

Métodos Cuantitativos

II

Modelos Autorregresivos con Medias Móviles

Alejandro Mosiño - *Universidad de Guanajuato*

v. 2026.02.09

Introducción

Estacionariedad (1/2)

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

1. Su media es constante:

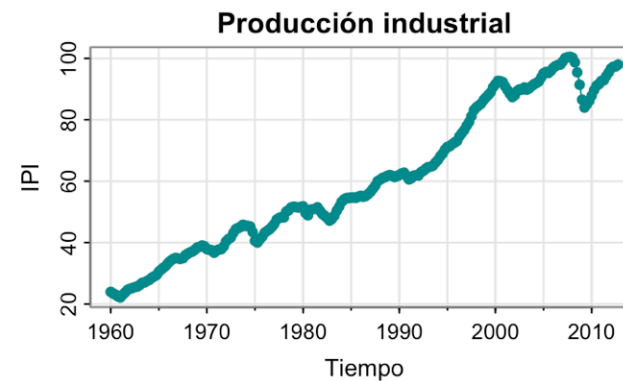
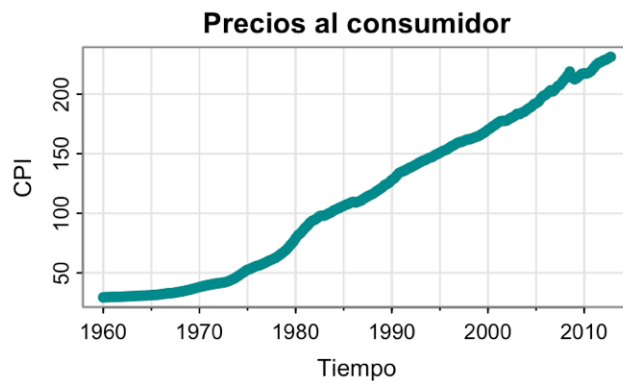
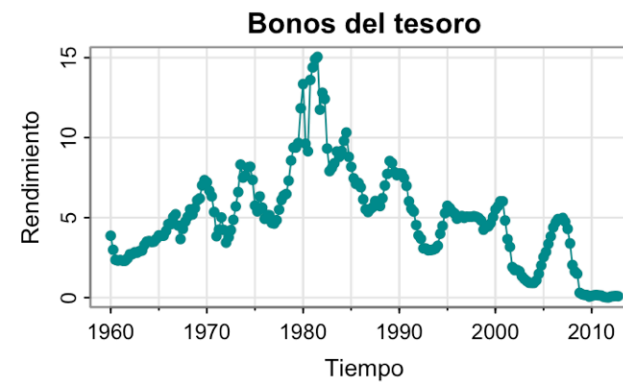
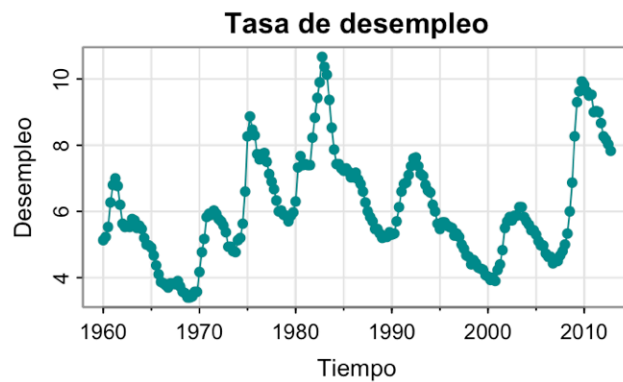
$$\begin{aligned}\mathbb{E}(y_t) &= \mu_{y_t} \\ &= \mu\end{aligned}$$

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

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

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(q):

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

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:

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

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:

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

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

Estacionariedad e Invertibilidad (1/2)

Usando la notación en términos del operador de rezagos, es posible demostrar que un proceso ARMA(p, q) es:

1. **Estacionario** si las raíces de $\Phi(L) = 0$ están *fuera* del círculo unitario ($|L| > 1$).
2. **Invertible** si las raíces de $\Theta(L) = 0$ están *fuera* del círculo unitario ($|L| > 1$).

Nota

- Un proceso AR estacionario puede reescribirse como un MA(∞).
- Un proceso MA invertible puede reescribirse como un AR(∞).

Estas condiciones garantizan que el modelo tenga una representación estable y única.

Estacionariedad e Invertibilidad (2/2)

Aunque en la práctica no trabajamos explícitamente con representaciones $AR(\infty)$ o $MA(\infty)$, su existencia es fundamental porque permite:

- Calcular momentos poblacionales (medidas de dependencia serial).
- Construir funciones impulso–respuesta.
- Derivar la *función de autocorrelación parcial*.

Descomposición de Wold y medidas de dependencia serial

Descomposición de Wold (1/3)

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 las funciones promedio, de autocovarianza y de autocorrelación. Para esto, conviene expresar el proceso en su representación MA(∞). Luego, hacemos uso de la **descomposición de Wold**.

Descomposición de Wold (2/3)

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

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

donde ε_t es un ruido blanco.

Descomposición de Wold (3/3)

Entonces, si y_t es una serie estacionaria:

$$\mathbb{E}(y_t) = \mu.$$

Y:

$$\gamma_y(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

Descomposición de Wold: MA(1)

Como ejemplo, consideremos un proceso MA(1):

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

Este proceso cumple naturalmente con la representación de Wold, con:

$$\psi_0 = 1, \quad \psi_1 = \theta_1.$$

Descomposición de Wold: AR(1)

Ahora, consideremos un proceso AR(1):

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

Si $|\phi_1| < 1$, el proceso tiene representación MA(∞):

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

Entonces, en este caso:

$$\psi_i = \phi_1^i.$$

Descomposición de Wold: Ejercicio

Usando la descomposición de Wold, muestra que, para un proceso MA(1):

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

Y muestra que, para un proceso AR(1) con $|\phi_1| < 1$:

$$\gamma(h) = \phi_1^{|h|} \frac{\sigma^2}{1 - \phi_1^2}, \quad \rho(h) = \phi_1^{|h|}.$$

Función de autocorrelación (FAC): identificación empírica

A nivel empírico, la función de autocorrelación (FAC) permite identificar patrones iniciales en la dinámica de la serie. En particular:

Modelo	Patrón FAC
AR(p)	Decae gradualmente
MA(q)	Se corta después del rezago q
ARMA(p,q)	Decae gradualmente

La FAC permite detectar la presencia de memoria en la serie y orientar la especificación del modelo.

La función de impulso respuesta

Función de Impulso-Respuesta (1/2)

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

- 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)

Si el proceso admite representación $MA(\infty)$:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

entonces la IRF es simplemente:

$$IRF(j) = \psi_j.$$

Esta nos indica el efecto que tiene un shock (*normalizado a una unidad*) en el periodo s -ésimo, ε_s , sobre la variable y_t .

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: 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: Ejemplo (1/2)

Ejercicio

Considera el proceso AR(2):

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

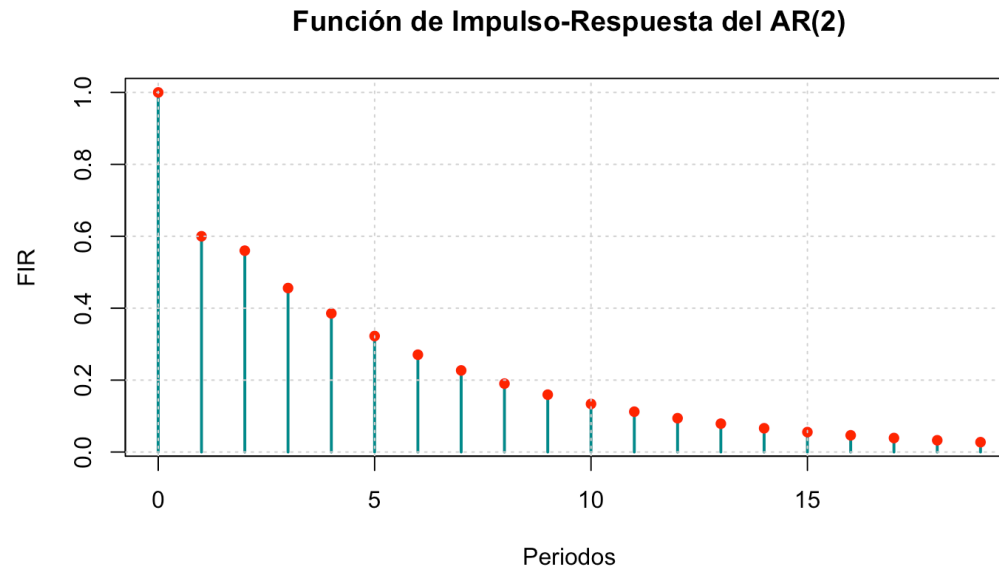
1. Muestra que la FIR es:

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 autocorrelación parcial

La función de autocorrelación parcial (1/4)

Considera un proceso $AR(p)$:

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

Este puede reescribirse como:

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

La función de autocorrelación parcial (2/4)

Lo anterior indica que, *si nuestro proceso es realmente un $AR(p)$* todos los coeficientes:

$$\phi_{p+1} = \phi_{p+2} = \cdots = \phi_{p+m} \\ = 0$$

En palabras: 1) la FACP es distinta de cero hasta el rezago p ; 2) para rezagos mayores, la FACP es cero.

Esta observación da origen a la definición de **función de autocorrelación parcial**.

La función de autocorrelación parcial (3/4)

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 FACP mide la correlación entre y_t y y_{t-k} una vez controlado el efecto lineal de los rezagos intermedios.

La función de autocorrelación parcial (4/4)

Para encontrar la FACP podemos utilizar la siguiente fórmula de *correlación condicional*:

$$\text{Corr}(X, Y \mid Z) = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1 - \rho_{XZ}^2)(1 - \rho_{YZ}^2)}}.$$

La función de autocorrelación parcial: Ejemplos

Podemos mostrar que, para un proceso AR(1), la función de autocorrelación parcial es:

$$\phi_{11} = \rho_1$$

Y, para un proceso AR(2):

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

La función de autocorrelación parcial: Estimación

En la práctica, la FACP no es más que una secuencia de coeficientes estimados por el método de los *mínimos cuadrados ordinarios*.

La función de autocorrelación parcial: identificación empírica

Al igual que la FAC, la función de autocorrelación parcial permite identificar patrones iniciales en la dinámica de la serie. En este caso tenemos:

Modelo	Patrón FACP
AR(p)	Se corta después del rezago p
MA(q)	Decae gradualmente
ARMA(p,q)	Decae gradualmente

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

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

En ausencia de parámetros, usamos las observaciones $\{y_t\}_{t=1}^T$ para encontrar:

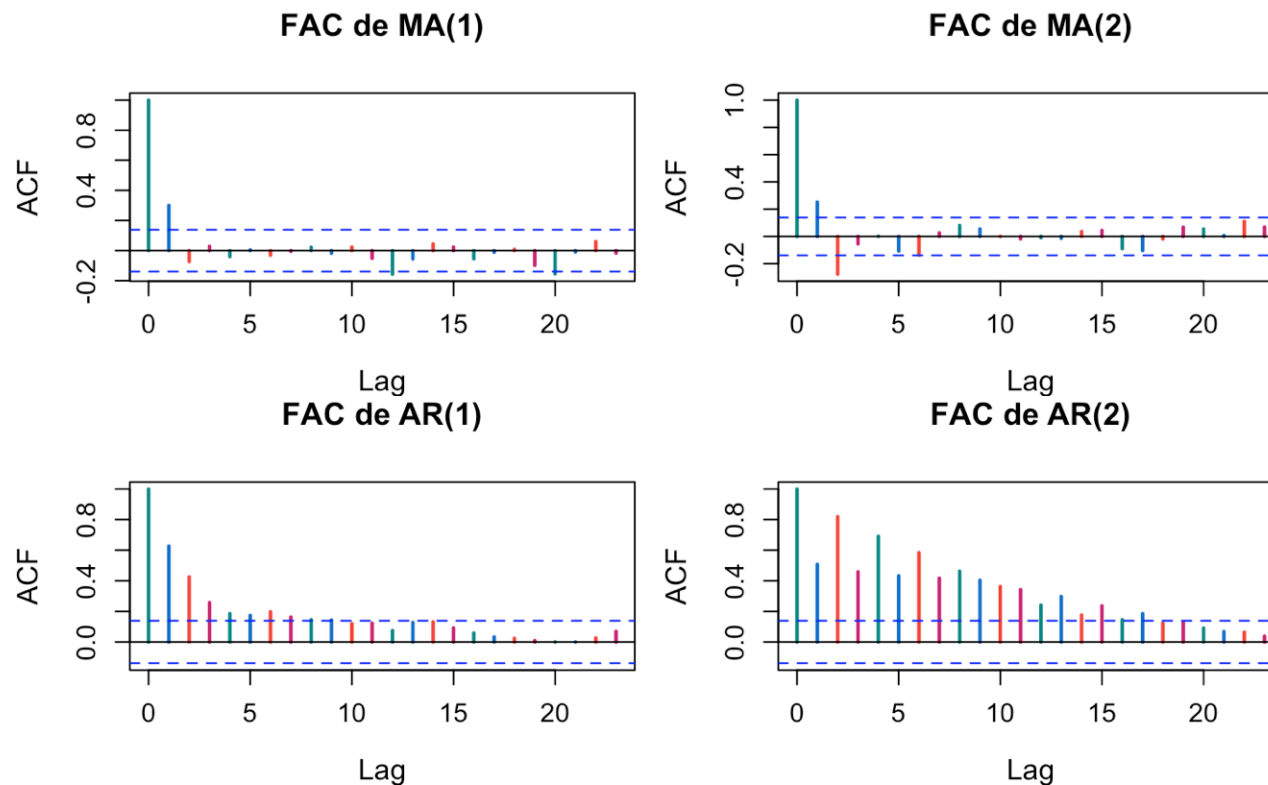
- Promedio muestral
- Autocovarianza muestral
- Autocorrelación muestral
- Autocorrelación parcial

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

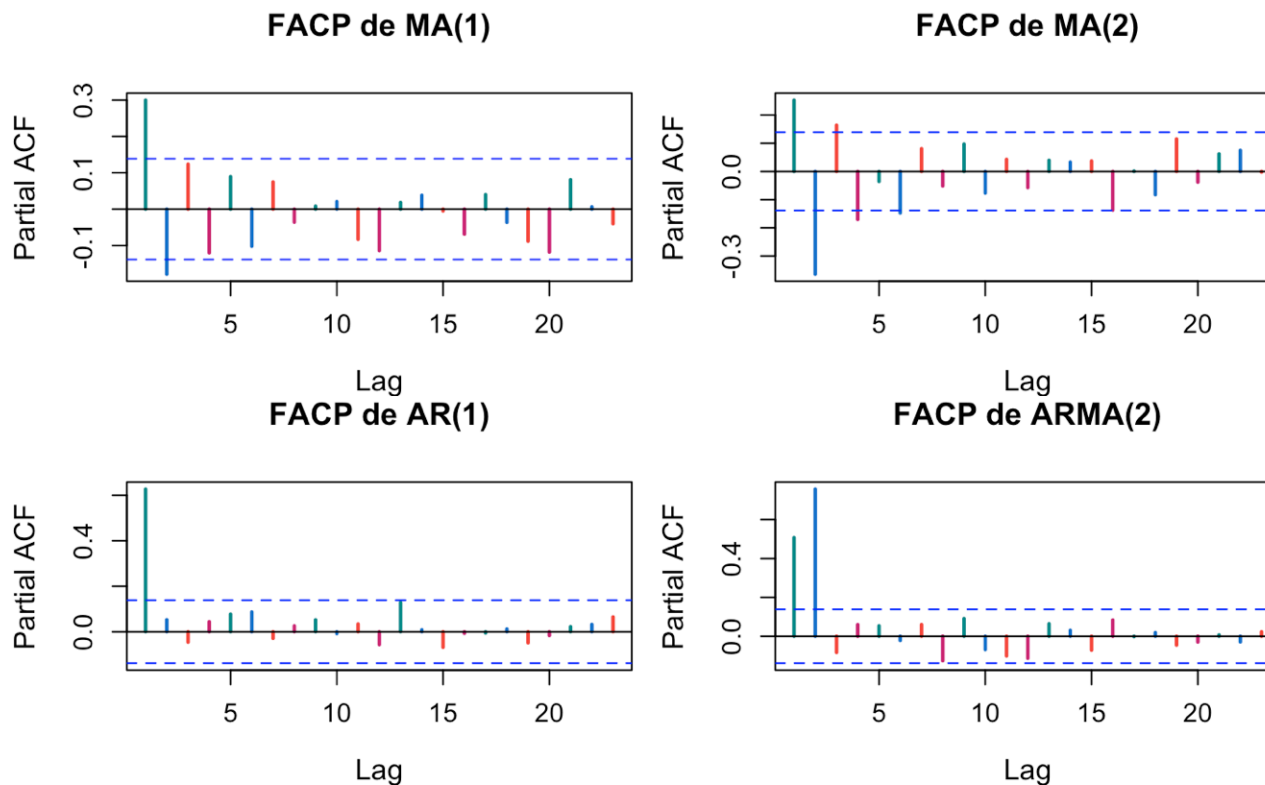
La *función de autocorrelación* puede ayudarnos a saber si el modelo adecuado es un AR/ARMA o un MA. Si el modelo es un MA, la función de autocorrelación nos indica su orden.



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.



Si la FACP presenta corte abrupto, tenemos evidencia de componente AR(p).

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):

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

- 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.

Estimación de un $ARMA(p,q)$

Nota

Una vez que hemos identificado el orden del proceso, podemos estimar este usando el **método de máxima verosimilitud**.

Pronósticos

Pronósticos usando ARMA(p, q)

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): Ejemplo

Ejemplo

Consideremos un proceso AR(1):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t, \quad |\phi_1| < 1$$

$$\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2)$$

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**, $\text{Var}[y_{T+h} - \mathbb{E}(y_{T+h} \mid \Omega_T)]$.

Ejemplo

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

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

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

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

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

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

Estimación y pronósticos: Ejemplo

En el siguiente ejemplo usamos un índice para el empleo en Canadá que va desde el primer trimestre de 1962 al cuarto trimestre de 1992. Usamos un modelo $AR(2)$ para pronosticar los cuatro trimestres de 1993 y 1994.



