Taller #5
Econometría 06169

Profesor: Julio César Alonso
Nota: Este taller debe ser entregado en papel y escrito en computador. No se revisarán trabajos escritos a mano.

1. Un empresario textil planea emplear un modelos con el fin de estimar la productividad de sus plantas (P_i), en función de la antigüedad de la maquinaria que alquila (A_i), la experiencia de los trabajadores contratados (E_i) y la calidad del combustible utilizado (D_{ji}), siendo esta última una variable ficticia que toma los valores: $D_{1i} = 1$ si el combustible es de calidad baja, cero en caso contrario; $D_{2i} = 1$ si el combustible es de calidad media, cero en caso contrario; y $D_{3i} = 1$ si el combustible es de calidad alta, cero en caso contrario.
   Para tal fin, el empresario lo ha contratado a usted para que realice las estimaciones, para ello el le propone los siguientes modelos:

$$P_i = b_0 + b_1 A_i + b_2 E_i + b_3 D_{1i} + b_4 D_{2i} + b_5 D_{3i} + e_i \quad (1)$$

$$P_i = b_1 A_i + b_2 E_i + b_3 D_{1i} + b_4 D_{2i} + b_5 D_{3i} + e_i \quad (2)$$

$$P_i = b_0 + b_1 A_i + b_2 E_i + b_3 D_{1i} + e_i \quad (3)$$

   a) ¿Cuál modelo sería más apropiado, si la estimación se lleva a cabo por MCO?
   b) Interprete los coeficientes de las variables en el modelo que usted escogió.
   c) ¿Se podría plantear un modelo alternativo? Si es posible interprete los coeficientes del modelo alternativo, y explique que ventajas o desventajas tendría estimar este nuevo modelo frente al escogido inicialmente por usted.

2. Una empresa de autobuses intermunicipales desea estimar la demanda de sus tiquete (D_t), en función del precio de los mismos (p_t) y de la calidad del servicio, evaluada a través de los gastos que la empresa realiza para la mejora del mismo (M_t). Para ello dispone de datos de los últimos 50 trimestres que se encuentran en el archivo T5-2.xls en la página Web..
   a) Plantee el modelo (con intercepto), estímelo[1] y discuta cualquier posible problema en la estimación.
   b) Estudios previos en la industria permiten asegurar con mucha confianza que el parámetro correspondiente a p_t posee un valor poblacional de -0.5. A partir de esta información y del modelo[2] de la parte a), estime los demás coeficientes.

3. Con la información que esta disponible en el archivo T5-3.xls, responda las siguientes preguntas:

   a) Estime el modelo de regresión $y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + e_i$, reporte sus resultados en una tabla y analice cuidadosamente sus resultados.
   b) ¿Considera conveniente suprimir la variable X_3 del modelo de regresión? ¿Y por qué no X_2? Justifique su respuesta.

---

[1] Reporte los resultados en una tabla.
[2] ¡¡No de la ecuación estimada!!

Taller #5

Respuestas Sugeridas
Econometría 06169

Profesor: Julio César Alonso
Nota: Este taller debe ser entregado en papel y escrito en computador. No se revisarán trabajos escritos a mano.

1. Un empresario textil planea emplear un modelo con el fin de estimar la productividad de sus plantas (P_i), en función de la antigüedad de la maquinaria de alquila (A_i), la experiencia de los trabajadores contratados (E_i) y la calidad del combustible utilizado (D_{ji}), siendo esta última una variable ficticia que toma los valores: $D_{1i} = 1$ si el combustible es de calidad baja, cero en caso contrario; $D_{2i} = 1$ si el combustible es de calidad media, cero en caso contrario; y $D_{3i} = 1$ si el combustible es de calidad alta, cero en caso contrario.
   Para tal fin, el empresario lo ha contratado a usted para que realice las estimaciones, para ello el le propone los siguientes modelos:

$$P_i = \beta_0 + \beta_1 A_i + \beta_2 E_i + \beta_3 D_{1i} + \beta_4 D_{2i} + \beta_5 D_{3i} + \varepsilon_i \quad (1)$$

$$P_i = \beta_1 A_i + \beta_2 E_i + \beta_3 D_{1i} + \beta_4 D_{2i} + \beta_5 D_{3i} + \varepsilon_i \quad (2)$$

$$P_i = \beta_0 + \beta_1 A_i + \beta_2 E_i + \beta_3 D_{1i} + \varepsilon_i \quad (3)$$

   a) ¿Cuál modelo sería más apropiado, si la estimación se lleva a cabo por MCO?

Claramente el modelo (1) presenta multicolinealidad perfecta, pues $1 = D_{1i} + D_{2i} + D_{3i}$; el tercer modelo no capturaría el efecto sobre la productividad del combustible de calidad alta o media, pues no incluye al menos una de estas dos variables dummy (Esto puede ser fácilmente mostrado al calcular el valor esperado de este modelo). Finalmente, noten que el modelo (2) no presenta multicolinealidad y además captura el efecto de las tres calidades de combustible sobre la productividad. Por tanto el modelo escogido debe ser el modelo (2).

   b) Interprete los coeficientes de las variables en el modelo que usted escogió.

Antes de interpretar los significados, es útil encontrar el valor esperado de la productividad bajo las diferentes posibles calidades de combustible. Se puede demostrar fácilmente que

$$E[P_i] = \begin{cases} \beta_1 A_i + \beta_2 E_i + \beta_3 & \text{si combustible de calidad baja} \\ \beta_1 A_i + \beta_2 E_i + \beta_4 & \text{si combustible de calidad media} \\ \beta_1 A_i + \beta_2 E_i + \beta_5 & o.w. \end{cases}$$

Así, los parámetros del modelo pueden ser interpretados de la siguiente forma:

- β_1 representa el cambio en la productividad provocado por un aumento en una unidad en la antigüedad de la maquinaria alquilada;
- β_2 representa el cambio en la productividad provocado por un aumento en una unidad en la experiencia de los trabajadores contratados;
- β_3, β_4, y β_5 representan la parte de la productividad que no es explicada ni por la antigüedad de las máquinas alquiladas, ni por la experiencia de los trabajadores contratados cuando se emplea combustible de calidad baja, media y alta, respectivamente.

c) ¿Se podría plantear un modelo alternativo? Si es posible interprete los coeficientes del modelo alternativo, y explique que ventajas o desventajas tendría estimar este nuevo modelo frente al escogido inicialmente por usted

Un modelo alternativo sería:

$$P_i = \beta_0 + \beta_1 A_i + \beta_2 E_i + \beta_3 D_{1i} + \beta_4 D_{2i} + \varepsilon_i \qquad (4)$$

Se puede demostrar fácilmente que

$$E[P_i] = \begin{cases} (\beta_0 + \beta_3) + \beta_1 A_i + \beta_2 E_i & \text{si combustible de calidad baja} \\ (\beta_0 + \beta_4) + \beta_1 A_i + \beta_2 E_i & \text{si combustible de calidad media} \\ \beta_0 + \beta_1 A_i + \beta_2 E_i & o.w. \end{cases}$$

en este caso la interpretación de los coeficientes es:

- β_1 representa el cambio en la productividad provocado por un aumento en una unidad en la antigüedad de la maquinaria alquilada;
- β_2 representa el cambio en la productividad provocado por un aumento en una unidad en la experiencia de los trabajadores contratados;
- β_0 representan la parte de la productividad que no es explicada ni por la antigüedad de las máquinas alquiladas ni por la experiencia de los trabajadores contratados cuando se emplea combustible de calidad alta;
- β_3 representan la diferencia entre usar combustible de calidad alta y calidad baja en la parte de la productividad que no es explicada ni por la antigüedad de las máquinas alquiladas ni por la experiencia de los trabajadores contratados;
- y β_4 representan la diferencia entre usar combustible de calidad alta y media baja en la parte de la productividad que no es explicada ni por la antigüedad de las máquinas alquiladas ni por la experiencia de los trabajadores contratados.

Es importante darse cuenta que estimar este ultimo modelo tiene una ligera ventaja sobre el modelo inicial, la ventaja es que con este último modelo se puede interpretar el R^2 mientras que en el modelo (2) no se puede, por carecer de un intercepto.

2. Una empresa de autobuses intermunicipales desea estimar la demanda de sus tiquetes (D_t), en función del precio de los mismos (p_t) y de la calidad del servicio, evaluada a través de los gastos que la empresa realiza para la mejora del mismo (M_t). Para ello dispone de datos de los últimos 50 trimestres que se encuentran en el archivo T5-2.xls en la página Web..
   a) Plantee el modelo (con intercepto), estímelo[1] y discuta cualquier posible problema en la estimación.

El modelo a estimar es:

$$D_t = \gamma_0 + \gamma_1 p_t + \gamma_2 M_t + \varepsilon_t . \qquad \text{(Modelo 1)}$$

Empleando los datos suministrados y EasyReg, se obtiene los resultados reportados en la Tabla 1. Noten que el F global (102.98) y el R^2 (.81) son relativamente altos. Esto debe brindar sospechas de la presencia de multicolinealidad. Así es necesario explorar si existe en efecto o no problemas de multicolinealidad en este modelo estimado.

Para constatar la existencia de multicolinealidad podemos encontrar el determinante de la matriz de correlación entre las variables explicativas ($|R|$) en este caso:

[1] Reporte los resultados en una tabla.

$$|R| = \left| \begin{bmatrix} 1 & 0.8315 \\ 0.8315 & 1 \end{bmatrix} \right| = 0.3086$$

Noten que este determinante es relativamente cercano a cero, mostrando indicios de multicolinealidad, al igual es importante notar que existe una fuerte correlación entre las dos variables dependientes (0.83).

Además podemos constatar este resultado observando la matriz de correlaciones entre los parámetros estimados. En este caso tenemos que esta matriz de correlación es:

$$\begin{bmatrix} 1 & -.8315 & .05411 \\ & 1 & -.5911 \\ & & 1 \end{bmatrix}$$

donde la primera columna y fila corresponde al estimador de γ_2, y la segunda fila y columna corresponde al estimador de γ_1. Claramente existe una fuerte correlación entre los estimadores de γ_1 y γ_2. Así, existe suficiente evidencia para concluir que existe multicolinelalidad en el modelo.

Tabla 1

	VARIABLE DEPENDIENTE: D_t Estadísticos t entre paréntesis	
	Ecuación 1	Ecuación 2
	Datos trimestrales MCO	Datos trimestrales MCR
constante	1,319.9413	3,195.9042
	(3.64) ***	(6.81) ***
M_t	0.0531	0.2365
	(0.17)	(8.74) ***
P_t	3.1242	-0.5000
	(7.45) ***	
R^2	0.81420	0.61392
F	102.980 ***	76.330 ***
# de Obs.	50	50

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
MCO: Mínimos Cuadrados Ordinarios
MCR: Mínimos Cuadrados Restringidos (β_2=0)

b) Estudios previos en la industria permiten asegurar con mucha confianza que el parámetro correspondiente a p_t posee un valor poblacional de -0.5. A partir de esta información y del modelo[2] de la parte a), estime los demás coeficientes.

Conociendo esta información, el modelo se puede rescribir como $D_t + 0.5p_t = \gamma_0 + \gamma_2 M_t + \varepsilon_t$, es decir el nuevo modelo será $w_t = \gamma_0 + \gamma_2 M_t + \varepsilon_t$, donde

[2] ¡¡No de la ecuación estimada!!

$w_t = D_t + 0.5 p_t$. Este nuevo modelo estimado por mínimos cuadrados ordinarios creando una nueva variable. Noten que esto es equivalente a emplear los mínimos cuadrados restringidos donde $R = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ y $C = \begin{bmatrix} -0.5 \end{bmatrix}$. Así el método empleado es el Mínimos Cuadrados Restringidos. Los resultados de esta estimación se reportan en la Tabla 1.

3. Con la información que esta disponible en el archivo T5-3.xls, responda las siguientes preguntas:

a) Estime el modelo de regresión $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$, reporte sus resultados en una tabla y analice cuidadosamente sus resultados.

Los resultados de l estimación de este modelo se reportan en la Tabla 2. Noten que X_3 no es significativa para el modelo (individualmente). Noten que el F global (11,620.21) y el R^2 (.999) son muy altos. Esto debe brindar sospechas de la presencia de multicolinealidad. Así es necesario explorar si existe en efecto o no problemas de multicolinealidad en este modelo estimado.

Tabla 2.

VARIABLE DEPENDIENTE: y_t Estadísticos t entre paréntesis	
	MCO
constante	1,957.7781
	(2.50) **
$X2_t$	0.3813
	(100.77) ***
$X3_t$	0.0016
	(0.52)
R^2	0.99936
F	11,620.210 ***
# de Obs.	18

(*) nivel de significancia: 10%
(**) nivel de significancia: 5%
(***) nivel de significancia: 1%
MCO: Mínimos Cuadrados Ordinarios

Para constatar la existencia de multicolinealidad podemos encontrar el determinante de la matriz de correlación entre las variables explicativas ($|R|$) en este caso:

$$|R| = \left| \begin{bmatrix} 1 & 0.7481127 \\ 0.7481127 & 1 \end{bmatrix} \right| = 0.44$$

Noten que este determinante es relativamente cercano a cero, mostrando indicios de multicolinealidad.

Además podemos constatar este resultado observando la matriz de correlaciones entre los parámetros estimados. En este caso tenemos que esta matriz de correlación es:

$$\begin{bmatrix} 1 & -.74811 & .60449 \\ & 1 & -.04271 \\ & & 1 \end{bmatrix}$$

donde la primera columna y fila corresponde al estimador de β_2, y la segunda fila y columna corresponde al estimador de β_3. Este test no muestra una correlación muy marcada entre los estimadores de β_2 y β_3.

Así, siendo un poco conservador, uno podría concluir que existe cierto grado de multicolinealidad, que si bien no es muy fuerte, si existe.

b) ¿Considera conveniente suprimir la variable X_3 del modelo de regresión? ¿Y por qué no X_2? Justifique su respuesta.

Noten que realmente no existe ninguna razón por la cual uno pueda descartar la variable X_2 o X_3, como lo discutimos en clase, la teoría económica debería ser nuestra guía para incluir o excluir variables de un modelo. En este caso no sabemos que representa cada una de las variables involucradas, por tanto no existe forma para tomar una decisión basada en la teoría para descartar una u otra variable. Por tanto deberíamos ser indiferentes en cual variable se descarta, bajo la salvedad que esto tendrá implicaciones sobre las bondades del estimador que veremos en próximas semanas.