

Econometría II

La autocorrelación

Alejandro Mosiño - *Universidad de Guanajuato*

v. 2026.02.27

¿Qué es la autocorrelación?

¿En qué consiste la autocorrelación?

Uno de los problemas que podemos encontrar en datos de *series temporales* es el de **autocorrelación**. Formalmente, existe autocorrelación cuando:

$$\begin{aligned}\text{cov}(u_i, u_j) &= \mathbf{E}(u_i u_j) \\ &= \sigma_{ij},\end{aligned}$$

para $i \neq j, i, j = 1, 2, \dots, n$.

Por lo tanto, la matriz de covarianzas de los errores se ve como:

$$\mathbf{\Omega} = \begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma^2 & \sigma_{23} & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \sigma_{3n} & \dots & \sigma^2 \end{pmatrix}$$

¿Cuáles son las consecuencias de la autocorrelación?

(Al igual que en el caso de la heteroscedasticidad) En presencia de autocorrelación:

- Los estimadores de MCO *siguen siendo* insesgados y consistentes.
- Los estimadores de MCO *dejan de ser* eficientes.
- Los resultados de las pruebas t y F *dejan de ser* confiables.

Detección de la autocorrelación

Detección de la autocorrelación (1/2)

Consideremos el modelo de regresión simple:

$$\text{Remuneración}_t = \beta_1 + \beta_2 \text{Productividad}_t + u_t$$

Donde medimos la relación que existe entre los índices de producción por hora y remuneración real por hora de 1960 a 2005 en los Estados Unidos. (La base de los índices es 1992=100.)

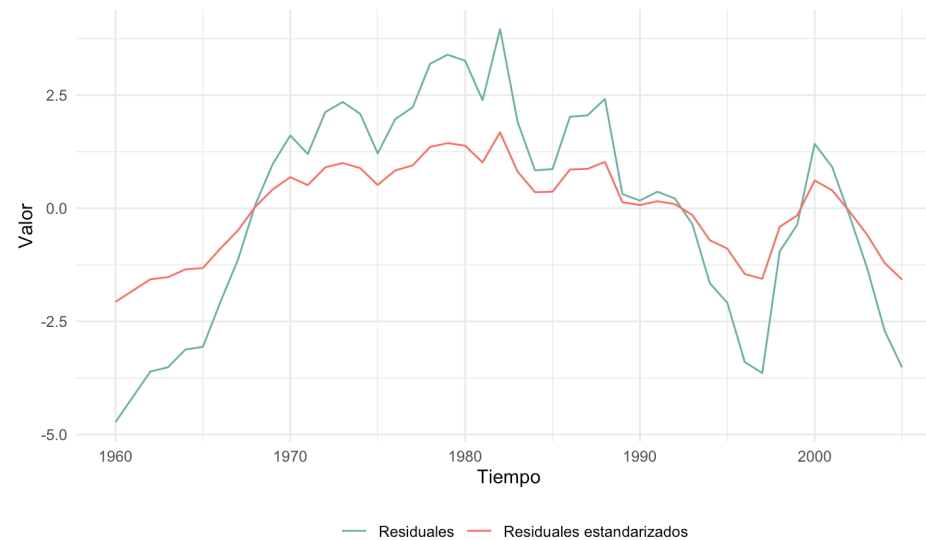
Detección de la autocorrelación (2/2)

Ejercicio

Utiliza los datos del archivo [E204-productivity.csv](#). Estima el modelo utilizando el método de los MCO. Explica tus resultados.

Método gráfico

Dado que la autocorrelación se refiere a la relación que existe entre los errores del modelo, u_t , una gráfica de los residuales, e_t , (o de los residuales estandarizados) contra el tiempo nos permite detectar la existencia de patrones identificables a simple vista.



El correlograma (1/5)

Otra posible evidencia de la presencia de autocorrelación es el **correlograma**. Este no es más que la gráfica de la **función de autocorrelación** de los errores del modelo. Esta se calcula como:

$$r_k = \frac{Cov(u_t, u_{t-k})}{Var(u_t)}$$

para todo $k = 1, 2, \dots$

El correlograma (2/5)

Ejercicio

Supongamos que los errores del modelo siguen un proceso autorregresivo de orden 1, AR(1). Es decir:

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

para $\varepsilon \sim \square(0, \sigma^2)$ y $|\rho| < 1$. Además, suponemos que ε_t y u_t son independientes y que el proceso es **estacionario**.
Calcula:

- El valor esperado de u_t y su varianza, suponiendo que los errores son homoscedásticos, $\text{var}(u_t) = \lambda_0$.
- La k -ésima autocovarianza de los errores, λ_k .
- La k -ésima autocorrelación de los errores, r_k .

Para cualquier valor de $|\rho| < 1$, grafica en un mismo diagrama la k -ésima correlación para $k = 0, 1, 2, 3, 4$.

El correlograma (3/5)

Cuando no conocemos el proceso que siguen los errores del modelo, la función de autocorrelación tiene que estimarse:

$$\begin{aligned}\hat{r}_k &= \frac{\text{Cov}(e_t, e_{t-k})}{\text{Var}(e_t)} \\ &= \frac{\sum_{t=1}^{T-k} (e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^T (e_t - \bar{e})^2},\end{aligned}$$

para $k = 1, 2, \dots$

El correlograma (4/5)

La estimación de la función de autocorrelación también nos permite calcular intervalos de confianza. Estos se calculan tomando en cuenta que:

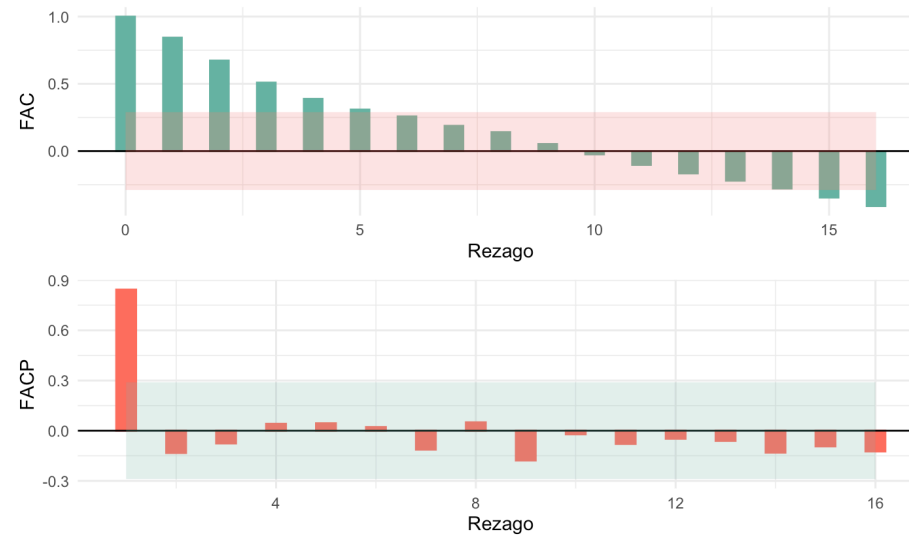
$$\sqrt{T}r_k \sim \mathcal{N}(0, 1).$$

Entonces, por ejemplo, el valor estimado de r_k , $\hat{r}_k$, es significativo al 5% si:

$$-1.96 \geq \sqrt{T}r_k \geq 1.96$$

El correlograma (5/5)

En el ejemplo anterior tenemos:



El correlograma muestra evidencia de autocorrelación entre varios retardos de los errores.

La Q de Ljung-Box (1/2)

El estadístico Q es una prueba estadística cuya hipótesis es:

H_0 : No existe correlación de orden k

H_1 : Existe correlación de orden k

El estadístico Q se calcula como:

$$Q = T(T + 2) \sum_{j=1}^k \frac{r_j^2}{T - j}$$

Y, si la hipótesis nula es correcta, $Q \sim \chi^2$ con k grados de libertad.

La Q de Ljung-Box (2/2)

En el ejemplo anterior, para el primer retardo tenemos:

```
##  
## Box-Ljung test  
##  
## data:  data$res  
## X-squared = 34.93, df = 1, p-value = 3.418e-09
```

Y para el segundo:

```
##  
## Box-Ljung test  
##  
## data:  data$res  
## X-squared = 57.684, df = 2, p-value = 2.98e-13
```

Prueba de Breusch-Godfrey (1/3)

Supongamos que el modelo es:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{u}$$
$$u_t = \gamma + \sum_{i=2}^k \delta_i x_{it} + \sum_{j=1}^p \rho_j u_{t-j}$$

En la prueba de Breusch-Godfrey se plantea la siguiente hipótesis:

$$H_0 : \mathbb{E}(u_t u_\tau) = 0 \text{ para todo } t \neq \tau$$

$$H_1 : \mathbb{E}(u_t u_\tau) \neq 0 \text{ para al menos un } t \neq \tau$$

Prueba de Breusch-Godfrey (2/3)

Los pasos para realizar la prueba de Breusch-Godfrey son los siguientes:

1. Estimar el modelo por MCO y obtener los residuales estimados, e_t .
2. Estimar por MCO la siguiente expresión:

$$e_t = \gamma + \sum_{i=2}^k \delta_i x_{it} + \sum_{j=1}^p \rho_j e_{t-j} + \epsilon_t$$

3. Calcular el estadístico $BG = T \times R^2$ y compararlo contra una distribución χ^2 con p grados de libertad.

Prueba de Breusch-Godfrey (3/3)

En el ejemplo, para $p = 1$ tenemos:

```
##  
## Breusch-Godfrey test for serial correlation of order up to 1  
##  
## data:  mod  
## LM test = 34.618, df = 1, p-value = 4.012e-09
```

Prueba de Durbin-Watson (1/4)

La **prueba de Durbin-Watson** es una de las pruebas más conocidas, pero a la vez una de las más limitadas para detectar la autocorrelación. La prueba supone que:

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

donde ε_t es un ruido blanco. La hipótesis de la prueba es:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Prueba de Durbin-Watson (2/4)

Los pasos para realizar la prueba de Durbin-Watson son los siguientes:

1. Estimar el modelo por MCO y recuperar los residuales estimados, e_t .
2. Calcular el estadístico DW :

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Nota que $0 \leq DW \leq 4$.

Prueba de Durbin-Watson (3/4)

Para estimar un modelo con heteroscedasticidad multiplicativa seguimos los siguientes pasos:

3. Utilizar los valores críticos de Durbin-Watson, D_L y D_U :

- Si $DW < D_L$: *Se rechaza H_0* . Evidencia de autocorrelación positiva.
- Si $D_L < DW < D_U$: *No se puede tomar decisión*.
- Si $DW > 4 - D_L$: *Se rechaza H_0* . Evidencia de autocorrelación negativa.

Prueba de Durbin-Watson (4/4)

Como regla de oro, *se espera que el estadístico de DW tenga un valor cercano a 2* para rechazar la hipótesis de autocorrelación.

En nuestro ejemplo:

```
## lag Autocorrelation D-W Statistic p-value
## 1      0.8437328      0.1738878      0
## Alternative hypothesis: rho != 0
```

Medidas correctivas de la autocorrelación

Medidas correctivas de la autocorrelación (1/2)

Si logramos mostrar que nuestro modelo presenta problemas de autocorrelación, el siguiente paso es aplicar alguna medida correctiva.

Si conocemos la matriz de varianzas y covarianzas de los errores, aplicamos el método de los MCG. Sin embargo, dado que el término de error es desconocido, la aplicación de este método es inviable.

Medidas correctivas de la autocorrelación (2/2)

Si logramos mostrar que nuestro modelo presenta problemas de autocorrelación, el siguiente paso es aplicar alguna medida correctiva:

- Si no conocemos la matriz de varianzas y covarianzas de los errores, pero podemos “adivinar” su forma, aplicamos el método de los **MCG Factibles** (MCGF).
- Si no podemos adivinar la forma de la matriz, estimamos el modelo por MCO, pero utilizando el procedimiento de **Newey-West** para calcular **varianzas robustas** (varianzas asintóticamente válidas).

El método de los MCGF: Método de Cochrane-Orcutt

(1/6)

Imaginemos que “adivinamos” que los errores del modelo siguen un proceso autorregresivo de orden 1, $AR(1)$. Es decir:

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

Esta ecuación se estima por MCO usando e_t como estimador de u_t . Con el valor estimado de ρ podemos transformar las variables en nuestro modelo calculando para toda variable z_t (incluyendo la constante):

$$\tilde{z}_t = z_t - \rho z_{t-1}$$

El método de los MCGF: Método de Cochrane-Orcutt

(2/6)

Aplicando esta transformación obtenemos:

```
##  
## Call:  
## lm(formula = tRcompb ~ tconst + tProdb - 1, data = data)  
##  
## Coefficients:  
## tconst tProdb  
## 40.7892  0.5903
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 0.2062101 1.564887 0.082  
## Alternative hypothesis: rho != 0
```

El método de los MCGF: Método de Cochrane-Orcutt (3/6)

Ejercicio

Responde lo siguiente:

- ¿Presenta el modelo transformado problemas de autocorrelación?
- ¿Qué pasa si el modelo se transforma utilizando $\rho = 1$?

El método de los MCGF: Método de Cochrane-Orcutt (4/6)

Ejercicio

Muestra que, si ρ es conocido, los procedimientos de MCGF y MCG son equivalentes.

El método de los MCGF: Método de Cochrane-Orcutt

(5/6)

Nota

La transformación de los datos depende de la forma del proceso u_t . Por ejemplo, si u_t sigue un proceso $AR(2)$ tenemos:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

Esta ecuación se estima por MCO usando e_t como estimador de u_t .

En este caso, se usan los valores estimados para ρ_1 y ρ_2 para transformar las variables en nuestro modelo (incluyendo la constante):

$$\tilde{z}_t = z_t - \rho_1 z_{t-1} - \rho_2 z_{t-2}$$

El método de los MCGF: Método de Cochrane-Orcutt (6/6)

Ejercicio

Estima el modelo por MCGF bajo el supuesto que el error sigue un proceso $AR(2)$. Comenta.

Varianzas robustas de Newey-West

Si no conocemos la forma de la matriz de covarianzas de los errores, podemos estimar el modelo por el método de los MCO, pero modificando los errores estándar para que estos tomen cuenta la autocorrelación.

En nuestro ejemplo, si usamos los **errores estándar robustos de Newey-West** obtenemos:

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 32.741903   3.345565  9.7867 1.294e-12 ***
## Prodb       0.670406   0.037003 18.1174 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Naturalmente, los residuales siguen presentando autocorrelación, pero los estimadores son insesgados y las pruebas t y F son válidas.

Práctica

Práctica

Utiliza los datos: [E205-usconsumption](#). El objetivo es determinar el gasto real en consumo en el periodo 1947-2000 en los Estados Unidos

Usando la variable consumption (Real Consumption Expenditure) como dependiente y el resto de las variables como explicativas, realiza un análisis completo del problema de la autocorrelación.