

## Uso de la función de correlación cruzada en la identificación de modelos ARMA

### Use of the Crosscorrelation Function in the Identification of ARMA Models

ELKIN CASTAÑO<sup>1,2,a</sup>, JORGE MARTÍNEZ<sup>3,b</sup>

<sup>1</sup>ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA,  
MEDELLÍN, COLOMBIA

<sup>2</sup>FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD DE ANTIOQUIA, MEDELLÍN, COLOMBIA

<sup>3</sup>DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE  
COLOMBIA, BOGOTÁ, COLOMBIA

---

#### Resumen

La función de correlación cruzada muestral (FCCM) ha sido empleada para estudiar la fortaleza y la dirección de la relación lineal entre dos procesos estocásticos conjuntamente estacionarios. Rosales (2004) y Castaño (2005) muestran que dicha función, calculada entre el proceso estacionario y los residuales de un modelo preliminar estimado, puede ser empleada como un diagnóstico adicional en la identificación de un modelo apropiado  $ARMA(p, q)$  para este proceso. El propósito de este trabajo es mostrar que la FCCM entre los residuales de un modelo preliminar, aunque no sea correcto, y la serie de tiempo estacionaria, contiene información relevante del modelo adecuado y, por tanto, puede ser usado como un diagnóstico adicional en la formulación y construcción de modelos ARMA (Autoregressive-Moving Average). El procedimiento propuesto se ilustra con series reales y simuladas.

**Palabras clave:** proceso ARMA, función de autocorrelación, función de autocorrelación parcial, función de autocorrelación cruzada, identificación.

#### Abstract

The sample cross-correlation function (SCCF) has been used to study the strength and direction of the linear relation between two jointly stationary stochastic processes. Rosales (2004) and Castaño (2005) show that the cross-correlation function between a stationary process and the residuals of an estimated preliminary model can be used as an additional diagnostic tool, for the identification of an appropriate  $ARMA(p, q)$  model, for the generating process of the series. The purpose of this article is to show that the

---

<sup>a</sup>Profesor asociado, profesor titular. E-mail: elkinvc@gmail.com

<sup>b</sup>Profesor asociado. E-mail: jmartinezc@unal.edu.co

FCCM between a series and the residual of a preliminary model to describe it, not necessarily correct, contains relevant information of the correct model and for this reason it can be used as a diagnostic tool for the construction of ARMA models. The procedure is illustrated with real and simulated series.

**Key words:** ARMA process, Autocorrelation function, Partial autocorrelation function, Cross-correlation function, Identification.

## 1. Introducción

En la literatura de series de tiempo, para la identificación de un modelo autorregresivo y de medias móviles estacionario e invertible,  $ARMA(p, q)$ , generalmente se hace uso de las funciones de autocorrelación y autocorrelación parcial muestrales (denotadas por sus siglas en inglés como SACF y SPACF, respectivamente) para tratar de obtener una especificación adecuada de los órdenes  $p$  y  $q$  del proceso (Box & Jenkins 1976). Para procesos puros  $AR(p)$  y  $MA(q)$ , estas funciones generalmente suelen proporcionar suficiente información para una correcta identificación. Sin embargo, cuando el proceso es mixto, es decir, posee componentes autorregresivas y de medias móviles, la información proporcionada por la SACF y la SPACF generalmente es insuficiente para lograr la identificación correcta de  $p$  y  $q$ . Debido a esto, se sugiere utilizar herramientas como la función de autocorrelación extendida muestral (ESACF), propuesta por Tsay & Tiao (1984); el método correlación canónica más pequeña (SCAN), propuesto también por Tsay & Tiao (1985); el método de “la esquina”, propuesto por Beguin et al. (1980); el criterio de información (AIC), desarrollado por Akaike (1974) y el criterio de información bayesiano (BIC), sugerido por Schwarz (1978).

La función de correlación cruzada muestral (FCCM) ha sido empleada para estudiar la fortaleza y la dirección de la relación entre dos procesos estocásticos conjuntamente estacionarios. Rosales (2004) y Castaño (2005) muestran que dicha función, calculada entre el proceso estacionario y los residuales de un modelo preliminar estimado (aunque no sea correcto), puede ser empleada como un diagnóstico adicional en la identificación de un modelo apropiado  $ARMA(p, q)$  para la representación del proceso.

En el presente documento se explora el uso de la FCCM como herramienta complementaria para la identificación y el diagnóstico de un modelo para una serie dada. En la segunda sección se revisan los conceptos preliminares empleados en el trabajo, se presenta la caracterización de un proceso  $ARMA(p, q)$  por medio de la FCC, se prueba una proposición que permite hacer operativos los resultados obtenidos en la caracterización anterior y se ilustran estos resultados con series simuladas. En la tercera sección; se presenta un ejemplo con una serie real ampliamente conocida con el cual se muestra la utilidad de la metodología propuesta. Finalmente, en la última sección, se presentan las conclusiones y recomendaciones que se derivan de este trabajo.

## 2. Empleo de la FCC en la identificación de series estacionarias

En esta sección se mostrará que todo modelo ARMA( $p, q$ ), estacionario e invertible, puede ser caracterizado en términos de la FCC entre el proceso y su término de error aleatorio. Además, se presenta un procedimiento que permite hacer operativos los resultados obtenidos.

### 2.1. Conceptos preliminares: la función de correlación cruzada

Considere dos procesos conjuntamente estacionarios  $x_t$  y  $y_t$ , para  $t = 0, \pm 1, \pm 2, \dots$ . La covarianza cruzada de orden  $k$  entre  $x_t$  y  $y_t$  está definida como (por ejemplo, Box & Jenkins 1976, Wei 1990):

$$\gamma_{xy}(k) = E[(X_t - \mu_x)(y_{t+k} - \mu_y)]$$

para  $k = 0, \pm 1, \pm 2, \dots$ .

Como función de  $k$ ,  $\gamma_{xy}(k)$  es llamada la función de covarianza cruzada entre  $x_t$  y  $y_t$ .

La estandarización de  $\gamma_{xy}(k)$  produce la función de correlación cruzada (FCC)

$$\rho_{xy}(k) = \gamma_{xy}(k) / (\sigma_x \sigma_y)$$

para  $k = 0, \pm 1, \pm 2, \dots$ , donde  $\sigma_x$  y  $\sigma_y$  son las desviaciones estándar de los procesos  $x_t$  y  $y_t$ .

La FCC mide no solamente la fortaleza de la relación, sino también su dirección. Esta última propiedad es útil para identificar variables causales. Por esta razón, es importante examinar la FCC tanto para los valores positivos de  $k$  como para los negativos. Para valores negativos de  $k$ , la FCC describe la influencia lineal de los valores pasados de  $y_t$  sobre  $x_t$ . Para valores positivos de  $k$ , la FCC indica la influencia lineal de los valores pasados de  $x_t$  sobre  $y_t$ . El gráfico de la FCC contra  $k$ , llamado correlograma cruzado, es útil para visualizar estas relaciones.

Dada una realización de  $n$  periodos del proceso estacionario bivalente  $x_t, y_t$ , la FCC se estimada con la función de correlación cruzada muestral (FCCM):

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{\hat{\sigma}_x \hat{\sigma}_y}$$

para  $k = 0, \pm 1, \pm 2, \dots$ , donde

$$\begin{aligned} \hat{\gamma}_{xy}(k) &= \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), & \text{si } k \geq 0 \\ &= \frac{1}{n} \sum_{t=1-k}^n (x_t - \bar{x})(y_{t+k} - \bar{y}), & \text{si } k < 0 \end{aligned}$$

y donde  $\hat{\sigma}_x = [\hat{\gamma}_{xx}(0)]^{1/2}$ ,  $\hat{\sigma}_y = [\hat{\gamma}_{yy}(0)]^{1/2}$ ,  $\bar{x}$  y  $\bar{y}$  son las desviaciones estándar y las medias muestrales de las series  $x_t$  y  $y_t$ , respectivamente.

Con los supuestos de normalidad, que la serie  $x_t$  es ruido blanco y que las series  $x_t$  y  $y_t$  son incorrelacionadas, Bartlett (1985) probó que

$$\text{Var} [\hat{\rho}_{xy}(k)] \approx (n - k)^{-1}$$

Por tanto, cuando la serie  $x_t$  es ruido blanco y hay normalidad, podemos contrastar la hipótesis que las dos series tienen correlación cruzada nula comparando  $\hat{\rho}_{xy}(k)$  con su error estándar aproximado  $\sqrt{(n - k)^{-1}}$ .

## 2.2. Caracterización de un proceso ARMA( $p, q$ ) por medio de la FCC

**Proposición 1.** *Suponga que el proceso  $\{Z_t\}$  sigue un modelo ARMA( $p, q$ ) estacionario e invertible bajo las condiciones de regularidad de la forma*

$$\phi(B)Z_t = \theta(B)a_t$$

con un proceso de ruido blanco  $\{a_t\}$  de media cero y varianza  $\sigma_a^2$ . Sea  $\rho_{aZ}(k)$  la función de correlación cruzada entre el ruido blanco  $a_t$  y  $Z_t$ . Entonces  $\rho_{aZ}(k)$  es tal que:

- i) Si  $k \geq 0$ , en general  $\rho_{aZ}(k) \neq 0$
- ii) Si  $k < 0$   $\rho_{aZ}(k) = 0$
- iii) La forma de  $\rho_{aZ}(k)$ , para  $k \geq 0$ , permite caracterizar el modelo ARMA( $p, q$ ).

**Demostración.**

- i) Por definición

$$\rho_{aZ}(k) = \frac{\gamma_{aZ}(k)}{[\text{Var}(a_t)\text{Var}(Z_t)]^{1/2}}$$

Como  $Z_t$ , es estacionario, este proceso puede escribirse como

$$Z_t = \psi(B)a_t$$

siendo  $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots$ , donde  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ . Entonces,

$$\gamma_{aZ}(k) = E[a_t Z_{t+k}] = E[a_t \psi(B)a_{t+k}] = \sum_{j=0}^{\infty} \psi_j E[a_t a_{t+k-j}] = \psi_k \sigma_a^2 \neq 0,$$

para el menos un  $k \geq 0$ ; por tanto,  $\rho_{aZ}(k) \neq 0$  para al menos un  $k \geq 0$

- ii) Si  $k < 0$ , entonces

$$\gamma_{aZ}(k) = E[a_t Z_{t+k}] = E[a_t \psi(B)a_{t+k}] = \sum_{j=0}^{\infty} \psi_j E[a_t a_{t+k-j}] = 0$$

ya que no existirán subíndices comunes en  $E[a_t a_{t+k-j}]$ . Por tanto,  $\rho_{aZ}(k) = 0$  si  $k < 0$ .

iii) La derivación de los resultados teóricos se encuentran en Rosales (2004), los cuales se pueden resumir de la siguiente manera:

- Si el proceso es un  $MA(q)$ , el lado derecho se anula a partir de  $k = q + 1$ .
- Si el proceso es un  $AR(p)$ , entonces, dependiendo de las raíces de  $\phi(B) = 0$ , el lado derecho es una combinación de decrecimientos exponenciales y/o mezcla de ondas sinusoidales amortiguadas.
- Si el proceso es un  $ARMA(p, q)$ , el lado derecho es una combinación de decrecimientos exponenciales y/o mezcla de ondas sinusoidales amortiguadas, a partir de  $k = q$ , dependiendo de las raíces de  $\phi(B) = 0$  y de  $\theta(B) = 0$ . Para valores de  $k$  menores o iguales a  $q$  no hay un patrón definido.
- En todos los casos anteriores  $\rho_{aZ}(k) = 0$ , si  $k < 0$ . □

A pesar de que la proposición anterior permite caracterizar el comportamiento de un modelo  $ARMA(p, q)$  en términos de su FCC, en la práctica su empleo requiere del conocimiento de una realización del ruido blanco del modelo, que en general no es observable.

La siguiente proposición justifica el uso de la FCC entre el proceso y los residuales de un modelo preliminar ajustado a los datos, como una herramienta de diagnóstico adicional en la identificación del modelo  $ARMA(p, q)$  que genera los datos.

**Proposición 2.** *Suponga que el proceso  $\{Z_t\}$  de desviaciones de la media sigue un modelo  $ARMA(p_1, q_1)$  estacionario e invertible bajo las condiciones de regularidad, con proceso de ruido blanco  $\{a_t\}$  de media cero y varianza  $\sigma_a^2$ . Sea  $b_t$  el proceso resultante de filtrar  $Z_t$  usando un modelo  $ARMA(p_2, q_2)$  y sea  $\rho_{bZ}(k)$  la función de correlación cruzada entre el proceso filtrado,  $b_t$ , y  $Z_t$ . Si  $\rho_{aZ}(k)$  es la función de correlación cruzada entre el ruido blanco  $a_t$  y  $Z_t$ , entonces la función de correlación cruzada  $\rho_{bZ}(k)$  es tal que:*

- i) Si  $k \geq 0$ ,  $\rho_{bZ}(k)$  contiene el patrón de comportamiento de  $\rho_{aZ}(k)$ . Es decir,  $\rho_{bZ}(k)$  “conserva la memoria” del verdadero proceso  $ARMA(p_1, q_1)$ .
- ii) Si  $k < 0$ ,  $\rho_{bZ}(k)$  es, en general, una función no nula.

**Demostración.**

i) Si el verdadero proceso de  $Z_t$  es un  $ARMA(p_1, q_1)$ , entonces

$$Z_t = \frac{\theta_1(B)}{\phi_1(B)} a_t$$

donde:  $\phi_1(B) = 1 - \phi_{11}B - \phi_{12}B^2 - \dots - \phi_{1p_1}B^{p_1}$  y  $\theta_1(B) = 1 - \theta_{11}B - \theta_{12}B^2 - \dots - \theta_{1q_1}B^{q_1}$  y donde  $a_t$  es un proceso de ruido blanco de media cero y varianza constante  $\sigma_a^2$ .

Como  $Z_t$  es estacionario, puede escribirse en términos del proceso ruido blanco como

$$Z_t = \psi_1(B)a_t \quad (1)$$

donde  $\psi_1(B) = \frac{\theta_1(B)}{\phi_1(B)} = 1 + \psi_{11}B + \psi_{12}B^2 + \psi_{13}B^3 + \dots = \sum_{j=0}^{\infty} \psi_{1j}B^j$ , con

$$\sum_{j=0}^{\infty} |\psi_{1j}| < \infty, \psi_{10} = 1.$$

Ahora bien, suponga que  $Z_t$  es filtrado usando un modelo ARMA( $p_2, q_2$ ), entonces

$$\frac{\phi_2(B)}{\theta_2(B)}Z_t = b_t$$

o

$$Z_t = \frac{\theta_2(B)}{\phi_2(B)}b_t$$

donde:  $\phi_2(B) = 1 - \phi_{21}B - \phi_{22}B^2 - \dots - \phi_{2p_2}B^{p_2}$  y  $\theta_2(B) = 1 - \theta_{21}B - \theta_{22}B^2 - \dots - \theta_{1q_2}B^{q_2}$ .

Como  $Z_t$  es estacionario, también puede escribirse en términos del proceso de medias móviles

$$Z_t = \psi_2(B)b_t \quad (2)$$

donde  $\psi_2(B) = \frac{\theta_2(B)}{\phi_2(B)} = 1 + \Psi_{21}B + \psi_{22}B^2 + \psi_{23}B^3 + \dots = \sum_{j=0}^{\infty} \psi_{2j}B^j$  con

$$\sum_{j=0}^{\infty} |\psi_{2j}| < \infty, \psi_{20} = 1.$$

Observe que  $b_t$  no es necesariamente un proceso de ruido blanco, ya que de

$$b_t = \frac{\psi_1(B)}{\psi_2(B)}a_t \quad (3)$$

donde, en general,  $\frac{\psi_1(B)}{\psi_2(B)} \neq 1$ .

Consideremos ahora la FCC entre  $b_t$  y  $Z_t$ :

$$\rho_{bZ}(k) = \frac{\gamma_{bZ}(k)}{[\text{Var}(b_t)\text{Var}(Z_{t+k})]^{1/2}} \quad (4)$$

para  $k = 0 \pm 1, \pm 2, \dots$ , donde  $\gamma_{bZ}(k) = E(b_t Z_{t+k})$  para  $k = 0 \pm 1, \pm 2, \dots$ , es la función de covarianza cruzada entre  $b_t$  y  $Z_t$ .

De (3) se obtiene que

$$b_t = \frac{\psi_1(B)}{\psi_2(B)}a_t = \psi_3(B)a_t = \left(1 + \psi_{31}B + \psi_{32}B^2 + \psi_{33}B^3 + \dots\right)a_t = \left(\sum_{j=0}^{\infty} \psi_{3j}B^j\right)a_t$$

donde, debido a la estacionalidad de  $Z_t$ ,  $\sum_{j=0}^{\infty} |\psi_{3j}| < \infty$ , con  $\psi_{30} = 1$ .

Entonces, puesto que  $\sum_{j=0}^{\infty} |\psi_{3j}| < \infty$

$$\begin{aligned} \gamma_{bZ}(k) &= E [b_t Z_{t+k}] = E [Z_{t+k} \psi_3(B) a_t] \\ &= E \left[ Z_{t+k} \left( 1 + \sum_{j=1}^{\infty} \psi_{3j} B^j \right) a_t \right] \\ &= E [a_t Z_{t+k}] + \sum_{j=1}^{\infty} \psi_{3j} E [a_{t-j} Z_{t+k}] \\ &= \gamma_{aZ}(k) + \sum_{j=1}^{\infty} \psi_{3j} \gamma_{aZ}(k+j) \end{aligned} \tag{5}$$

además,

$$Var(b_t) = Var(\psi_3(B) a_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_{3j}^2 \tag{6}$$

Reemplazando (5) y (6) en (4), se tiene

$$\begin{aligned} \rho_{bZ}(k) &= \frac{\gamma_{bZ}(k)}{[Var(b_t) Var(Z_{t+k})]^{1/2}} = \\ &= \frac{\gamma_{aZ}(k) + \sum_{j=1}^{\infty} \psi_{3j} \gamma_{aZ}(k+j)}{\left[ \sigma_a^2 \sum_{j=0}^{\infty} \psi_{3j}^2 Var(Z_{t+k}) \right]^{1/2}} = \frac{\gamma_{aZ}(k) + \sum_{j=1}^{\infty} \psi_{3j} \gamma_{aZ}(k+j)}{\left[ \sum_{j=0}^{\infty} \psi_{3j}^2 \right]^{1/2} [Var(a_t) Var(Z_t)]^{1/2}} = \\ &= \frac{\gamma_{aZ}(k)}{\left[ \sum_{j=0}^{\infty} \psi_{3j}^2 \right]^{1/2} [Var(a_t) Var(Z_t)]^{1/2}} + \frac{\sum_{j=1}^{\infty} \psi_{3j} \gamma_{aZ}(k+j)}{\left[ \sum_{j=0}^{\infty} \psi_{3j}^2 \right]^{1/2} [Var(a_t) Var(Z_t)]^{1/2}} = \\ &= c \frac{\gamma_{aZ}(k)}{[Var(a_t) Var(Z_t)]^{1/2}} + c \sum_{j=1}^{\infty} \psi_{3j} \frac{\gamma_{aZ}(k+j)}{[Var(a_t) Var(Z_t)]^{1/2}} \end{aligned}$$

Por tanto,

$$\rho_{bZ}(k) = c \rho_{aZ}(k) + c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j) \tag{7}$$

donde  $c = \frac{1}{\left[ \sum_{j=0}^{\infty} \psi_{3j}^2 \right]^{1/2}}$  y  $0 < c \leq 1$ .

La ecuación (7) muestra la relación general entre  $\rho_{bZ}(k)$  y  $\rho_{aZ}(k)$ . La función  $\rho_{bZ}(k)$  contiene el patrón de comportamiento de  $\rho_{aZ}(k)$  a través de  $c \rho_{aZ}(k)$ .

Sin embargo, el comportamiento de  $\rho_{aZ}(k)$  puede ser distorsionado por el término  $c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j)$ , donde los coeficientes  $\psi_{3j}$  se obtienen igualando los coeficientes de las mismas potencias de  $B$  en ambos lados de la ecuación

$$\psi_2(B)\psi_3(B) = \psi_1(B)$$

En general,  $\psi_{3j}$  está dado por

$$\begin{aligned} \psi_{3j} &= -\psi_{21}\psi_{3,j-1} - \psi_{22}\psi_{3,j-2} - \cdots - \psi_{2,j-1}\psi_{31} - \psi_{2,j} + \psi_{1,j} \\ \psi_{3j} &= (\psi_{1,j} - \psi_{2,j}) - \sum_{i=1}^{j-1} \psi_{2,i}\psi_{3,j-1} \end{aligned} \quad (8)$$

para  $j = 1, 2, 3, \dots$

Esta solución indica que

- a) Si  $\psi_{1j} = \psi_{2j}$ , para  $j \geq 0$ , es decir, el modelo con el cual se filtra  $Z_t$  es el modelo generador de los datos de  $Z_t$ , entonces, usando (8)  $\psi_{3j} = 0$  para  $j > 0$ ,  $c = 1$ . Por tanto, como habría de esperarse

$$\rho_{bZ}(k) = \rho_{aZ}(k)$$

para  $j = 1, 2, 3, \dots$

- b) Si  $\psi_{1j} \approx \psi_{2j}$ , para  $j > 0$ , es decir, el modelo por el cual se filtra  $Z_t$  es un modelo aproximado al modelo generador de los datos, entonces  $\psi_{3j} \approx 0$ , para  $j > 0$ , y

$$\rho_{bZ}(k) \approx \rho_{aZ}(k)$$

para  $k \geq 1$ . Es decir,  $\rho_{bZ}(k)$  presenta aproximadamente el patrón de comportamiento de  $\rho_{aZ}(k)$ . Por ejemplo, suponga que el proceso  $Z_t$  es generado por  $Z_t = a_t - \theta_1 a_{t-1}$  y que el modelo para filtrar a  $Z_t$  es  $Z_t = b_t - \delta_1 b_{t-1} - \delta_2 b_{t-2}$ . Empleando la ecuación (7), y la caracterización de un MA(1) por medio de la FCC, si  $k \geq 1$

$$\rho_{bZ}(k) = c\rho_{aZ}(k)$$

ya que  $\rho_{aZ}(k+j) = 0$  para  $j = 1, 2, 3, \dots$ . En este caso,  $\rho_{bZ}(k)$  es solamente una contracción de  $\rho_{aZ}(k)$  y  $\rho_{bZ}(k)$  tiene la misma apariencia que  $\rho_{aZ}(k)$ .

Si  $k = 0$ ,

$$\rho_{bZ}(0) = c\rho_{aZ}(0) + c\psi_{31}\rho_{aZ}(1)$$

donde, de la ecuación (8),  $\psi_{31} = \psi_{11} - \psi_{21} = \theta_1 - \delta_1$ . Por tanto, cuanto más cerca estén  $\theta_1$  y  $\delta_1$  habrá menor distorsión. Resumiendo, si  $k \geq 0$ , se observa que  $\rho_{bZ}(k) \approx \rho_{aZ}(k)$ .

- c) Si  $\psi_{1j} \neq \psi_{2j}$  y  $k = 0, 1, 2, 3, \dots$ ,  $\rho_{bZ}(k)$  aún presenta aproximadamente el patrón de comportamiento de  $\rho_{aZ}(k)$  aunque el término  $c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j)$

$j$ ) puede llegar a ser grande y, en algunos casos, puede distorsionar el comportamiento de  $\rho_{aZ}(k)$ . Sin embargo, es posible establecer una cota para  $c \sum_{j=1}^{\infty} \Psi_{3j} \rho_{aZ}(k+j)$ . Partiendo de la expresión (7), se obtiene que

$$\begin{aligned}
 c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j) &= \rho_{bZ}(k) - c\rho_{aZ}(k) \\
 \left| c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j) \right| &= |\rho_{bZ}(k) - c\rho_{aZ}(k)| \\
 &\leq |\rho_{bZ}(k)| + |c\rho_{aZ}(k)| \leq 1 + c
 \end{aligned}$$

Por tanto  $c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j)$  está acotado inferiormente por  $-(1+c)$  y superiormente por  $1+c$ .

Cuando el verdadero proceso es un MA( $q$ ) entonces  $\rho_{aZ}(k) = 0$  si  $k > q$ ; en este caso, de acuerdo con (7), se tiene que

$$\rho_{bZ}(k) = 0 \quad \text{si } k > q$$

lo cual indica que para esta clase de procesos la función de autocorrelación cruzada entre  $b_t$  y  $Z_t$  desaparece a partir del verdadero orden  $q$  y el empleo de cualquier modelo para filtrar  $Z_t$  no distorsiona el tipo de proceso ni su orden  $q$ .

ii) Considere  $k < 0$  en la ecuación (7),

$$\rho_{bZ}(k) = c\rho_{aZ}(k) + c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j)$$

Entonces, puesto que para el verdadero proceso  $\rho_{aZ}(k) = 0$  si  $k < 0$ ,

$$\rho_{bZ}(k) = c \sum_{j=1}^{\infty} \psi_{3j} \rho_{aZ}(k+j) \neq 0$$

donde, en general,  $\rho_{aZ}(k+j) \neq 0$  si  $k+j \geq 0$ . Por tanto,  $\rho_{bZ}(k)$  presentará valores no nulos si  $k < 0$ . □

El siguiente ejemplo ilustra los resultados de la proposición 2.

**Ejemplo 1.** Considere el modelo

$$Z_t = (1 + 0.7B^4)a_t \tag{9}$$

donde  $\{a_t\}$  es un proceso de ruido blanco  $N(0,1)$ . A continuación se presentan los correlogramas cruzados entre  $Z_t$  y la serie  $Z_t$  filtrada por diferentes modelos ARMA.

- i)  $Z_t$  es filtrada usando el modelo  $Z_t = (1 + 0.7B^4)a_t$ , es decir, usando el verdadero proceso generador de datos.

Observando la figura 1, se concluye que el modelo que generó los datos es un MA(4) con restricciones de nulidad para los tres primeros coeficientes, el cual es el verdadero modelo. Esta información se encuentra representada en la parte derecha del correlograma, la cual es no nula únicamente para el rezago  $k = 4$ , y en la nulidad de la parte izquierda de dicha función.

- ii)  $Z_t$  filtrada usando un modelo un poco diferente al verdadero,  $Z_t = (1 + 0.4B^4)b_t$ .

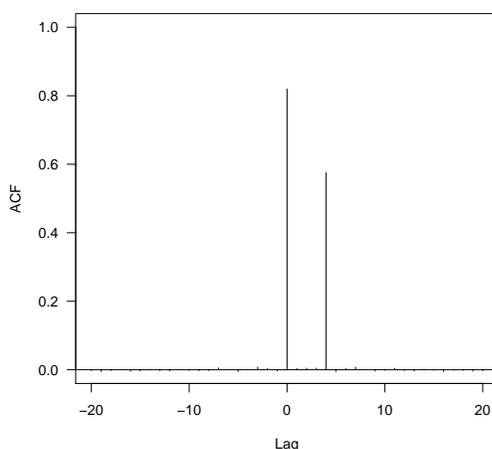


FIGURA 1: Modelo  $Z_t = (1 + 0.7B^4)a_t$ .

La figura 2 indica que el modelo que generó los datos es un MA(4) con restricciones de nulidad para los tres primeros coeficientes, el cual es el verdadero modelo. Sin embargo, el lado izquierdo no nulo señala que existe alguna especificación errónea del modelo, relacionada con la magnitud del coeficiente del operador  $B^4$ .

- iii)  $Z_t$  filtrada usando un modelo  $Z_t = 1/(1 - 0.5B^4)b_t$

Aunque el modelo utilizado es un AR(4), la parte derecha de la FCC, mostrada en la figura 3, conserva la memoria señalando claramente un MA(4) como proceso generador. Sin embargo, la parte izquierda no nula indica que el AR(4) no es el modelo verdadero.

- iv)  $Z_t$  es filtrada usando el modelo  $Z_t = 1/(1 - 0.4B)b_t$ ,

La parte derecha del correlograma de la figura 4 conserva la memoria del MA(4), aunque con alguna distorsión, y la parte izquierda no nula señala que el modelo AR(1) no es el verdadero.

Los siguientes casos presentan resultados similares en el sentido de que la parte derecha del correlograma ( $k > 0$ ) siempre conserva, en mayor o menor

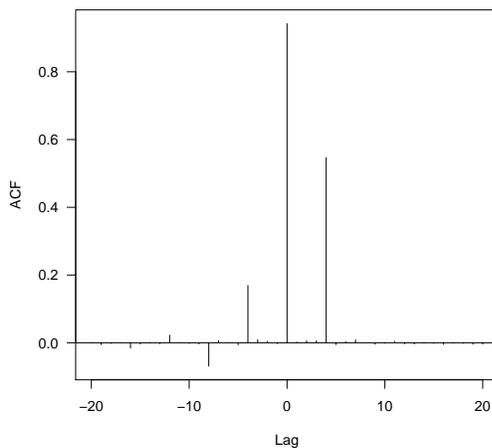


FIGURA 2: Modelo  $Z_t = (1 + 0.4B^4)b_t$ .

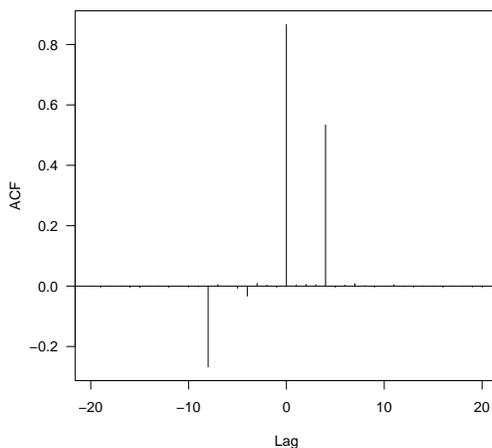
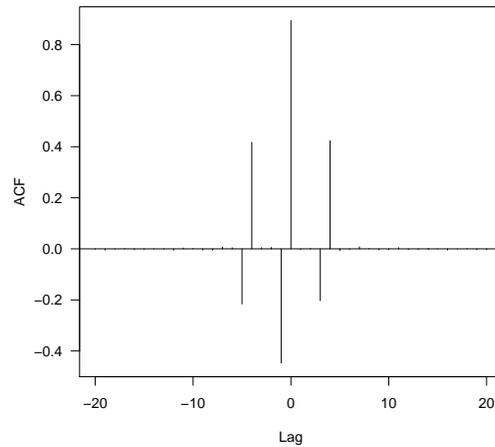
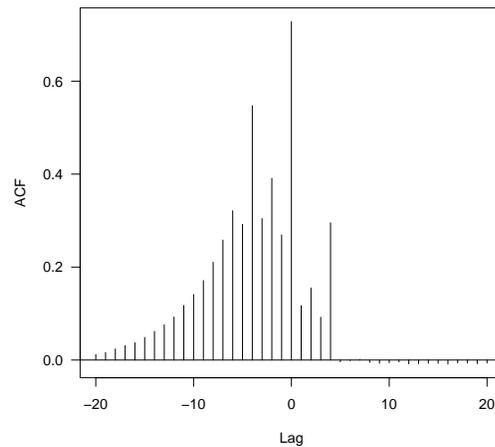


FIGURA 3: Modelo  $Z_t = 1/(1 - 0.5B^4)b_t$ .

grado, la memoria del verdadero MA(4), y la parte izquierda ( $k < 0$ ) no nula señala lo inadecuado del modelo empleado (frente al verdadero modelo) para filtrar la serie.

- v)  $Z_t$  es filtrada usando el modelo  $Z_t = (1 - 0.7B - 0.1B^2) / (1 - 0.4B + 0.2B^2)b_t$  (figura 5).
- vi)  $Z_t$  es filtrada usando el modelo  $(1 - B)(1 - B^{12})Z_t = 1/(1 - 0.6B)(1 + 0.5B^{12})b_t$  (figura 6).
- vii)  $Z_t$  es filtrada usando el modelo  $(1 - B)Z_t = -0.51 + b_t$  (figura 7).
- viii)  $Z_t$  es filtrada usando el modelo  $Z_t = (1 - 0.3B^2)b_t$  (figura 8).

FIGURA 4: Modelo  $Z_t = 1/(1 - 0.4B)b_t$ .FIGURA 5: Modelo  $Z_t = (1 - 0.7B - 0.1B^2)/(1 - 0.4B + 0.2B^2)b_t$ .

### 2.3. Aplicación de la proposición 2 a la identificación de un modelo ARMA( $p, q$ )

La proposición 2 puede ser empleada para construir un procedimiento que ayude a la identificación de un modelo ARMA( $p, q$ ). A continuación se presentan los pasos que deben seguirse.

- i) Se identifica un modelo preliminar, se estiman sus parámetros y se calcula la FCCM entre los residuales del modelo estimado y la serie  $Z_t$ ,  $\hat{\rho}_{bZ}(k)$ .
- ii) Para  $k \geq 0$ , el patrón de comportamiento de  $\hat{\rho}_{bZ}(k)$  indicará el modelo adecuado para  $Z_t$  y permitirá revisar si el modelo preliminar es correcto.

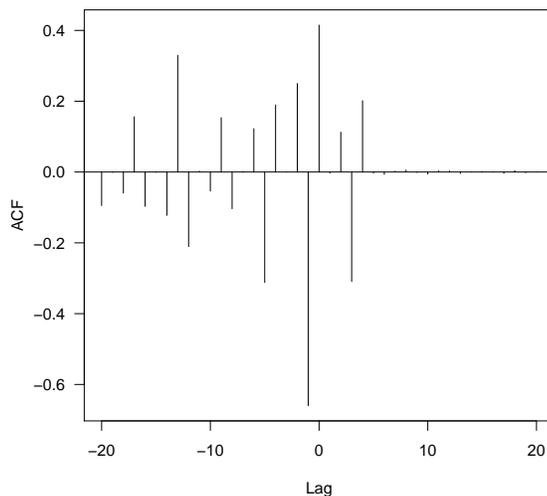


FIGURA 6: Modelo  $(1 - B)(1 - B^{12})Z_t = 1/(1 - 0.6)(1 + 0.5B^{12})b_t$ .

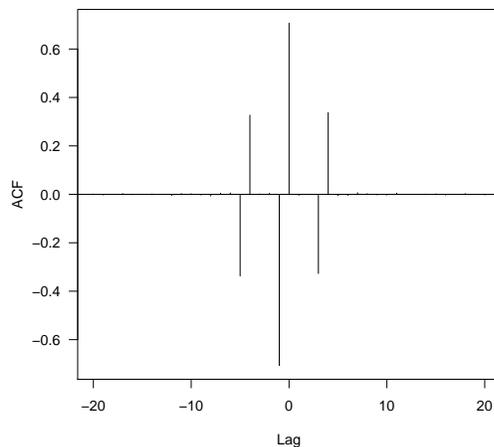


FIGURA 7: Modelo  $(1 - B)Z_t = -0.51 + b_t$ .

- iii) Para  $k < 0$ , el comportamiento no nulo de  $\hat{\rho}_{bZ}(k)$  señalará que el modelo preliminar no es adecuado.
- iv) Cuando para  $k < 0$ ,  $\hat{\rho}_{bZ}(k)$  es nula, la evidencia sobre el modelo incorrecto se presentará en el comportamiento  $\rho_{bZ}(k)$  para  $k \geq 0$ .

Claramente, cuando las series son no estacionarias homogéneas todos los resultados obtenidos se cumplen para los procesos apropiadamente diferenciados. El siguiente ejemplo ilustra el procedimiento de identificación.

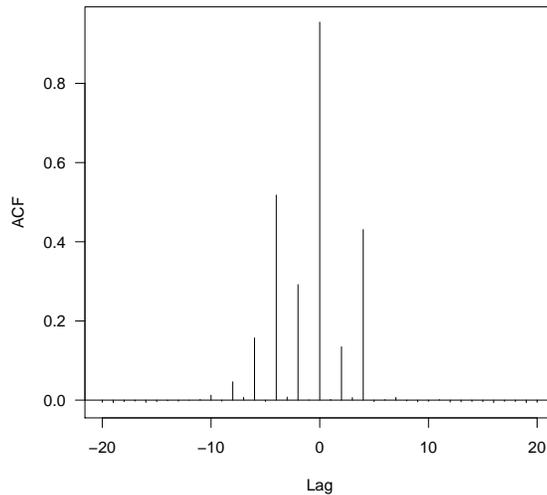


FIGURA 8: Modelo  $Z_t = (1 - 0.3B^2)b_t$ .

### 3. Empleo de la FCC en la identificación de series: un ejemplo real

Considere la serie mensual del total de pasajeros en vuelos internacionales (Serie G de Box & Jenkins 1976), estabilizada en varianza usando el logaritmo natural, cuya representación aparece en la figura 9.

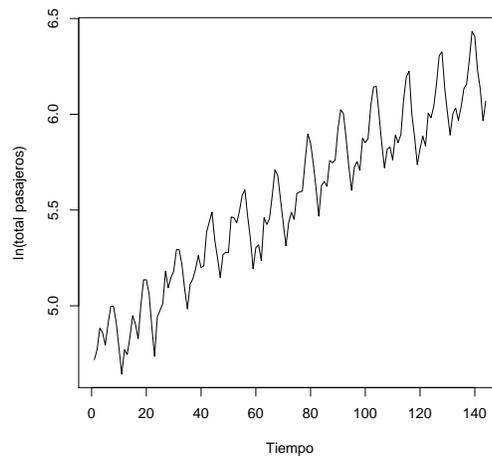


FIGURA 9: Logaritmo natural del total de pasajeros.

Como la serie presenta tendencia y estacionalidad, con periodo estacional  $S = 12$ , se necesita aplicar los operadores  $(1 - B^{12})$  o  $(1 - B)(1 - B^{12})$  para desestacionalizarla y eliminar su tendencia.

Partiendo de la aplicación del operador  $(1 - B)(1 - B^{12})$  a la serie y de un modelo estacional preliminar ARMA(1, 1)  $\times$  (1, 1) para la serie diferenciada de la forma

$$(1 - \phi_1 B)(1 - \Phi_{12} B^{12}) Y_t = (1 - \theta_1 B)(1 - \Theta_{12} B^{12}) a_t$$

donde  $Y_t$  es la serie de los logaritmos doblemente diferenciada, se obtuvo la estimación de los parámetros del modelo mostrados en la tabla 1, empleando el paquete estadístico SCA (Scientific Computing Associates Corp.)

TABLA 1: Estimación de modelo preliminar.

Parámetro	Estimación	Error Est.	$t$
$\theta_1$	0.4455	0.1762	2.52
$\Theta_{12}$	0.2147	0.1258	1.71
$\phi_1$	-0.0221	0.1954	-0.11
$\Phi_{12}$	-0.3439	0.1274	-2.70

Claramente, en la tabla 1 se observa que el modelo estimado parece no ser adecuado.

La FCCM entre  $Z_t$  y los residuales  $\hat{a}_t$  de este modelo se presenta en la figura 10.

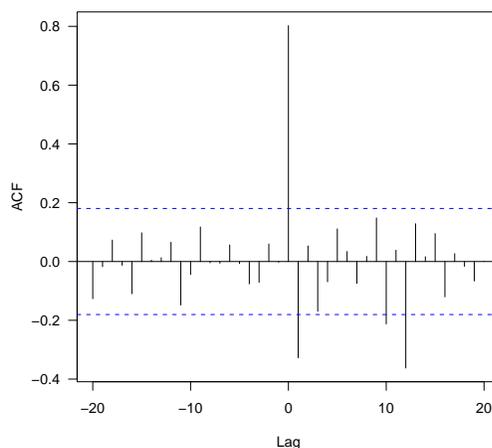


FIGURA 10: FCCM del modelo preliminar.

La figura 10 indica que la selección del modelo preliminar es incorrecta, pues no parecen existir componentes autorregresivos en el modelo (se manifestarían por medio de mezclas de decrecimientos exponenciales y/u ondas sinusoidales en la parte derecha de la función). Ahora bien, fácilmente se puede probar que la FCC entre  $a_t$  y  $Z_t$  para el modelo estacionario e invertible

$$Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12}) a_t \tag{10}$$

es

$$\begin{aligned}
 \rho_{a\Delta Z}(k) &= 0 && \text{si } k < 0 \\
 &= \frac{-\theta_1}{(1 + \theta_1^2 + \Theta_1^2 + \theta_1^2\Theta_1^2)} && \text{si } k = 0 \\
 &= 0 && \text{si } k = 1 \\
 &= \frac{-\Theta_1}{(1 + \theta_1^2 + \Theta_1^2 + -\theta_1^2\Theta_1^2)} && \text{si } 1 < k < 12 \\
 &= \frac{\theta_1\Theta_1}{(1 + \theta_1^2 + \Theta_1^2 + -\theta_1^2\Theta_1^2)} && \text{si } k = 13 \\
 &= 0 && \text{si } k > 13
 \end{aligned}$$

Este patrón teórico es consistente con el patrón exhibido en la figura 10, el cual señala la posibilidad de un modelo MA(1) para la parte no estacional de la serie y un MA(1) para la parte estacional. Este es el modelo que tradicionalmente se ha identificado para esta serie. La FCCM entre la serie y los residuos del modelo resultantes se muestra en la figura 11.

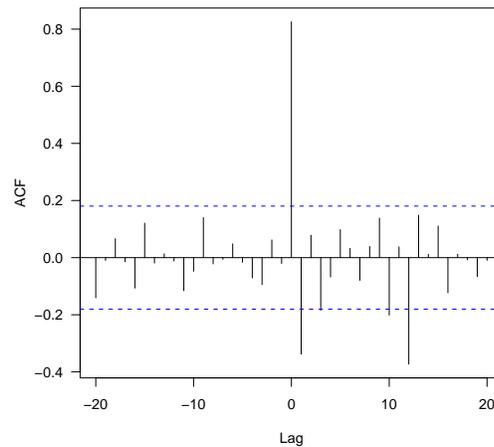


FIGURA 11: FCCM del modelo final.

Como antes, la parte derecha identifica el modelo estimado, y la parte izquierda parece ser nula, indicando que la especificación es correcta y consistente con el modelo (10).

La identificación incorrecta de los órdenes de diferenciación produce un correlograma cruzado para el modelo preliminar, como se observa en la figura 12.

En este caso, es clara la incorrecta especificación del modelo y la necesidad de diferenciar usando el operador completo  $(1 - B)(1 - B^{12})$ . Finalmente, es importante anotar que este diagnóstico de especificación puede ser obtenido fácilmente,

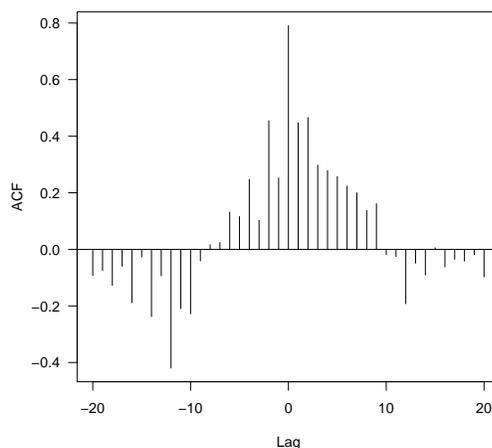


FIGURA 12: FCCM para la serie diferenciada solo con  $(1 - B^{12})$ .

pues la mayoría de programas para el análisis de series de tiempo disponen de una opción para calcular la FCCM.

## 4. Conclusiones

La función de correlación cruzada muestral (FCCM) entre una serie de tiempo y los residuos de un modelo propuesto para representar dicha serie constituye una herramienta de gran utilidad para verificar si este modelo puede considerarse apropiado para la serie en consideración.

Si los elementos del lado izquierdo del croscorelograma no muestran ninguna estructura y los elementos del lado derecho reproducen la estructura de autocorrelación de la serie, el modelo propuesto es correcto; en caso contrario, el lado derecho del autocorelograma contiene elementos de la estructura de autocorrelación del modelo correcto que permiten su identificación.

El procedimiento presentado en este artículo es de gran utilidad en el análisis de series de tiempo univariadas porque se puede iniciar con modelo muy aproximado, incluso errado, y se logra determinar si el modelo propuesto es correcto. En caso contrario se tienen nuevos elementos que permiten una identificación correcta del modelo buscado.

[Recibido: mayo de 2008 — Aceptado: octubre de 2008]

## Referencias

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **AC-19**, 716–723.

- Bartlett, M. S. (1985), *Stochastic Processes*, Cambridge University Press, Cambridge, USA.
- Beguín, J. M., Gouriéroux, C. & Monfort, A. (1980), *Identification of a Mixed Autoregressive-Moving Average Process: The Corner Method*, Time Series, Amsterdam, Netherlands.
- Box, G. E. P. & Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, USA.
- Castaño, E. (2005), La función de correlación cruzada en series no estacionarias: identificación, tendencias determinísticas y raíces unitarias, Tesis de maestría, Universidad Nacional de Colombia, Sede Medellín, Facultad de Ciencias, Escuela de Estadística.
- Rosales, L. F. (2004), La función de correlación cruzada como elemento de diagnóstico para los modelos ARMA(p,q), Trabajo de grado, Universidad Nacional de Colombia, Sede Medellín, Facultad de Ciencias, Escuela de Estadística.
- Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Ann. Statist* **6**(2), 461–464.
- Tsay, R. S. & Tiao, G. C. (1984), 'Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Non-stationary ARMA Models', *Journal of the American Statistical Association* **79**, 84–96.
- Tsay, R. S. & Tiao, G. C. (1985), 'Use of Canonical Analysis in Time Series Model Identification', *Biometrika* **72**, 299–315.
- Wei, W. W. S. (1990), *Time Series Analysis, Univariate and Multivariate Methods*, Addison-Wesley, California, USA.