Universidad Carlos III de Madrid

César Alonso

ECONOMETRÍA

Tema 8: AUTOCORRELACIÓN

Índice

1. Introducción

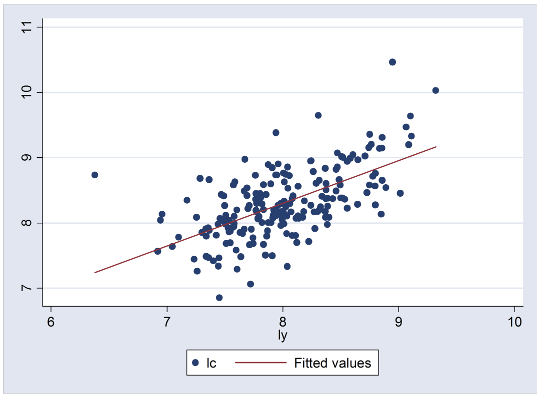
- El término **autocorrelación** (o *correlación serial*) hace referencia a situaciones donde las observaciones de la variable dependiente no son extraídas independientemente.

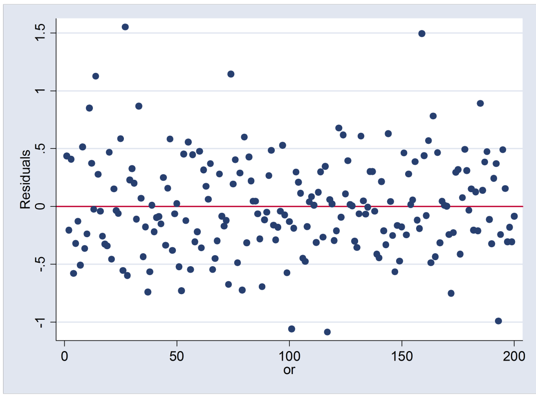
- Este fenómeno es muy habitual en el caso de datos de series temporales.

- Por el contrario, dicho fenómeno no se presenta en el caso de datos de sección cruzada, en que las unidades que los componen son independientes entre sí.

Ejemplo (Datos de sección cruzada):

- Datos de sección cruzada de 200 familias españolas formadas por parejas con o sin hijos seleccionadas aleatoriamente de la EPF de 1990-91.
- Regresión simple del logaritmo del consumo per cápita sobre el logaritmo de la renta disponible per cápita.
- Gráfico de la izquierda: logaritmo del consumo per capita sobre el logaritmo de la renta disponible per capita
  así como la recta de la regresión lineal correspondiente.
- Gráfico de la derecha: residuos de dicha regresión frente al número de orden de cada observación (que es arbitrario en el caso de datos de sección cruzada).
  - Como los residuos son los análogos muestrales de los errores de la regresión poblacional,
    si los errores fueran independientes no deberíamos encontrar ningún patrón.





- En el caso de datos de series temporales (para los que el orden de ocurrencia de las observaciones es relevante), la existencia de autocorrelación es bastante frecuente, debido entre otras cosas a la **dependencia temporal** causada por la inercia que presentan los datos económicos.

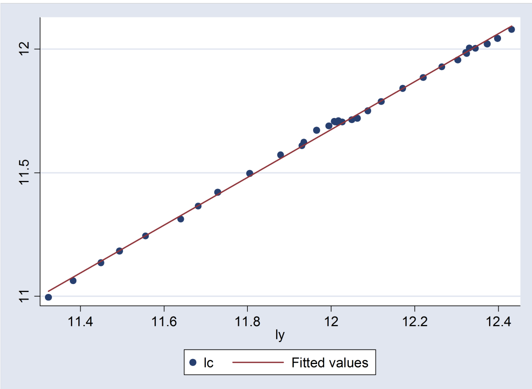
- Es bastante plausible que los **shocks o perturbaciones** que afectan a las variables económicas de interés presenten dependencia temporal. *Ejemplos:*

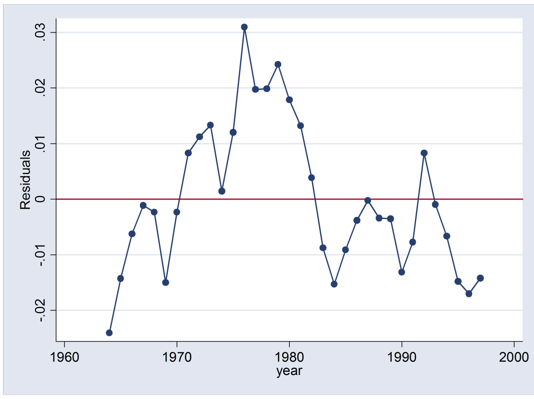
  - Queremos analizar la relación entre desemplo e inflación entre 1950 y 1990.
    Todos los factores inobservables que pueden afectar al desempleo pero no se incluyen en dicha relación se recogen en el término de error.
    Algunas de estos factores inobservables pueden presentar dependencia temporal: en particular, los **shocks energéticos** de los años 70.
    Dicho shock energético puede originar un desempleo mayor que el predicho por el modelo (y por lo tanto, un error positivo grande).
    Dado que los efectos del shock energético no se amortiguan rápidamente, los errores de los años siguientes serán también positivos.
    La magnitud de dichos errores a causa del shock energético se irá reduciendo y eventualmente pasados unos períodos tendremos errores próximos a cero (o incluso negativos).

  - Suponga que queremos analizar la relación entre índice bursátil y crecimiento.
    Entre los factores inobservables que pueden afectar al índice bursátil, están el grado de confianza de los agentes económicos (que se ve afectado por noticias no anticipadas).
    Dichas expectativas pueden originar un valor del índice bursátil inferior al predicho por el modelo si la confianza de los agentes empeora.
    Es de esperar que el grado de confianza de los agentes presente una cierta inercia.

Ejemplo (Datos de series temporales):

- Datos anuales para la economía española entre 1964 y 1997 de consumo agregado y PIB a precios de mercado, en millones de euros constantes de 1986.
- Regresión simple del logaritmo del consumo sobre el logaritmo del PIB.
- Gráfico de la izquierda: logaritmo del consumo sobre el logaritmo del PIB así como la recta de la regresión lineal correspondiente.
- Gráfico de la derecha: residuos de dicha regresión frente al número de orden de cada observación (es decir, el año, **que no es arbitrario**).
  - Interpretando los residuos como los análogos muestrales de los errores de la relación entre el logaritmo del consumo y el logaritmo del PIB, *si los errores fueran independientes no deberíamos encontrar ningún patrón* a lo largo del tiempo.
  - Sin embargo, los residuos muestran períodos en que son predominantemente negativos y períodos en que son predominantemente positivos.





- Cuando las observaciones están correlacionadas, es posible que el estimador de mínimos cuadrados deje de ser el estimador óptimo, y que la formulación convencional de su error estándar puede ser inapropiada.

- Por conveniencia, trabajaremos con la regresión simple, donde solamente se condiciona en una variable.

- A la hora de indexar las observaciones *utilizaremos* t, s ó $t-j$ *como subíndices* en lugar de i, h, para resaltar que el análisis que vamos a efectuar es particularmente relevante con series temporales.

2. El modelo de regresión con series temporales

- Supongamos que disponemos de una serie temporal de (Y_t, X_t) $(t = 1, \ldots, T)$, es decir: observamos T observaciones consecutivas de Y_t y X_t.
- Supondremos que se verifican todos los supuestos del modelo de regresión clásico, excepto el que se refiere a la independencia de las observaciones.
- Dada una serie temporal de T observaciones, podemos escribir el modelo como

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (t = 1, \ldots, T)$$

- **Supuestos del modelo de regresión con observaciones dependientes**:

  1. Linealidad en parámetros $(Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t)$
  2.
     (i) $E(\varepsilon_t | X_1, \ldots, X_T) = E(\varepsilon_t | X_t) \quad \forall t$
     Este supuesto se cumplía siempre con observaciones independientes (datos de sección cruzada).
     Pero con datos de series temporales, este supuesto es restrictivo,
     porque establece que la media de la perturbación sólo se ve afectada por el valor contemporáneo de X.

     (ii) $E(\varepsilon_t | X_t) = 0 \quad \forall t$

  Los supuestos 2.(i) y 2.(ii) establecen por tanto conjuntamente que

$$E(\varepsilon_t | X_1, \ldots, X_T) = E(\varepsilon_t | X_t) = 0 \quad \forall t$$

  En particular, nótese que exigimos que la perturbación no esté correlacionada con valores pasados, presente ni futuros de X.
  Esto supone que descartamos la posibilidad de que las perturbaciones puedan afectar a valores futuros de X.
  Esta implicación, que supone **exogeneidad estricta** de X, es muy restrictiva.

Supone que aunque los shocks o perturbaciones afectan por definición a Y, nunca tienen efecto sobre X. (lo que parece poco plausible)

Implicaciones:

- $E(\varepsilon_t) = 0 \quad \forall t$ (por la ley de esperanzas iteradas).
- $C(X_t, \varepsilon_t) = 0 \quad \forall t$
- Por los supuestos 1. y 2.,

$$E(Y_t | X_1, \ldots, X_T) = E(Y_t | X_t) = \beta_0 + \beta_1 X_t$$

  es decir, la FEC es lineal.

3.

(i) $V(\varepsilon_t | X_1, \ldots, X_T) = V(\varepsilon_t | X_t) = \sigma^2 \quad \forall t$
(**Homocedasticidad condicional**)

(ii) $C(\varepsilon_t, \varepsilon_{t-j} | X_1, \ldots, X_T) = C(\varepsilon_t, \varepsilon_{t-j} | X_t, X_{t-j}) = \sigma_{t,t-j} \quad \forall t, j \quad (j \neq 0)$
(**Autocorrelación condicional** o **correlación serial condicional**)
Este supuesto implica que la covarianza entre dos perturbaciones ocurridas en distintos períodos (condicional a las realizaciones de X en dichos períodos) puede ser distinta de cero.

(iii) $\sigma_{t,t-j} = \gamma_j \quad \forall t, j \quad (j \neq 0)$
Este supuesto implica que la covarianza entre dos perturbaciones ocurridas en distintos períodos (condicional a las realizaciones de X en dichos períodos) sólo depende del tiempo j transcurrido entre ambos períodos, pero no del período t considerado.
Es decir: la covarianza condicional entre las perturbaciones ocurridas en 1980 y 1985 es la misma que la covarianza condicional entre las perturbaciones ocurridas en 1990 y 1995.
Detrás de este supuesto está la **condición de estacionariedad (en covarianza)**, que establece que todos los momentos condicionales de primer y segundo orden (medias, varianzas y covarianzas) no dependen del momento del tiempo considerado.

De hecho, la condición de estacionariedad está también implícita en el supuesto 2. y en el supuesto 3.(i).

Intuitivamente, la estacionariedad supone que la relación entre las variables es relativamente estable a lo largo del tiempo. De lo contrario, los parámetros de la relación variarían a lo largo del tiempo, lo que nos impediría inferir cómo afectan cambios de una variable sobre el valor medio de la otra.

Los supuestos 3.(i), 3.(ii) y 3.(iii) establecen por tanto conjuntamente que

$$V\left(\varepsilon_t \mid X_1, \ldots, X_T\right) = V\left(\varepsilon_t \mid X_t\right) = \sigma^2 \quad \forall t$$

$$C\left(\varepsilon_t, \varepsilon_{t-j} \mid X_1, \ldots, X_T\right) = C\left(\varepsilon_t, \varepsilon_{t-j} \mid X_t, X_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$$

Implicaciones (por la ley de esperanzas iteradas):

$$V\left(\varepsilon_t\right) = \sigma^2 \quad \forall t$$

$$C\left(\varepsilon_t, \varepsilon_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$$

3. Consecuencias sobre la estimación MCO

- ¿Cuáles son las implicaciones de la autocorrelación sobre las propiedades de la regresión lineal MC muestral de Y sobre X, esto es $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{X}$?

- Recordemos que, dada una muestra, la pendiente MCO se expresa como

$$\widehat{\beta}_1 = \sum_t c_t Y_t$$

donde, definiendo la variable explicativa en desviaciones con respecto a su media, $x_s = X_s - \overline{X}$,

$$c_t = x_t / \sum_s x_s^2$$

- Además, sabemos que

$$\sum_t c_t = 0, \quad \sum_t c_t X_t = 1, \quad \sum_t c_t^2 = 1/\sum_s x_s^2$$

Por tanto, $\widehat{\beta}_1$ tiene esperanza

$$\begin{aligned} E\left(\widehat{\beta}_1 \middle| X_1, \ldots, X_T\right) &= \sum_t c_t E\left(Y_t | X_t\right) = \sum_t c_t \left(\beta_0 + \beta_1 X_t\right) \\ &= \beta_0 \sum_t c_t + \beta_1 \sum_t c_t X_t = \beta_1 \end{aligned}$$

En consecuencia, con los supuestos arriba citados, el estimador MCO sigue siendo **insesgado**.

- Además, es fácil comprobar que

$$\operatorname{p\,lím} \widehat{\beta}_1 = \beta_1,$$

y por tanto el estimador MCO $\widehat{\beta}_1$ es un estimador **consistente** de β_1. Para ello, nótese que

$$\widehat{\beta}_1 = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t Y_t}{\frac{1}{T}\sum_{s=1}^{T} x_s^2} = \frac{S_{XY}}{S_X^2}.$$

$$\beta_1 = \frac{C(X_t, Y_t)}{V(X_t)}$$

Como S_{XY}, S_X^2 son estimadores consistentes de $C(X_t, Y_t)$, $V(X_t)$, respectivamente, el cociente entre S_{XY} y S_X^2 estima consistentemente β_1.

- **Sin embargo**, la varianza del estimador tiene ahora la forma

$$V\left(\widehat{\beta}_1 \middle| X_1,\ldots,X_T\right) = \sum_j \sum_t c_t c_{t-j} \sigma_{t,t-j} = \sigma^2 \sum_t c_t^2 + 2 \sum_j \sum_{t>j} c_t c_{t-j} \sigma_{t,t-j}$$

y por tanto, dado que $c_t c_{t-j} = x_t x_{t-j} \Big/ \left(\sum_s x_s^2\right)^2 = x_t x_{t-j} \Big/ (TS_X^2)^2$,

$$\begin{aligned} V\left(\widehat{\beta}_1\right) &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \sum_{t>j} x_t x_{t-j} \sigma_{t,t-j}\right] \\ &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \sum_{t>j} x_t x_{t-j} \gamma_j\right] \\ &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \gamma_j \sum_{t>j} x_t x_{t-j}\right] \end{aligned}$$

Ejemplo Es interesante verificar la obtención de la expresión para la varianza, desarrollando explícitamente un caso sencillo. Para $T = 3$, tenemos $\widehat{\beta}_1 = c_1 Y_1 + c_2 Y_2 + c_3 Y_3$, con lo cual

$$\begin{aligned} V\left(\widehat{\beta}_1 \middle| X_1, X_2, X_3\right) &= c_1^2 \sigma_{11} + c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} \\ &+ c_2 c_1 \sigma_{21} + c_2^2 \sigma_{22} + c_2 c_3 \sigma_{23} \\ &+ c_3 c_1 \sigma_{31} + c_3 c_2 \sigma_{32} + c_3^2 \sigma_{33} \end{aligned}$$

Como $\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma^2$ y $\sigma_{12} = \sigma_{21} = \gamma_1$, etc., esta expresión queda

$$\begin{aligned} V\left(\widehat{\beta}_1 \middle| X_1, X_2, X_3\right) &= \sigma^2\left(c_1^2 + c_2^2 + c_3^2\right) + 2\left(c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} + c_2 c_3 \sigma_{23}\right) \\ &= \frac{\sigma^2}{TS_X^2} + 2\left(c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} + c_2 c_3 \sigma_{23}\right) \\ &= \frac{\sigma^2}{TS_X^2} + 2\left[\left(c_1 c_2 + c_2 c_3\right)\gamma_1 + c_1 c_3 \gamma_2\right]. \end{aligned}$$

- El hecho de que en general $\gamma_j \neq 0$ ($j \neq 0$) hace que $V\left(\widehat{\beta}_1 \middle| X_1,\ldots,X_T\right) \neq \sigma^2 / (TS_X^2)$, de manera que la varianza no viene dada por la expresión convencional.

- El programa de ordenador no sabe que el modelo ha cambiado, por lo que continuará calculando los valores

$$s^2 = \sum_{t=1}^{T} e_t^2 \Big/ (T-2)$$
$$s_{\widehat{\beta}_1}^2 = s^2 \big/ \left(T S_X^2\right)$$

- Claramente, el estimador convencional de la varianza no estima apropiadamente $\sigma_{\widehat{\beta}_1}^2$, y por tanto el error estándar convencional es inapropiado.
  Consecuentemente, los intervalos de confianza y los contrastes de hipótesis habituales son también incorrectos.

- **Eficiencia**
  MCO **no** es, en este contexto, el Estimador Lineal Insesgado de Mínima Varianza.
  (Dado que dicho estimador ignora la estructura de autocorrelación de los errores).

4. Inferencia robusta a autocorrelación: Estimador de la varianza de Newey-West

- La idea de obtener estimadores de la varianza del estimador MCO que sean robustos a autocorrelación se basa en la misma filosofía que la propuesta de Eicker-White para el caso de heterocedasticidad condicional.

- En este sentido, la forma natural de estimar el término adicional

$$\sum_j \sum_{t>j} x_t x_{t-j} \gamma_j = \sum_j \gamma_j \sum_{t>j} x_t x_{t-j}$$

  sería, denotando los residuos del modelo como $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$, $(t = 1, \ldots, T)$:

$$\sum_{j=1}^{J} \frac{1}{T-j} \sum_{t\ (t>j)} x_t x_{t-j} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j}.$$

  que, utilizando el supuesto de que $C\left(\varepsilon_t, \varepsilon_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$, se simplificaría a

$$\sum_{j=1}^{J} \frac{1}{T-j} \left(\sum_{t\ (t>j)} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j} \right) \left(\sum_{t\ (t>j)} x_t x_{t-j} \right).$$

- Un problema es que en principio el sumatorio en j debería ir desde 1 hasta ∞ para recoger todas las posibles covarianzas no nulas, pero en la práctica esto queda limitado por el tamaño muestral T.

  De hecho, J suele escogerse de forma arbitraria, siendo un criterio habitual

$$J = \text{int}\left[(T)^{1/4}\right]$$

  donde $\text{int}(w)$ es la función que determina la parte entera de w.

  A efectos prácticos, la elección se supedita básicamente a la frecuencia de los datos y al número de datos de que disponemos.

  - Datos anuales: J entre 1 y 3.
  - Datos trimestrales: J entre 4 y 12.
  - Datos mensuales: J entre 12 y 36.

- El principal problema del estimador resultante

$$\widetilde{V}\left(\widehat{\beta}_1\right)=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\sum_{j=1}^{J}\frac{1}{T-j}\left(\sum_{t\ (t>j)}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}\right)\left(\sum_{t\ (t>j)}x_tx_{t-j}\right)$$

  es que, para una muestra dada, no está garantizado que la estimación de $\widetilde{V}\left(\widehat{\beta}_1\right)$ sea positiva.

- Newey y West (1987) propusieron un estimador consistente de la varianza que resuelve este problema.

  Para ello, debemos fijar un valor entero J que determina el número de períodos pasados los cuales la dependencia temporal de los errores es nula o poco importante.

  El estimador propuesto sería

$$s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\sum_{j=1}^{J}\left(\frac{j}{J+1}\right)\frac{1}{T-j}\left(\sum_{t\ (t>j)}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}\right)\left(\sum_{t\ (t>j)}x_tx_{t-j}\right)$$

  Por ejemplo:

  - $J=1$

$$s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\left[\frac{1}{2}\left(\frac{1}{T-1}\sum_{t=2}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}\right)\left(\sum_{t=2}^{T}x_tx_{t-1}\right)\right].$$

  - $J=2$

$$\begin{aligned}s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}&\left[\frac{1}{2}\left(\frac{1}{T-1}\sum_{t=2}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}\right)\left(\sum_{t=2}^{T}x_tx_{t-1}\right)\right.\\&\left.+\frac{2}{3}\left(\frac{1}{T-2}\sum_{t=3}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-2}\right)\left(\sum_{t=3}^{T}x_tx_{t-2}\right)\right]\end{aligned}$$

- Newey y West prueban que este estimador es consistente para elecciones de J arbitrarias siempre que J crezca con el tamaño muestral T (intuitivamente, cuanto más larga sea la historia de que dispongamos acerca de los datos, más flexibles podemos ser respecto al patrón de autocorrelación).

- En la práctica, la mayor parte de los programas econométricos incorporan con la estimación MCO o de VI la opción de errores estándar robustos a autocorrelación.

  Algunos de estos programas, como E-Views, toman un valor por defecto de J (es decir, el máximo de períodos relevantes para el cálculo de covarianzas no nulas).

- Además, en el caso de datos series temporales **es posible que existan a la vez heterocedasticidad y autocorrelación**, existiendo un estimador de la varianza robusto a ambas situaciones.

  En este sentido, la mayoría de los programas de ordenador incorporan esta posibilidad.

5. Contrastes de autocorrelación

- Suponga que sospecha de la presencia de autocorrelación, y que desea informarse acerca de esta posibilidad a partir de los datos muestrales.

- Además, suponga que cree que de haber autocorrelación, se caracteriza por ser autocorrelación de primer orden, es decir, $E\left(\varepsilon_t \varepsilon_{t-1}\right) \neq 0$. Ello supondría que si observáramos la serie temporal de ε_t $(t = 1, \ldots, T)$, y efectuáramos la regresión

$$\varepsilon_t = \theta_1 \varepsilon_{t-1} + v_t,$$

el contraste de la hipótesis nula $H_0 : E\left(\varepsilon_t \varepsilon_{t-1}\right) = 0$ sería equivalente a contrastar $H_0 : \theta_1 = 0$.
(Como $E\left(\varepsilon_t\right) = 0$ para todo t, en principio la proyección lineal de ε_t sobre ε_{t-1} no requiere constante).

- Sin embargo, no observamos ε_t, pero sí los residuos del modelo, $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$, que son los análogos muestrales de ε_t.

- Efectuando la regresión lineal MC de $\widehat{\varepsilon}_t$ sobre $\widehat{\varepsilon}_{t-1}$ sin constante, la salida del ordenador mostrará una pendiente estimada

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t}{\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1}^2}$$

y un error estándar de la pendiente que denominaremos $s_{\widehat{\rho}}$.
Para contrastar $H_0 : \theta = 0$, se puede demostrar que $t_\theta = \widehat{\rho}/s_{\widehat{\rho}}$ tiene una distribución aproximada $N(0, 1)$ cuando la hipótesis nula es cierta.
Si t_θ es grande en valor absoluto, tendremos evidencia contra $\theta = 0$ y a favor de la existencia de autocorrelación de primer orden.

- El contraste de autocorrelación de primer orden más popular (que aparece en la mayoría de las salidas de programas econométricos) es el *contraste de Durbin-Watson* d.

  Este contraste se relaciona con el contraste de autocorrelación de primer orden que hemos visto anteriormente en la forma

  $$d \simeq 2(1 - \widehat{\rho}).$$

  Sin embargo, este contraste tiene el inconveniente de que los valores críticos del estadístico dependen del tamaño muestral.

  Además, su interpretación es contraintuitiva, porque valores *pequeños* del estadístico de Durbin-Watson sugieren *alta* autocorrelación positiva, y viceversa.

- En la práctica, podemos contrastar la existencia de autocorrelación de orden superior.

  Por ejemplo, si queremos contrastar la posibilidad de que exista autocorrelación hasta un orden q, podemos efectuar la regresión

  $$\widehat{\varepsilon}_t = \theta_0 + \theta_1 \widehat{\varepsilon}_{t-1} + \theta_2 \widehat{\varepsilon}_{t-2} + \cdots + \theta_q \widehat{\varepsilon}_{t-q} + v_t,$$

  donde $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$ son los residuos MCO, y contrastar la hipótesis nula $H_0 : \theta_1 = \theta_2 = \cdots = \theta_q = 0$.

  Intuitivamente, este contraste es en la práctica un contraste de regresión.