

## Capítulo 16

# Construcción de modelos de regresión dinámica entre variables estacionarias

**Gwilym M. Jenkins** (1933- 1982)

Estadístico británico. Fue profesor en Imperial College, Londres y en la Universidad de Lancaster. Escribió con George Box el libro de series temporales más influyente de la segunda mitad del siglo XX, basado en sus investigaciones conjuntas sobre procesos ARIMA. Abandonó la Universidad de Lancaster en 1974 para formar su propia empresa y murió de cáncer a los 49 años.

### 16.1 Introducción

Un modelo de regresión dinámica es una representación de la relación entre dos variables que observamos a lo largo del tiempo. Los modelos de regresión clásicos se inventaron para representar la relación entre variables estáticas y el nombre de regresión proviene de los estudios que Galton realizó a finales del siglo XIX para investigar cómo se transmiten los caracteres genéticos entre generaciones. Galtón encontró que existe una relación lineal entre la estaturas de los hijos y las de sus padres y que si construimos la recta que proporciona las estaturas medias de los hijos dadas las estaturas de sus padres esta recta muestra lo que él denominó una regresión a la media: los padres altos tienen, en promedio, hijos altos y los padres bajos, en promedio, hijos bajos, pero los hijos de padres muy altos o muy bajos están más próximos a la media de la población que sus padres. Desde entonces se denominan ecuaciones de regresión a las ecuaciones que permiten prever el valor esperado de una variable cuando conocemos los valores de otras variables relacionadas con la que se desea prever. El modelo de regresión lineal simple relaciona la media de una variable dependiente,  $y$ , con una variable explicativa,  $x$ , mediante:

$$E(y|x_i) = \beta_0 + \beta_1 x_i. \quad (16.1)$$

Si suponemos que la distribución de la variable  $y$  condicionada a un valor fijo de la variable  $x$  tiene varianza constante, que no depende de  $x$ , el modelo anterior puede escribirse en función de las dos variables como:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (16.2)$$

donde las variables  $u_i$ , que son no observadas, se denominan las perturbaciones o innovaciones del modelo y tienen media cero, varianza constante y son independientes entre sí.

Al intentar aplicar este modelo a dos variables que son series temporales nos encontramos con dos problemas principales. En primer lugar, la ecuación (16.1) supone que la relación entre las dos variables es instantánea, cuando entre variables dinámicas la relación puede transmitirse con ciertos retardos. Por ejemplo, la relación puede no darse entre  $x_t$  e  $y_t$ , sino entre  $x_{t-k}$  y  $y_t$ , con un retardo  $k$  desconocido. En otros casos el valor de  $y_t$  puede depender no sólo del valor  $x_t$  sino también de los valores de  $x_{t-1}, \dots, x_{t-h}$ . El lector habrá seguramente advertido que si tratamos de aumentar la temperatura del agua en la ducha abriendo la

llave del agua caliente, la respuesta del sistema es gradual y la temperatura tarda unos segundos en alcanzar el nivel deseado. En este sistema, si llamamos  $y_t$ , a la temperatura del agua que sale de la ducha en el segundo  $t$  y  $x_t$ , a la apertura de la llave en ese instante, un aumento de  $x_t$  en  $t = t_0$  produce efectos en la respuesta  $y_t$  en los instantes  $t_0 + 1, t_0 + 2, \dots, t_0 + k$ , aumentando la temperatura gradualmente durante  $k$  segundos hasta un cierto nivel. Como consecuencia, la temperatura en  $t$  (la variable  $y_t$ ) depende no sólo de la apertura de la llave en  $t$ , variable  $x_t$ , sino de la apertura de la llave en los instantes anteriores, es decir, de las variables  $x_{t-1}, \dots, x_{t-k}$ . Como segundo ejemplo, si  $y_t$  representa los pasajeros transportados en un vuelo el día  $t$  y  $x_t$  el precio del billete en ese día, un aumento del precio hoy no tendrá efectos sobre los pasajeros transportados hoy, ya que todos o la gran mayoría de los pasajeros habrían comprado los billetes antes de la subida. Sin embargo, es esperable que como consecuencia de la subida se manifieste una disminución de la demanda del vuelo. En consecuencia la variable  $y_t$  dependerá no sólo de  $x_{t-1}$  sino también de los valores previos  $x_{t-2}, \dots, x_{t-h}$ .

En segundo lugar, la ecuación (16.2) supone que la parte de la variable respuesta no explicada por las variables independientes  $x$ , es un proceso  $u$  de variables independientes. Esta hipótesis es natural para datos estáticos, pero no suele verificarse con datos dinámicos. Si pensamos de nuevo en el ejemplo de la temperatura del agua en la ducha, la ecuación (16.1) indica que cuando  $x_t$  (apertura de la llave) permanece constante, la temperatura del agua debe variar aleatoriamente de forma imprevisible. Sin embargo, nuestra experiencia diaria nos indica que, afortunadamente, la temperatura evoluciona siempre con inercia, de manera que la temperatura en un instante,  $y_t$  será próxima a la temperatura en el instante anterior,  $y_{t-1}$ , y ésta a  $y_{t-2}$ , etc. De la misma forma, si el número de pasajeros transportados en un vuelo depende del día de la semana tendremos autocorrelación en las innovaciones.

En este capítulo vamos a estudiar modelos para relacionar variables a lo largo del tiempo que tienen en cuenta estos dos problemas. Comenzaremos en este capítulo suponiendo que las variables que queremos relacionar son estacionarias y en el siguiente generalizaremos estos resultados para variables no estacionarias. Estos modelos se conocen en Ingeniería como modelos de función de transferencia y en Economía, como modelos econométricos dinámicos. Describen cómo se transmiten los efectos entre variables y, por lo tanto, son un instrumento imprescindible en economía, ingeniería, biología, etc. para evaluar respuestas dinámicas. Si las variables explicativas son controlables, estos modelos permiten simular y evaluar políticas alternativas. Si no lo son, ofrecen la posibilidad de estudiar cómo ciertos «escenarios» (definidos por posibles evoluciones de la variable explicativa) afectan a la variable respuesta.

La utilidad de estos modelos para la predicción es distinta según sea o no instantánea la relación entre la variable explicativa y la respuesta. Cuando la relación es instantánea para prever  $y_t$  es necesario conocer  $x_t$ . Si esta variable es desconocida, será necesario construir un modelo univariante para  $x_t$ , y es posible que la mejora en la predicción que aporte el modelo dinámico respecto al univariante para  $y_t$  sea pequeña. En este caso el modelo es más adecuado para la predicción condicional de la respuesta en función de posibles valores de la variable explicativa. Si la relación no es instantánea y una variación en  $x_t$  afecta a  $y_{t+k}$  ( $k > 0$ ), diremos que  $x_t$  es un indicador avanzado de la respuesta, y su conocimiento mejorará las predicciones para  $y_t$ .

## 16.2 Relaciones entre dos series estacionarias: las funciones de covarianzas y correlaciones cruzadas

### 16.2.1 La función de covarianzas cruzadas

Supongamos que tenemos dos procesos estacionarios  $x_t, y_t$ , con medias  $\mu_x, \mu_y$ . Vamos a introducir medidas de su dependencia lineal para distintos retardos. Para ello introduciremos la *función de covarianzas cruzadas*, que se define mediante:

$$\gamma_{xy}(t, t+k) = E\{(x_t - \mu_x)(y_{t+k} - \mu_y)\} = \gamma_{yx}(t+k, t) \quad (16.3)$$

Observemos que esta función indica el orden en que se toman las variables y los instantes que corresponde a cada una de ellas. La primera variable, que es la que aparece en primer lugar en el subíndice,  $x$ , se supone

en  $t$  y la segunda,  $y$ , en  $t+k$ . Por tanto  $\gamma_{xy}(t, t+k)$  mide la relación lineal entre las dos variables  $x_t$  y  $y_{t+k}$ . El valor de  $k$  puede ser positivo o negativo.

Diremos que ambos procesos son *conjuntamente estacionarios* si cada uno de ellos es estacionario y las covarianzas cruzadas sólo dependen del retardo entre las variables,  $k$ , y no del instante inicial considerado. Entonces escribimos la función de covarianzas cruzadas como:

$$\gamma_{xy}(t, t+k) = \gamma_{xy}(k), \quad k = 0, \pm 1, \pm 2, \dots$$

La función de covarianzas cruzadas para procesos conjuntamente estacionarios tiene las propiedades siguientes:

1. Se verifica  $\gamma_{xy}(k) = \gamma_{xy}(-k)$ . En efecto:

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

y concluimos que la función  $\gamma_{xy}(k)$ , ( $k = 0, \pm 1, \dots$ ) resume toda la dependencia lineal, no siendo necesaria calcular la  $\gamma_{yx}$ .

2. Para  $k > 0$  los coeficientes  $\gamma_{xy}(k)$  representan como los valores de  $x_t$  influyen en los valores futuros de  $y_t$ , lo que interpretamos diciendo la parte positiva de esta función representa la relación «causal» de  $x$  hacia  $y$ .
3. Para  $k < 0$  los coeficientes  $\gamma_{xy}(k)$  representan como los valores de  $y_t$  influyen en los valores futuros de  $x_t$ , lo que interpretamos diciendo que la parte negativa de esta función representa la relación «causal» de  $y$  hacia  $x$ .

Diremos que  $x_t$  influye linealmente sobre  $y_t$  si existe algún valor  $\gamma_{xy}(k)$  no nulo para  $k > 0$ . Análogamente, diremos que  $y_t$  influye linealmente sobre  $x_t$  si existen valores no nulos en  $\gamma_{xy}(-k)$  para  $k > 0$ . Entre dos series podemos encontrar causalidad empírica de este tipo en una sola dirección o puede existir en las dos direcciones y entonces decimos que existe realimentación o una relación recíproca entre ambas series.

En adelante, para simplificar la exposición, diremos normalmente que dos variables son estacionarias para indicar que son conjuntamente estacionarias.

### Ejemplo 16.1

Supongamos los dos procesos estacionarios siguientes de media cero:

$$x_t = a_t$$

$$y_t = \phi y_{t-1} + a_t - \theta a_{t-1}$$

donde  $a_t$  esta incorrelada con  $y_{t-k}$  para cualquier  $k$  positivo. Vamos a calcular su función de covarianzas cruzadas.

Aplicando la definición de esta función como  $E(a_t y_{t-k}) = 0$ , para  $k > 0$ , y  $E(a_t a_{t-k}) = 0$  para  $k \neq 0$ , tenemos que:

$$\begin{aligned} \gamma_{xy}(0) &= E(x_t y_t) &= E\{a_t (\phi y_{t-1} + a_t - \theta a_{t-1})\} &= E\{a_t^2\} = \sigma^2 \\ \gamma_{xy}(1) &= E(x_t y_{t+1}) &= E\{a_t (\phi y_t + a_{t+1} - \theta a_t)\} &= \sigma^2 (\phi - \theta) \\ \gamma_{xy}(2) &= E(x_t y_{t+2}) &= E\{a_t (\phi y_{t+1} + a_{t+2} - \theta a_{t+1})\} &= \phi E\{a_t y_{t+1}\} = \\ & & & \phi E\{a_t (\phi y_t + a_{t+1} - \theta a_t)\} = \phi \sigma^2 (\phi - \theta) \\ \gamma_{xy}(k) &= E(x_t y_{t+k}) &= \phi E\{x_t y_{t+k-1}\} &= \phi \gamma_{xy}(k-1) \quad k \geq 2 \end{aligned}$$

La función de covarianzas cruzadas  $\gamma_{xy}(k)$  tiene todos los términos positivos para  $k > 0$ . Para  $k$  negativo:

$$\gamma_{xy}(-1) = \gamma_{xy}(1) = E\{x_t y_{t-1}\} = 0$$

$$\gamma_{xy}(-k) = 0$$

Concluimos que la función de covarianzas cruzadas es nula para  $k < 0$ , y, por tanto,  $x_t$  influye sobre  $y_t$ , pero no al contrario.

## 16.2.2 Función de autocorrelación cruzada

Definimos la función de correlación cruzada  $\rho_{xy}(k)$  de dos procesos estocásticos conjuntamente estacionarios  $x_t$ ,  $y_t$  por:

$$\boxed{\rho_{xy} = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y}} \quad (k = 0, \pm 1, \pm 2, \dots)$$

Esta función es la estandarización de  $\gamma_{xy}(k)$  y, en consecuencia, tiene propiedades análogas. Se verifica que

1.  $|\rho_{xy}(k)| \leq 1$ ;
2.  $\rho_{xy}(k) = \rho_{yx}(-k)$  y la función no es simétrica alrededor del origen. Para  $k > 0$  mide la relación desde  $x$  hacia  $y$ , mientras que para  $k < 0$  desde  $y$  hacia  $x$ .

Los modelos que se estudian en este capítulo suponen que esta función es nula para  $k < 0$ , es decir que existe relación unidireccional de la serie  $x$  hacia la serie  $y$  pero no al contrario.

### Ejemplo 16.2

Calcular la función de correlación cruzada para las series del ejemplo 16.1.

La varianza de  $x_t$  es directamente:

$$\text{Var}(x_t) = \text{Var}(a_t) = \sigma^2$$

y la varianza de un ARMA(1,1) se calculó en la sección 5.7 donde se obtuvo:

$$\gamma_y(0) = \sigma^2 \frac{1 + \theta^2 - 2\theta\phi}{1 - \phi^2}$$

Como  $\gamma_{xy}(k)$  es nula, para  $k < 0$  también lo será  $\rho_{xy}(k)$ . Para  $k > 0$ , tendremos que :

$$\begin{aligned} \rho_{xy}(0) &= \frac{\sqrt{1 - \phi^2}}{\sqrt{1 + \theta^2 - 2\phi\theta}} \\ \rho_{xy}(1) &= \frac{(\phi - \theta) \sqrt{1 - \phi^2}}{\sqrt{1 + \theta^2 - 2\phi\theta}} \\ \rho_{xy}(2) &= \phi \frac{(\phi - \theta) \sqrt{1 - \phi^2}}{\sqrt{1 + \theta^2 - 2\phi\theta}} = \phi \rho_{xy}(1) \\ \rho_{xy}(k) &= \phi \rho_{xy}(k-1) \quad k \geq 2 \end{aligned}$$

### 16.2.3 Dependencia lineal entre dos procesos y covarianzas cruzadas

En general la relación lineal entre dos procesos estocásticos estacionarios de media cero puede representarse como:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + n_t \quad (16.4)$$

donde los coeficientes  $v_i$  describen la relación dinámica entre las dos series y el proceso  $n_t$ , que es también estacionario, recoge el efecto de todas las otras variables que pueden tener efecto sobre  $y_t$ . Esta ecuación puede escribirse de forma más compacta utilizando el operador de retardo:

$$y_t = v(B)x_t + n_t \quad (16.5)$$

donde

$$v(B) = v_0 + v_1 B + v_2 B^2 + \dots$$

se denomina la función de transferencia y a los coeficientes  $v_i$  se les denomina función de respuesta a impulsos, como estudiamos en el capítulo 12,  $n_t$  es la perturbación que seguirá un proceso estacionario. La función de

transferencia puede en teoría tener un número indefinido de coeficientes y se parametriza, de manera similar a como hemos visto para modelos ARMA, mediante un cociente de polinomios. Supondremos que

$$v(B) = \frac{w_m(B)}{\delta_a(B)} B^b$$

donde

$$w_m(B) = w_0 + w_1 B + \dots + w_m B^m$$

es el numerador de la función de transferencia que juega un papel similar a la parte MA de un modelo ARMA, y

$$\delta_a(B) = 1 - \delta_1 B - \dots - \delta_a B^a$$

es la parte AR. El término  $B^b$  tiene en cuenta que la relación puede establecerse con un retardo inicial  $b$ . En la práctica los órdenes  $m$  y  $a$  son pequeños, no mayores que 3.

Es intuitivo que las covarianzas cruzadas entre ambas variables –que miden la dependencia lineal entre ellas– deben estar relacionadas con la función  $v(B)$ . Para encontrar esta relación, multiplicando por  $x_{t-k}$  la ecuación (16.4). Tomando esperanzas, se obtiene

$$\gamma_{xy}(k) = E(x_{t-k} y_t) = v_0 \gamma_{xx}(k) + v_1 \gamma_{xx}(k-1) + \dots$$

Esta expresión indica que las covarianzas cruzadas entre las dos variables contienen efectivamente los coeficientes de la relación entre ambas para todos los retardos, pero están «contaminadas» por la dinámica de la serie explicativa  $x_t$ . Sin embargo, si  $x_t$  fuese un proceso de ruido blanco, como entonces  $\gamma_{xx}(k) = 0$ ,  $\forall k \neq 0$ , se verificaría que:

$$\gamma_{xy}(k) = v_k \gamma_{xx}(0)$$

y podríamos conocer los coeficientes  $v$  a partir de las funciones de covarianzas mediante:

$$v_k = \frac{\gamma_{xy}(k)}{\gamma_{xx}(0)} \quad (16.6)$$

Podemos interpretar estos resultados observando que (16.4) es similar a una ecuación de regresión múltiple con regresores  $x_t, x_{t-1}, x_{t-2}, \dots$  y perturbación  $n_t$ . Cuando las variables explicativas están incorreladas, la estimación del coeficiente de cada variable mediante regresión múltiple es el mismo que en la regresión simple. Por tanto el coeficiente  $v_k$  de  $x_{t-k}$  en (16.4) debe ser el mismo que en la ecuación:

$$y_t = v_k x_{t-k} + n_t$$

donde  $v_k$  se estima por el cociente de la covarianza entre la variable dependiente y el regresor y la varianza del regresor. Este es el resultado (16.6), ya que si  $x_t$  es ruido blanco los regresores en (16.4) son ortogonales.

En conclusión, únicamente cuando la variable explicativa es ruido blanco, la covarianza cruzada de orden  $k$  entre las series es proporcional a la función de respuesta a impulsos de orden  $k$ . En otro caso la covarianza cruzada de orden  $k$  incluye información sobre la dependencia para todos los retardos y no solo para el  $k$ .

## 16.3 Estimación de la función de covarianzas cruzadas

Supongamos que tenemos dos series temporales estacionarias,  $x_t, y_t$  y queremos investigar su dependencia. Si suponemos que las series son realizaciones de tamaño  $T$  de dos procesos conjuntamente estacionarios,  $x_t, y_t$ , podemos estimar las covarianzas cruzadas entre ellos a partir de sus covarianzas cruzadas muestrales. Recordemos que en el capítulo 3 vimos que las autocovarianzas teóricas de un proceso  $z_t$  se estiman mediante las autocovarianzas muestrales:

$$\hat{\gamma}_z(k) = \frac{1}{T} \sum_{t=1}^{T-k} (z_t - \bar{z})(z_{t+k} - \bar{z})$$

De la misma forma, definimos las covarianzas cruzadas muestrales mediante:

$$\hat{\gamma}_{xy}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(y_{t+k} - \bar{y})$$

y las correlaciones cruzadas muestrales mediante:

$$r_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{s_x s_y}$$

donde  $s_x = \hat{\gamma}_x^{1/2}(0)$ ,  $s_y = \hat{\gamma}_y^{1/2}(0)$ . Las propiedades muestrales de las funciones de correlación cruzada son complicadas. Sin embargo, si uno de los procesos es *ruido blanco*, puede demostrarse –(véase Bartlett (1973))– que si *ambos procesos no están correlados entre sí*, las covarianzas entre los coeficientes de correlación muestras  $r_{xa}(k)$  y  $r_{xa}(k+h)$  pueden calcularse aproximadamente por:

$$\begin{aligned} \text{Cov}[r_{xa}(k)r_{xa}(k+h)] &\simeq \frac{\rho_x(h)}{(T-k)} \\ \text{Var}[r_{xa}(k)] &\simeq \frac{1}{T-k} \end{aligned} \quad (16.7)$$

Estas dos expresiones implican que:

$$\rho[r_{xa}(k)r_{xa}(k+h)] \simeq \rho_x(h) \quad (16.8)$$

Es decir, aunque  $x_t$  y  $a_t$  no estén correlados entre sí (por tanto,  $\rho_{xa}(k) \equiv 0$ ), los coeficientes estimados  $r_{xa}(k)$  si lo estarán; la función de correlación cruzada estimada variará alrededor de cero con varianza dada por (16.7) pero, según (16.8), al estar sus coeficientes relacionados de la misma forma que las autocorrelaciones de  $x_t$  tenderá a reproducir el correlograma de  $X_t$ .

Como hemos visto las funciones de covarianzas cruzadas son difíciles de interpretar cuando las series no son ruido blanco. Por estas razones la identificación de la relación entre las variables se realiza estimando los coeficientes de la función de transferencia como veremos a continuación y la función de covarianzas cruzadas sólo se utiliza para contrastar que los residuos del modelo estimado son ruido blanco, como veremos en la sección de Diagnóstico del modelo.

## 16.4 Estimación de los coeficientes de regresión

Dada una muestra de las variables  $(x_t, y_t)$  de tamaño  $T$ , vamos a ver cómo estimar los coeficientes de la relación dinámica entre las dos variables estacionarias,  $y_t$ ,  $x_t$  cuya relación va es en la dirección de  $x$  hacia  $y$  sin que exista realimentación en sentido contrario. Para simplificar la exposición vamos a suponer en esta sección que sólo el primer coeficiente  $v$  es distinto de cero en la relación (16.5), de manera que podemos escribir:

$$y_t = \beta_0 + v_1 x_t + n_t$$

Esta ecuación indica que hemos dividido todas las causas que influyen en la variable respuesta,  $y_t$  en dos grupos. El primero, es el efecto de la variable explicativa,  $x_t$ . El segundo, incluye el efecto de todas las demás variables que afectan a los valores de  $y_t$  que es la perturbación. Vamos a considerar primero que resultados se obtienen si suponemos incorrelación en las perturbaciones, para analizar después el caso general.

### 16.4.1 Estimación suponiendo perturbaciones incorreladas

Supongamos que admitimos que la perturbación es ruido blanco normal, es decir que  $n_t = u_t$ , donde las  $u_t$  son variables normales con  $E(u_t) = 0$ ,  $E(u_t u_{t-k}) = 0$  para  $k \neq 0$ , y  $E(u_t^2) = \sigma^2$ . Entonces el modelo es similar al de regresión clásico y sus parámetros,  $\beta_0, v_1$  y  $\sigma^2$ , pueden estimarse imponiendo normalidad de las

distribuciones condicionadas por máxima verosimilitud, que equivale a mínimos cuadrados. La condición de mínimos cuadrados es minimizar

$$\min \sum_{t=1}^T e_t^2 = \min \sum_{t=1}^T (y_t - (\hat{\beta}_0 + \hat{v}_1 x_t))^2$$

donde los residuos estimados se definen como  $e_t = y_t - (\hat{\beta}_0 + \hat{v}_1 x_t)$ . Derivando respecto al parámetro  $\beta_0$  se obtiene la ecuación:

$$\sum_{t=1}^T e_t = \sum_{t=1}^T (y_t - \hat{\beta}_0 - \hat{v}_1 x_t) = 0$$

que indica que los residuos, como las perturbaciones,  $u_t$ , que estiman, deben de tener media cero. Derivando respecto a  $v_1$  e igualando a cero:

$$\sum_{t=1}^T e_t x_t = \sum_{t=1}^T (y_t - \hat{\beta}_0 - \hat{v}_1 x_t) x_t = 0$$

ecuación que indica que los residuos deben estar incorrelados con la variable explicativa  $x_t$ . Estas dos ecuaciones conducen a la recta estimada

$$\hat{y}_t = \hat{\beta}_0 + \hat{v}_1 x_t,$$

con coeficientes estimados:

$$\hat{\beta}_0 = \bar{y} - \hat{v}_1 \bar{x},$$

donde  $\bar{y} = \sum y_t / T$  y  $\bar{x} = \sum x_t / T$  son las medias muestrales de las variables y :

$$\hat{v}_1 = \frac{\hat{\gamma}_{xy}(0)}{s_x^2} \quad (16.9)$$

donde  $\hat{\gamma}_{xy}(0) = \sum_{t=1}^T (y_t - \bar{y})(x_t - \bar{x}) / T$  es la covarianza cruzada muestral de orden cero definida en la sección anterior y  $s_x^2 = \hat{\gamma}_x(0) = \sum_{t=1}^T (x_t - \bar{x})^2 / T$ , la varianza de la variable  $x_t$ . La interpretación de la primera condición es que la recta estimada debe pasar por el centro de los datos y la de la segunda, que la pendiente de la recta es la covarianza estandarizada. Sustituyendo la expresión de  $\hat{\beta}_0$ , la ecuación de regresión estimada puede escribirse como

$$\hat{y}_t = \bar{y} + \hat{v}_1 (x_t - \bar{x})$$

que indica que las desviaciones de la respuesta a su media son proporcionales a las desviaciones de la variable explicativa a la suya. En la hipótesis de que el proceso  $u_t$  es de ruido blanco, los estimadores obtenidos son centrados de varianza mínima. Se demuestra que la varianza de la estimación de la pendiente es

$$\text{Var}(\hat{v}_1) = \frac{\sigma^2}{T s_x^2} \quad (16.10)$$

Para realizar un contraste de que la relación encontrada es significativa necesitamos un estimador de  $\sigma^2$ . El estimador centrado habitual es

$$\hat{s}_R^2 = \frac{\sum e_t^2}{T - 2}$$

y para contrastar que hay relación lineal entre las dos variables en la población, es decir, que el parámetro  $v_1$  es distinto de cero, se compara su estimador con su desviación típica estimada, utilizando en lugar de  $\sigma^2$  el estimador  $\hat{s}_R^2$ . En la hipótesis de que este parámetro es cero en la población, de manera que no existe relación entre las variables, el estadístico:

$$t = \frac{\hat{v}_1 s_x \sqrt{T}}{\hat{s}_R} \quad (16.11)$$

se distribuirá por la normalidad como una t de Student con n-2 grados de libertad. Rechazaremos que no existe relación cuando este estadístico este significativamente alejado de cero.

### 16.4.2 Consecuencias de la autocorrelación en la perturbación

El análisis anterior parte de la hipótesis central de que el proceso  $u_t$  es un proceso de ruido blanco, lo que no es cierto en general cuando relacionamos series temporales. Supongamos que la variable  $u_t$  tiene autocorrelación. Entonces, los estimadores del modelo de regresión clásico dejan de ser eficientes y la varianza del estimador de la pendiente (16.9) puede ser mucho mayor que la calculada por la fórmula (16.10). Se demuestra en el apéndice 16.1 que cuando relacionamos variables estacionarias y la perturbación tiene estructura dinámica, la varianza del estimador de mínimos cuadrados (16.9) es

$$var(\hat{v}_1) = E(var(\hat{v}_1^{IND})) \left[ 1 + 2 \sum_{t=1}^{T-1} (T-t) \rho_u(t) E(r_{xx}(t)) \right] \quad (16.12)$$

donde  $var(\hat{v}_1)$  es la varianza del estimador de la pendiente (16.9) cuando hay autocorrelación,  $var(\hat{v}_1^{IND})$  la expresión de la varianza (16.10) calculada suponiendo que la perturbación del modelo es ruido blanco,  $\rho_u(t)$  es el coeficiente de autocorrelación de orden  $t$  de los residuos y  $r_{xx}(t)$  el coeficiente de autocorrelación de orden  $t$  de la variable  $x$ . Esta ecuación indica que podemos encontrarnos en las tres situaciones siguientes:

(1) La variable  $x_t$  es ruido blanco incorrelado con la perturbación  $u_t$ . Entonces, sustituyendo en (16.12)  $E(r_{xx}(t))$  por cero tenemos que  $var(\hat{v}_1) = E(var(\hat{v}_1^{IND}))$  y el estimador tradicional de regresión es eficiente. Observemos que este resultado no depende de que  $u_t$  tenga o no autocorrelación y está garantizado por la falta de dinámica en la variable explicativa.

(2) La variable  $x_t$  tiene estructura dinámica, pero el ruido  $u_t$  es ruido blanco. De nuevo sustituyendo en (16.12)  $\rho_u(t)$  por cero  $var(\hat{v}_1) = E(var(\hat{v}_1^{IND}))$  y el estimador es eficiente.

(3) Tanto la variable  $x_t$  como el ruido tienen autocorrelación positiva. Entonces, la varianza del estimador de la pendiente puede ser mucho mayor que la calculada con la fórmula (16.10) bajo la hipótesis de independencia.

La implicación de este resultado es que cuando tanto la variable explicativa,  $x_t$ , como la perturbación,  $u_t$  tienen autocorrelación, podemos fácilmente encontrar relación significativa entre variables independientes. En efecto, el estimador (16.9) tendrá una varianza que puede ser mucho mayor que la dada por la fórmula (16.10), y al contrastar la presencia de relación entre las dos variables mediante el estadístico  $t$  tradicional con (16.11), como la varianza está muy subestimada, podemos concluir fácilmente que existe relación entre dos variables que en realidad son independientes.

Como ilustración la figura 16.1 muestra la distribución del estadístico  $t$  (16.11) obtenida simulando 2000 procesos independientes AR(1) con el mismo valor del parámetro y tamaño muestral igual a 100. Se observa que cuando el parámetro es .7 el estadístico  $t$  será mayor que 2 en valor absoluto aproximadamente el 25% de las veces, y en estos casos concluiremos que existe relación entre las variables, aunque estas sean independientes. Si el parámetro es .95 este porcentaje de error aumenta al 63%, es decir, es bastante probable que encontremos relación entre dos variables independientes.

Estos resultados sugieren que como no podemos saber a priori si existe o no autocorrelación en la perturbación, siempre que la variable explicativa,  $x_t$ , tenga dinámica y no sea un proceso de ruido blanco, debemos evitar estos problemas permitiendo que la perturbación pueda tener estructura dinámica. Vamos a indicar cómo realizar entonces la estimación del modelo.

### 16.4.3 Mínimos cuadrados generalizados

Un procedimiento teórico de resolver el problema si conocemos la estructura de autocorrelación de las perturbaciones es transformar el problema de regresión de manera que la perturbación no tenga autocorrelación y aplicar después las formulas tradicionales al modelo transformado. Esto es lo que se conoce como utilizar mínimos cuadrados generalizados. Desgraciadamente, es raro que conozcamos la estructura de autocorrelación de las perturbaciones, por lo que este procedimiento no es aplicable directamente. Sin embargo, si suponemos un modelo ARMA inicial para las perturbaciones este procedimiento podría aplicarse iterativamente como explicamos a continuación.

Vamos a suponer inicialmente que conocemos la estructura de autocorrelación en la perturbación que es equivalente a suponer que conocemos el modelo que sigue. Supongamos el caso más simple donde el modelo



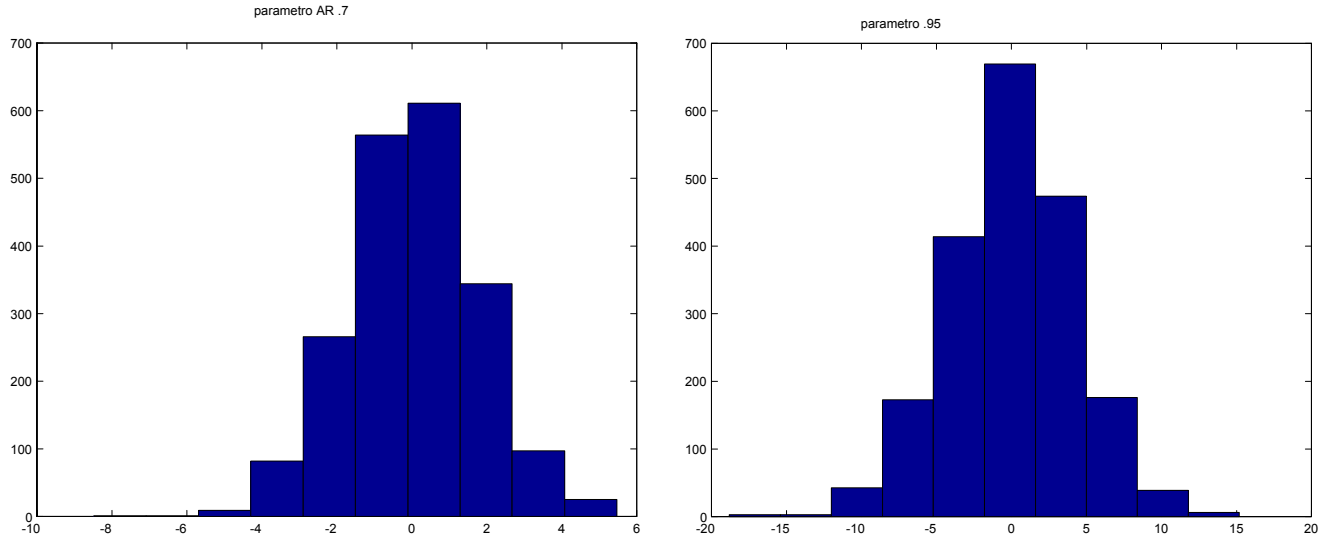


Figura 16.1: Distribución del estadístico  $t$  obtenido al regresar dos procesos estacionarios AR independientes. En el primer caso el coeficiente de ambos es .7 y en el segundo .95

es

$$y_t = \beta_0 + v_1 x_t + n_t$$

y la perturbación sigue el proceso AR(1)

$$n_t = \rho n_{t-1} + a_t$$

donde  $a_t$  es ruido blanco. Multiplicando la ecuación del modelo por el operador  $(1 - \rho B)$ , tenemos que

$$(1 - \rho B)y_t = (1 - \rho)\beta_0 + v_1(1 - \rho B)x_t + a_t$$

y definiendo las nuevas variables  $\tilde{y}_t = y_t - \rho y_{t-1}$ ,  $\tilde{x}_t = x_t - \rho x_{t-1}$ , podemos aplicar mínimos cuadrados a la regresión

$$\tilde{y}_t = c + v_1 \tilde{x}_t + a_t$$

entre  $\tilde{y}_t$  y  $\tilde{x}_t$  y obtener  $v_1$ . Esto equivale a aplicar la transformación

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \dots \\ \tilde{y}_T \end{bmatrix} = \begin{bmatrix} 1 & 0.. & 0 & 0 \\ -\rho & 1.. & 0 & 0 \\ \dots & 0.. & 1 & 0 \\ 0 & 0.. & -\rho & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{bmatrix}$$

que, como se justifica en el apéndice 16.2, es aproximadamente el resultado de aplicar el método general de mínimos cuadrados generalizados. De esta manera obtenemos en la pendiente el coeficiente  $v_1$  y

Si no conocemos el modelo el procedimiento anterior podría aplicarse iterativamente como sigue:

(1) Comenzamos suponiendo que  $n_t$  es ruido blanco y estimamos los parámetros de la ecuación por mínimos cuadrados. Con estos valores obtenemos la perturbación estimada,  $\hat{n}_t = y_t - \hat{\beta}_0 + \hat{v}_1 x_t$

(2) Se estima un modelo para la serie de perturbaciones estimada, sea  $\hat{n}_t = (1 - \hat{\rho}B)a_t$ .

(3) Se transforma la ecuación para que tenga residuos incorrelados mediante  $\tilde{y}_t = y_t - \hat{\rho}y_{t-1}$ ,  $\tilde{x}_t = x_t - \hat{\rho}x_{t-1}$ , y se realiza una regresión entre estas dos variables para obtener nuevos coeficientes.

(4) Se iteran las etapas (2) y (3) hasta obtener convergencia, de manera que los parámetros no cambien de una iteración a la siguiente.

## 16.5 Construcción de modelos de regresión dinámica

La construcción de un modelo de regresión dinámica entre dos variables estacionarias se realiza en las etapas de identificación, estimación y diagnóstico como se explica a continuación.

Como no podemos saber a priori si la perturbación está o no incorrelada, siempre que la variable  $x_t$  tenga estructura de autocorrelación conviene formular el modelo suponiendo perturbaciones autocorreladas. En efecto, si no existiese autocorrelación en la perturbación y la variable  $x_t$  fuese capaz de identificar toda la dinámica de la  $y_t$  esperamos detectarlo en la estimación con coeficientes no significativos en el modelo del ruido. Sin embargo, si existe autocorrelación y no la tenemos en cuenta los estimadores de los coeficientes  $v_i$  serán muy poco precisos y con desviaciones típicas estimadas erróneas, con lo que podemos fácilmente llegar a conclusiones equivocadas respecto a la relación entre las variables.

Para concretar supongamos que la variable  $y_t$  sigue un proceso ARMA  $(p_y, q_y)$  y que la variable  $x_t$  siga también un proceso ARMA  $(p_x, q_x)$ . Entonces, formularemos el modelo de regresión dinámica

$$y_t = \beta_0 + \frac{w_m(B)}{\delta_a(B)} B^b x_t + n_t \quad (16.13)$$

donde la perturbación,  $n_t$ , sigue proceso:

$$\phi(B)n_t = \theta(B)a_t \quad (16.14)$$

Las ecuaciones (16.13) y (16.19) definen el modelo dinámico.

### 16.5.1 Identificación

La identificación del modelo supone decidir los órdenes de los operadores  $m, a, b$  de la función de transferencia y  $p, q$  de la perturbación. En la práctica lo más simple es suponer que el modelo de la perturbación es el mismo que el de variable  $y_t$ . La justificación de esta elección es que si no hay relación entre las variables y los coeficientes de la función de transferencia son cero entonces  $y_t$  seguirá el proceso de la perturbación. En general, el modelo final para  $n_t$  será más simple que el de  $y_t$ , ya que es muy posible que  $x_t$  explique parte de la dinámica de la respuesta, pero esta formulación inicial nos asegura que recogemos toda la posible dinámica de la variable respuesta  $y_t$ .

La estimación de la función de transferencia se realiza estimado el modelo:

$$y_t = \sum v_i x_{t-i} + n_t \quad (16.15)$$

y estudiando la estructura de los coeficientes  $v_i$ . No conviene seguir reglas estrictas de si los coeficientes estimados son significativos o no, ya que normalmente existirá alta multicolinealidad en la estimación y los errores estándar de los coeficientes serán altos. Conviene considerar todos los coeficientes con valores de  $t$  mayores que uno y buscar la estructura que tienen, de forma similar a como se hizo con el correlograma o fas de un proceso ARMA. En concreto:

- Si los primeros  $b$  coeficientes  $\hat{v}_i$  parecen iguales a cero entonces la función empieza en el retardo  $b$ .
- Si hay  $m$  coeficientes aparentemente significativos sin una pauta característica de variación, tomar ese grado para el numerador de la función.
- Si se observa una estructura de decrecimiento geométrico en los coeficientes, incluir un término de orden uno en el denominador,  $a = 1$ . Si el decrecimiento es sinusoidal suponer  $a = 2$ .

De esta manera se establece la función

$$\frac{w_m(B)}{\delta_a(B)} B^b$$

Una alternativa propuesta originalmente por Box y Jenkins es el preblanqueo que se expone en el apéndice 16.3 No se utiliza mucho actualmente porque es más fácil estimar directamente y no puede extenderse fácilmente con muchas variables



Una vez obtenido  $\hat{y}_t^*$  el proceso de inercia  $n_t$ , se calcula por:

$$\hat{n}_t = \hat{y}_t - \hat{y}_t^*$$

y si  $\hat{n}_t$  sigue un proceso ARMA(p, q),  $\hat{a}_t$ , se obtendrá con

$$\hat{a}_t = \hat{\theta}_1 \hat{a}_{t-1} + \dots + \hat{\theta}_q \hat{a}_{t-q} + \hat{n}_t - \hat{\phi}_1 \hat{n}_{t-1} - \dots - \hat{\phi}_p \hat{n}_{t-p}$$

4. Sea  $\underline{\beta}' = (\underline{w}' \underline{\delta}' \underline{\theta}' \underline{\phi}')$  el vector de parámetros del modelo y  $\underline{\beta}_0$  el valor inicial, utilizado en los cálculos anteriores. Sea

$$S(\underline{\beta}) = \sum \hat{a}_t^2(\underline{\beta})$$

y  $S(\underline{\beta}_0)$  el valor inicial. Calcularemos otro valor  $\underline{\beta}_1$  que disminuya la suma de cuadrados  $S(\underline{\beta}_1)$  mediante un algoritmo tipo Gauss-Newton e iteraremos hasta obtener convergencia.

Como en el caso univariante las varianzas y covarianzas de los estimadores se obtendrán a partir de la matriz de segundas derivadas en el máximo.

La maximización de la función de verosimilitud exacta se efectúa con estos mismos principios. Normalmente se estiman varios modelos modificando los parámetros y se elige el más adecuado con el criterio BIC.

### 16.5.3 Contrastes Diagnósticos

Una vez seleccionado un modelo entre los estimados antes de aceptarlo conviene efectuar contrastes diagnósticos para comprobar que no presenta deficiencias detectables que pueden sugerir reformulaciones. Clasificaremos los contrastes que pueden aplicarse en: (a) sobre los parámetros estimados, y (b) sobre los residuos.

#### Contrastes sobre los parámetros

Debe comprobarse, en primer lugar, que no existen correlaciones altas entre los parámetros estimados, lo que sugeriría una mala definición de la situación de estimación y en segundo, que cada uno de ellos es significativamente distinto de cero, comparando el valor obtenido con la desviación estándar estimada. Hay que recordar que estos contrastes son sólo válidos asintóticamente, por lo que conviene fijar el nivel de significación en función de la interpretación lógica del modelo, especialmente para los parámetros de la función de transferencia: si su signo y magnitud coinciden con lo esperado conviene mantener parámetros aunque su estadístico  $t(\hat{\beta}/\hat{s}(\hat{\beta}))$  sea tan sólo de uno.

Una vez definida la estructura del modelo debe comprobarse si su ganancia concuerda con el comportamiento esperado del sistema. Conviene también examinar:

1. Si el modelo puede simplificarse eliminando operadores con valores próximos en el numerador y en el denominador.
2. Si las raíces de los polinomios autorregresivos cumplen las condiciones de estabilidad o estacionariedad adecuada.
3. Si los efectos de los distintos operadores del ruido y de la función de transferencia ofrecen una interpretación coherente con el comportamiento cualitativo del sistema que estamos modelando.

#### Contrastes sobre los residuos

Consisten, como siempre, en comprobar si los residuos son una secuencia de variables aleatorias normales con varianza constante e independientes. Esto supone realizar contrastes para confirmar:

1. Normalidad.
2. Que su media es cero.

3. Que están incorrelados entre sí.

Estos contrastes son análogos al caso univariante, y se reducen a comprobar si la serie  $a_t$  puede considerarse ruido blanco.

Para diseñar contrastes de errores de especificación, veamos las consecuencias de una especificación incorrecta. Supongamos que el modelo verdadero es:

$$y_t = b(B) x_t + \psi(B) a_t^*$$

pero erróneamente estimamos:

$$y_t = \hat{v}_1(B) x_t + \hat{\psi}_1(B) \hat{a}_t$$

Entonces:

$$v(B)x_t + \psi(B)a_t^* = \hat{v}_1(B)x_t + \hat{\psi}_1(B)\hat{a}_t$$

y los residuos estimados  $\hat{a}_t$  verificarán:

$$\hat{a}_t = \frac{v(B) - \hat{v}_1(B)}{\hat{\psi}_1(B)} x_t + \frac{\psi(B)}{\hat{\psi}_1(B)} a_t^*$$

Esta expresión indica que podemos encontrarnos en los casos siguientes:

1. Si  $\psi(B) \neq \hat{\psi}_1(B)$  y  $v(B) \neq \hat{v}_1(B)$  los residuos  $\hat{a}_t$  estarán autocorrelados entre sí y correlados con los valores de  $x_t$ . Este efecto se detectará estudiando la función de correlación cruzada entre los residuos del modelo y la variable y forman explicativa.
2. Si la función de transferencia es correcta, aproximadamente  $v(B) = \hat{v}_1(B)$ , pero el modelo del proceso de inercia es equivocado, habrá autocorrelación en  $\hat{a}_t$ , pero no existirá correlación cruzada entre  $\hat{a}_t$  y  $x_t$ .
3. Si la función de transferencia es incorrecta, aunque el modelo del ruido sea adecuado, también observaremos correlación cruzada entre los residuos y  $x_t$  y autocorrelación entre los residuos. En efecto, si la función de transferencia no es correcta y si esta bien especificado el modelo del ruido, los residuos pueden escribirse como

$$\hat{a}_t = \frac{v(B) - \hat{v}_1(B)}{\hat{\psi}_1(B)} x_t + a_t^*$$

y si la  $x_t$  no es ruido blanco esta ecuación implica autocorrelación entre los residuos estimados  $\hat{a}_t$ .

En resumen, la función de correlación cruzada entre los residuos estimados  $\hat{a}_t$  y la variable explicativa  $x_t$  es muy útil para detectar posibles errores en el modelo. Si encontramos autocorrelación en los residuos estimados pero estos están incorrelados con la variable  $x_t$  esto sugiere que el modelo del ruido no es adecuado y debe reformularse, pero la función de transferencia es correcta. Si encontramos correlaciones cruzadas entre los residuos estimados y la  $x_t$  esto sugiere que la función de transferencia no está bien especificada. Además, en este caso es posible que necesitemos también reformular el modelo del ruido.

Dado que la función de correlación entre  $\hat{a}_t$  y  $x_t$  puede no ser muy clara debido a la propia estructura de dependencia de  $x_t$ , como ya comentamos en el caso general, es más operativo estudiar la correlación entre  $\hat{a}_t$  y  $a_t$ , siendo  $a_t$  los residuos del modelo univariante de  $x_t$ , ya que, al ser las  $a_t$  ortogonales, la imagen será mucho más clara.

## 16.6 El modelo con varias variables explicativas

La metodología que hemos descrito en las páginas anteriores se generaliza sin dificultad para varias series explicativas. El modelo será

$$y_t = \frac{w_1(B)B^{b_1}}{\delta_1(B)}x_{1t} + \dots + \frac{w_k(B)B^{b_k}}{\delta_k(B)}x_{kt} + n_t \quad (16.16)$$

La identificación del modelo se realiza como en el caso anterior. Se toma como modelo inicial para la perturbación el modelo de la respuesta y se estima un modelo donde se supone una estructura simple inicial para la respuesta. Una alternativa simple es suponer inicialmente que los numeradores y los denominadores son de orden uno, es decir del tipo

$$\frac{w_0 + w_1B}{1 - \delta B}$$

y en función de los coeficientes encontrados significativos ir reformulando hasta obtener la estructura adecuada. La estimación y la diagnosis se realiza con los mismos principios.

### Apéndice 16.1 Varianza de la estimación de mínimos cuadrados con autocorrelación

El estimador de la pendiente es

$$\hat{v}_1 = \sum \frac{(x_t - \bar{x})(y_t - \bar{y})}{Ts_x^2}$$

y sustituyendo en esta ecuación  $y_t = \bar{y} + v_1(x_t - \bar{x}) + u_t$ , se obtiene que

$$\hat{v}_1 = v_1 + \sum \frac{(x_t - \bar{x})u_t}{Ts_x^2}$$

Supongamos que  $u_t$  es independiente de  $x_t$  y tiene esperanza cero. Tomando esperanzas en esta expresión comprobamos que  $E(\hat{v}_1) = v_1$  y el estimador es centrado. Para calcular su varianza escribamos:

$$(\hat{v}_1 - v_1)^2 = \left( \sum \frac{(x_t - \bar{x})u_t}{Ts_x^2} \right)^2$$

Para calcular la varianza vamos a tomar primero esperanzas en esta ecuación condicionado a  $Ts_x^2$ , es decir, vamos a promediar sobre todas las secuencias que conducen al mismo valor de  $Ts_x^2$ . Entonces:

$$\begin{aligned} E((\hat{v}_1 - v_1)^2 | Ts_x^2) &= \frac{1}{Ts_x^2} E \left( \sum_{t=1}^T \frac{(x_t - \bar{x})^2}{Ts_x^2} u_t^2 + 2 \sum_{t=1}^{T-1} \frac{(x_t - \bar{x})(x_{t+1} - \bar{x})}{Ts_x^2} u_t u_{t+1} + \dots \right. \\ &\quad \left. 2 \sum_{t=1}^{T-k} \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2} u_t u_{t+k} + \dots + 2 \frac{(x_1 - \bar{x})(x_T - \bar{x})}{Ts_x^2} u_1 u_T \right) \end{aligned}$$

Esta ecuación indica que esta esperanza condicionada depende de toda la dinámica de ambas variables. Llamando  $r_{xx}(k) = \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) / Ts_x^2$  a los coeficientes de autocorrelación muestral de la variable  $x_t$  y  $\rho_u(k)$  a los coeficientes de autocorrelación teóricos del proceso  $u_t$ , como:

$$E \left[ \sum_{t=1}^{T-k} \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2} u_t u_{t+k} \right] = \sum_{t=1}^{T-k} E \left[ \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2} \right] E(u_t u_{t+k}) = \sigma_u^2 \rho_u(k) E(r_{xx}(k))$$

Tenemos que

$$E((\hat{v}_1 - v_1)^2 | Ts_x^2) = \frac{\sigma_u^2}{Ts_x^2} [1 + 2(T-1)\rho_u(1)E(r_{xx}(1)) + \dots + 2\rho_u(T-1)E(r_{xx}(T-1))]$$

y llamando

$$var(\hat{v}_1^{IND}) = \frac{\sigma_u^2}{Ts_x^2}$$

a la varianza bajo independencia, tenemos que, tomando esperanzas ahora respecto a  $Ts_x^2$  y utilizando que  $E(y) = E_x E_{y|x}(y|x)$ , obtenemos finalmente que

$$\text{var}(\hat{v}_1) = E(\text{var}(\hat{v}_1^{IND})) [1 + 2(T-1)\rho_u(1)E(r_{xx}(1)) + \dots + 2\rho_u(T-1)E(r_{xx}(T-1))]$$

Para interpretar esta ecuación consideremos los casos siguientes:

(1) La variable  $x_t$  es ruido blanco, pero el ruido  $u_t$  tiene autocorrelación. Entonces,  $\text{var}(\hat{v}_1) = E(\text{var}(\hat{v}_1^{IND}))$  y el estimador es eficiente.

(2) La variable  $x_t$  tiene estructura dinámica pero el ruido  $u_t$  es ruido blanco. De nuevo  $\text{var}(\hat{v}_1) = E(\text{var}(\hat{v}_1^{IND}))$  y el estimador es eficiente.

(3) Tanto la variable  $x_t$  como el ruido tienen autocorrelación positiva. Entonces la varianza del estimador de la pendiente puede ser mucho mayor que la calculada en promedio bajo la hipótesis de independencia.

Por ejemplo, supongamos para simplificar que ambos procesos siguen procesos AR(1) independientes:

$$x_t = \phi x_{t-1} + \epsilon_t$$

y

$$u_t = \rho u_{t-1} + \zeta_t.$$

Entonces  $\rho_u(k) = \rho^k$  y tomando aproximadamente  $E(r_x(k)) \approx \phi^k$  tenemos que

$$\text{var}(\hat{v}_1) = E(\text{var}(\hat{v}_1^{IND})) \left[ 1 + 2 \sum_{k=1}^{T-1} (T-k) \rho^k \phi^k \right]$$

Los términos del sumatorio son una suma de las progresiones geométricas  $a + a^2 + \dots + a^j$  con  $a = \rho\phi$  y puede comprobarse que

$$\sum_{k=1}^{T-1} (T-k)a^k = \frac{1}{(1-a)^2} [a(1-a) - (1-a^T)]$$

y despreciando para simplificar el término  $\rho^T \phi^T$ , resulta que

$$\text{var}(\hat{v}_1) = E(\text{var}(\hat{v}_1^{IND})) \left[ \frac{1 + \rho\phi}{1 - \rho\phi} \right]$$

Por ejemplo, si  $\rho = \phi = .9$  la varianza real del estimador de mínimos cuadrados es 9.5 veces la calculada suponiendo independencia. El estimador es muy ineficiente y está sesgado a encontrar relaciones aunque no existan.

### Apéndice 16.2 Mínimos cuadrados generalizados

Para explicar el procedimiento escribiremos el modelo de regresión en su forma general

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}$$

donde  $\mathbf{Y}$  es un vector de dimensión  $T$  de la variable dependiente, y  $\mathbf{X}$  es una matriz  $T \times p$  de variables explicativas y  $\boldsymbol{\beta}$  es un vector de parámetros. La primera columna de  $\mathbf{X}$  contiene un vector de unos y las siguientes los valores de las variables dependientes, que incluyen valores actuales y retardados de cada una de las variables. Si las perturbaciones de esta ecuación estuviesen incorreladas el estimador mínimo cuadrático de esta ecuación es

$$\hat{\boldsymbol{\beta}}_{MC} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Sin embargo, con series temporales es esperable que las perturbaciones,  $\mathbf{N}$ , no sean ruido blanco y tengan una estructura de dependencia definida por su matriz de varianzas covarianzas,

$$E[\mathbf{N}\mathbf{N}'] = \sigma^2 \mathbf{G} = \sigma^2 \begin{bmatrix} 1 & \varrho_1 & \varrho_2 & \cdots & \varrho_{T-1} \\ \varrho_1 & 1 & \varrho_1 & & \varrho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho_{T-1} & \varrho_{T-2} & \varrho_{T-3} & \cdots & 1 \end{bmatrix}$$

Supongamos que la matriz  $\mathbf{G}$  fuese conocida. Entonces podríamos transformar el problema en otro con nuevas perturbaciones incorreladas y aplicar después mínimos cuadrados a la ecuación transformada. Como toda matriz definida positiva admite una representación como  $\mathbf{G} = \mathbf{L}\mathbf{L}'$  donde la matriz  $\mathbf{L}$  es triangular, que se conoce como descomposición de Cholesky, podemos transformar el modelo mediante

$$\mathbf{L}^{-1}\mathbf{Y} = \mathbf{L}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{L}^{-1}\mathbf{N}$$

y llamando  $\tilde{\mathbf{Y}} = \mathbf{L}^{-1}\mathbf{Y}$  y  $\tilde{\mathbf{X}} = \mathbf{L}^{-1}\mathbf{X}$  a las variables transformadas y  $\mathbf{E} = \mathbf{L}^{-1}\mathbf{N}$  al nuevo vector de perturbaciones que tendrán matriz de varianzas covarianzas

$$E[\mathbf{E}\mathbf{E}'] = E[\mathbf{L}^{-1}\mathbf{N}\mathbf{N}\mathbf{L}^{-1'}] = \mathbf{L}^{-1}E[\mathbf{N}\mathbf{N}']\mathbf{L}^{-1'} = \sigma^2\mathbf{I}$$

y tendrá perturbaciones incorreladas. Aplicando mínimos cuadrados a las variables  $\tilde{\mathbf{Y}}$ ,  $\tilde{\mathbf{X}}$ , el estimador obtenido para  $\boldsymbol{\beta}$ , será:

$$\hat{\boldsymbol{\beta}}_G = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}$$

y sustituyendo las variables transformadas por las originales:

$$\hat{\boldsymbol{\beta}}_G = (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}^{-1}\mathbf{Y} \quad (16.17)$$

que llamaremos *estimador de mínimos cuadrados generalizados* o estimador *MCG*.

Por ejemplo, supongamos que el ruido sigue el proceso ARMA

$$\mathbf{N} = \boldsymbol{\Psi}\mathbf{A} \quad (16.18)$$

donde

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0.. & 0 & 0 \\ \psi_1 & 1.. & 0 & 0 \\ \dots & 0.. & 1 & 0 \\ \psi_T & \psi_{k..} & \psi_1 & 1 \end{bmatrix}$$

es la matriz triangular de coeficientes de la representación MA( $\infty$ ) del proceso y  $\mathbf{A} = (a_1, \dots, a_T)'$  es un proceso de ruido blanco. El proceso también puede escribirse como

$$\boldsymbol{\Pi}\mathbf{N} = \mathbf{A}$$

donde  $\boldsymbol{\Pi} = \boldsymbol{\Psi}^{-1}$  es la matriz autorregresiva. La matriz de covarianzas de  $\mathbf{N}$  será

$$E[\mathbf{N}\mathbf{N}'] = \boldsymbol{\Psi}\boldsymbol{\Psi}'\sigma^2$$

y la inversa de  $\boldsymbol{\Psi}^{-1}$  es  $\boldsymbol{\Pi}$ . Por tanto la transformación necesaria para un proceso AR(1) mediante su matriz de coeficientes AR que tendrá unos en la diagonal y el coeficiente  $\phi$  en la siguiente subdiagonal y el resto serán ceros. Es decir

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \dots \\ \tilde{y}_T \end{bmatrix} = \begin{bmatrix} 1 & 0.. & 0 & 0 \\ -\phi & 1.. & 0 & 0 \\ \dots & 0.. & 1 & 0 \\ 0 & 0.. & -\phi & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{bmatrix}$$

En análisis anterior es aproximado para los primeros valores ya que la ecuación (16.18) no genera el proceso exactamente, sino sólo después de pasar los primeros retardos. Por ejemplo, para un MA(1) todos los  $\psi_j$  son cero para  $j \geq 2$  y la ecuación que resulta para  $u_1$  es  $u_1 = a_1$  y sólo a partir de  $j = 2$  obtenemos  $u_2 = a_2 + \psi a_1$ . Por tanto la matriz de varianzas y covarianzas teórica de un MA(1) no es exactamente  $\boldsymbol{\Psi}\boldsymbol{\Psi}'\sigma^2$  pero si aproximadamente, si despreciamos la primera observación.

### Apéndice 16. 3 Estimación de la función de transferencia mediante preblanqueo.

Supongamos dos procesos relacionados por la ecuación:

$$y_t = v(B)x_t + n_t \quad (16.19)$$

donde  $x_t$  sigue el proceso ARMA:



$$\phi_x(B) x_t = \theta_x(B) \alpha_t$$

siendo  $\alpha_t$  un proceso de ruido blanco con varianza  $\sigma_\alpha^2$ . Se verifica que:

$$\alpha_t = \theta_x^{-1}(B) \phi_x(B) x_t = \psi^{-1}(B) x_t$$

Si aplicamos este operador  $\psi^{-1}(B)$  a los dos miembros de la relación (16.19), tendremos que:

$$\psi^{-1}(B) y_t = v(B) \psi^{-1}(B) x_t + \psi^{-1}(B) n_t$$

y llamando  $\beta_t = \psi^{-1}(B) y_t$ ,  $\varepsilon_t = \psi^{-1}(B) n_t$ :

$$\boxed{\beta_t = v(B) \alpha_t + \varepsilon_t} \quad (16.20)$$

resultando que existe la misma función de transferencia entre  $\beta_t$  y  $\alpha_t$  que entre  $y_t$  y  $x_t$ . La diferencia principal es que en este último caso la variable explicativa,  $\alpha_t$ , es ruido blanco. En consecuencia, multiplicando ambos miembros de (16.20) por  $\alpha_{t-k}$  y tomando esperanzas:

$$E[\beta_t \alpha_{t-k}] = v_k \sigma_\alpha^2 = \gamma_{\alpha\beta}(k)$$

ya que  $\varepsilon_t$  estará incorrelado con  $\alpha_{t-k}$ , al estarlo, por hipótesis,  $n_t$  y  $x_t$ . Entonces,

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2}$$

y llamando  $\rho_{\alpha\beta}(k)$  a la función de correlación cruzada entre ambas series:

$$\gamma_{\alpha\beta}(k) = \rho_{\alpha\beta}(k) \times \sigma_\alpha \sigma_\beta$$

se obtiene finalmente:

$$\boxed{v_k = \rho_{\alpha\beta}(k) \frac{\sigma_\beta}{\sigma_\alpha}} \quad (16.21)$$

Para obtener una estimación preliminar de  $\hat{v}_k$  se sustituyen los coeficientes de correlación teóricos  $\rho_{\alpha\beta}(k)$  por los muestrales,  $r_{\alpha\beta}(k)$ , y las varianzas teóricas por las estimadas. Con ello resulta:

$$\hat{v}_k = r_{\alpha\beta}(k) \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}$$

y aunque esta estimación por momentos no es eficiente, la representación de  $\hat{v}_k$  en función de  $k$  puede servirnos para obtener una identificación inicial de los órdenes  $(m, a, b)$ .

Este procedimiento se conoce como el *método del preblanqueado* de las series y es debido a Box y Jenkins. Tiene el inconveniente de que no puede extenderse cuando existe más de un regresor.