

Prácticas Tema 5. Ampliaciones del Modelo lineal básico

Ana J. López y Rigoberto Pérez

Dpto. Economía Aplicada, Universidad de Oviedo

PRÁCTICA 5.1. Se ha examinado la evolución reciente de las ventas de un dispositivo electrónico (Y , en miles de unidades) y sus gastos en publicidad (X , en millones de euros) obteniéndose los siguientes resultados:

	X_t	Y_t
2000	200	45
2001	250	45
2002	300	52
2003	325	50
2004	350	50
2005	320	50
2006	340	60
2007	400	75
2008	410	72
2019	420	75
2010	430	80
2011	425	82

- a) Estimar un modelo lineal que explique las ventas en función de los gastos en publicidad, estudiando su significación y su capacidad explicativa.
- b) Contrastar la hipótesis de no autocorrelación, comentando los resultados.

SOLUCIÓN:

- a) Estimar un modelo lineal que explique las ventas en función de los gastos en publicidad, estudiando su significación y su capacidad explicativa.

El modelo propuesto es $Y_t = \beta_1 + \beta_2 X_t$ y a partir de la información disponible se obtiene:

$$n = 12 \qquad \sum_t Y_t = 4170 \qquad \sum_t X_t = 733$$

$$\sum_t Y_t^2 = 1508650 \quad \sum_t X_t^2 = 47151 \quad \sum_t X_t Y_t = 265670$$

con lo cual se llega a los siguientes coeficientes de regresión para el modelo simple:

$$\hat{\beta}_2 = \frac{S_{XY}}{S_X^2} = \frac{912,708}{198,076} = 4,6078$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 347,5 - (4,6078)(61,08) = 66,0365$$

Para llevar a cabo el contraste básico de significación del modelo, es decir, de la hipótesis $H_0 : \beta_2 = 0$ se construye el estadístico t de student que en este caso proporciona el resultado $\frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{4,6078}{\sqrt{0,3832}} = 7,44$, al que se ha llegado calculando previamente:

$$S^2 = \frac{\sum_t \hat{u}_t^2}{n-2} = \frac{9107,41}{10} = 910,741$$

$$S_{\hat{\beta}_2}^2 = \frac{S^2}{\sum_t (X_t - \bar{X})^2} = \frac{910,741}{2376,916} = 0,3832$$

A partir de estos resultados se llega a un nivel crítico muy reducido que conduce al rechazo de la hipótesis y valida como significativa la variable X

$$p = P(|t_{10}| > 7,44) \approx 0$$

Por lo que se refiere a la capacidad explicativa, el coeficiente de determinación proporciona un resultado bastante adecuado, ya que se concluye que más del 84% de los cambios en las ventas se explican por el gasto publicitario

$$R^2 = 1 - \frac{\sum_t \hat{u}_t^2}{\sum_t (Y_t - \bar{Y})^2} = 0,8471$$

b) Contrastar la hipótesis de no autocorrelación, comentando los resultados

Para contrastar este supuesto se calcula el estadístico Durbin Watson, que viene dado por la expresión:

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

y puede ser calculado a partir de la información de la tabla:

	Y_t	$\hat{Y}_t$	$\hat{u}_t$	$\hat{u}_t^2$	$(\hat{u}_t - \hat{u}_{t-1})^2$
2000	200	259,57	-59,57	3548,19	—
2001	250	273,39	-23,39	547,1	1308,73
2002	300	305,65	-5,65	31,87	314,88
2003	325	296,43	28,57	816,27	1170,72
2004	350	296,43	53,57	2869,8	625
2005	320	296,43	23,57	555,57	900
2006	340	342,51	-2,51	6,29	680,09
2007	400	411,63	-11,63	135,16	83,14
2008	410	397,80	12,20	148,78	567,56
2009	420	411,63	8,37	70,12	14,62
2010	430	434,67	-4,67	21,77	170,02
2011	425	443,88	-18,88	356,49	202,09

que proporciona el resultado:

$$d_{DW} = \frac{\sum_t (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_t \hat{u}_t^2} = \frac{6036,86}{9107,41} = 0,66$$

Al ser este valor próximo a 0 nos llevaría a concluir que existe evidencia de autocorrelación positiva entre las perturbaciones, hecho que podría estar reflejando problemas de especificación en el modelo.

PRACTICA 5.2. Se dispone de información regional relativa al gasto sanitario (Y , en millones de euros), la renta (R , en millones de euros) y la población de 65 o más años (P_{65} , en %). A partir de una muestra de 51 regiones se ha estimado el modelo mínimo cuadrático:

$$\hat{Y}_i = -3,35 + 0,142 R_i + 0,306 P_i^{65}$$

(1,4) (0,0018) (0,1084)

con $R^2 = 0,992$

- a) Realizar los contrastes individuales de significación.
- b) Se plantea la duda de incorporar como variable explicativa en el modelo la población total (P). Para ello se lleva a cabo el test de inclusión de variables omitidas obteniéndose como discrepancia muestral $F^* = 6,307$. Enunciar la hipótesis nula del contraste determinando cuál será la conclusión.
- c) Una vez modificada la especificación del modelo se decide analizar la presencia de heterocedasticidad mediante el test de White siendo el coeficiente de determinación de la regresión auxiliar 0,778. ¿Cuál es la conclusión del contraste?
- d) Si se asume que la dispersión es proporcional a la población, indicar qué transformación sería aconsejable en el modelo de gasto sanitario.

SOLUCIÓN:

- a) Realizar los contrastes individuales de significación.

Las hipótesis recogen la nulidad de los parámetros β_2 y β_3 , respectivamente y las discrepancias en las que se basa el contraste siguen distribución t de Student con 48 grados de libertad.

En el caso del contraste de la hipótesis $H_0 : \beta_2 = 0$ la discrepancia viene dada por $\frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{0,142}{0,0018} = 78,888$, valor que lleva asociado un nivel crítico $p = P(|t_{48}| > 78,88) \approx 0$ y conduce al rechazo de la hipótesis. Así pues, la variable renta regional es relevante para explicar el correspondiente gasto sanitario.

De modo análogo se contrasta el supuesto $H_0 : \beta_3 = 0$ siendo en este caso la discrepancia $\frac{\hat{\beta}_3}{S_{\hat{\beta}_3}} = \frac{0,306}{0,1884} = 2,822$, con nivel crítico $p = P(|t_{48}| > 2,822) = 0,006$ que es suficientemente reducido para rechazar la hipótesis nula y permite afirmar que la variable población mayor de 65 años también es relevante para explicar el correspondiente gasto sanitario.

- b) Se plantea la duda de incorporar como variable explicativa en el modelo la población total (P). Para ello se lleva a cabo el test de inclusión de variables omitidas obteniéndose como discrepancia muestral $F^* = 6,307$. Enunciar la hipótesis nula del contraste determinando cuál será la conclusión.

La hipótesis nula de este contraste es que la variable omitida no resulta relevante, por lo tanto se contrastaría la nulidad de su coeficiente β_4 .

Así pues se contrastaría $H_0 : \beta_4 = 0$ frente a $H_1 : \beta_4 \neq 0$ y la resolución de este test es similar a la de un contraste de restricciones de parámetros. Más concretamente, en este caso se proporciona la discrepancia muestral que ha sido obtenida mediante el estadístico $\left(\frac{\hat{\mathbf{u}}_R' \hat{\mathbf{u}}_R - \hat{\mathbf{u}}' \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} \right) \left(\frac{n-k}{r} \right) \approx F_{n-k}^r$ siendo r el número de restricciones (1 en este caso) y $n-k$ los grados de libertad (47 en este caso).

El nivel crítico asociado a la discrepancia muestral es: $p = P(F_{47}^1 > 6,307) = 0,015$ que es suficientemente bajo para rechazar la hipótesis nula. Por lo tanto, la variable población total es relevante y debe ser incluida en la especificación del modelo.

c) Una vez modificada la especificación del modelo se decide analizar la presencia de heterocedasticidad mediante el test de White siendo el coeficiente de determinación de la regresión auxiliar 0,778. ¿Cuál es la conclusión del contraste?

La hipótesis nula del test de White establece que las perturbaciones son homocedásticas. Este contraste se basa en una regresión auxiliar de los residuos cuadráticos en función de las variables explicativas, sus cuadrados y sus productos cruzados, de modo que cuanto mejor sea la bondad de este ajuste, mayor será la evidencia muestral para el rechazo de la hipótesis nula.

El estadístico del contraste se calcula como $nR^2 \approx \chi_m^2$, siendo m el número de regresores. En este caso el producto de $n=51$ y $R^2 = 0,778$ proporciona el resultado 39,678 que lleva asociado que lleva asociado un nivel crítico: $p = P(\chi_9^2 > 39,678) \approx 0$ que conduce al rechazo de la hipótesis nula de homocedasticidad.

d) Si se asume que la dispersión es proporcional a la población, indicar qué transformación sería aconsejable en el modelo de gasto sanitario

Se asume en este caso $E(u_i^2) = \sigma_i^2 = \sigma^2 P_i$ y se trata de buscar ponderaciones que permitan transformar estas perturbaciones en homocedásticas, es decir, $E(W_i u_i)^2 = \sigma^2$. Bajo este requisito se obtiene la ponderación $W_i^2 E(u_i^2) = \sigma^2 \Rightarrow W_i^2 = \frac{\sigma^2}{E(u_i^2)} = \frac{\sigma^2}{\sigma^2 P_i} \Rightarrow W_i = \sqrt{\frac{1}{P_i}}$ que debería ser utilizada para estimar el modelo mediante mínimos cuadrados ponderados.

PRACTICA 5.3. A partir de la información proporcionada por 30 establecimientos de una empresa multinacional se ha estimado un modelo para explicar los salarios base (Y , en miles de dolares corregidos a Paridades de Poder Adquisitivo, PPA) en función de los índices de precios (X , en %).

$$\hat{Y}_i = 1,858 + 0,8 X_i$$

(0,044)
(0,01)

Sabiendo que el test de White ha proporcionado el resultado $nR^2 = 10,8$

a) Indicar cuál ha sido el método seguido para su obtención, estudiando la presencia de heteroscedasticidad.

b) Si se sospecha que la varianza es inversamente proporcional a los precios, indicar cómo debería ser estimado el modelo.

SOLUCIÓN:

a) Indicar cuál ha sido el método seguido para su obtención, estudiando la presencia de heteroscedasticidad.

En el método propuesto por White, basado en la regresión de los residuos cuadráticos $\hat{u}_i^2$ se ha obtenido una discrepancia que adopta el valor 10,8. Teniendo en cuenta que esta discrepancia se obtiene mediante la expresión $nR^2 \simeq \chi_2^2$ (cuyos grados de libertad vienen dados por el número de variables explicativas o regresores del modelo auxiliar) se observa que el nivel crítico asociado es reducido, $p = P(\chi_2^2 > 10,8) = 0,0045$ y por tanto conduce al rechazo de la hipótesis nula planteada, aconsejando alguna transformación en el modelo para garantizar la homocedasticidad.

b) Si se sospecha que la varianza es inversamente proporcional a los precios, indicar cómo debería ser estimado el modelo.

Para decidir qué tipo de ponderación conviene utilizar, asumimos una estructura del tipo $\sigma_i^2 = \frac{\sigma^2}{X_i}$ y buscamos los pesos W_i que garanticen: $E(W_i u_i)^2 = \sigma^2$.

Para que esto ocurra:

$$E(W_i u_i)^2 = \sigma^2 \Rightarrow W_i^2 E(u_i^2) = \sigma^2 \Rightarrow W_i^2 = \frac{\sigma^2}{E(u_i^2)} = \frac{\sigma^2}{\frac{\sigma^2}{X_i}} = X_i \Rightarrow W_i = \sqrt{X_i}$$

Es decir, las ponderaciones adecuadas serían las raíces cuadradas de los índices de precios.

PRACTICA 5.4. Se han estimado modelos para explicar la tasa de empleo (E) en función del crecimiento del PIB (C), obteniéndose los resultados resumidos en la tabla:

Ambito muestral	Modelo estimado	$\hat{\mathbf{u}}'\hat{\mathbf{u}}$
Unión Europea Ampliada UE-27	$\hat{E}_i = 61,8 + 0,37C_i$	857,4
Unión Europea UE-15	$\hat{E}_i = 62,8 + 0,45C_i$	611,9
Países de las últimas ampliaciones (12)	$\hat{E}_i = 55,9 + 1,07C_i$	156,9

Justificar si, con esta información, se podría asumir la existencia de una estructura única en el comportamiento del empleo.

SOLUCIÓN:

Para contrastar la hipótesis de constancia de los parámetros en el modelo lineal propuesto se hace uso del test de Chow. La hipótesis nula del contraste puede ser expresada como

$$H_0 : \beta_{[n1]} = \beta_{[n2]} = \beta \quad \sigma_{[n1]}^2 = \sigma_{[n2]}^2 = \sigma^2$$

La discrepancia viene dada por la expresión:

$$\frac{\hat{u}'\hat{u} - (\hat{u}'_1\hat{u}_1 + \hat{u}'_2\hat{u}_2)}{\hat{u}'_1\hat{u}_1 + \hat{u}'_2\hat{u}_2} \left(\frac{n_1 + n_2 - 2k}{k} \right) \simeq F_{n_1+n_2-2k}^k$$

donde se compara la suma de los residuos cuadráticos del modelo global con las sumas de residuos cuadráticos asociados a los modelos estimados sobre las submuestras.

Sustituyendo la información muestral disponible se llega al resultado:

$$\frac{\hat{u}'\hat{u} - (\hat{u}'_1\hat{u}_1 + \hat{u}'_2\hat{u}_2)}{\hat{u}'_1\hat{u}_1 + \hat{u}'_2\hat{u}_2} \left(\frac{n_1 + n_2 - 2k}{k} \right) = \frac{857,4 - (611,9 + 156,9)}{611,9 + 156,9} \left(\frac{27 - 4}{2} \right) = 1,325$$

que lleva asociado el nivel crítico

$$p = P(F_{23}^2 > 1,325) = 0,28$$

por lo que no existe evidencia muestral para el rechazo de la hipótesis nula de constancia de parámetros (o equivalentemente estabilidad estructural)
