

Taller 8: Otros problemas econométricos
Econometría 06216
14-03-2011

Profesor: Carlos Giovanni González Espitia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 21 de marzo de 2011. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**(no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos**.

El alcalde de la ciudad Kalima, el Dr. Espina, está interesado en analizar el comportamiento de la producción industrial en la región. Para ello, ha encargado al Grupo de Investigaciones Económicas de la alcaldía del cual usted hace parte para la realización de este análisis. La Dra. Sandra, quien encabeza la investigación, cree que la producción industrial en la región (Y_t) dada en millones de pesos está explicada por el número de trabajadores del sector industrial (L_t); el consumo de bienes industriales (C_t) en millones de pesos y la inversión privada en el sector (I_t) en millones de pesos. Para ello, cuenta con datos para los últimos 17 años que se encuentran adjuntos en el archivo T8-11-01.xls.

Pregunta 1

- a. Escriba correctamente el modelo econométrico a estimar .
- b. Interprete a priori los coeficientes del modelo justificando el signo esperado.
- c. Estime el modelo (1) y reporte sus resultados en una tabla.

Pregunta 2

La Dra. Sandra lo designa a usted para hacer la presentación de los resultados al alcalde. Sin embargo, usted que se destacó en su curso de econometría sospecha que puede existir un

Econometría *Sem01-2011*

problema econométrico en el modelo. Realice las pruebas para identificar formalmente los problemas econométricos (multi, hetero, auto y otros) que puedan existir.

Pregunta 3

- a. En caso de encontrar algún problema econométrico, corríjalo (si es posible) por el método más adecuado.
- b. ¿Qué puede concluir?

Pregunta 4

La compañía Manzana mordida está interesada en conocer el comportamiento de la demanda en los Estados Unidos de su último procesador de computadores el cual fue lanzado al mercado en 2010: el procesador ipad. Para ello, el CEO de la corporación conformó un grupo investigativo con los mejores econométricos de la ciudad con el fin de obtener los resultados más aproximados y confiables posibles. El director de la investigación, basado en sus conocimientos en el campo de la informática, sugiere que la cantidad demandada del nuevo procesador está en función del precio del procesador de la competencia que posee similares características al nuevo procesador de la empresa; del precio en el mercado de los laptop que están utilizando el nuevo procesador, del precio del mercado de memorias USB de alta capacidad de almacenamiento, y de la cantidad de estudiantes universitarios cuya carrera está relacionada con sistemas. Usted posee datos de corte transversal de una muestra recogida durante la investigación que está disponible en el archivo adjunto T8-11-01-2.xls. Con base en ello responda las siguientes preguntas.

- a. Escriba correctamente el modelo a estimar que sugiere el director de la investigación.
- b. Interprete a priori los coeficientes del modelo anterior.
- c. Estime el modelo (2) y repórtelo en una tabla

Pregunta 5

- a. Suponiendo que no existe ninguno de los tres problemas econométricos más comunes (multi, hetero y auto), ¿Cree usted que existe algún otro problema econométrico? Argumente desde su punto de vista por qué cree que existiría o no otro problema econométrico.

Pregunta 6

- a. Solucione (de ser posible) el problema econométrico encontrado.
- b. Ahora bien, el alcalde afirma: “en las encuestas realizadas algunas tiendas no quisieron revelar el valor real de la cantidad demandada de procesadores ipad”. Que implicaciones tendría para la investigación esta afirmación de ser cierta.

Taller 8: Otros problemas econométricos
Econometría 06216
14-03-2011

Profesor: Carlos Giovanni González Espítia

Monitores: Adriana Caicedo – Jessica Echeverry – Samir Aristizábal

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 21 de marzo de 2011. **Sólo se calificarán talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**(no se recibirán talleres después de esa hora y fecha límite).

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo en pareja. Por tanto el taller debe reflejar **únicamente** el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos **con todos ellos**.

El alcalde de la ciudad Kalima, el Dr. Espina, está interesado en analizar el comportamiento de la producción industrial en la región. Para ello, ha encargado al Grupo de Investigaciones Económicas de la alcaldía del cual usted hace parte para la realización de este análisis. La Dra. Sandra, quien encabeza la investigación, cree que la producción industrial en la región (Y_t) dada en millones de pesos está explicada por el número de trabajadores del sector industrial (L_t); el consumo de bienes industriales (C_t) en millones de pesos y la inversión privada en el sector (I_t) en millones de pesos. Para ello, cuenta con datos para los últimos 17 años que se encuentran adjuntos en el archivo T8-11-01.xls.

Pregunta 1

- a. Escriba correctamente el modelo econométrico a estimar .

$$Y_t = \beta_1 + \beta_2 L_t + \beta_3 C_t + \beta_4 I_t + \varepsilon_t(1)$$

$t = 1, 2, \dots, 17.$

- b. Interprete a priori los coeficientes del modelo justificando el signo esperado.

β_1 = Representa la producción industrial que no depende del número de trabajadores del sector industrial, del consumo de bienes industriales ni de la inversión privada en el sector. Se espera signo positivo.

β_2 = Una variación de un trabajador en el sector industrial, genera en promedio una variación de β_2 millones de pesos en la producción industrial. Se espera signo positivo.

β_3 = Una variación de un millón de pesos en la inversión privada, genera, en promedio una variación de β_3 millones de pesos en la producción industrial. Se espera signo positivo.

- c. Estime el modelo (1) y reporte sus resultados en una tabla.

Tabla 1: Estimación del comportamiento de la producción industrial

Variable dependiente: (Y_t)		
MCO		
Variable	Parámetro	Significancia
Constante (β_1)	-10067.20 (-0.505)	
L_t (β_2)	-0.8870643 (-0.471)	
C_t (β_3)	12.6499358 (2.718)	***
I_t (β_4)	-0.1120910 (-0.069)	
R ²	0.6485	
R ² -ajustado	0.5674	
n	17	
F	8	***

(*) Nivel de significancia: 10%

(**) Nivel de significancia: 5%

(***) Nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

Fuente: Datos obtenidos en Easyreg

Pregunta 2

La Dra. Sandra lo designa a usted para hacer la presentación de los resultados al alcalde. Sin embargo, usted que se destacó en su curso de econometría sospecha que puede existir un problema econométrico en el modelo. Realice las pruebas para identificar formalmente los problemas econométricos (multi, hetero, auto y otros) que puedan existir.

Dado que el número de observaciones no es lo suficientemente grande ($n < 20$) y el Teorema del Límite Central sólo aplica para muestras grandes, es necesario **probar si los errores se distribuyen normalmente**.

Para ello, emplearemos el test de **Jarque-Bera**. Este test tiene la siguiente hipótesis nula y alterna:

Ho: Los errores se distribuyen normalmente

Ha: No Ho

Donde el criterio de rechazar Ho es que $JB > \chi_2^2$

De este modo, a partir de la salida de EasyReg obtuvimos que:

JB = 22.106367

Con un p-valor = 0.00002

Así, con un nivel de confianza del 99% podemos rechazar la Hipótesis nula ($p\text{-valor} < \alpha$). Por lo tanto, la única opción sería aumentar la muestra, pero como en este caso no es posible, los datos o son los adecuados para aplicar todos los tests vistos en clase, exceptuando aquellos en cuya aplicación no se requiere asumir un comportamiento normal de los residuos.

A continuación se presentan las pruebas que pueden realizarse sin tener que asumir un comportamiento sobre la distribución de los errores:

- **Multicolinealidad**

En primer lugar, tenemos que no se perciben síntomas de multicolinealidad No perfecta en el modelo, pues a pesar de que los valores t son bajos, el F y R^2 no son altos.

Sin embargo, para determinar con más exactitud, si existe algún problema de multicolinealidad, realizaremos las pruebas estudiadas en clase, es decir, la matriz de correlación de las variables independientes, la prueba Kappa y la matriz de correlaciones de los coeficientes estimados.

1. Matriz de correlación de las X's

El determinante de la matriz de correlación $|R|$, empleando los valores propios, es:

$$|R| = 2,98748 \times 0,01207 \times 0,000442 = 0,00001597$$

Dado que este resultado tiende a cero, puede darse un problema de multicolinealidad serio.

2. Prueba Kappa

$$k(X) = \frac{\sqrt{\lambda_{Max}}}{\sqrt{\lambda_{Min}}} = \sqrt{\frac{2,98748493}{0,0004429429}} = 82,1256$$

Como se puede percibir, el valor de $\kappa(X)$ es mayor a 30 lo cual permite intuir que hay un problema de multicolinealidad serio.

3. Matriz de correlaciones entre los β 's estimados

	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_1$
$\widehat{\beta}_2$	1	-0,9750	1,7319	-0,2870
$\widehat{\beta}_3$		1	-2,6739	0,3770
$\widehat{\beta}_4$			1	-2,7033
$\widehat{\beta}_1$				1

Podemos percibir que el valor absoluto de la correlación entre los coeficientes del consumo de bienes industriales (C_t) y del número de trabajadores del sector industrial (L_t) es mayor a 0,8. Al igual que la correlación entre los coeficientes de la inversión privada en el sector (I_t) y del número de trabajadores del sector industrial (L_t); la correlación entre los coeficientes de la inversión privada en el sector (I_t) y del consumo de bienes industriales (C_t); y la correlación entre el intercepto y la inversión privada en el sector (I_t). Por lo tanto podemos inferir que si hay presencia de un problema de multicolinealidad grave. Sin embargo, esta prueba tiene validez si hay homoscedasticidad y no se presenta autocorrelación en la regresión. Por ello, no se puede concluir con completa confianza sobre esta prueba sin realizar otros análisis.

Conclusión: todas las anteriores pruebas brindan evidencia de que existe un problema de multicolinealidad en el modelo.

- **Autocorrelación**

Para este problema, contamos con una serie de pruebas, sin embargo, no todas pueden ser aplicadas pues algunas suponen una distribución normal para los errores y como ya probamos los errores no cumplen este supuesto. De este modo, sólo podremos aplicar **la prueba de Rachas** que es una prueba no paramétrica.

Prueba de Rachas

En esta prueba la hipótesis nula es la existencia de no autocorrelación versus la hipótesis alterna de autocorrelación:

Ho: $\rho = 0$

Ha: $\rho \neq 0$

Primero, contamos el número de errores con signo positivo (N_+) y con signo (N_-) y tenemos que

$N_+ = 8$

$N_- = 9$

Luego, tenemos que el número de rachas (k) es:

$k = 11$

Como N_+ y N_- son menores que 20, entonces se puede construir un intervalo de confianza del 95% para el número de rachas "razonable" bajo la hipótesis nula a partir de los valores críticos. La hipótesis nula se puede rechazar si el número de rachas observadas no están contenidas en el intervalo de confianza

En este caso el límite inferior de la prueba de Rachas a un nivel de confianza del 95% es 5 y el límite superior es 14. Dado que el número de Rachas que habíamos obtenido es igual a 11 y este cae dentro del intervalo [5, 14], no hay suficiente evidencia para rechazar la hipótesis nula de no autocorrelación con un nivel de confianza del 95%. No obstante, el intervalo razonable para el número de rachas sobre el cual podría concluirse no auto-correlación es de 9 a 22 rachas. En este caso el número de rachas es 11, por ello, note que la decisión se toma casi al margen del intervalo, por lo que sería necesario realizar otras pruebas con el fin de detectar auto-correlación y dado que el supuesto de normalidad no se cumple esto no es posible.

- **Heteroscedasticidad**

Para probar la existencia de este problema contamos con tres pruebas: la prueba de Goldfeld y Quandt, la prueba de Breusch – Pagan y la prueba de White. Sin embargo, las tres no pueden ser efectuadas, ya que las dos primeras necesitan el supuesto de la distribución normal de los errores para aplicarlas al modelo y como vimos en un comienzo, a través del test de Jarque-Bera, los errores no se distribuyen normalmente. Por ello, aplicaremos sólo la prueba de White para determinar la existencia o no de heteroscedasticidad.

Prueba de White

La hipótesis a contrastar es:

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_a: \text{No } H_0$$

Para realizar esta prueba, primero corremos el modelo original para encontrar la serie de residuos $\hat{\varepsilon}_i$. Luego, corremos la siguiente regresión auxiliar:

$$\varepsilon_i^2 = \beta_1 + \beta_2 L_t + \beta_3 C_t + \beta_4 I_t + \beta_5 L_t^2 + \beta_6 C_t^2 + \beta_7 I_t^2 + \beta_8 L_t * C_t + \beta_9 L_t * I_t + \beta_{10} C_t * I_t + \varepsilon_i \quad i = 1, 2, \dots, 17$$

El Estadístico que se debe utilizar es:

$$W_\alpha = nR^2$$

Dado que se obtuvo $R^2 = 0,9954$, el estadístico de White en este caso es:

$$W_\alpha = 17 * 0,9954 = 16,92$$

El criterio de rechazo de H_0 es: $W_\alpha = \chi_{g(\alpha)}^2$

En este caso g es igual a 9 (# de parámetros de la regresión auxiliar menos 1), por ello, con un nivel de significancia del 1% el valor crítico es 21,66.

Así, dado que $W_\alpha = 16,92 < 21,66$, con un nivel de confianza del 99% no puede rechazar la hipótesis nula de que existe homoscedasticidad (lo mismo ocurre al 95% y 90% de confianza)

Pregunta 3

- a. En caso de encontrar algún problema econométrico, corríjalo (si es posible) por el método más adecuado.

El problema de normalidad no puede ser solucionado dado que no podemos aumentar la muestra, lo que condiciona la no solución de los otros problemas econométricos. Note que es posible solucionar el problema de multicolinealidad eliminando una de las variables explicativas, pero esta acción no serviría de nada si no es posible garantizar la existencia de un error distribuido normalmente.

- b. ¿Qué puede concluir?

Si no se puede garantizar un error distribuido normalmente, no puede garantizarse estimadores MELI a través del método MCO, por lo tanto, no se puede concluir nada a partir del modelo estimado y no tiene sentido hablar de inferencia a partir de estos resultados

Pregunta 4

La compañía Manzana mordida está interesada en conocer el comportamiento de la demanda en los Estados Unidos de su último procesador de computadores el cual fue lanzado al mercado en 2010: el procesador ipad. Para ello, el CEO de la corporación conformó un grupo investigativo con los mejores econométricos de la ciudad con el fin de obtener los resultados más aproximados y confiables posibles. El director de la investigación, basado en sus conocimientos en el campo de la informática, sugiere que la cantidad demandada del nuevo procesador está en función del precio del procesador de la competencia que posee similares características al nuevo procesador de la empresa; del precio en el mercado de los laptop que están utilizando el nuevo procesador, del precio del mercado de memorias USB de alta capacidad de almacenamiento, y de la cantidad de estudiantes universitarios cuya carrera está relacionada con sistemas. Usted posee datos de corte transversal de una muestra recogida durante la investigación que está disponible en el archivo adjunto T8-11-01-2.xls. Con base en ello responda las siguientes preguntas.

- a. Escriba correctamente el modelo a estimar que sugiere el director de la investigación.

El modelo a estimar es

$$D_i = \delta_1 + \delta_2 P_{si} + \delta_3 P_{ci} + \delta_4 U_i + \delta_5 E_i + \varepsilon_i \quad i = 1, 2, \dots, 15$$

- b. Interprete a priori los coeficientes del modelo anterior.

δ_1 = Representa la demanda del procesador de Inter que no depende del precio del procesador de la competencia que posee similares características al nuevo procesador de la empresa; del precio de mercado de los laptop que están utilizando el nuevo procesador, del precio de mercado de memorias USB de alta capacidad de almacenamiento ni de la cantidad de estudiantes universitarios cuya carrera está relacionada con sistemas.

δ_2 = Una variación de una unidad monetaria en el precio del procesador de la competencia, genera en promedio una variación de δ_2 unidades monetarias en la demanda del último procesador de Inter.

δ_3 = Una variación de una unidad monetaria en el precio en el mercado de los laptop que utilizan este procesador, genera en promedio una variación de δ_3 unidades monetarias en la demanda del último procesador de Inter.

δ_4 = Una variación de una unidad monetaria en el precio en el mercado de las USB de gran capacidad de almacenamiento, genera en promedio una variación de δ_4 unidades monetarias en la demanda del último procesador de Inter.

δ_5 = Una variación de una unidad en la cantidad de estudiantes universitarios cuya carrera está relacionada con sistemas, genera en promedio una variación de δ_5 unidades monetarias en la demanda del último procesador de Inter.

- c. Estime el modelo (2) y repórtelo en una tabla

Tabla 2: Estimación del comportamiento de la demanda de procesadores ipad

Variable dependiente: (D_t)		
MCO		
Variable	Parámetro	Significancia
Constante	1728,04 (6,260)	***
P_{si}	83,6917 (1,095)	
P_{ci}	0,08115 (0,402)	
U_i	0,1358 (0,015)	
E_i	-0,0755 (-0,992)	
R^2	0,9612	
R^2 -ajustado	0,9456	
n	15	
F	61,88	***

(*) Nivel de significancia: 10%

(**) Nivel de significancia: 5%

(***) Nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

Fuente: Datos obtenidos en Easyreg

Pregunta 5

- a. Suponiendo que no existe ninguno de los tres problemas econométricos más comunes (multi, hetero y auto), ¿Cree usted que existe algún otro problema econométrico? Argumente desde su punto de vista por qué cree que existiría o no otro problema econométrico.

Primero, dado que tenemos una muestra pequeña, es necesario comprobar a través de la prueba Jarque-Bera si existe un problema de normalidad, pues no podemos en este caso apoyarnos en el teorema del límite central para hacer inferencia de los resultados.

Entonces la hipótesis nula y alterna es:

Ho: Los errores se distribuyen normalmente

Ha: No Ho

Donde el criterio de rechazar Ho es que $JB > \chi_2^2$

De este modo, a partir de la salida de EasyReg se obtiene:

JB = 0,4980

Con un p-valor = 0.779

De esta forma no podemos rechazar la hipótesis nula de normalidad de los errores, por lo que **podemos concluir que no existe un problema de normalidad de los errores y podremos hacer inferencia de acuerdo a los resultados de las estimaciones.**

Por otro lado, se puede inferir del enunciado que se está incurriendo en un claro problema de **omisión de una variable relevante**. De acuerdo a la teoría económica, y como es sabido por todos, la demanda de cualquier bien (en este caso un procesador de computadora) siempre está explicada en alguna medida por su propio precio. Ahora bien, solo si esta variable “Precio del procesador Ipad” está correlacionada con las variables incluidas en el modelo, tendremos realmente un problema grave. En este caso, se puede afirmar que esta variable estaría correlacionada con las variables que representan el precio del bien sustituto y el precio del bien complementario, y por tanto existe un problema en cuanto los estimadores son sesgados, inconsistentes e ineficientes, a la luz de este tipo de problema econométrico.

De igual forma, vemos que el investigador incluye una variable que mide el precio de una USB que, en este caso, no explicaría el comportamiento de la demanda de un procesador para computadoras. Es posible decir que se estaría incurriendo en un problema de **inclusión de una variable irrelevante**, lo cual implica que aunque los estimadores siguen siendo insesgados y consistentes, las varianzas de los betas son ahora más grandes y por tanto no serían eficientes.

Pregunta 6

- a. Solucione (de ser posible) el problema econométrico encontrado.

De acuerdo a lo visto en clase, existen muchos métodos que nos ayudarían a mitigar el problema causado por la omisión de una variable relevante y la inclusión de una variable irrelevante. Sin embargo, la solución más pertinente y que genera mayor confiabilidad es resolver el problema por medio de **la teoría económica**. En este caso, la solución radicaría en **incluir la variable “Precio del procesador Ipad”**.

- b.** Ahora bien, el alcalde afirma: “en las encuestas realizadas algunas tiendas no quisieron revelar el valor real de la cantidad demandada de procesadores ipad”. Que implicaciones tendría para la investigación esta afirmación de ser cierta.

Posiblemente existirá un error de medición en la variable dependiente que es la cantidad demandada de procesadores ipad. No implica ningún problema en los estimadores MCO, si existiera el error de medición aun son MELI.