

ANUARIO DE PSICOLOGÍA
Núm. 43 - 1989 (4)

REGRESIÓN DE SERIES DE TIEMPO
CON MEDICIONES
IGUALMENTE ESPACIADAS

GUILLERMO VALLEJO SECO
Departamento de Psicología y Filosofía
Universidad de Oviedo

Guillermo Vallejo Seco
Departamento de Psicología y Filosofía
Universidad de Oviedo
Aniceto Sala, s/n.
30005 Oviedo

Introducción

La meta del presente escrito es efectuar una breve revisión de lo que a nuestro entender constituye una de las herramientas de trabajo más ampliamente utilizadas dentro de la investigación cuantitativa actual, nos estamos refiriendo, lógicamente, al modelo lineal general. Como es de todos conocido, un modelo lineal consta de dos partes: una ecuación modelo y un conjunto de hipótesis básicas que especifican la naturaleza de los componentes aleatorios y cualquier limitación que los parámetros deban de satisfacer (Kirk, 1982). Para proceder a la estimación de los parámetros, aspecto éste que sólo puede ser efectuado desde los datos recogidos, es muy usual asumir que los factores incluidos en el término de error están minimizados bajo los supuestos de media nula, varianza constante y covarianzas nulas. Cuando las hipótesis de ausencia de dependencia entre los valores del error y de varianza constante se cumplen, los estimadores de los parámetros tienen varianza mínima y son insesgados, por tanto, son estimadores insesgados lineales óptimos. Sin embargo, en el caso que nos encontremos con dependencia entre los errores y/o varianza cambiante, los estimadores siguen siendo insesgados, ya que esta propiedad no depende de la forma de la matriz de varianzas-covarianzas de los errores, pero son ineficientes; y por consiguiente, las estimaciones de las varianzas serán sesgadas (Gottman, 1981; Berry y Lewis-Beck, 1986).

Por lo tanto, a raíz de lo dicho, está clara la necesidad de ser cautelosos cuando se trabaja con este procedimiento tanto con datos registrados temporalmente, como con datos registrados transversalmente. Finalmente, nosotros con el fin de que la relajación en los supuestos de varianza constante y de covarianzas nulas —centrándonos en especial en el caso donde las varianzas sean homocedásticas pero exista autocorrelación—, no atenten contra la validez de las estimaciones estadísticas y de las inferencias causales, presentamos dos de las soluciones más ampliamente recomendadas por la mayor parte de los investigadores que han trabajado este tipo de cuestiones (Hibbs, 1974; Kelejian y Oates, 1974; Pindyck y Rubinfeld, 1976; Wonnacott y Wonnacott, 1982; entre otros). Una por mínimos cuadrados generalizados y la otra mediante transformaciones y diferencias generalizadas.

El modelo lineal general

Supongamos que el comportamiento de una variable criterio «y» se consi-

dera que puede ser adecuadamente explicado mediante una relación lineal de «h-1» variables predictoras «X», más un término de error, «ε», que recoge el efecto conjunto de otras variables no directamente implicadas en el modelo, cuyo efecto individual no resulta relevante:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{h-1} x_{i,h-1} + \varepsilon_i \quad \begin{matrix} (i=1, \dots, n) \\ (j=1, \dots, h-1) \end{matrix} \quad (1)$$

El modelo de la ecuación (1) representa el siguiente sistema de ecuaciones (esto es, aplicándolo a una realidad concreta, partiremos de N observaciones para el conjunto de variables observables implicadas):

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_{h-1} x_{1,h-1} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_{h-1} x_{2,h-1} + \varepsilon_2 \\ &\vdots \\ y_N &= \beta_0 + \beta_1 x_{N1} + \beta_2 x_{N2} + \dots + \beta_{h-1} x_{N,h-1} + \varepsilon_N \end{aligned} \quad (2)$$

De lo expuesto, resulta inmediato expresar este sistema en forma matricial, denominando y al vector ($N \times 1$) de los valores de la variable criterio, X a la matriz de constantes ($N \times h$) de los valores de las variables predictoras, β al vector ($h \times 1$) de parámetros y ε al vector ($N \times 1$) de errores aleatorios:

$$y = X\beta + \varepsilon$$

Expresión que queda definida en base a las siguientes matrices:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,h-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2,h-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{N,h-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{h-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (3)$$

$$\begin{matrix} y & = & X & + & \varepsilon \\ (N \times 1) & & (N \times h) & & (h \times 1) & & (N \times 1) \end{matrix}$$

Para proceder a la estimación del vector de coeficientes β , hemos de establecer algunas asunciones sobre el modo en que se han obtenido las observaciones en el modelo de la ecuación (1). Estos supuestos son decisivos en el proceso de estimación, así que comenzaremos con el conjunto más sencillo para pasar en los próximos apartados a relajar estos supuestos básicos. El conjunto más simple de supuestos importantes es este:

1) Las variables independientes son fijas, no existiendo entre ellas relación lineal.

2) El término de error tiene media 0 y varianza constante para todas las observaciones.

3) No existe correlación entre los errores correspondientes a observaciones diferentes.

Más formalmente los supuestos anteriores pueden expresarse como sigue:

$$\begin{aligned} y &= X\beta + \varepsilon \\ E(\varepsilon) &= 0 \\ \text{COV}(\varepsilon) &= \sigma^2 I \\ X &\text{ matriz fija con rango } (X) = p(X) = h < N: \text{hipótesis de rango pleno.} \end{aligned} \quad (4)$$

Para una mejor comprensión de estas expresiones, vamos a desarrollarlas en el siguiente modelo de regresión simple:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (5)$$

La hipótesis $E(\varepsilon)=0$ implica que

$$E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_N \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ \vdots \\ E(\varepsilon_N) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \quad (6)$$

De forma similar, la hipótesis $\text{COV}(\varepsilon)=\sigma^2 I$ significa que

$$\text{COV}(\varepsilon)=E(\varepsilon\varepsilon') = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1 \varepsilon_2) \dots E(\varepsilon_1 \varepsilon_N) \\ E(\varepsilon_2 \varepsilon_1) E(\varepsilon_2^2) & \dots E(\varepsilon_2 \varepsilon_N) \\ \vdots & \vdots \\ \vdots & \vdots \\ E(\varepsilon_N \varepsilon_1) E(\varepsilon_N \varepsilon_2) \dots E(\varepsilon_N^2) \end{bmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & 0 \dots 0 \\ 0 & 1 \dots 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 0 \dots 1 \end{bmatrix} \quad (7)$$

Es decir, $E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 & \text{para } i=j \\ 0 & \text{para } i \neq j \end{cases}$

Además, se sigue que el valor esperado de y es igual a $X\beta$ y la varianza de y es igual a $\sigma^2 I$.

$$\begin{aligned} E(y) &= E(X\beta + \varepsilon) = X\beta + E(\varepsilon) = X\beta \\ V(y) &= V(X\beta + \varepsilon) = X\beta + V(\varepsilon) = \sigma^2 I \end{aligned} \quad (8)$$

Por último, la hipótesis de matriz X fija supone la consideración en el modelo de la variable predictora como variable no aleatoria y $p(X) = h < N$ implica, en nuestro caso, que existen determinantes no nulos que garantizan la invertibilidad de $X'X$.

Estimación de los parámetros del modelo

Una vez que tenemos determinado el número de variables que integran el modelo, así como las relaciones matemáticas existentes entre todos sus componentes, se considera concluida la fase denominada de especificación y podemos pasar a la fase de estimación para descubrir los parámetros que relacionan mejor la variable dependiente con las independientes.

De entre los diferentes métodos o criterios de estimación estadística, los más habituales son, como es conocido, el de mínimos cuadrados y el de máxima verosimilitud. Aquí, teniendo presente cómo están registrados los datos y que no hay observaciones extraviadas, nos vamos a centrar solamente en el primero de los dos procedimientos, aunque se puede comprobar cómo en el caso del método básico de la regresión lineal los resultados a la hora de estimar los parámetros del modelo coinciden; ya que en los dos métodos se trata de hacer mínima la suma de las desviaciones al cuadrado a la hora de predecir β desde $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{h-1} x_{i,h-1}$.

Los estimadores mínimo-cuadráticos ordinarios (MCO), $\hat{\beta}$, serán aquéllos que hagan mínima la suma de las diferencias al cuadrado entre los valores reales y estimados de la variable dependiente, o lo que es lo mismo, la suma de los cuadrados de los residuos o estimaciones de los valores observables del error aleatorio.

Así pues, de acuerdo con este procedimiento deseamos determinar el vector $\hat{\beta}$ estimador del vector β que haga mínima la suma de cuadrados del error, formalmente.

$$\begin{aligned}
 S &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 & (9) \\
 &= \sum_{i=1}^N [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_{h-1} x_{i,h-1})]^2 = \sum_{i=1}^N \hat{\varepsilon}_i^2 = \text{mínimo}
 \end{aligned}$$

En notación matricial

$$\begin{aligned}
 S &= \hat{\varepsilon}'\hat{\varepsilon} = (\mathbf{y} - \mathbf{y})'(\mathbf{y} - \mathbf{y}) \\
 &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) & (10) \\
 &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}
 \end{aligned}$$

Para hallar el valor de $\hat{\beta}$ que minimiza la suma de los cuadrados de los residuos derivamos la ecuación (10) con respecto a $\hat{\beta}$:

$$\begin{aligned} \theta \varepsilon' / \theta \hat{\beta} &= 0 - 2X'y + 2X'X\hat{\beta} \\ &= -X'y + X'X\hat{\beta} \end{aligned} \quad (11)$$

Igualando a cero y reordenando los términos en (11) obtenemos la ecuación matricial que da el conjunto de parámetros estimados $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1}$ que minimiza $\varepsilon'\varepsilon$ ó $\sum_{i=1}^N (y_i - \hat{y}_i)^2$ es:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (12)$$

Propiedades de los estimadores de los parámetros mínimo-cuadráticos

Considerando que este apartado tiene una importancia capital en la comprensión de los problemas que hemos presentado en la introducción, nuestro objetivo en él es desarrollar de forma clara y precisa las propiedades deseables que deben reunir los estimadores mínimo-cuadráticos.

El primer aspecto a observar de los estimadores mínimo-cuadráticos es el de linealidad. Por lineales se entiende que son funciones lineales del vector de observaciones como se demostró en la ecuación (12). (téngase presente que establecimos que las variables predictoras permanecen fijas de muestra a muestra).

El segundo aspecto a probar de los estimadores mínimo-cuadráticos es la condición de insesgadez. Wonnacott y Wonnacott (1982, p. 60) nos dicen «que un estimador insesgado es el que, en promedio, coincide con el valor del parámetro». O sea, la esperanza matemática de $\hat{\beta} = \beta$. Naturalmente, los estimadores no tienen por qué coincidir en cada aplicación concreta con los parámetros teóricos, sino que dependerán de los errores:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon \end{aligned} \quad (13)$$

es decir, la diferencia entre el estimador y el parámetro será una función lineal de los términos de error del modelo y constituye la llamada varianza muestral.

Sin embargo, si se repite el proceso de muestreo, en media, coinciden $\hat{\beta}$ y β , por lo que $\hat{\beta}$ resulta un estimador insesgado de β .

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1}X'y] \\ &= \beta + (X'X)^{-1}X'E(\varepsilon), \text{ ya que } X \text{ es fija} \\ &= \beta; \text{ ya que por hipótesis } E(\varepsilon) = 0 \end{aligned} \quad (14)$$

Por su parte, la matriz de dispersión o de varianzas-covarianzas de los estimadores MCO es definida como: $COV(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$

$$\begin{aligned}
 \text{COV}(\hat{\beta}) &= (\hat{\beta} - \beta) \\
 &= [(X'X)^{-1}X'y - \beta] = [(X'X)^{-1}X'(X\beta + \epsilon) - \beta] \\
 &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon - \beta \\
 &= \beta + (X'X)^{-1}X'\epsilon - \beta \\
 &= (X'X)^{-1}X'\epsilon
 \end{aligned}$$

$$\text{COV}(\hat{\beta}) = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \quad (15)$$

La anterior expresión implica que el cálculo de la matriz de dispersión exige disponer del valor de la varianza poblacional, σ^2 , (y, por tanto, desconocida) del error aleatorio. Como consecuencia, deberemos tratar de establecer un estimador para σ^2 , aspecto éste que abordaremos posteriormente. Hasta ahora hemos demostrado que los estimadores mínimo-cuadráticos son lineales e insesgados, pero como la insesgaredad no implica nada con respecto a la dispersión de la distribución del estimador debemos también investigar si entre todos los estimadores de esta clase estos estimadores son los de menor varianza; pues, estimadores insesgados, pero con varianza elevada, producirán estimaciones muy alejadas del verdadero objetivo. Por ello es deseable que los estimadores mínimo-cuadráticos tengan varianza mínima. En definitiva, se trata de demostrar que los estimadores $\hat{\beta}$, además de ser lineales e insesgados, poseen una varianza menor que cualquier otro estimador lineal e insesgado y, por tanto, que son estimadores lineales insesgados óptimos (ELIO) tal y como se establece en el teorema de Gauss-Markov.

Si β^* es cualquier otro estimador de β diferente de $\hat{\beta}$ se puede mostrar que la matriz de dispersión del estimador β^* es mayor o a lo sumo igual que la matriz de dispersión de $\hat{\beta}$, formalmente:

$$\text{COV}(\beta^*) \geq \text{COV}(\hat{\beta}) \quad (16)$$

La desigualdad ha de interpretarse como que $\text{COV}(\beta^*) - \text{COV}(\hat{\beta})$ es una matriz semidefinida positiva. En particular, los elementos de la diagonal de $\text{COV}(\beta^*) - \text{COV}(\hat{\beta})$ son todos ≥ 0 .

A efectos de demostrar dicha propiedad, de la clase más general de los estimadores lineales, partamos de cualquier β^* arbitrario:

$$\beta^* = [(X'X)^{-1}X' + P]y \quad (17)$$

donde P será una matriz cualquiera ($h \times n$) que en caso de anularse implica que $\beta^* = \hat{\beta}$. Desarrollando la expresión anterior en función del error aleatorio tenemos:

$$\begin{aligned}
 \beta^* &= [(X'X)^{-1}X' + P](X\beta + \epsilon) \\
 &= \beta + (X'X)^{-1}X'\epsilon + PX\beta + P\epsilon
 \end{aligned} \quad (18)$$

si imponemos ahora la restricción de que sea β^* insesgado

$$\begin{aligned}
 E(\beta^*) &= \beta + (X'X)^{-1}X'E(\epsilon) + PX\beta + E(P\epsilon) \\
 &= \beta + PX\beta \\
 &= \beta; \text{ luego } PX = 0
 \end{aligned}
 \tag{19}$$

Así pues, si β^* es insesgado, $E(\beta^*) = \beta$, y puesto que esta relación es cierta para cualquier valor posible de β , PX debe ser igual a cero.

Calculamos seguidamente la matriz de varianza-covarianza de cualquier estimador lineal insesgado:

$$\begin{aligned}
 \text{COV}(\beta^*) &= E[(\beta^* - \beta)(\beta^* - \beta)']; \text{ donde } \text{COV}(\beta^*) = [(X'XZ)^{-1}X' + P]E \\
 &= [(X'X)^{-1}X' + P]E(\epsilon\epsilon') [X(X'X)^{-1} + P'] \\
 &= \sigma^2[(X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'P' + PX(X'X)^{-1} + PP'] \\
 &= \sigma^2 [(X'X)^{-1} + PP']
 \end{aligned}
 \tag{20}$$

dado que $(X'X)^{-1}X'X = I$ y los otros dos sumandos resultan nulos por incluir como factores PX o $X'P'$.

Por tanto, $\text{COV}(\beta^*) - \text{COV}(\beta) = \sigma^2 PP'$ que es una matriz semidefinida positiva. Esto prueba que $\hat{\beta}$ es el estimador insesgado óptimo de β , ya que como se puede comprobar el menor valor de $\text{COV}\hat{\beta}^*$ se obtiene para $PP' = 0$; es decir, precisamente para $\hat{\beta}^* = \beta$.

En definitiva, hemos demostrado que los β^{MCO} son lineales e insesgados óptimos, por lineales se entiende que son funciones lineales de y , insesgados porque en promedio coinciden con el parámetro $[E(\beta) = \beta]$ y óptimos porque tales estimadores poseen una varianza menor que cualquier otro estimador lineal insesgado.

Debe de tenerse presente que para que la última igualdad pueda deducirse debe partirse de que las x son fijas o al menos tienen una distribución independiente de ϵ , así como de aceptar las restantes hipótesis sobre el comportamiento de ϵ . Sólo sobre estos supuestos puede decirse que los estimadores obtenidos son ELIO.

Por último, cuando se amplía a la población, los estimadores mínimo-cuadráticos deben de ser consistentes. Esto puede definirse diciendo que $\hat{\beta}$ es un estimador consistente de β , si su error cuadrático medio $E(M) = (\hat{\beta} - \beta)^2$, tiende a 0 cuando n tiende a infinito.

Estimador insesgado MCO de la varianza de los errores aleatorios

Como los valores de ϵ no pueden obtenerse directamente y deben derivarse desde el modelo ajustado, parece admisible que el estimador σ^2 se base en la suma de cuadrados de los residuales $\hat{\epsilon}'\hat{\epsilon}$. Para definir los residuales del modelo mínimo-cuadrático pondremos $\hat{\epsilon}$ en función de ϵ sustituyendo en $\hat{\epsilon} = y - X\hat{\beta}$, y por $X\hat{\beta} + \epsilon$ y $\hat{\beta}$ por $(X'X)^{-1}X'y$ y como sigue:

$$\begin{aligned}
\hat{\varepsilon} &= y - X\hat{\beta} = X\beta + \varepsilon - X(X'X)^{-1}X'y \\
&= X\beta + \varepsilon - X(X'X)^{-1}X'(X\beta + \varepsilon) \\
&= X\beta + \varepsilon - X(X'X)^{-1}X'X\beta - X(X'X)^{-1}X'\varepsilon \\
&= X\beta + \varepsilon - XI\beta - X(X'X)^{-1}X'\varepsilon \\
&= \varepsilon - X(X'X)^{-1}X'\varepsilon \\
&= [I_n - X(X'X)^{-1}X']\varepsilon \\
&= M\varepsilon
\end{aligned} \tag{21}$$

Por lo tanto, nos quedan los residuos como una función lineal de los errores aleatorios a través de una matriz cuadrada M de orden n y que presenta las siguientes propiedades importantes.

a. *Simetría*, ya que la traspuesta es igual a ella misma.

$$M = M' = I_n - X(X'X)^{-1}X' \tag{22}$$

b. *Idempotencia*, ya que multiplicada por ella misma resulta ella misma

$$M = MM' = M^2 = M \tag{23}$$

por tanto, la suma de los cuadrados de los residuales es:

$$\begin{aligned}
\hat{\varepsilon}'\hat{\varepsilon} &\doteq \varepsilon'M'M\varepsilon \\
&= \varepsilon'M\varepsilon \text{ puesto que } M \text{ es simétrica e idempotente} \\
&= \varepsilon'[(I_n - X(X'X)^{-1}X')]\varepsilon
\end{aligned} \tag{24}$$

c. *Semidefinida positiva* ya que como se ha demostrado en la ecuación anterior $\varepsilon'\varepsilon = \varepsilon'M\varepsilon$. Pero la suma de cuadrados de los residuos $\varepsilon'\varepsilon \geq 0$. De aquí que M sea semidefinida positiva por definición y, por tanto, el valor esperado de $\varepsilon'M\varepsilon = \sigma^2$ traza de (M) , aspecto éste que resulta inmediato comprobar de la demostración que siguiendo a Pulido (1983) mostramos a continuación:

$$\begin{aligned}
E \left[\begin{matrix} \varepsilon_1 & \dots & \varepsilon_n \end{matrix} \right] \begin{bmatrix} m_{11} & \dots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \dots & m_{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} &= \\
= E(\varepsilon_1 \sum \varepsilon_j m_{j1} + \dots + \varepsilon_n \sum \varepsilon_j m_{jn}) &= \sigma^2 m_{11} + \dots + \sigma^2 m_{nn} = \sigma^2 \text{ traza } (M)
\end{aligned} \tag{25}$$

Aprovechando ahora las propiedades de las trazas, y en particular que $\text{traza}(AB) = \text{traza}(BA)$; traza de $(ABC) = \text{traza}(CAB) = \text{traza}(BCA)$, pueden definirse ambas operaciones tomando esperanzas matemáticas de ambos miembros:

$$\begin{aligned}
 E(\hat{\varepsilon}'\hat{\varepsilon}) &= E(\varepsilon'MM\varepsilon) \\
 &= \sigma^2 \text{traza}(M) \\
 &= \sigma^2 \{n - \text{traza} [(X'X)^{-1} X'X]\} = \sigma^2 \{n - \text{traza}(I_h)\} \\
 &= \sigma^2 (N - h)
 \end{aligned}
 \tag{26}$$

donde h es el número de variables predictoras más el término constante y N el número de sujetos. Así pues, un estimador muestral insesgado de la varianza del error es dado por:

$$E(s^2) = E\left[\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N-h}\right] = \sigma^2 \tag{27}$$

El estimador por mínimos cuadrados generalizados

En el desarrollo del modelo lineal $y = X\beta + \varepsilon$ para proceder a la estimación del vector de los parámetros β establecimos como hipótesis básicas del comportamiento de los errores la constancia de las varianzas y la nulidad de las covarianzas, lo cual conlleva a reducir la matriz de dispersión de varianzas-covarianzas al escalar $\sigma^2 I$, o sea, a un único valor, dado que:

$$E(\varepsilon\varepsilon') = \sigma^2 I_n \tag{28}$$

Ahora bien, si la matriz de varianzas-covarianzas no cumple las dos asunciones básicas contenidas en la expresión $E(\varepsilon\varepsilon') = \sigma^2 I$, la aplicación de la solución mínimo-cuadrática obtenida en la ecuación (12) continuará produciendo estimaciones lineales e insesgadas, pero ineficientes, ya que los estimadores efectuados de la varianza de $\hat{\beta}$ cuando la matriz de las varianzas-covarianzas no es escalar estarán sesgadas por un doble motivo. Veamos esto como sigue:

$$\begin{aligned}
 \text{VAR}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\
 &= E[(X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1}] \\
 &= E[(X'X)^{-1} X' (\sigma^2 \Omega) X (X'X)^{-1}] \\
 \text{VAR}(\hat{\beta}) &= \sigma^2 (X'X)^{-1} X' \Omega X (X'X)^{-1}
 \end{aligned}
 \tag{29}$$

por tanto, si tenemos en cuenta la estimación de la varianza de $\hat{\beta}$ obtenida en la ecuación (15) cuando la matriz de covarianzas era escalar podemos comprobar que

$$(X'X)^{-1} \neq (X'X)^{-1} X' \Omega X (X'X)^{-1} \tag{30}$$

pero, añadamos además

$$\begin{aligned}
 E(\hat{\epsilon}'\hat{\epsilon}) &= \text{tr}(M\epsilon\epsilon') \\
 &= \sigma^2 \text{tr}[\Omega - X(X'X)^{-1}X'\Omega] \\
 &= \sigma^2 \text{tr}\Omega - \text{tr}(X'X)^{-1}X'\Omega X \\
 &= \sigma^2 N - \text{tr}[(X'X)^{-1}X'\Omega X]
 \end{aligned} \tag{31}$$

por tanto, haciendo uso de la estimación de la varianza del error obtenida con la ecuación (26) cuando la matriz de covarianzas era escalar podemos verificar que

$$\text{Tr}[(X'X)^{-1}X'X] \neq \text{Tr}[(X'X)^{-1}X'\Omega X] \tag{32}$$

En consecuencia, vemos que los mínimos cuadrados ordinarios cuando la matriz de covarianzas no se reduce a un único escalar producirán estimaciones sesgadas, por defecto o por exceso, con lo cual los contrastes habituales carecen de validez.

Vistas las implicaciones que se derivan del incumplimiento de las asunciones de varianza constante y de covarianzas nulas, el problema que se nos plantea es como conducirnos cuando esto ocurra. Para ello en este apartado el modelo de mínimos cuadrados ordinarios va a ser generalizado a situaciones en las cuales se considere la existencia de una matriz de covarianzas de los errores que pudiera tener tantos elementos distintos en la diagonal como fuera de ella. Para ello revisamos la expresión $\sigma^2 I$ del modelo básico y la sustituimos por otra que nos permita el manejar la existencia de heterocedasticidad y/o de autocorrelación. En particular, la expresión revisada la hemos expresado anteriormente como sigue:

$$E(\epsilon\epsilon') = \sigma^2 \Omega \tag{33}$$

donde σ^2 es desconocida, pero Ω es una matriz conocida de orden $n \times n$, simétrica y definida positiva.

Así pues, la cuestión que se nos plantea ahora es cómo estimar el vector de parámetros β en el modelo lineal generalizado.

$$y = X\beta + \epsilon \tag{34}$$

manteniendo las hipótesis

$$E(\epsilon) = 0 \tag{35}$$

$$X \text{ matriz fija con rango pleno } (h) < N \tag{36}$$

pero añadiendo

$$E(\epsilon\epsilon') = \sigma^2 \Omega \tag{37}$$

Este problema se ha enfocado de varias formas equivalentes, una de las más sencillas es la siguiente.

De acuerdo con el teorema de descomposición de Cholesky si una matriz D es definida positiva, existe una matriz P , no singular, tal que $PP' = D$. En el

caso de Ω podemos encontrar una matriz P que multiplicada por P' sea igual a Ω^{-1} .

De este modo, premultiplicando el modelo $y = X\beta + \varepsilon$ por P se obtiene

$$Py = (PX)\beta + (P\varepsilon) \quad (38)$$

con lo cual se puede comprobar que la nueva matriz de covarianza para los residuales transformados es σ^2 veces la matriz de identidad, formalmente

$$E(\varepsilon_i \varepsilon_j') = \sigma^2 I \quad (\text{ver ecuación 81}) \quad (39)$$

y el estimador β que cumple con las propiedades del teorema de Gauss-Markov es ahora

$$\begin{aligned} \tilde{\beta} &= (X'P'PX)^{-1}X'P'Py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega y \end{aligned}$$

Asumiendo la normalidad de los errores hubiésemos llegado a la misma solución de haber partido de la función de probabilidad normal multivariada

$$L(\tilde{\beta}) = \frac{1}{(2\pi\sigma^2)^{n/2} |\Omega|^{1/2}} e^{-(1/2)(y-X\tilde{\beta})'\Omega^{-1}(y-X\tilde{\beta})} \quad (41)$$

con lo que se puede verificar que en el modelo lineal generalizado las estimaciones máximo-verosímiles del vector de parámetros β son idénticas a las estimaciones mínimo-cuadráticas generalizadas, ya que esta función alcanza su máximo con respecto a β , cuando la magnitud del exponente es mínimo. Baste para ello repetir la demostración realizada en las ecuaciones (10) y (11) para comprobar que el vector de derivadas parciales en este caso es $0 - 2X'\Omega^{-1}y + 2X'\Omega^{-1}X\tilde{\beta}$, con lo que igualando a cero y reordenando los términos obtendremos el mismo estimador mínimo cuadrático generalizado que es la ecuación (40).

Propiedades y características de los estimadores mínimo-cuadráticos generalizados

El nuevo estimador definido en la ecuación (40) es el estimador por mínimos cuadrados generalizados o estimador de Aitken y coincide con el estimador de mínimos cuadrados ordinarios en el caso particular de que $\Omega = I$.

Ahora, utilizando el subíndice MCG para simbolizar mínimos cuadrados generalizados, vamos a desarrollar las propiedades que deben reunir los estimadores MCG (utilizando argumentos análogos a los empleados al comentar las propiedades de los estimadores MCO) para que se cumpla que el mejor estimador lineal insesgado del verdadero vector de parámetros β en el modelo de las ecuaciones (34) a (37) es el estimador MCG($\tilde{\beta}$).

Para empezar el estimador β^{MCG} es insesgado como se puede verificar en la demostración que sigue:

$$\begin{aligned}
 E(\beta^{MCG}) &= E(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta + \epsilon)] \\
 &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + (X'\Omega^{-1})^{-1}X'\Omega^{-1}\epsilon] \\
 &= \beta + E[(X'\Omega^{-1}X)^{-1}\epsilon] \\
 &= \beta + [(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\epsilon)] \\
 E(\beta^{MCG}) &= \beta
 \end{aligned}
 \tag{42}$$

Por su parte Aitken (1935) ha demostrado que el estimador (β^{MCG}) que tiene varianza mínima será:

$$\begin{aligned}
 COV(\beta^{MCG}) &= E[(\bar{\beta} - \beta)(\bar{\beta} - \beta)'] \\
 &= E\{[\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon - \beta][\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon - \beta]'\} \\
 &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(\epsilon\epsilon')\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\
 &= [(X'\Omega^{-1})^{-1}X'\Omega^{-1}E(\epsilon\epsilon')\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\
 &= \sigma^2(X'\Omega^{-1}X)^{-1}
 \end{aligned}
 \tag{43}$$

Los estimadores β^{MCG} no sólo serán más eficientes que los β^{MCO} —ya que la $COV(\beta)$ en el caso de que Ω no sea una matriz escalar estará sesgada, dado que, por un lado, $(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq (X'X)^{-1}$ (véase ecuación 29) y, por otro, $E(\hat{\sigma}^2) \neq \sigma^2$ (véase ecuación 31)— sino que también puede demostrarse que los estimadores por mínimos cuadrados generalizados son, como establece el teorema de Gauss-Markov, los que minimizan la varianza de cada estimador; es decir, que los estimadores β^{MCG} pasan a ser ahora ELIO, al igual que ocurría con los estimadores β^{MCO} anteriormente cuando los supuestos básicos del modelo de la regresión no se relajaban.

Como hicimos en el apartado análogo de más arriba, para demostrar que $\bar{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ es un estimador de varianza mínima, trataremos de ver si es posible mejorarlo, por supuesto sin introducir sesgo. Es decir, trataremos de ver si es posible encontrar un estimador lineal que sea más eficiente, pero que a la vez cumpla con la condición de insesgado.

Para tales fines partiremos de cualquier otro estimador lineal (β^*) de β diferente de $\bar{\beta}$:

$$\beta^* = [(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + P]y
 \tag{44}$$

donde P la supuesta mejora a determinar será como en el caso anterior del modelo básico de la regresión una matriz cualquiera que en caso de anularse implica que $\beta^* = \bar{\beta}$. Seguidamente mostraremos esto desarrollando la expresión anterior en función del error aleatorio.

$$\begin{aligned}
 \beta^* &= [(X'\Omega^{-1}X)^{-1}\Omega^{-1} + P] [X\beta + \epsilon] \\
 &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + (X'\Omega^{-1}X)^{-1}X\Omega^{-1}\epsilon + PX\beta + P\epsilon \\
 &= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon + PX\beta + P\epsilon \\
 E(\beta^*) &= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\epsilon) + PX\beta + PE(\epsilon) \\
 &= \beta + PX\beta \\
 &= \beta
 \end{aligned} \tag{45}$$

Ahora bien, con el fin de no cambiar el valor esperado de nuestro estimador insesgado ($\tilde{\beta}$), y también de evitar que haya sesgos en cualquier otro estimador lineal de β , $PX\beta$ debe ser igual a cero para todos los posibles vectores β (Wonnacott y Wonnacott, 1982).

Así pues, si β^* es insesgado, $E(\beta^*) = \beta$, y puesto que esta relación es cierta para cualquier posible valor de β , tenemos, por tanto que $PX = 0$. Por su parte, la matriz de covarianzas será:

$$\begin{aligned}
 COV(\beta^*) &= E[(\beta^* - \beta)(\beta^* - \beta)'] \text{ donde } COV(\beta^*) = [(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + P]\epsilon \\
 &= [(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + P]E(\epsilon\epsilon') [\Omega^{-1}X(X'\Omega^{-1}X)^{-1} + P'] \\
 &= \sigma^2[(X'\Omega^{-1}X)^{-1} + PP']
 \end{aligned} \tag{46}$$

dado que $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X = I$ es el producto de una matriz por su inversa y los otros dos términos se hacen cero por incluir como factores a PX o $X'P'$. Por tanto, $COV(\beta^*) - COV(\tilde{\beta}) = \sigma^2 PP'$, que es una matriz semidefinida positiva. Esto prueba que $\tilde{\beta}$ es un estimador máximo verosímil de β , por lo que su varianza es mínima frente a todos los estimadores insesgados (Johnston, 1984; Maddala, 1977).

Estimador insesgado MCG de la varianza de los errores aleatorios

A diferencia del modelo de la regresión MCO, el de MCG produce un estimador insesgado de σ^2 cuando las varianzas son heterocedásticas y/o los errores están correlacionados. Esto es demostrado de forma análoga a como lo hicimos en el apartado equivalente anterior. Los residuales en un modelo generalizado son definidos:

$$\begin{aligned}
 \tilde{\epsilon} &= y - X\tilde{\beta} = y - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\
 &= (X\beta + \epsilon) - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta + \epsilon) \\
 &= (X\beta + \epsilon) - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon \\
 &= X\beta + \epsilon - X\beta - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon \\
 &= \epsilon - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon \\
 &= [I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]\epsilon \\
 &= \tilde{M}\epsilon
 \end{aligned} \tag{47}$$

Por lo tanto, la suma de los cuadrados de los residuales minimizada por MCG puede ser expresada como sigue:

$$\begin{aligned}
 (y - X\tilde{\beta})' \Omega^{-1} (y - X\tilde{\beta}) &= \tilde{\varepsilon}' \Omega^{-1} \tilde{\varepsilon} \\
 \varepsilon \Omega^{-1} \tilde{\varepsilon} &= (\tilde{M}\varepsilon)' \Omega^{-1} (\tilde{M}\varepsilon) \\
 \tilde{\varepsilon} \Omega^{-1} \tilde{\varepsilon} &= \varepsilon' \tilde{M}' \Omega^{-1} \tilde{M} \varepsilon = \varepsilon' \Omega^{-1} \tilde{M} \varepsilon
 \end{aligned}
 \tag{48}$$

De aquí se sigue que el valor esperado para la suma de los cuadrados de los residuales nos da el estimador MCG para suma de cuadrados de los errores aleatorios:

$$\begin{aligned}
 E(\tilde{\varepsilon}' \Omega^{-1} \tilde{\varepsilon}) &= E(\varepsilon' \Omega^{-1} \tilde{M} \varepsilon) \\
 &= E \text{ traza } (\varepsilon' \Omega^{-1} \tilde{M} \varepsilon) \\
 &= E \text{ traza } (\tilde{M}' \varepsilon' \Omega^{-1}), \text{ por ecuación (25)} \\
 &= \sigma^2 \text{ traza } M \\
 &= \sigma^2 \text{ traza } [I_n - X (X' \Omega^{-1})^{-1} X' \Omega^{-1}] \\
 &= \sigma^2 (N - h)
 \end{aligned}
 \tag{49}$$

Así pues, un estimador insesgado MCG de la varianza del error es obtenido por:

$$\sigma^2 = \frac{\tilde{\varepsilon} \Omega^{-1} \tilde{\varepsilon}}{N - h}
 \tag{50}$$

Las propiedades de los estimadores de los parámetros, así como de los errores aleatorios establecen que el procedimiento de MCG nos proporciona lo que es en teoría una óptima solución para el problema creado por el hecho de que la matriz de varianzas-covarianzas no reduzca su valor a un único valor σ^2 . Por último, en lo que resta del presente trabajo ilustraremos cómo se puede hacer uso del procedimiento descrito en los apartados anteriores, en el caso en que los supuestos básicos del modelo de la regresión lineal de media nula y varianza constante se mantengan, pero por el propio tipo de registros (de corte longitudinal) el de covarianzas nulas se vea incumplido.

Dependencia serial

Si bien es verdad que en muchos estudios dentro de las ciencias sociales y comportamentales, generalmente basados en datos de sección cruzada o de corte transversal, la hipótesis de varianza constante es poco realista, no lo es menos que también existen otras muchas situaciones, por cierto muy frecuentes dentro del ámbito psicológico, donde lo que resulta difícil mantener no es ya la hipótesis de homocedasticidad, sino la ausencia de dependencia entre los errores. Produciéndose de esta forma problemas de estimación estadística y con ello, como es lógico, de inferencia causal. En general, podemos decir que el incumplimiento del supuesto de varianza constante, es un problema encontrado habitualmente

cuando se trabaja con datos de sección cruzada (sobre todo en investigaciones no experimentales), más que con datos registrados temporalmente, mientras que el incumplimiento de la independencia serial de los errores es un problema encontrado habitualmente cuando se trabaja con datos de corte longitudinal, aunque la dependencia serial también puede ocurrir con datos transversales como ocurre en el caso en que sean las variables omitidas las responsables de la autocorrelación (Maddala, 1977) (Berry y Lewis-Beck, 1986).

Como es frecuentemente señalado dentro de la investigación econométrica y predictiva (Maddala, 1977; Box y Jenkins, 1970, 1976), los supuestos más comunes sobre la autocorrelación de los residuales son que forman:

1. Un proceso autorregresivo (AR).
2. Un proceso de media móvil (MA).
3. Un proceso mixto autorregresivo y de media móvil (ARMA).

En este ensayo, para ilustrar el problema consideraremos que la dependencia serial se debe a un proceso autorregresivo de primer orden AR(1). Somos conscientes de las limitaciones que presenta la simplicidad de este supuesto. Ya que, como Engle (1973) ha mostrado, no está claro que los procedimientos o soluciones que vamos a presentar en las líneas siguientes sean superiores al procedimiento MCO en el caso de que los modelos de dependencia serial difieran del presentado aquí, Considérese el modelo

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_{h-1} x_{t,h-1} + \varepsilon_t \quad (51)$$

en donde suponemos que el error aleatorio ε_t sigue un proceso autorregresivo de primer orden (cada residual depende únicamente de su propio valor previo y de un componente aleatorio de ruido blanco).

$$\varepsilon_t = p\varepsilon_{t-1} + v_t; \text{ por lo que } \text{COV}(\varepsilon) = \sigma_v^2 \Omega \quad (52)$$

en donde p está comprendido entre 1 y -1 y v_t satisface los supuestos de los errores aleatorios del modelo básico de la regresión; esto es, $E(v_t) = 0, \text{VAR}(v_t) = \sigma^2$ para todo t y $\text{COV}(v_t, v_s) = 0$, para todo $t \neq s$.

Entonces tenemos

$$\begin{aligned} \varepsilon_t &= p\varepsilon_{t-1} + v_t \\ &= p(p\varepsilon_{t-2} + v_{t-1}) + v_t \\ &= p^2(p\varepsilon_{t-3} + v_{t-2}) + pv_{t-1} + v_t \\ &= \dots \\ &= v_t + pv_{t-1} + p^2 v_{t-2} + \dots + p^n v_{t-n} \end{aligned} \quad (53)$$

y, por tanto,

$$E(\varepsilon_t) = 0; \text{ ya que } E(v_t) = 0 \text{ para todo } t.$$

Además, $E(\varepsilon_{t-1} v_t) = 0$, porque los v_t son serialmente independientes y ε_{t-1} , incluye v_{t-1}, v_{t-2}, \dots , etc., pero no v_t .

A partir de aquí la varianza de ε_t para el modelo autorregresivo de primer orden es obtenida como sigue:

$$\begin{aligned}\sigma_\varepsilon^2 &= E(\varepsilon_t^2) = E(p\varepsilon_{t-1} + v_t)^2 \\ &= p^2 E(\varepsilon_{t-1}^2) + E(v_t^2) + 2pE(\varepsilon_{t-1}v_t) \\ &= p^2\sigma_\varepsilon^2 + \sigma_v^2\end{aligned}\quad (54)$$

dado que según el supuesto de estacionariedad la varianza de ε_t es la misma para todo t

$$\begin{aligned}\sigma_\varepsilon^2 - p^2\sigma_\varepsilon^2 &= \sigma_v^2 \\ \sigma_\varepsilon^2(1 - p^2) &= \sigma_v^2 \\ \sigma_\varepsilon^2 &= \frac{\sigma_v^2}{1 - p^2}\end{aligned}\quad (55)$$

Por otra parte, la autovarianza entre ε_t y ε_{t-1} es obtenida

$$\begin{aligned}E(\varepsilon_t\varepsilon_{t-1}) &= E[(p\varepsilon_{t-1} + v_t)\varepsilon_{t-1}] \\ &= pE(\varepsilon_{t-1}^2) + E(\varepsilon_{t-1}v_t) \\ &= p\sigma_\varepsilon^2\end{aligned}\quad (56)$$

ya que como especificamos anteriormente ε_t tiene varianza constante independientemente del tiempo y $E(\varepsilon_{t-1}v_t) = 0$.

Análogamente,

$$\begin{aligned}E(\varepsilon_t\varepsilon_{t-2}) &= E[(p\varepsilon_{t-1} + v_t)\varepsilon_{t-2}] \\ &= E[(p^2\varepsilon_{t-2} + pv_{t-1} + v_t)\varepsilon_{t-2}] \\ &= p^2E(\varepsilon_{t-2}^2) + pE(\varepsilon_{t-2}v_{t-1}) + E(\varepsilon_{t-2}v_t) \\ &= p^2\sigma_\varepsilon^2 \text{ por las mismas asunciones anteriores.}\end{aligned}\quad (57)$$

Y, en general,

$$E(\varepsilon_t\varepsilon_{t-s}) = p^s\sigma_\varepsilon^2 \quad (58)$$

De este modo, podemos comprobar cómo la hipótesis de independencia serial entre los residuos del modelo de la ecuación (51) no es respetada. A la misma conclusión hubiésemos llegado en el caso de que los modelos más complejos de dependencia serial hubieran sido contemplados.

Una vez que tenemos derivadas las funciones de varianza y autovarianza es fácil derivar la función de autocorrelación entre ε_t y ε_{t-s} para un modelo AR(1).

$$\rho_s = \frac{\text{COV}(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{\text{VAR}(\varepsilon_t)} \sqrt{\text{VAR}(\varepsilon_{t-1})}} = \frac{p^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2} = p^s \quad (59)$$

por consiguiente, la función de autocorrelación de retardo s cuando el término

de error de la ecuación (51) sigue un proceso AR(1) coincide con la s-ésima potencia del parámetro p , con lo que el correlograma de la serie ϵ_t decaerá exponencialmente conforme incrementamos los intervalos temporales, dado que $p < 1$. El correlograma es una representación gráfica de la evolución de los coeficientes de autocorrelación y juega un papel central en la identificación del modelo de dependencia serial.

La expresión $E(\epsilon\epsilon') = \sigma_\epsilon^2 \Omega$ podemos ahora escribirla en la forma

$$E(\epsilon\epsilon') = \sigma_\epsilon^2 \begin{bmatrix} 1 & p & p^2 & p^3 & \dots & p^{t-1} \\ p & 1 & p & p^2 & \dots & p^{t-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p^{t-1} & p^{t-2} & p^{t-3} & p^{t-3} & \dots & 1 \end{bmatrix} \quad (60)$$

Problemas derivados de la autocorrelación de primer orden [AR(1)]

Si aplicamos el método MCO a un modelo en que los residuos siguen un esquema AR(1) obtendremos estimadores MCO no sesgados, pero menos eficientes que los obtenidos al utilizar un método que tenga en cuenta la autocorrelación, como por ejemplo el de MCG. Además, si aplicamos el procedimiento de los MCO para obtener las varianzas muestrales de los coeficientes de regresión es probable que obtengamos una importante subestimación de los mismos. De esta forma, si aplicamos una solución por MCG, lo que ganamos es una mayor eficacia al disponer de unas varianzas correctamente estimadas (Maddala, 1977).

Para demostrar esto recordemos que

$$E(\hat{\beta}) = \beta + [(X'X)^{-1}X'E(\epsilon)] \quad (61)$$

$$= \beta$$

y que la varianza de $\hat{\beta}$ en el caso de ausencia de autocorrelación era

$$VAR(\hat{\beta}) = \sigma_\epsilon^2 (X'X)^{-1} = \frac{\sigma_\epsilon^2}{\sum_{t=1}^N x_t^2} \quad (62)$$

donde x_t son desviaciones de la media, con lo que $X' = [x_1, \dots, x_n]$. Cuando los errores están autocorrelacionados la $VAR(\hat{\beta})$ era como vimos en la ecuación (29) igual a

$$VAR(\hat{\beta}^{MCO}) = \sigma_\epsilon^2 (X'X)^{-1}X'\Omega X(X'X)^{-1} \quad (63)$$

Seguidamente pasamos a ilustrar el error cometido al utilizar la $VAR(\hat{\beta})$ de la ecuación (15) en lugar de la $VAR(\hat{\beta})$ de la ecuación (29) con el modelo de

la regresión simple, para ello se han seguido los trabajos de Johnston (1984) y Rao y Griliches (1960).

$$\begin{aligned} y_t &= \beta_1 X_t + \varepsilon_t \\ \varepsilon_t &= p\varepsilon_{t-1} + v_t \\ x_t &= rx_{t-1} + w_t \end{aligned} \quad (64)$$

en donde v_t y w_t cumplen las hipótesis postuladas anteriormente por los errores aleatorios y p y r se encuentran comprendidos entre 1 y -1.

$$\begin{aligned} \text{VAR}(\hat{\beta}^{\text{MCO}}) &= \sigma_\varepsilon^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} \\ &= \frac{\sigma_\varepsilon^2}{\sum_{t=1}^N x_t^2} [x_1 \dots x_t] \begin{bmatrix} 1 & p & p^2 & \dots & p^{t-1} \\ p & 1 & p & \dots & p^{t-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p^{t-1} & p^{t-2} & p^{t-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_t \end{bmatrix} \frac{1}{\sum_{t=1}^N x_t^2} \\ &= \frac{\sigma_\varepsilon^2}{\sum_{t=1}^N x_t^2} \left[1 + 2p \frac{\sum_{t=1}^{N-1} x_t x_{t+1}}{\sum_{t=1}^N x_t^2} + 2p^2 \frac{\sum_{t=1}^{N-2} x_t x_{t+2}}{\sum_{t=1}^N x_t^2} + \dots + 2p^{n-1} \frac{x_1 x_t}{\sum_{t=1}^N x_t^2} \right] \end{aligned} \quad (65)$$

en el caso de que t sea elevado la expresión entre paréntesis es aproximadamente igual

$$\begin{aligned} (2 + 2pr + 2p^2r^2 + \dots - 1) &= \frac{1 + pr}{1 - pr} \\ &= \frac{\sigma_\varepsilon^2}{\sum_{t=1}^N x_t^2} \left[\frac{1 + pr}{1 - pr} \right] \end{aligned} \quad (66)$$

(Répase que los términos que contienen x dentro del paréntesis son estimaciones de los coeficientes de autocorrelación de x)

De esta forma, nos encontramos que si se aplica el procedimiento de los MCO para estimar la $\text{VAR}(\hat{\beta})$, no tenemos en cuenta el término que va en paréntesis; con lo que si p es positivo y la variable independiente está autocorrelacionada la expresión $[1 + pr/1 - pr]$ será casi seguro mayor que uno, subestimándose así la verdadera varianza de $\hat{\beta}$. Cuando la variable independiente no está autocorrelacionada ($r \cong 0$), como ocurriría con un diseño de series temporales interrump-

pidas en el cual el tratamiento se presentase y se extrajese aleatoriamente, lo cual no suele ocurrir, el sesgo en la estimación de la varianza de $\hat{\beta}$ estaría minimizado, aun cuando el valor de la autocorrelación de los errores fuese importante (Judd y Kenny, 1981; Cook, Dintzer y Mark, 1980).

Sin embargo, como señala Hibbs (1974), la situación puede resultar aún peor que la referida, pues $\sigma_{\hat{\epsilon}}^2$ también infraestima el verdadero valor σ_{ϵ}^2 . Para demostrar esto recordemos que

$$E(\hat{\epsilon}'\hat{\epsilon}) = \sigma_{\epsilon}^2 [n - \text{tr} [(X'X)^{-1}X'X]] = \sigma^2 (n - h) \quad (\text{véase ecuación 29}) \quad (67)$$

mientras que en el caso en que los errores están serialmente correlacionados el valor esperado de $\epsilon'\epsilon$ era

$$E(\hat{\epsilon}'\hat{\epsilon}) = \sigma_{\epsilon}^2 n - \text{tr} [(X'X)^{-1}X'\Omega X] \quad (\text{véase ecuación 31}) \quad (68)$$

por lo tanto, si partimos del mismo modelo de regresión simple y aplicamos el mismo procedimiento seguido anteriormente para calcular el sesgo que se comete al aplicar mínimos cuadrados ordinarios para estimar la $\text{VAR}(\beta)$, el valor esperado de $(\hat{\epsilon}'\hat{\epsilon})$ no es aproximadamente igual a $\sigma^2(n-1)$ sino a

$$E(\hat{\epsilon}'\hat{\epsilon}) \cong \sigma_{\epsilon}^2 \left[n - \frac{1 + \rho r}{1 - \rho r} \right] \quad (69)$$

En consecuencia, en casos como el considerado, existe una doble infravaloración de las varianzas muestrales si se aplica el procedimiento de los mínimos cuadrados ordinarios sin corregirse la autocorrelación.

Pruebas usuales para estimar la dependencia serial

La presencia de dependencia serial en los residuales plantea un problema tan serio para el uso de los mínimos cuadrados ordinarios que es de la máxima urgencia antes de analizar los datos ejecutar alguna prueba para detectar su presencia. Además, siempre que nos sea posible, una vez que hemos verificado la presencia de autocorrelación deberíamos de llevar a cabo una serie de pruebas más detalladas para poder ajustar lo más exactamente posible el tipo y el orden de correlación serial. Como ya hemos señalado, la dependencia serial no siempre sigue el mismo proceso, de aquí que lo que hagamos con ella dependerá de cual sea nuestra hipótesis acerca de ella. Los procesos que pueden ser generados por la autocorrelación pueden ser diversos (autorregresivo, de media móvil, etc.); no obstante, en el caso de que dispongamos de muestras reducidas en exceso, una forma muy sencilla, pero muy común es asumir que los residuales siguen un proceso autorregresivo de primer orden. Esto es análogo a presuponer, por ejemplo,

que la varianza (σ_1^2 es proporcional a la puntuación correspondiente al cuadrado, es decir $\sigma_1^2 = \sigma_1^2 X_{n1}^2$).

En la actualidad, diversas medidas de la dependencia serial se hallan disponibles. En econometría se emplea con mucha frecuencia un test paramétrico que es válido para muestras reducidas y un alto número de regresores conocido con el nombre «d» de Durbin-Watson y definido como:

$$d = \frac{\sum_{t=2}^N (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^N \hat{\epsilon}_t^2} \quad (70)$$

Con todo esta medida, aunque útil, no es tan general como otra medida de dependencia denominada autocorrelación; la cual para retardos de orden uno es definida como:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^N (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (71)$$

Nosotros por razones metodológicas y prácticas recomendamos computar la autocorrelación desde los residuales y no desde las observaciones directas. Dado que la media de ϵ_t es cero, la ecuación anterior nos queda definida como sigue:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^N \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^N \hat{\epsilon}_t^2} \quad (72)$$

Para una justificación extensa de por qué recomendamos computar la autocorrelación desde los residuales obtenidos por MCO véase Vallejo, 1986a; Gorsuch, 1983; Levin, Marascuilo y Hubert, 1978.

Si desarrollásemos la relación existente entre «d» y el coeficiente de correlación serial de primer orden ρ , viene dada por

$$d \cong 2(1 - \hat{\rho}) \quad (73)$$

Esta última expresión nos indica que el estadístico «d» varía entre 0 y 4, de manera que cuando ρ se aproxima a 0 «d» se aproxima a 2, cuando $\hat{\rho}$ se aproxima a -1 «d» se aproxima a 4 y cuando $\hat{\rho}$ se aproxima a 1 «d» se aproxima a 0.

Dado que la distribución del estadístico «d» depende del número de varia-

bles independientes del valor, que tomen éstas y del tamaño de la muestra, lo anterior constituye una aproximación. No obstante, dada una muestra de tamaño N y h variables independientes bajo la hipótesis nula de que las perturbaciones aleatorias del modelo están distribuidas independientemente, el estadístico « d » siempre está comprendido dentro de los límites de otros dos estadísticos para un determinado nivel de significación de « d », denominados límite superior e inferior (en adelante d_s y d_i). Como se muestra en la figura 1 con este procedimiento para probar la significación del estadístico « d » existen tres áreas de decisión, dos de las cuales permiten hacer inferencias acerca de la significación de la autocorrelación existente, mientras que en una tercera que se encuentra entre las dos anteriores no hay información suficiente para aceptar o rechazar la hipótesis nula.

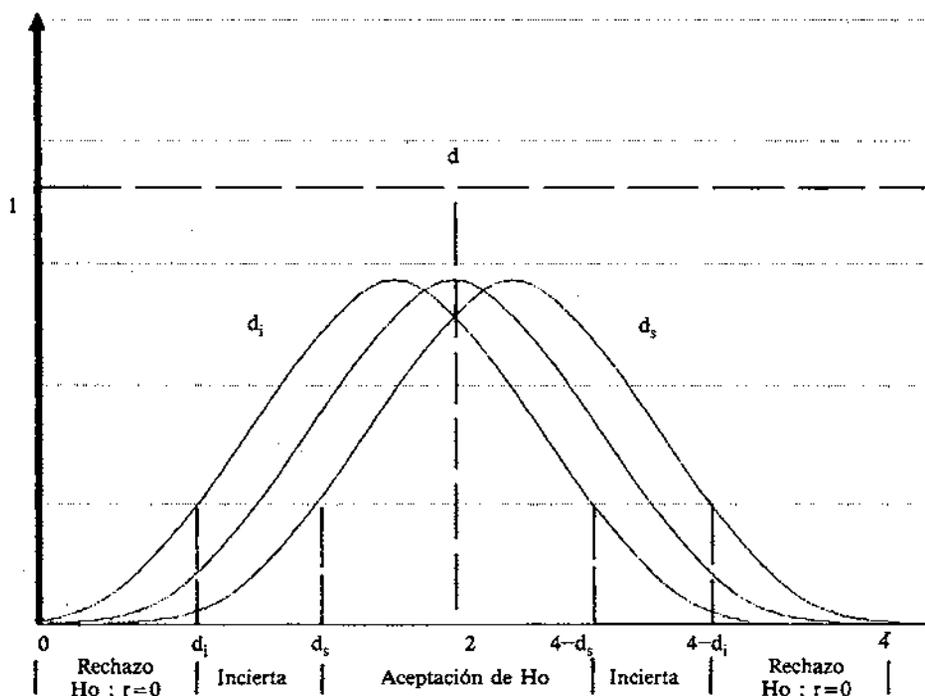


FIGURA 1. PUNTOS CRÍTICOS DEL CONTRASTE DE DURBIN-WATSON.

Kmenta (1971, p. 295) resume del modo siguiente las reglas de decisión que nos permiten probar la hipótesis nula (H_0) de no autocorrelación frente a la hipótesis alternativa (H_1) de correlación serial positiva.

Si « d » < « d_i » rechazamos H_0 y aceptamos H_1

Si « d_s » < « d » < « $4-d_s$ » aceptamos H_0 y rechazamos H_1

Si « d_1 » \leq « d » \leq « d_S » La decisión cae dentro de la zona incierta

Las reglas de decisión para probar H_0 de no autocorrelación frente a la H_1 de correlación serial negativa son:

Si « $4-d_1$ » $<$ « d » Rechazamos H_0 y aceptamos H_1

Si « d_S » $<$ « d » $<$ « $4-d$ » Aceptamos H_0 y rechazamos H_1

Si « $4-d_S$ » \leq « d » \leq « $4-d_1$ » La decisión cae dentro de la zona incierta.

Los valores críticos de « d_1 » y « d_S » para probar la hipótesis nula de ausencia de autocorrelación ($H_0: \hat{\rho} = 0$) frente a las alternativas ($H_1: \hat{\rho} > 0$ y $H_1: \hat{\rho} < 0$) pueden consultarse en las tablas de Savin y White (1977).

Estas tablas nos proporcionan los valores críticos de « d_1 » y « d_S » a un nivel de significación del 5% y 1% para muestras entre 6 y 200 observaciones que varían según la complejidad del modelo. Si sólo es estimado un parámetro exceptuando la constante de intercepción, entonces la primera columna de la tabla debería de ser usada. Las tablas referidas permiten evaluar el estadístico « d » hasta con veintiún regresores.

Finalmente, señalar que también existen una serie de pruebas no paramétricas basadas en el signo de los residuales. Algunas de las más comunes son un test de χ^2 «tabla de contingencia de 2×2 » de signos positivos y negativos en el tiempo « T » y « $T-1$ » de Griliches y otros (1962) y el test de cambio de signo sugerido por Geary (1963). Esta prueba implica un simple recuento del número de cambios de signo de los residuales obtenidos por el procedimiento de los mínimos cuadrados. En orden a probar la hipótesis nula se deben computar un número máximo y mínimo de cambios de signo. Para contrastar estos cambios existen unas tablas de referencia con valores críticos al 5% y 1% elaborados por Habibagahi y Pratschke (1972).

Estimación en presencia de autocorrelación

Una vez que hemos presentado los problemas que se derivan de la presencia de autocorrelación, así como los procedimientos para detectarla, debe procederse a eliminar la autocorrelación de nuestras observaciones. Para ello dos soluciones van a ser ilustradas: Solución por MCG y solución mediante diferencias generalizadas.

Consideremos en primer lugar el procedimiento de los MCG. Para ello partamos del modelo

$$y = \beta X + \varepsilon$$

$$\text{con } E(\varepsilon) = 0 \text{ y } E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \Omega \quad \left| \begin{array}{l} E(\varepsilon_t \varepsilon_{t-s}) = \sigma_\varepsilon^2 ; s = 0 \text{ homocedasticidad} \\ E(\varepsilon_t \varepsilon_{t-s}) = \sigma_\varepsilon^2 \rho^s ; s \neq 0 \text{ autocovarianza} \end{array} \right. \quad (74)$$

La aplicación del procedimiento MCG para un proceso AR(1) exige partir de una matriz de covarianza para los errores correlacionados que toma, como se desprende de la función de autocorrelación $p_s = p^s$, la siguiente forma:

$$E(\epsilon\epsilon') = \sigma_\epsilon^2 \begin{bmatrix} 1 & p & p^2 & \dots & p^{t-1} \\ p & 1 & p & \dots & p^{t-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p^{t-1} & p^{t-2} & p^{t-3} & \dots & 1 \end{bmatrix} = \sigma_\epsilon^2 \Omega \tag{75}$$

Con la matriz de covarianza para ϵ no es escalar los MCO no son eficientes y, por consiguiente, los estimadores ELIO de β se obtienen al aplicar MCG. Así pues, si se sabe que el error sigue un proceso AR(1) y conocemos el valor de p , el procedimiento de los MCG puede ser aplicado directamente sin pérdida de eficacia. Sin embargo, recuérdese que la solución MCG implica:

$$\begin{aligned} \tilde{\beta} &= (X'\Omega^{-1}X)^{-1} \Omega^{-1}y \\ \text{VAR}(\tilde{\beta}) &= \sigma_\epsilon^2 (X'\Omega^{-1}X)^{-1} \\ \sigma_\epsilon^2 &= \frac{\tilde{\epsilon}'\Omega^{-1}\tilde{\epsilon}}{N - k} \end{aligned} \tag{76}$$

y puesto que nosotros sólo disponemos de Ω necesitamos por ello calcular Ω^{-1} . Por consiguiente, dado que $\Omega\Omega^{-1} = I$ es prontamente verificado que:

$$\Omega^{-1} = \frac{1}{1-p^2} \begin{bmatrix} 1 & -p & 0 & \dots & 0 & 0 & 0 \\ -p & 1+p^2 & -p & \dots & 0 & 0 & 0 \\ 0 & -p & 1+p^2 & \dots & 0 & 0 & 0 \\ \vdots & \dots & \dots & \ddots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -p & 1+p^2 & -p \\ 0 & 0 & 0 & 0 & -p & 1 & 1 \end{bmatrix} \tag{77}$$

por lo que los estimadores lineales insesgados óptimos de β estarán ahora asegurados mediante la inserción de Ω^{-1} en las soluciones MCG anteriores.

Un modelo alternativo, y tal vez más conveniente, ya que evita las pesadas operaciones con matrices, consiste en operar transformaciones en el modelo de la ecuación anterior, de modo que la matriz de covariaciones de los errores transformados sea escalar, satisfaciendo con ello el modelo de MCO. Esto es, se trata de buscar una matriz de transformación P , tal que:

$$\begin{aligned} p'p &= \Omega^{-1} \\ p \Omega p' &= I \end{aligned} \tag{78}$$

(la matriz de transformación P con las propiedades descritas existe siempre que Ω sea una matriz definida positiva, ver Johnston (1984, cap. 8)).

Así pues, si premultiplicamos el modelo $y = \beta X + \varepsilon$ por la matriz de transformación no singular P , y aplicamos mínimos cuadrados a las variables transformadas, los estimadores obtenidos son equivalentes a las estimaciones MCG.

En efecto, consideremos la matriz que cumple con la relación $P'P = \Omega^{-1}$, entonces P de orden $(t \times t)$:

$$P_1 = \begin{bmatrix} \sqrt{1-p^2} & 0 & 0 & \dots & 0 & 0 \\ -p & 1 & 0 & \dots & 0 & 0 \\ 0 & -p & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & -p & 1 \end{bmatrix} \quad (79)$$

se puede comprobar fácilmente que la relación

$$Py = PX\beta + P\varepsilon \quad (80)$$

tiene una matriz de covarianza para el término de error que satisface las asunciones del modelo MCO

$$\begin{aligned} E(P\varepsilon\varepsilon'P') &= P\sigma_\varepsilon^2\Omega P' \\ &= \sigma_\varepsilon^2 P\Omega P' \\ &= \sigma_\varepsilon^2 I \end{aligned} \quad (81)$$

Así pues, los MCO se pueden aplicar a los datos transformados $Py = Y_*$ y $PX = X_*$ que son:

$$Y_* = \begin{bmatrix} \sqrt{1-p^2} & y_1 \\ y_2 - p & y_1 \\ \vdots & \vdots \\ y_t - p & y_{t-1} \end{bmatrix}; X_* = \begin{bmatrix} \sqrt{1-p^2} & \sqrt{1-p^2}x_{1,1} & \sqrt{1-p^2}x_{1,h-1} \\ 1-p & x_{2,1} - px_{1,1} & x_{2,h-1} - px_{1,h-1} \\ \vdots & \vdots & \vdots \\ 1-p & x_{t,1} - px_{t-1,1} & x_{t,h-1} - px_{t-1,h-1} \end{bmatrix} \quad (82)$$

Con lo que se puede ahora comprobar que las puntuaciones quedan en forma de diferencias generalizadas y, que la aplicación de los mínimos cuadrados ordinarios a estas diferencias generalizadas coincide con el estimador MCG del modelo original.

$$\begin{aligned}
 \beta^{\text{MCO}} &= (X_*' X_*)^{-1} X_*' Y_* \\
 &= [(PX)' (PX)]^{-1} (PX)' PY \\
 &= [X' (P'P) X]^{-1} X' (P'P) Y \\
 \beta^{\text{MCG}} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y
 \end{aligned}
 \tag{83}$$

Idealmente, para proceder directamente con la solución MCG, haciendo así la transformación de las variables innecesarias, necesitamos conocer el proceso generador del error y su valor. En nuestro caso particular el modelo de dependencia serial es autorregresivo de primer orden y el verdadero valor del parámetro generador de errores autocorrelacionados; sin embargo, en la mayor parte de los casos, no se conoce ni el orden de la estructura de la autocorrelación, ni el valor del parámetro y, por tanto, las estimaciones mínimo cuadráticas generalizadas no se pueden calcular directamente, a no ser que introduzcamos técnicas de estimación no lineales. Para evitar la dificultad que supone tal introducción han sido sugeridos procedimientos de estimación de dos etapas (Gotman, 1981). La primera etapa incluye la obtención de una estimación p y con ello de la forma de Ω mediante el método de los mínimos cuadrados. En la segunda estimamos el valor de β usando el método de los mínimos cuadrados generalizados. Este procedimiento es denominado por Hibbs (1974) estimación pseudo-generalizada.

Por último, hacer hincapié en que el procedimiento en dos etapas de las diferencias generalizadas también exige conocer el valor de p para transformar las observaciones originales y aplicar posteriormente MCO a los datos transformados; aspecto éste para el cual se han propuesto distintos métodos (véase Vallejo, 1986b para una presentación de los procedimientos más difundidos).

Conclusiones finales

Si bien son muchos los aspectos merecedores de un breve comentario final, en nuestra opinión hay dos que deben ser resaltados, pues además de haber capitalizado gran parte de nuestro trabajo, forman parte del doble objetivo que nos proponíamos en su inicio.

En primer lugar, señalar que siempre que nuestros datos sigan el modelo $y = X\beta + \varepsilon$ y la matriz de covarianzas no se reduzca a un único escalar (ausencia de varianzas constantes y/o de covarianzas nulas) con la aplicación del procedimiento de los MCO aparecen una serie de problemas que podemos resumir como sigue:

a) Los estimadores obtenidos por MCO aunque insesgados, ya que la matriz de covarianzas Ω no aparece en el cálculo de los valores esperados, son menos eficientes que en el caso de que la matriz de covarianzas fuese escalar.

b) Como consecuencia, las varianzas muestrales de los estimadores estarán sesgadas, por lo general se hallarán subestimados, aunque esto como es lógico dependerá de la forma de Ω . En cualquier caso, el procedimiento ya no es válido.

do, como tampoco lo son las predicciones que efectuemos, ni los contrastes habituales de t y F al asumirse una distribución que no cumple con las propiedades básicas.

En segundo lugar, siempre que los datos sigan un modelo $y = X\beta + \varepsilon$ y la matriz de covarianza no sea escalar, el procedimiento de los mínimos cuadrados generalizados consigue que las estimaciones realizadas a través de él gocen de las mismas propiedades ELIO que las de los estimadores MCO cuando la matriz de covarianzas de los errores es escalar.

RESUMEN

Actualmente un fenómeno cada vez más frecuente es el empleo de procedimientos analíticos para indagar acerca de los efectos de las intervenciones en estudios de carácter longitudinal. Por este motivo, en el presente trabajo además de considerar el problema de obtener estimadores eficientes para los parámetros de un sistema de M ecuaciones de regresión para medidas seriales igualmente espaciadas, mostramos cómo la significación estadística de los coeficientes de regresión es mucho más realista cuando la estructura del error es aproximadamente modelada mediante algún proceso ARMA.

SUMMARY

Presently, the use of analytic procedures to study the effects of interventions in longitudinal designs is becoming increasingly frequent. That's why in this paper, besides to consider the problem of obtaining efficient estimators for a system with M regression equations for equally spaced serial measurements, we show how the statistical significance of regression coefficients is much more realistic when the error structure is approximately modeled using an ARMA process.

REFERENCIAS

- Aiken, A.C. (1935). On least squares and linear combination of observation. *Proceedings of the Royal Society*, Vol. 35, 42-48.
- Berry, W.D. and Lewis-Beck, M.S. (1986). Interrupted time series. En Berry and Lewis-Beck (Eds.): *New tools for Social Scientist*. London: Sage, 209-241.
- Box, G.E.P. y Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Box, G.E.P. y Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. Segunda edición revisada. San Francisco: Holden-Day.
- Cook, T.; Dintzer, L. y Mark, M. (1980). The causal analysis of concomitant times series. En L. Bickman (Ed.). *Applied Social Psychology Annual*. London: Sage, 93-135.
- Engle, R.F. (1973). Specification of the disturbance for efficient estimation. *Econometrica*, 41.

- Geary, R. (1970). Relative efficiency of count sign changes for assessing residual autoregression in least squares regression. *Biometrika*, 57, 123-127.
- Goldfeld, S.M. y Quandt, R.E. (1972). *Nonlinear Methods in Econometrics*. North-Holland.
- Gorsuch, R.L. (1983). Three methods for analysing limited time series (N=1) data. *Behavioral Assessment*, 5, 141-154.
- Gottman, J.M. (1981). *Time series analysis: A comprehensive introduction for social scientists*. Cambridge University Press.
- Habibagahi, H. y Pratschke, J. (1972). A comparison on the power of the Von Neumann ratio Durbin-Watson and Geary tests *The review of Economics and Statistics*. 179-185.
- Hibbs, D.A. (1974). Problems of statistical estimation and causal inference in time-series regression models. En H.2. Costner (Ed.), *Sociological methodology*. San Francisco: Josse y Bass.
- Johnston, J. (1984). *Econometric Methods*. McGraw-Hill (Traducción de la obra por la editorial Vicens Vives, 1987).
- Judd, Ch. y Kenny, D. (1981). Estimating the effects of social interventions. *Cambridge University Press*.
- Kelejian, H. y Oates, W.E. (1974): *Introduction to Econometrics: Principles and Applications*. Harper and Row.
- Kirk, R.E. (1982). *Experimental Design*. Brooks-Cole.
- Kmenta, J. (1971). *Elements of econometrics*. N. York: Mcmillan (traducción de la obra en la editorial Vicens Vives, 1980).
- Levin, J.R.; Marascuilo, L.A. y Hubert, L.J. (1978). N=nonparametric randomization tests. En Kratochwill (Ed.): *Single subject research: Strategies for Evaluating Change*. Academic Press.
- Maddala, G.S. (1977). *Econometrics*. N. York: McGraw-Hill (traducido de la primera edición en inglés por la editorial McGraw-Hill).
- Pindyck, R.S. y Rubinfeld, D.L. (1976). *Econometric Models and Economic Forecasts*. Mc Graw-Hill.
- Pulido, A. (1986). *Modelos Económicos*. Editorial Pirámide.
- Rao, P. y Griliches, Z. (1969). Small-sample properties of several two stage regression methods in the autocorrelated errors. *Journal of the American Statistical Association*, 64, 253-272.
- Savin, N.E. y White, K.J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, 45, 1989-1996.
- Vallejo, G. (1986). Aplicación de análisis de series temporales con diseños N=1: Consideraciones generales. *Revista Española de Terapia del Comportamiento*, 4 (1), 1-30.
- Vallejo, G. (1986). Procedimientos simplificados de análisis de las series temporales interrumpidas: Modelos estáticos. *Revista Española de Terapia del Comportamiento*, 4 (2), 114-148.
- Wonnacott, R.J. y Wonnacott, T.H. (1980). *Econometrics*. John-Wiley (traducción de la segunda edición por la editorial Aguilar, 1982).

