

# Autocorrelación

Román Salmerón Gómez

Universidad de Granada



Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

**Naturaleza del problema**

**Causas y consecuencias de la autocorrelación**

**Procedimientos de Detección**

**Estimación en los modelos con autocorrelación**

Contenidos

Naturaleza del problema

Autocorrelación

Causas y  
consecuencias de la  
autocorrelación

Procedimientos de  
Detección

Estimación en los  
modelos con  
autocorrelación

# Naturaleza del problema

En el modelo lineal general,  $y = X\beta + u$ , se supone que la perturbación aleatoria es tal que  $E[u] = 0_{n \times 1}$  y  $Var(u) = E[u \cdot u^t] = \sigma^2 I_{n \times n}$ , lo cual implica que:

- $E[u_t] = 0, \forall t \in \{1, \dots, n\}$ .
- $E[u_t^2] = Var(u_t) = \sigma^2, \forall t \in \{1, \dots, n\}$  (varianza constante = homocedasticidad).
- $E[u_i \cdot u_j] = Cov(u_i, u_j) = 0, \forall i \neq j \in \{1, \dots, n\}$  (incorrelación).

Cuando se incumple el supuesto de incorrelación, es decir, la covarianza de la perturbación aleatoria es no nula para dos instantes de tiempo distintos,  $E[u_i \cdot u_j] \neq 0, \forall i \neq j$  o, equivalentemente,  $E[u_t \cdot u_{t-k}] \neq 0, \forall k > 0$ , se dice que hay autocorrelación.

En tal caso, los elementos de fuera de la diagonal principal de la matriz de varianzas-covarianzas no son todos nulos (hay al menos un elemento no nulo).

Este problema aparece especialmente cuando se disponen de datos de series temporales, es decir, cuando se disponen de observaciones que miden una variable de una entidad (individuos, familias, empresas, etc.) a lo largo del tiempo.

## Contenidos

### Naturaleza del problema

Autocorrelación

### Causas y consecuencias de la autocorrelación

### Procedimientos de Detección

### Estimación en los modelos con autocorrelación

Así, por ejemplo, al estudiar la productividad de una empresa, la aparición de una nueva máquina en un momento determinado, además de producir un efecto en dicho instante de tiempo, lo producirá también en sucesivos. No es factible pensar que dicho efecto vaya a desaparecer en instantes de tiempo sucesivos. Esto implicará que las perturbaciones en ambos momentos están correlacionados entre sí.

Uriel, E. y otros (1990). Econometría. El modelo lineal.

Similar interpretación merece el fichaje de un jugador estrella por un equipo de fútbol. El incremento (deseable) en ventas de camisetas con el dorsal que llevará el nuevo jugador, además de afectar al momento del anuncio de su fichaje, se presupone que influirá también en la venta en el futuro.

... este es mío

Consideraremos entonces el modelo lineal general

$$y_{n \times 1} = X_{n \times k} \cdot \beta_{k \times 1} + u_{n \times 1},$$

tal que  $E[u] = 0_{n \times 1}$  y  $Var(u) = \sigma^2 \cdot \Omega_{n \times n}$  donde  $\Omega$  es una matriz cuya diagonal principal es constantemente 1 y de los elementos que no pertenecen a dicha diagonal principal hay al menos un elemento no nulo.

Al igual que en el tema anterior, se dice que el modelo tiene una matriz de varianzas-covarianzas no escalar o con perturbaciones no esféricas (ya que la matriz de varianzas-covarianzas de la perturbación aleatoria es distinta al producto de una constante por la matriz diagonal).

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Ejemplo

Procedimientos de Detección

Estimación en los modelos con autocorrelación

# Causas y consecuencias de la autocorrelación

# Causas y consecuencias de la autocorrelación

## Contenidos

### Naturaleza del problema

### Causas y consecuencias de la autocorrelación

### Ejemplo

### Procedimientos de Detección

### Estimación en los modelos con autocorrelación

Las principales causas que provocan autocorrelación en un modelo lineal son:

- Existencia de variables endógenas retardadas.
- Omisión de variables relevantes: la perturbación aleatoria contendrá a la variable excluida ocasionando un patrón de correlación.
- Si se especifica una relación funcional errónea (por ejemplo, una relación lineal cuando no lo es), el término de perturbación captará tal efecto provocando autocorrelación en el modelo.
- Las técnicas de manipulación de datos (interpolación, promedios, etc.) pueden introducir un patrón sistemático en el modelo que conduzca a la autocorrelación.
- Naturaleza del fenómeno: con datos correspondientes a series de tiempo (se observa una variable a lo largo del tiempo) es probable que observaciones sucesivas sean dependientes entre sí provocando autocorrelación.

La presencia de autocorrelación en un modelo lineal provoca, al igual que en el caso de la heteroscedasticidad, que los estimadores obtenidos no sean óptimos (aunque si sean lineales e insesgados).



Supongamos que el modelo correcto para explicar  $Y_t$  es

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

y que erróneamente se especifica el modelo

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + v_t.$$

En tal caso, el término de error del segundo modelo dependerá de la variable omitida (por ejemplo,  $v_t = \beta_3 X_{3t} + u_t$ ). Tal relación conducirá a la presencia de autocorrelación en el modelo:

$$\begin{aligned} Cov(v_t, v_{t-k}) &= E[v_t \cdot v_{t-k}] = E[(\beta_3 X_{3t} + u_t) \cdot (\beta_3 X_{3t-k} + u_{t-k})] \\ &= \beta_3^2 X_{3t} X_{3t-k} + \beta_3 X_{3t} E[u_{t-k}] \\ &\quad + \beta_3 X_{3t-k} E[u_t] + E[u_t \cdot u_{t-k}] \\ &= \beta_3^2 X_{3t} X_{3t-k} \neq 0, \forall k > 0, \end{aligned}$$

donde se ha tenido en cuenta que la perturbación aleatoria del primer modelo tiene media cero y está incorrelada.

Contenidos

Naturaleza del problema

Causas y  
consecuencias de la  
autocorrelación

**Procedimientos de  
Detección**

Métodos gráficos

Métodos analíticos

Estimación en los  
modelos con  
autocorrelación

# Procedimientos de Detección

Para detectar la autocorrelación en un modelo lineal múltiple disponemos de distintos procedimientos.

En primer lugar usaremos métodos gráficos a partir de los cuales intentaremos intuir cuáles son las variables que provocan la existencia de autocorrelación en el modelo. Ya que las perturbaciones aleatorias no son observables, usaremos los residuos de la estimación por MCO (al igual que para detectar la heteroscedasticidad). Concretamente, analizaremos el gráfico temporal de los residuos y el gráfico de dispersión de los mismos frente a algún retardo suyo.

Puesto que tomar una decisión a partir de un procedimiento gráfico no es muy adecuado ya que son fácilmente manipulables y sería totalmente subjetiva, recurriremos a métodos analíticos para determinar la presencia de heteroscedasticidad en el modelo. De todos los métodos analíticos disponibles, el más utilizado, y que estudiaremos, es el de Durbin-Watson. Además, como este método no es adecuado cuando existen variables retardadas como explicativas (ya que entonces tiende a indicar ausencia de autocorrelación), se estudia entonces el contraste  $h$  de Durbin. También usaremos el contraste de Ljung-Box.

Dentro de los procedimientos gráficos consideraremos:

**Gráfico temporal de los residuos:** si los residuos están incorrelados deben distribuirse de forma aleatoria alrededor de su media (el cero). Sin embargo, si están correlacionados:

- observaremos rachas de residuos por debajo y por encima de la media (correlación positiva).
- observaremos una alternancia en el signo de los residuos (correlación negativa).

**Gráficos de dispersión:** el gráfico de dispersión de los residuos,  $e_t$ , frente algún retardo suyo,  $e_{t-k}$  (normalmente se considera  $k = 1$ ), puede mostrar la existencia de autocorrelación positiva (tendencia creciente en el gráfico) o negativa (tendencia decreciente).

**Funciones de autocorrelación y autocorrelación parcial:** Si dichas funciones, ACF y ACP, corresponden a ruido blanco, se concluirá que la perturbación es incorrelada.

## Contenidos

Naturaleza del problema

## Causas y consecuencias de la autocorrelación

## Procedimientos de Detección

## Métodos gráficos

## Métodos analíticos

Estimación en los modelos con autocorrelación

Consideremos los datos de la tabla 1 para ajustar un modelo que analice el consumo de energía eléctrica (en miles de TEP) a partir del PIB a precios constantes (millones de euros).

Dada la naturaleza del problema (serie temporal), tal y como se ha indicado, sospechamos la posible presencia de autocorrelación en el modelo. Por tal motivo, en primer lugar usaremos los métodos gráficos para intentar detectarla. Usaremos con tal objetivo los residuos de la estimación por MCO del modelo:

$$\hat{C}_t = -6234'54 + 0'0426873 \cdot PIB_t,$$

$$R^2 = 0'992408$$

Tabla 1: Datos observados

Año	Consumo	PIB
1987	9427	355312
1988	9876	373412
1989	10410	391443
1990	10974	406252
1991	11372	416582
1992	11488	420462
1993	11569	416126
1994	11999	426041
1995	12462	437787
1996	12827	448457
1997	13331	466513
1998	14290	486785
1999	15364	507346
2000	16309	528714
2001	17282	543746
2002	17756	554852

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

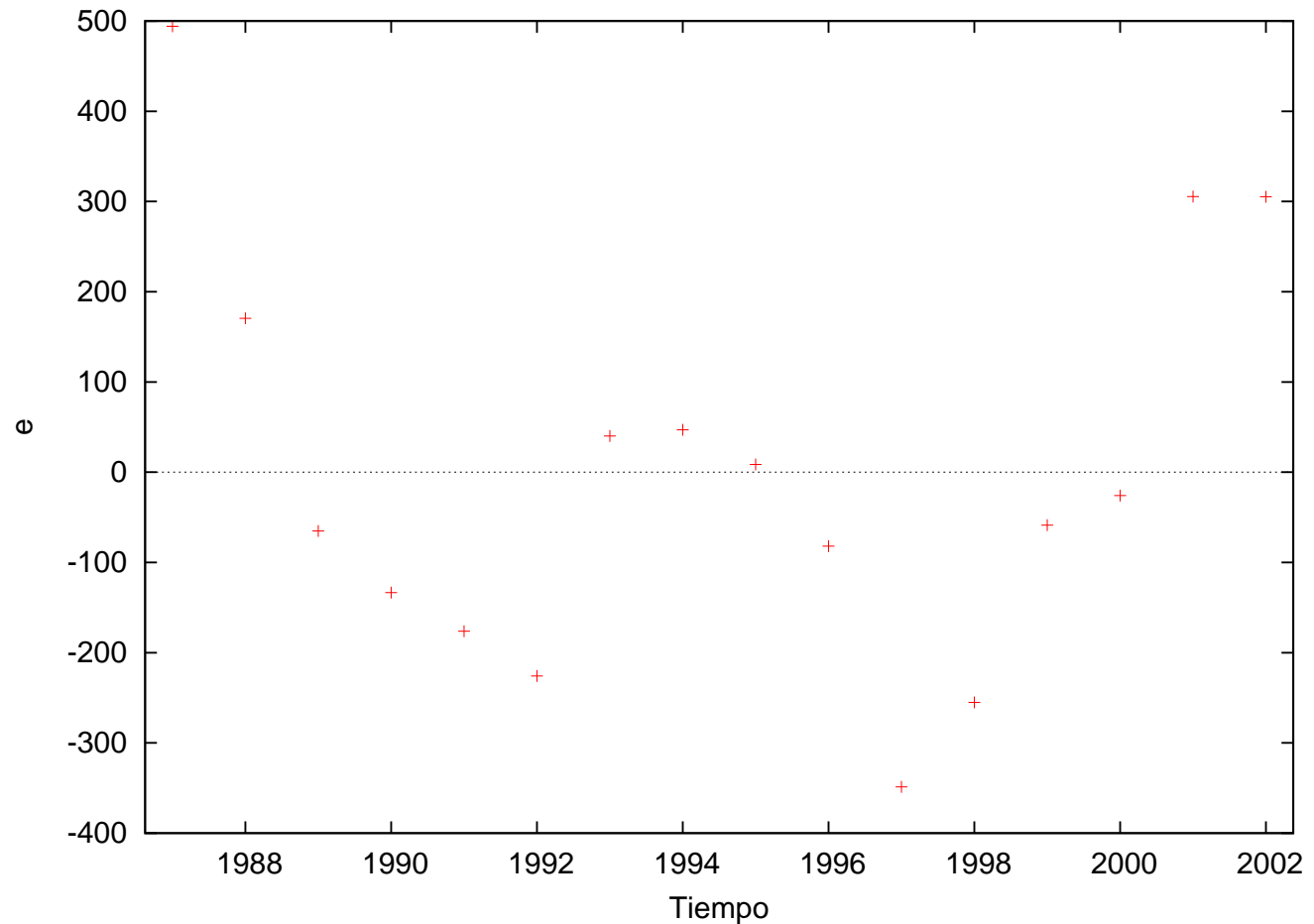


Figura 1: Gráfico temporal de los residuos

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

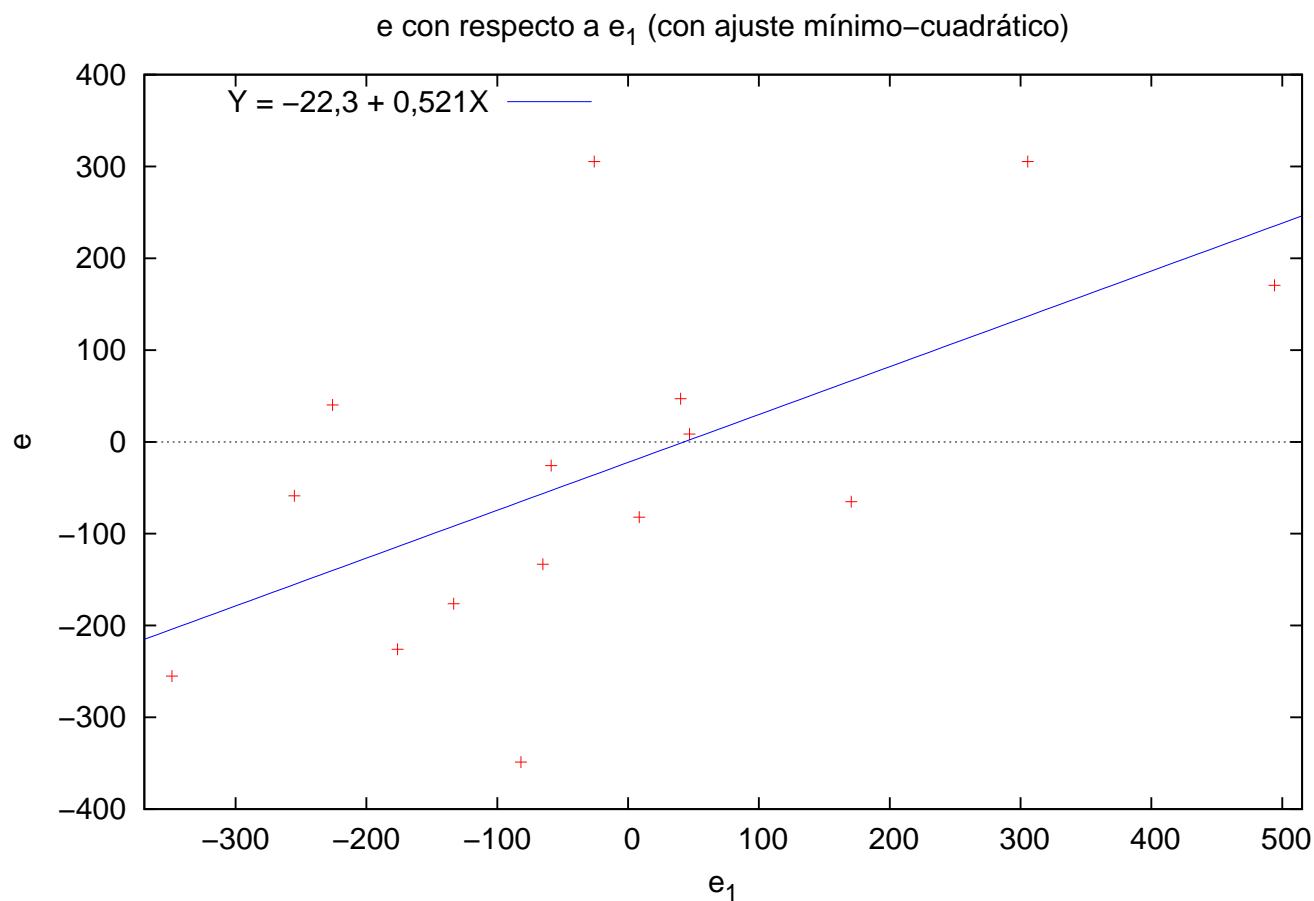


Figura 2: Gráfico de dispersión

## Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

En el gráfico de la figura 1 observamos rachas de residuos por encima y por debajo de la media (cero), mientras que en el de la figura 2 observamos una tendencia claramente creciente. Por tanto, podemos pensar que hay presencia de autocorrelación positiva en el modelo.

Para confirmar este hecho recurriremos a los procedimientos analíticos del contraste de Durbin-Watson y de Ljung-Box.



# Contraste de Durbin-Watson

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

Supongamos que la autocorrelación de la perturbación aleatoria viene definida por un proceso autorregresivo de primer orden, esto es:

$$u_t = \rho u_{t-1} + v_t, \quad v_t \sim N(0, \sigma^2), \forall t.$$

Luego para contrastar si realmente hay autocorrelación en el modelo hay que plantear los siguientes contrastes de hipótesis:

$$H_0 : \rho = 0 \text{ (incorrelación)}$$

$$H_0 : \rho = 0 \text{ (incorrelación)}$$

$$H_1 : \rho > 0 \text{ (correlación positiva)}$$

$$H_1 : \rho < 0 \text{ (correlación negativa)}$$

Para tomar una decisión en los contrastes anteriores utilizaremos el estadístico de Durbin-Watson, que se define como:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}, \quad (1)$$

donde  $e$  denota a los residuos del modelo estimado por MCO.

# Aproximación al estadístico de Durbin-Watson

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

Teniendo en cuenta que para un número de observaciones,  $n$ , grande podemos

establecer que  $\sum_{t=2}^n e_t^2 \simeq \sum_{t=2}^n e_{t-1}^2 \simeq \sum_{t=1}^n e_t^2$ , y denotando  $\rho = \frac{\sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2}$ , es

claro que:

$$\begin{aligned} d &= \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2} \\ &\simeq \frac{2 \cdot \sum_{t=2}^n e_t^2 - 2 \sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2} = 2 \cdot (1 - \rho). \end{aligned}$$

Obsérvese que  $\rho$  corresponde a la pendiente (y coeficiente de correlación) de la regresión  $e_t = \rho e_{t-1} + \epsilon_t$ .

# Interpretación estadístico de Durbin-Watson

## Contenidos

### Naturaleza del problema

### Causas y consecuencias de la autocorrelación

### Procedimientos de Detección

#### Métodos gráficos

#### Métodos analíticos

### Estimación en los modelos con autocorrelación

Teniendo en cuenta que  $\rho$  es el coeficiente de correlación de los residuos, se tiene que:

- habrá correlación negativa en los residuos cuando  $\rho$  esté próximo a -1, lo cual se traduce en que  $d$  sea próximo a 4.
- habrá incorrelación en los residuos cuando  $\rho$  esté próximo a 0, lo cual se traduce en que  $d$  sea próximo a 2.
- habrá correlación positiva en los residuos cuando  $\rho$  esté próximo a 1, lo cual se traduce en que  $d$  sea próximo a 0.

Pero, ¿cómo de próximo a los valores 0, 2 y 4 se ha de estar? Durbin y Watson encontraron unas cotas,  $d_L$  y  $d_U$ , tales que:

- si  $d < d_L$ , entonces habrá autocorrelación positiva.
- si  $d_U < d < 4 - d_U$ , entonces hay incorrelación.
- si  $d > 4 - d_L$ , entonces hay autocorrelación negativa.
- en cualquier otro caso el contraste no es concluyente.

# Cotas del estadístico de Durbin-Watson

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Métodos gráficos

Métodos analíticos

Estimación en los modelos con autocorrelación

Representando gráficamente la información anterior se obtiene:

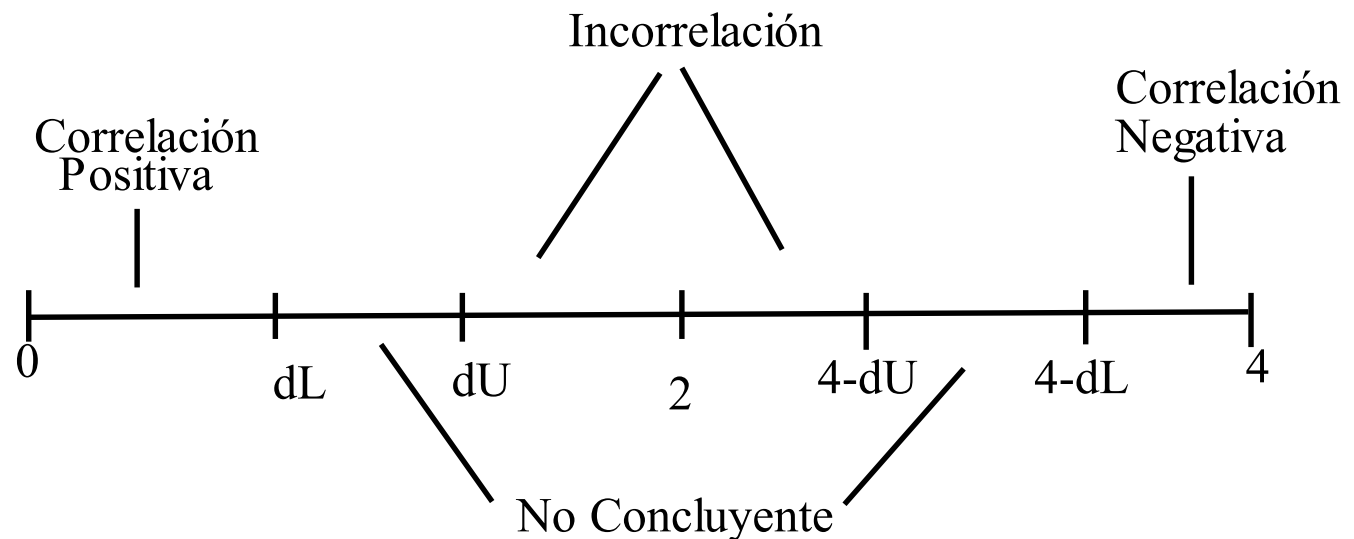


Figura 3: Cotas del estadístico de Durbin-Watson

Considerando el modelo que ajusta el consumo energético en función del PIB, tras estimarlo por MCO

$$\hat{C}_t = -6234'54 + 0'0426873 \cdot PIB_t, \quad R^2 = 0'992408$$

(451'562) (0'0009978)

y calcular sus residuos, podemos obtener la siguiente información:

Año	$e_t$	$e_{t-1}$	$e_t^2$	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	$e_t \cdot e_{t-1}$
1987	494'1584		244192'5			
1988	170'5190	494'1584	29076'7	-323'6394	104742'4	84263'40
1989	-65'1749	170'5190	4247'8	-235'6939	55551'6	-11113'56
1990	-133'3305	-65'1749	17777'0	-68'1556	4645'2	8689'80
1991	-176'2899	-133'3305	31078'1	-42'9594	1845'5	23504'82
1992	-225'9164	-176'2899	51038'2	-49'6266	2462'8	39826'78
1993	40'1755	-225'9164	1614'1	266'0919	70804'9	-9076'31
1994	46'9314	40'1755	2202'6	6'7558	45'6	1885'49
1995	8'5268	46'9314	72'7	-38'4045	1474'9	400'18
1996	-81'9462	8'5268	6715'2	-90'4730	8185'4	-698'74
1997	-348'7073	-81'9462	121596'8	-266'7611	71161'5	28575'23
1998	-255'0634	-348'7073	65057'3	93'6439	8769'2	88942'46
1999	-58'7561	-255'0634	3452'3	196'3073	38536'6	14986'52
2000	-25'8974	-58'7561	670'7	32'8587	1079'7	1521'63
2001	305'4278	-25'8974	93286'1	331'3252	109776'4	-7909'78
2002	305'3431	305'4278	93234'4	-0'0847	0'0072	93260'26
Total			765312'5		479089'5	357073'5

A partir de dicha información se puede calcular el valor del estadístico de Durbin-Watson:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{479089'5}{765312'5} = 0'6260051,$$

y de su aproximación:

$$\begin{aligned} d &\simeq 2 \cdot \left( 1 - \frac{\sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2} \right) = 2 \cdot \left( 1 - \frac{357073'5}{765312'5} \right) \\ &= 2 \cdot (1 - 0'4665722) = 2 \cdot 0'5334278 = 1'066856. \end{aligned}$$

Comparando estos valores con las cotas de Durbin-Watson para 16 observaciones al 5% de significación,  $d_L = 1'1062$  y  $d_U = 1'3709$ , se tiene que hay autocorrelación positiva ya que  $d < d_L$ .

El contraste de Durbin-Watson es válido cuando la autocorrelación de los errores es autorregresiva de orden 1 y cuando la regresión no incluye entre las variables explicativas algún retardo de la variable dependiente.

En este segundo caso se recurre al contraste  $h$  de Durbin. Esto es, en modelos del tipo

$$Y_t = \alpha Y_{t-1} + \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t,$$

donde  $u_t = \rho u_{t-1} + v_t$ , para contrastar

$$H_0 : \rho = 0 \text{ (incorrelación)}$$

$$H_1 : \rho \neq 0 \text{ (correlación)}$$

utilizaremos el estadístico  $h = \rho \cdot \sqrt{\frac{n}{1 - n \cdot \widehat{Var}(\hat{\alpha})}} \sim N(0, 1)$ , de forma que se rechazará la hipótesis nula si  $|h| > Z_{1 - \frac{\alpha}{2}}$ .

Si los residuos son independientes, sus primeras  $m$  autocorrelaciones son cero, para cualquier valor de  $m$ . El contraste de Ljung-Box contrasta, entonces, la hipótesis nula de que las primeras  $m$  autocorrelaciones,  $\rho_m$ , son cero. Esto es:

$$\left. \begin{array}{l} H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0 \\ H_1 : \exists i \in \{1, 2, \dots, m\} \text{ tal que } \rho_i \neq 0 \end{array} \right\}$$

Se rechaza la hipótesis nula (de incorrelación) si  $Q_{LB} = n(n+2) \sum_{s=1}^m \frac{r(s)^2}{n-s} > \chi_{m-1}^2(1-\alpha)$  donde

$$r(s) = \frac{\sum_{t=s+1}^n e_t e_{t-s}}{\sum_{t=1}^n e_t^2},$$

es el coeficiente de autocorrelación muestral de orden  $k$ .

Si las observaciones son independientes (incorrelación), los coeficientes  $r(s)$  serán próximos a cero, por lo que no se rechazaría la hipótesis nula. Por otro lado, el valor de  $m$  puede ser fijado arbitrariamente aunque no debe de ser grande.



Consideremos el modelo que ajusta el consumo energético en función del PIB para aplicar el anterior contraste para  $m = 3$ .

De este ejemplo ya sabemos que  $\sum_{t=1}^n e_t^2 = 765312'5$ . Además, a partir de los residuos podemos construir la siguiente tabla:

Año	$e_t$	$e_{t-1}$	$e_{t-2}$	$e_{t-3}$	$e_t \cdot e_{t-1}$	$e_t \cdot e_{t-2}$	$e_t \cdot e_{t-3}$
1987	494'1584						
1988	170'5190	494'1584			84263'40		
1989	-65'1749	170'5190	494'1584		-11113'56	-32206'724	
1990	-133'3305	-65'1749	170'5190	494'1584	8689'80	-22735'383	-65886'386
1991	-176'2899	-133'3305	-65'1749	170'5190	23504'82	11489'676	-30060'777
1992	-225'9164	-176'2899	-133'3305	-65'1749	39826'78	30121'546	14724'078
1993	40'1755	-225'9164	-176'2899	-133'3305	-9076'31	-7082'534	-5356'619
1994	46'9314	40'1755	-225'9164	-176'2899	1885'49	-10602'572	-8273'531
1995	8'5268	46'9314	40'1755	-225'9164	400'18	342'568	-1926'343
1996	-81'9462	8'5268	46'9314	40'1755	-698'74	-3845'849	-3292'229
1997	-348'7073	-81'9462	8'5268	46'9314	28575'23	-2973'357	-16365'321
1998	-255'0634	-348'7073	-81'9462	8'5268	88942'46	20901'476	-2174'874
1999	-58'7561	-255'0634	-348'7073	-81'9462	14986'52	20488'681	4814'839
2000	-25'8974	-58'7561	-255'0634	-348'7073	1521'63	6605'478	9030'612
2001	305'4278	-25'8974	-58'7561	-255'0634	-7909'78	-17945'746	-77903'453
2002	305'3431	305'4278	-25'8974	-58'7561	93260'26	-7907'592	-17940'769
Total					357073'5	-15350'3338	-200610'77

## Contenidos

Naturaleza del problemaCausas y consecuencias de la autocorrelaciónProcedimientos de Detección

## Métodos gráficos

## Métodos analíticos

Estimación en los modelos con autocorrelación

A partir de la cual podemos obtener:

$$r(1) = \frac{\sum_{t=2}^n e_t \cdot e_{t-1}}{\sum_{t=1}^n e_t^2} = \frac{357073'5}{765312'5} = 0'4666,$$

$$r(2) = \frac{\sum_{t=3}^n e_t \cdot e_{t-2}}{\sum_{t=1}^n e_t^2} = \frac{-15350'3338}{765312'5} = -0'0201,$$

$$r(3) = \frac{\sum_{t=4}^n e_t \cdot e_{t-3}}{\sum_{t=1}^n e_t^2} = \frac{-200610'77}{765312'5} = -0'2622.$$

De forma que:

$$\begin{aligned}
 Q_{LB} &= n(n+2) \sum_{s=1}^m \frac{r(s)^2}{n-s} = 16 \cdot 18 \cdot \left( \frac{r(1)^2}{15} + \frac{r(2)^2}{14} + \frac{r(3)^2}{13} \right) \\
 &= 288 \cdot \left( \frac{0'21771556}{15} + \frac{0'00040401}{14} + \frac{0'06874884}{13} \right) \\
 &= 288 \cdot (0'01452 + 0'000028 + 0'0053) = 288 \cdot 0'01983 \\
 &= 5'7115.
 \end{aligned}$$

Como  $Q_{LB} = 5'7115 \not\geq 5'991 = \chi_2^2(0'95)$ , no rechazo la hipótesis nula de incorrelación.

Adviértase que para  $m = 2$ , se tiene que  $Q_{LB} = 4'1885 > 3'841 = \chi_1^2(0'95)$ . Por tanto, en tal caso si se rechazaría la hipótesis nula de incorrelación.

Contenidos

Naturaleza del problema

Causas y  
consecuencias de la  
autocorrelación

Procedimientos de  
Detección

Estimación en los  
modelos con  
autocorrelación

Estimación bajo  
autocorrelación

Ejemplo

# Estimación en los modelos con autocorrelación

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

Consideremos el modelo lineal uniecuacional múltiple

$$y = X\beta + u,$$

donde  $E[u] = 0_{n \times 1}$ ,  $E[u_t^2] = \sigma^2$ ,  $t = 1, \dots, n$  (homocedasticidad) y  $E[u_i \cdot u_j] = \sigma_{ij}$ ,  $\forall i \neq j \in \{1, \dots, n\}$  (autocorrelación).

Por tanto, la matriz de varianzas-covarianzas de la perturbación aleatoria será

$$\text{Var}(u) = E[u \cdot u^t] = \begin{pmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \cdot \Omega.$$

Para resolver el problema de autocorrelación hay que aplicar el método de mínimos cuadrados generalizados al modelo, para lo cual debemos obtener una matriz no estocástica  $P$  tal que  $P^t \cdot P = \Omega^{-1}$ .

Para conseguir el objetivo anterior supondremos que la estructura de autocorrelación viene marcada por un proceso autorregresivo de orden 1, esto es,  $u_t = \rho \cdot u_{t-1} + v_t$ .

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

En tal caso:

$$\begin{aligned} Cov(u_t, u_{t-k}) &= E[u_t \cdot u_{t-k}] = E[(\rho \cdot u_{t-1} + v_t) \cdot u_{t-k}] \\ &= \rho \cdot E[u_{t-1} \cdot u_{t-k}] + E[v_t \cdot u_{t-k}] \\ &= \rho \cdot E[u_{t-1} \cdot u_{t-k}], \end{aligned}$$

por lo que repitiendo el cálculo de forma iterativa:

$$Cov(u_t, u_{t-k}) = \rho^k \cdot \sigma^2, \quad k > 0.$$

Entonces, la matriz de varianzas covarianzas de la perturbación aleatoria viene determinada por la matriz:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.$$

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

Adviértase que la suposición de que la perturbación aleatoria sigue el esquema de un proceso AR(1) es fundamental para que se pueda estimar el modelo. Teniendo en cuenta la matriz  $\Omega$  original, tendríamos que estimar  $\frac{n^2-n}{2} + 1 + k$  parámetros:  $\frac{n^2-n}{2}$  de la matriz  $\Omega$ , 1 correspondiente a  $\sigma^2$  y  $k$  de los coeficientes de los regresores. Por tanto, se tendrían más parámetros a estimar que observaciones. Cuestión que no es viable.

Sin embargo, a partir de la suposición realizada sobre la perturbación aleatoria habría que estimar  $k + 2$  parámetros: 1 de la matriz  $\Omega$  (que es  $\rho$ ), 1 correspondiente a  $\sigma^2$  y  $k$  de los coeficientes de los regresores. Por lo que la estimación si sería viable.

Entonces, teniendo en cuenta que:

$$\Omega^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1 + \rho & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1 + \rho & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 + \rho & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 0 \end{pmatrix},$$

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

la matriz buscada es para realizar la transformación es:

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

El modelo transformado a partir de la matriz  $P$  anterior,  $y^* = X^*\beta + u^*$ , es tal que

$$y^* = Py = \begin{pmatrix} \sqrt{1 - \rho^2} \cdot Y_1 \\ Y_2 - \rho \cdot Y_1 \\ \vdots \\ Y_n - \rho \cdot Y_{n-1} \end{pmatrix}, \quad u^* = Pu = \begin{pmatrix} \sqrt{1 - \rho^2} \cdot u_1 \\ u_2 - \rho \cdot u_1 \\ \vdots \\ u_n - \rho \cdot u_{n-1} \end{pmatrix},$$



# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

$$X^* = PX = \begin{pmatrix} \sqrt{1 - \rho^2} & \sqrt{1 - \rho^2} \cdot X_{21} & \cdots & \sqrt{1 - \rho^2} \cdot X_{k1} \\ 1 - \rho & X_{22} - \rho \cdot X_{21} & \cdots & X_{k2} - \rho \cdot X_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 - \rho & X_{2n} - \rho \cdot X_{2n-1} & \cdots & X_{kn} - \rho \cdot X_{kn-1} \end{pmatrix},$$

por lo que las nuevas observaciones responden a las expresiones:

$$Y_t^* = \begin{cases} \sqrt{1 - \rho^2} \cdot Y_1 & t = 1 \\ Y_t - \rho \cdot Y_{t-1} & t > 1 \end{cases}, \quad X_{it}^* = \begin{cases} \sqrt{1 - \rho^2} \cdot X_{i1} & t = 1 \\ X_{it} - \rho \cdot X_{it-1} & t > 1 \end{cases},$$

con  $i = 1, \dots, k$ .

Ahora bien, para obtener estos valores del modelo transformado es necesario estimar  $\rho$ , pero para esto hay que estimar también los coeficientes. ¿Cómo resolver este problema? Mediante el proceso iterativo de Prais-Winsten que describimos a continuación.

Proceso iterativo de Prais-Winsten:

1. Estimar el modelo original por MCO.
2. Utilizar los residuos del ajuste anterior para estimar  $\rho$ .
3. Utilizar la estimación de  $\rho$  para obtener  $Y_t^*$  y  $X_{it}^*$ .
4. Estimar el modelo transformado por MCO volviendo al paso 1.

Repetir este proceso hasta que la diferencia entre dos estimaciones consecutivas de  $\rho$  sea menor que un valor prefijado (de orden  $10^{-3}$  normalmente).

En los casos en los que se disponga de un número de observaciones suficientemente grande se puede despreciar la primera observación (perdiéndola) transformando los datos según

$$Y_t^* = Y_t - \rho \cdot Y_{t-1}, \quad X_{it}^* = X_{it} - \rho \cdot X_{it-1},$$

para  $t > 1$  e  $i = 1, \dots, k$ . En tal caso se puede establecer un proceso iterativo similar al anterior que recibe el nombre de Cochrane-Orcutt.

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

Finalmente, comprobemos que con esta transformación realmente se ha eliminado el problema en el modelo.

Para  $t = 1$ , se tiene que  $u_1^* = \sqrt{1 - \rho^2} \cdot u_1$ , por lo que:

$$\begin{aligned} E[u_1^*] &= \sqrt{1 - \rho^2} \cdot E[u_1] = 0, \\ Var(u_1^*) &= E[(1 - \rho^2) \cdot u_1^2] = (1 - \rho^2) \cdot E[u_1^2] = (1 - \rho^2) \cdot \sigma^2, \\ Cov(u_2^*, u_1^*) &= E \left[ (u_2 - \rho u_1) \cdot \sqrt{1 - \rho^2} \cdot u_1 \right] \\ &= \sqrt{1 - \rho^2} E[u_2 u_1] - \rho \sqrt{1 - \rho^2} \cdot E[u_1^2] \\ &= \rho \sqrt{1 - \rho^2} \cdot \sigma^2 - \rho \sqrt{1 - \rho^2} \cdot \sigma^2 = 0. \end{aligned}$$

Mientras que para  $t > 1$ , se tiene que  $u_t^* = u_t - \rho u_{t-1}$  y entonces:

$$\begin{aligned} E[u_t^*] &= E[u_t] - \rho E[u_{t-1}] = 0, \\ Var(u_t^*) &= E[u_t^2 + \rho^2 u_{t-1}^2 - 2\rho u_t u_{t-1}] =, \\ &= E[u_t^2] + \rho^2 E[u_{t-1}^2] - 2\rho E[u_t u_{t-1}] = \\ &= \sigma^2 + \rho^2 \sigma^2 - 2\rho \sigma^2 = \sigma^2 - \rho^2 \sigma^2 = (1 - \rho^2) \cdot \sigma^2, \end{aligned}$$

# Estimación bajo autocorrelación

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

$$\begin{aligned} Cov(u_t^*, u_{t-1}^*) &= E[u_t^* u_{t-1}^*] = E[(u_t - \rho u_{t-1}) \cdot (u_{t-1} - \rho u_{t-2})] \\ &= E[u_t u_{t-1}] - \rho E[u_t u_{t-2}] - \rho E[u_{t-1}^2] + \rho^2 E[u_{t-1} u_{t-2}] \\ &= \rho \sigma^2 - \rho \rho^2 \sigma^2 - \rho \sigma^2 + \rho^2 \rho \sigma^2 = 0. \end{aligned}$$

Por tanto, el modelo transformado tiene una perturbación aleatoria con media cero, varianza constante,  $(1 - \rho^2) \cdot \sigma^2$ , e incorrelada. Luego el problema ha sido resuelto.

## Contenidos

## Naturaleza del problema

## Causas y consecuencias de la autocorrelación

## Procedimientos de Detección

## Estimación en los modelos con autocorrelación

## Estimación bajo autocorrelación

## Ejemplo

Dado el modelo que analiza el consumo energético a partir del PIB, en el que sabemos que hay autocorrelación positiva, vamos a transformarlo de forma adecuada para subsanar dicho problema.

A partir de la estimación por MCO del modelo original y considerando los residuos de dicho ajuste, obtenemos una estimación de  $\rho$  sin más que ajustar el modelo de  $e_t$  frente a  $e_{t-1}$ , obteniéndose:

$$\hat{e}_t = 0'531275 \cdot e_{t-1}, \quad R^2 = 0'364016 \\ (0'18768)$$

Es decir,  $\rho = 0'531275$ .

Como se dispone de pocas observaciones, transformaremos los datos conforme al proceso de Prais-Winsten:

$$C_1^* = \sqrt{1 - 0'531275^2} \cdot C_1 = 0'8471994 \cdot C_1, \quad C_t^* = C_t - 0'531275 \cdot C_{t-1}, \quad t > 1,$$

$$cte_1^* = \sqrt{1 - 0'531275^2} = 0'8471994, \quad cte_t^* = 1 - 0'531275 = 0'468725, \quad t > 1,$$

$$PIB_1^* = \sqrt{1 - 0'531275^2} \cdot PIB_1 = 0'8471994 \cdot PIB_1,$$

$$PIB_t^* = PIB_t - 0'531275 \cdot PIB_{t-1}, \quad t > 1.$$

## Contenidos

## Naturaleza del problema

## Causas y consecuencias de la autocorrelación

## Procedimientos de Detección

## Estimación en los modelos con autocorrelación

## Estimación bajo autocorrelación

## Ejemplo

Los datos del modelo transformado serán:

Año (t)	$C_t^*$	$PIB_t^*$	$cte_t^*$
1987	7986'549	301020'1	0'8471994
1988	4867'671	184643'6	0'4687250
1989	5163'128	193058'5	0'4687250
1990	5443'427	198288'1	0'4687250
1991	5541'788	200750'5	0'4687250
1992	5446'341	199142'4	0'4687250
1993	5465'713	192745'1	0'4687250
1994	5852'680	204963'7	0'4687250
1995	6087'231	211442'1	0'4687250
1996	6206'251	215871'7	0'4687250
1997	6516'336	228259'0	0'4687250
1998	7207'573	238938'3	0'4687250
1999	7772'080	248729'3	0'4687250
2000	8146'491	259173'8	0'4687250
2001	8617'436	262853'5	0'4687250
2002	8574'505	265973'3	0'4687250

A partir de los valores transformados se obtiene la estimación:

$$\hat{C}_t^* = -6004'45 + 0'042286 \cdot PIB_t^*, \quad R^2 = 0'999307$$

$$(622'719) \quad (0'00136949)$$

y un valor del estadístico de Durbin-Watson,  $d$ , de 1'208017. Puesto que  $d_L < d < d_U$ , el contraste de Durbin-Watson no es concluyente.

A continuación, siguiendo el proceso iterativo, habría que volver a estimar  $\rho$  a partir de este modelo transformado y volver a transformar los datos. Y así sucesivamente hasta que el proceso converja.

En este caso es necesario realizar 12 iteraciones para la convergencia obteniéndose la estimación:

$$\hat{C}_t^* = -5794'48 + 0'0419213 \cdot PIB_t^*, \quad R^2 = 0'996197$$

$$(785'107) \quad (0'001717)$$

y un valor del estadístico de Durbin-Watson,  $d$ , de 1'400435. Puesto que  $d_U < d < 4 - d_U$ , el contraste de Durbin-Watson confirma la ausencia de autocorrelación en el modelo.

Al calcular la aproximación del estadístico de Durbin-Watson se obtuvo que  $\rho = 0'4665722$ , mientras que al hacer la regresión de  $e_t$  frente a  $e_{t-1}$  hemos obtenido que  $\rho = 0'531275$ . Si ambos métodos de obtener el valor de  $\rho$  son equivalentes, ¿a qué se debe esta diferencia?

Si partimos de la regresión de los residuos se tiene que:

$$\begin{aligned} \rho &= \frac{Cov(e_{t-1}, e_t)}{Var(e_{t-1})} = \frac{\frac{1}{T} \cdot \sum_{t=2}^n e_t e_{t-1} - \left( \frac{1}{T} \cdot \sum_{t=1}^n e_t \right) \cdot \left( \frac{1}{T-1} \cdot \sum_{t=2}^n e_{t-1} \right)}{\frac{1}{T-1} \cdot \sum_{t=2}^n e_{t-1}^2 - \left( \frac{1}{T-1} \cdot \sum_{t=2}^n e_{t-1} \right)^2} \\ &= \frac{23804 - (-3'4106 \cdot 10^{-13}) \cdot (-20'356)}{44805 - 20'356^2} = 0'5362393 \simeq 0'531275. \end{aligned}$$

Luego la diferencia comentada se debe a que teóricamente se ha supuesto que

$$\sum_{t=2}^n e_{t-1}^2 \simeq \sum_{t=1}^n e_t^2 \text{ y que } \sum_{t=2}^n e_{t-1} \simeq \sum_{t=1}^n e_t = 0.$$



Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

- [1] Salmerón, R. (2012). El modelo lineal general mediante Gretl (online).
- [2] Notas sobre los Mínimos Cuadrados Generalizados (MCG).
- [3] Econometría en YouTube: <http://www.youtube.com/watch?v=xVRRUdk-N8Q>

Puedes encontrarlas en la dirección web:

<http://www.ugr.es/local/romansg/material/WebEco/index.html>

Contenidos

Naturaleza del problema

Causas y consecuencias de la autocorrelación

Procedimientos de Detección

Estimación en los modelos con autocorrelación

Estimación bajo autocorrelación

Ejemplo

- [1] Gujarati, D. (1997). *Econometría*. Ed. McGraw Hill. Capítulo 10.
- [2] Johnston, J. (1989). *Métodos de Econometría*. Ed. Vicens-Vives. Capítulo 8.
- [3] Novales, A. (1993). *Econometría*. McGraw Hill. Capítulo 7.
- [4] Uriel, E., Contreras, D., Moltó, M.L. y Peiró, A. (1990). *Econometría. El Modelo Lineal*. Editorial AC. Capítulo 11.
- [5] Wooldridge, J.M. (2005). *Introducción a la Econometría: Un enfoque moderno*. Thomson. Capítulo 12.