

## Autocorrelación

### 10.1. Introducción

En el modelo lineal general con perturbaciones no esféricas

$$(10.1) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Omega})$$

el estimador lineal insesgado de menor varianza del vector de parámetros  $\boldsymbol{\beta}$  es el estimador de mínimos cuadrados generalizados

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

con distribución muestral

$$\hat{\boldsymbol{\beta}}_{MCG} \sim N(\boldsymbol{\beta}, \sigma_u^2 (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1})$$

siendo el estimador insesgado del escalar  $\sigma_u^2$

$$\hat{\sigma}_{u,MCG}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{T - k}$$

en donde  $T$  y  $k$  son el número de observaciones y parámetros, respectivamente.

Un caso especial del modelo (10.1) aparece, con datos de series temporales, cuando relajamos el supuesto de **ausencia de autocorrelación** especificando un proceso estocástico estacionario e invertible para las perturbaciones. En este tema ampliamos el modelo lineal general

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t, \quad t = 1, \dots, T$$

especificando un proceso estocástico  $ARMA(p, q)$  para las perturbaciones

$$u_t = \phi_1 u_{t-1} + \cdots + \phi_p u_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

en donde suponemos que el polinomio autorregresivo  $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$  es estacionario y el polinomio de medias-móviles  $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$  es invertible; en otras palabras, las raíces de  $\phi(B) = 0$  y  $\theta(B) = 0$  son mayores que uno en valor absoluto.

Ahora, la matriz de varianzas y covarianzas de  $\mathbf{u}$  tendrá la forma

$$V(\mathbf{u}) = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} = \gamma_0 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{pmatrix} = \sigma_u^2 \boldsymbol{\Omega}$$

en donde las autocovarianzas  $\gamma_k = cov(u_t, u_{t-k})$  o autocorrelaciones  $\rho_k = \gamma_k / \gamma_0$  (y de aquí,  $\sigma_u^2 = \gamma_0$  y  $\boldsymbol{\Omega}$ ) dependerán de los parámetros  $\phi_i$  ( $i = 1, \dots, p$ ) y  $\theta_i$  ( $i = 1, \dots, q$ ).

Conviene notar que, cuando los coeficientes  $\phi_i$  y  $\theta_i$  son todos iguales a cero, la matriz  $V(\mathbf{u})$  será escalar  $V(\mathbf{u}) = \gamma_0 \mathbf{I}_T$ .

Son dos los objetivos principales que perseguimos en este tema. En primer lugar, presentamos algunos métodos que nos permiten no sólo detectar si los errores del modelo lineal general presentan autocorrelación, sino también identificar los órdenes  $p$  y  $q$  del proceso ARMA subyacente. En segundo lugar, describimos un método de estimación computacionalmente eficiente bajo autocorrelación, que evita la inversión de la matriz  $\mathbf{\Omega}$  para calcular el estimador MCG. Debemos tener en cuenta que la inversión de una matriz cuya dimensión depende del tamaño muestral es una operación que puede llevar mucho tiempo (*time-consuming*).

Conviene notar que la autocorrelación serial en las perturbaciones puede ser una consecuencia de un error de especificación debido a la omisión de variables relevantes que presentan autocorrelación. Por ejemplo, si el modelo correcto para explicar  $Y_t$  es

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

y erróneamente especificamos el modelo

$$Y_t = \beta_1 + \beta_2 X_{2t} + e_t$$

entonces el término de error  $e_t = \beta_3 X_{3t} + u_t$  depende la variable omitida  $X_{3t}$ . Si los datos de esta variable son dependientes en el tiempo, entonces el error  $e_t$  presentará autocorrelación.

Un problema similar aparece cuando erróneamente especificamos modelos estáticos en lugar de modelos dinámicos. Por ejemplo, consideremos el modelo de retardos distribuidos

$$(10.2) \quad Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_s X_{t-s} + u_t$$

que podemos ampliar incluyendo retardos de  $Y_t$

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_r Y_{t-r} + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_s X_{t-s} + u_t$$

y que se denomina modelo de retardos distribuidos con autocorrelación, denotado por  $ADL(r, s)$ . Si la verdadera relación entre las variables  $Y_t$  y  $X_t$  viene dada por un modelo  $ADL(r, s)$ , pero erróneamente especificamos el modelo estático

$$Y_t = \beta_0 X_t + e_t$$

entonces los errores  $e_t$  estarán serialmente correlacionados. A veces se utiliza la expresión **correlación serial impura** para hacer referencia a la autocorrelación causada por la omisión de variables relevantes, y se habla de **correlación serial pura** cuando el término de error sigue una proceso  $ARMA(p, q)$ .

## 10.2. Detección de autocorrelación

Como las perturbaciones del modelo lineal general no son observables, debemos basar los métodos de detección en el mejor estimador disponible de  $u_t$ : los residuos de la estimación de mínimos cuadrados. Tales métodos pueden ser de dos tipos: métodos gráficos y contrastes de hipótesis. Aunque los segundos se consideran a menudo más formales que los primeros en términos de propiedades estadísticas, veremos que el análisis

de diferentes gráficos basados en los residuos de un modelo no sólo es útil en la detección de autocorrelación, sino también para comprender la racionalidad y limitaciones de los contrastes estadísticos.

### 10.2.1. Métodos gráficos.

*10.2.1.1. Gráfico temporal.* El gráfico temporal de los residuos es una herramienta muy útil para detectar la presencia de autocorrelación. Si los residuos están incorrelacionados, entonces se deben distribuir aleatoriamente por encima y por debajo de su media, que será cero si el modelo incluye un término constante. En este caso, el conocimiento de los residuos previos a un instante  $t$  no nos ayuda a predecir si el residuo  $\hat{u}_t$  se encuentra por encima o por debajo de la media. Por el contrario, si los residuos están positivamente correlacionados, entonces observaremos en el gráfico temporal rachas de residuos por debajo y por encima de la media, situación que se ilustra en el primer gráfico de la figura 1. Cuando los residuos estén negativamente correlacionados, el gráfico temporal revelará una alternancia en el signo de los residuos, que es la situación mostrada en el segundo gráfico de la figura 1.

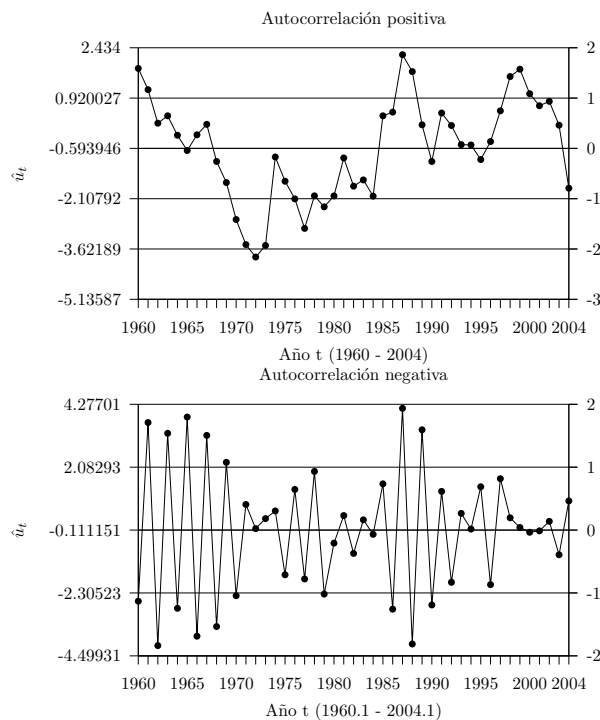


Figura 1: Gráfico temporal

*10.2.1.2. Gráficos de dispersión.* El gráfico de dispersión de  $\hat{u}_t$  frente a  $\hat{u}_{t-1}$  puede utilizarse para comprobar si existe autocorrelación positiva o negativa de primer orden. La figura 2 muestra los gráficos de dispersión correspondientes a los residuos dibujados en la figura 1. Este gráfico de dispersión puede construirse para otros retardos,  $\hat{u}_{t-k}$ .

*10.2.1.3. Función de autocorrelación simple muestral.* El gráfico de dispersión de  $\hat{u}_t$  frente a  $\hat{u}_{t-k}$  puede resumirse estimando la ecuación de regresión simple sin término constante

$$(10.3) \quad \hat{u}_t = \phi_k \hat{u}_{t-k} + e_t \quad t = k + 1, \dots, T$$

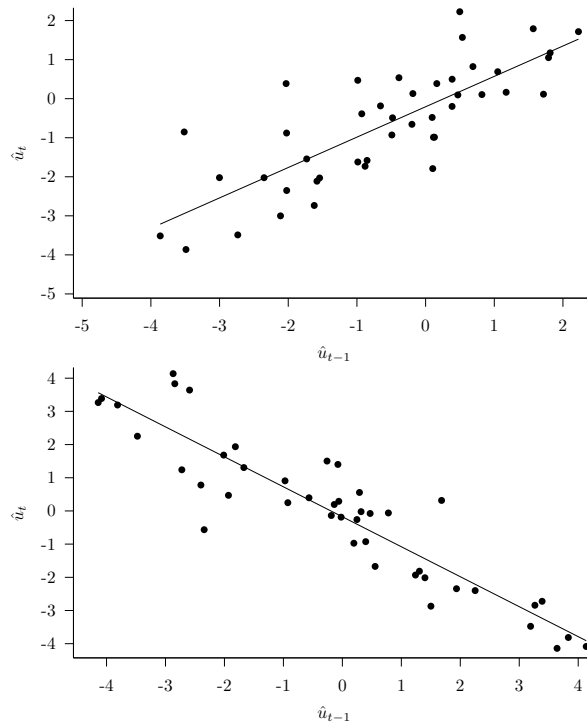


Figura 2: Gráfico de dispersión

El estimador de mínimos cuadrados del parámetro  $\phi_k$  es

$$\hat{\phi}_k = \frac{\sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k}}{\sum_{t=k+1}^T \hat{u}_{t-k}^2}$$

que aproximadamente coincide con el coeficiente de correlación simple muestral en el retardo  $k$

$$\hat{\phi}_k \simeq \hat{\rho}_k = \frac{\sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k}}{\sum_{t=1}^T \hat{u}_t^2}$$

que toma valores entre -1 y 1.

Estimando la ecuación de regresión simple (10.3) para diferentes retardos  $k = 1, 2, \dots, K$ , obtenemos una serie de coeficientes de correlación simple  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_K$ , cuya representación en términos de  $k$  se denomina función de autocorrelación simple (ACF) muestral o correlograma.

Para contrastar la hipótesis nula  $\phi_k = 0$  frente a la alternativa de que  $\phi_k \neq 0$  construimos la ratio  $t$

$$t = \frac{\hat{\phi}_k}{\sqrt{\text{Var}(\hat{\phi}_k)}}$$

que asintóticamente sigue una distribución  $N(0, 1)$ , siendo  $\text{Var}(\hat{\phi}_k) = 1/T$ . La hipótesis se rechaza si

$$\left| \frac{\hat{\phi}_k}{\sqrt{\text{Var}(\hat{\phi}_k)}} \right| > z_{\alpha/2} \Leftrightarrow |\hat{\phi}_k| > z_{\alpha/2} \sqrt{\text{Var}(\hat{\phi}_k)}$$

en donde  $z_{\alpha/2}$  es el valor crítico tal que  $\text{Prob}(N(0, 1) > z_{\alpha/2}) = \alpha/2$ .

Cuando  $T > 30$ , se usa la regla  $|\hat{\phi}_k| > 2/\sqrt{T}$ . La figura (3) muestra dos correlogramas con las bandas  $\pm 2/\sqrt{T}$ . Mientras que los coeficientes que se encuentran fuera de las bandas de significación se consideran estadísticamente distintos de cero, los que están dentro de bandas son estadísticamente iguales a cero. Se utiliza la expresión *el correlograma se corta en el retardo k* para indicar que las correlaciones se anulan a partir del retardo  $k$ . Vemos que el primer correlograma se corta en el retardo 5; y el segundo, en el retardo 6.

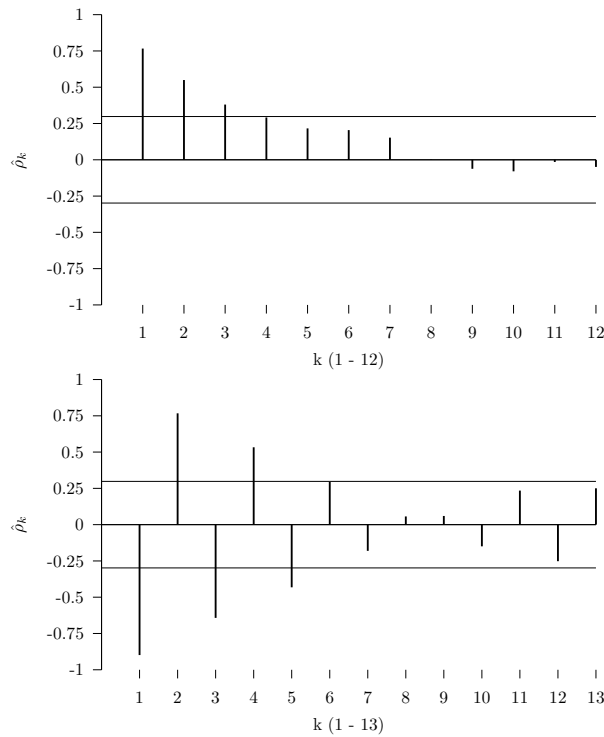


Figura 3: Función de autocorrelación simple muestral

*10.2.1.4. Función de autocorrelación parcial muestral.* La ACF muestral resume los gráficos de dispersión de  $\hat{u}_t$  frente a  $\hat{u}_{t-k}$ . Sus coeficientes son aproximadamente los coeficientes estimados en ecuaciones de regresión simple. Un complemento de esta función es la función de autocorrelación parcial (PACF) muestral que se construye estimando sucesivas ecuaciones de regresión múltiple:

$$\begin{aligned}\hat{u}_t &= \phi_{11}\hat{u}_{t-1} + e_t \\ \hat{u}_t &= \phi_{21}\hat{u}_{t-1} + \phi_{22}\hat{u}_{t-2} + e_t \\ \hat{u}_t &= \phi_{31}\hat{u}_{t-1} + \phi_{32}\hat{u}_{t-2} + \phi_{33}\hat{u}_{t-3} + e_t \\ &\vdots \\ \hat{u}_t &= \phi_{k1}\hat{u}_{t-1} + \dots + \phi_{kk}\hat{u}_{t-k} + e_t\end{aligned}$$

La serie de coeficientes  $\phi_{11}, \phi_{22}, \dots, \phi_{kk}$  se denomina PACF muestral. Esta terminología proviene de la interpretación de los coeficientes de la regresión múltiple como coeficientes

de regresión parcial. Los coeficientes  $\sqrt{T}\phi_{kk}$  también siguen asintóticamente una distribución  $N(0,1)$ . Por tanto, la hipótesis nula  $\phi_{kk} = 0$  se rechaza frente a la alternativa  $\phi_{kk} \neq 0$ , cuando  $|\hat{\phi}_{kk}| > 2/\sqrt{T}$ .

Conviene notar que la PACF muestral no sólo nos indica si existe autocorrelación, sino también el orden de la misma. Si los residuos siguen un proceso  $AR(k)$ , entonces la PACF muestral se corta en el retardo  $k$ . En la figura (4), las dos PACF se cortan en el primer retardo, lo que indica autocorrelación del tipo  $AR(1)$  con parámetro positivo en el primer caso y negativo en el segundo.

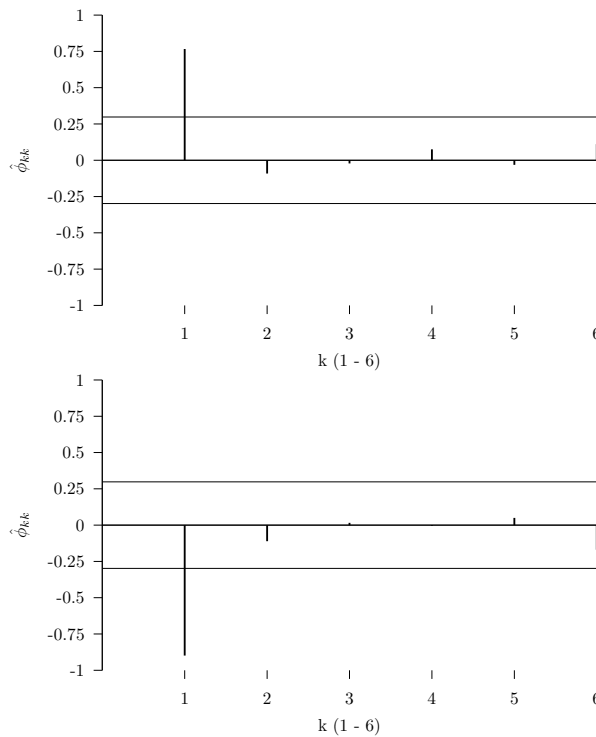


Figura 4: Función de autocorrelación simple

**10.2.2. Contrastes formales.**

10.2.2.1. *Test de rachas.* El problema a contrastar es

$$H_0 : u_t \sim iid$$

$$H_1 : u_t \not\sim iid$$

Bajo  $H_0$  las perturbaciones se distribuyen idéntica e independiente, mientras que bajo  $H_1$  las perturbaciones no se distribuyen idéntica e independientemente.

Tomando de referencia el primer gráfico temporal de la figura (1), podemos representar los residuos que se encuentran por encima de la media con el símbolo + y los que están por debajo con el símbolo -

+++++--++-----+++++--+++++--+++++--+++++--

Si definimos una racha como una serie de observaciones con el mismo signo, entonces tenemos  $n = 10$  rachas, de las cuales  $n_1 = 5$  son positivas y  $n_2 = 5$  son negativas.

El número de rachas  $n$  es una variable aleatoria asintóticamente normal con media

$$\mu = E(n) = \frac{2n_1n_2}{n} + 1$$

y varianza

$$\sigma^2 = V(v) = \frac{(\mu - 1)(\mu - 2)}{n - 1}$$

La región crítica del test de rachas es

$$\left| \frac{n - \mu}{\sigma} \right| > z_{\alpha/2}$$

10.2.2.2. *El contraste de Durbin-Watson.* Estos autores consideraron un proceso autorregresivo de primer orden AR(1) para las perturbaciones

$$u_t = \phi_1 u_{t-1} + \epsilon_t, \quad \epsilon_t \sim iidN(0, \sigma_\epsilon^2)$$

y formularon el problema de contraste

$$H_0 : \phi_1 = 0$$

$$H_1 : \phi_1 > 0$$

en donde la hipótesis nula de ausencia de correlación serial se contrasta frente a la alternativa de un lado de correlación positiva. La hipótesis alternativa puede reemplazarse por  $H_1 : \phi_1 < 0$ , si sospechamos que la autocorrelación de primer orden puede ser negativa.

DEFINICIÓN 86. *El estadístico de Durbin-Watson se define como la ratio entre la suma de los cuadrados de las diferencias sucesivas y la suma de cuadrados de los residuos*

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}$$

Observación 59. *El estadístico DW puede contemplarse como la ratio entre el cuadrado medio las diferencias sucesivas y la varianza de los residuos, también conocida con la ratio de von Neumann.*

PROPOSICIÓN 99. *El estadístico DW es aproximadamente igual a  $2(1 - \hat{\rho}_1)$ , en donde  $\hat{\rho}_1$  es el coeficiente de autocorrelación simple muestral en el retado 1.*

DEMOSTRACIÓN.

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2} = \frac{\sum_{t=2}^T \hat{u}_t^2 + \sum_{t=2}^T \hat{u}_{t-1}^2 - 2 \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2}$$

Cuando el número de datos  $T$  es grande, entonces las magnitudes

$$\sum_{t=2}^T \hat{u}_t^2 \simeq \sum_{t=2}^T \hat{u}_{t-1}^2 \simeq \sum_{t=1}^T \hat{u}_t^2$$

De modo que

$$DW \simeq \frac{2 \sum_{t=1}^T \hat{u}_t^2 - 2 \sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2} = 2 - 2 \frac{\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2} = 2(1 - \hat{\rho}_1)$$

□

PROPOSICIÓN 100. *El estadístico DW toma valores en el rango de 0 a 4, cumpliéndose que*

- $\hat{\rho}_1 = 1 \Rightarrow DW = 0$
- $\hat{\rho}_1 = 0 \Rightarrow DW = 2$
- $\hat{\rho}_1 = -1 \Rightarrow DW = 4$

De aquí, el estadístico DW tomará un valor próximo a cero cuando exista autocorrelación positiva alta, tomará un valor próximo a cuatro cuando exista autocorrelación negativa alta, y tomará un valor próximo a dos cuando no exista autocorrelación.

La distribución de  $DW$  depende de la matriz de regresores  $\mathbf{X}$ . Diferentes matrices  $\mathbf{X}$  conducen a diferentes distribuciones y, por tanto, a diferentes valores críticos para realizar el contraste. Durbin y Watson encontraron unas cotas  $d_L$  y  $d_U$ , independientes de la matriz  $\mathbf{X}$ , que dividen el segmento de 0 a 4 en las siguientes regiones:

1.  $DW < d_L \Rightarrow$ , autocorrelación positiva de primer orden
2.  $d_U < DW < 4 - d_U \Rightarrow$ , incorrelación
3.  $DW > 4 - d_L \Rightarrow$ , autocorrelación negativa de primer orden

Cuando el estadístico DW cae entre las cotas  $d_L$  y  $d_U$  o entre las cotas  $4 - d_U$  y  $4 - d_L$  el contraste no es concluyente.

El estadístico DW es sólo válido cuando

1. la ecuación de regresión contiene término constante,
2. la ecuación de regresión no incluye entre las variables explicativas algún retardo de la variable dependiente,
3. la autocorrelación de los errores es del tipo  $AR(1)$ .

*Observación 60.* El estadístico DW pone su atención en el primer coeficiente de la PACF, ignorando los restantes coeficientes. De aquí, cuando se rechaza  $H_0$  no podemos afirmar que el proceso es un  $AR(1)$ . Además, si no se rechaza  $H_0$  tampoco podemos concluir que no existe autocorrelación.

*Observación 61.* La distribución exacta de DW puede obtenerse mediante integración numérica, expresando el estadístico como una ratio de formas cuadráticas en variables normales

$$DW = \frac{\hat{\mathbf{u}}' \mathbf{A} \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} = \frac{\mathbf{u}' \mathbf{M} \mathbf{A} \mathbf{M} \mathbf{u}}{\mathbf{u}' \mathbf{M} \mathbf{u}}$$

en donde  $\mathbf{A}$  es una matriz tridiagonal

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}$$

*Observación 62.* Wallis extendió el estadístico de DW al proceso  $u_t = \phi_4 u_{t-4} + \epsilon_t$ , que es útil para series trimestrales.

**10.2.3. El contraste de Breush-Godfrey.** En trabajos independientes, estos autores especificaron un proceso  $AR(p)$  o  $MA(p)$  para los errores

$$u_t = \phi_1 u_{t-1} + \dots + \phi_p u_{t-p} + \epsilon_t$$

$$u_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_p \epsilon_{t-p}$$



y formularon el problema de contraste

$$H_0 : \phi_1 = \dots = \phi_p = 0$$

$$H_1 : \phi_i \neq 0$$

en donde los procesos  $AR(p)$  y  $MA(p)$  son alternativas localmente equivalentes bajo la hipótesis nula de ausencia de autocorrelación.

Los pasos para realizar el contraste son los siguientes:

1. Estimar la ecuación de regresión de interés

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

por mínimos cuadrados ordinarios, y obtener los residuos  $\hat{u}_t$ .

2. Estimar la ecuación de regresión auxiliar

$$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \dots + \alpha_k X_{kt} + \delta_1 \hat{u}_{t-1} + \dots + \delta_p \hat{u}_{t-p} + e_t$$

y obtener el  $R^2$ .

3. Calcular el estadístico de contraste

$$\chi = (T - p)R^2$$

que sigue una distribución  $\chi^2$  con  $p$  grados de libertad.

4. La hipótesis  $H_0$  se rechaza al nivel de significación  $\alpha$ , si  $\chi > c$ , donde  $c$  es el valor crítico para el cual  $Prob(\chi_p^2 > c) = \alpha$

A diferencia del estadístico de Durbin-Watson, este contraste puede aplicarse cuando la ecuación de regresión incluye retardos de la variable dependiente.

**10.2.4. Los estadísticos Q de Box-Pierce y Box-Ljung.** Estos autores consideraron el problema de decidir si los  $m$  primeros coeficientes de la ACF de los errores de un proceso  $ARMA(p, q)$  son conjuntamente iguales a cero (*portmanteau test*). Box y Pierce demostraron que el estadístico

$$Q = T \sum_{k=1}^m \hat{\rho}_k^2 \stackrel{a}{\sim} \chi_{m-p-q}^2$$

que fue posteriormente modificado por Ljung-Box para mejorar su comportamiento en muestras pequeñas

$$Q' = T(T+2) \sum_{k=1}^p \frac{\hat{\rho}_k^2}{T-k} \stackrel{a}{\sim} \chi_p^2$$

en donde vemos que cada coeficiente  $\hat{\rho}_k^2$  está ponderado por el inverso del número de datos que se usan en su estimación, de modo que el peso de cada correlación aumenta con el retardo  $k$ .

### 10.3. Mínimos cuadrados generalizados

Supongamos que en el modelo de regresión múltiple para datos de series temporales

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t, \quad t = 1, \dots, T$$

el término de error  $u_t$  sigue un proceso  $AR(1)$

$$u_t = \phi_1 u_{t-1} + \epsilon_t$$

en donde el error  $\epsilon_t$  es un proceso de ruido blanco:  $E(\epsilon_t) = 0$ ,  $E(\epsilon_t^2) = \sigma_\epsilon^2$  y  $E(\epsilon_t \epsilon_{t-k}) = 0 \forall k \neq 0$ . Sabemos que la función de autocovarianzas de un proceso AR(1) es  $\gamma_k = \phi_1^k \gamma_0$  con  $\gamma_0 = \sigma_\epsilon^2 / (1 - \phi_1^2)$ , de modo que la matriz de varianzas y covarianzas de  $\mathbf{u}$  puede escribirse como

$$E(\mathbf{u}\mathbf{u}') = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \begin{pmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{T-2} \\ \phi_1^2 & \phi_1 & 1 & \dots & \phi_1^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{T-1} & \phi_1^{T-2} & \phi_1^{T-3} & \dots & 1 \end{pmatrix} = \sigma_\epsilon^2 \mathbf{\Omega}$$

Si el parámetro  $\phi_1$  fuera conocido, entonces podríamos calcular el estimador de mínimos cuadrados generalizados usando la fórmula

$$\hat{\beta}_{MCG} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

que requiere invertir la matriz  $\mathbf{\Omega}$ . Puesto que la inversión de una matriz de orden  $T \times T$  puede ser una operación muy costosa en tiempo de cálculo, resulta conveniente disponer de una fórmula que nos permita crear directamente la matriz  $\mathbf{\Omega}^{-1}$  sin necesidad de invertir  $\mathbf{\Omega}$ . Podemos comprobar fácilmente que

$$\mathbf{\Omega}^{-1} = \begin{pmatrix} 1 & -\phi_1 & 0 & \dots & 0 & 0 \\ -\phi_1 & 1 + \phi_1^2 & -\phi_1 & \dots & 0 & 0 \\ 0 & -\phi_1 & 1 + \phi_1^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 + \phi_1^2 & -\phi_1 \\ 0 & 0 & 0 & \dots & -\phi_1 & 1 \end{pmatrix}$$

que es una matriz tridiagonal cuyos elementos de la diagonal principal son iguales a  $1 + \phi_1^2$  excepto el primero y el último elemento que son iguales a 1, y cuyos elementos de la primera subdiagonal por debajo y encima de la diagonal principal son iguales a  $-\phi_1$ ; los restantes elementos son iguales a cero.

Podemos reducir nuevamente el coste computacional usando la descomposición  $\mathbf{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$ , en donde

$$\mathbf{P} = \begin{pmatrix} \sqrt{1 - \phi_1^2} & 0 & 0 & \dots & 0 & 0 \\ -\phi_1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi_1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\phi_1 & 1 \end{pmatrix}$$

es una matriz bidiagonal inferior con elementos en la diagonal principal iguales a 1, excepto el primero que es igual a  $\sqrt{1 - \phi_1^2}$ , y elementos en la primera subdiagonal por debajo de la diagonal principal iguales a  $-\phi_1$ .

Expresando el estimador MCG en términos de  $\mathbf{P}$ , tenemos

$$\hat{\beta}_{MCG} = (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}$$

en donde

$$\mathbf{Py} = \begin{pmatrix} \sqrt{1 - \phi_1^2} Y_1 \\ Y_2 - \phi_1 Y_1 \\ \vdots \\ Y_T - \phi_1 Y_{T-1} \end{pmatrix} \quad \text{y} \quad \mathbf{PX} = \begin{pmatrix} \sqrt{1 - \phi_1^2} & \sqrt{1 - \phi_1^2} X_{21} & \dots & \sqrt{1 - \phi_1^2} X_{k1} \\ 1 - \phi_1 & X_{22} - \phi_1 X_{21} & \dots & X_{k2} - \phi_1 X_{k1} \\ \vdots & \vdots & \dots & \vdots \\ 1 - \phi_1 & X_{2T} - \phi_1 X_{2T-1} & \dots & X_{kT} - \phi_1 X_{kT-1} \end{pmatrix}$$

De aquí, podemos escribir el modelo de regresión transformado  $\mathbf{Py} = \mathbf{PX}\beta + \mathbf{Pu}$  como

$$\sqrt{1 - \phi_1^2} Y_1 = \beta_1 \sqrt{1 - \phi_1^2} + \beta_2 \sqrt{1 - \phi_1^2} X_{21} + \dots + \beta_k \sqrt{1 - \phi_1^2} X_{k1} + \epsilon_1$$

$$Y_t - \phi_1 Y_{t-1} = \beta_1 (1 - \phi_1) + \beta_2 (X_{2t} - \phi_1 X_{2t-1}) + \dots + \beta_k (X_{kt} - \phi_1 X_{kt-1}) + \epsilon_t$$

Las variables transformadas  $Y_t^* = Y_t - \phi_1 Y_{t-1}$  y  $X_{jt}^* = X_{jt} - \phi_1 X_{jt-1}$  ( $j = 1, \dots, k$ ) podemos calcularlas en cualquier programa econométrico con capacidad para transformar datos o en una hoja de cálculo. La primera observación, que recibe una transformación diferente, podríamos ignorarla si tenemos una muestra grande.

El procedimiento descrito tiene dos limitaciones. En primer lugar, no podemos extenderlo a otras formas de autocorrelación. Por ejemplo, no tenemos una fórmula cerrada para crear la inversa de un proceso  $AR(p)$  general. En segundo lugar, el cálculo del estimador *MCG* requiere conocer el parámetro  $\phi_1$ . Si disponemos de una estimación  $\hat{\phi}_1$ , entonces podemos estimar el estimador *MCG* como

$$\hat{\beta}_{MCG} = (\mathbf{X}' \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Omega}^{-1} \mathbf{y}$$

que es el estimador de mínimos cuadrados generalizados factibles (*MCGF*). Las propiedades estadísticas de  $\hat{\beta}_{MCG}$  son desconocidas en muestras finitas son desconocidas, y dependen de las propiedades asintóticas de  $\phi_1$  en muestras grandes. Si  $\hat{\phi}_1$  es un estimador consistente de  $\phi_1$ , entonces  $\hat{\beta}_{MCG}$  también es consistente.

**10.3.1. El procedimiento iterativo de Cochrane-Orcutt.** Cochrane y Orcutt propusieron un método simple para estimar iterativamente los parámetros  $\beta$  de un modelo de regresión y los parámetros  $\phi_i$  de un proceso  $AR(p)$ , que evita la inversión de la matriz  $\Omega$  y que ilustramos aquí considerando un modelo de regresión con dos variables explicativas

$$(10.4) \quad Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

y autocorrelación de tipo  $AR(1)$

$$(10.5) \quad u_t = \phi_1 u_{t-1} + \epsilon_t$$

La idea es transformar el modelo de regresión en un modelo cuyo término de error sea ruido blanco  $\epsilon_t = u_t - \phi_1 u_{t-1}$ . Dada la relación entre  $Y_t$  y  $u_t$ , podemos definir directamente las variables

$$Y_t^* = Y_t - \phi_1 Y_{t-1} \quad \text{y} \quad X_{jt}^* = X_{jt} - \phi_1 X_{jt-1}$$

y especificar la relación

$$(10.6) \quad Y_t^* = \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + \epsilon_t$$

en donde  $X_{1t}^* = (1 - \phi_1)$ .

*Observación 63.* Si el proceso  $AR(1)$  en términos del operador  $B$ ,  $(1 - \phi_1 B)u_t = \epsilon_t$ , vemos que multiplicando la ecuación (10.4) por el polinomio  $(1 - \phi_1 B)$  obtenemos la ecuación (10.6)

$$(1 - \phi_1 B)y_t = \beta_1(1 - \beta_1) + \beta_2(1 - \phi_1 B)X_{2t} + \beta_3(1 - \phi_1 B)X_{3t} + \epsilon_t$$

Nuestro objetivo es estimar los coeficientes  $\beta_i$  ( $i = 1, 2, 3$ ) de la ecuación (10.6) y el parámetro  $\phi_1$  de la ecuación (10.5). El problema consiste en que para estimar  $\beta_i$  ( $i = 1, 2, 3$ ) necesitamos una estimación de  $\phi_1$ , y para estimar  $\phi_1$  necesitamos estimaciones de  $\beta_i$  ( $i = 1, 2, 3$ ). La solución a este círculo vicioso propuesta por Cochrane y Orcutt es el siguiente procedimiento iterativo:

1. Usar una estimación preliminar de  $\phi_1$ , por ejemplo  $\phi_1 = 0$ , transformar los datos y estimar por MCO los coeficientes  $\beta_i$  ( $i = 1, 2, 3$ ) en la ecuación (10.6).
2. Usando estas estimaciones, podemos obtener los residuos  $\hat{u}_t$  en la ecuación (10.4).
3. Los residuos  $\hat{u}_t$  nos permiten encontrar una nueva estimación de  $\phi_1$  en (10.5).
4. Los pasos 1, 2 y 3 se repiten sucesivamente hasta alcanzar la convergencia, es decir, hasta que las estimaciones en dos repeticiones o iteraciones consecutivas prácticamente no cambien.

Los criterios de convergencia usuales son:

1. La estimación del parámetro  $\phi_1$  en dos iteraciones consecutivas no cambia,  $|\phi_1^{(j)} - \phi_1^{(j-1)}| < 10^{-5}$ , donde  $j$  indica la iteración.
2. La estimación de cada parámetro  $\beta$  en dos iteraciones consecutivas no cambia,  $|\beta_i^{(j)} - \beta_i^{(j-1)}| < 10^{-5}$ .
3. La estimación de la suma de cuadrados de los residuos  $\sum_{t=1}^T \hat{u}_t^2$  en dos iteraciones consecutivas no cambia,  $|(\sum_{t=1}^T \hat{u}_t^2)^{(j)} - (\sum_{t=1}^T \hat{u}_t^2)^{(j-1)}| < 10^{-5}$ .
4. Se alcanza un número máximo de repeticiones.

*Observación 64.* En general, si el término de error sigue un proceso  $AR(p)$ ,  $\phi(B)u_t = \epsilon_t$ , el modelo transformado se obtiene multiplicando el modelo de interés por el polinomio  $\phi(B)$ .

## 10.4. Predicción

Las formulas de la predicción puntual y por intervalo en presencia de autocorrelación también se ven afectadas. En el modelo lineal general

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

con errores de tipo  $ARMA(p, q)$ , la predicción puntual en el origen  $n$  al horizonte  $h$  viene dada por

$$\hat{y}_n(h) = \mathbf{x}'_{n+h} \hat{\boldsymbol{\beta}}_{MCG} + \hat{u}_n(h)$$

en donde, a diferencia de lo que ocurre en el modelo clásico,  $\hat{u}_n(h)$  será distinto de cero. Ahora el error de predicción

$$\hat{\epsilon}_n(h) = -\mathbf{x}'_{n+h} (\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta}) + (u_{n+h} - \hat{u}_n(h))$$

tiene momentos desconocidos. Si suponemos que los parámetros del proceso  $ARMA(p,q)$  son conocidos, entonces el error de predicción será insesgado y tendrá varianza

$$V(\hat{e}_n(h)) = \mathbf{x}'_{n+h} V(\hat{\beta}_{MCG}) \mathbf{x}_{n+h} + \sigma_\epsilon^2 (1 + \psi_1^2 + \dots + \psi_{h-1}^2)$$

en donde  $\psi_j$  son los coeficientes de la representación  $MA(\infty)$  del proceso  $ARMA(p, q)$ .

### 10.5. Resumen

1. El supuesto de ausencia de autocorrelación en las perturbaciones debe cuestionarse cuando trabajamos con datos de series temporales.
2. El proceso  $AR(1)$  es la forma de autocorrelación más empleada como alternativa al supuesto de ausencia de autocorrelación.
3. El análisis gráfico de los residuos de un modelo estimado por mínimos cuadrados es útil para detectar la presencia y forma de autocorrelación en los errores.
4. El estadístico de Durbin-Watson,  $DW = 2(1 - \hat{\rho}_1)$ , es el contraste más popular para detectar autocorrelación de primer orden, pero no está exento de limitaciones.
5. Algunos de los inconvenientes del contraste Durbin-Watson son subsanados por el contraste de Breusch-Godfrey, que puede interpretarse con un contraste de variables omitidas.
6. El procedimiento iterativo de Cochrane-Orcutt es un procedimiento computacionalmente eficiente para estimar modelos de regresión con perturbaciones de tipo  $AR$ , pero no es directamente aplicable cuando las perturbaciones siguen un esquema  $MA$ .
7. La detección de la autocorrelación no sólo es importante en la estimación y contraste de hipótesis, sino también en la predicción.

### Palabras clave

Autocorrelación	Test de rachas
Correlación serial pura	Contraste de Durbin-Watson
Correlación serial impura	Procedimiento iterativo
Análisis gráfico de residuos	Criterios de convergencia
Función de autocorrelación muestral	

### 10.6. Ejercicios

1. Use las tablas de Durbin-Watson para contrastar la hipótesis de autocorrelación de primer orden en los dos siguientes casos: (1)  $DW = 0,6$ ,  $k = 3$  y  $n = 50$ ; (2)  $DW = 3,2$ ,  $k = 5$  y  $n = 80$ , en donde  $k$  incluye el término constante.
2. Describa el procedimiento de Breusch y Godfrey para contrastar la hipótesis de que los errores de un modelo  $ADL(1,1)$  siguen un proceso  $AR(2)$ .
3. Describa el procedimiento iterativo de Cochrane-Orcutt en la estimación del modelo de regresión simple con autocorrelación del tipo  $AR(2)$ .
4. Describa el procedimiento iterativo de Cochrane-Orcutt en la estimación del modelo de regresión simple con autocorrelación del tipo  $MA(1)$  (Pista: aproxime el proceso  $MA(1)$  por un proceso  $AR(p)$ ).

5. Suponga que el modelo correcto para describir una variable  $Y_t$  es

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t$$

pero por error se especifica el modelo

$$Y_t = \alpha_1 + \alpha_2 X_t + v_t$$

$$v_t = \phi_1 v_{t-1} + \epsilon_t$$

- a) Explique la principal diferencia entre ambos modelos.
- b) Proponga un posible contraste para detectar este error de especificación.