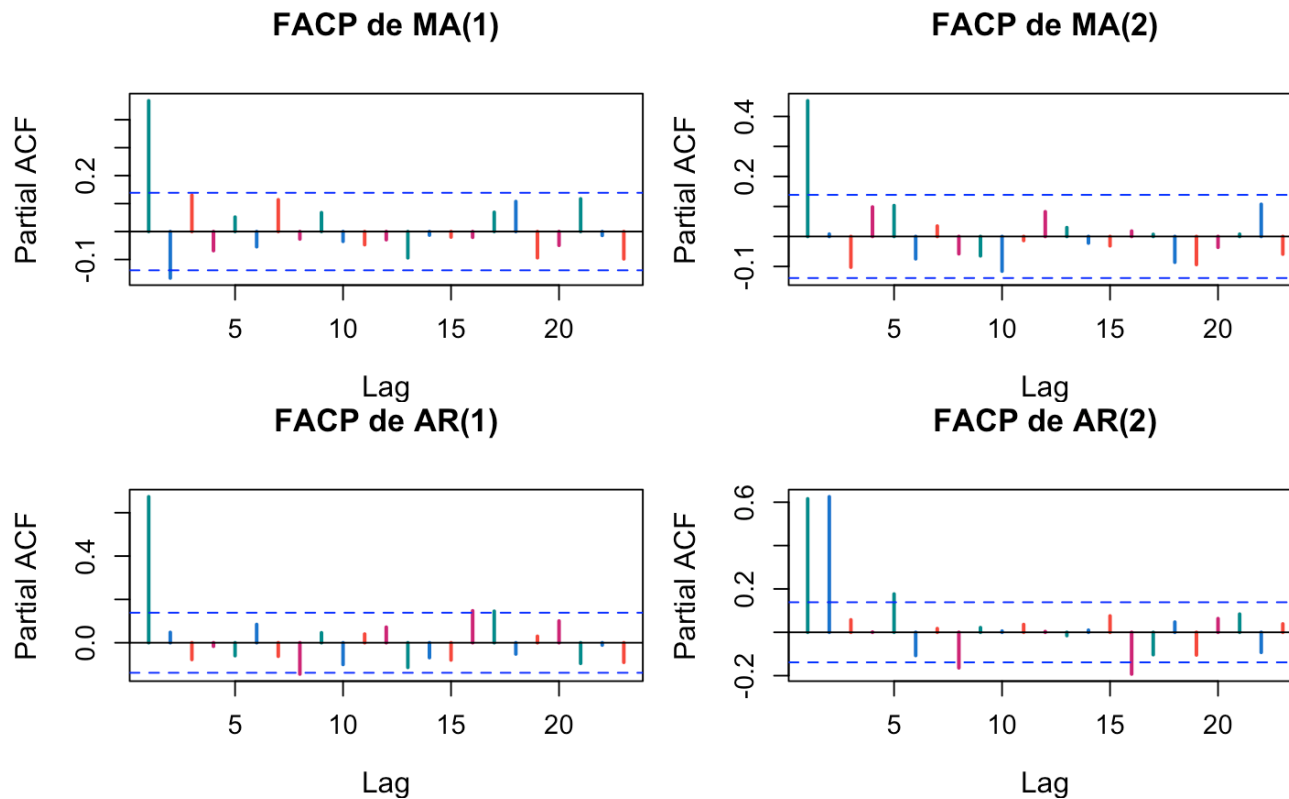
La *función de autocorrelación* puede ayudarnos a saber si el modelo adecuado es un AR/ARMA o un MA. Si el modelo es un MA, la función de autocorrelación nos indica su orden.



Identificación del orden de un $ARMA(p,q)$ (3/8)

Regla 2

Si el modelo resulta ser un $AR(p)$, la *función de autocorrelación parcial* nos indicará su orden.



Identificación del orden de un $ARMA(p,q)$ (4/8)

Regla 3

Los residuales del modelo seleccionado deben comportarse (estadísticamente) como un ruido blanco.

Identificación del orden de un ARMA(p,q) (5/8)

Para saber si el residual de un proceso se comporta (estadísticamente) como un ruido blanco, usamos el **estadístico Q** de *Ljung-Box*:

$$Q_{LB} = T(T + 2) \sum_{h=1}^m \frac{\rho_h^2}{T - h}$$

Probamos la hipótesis nula:

$$H_0 : \varepsilon_t \sim WN$$

$$H_1 : \varepsilon_t \not\sim WN$$

Si la hipótesis nula es correcta, se puede demostrar que:

$$Q_{LB} \sim \chi_m^2; \quad m \approx \sqrt{T}$$

Identificación del orden de un ARMA(p,q) (6/8)

Regla 4

Si tenemos más de un modelo, podemos seleccionar el mejor usando la función de verosimilitud. En general, el modelo con un *valor más alto* para la función de (log) verosimilitud es el mejor.

Esto también nos ayuda a elegir el orden de un proceso ARMA(p,q) cuando p y q son simultáneamente diferentes de cero.

Identificación del orden de un ARMA(p,q) (7/8)

Alternativamente, podemos elegir el modelo comparando:

- Los valores del **criterio de información de Akaike** (AIC):

$$AIC = \mathbb{E}[-2\ln(\cdot)] = T \ln(s^2) + 2k$$

- Los valores del **criterio de información de Schwartz** (SIC o BIC):

$$BIC = T \ln(s^2) + k \ln(T)$$

En general, buscamos el modelo que tenga el *menor valor posible* para estos criterios.

Identificación del orden de un ARMA(p,q) (8/8)

Regla 5

Si el AIC y el BIC seleccionan modelos diferentes, tomar en cuenta que:

- *AIC es eficiente*: selecciona el modelo que tenga el error de predicción más pequeño.
- *BIC es consistente*: selecciona el modelo correcto con probabilidad 1 si T es grande.

Pronósticos

Pronósticos usando ARMA(p,q) (1/3)

Una vez que hemos estimado los coeficientes de un proceso ARMA(p,q), podemos utilizar el modelo para realizar pronósticos.

Sea y_T el último valor disponible de y_t . Entonces, el mejor pronóstico para h periodos hacia el futuro de y_t , $\hat{y}_{T+h|T}$, $h = 1, 2, \dots$, es la **proyección lineal óptima**:

$$\begin{aligned}\hat{y}_{T+h|T} &= \mathbb{E}(y_{T+h} \mid \Omega_T), \\ \Omega_T &= \{y_T, y_{T-1}, y_{T-2}, \dots\}\end{aligned}$$

la cual, *en el caso de normalidad*, coincide con el valor esperado condicional.

Pronósticos usando ARMA(p,q) (2/3)

Ejemplo

Consideremos un proceso AR(1):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t, \quad |\phi_1| < 1$$
$$\varepsilon_t \sim iid \square(0, \sigma^2)$$

En este caso tenemos:

$$\begin{aligned}\mathbb{E}(y_{T+1} \mid \Omega_T) &= \phi_0 + \phi_1 y_T, \\ \mathbb{E}(y_{T+2} \mid \Omega_T) &= \phi_0(1 + \phi_1) + \phi_1^2 y_T, \\ \mathbb{E}(y_{T+3} \mid \Omega_T) &= \phi_0(1 + \phi_1 + \phi_1^2) + \phi_1^3 y_T, \\ &\vdots \\ \mathbb{E}(y_{T+h} \mid \Omega_T) &= \phi_0 \sum_{j=0}^{h-1} \phi_1^j + \phi_1^h y_T \\ &= \mu(1 - \phi_1^h) + \phi_1^h y_T\end{aligned}$$

Pronósticos usando ARMA(p,q) (3/3)

Ejemplo (cont'd)

Nota que, para un proceso AR(**1**), si consideramos un horizonte más largo:

$$\lim_{h \rightarrow \infty} \mathbb{E}(y_{T+h} \mid \Omega_T) = \mu,$$

donde:

$$\mu = \frac{\phi_0}{1 - \phi_1}.$$

Esto es, $\mathbb{E}(y_{T+h})$ *converge a la media no condicional* de y_t .

Varianza del pronóstico (1/2)

Por supuesto, la calidad del pronóstico y_{T+h} depende crucialmente de la **varianza del error del pronóstico**, $\mathbb{V}ar[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)]$.

Ejemplo

Consideremos un proceso AR(1). La varianza del error de pronóstico de y_t para h periodos hacia el futuro es:

$$\begin{aligned}\mathbb{V}ar[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)] &= \sigma^2 \sum_{j=0}^{h-1} \phi_1^{2j} \\ &= \sigma^2 \frac{1 - \phi_1^{2h}}{1 - \phi_1^2}\end{aligned}$$

Varianza del pronóstico (2/2)

Nota que, conforme h crece, tenemos que la varianza del pronóstico *converge a la varianza no condicional* de y_t :

$$\lim_{h \rightarrow \infty} \text{Var}[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)] = \frac{\sigma^2}{1 - \phi_1^2}.$$

Pronósticos, caso general

Recordemos que, si un proceso ARMA(p, q) es estacionario e invertible, existe una representación MA(∞) de la forma:

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \psi_0 = 1.$$

En este caso:

$$\begin{aligned} \text{Var}[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)] &= \sigma^2 \sum_{i=0}^{h-1} \psi_i^2 \\ \lim_{h \rightarrow \infty} \mathbb{E}(y_{T+h} \mid \Omega_T) &= \mu. \end{aligned}$$

Intervalos de predicción

Si $\varepsilon_t \sim iid \square(0, \sigma^2)$, un intervalo de confianza de $100(1 - \alpha)\%$ para y_{T+h} se calcula como:

$$\hat{y}_{T+h|T} \pm z_{1-\alpha/2} \sqrt{\mathbb{V}ar[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)]}.$$

Por ejemplo, si $\alpha = 0.05$, $z_{1-\alpha/2} = 1.96$. Entonces:

$$\hat{y}_{T+h|T} \pm 1.96 \sqrt{\mathbb{V}ar[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)]}.$$