Universidad Carlos III de Madrid
Econometría I
Examen Final, Convocatoria Extraordinaria, Curso 2009-2010. **SOLUCIONES**

6 de Septiembre de 2010

Duración del examen: 2 horas.
Nota importante: Alguna información contenida en las salidas es redundante.

PREGUNTA 1 *Una de las funciones de producción más utilizadas es la función CES (Constant Elasticity substitution) debido a que anida como casos particulares otras funciones de producción utilizadas habitualmente en la literatura empírica tales como la Cobb-Douglas o la Leontieff. La expresión de la función CES es la siguiente:*

$$Y = \gamma(\delta K^{\rho} + (1-\delta)L^{\rho})^{v/\rho}$$

donde Y, K y L denotan la producción, el capital y el trabajo, γ es el parámetro de eficiencia, δ es la proporción en que los dos factores entran en la función de producción, ρ es el parámetro que define la elasticidad de substitución y v es el parámetro que mide los rendimientos a escala, de manera que $v = 1$, $v > 1$ y $v < 1$ indican, respectivamente, rendimientos constantes, rendimientos crecientes y rendimientos decrecientes a escala.

Un investigador ha especificado un modelo econométrico para estimar la tecnología con datos de 25 empresas manufactureras basándose en la aproximación lineal de primer orden de la función CES (expresada en logaritmos):

$$\begin{aligned} \log Y &= \log\gamma + \delta v \log K + (1-\delta)v\log L - 1/2\rho\log(1-\delta)v[\log(K/L)]^2 + \epsilon \\ &= \beta_0 + \beta_1 \log K + \beta_2 \log L + \beta_3[\log(K/L)]^2 + \epsilon. \end{aligned}$$

(a) *Contrastar la hipótesis de rendimientos constantes a escala en la tecnología CES. Establece claramente la hipótesis nula y el método de contraste.*

La hipótesis nula es

$$H_0 : \beta_1 + \beta_2 = 1$$

y la alternativa es

$$H_0 : \beta_1 + \beta_2 \neq 1.$$

Si se sustituye la hipótesis nula en el modelo no restringido se obtiene el modelo restringido

$$\log Y = \beta_0 + \beta_1 \log K + (1-\beta_1)\log L + \beta_3[\log(K/L)]^2 + \epsilon$$

o lo que es lo mismo,

$$\log(Y/L) = \beta_0 + \beta_1 \log(K/L) + \beta_3[\log(K/L)]^2 + \epsilon.$$

Usando las Salidas 1 y 2, que estiman ambos modelos, obtenemos

$$F = \frac{SCE_r - SCE_{nr}}{SCE_{nr}}\frac{n-k-1}{1} = \frac{0.0154 - 0.0148}{0.0148}\frac{25-3-1}{1} = 0.851,$$

y como $q = 1$, $F = W$ y bajo la nula se puede aproximar por una χ^2_1, y se puede comprobar como F no es significativo a ningún nivel de significación habitual y la hipótesis nula de rendimientos constantes a escala no se rechaza.

(b) *Investigar la significatividad del parámetro β_3 en la estimación MCO.*

Usando la Salidas 1 y 3 podemos contrastar la hipótesis nula de

$$H_0 : \beta_3 = 0$$

en contra de la alternativa

$$H_0 : \beta_3 \neq 0.$$

Si se sustituye la hipótesis nula en el modelo no restringido se obtiene el modelo restringido

$$\log Y = \beta_0 + \beta_1 \log K + \beta_2 \log L + \epsilon.$$

Usando las Salidas 1 y 3, que estiman ambos modelos, obtenemos

$$F = \frac{SCE_r - SCE_{nr}}{SCE_{nr}} \frac{n-k-1}{1} = \frac{0.0161 - 0.0148}{0.0148} \frac{25-3-1}{1} = 1.845$$

y como $q = 1$, $F = W$ y bajo la nula se puede aproximar por una χ^2_1, y se puede comprobar como F no es significativo a ningún nivel de significación habitual y el parámetro β_3 no es significativo.

(c) *Considere ahora que el modelo verdadero para representar la tecnología de las empresas de un sector es*

$$\log(Y/L) = \beta_0 + \beta_1 \log(K/L) + \beta_2 [\log(K/L)]^2 + \epsilon,$$

en el que el término de error satisface los supuestos del modelo de regresión: $E(\epsilon|K,L) = 0, Var(\epsilon|K,L) = \sigma^2$*, donde* $\beta_2 > 0$*,* $Cov(\log(K/L), [\log(K/L)]^2) > 0$*.*

Si se omite $[\log(K/L)]^2$*, obtener el signo y magnitud del sesgo de inconsistencia (o sesgo asintótico) del estimador de* β_1 *en la regresión simple de* $\log(Y/L)$ *sobre* $\log(K/L)$*. ¿Cuál sería el sesgo en caso de heteroscedasticidad condicional?*

El sesgo asintótico será

$$Sesgo\left(\tilde{\beta}_1\right) = \beta_2 \frac{Cov\left(\log(K/L), [\log(K/L)]^2\right)}{Var\left(\log(K/L)\right)} > 0.$$

La heteroscedasticidad no afecta a la forma o la existencia del sesgo.

Salida 1.
Dependent Variable: $\log Y$
Method: Least Squares
Sample: 1 25
Included observations: 25

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	*-1.9602*	*2.2051*	*-0.89*	*0.384*
$\log K$	*0.6501*	*0.0303*	*21.47*	*0.000*
$\log L$	*0.5592*	*0.2075*	*2.69*	*0.013*
$[\log (K/L)]^2$	*0.0879*	-	-	-

R-squared 0.9912
Adjusted R-squared 0.9900
S.E. of regression 0.0266
Sum squared resid 0.0148

Salida 2. Dependent Variable: $\log(Y/L)$
Method: Least Squares
Sample: 1 25
Included observations: 25

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	*0.0155*	*0.0084*	*1.84*	*0.079*
$\log (K/L)$	*0.6262*	*0.0144*	*43.51*	*0.000*
$[\log (K/L)]^2$	*0.0379*	*0.0323*	*1.17*	*0.253*

R-squared 0.9921
Adjusted R-squared 0.9914
S.E. of regression 0.0264
Sum squared resid 0.0154

Salida 3. Dependent Variable: $\log Y$
Method: Least Squares
Sample: 1 25
Included observations: 25

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	*0.6394*	*1.1255*	*0.56*	*0.576*
$\log K$	*0.6130*	*0.0135*	*45.44*	*0.000*
$\log L$	*0.3214*	*0.1142*	*2.81*	*0.010*

R-squared 0.9905
Adjusted R-squared 0.9896
S.E. of regression 0.0271
Sum squared resid 0.0161

PREGUNTA 2 *Queremos entender cuales son los determinantes de la participación en actividad física. Tenemos información para adultos con edades que fluctúan entre 25 y 55 años. La variable ficticia sport, toma el valor uno si el individuo ha participado en alguna actividad física durante la anterior semana y cero en caso contrario. Como variables explicativas en nuestro modelo tenemos la variable binaria $female$, que toma el valor uno si el individuo es mujer, y las variables continuas edad (age), edad al cuadrado ($age2$), y años de educación ($yedu$).*

(a) *Interpretar el coeficiente de la variable $female$. Proponer un modelo donde el efecto de la educación y la edad sobre la decisión de realizar actividad física dependa del género. Comentar si este modelo es más general o no que estimar dos modelos lineales diferentes para hombres y para mujeres.*

Si el adulto es una mujer, tiene una probabilidad (estimada) 0.276 menor de participar en actividad física que si es hombre, todo lo demás igual (c.p.).

El modelo sería

$$\begin{aligned} sport &= \beta_0 + \beta_1 female + \beta_2 age + \beta_3 age^2 + \beta_4 yedu \\ &\quad + \beta_5 age * female + \beta_6 age^2 * female + \beta_7 yedu * female + u. \end{aligned}$$

Como todas las variables del modelo sin información sobre género (age, age^2, $yedu$) presentan efectos parciales diferentes para $female = 0$ y para $female = 1$, el modelo es equivalente a estimar dos modelos separados para las subpoblaciones de hombres y mujeres, respectivamente.

(b) *Obtener una expresión para la varianza condicional de sport en función de las variables $female$, age y $yedu$. Contrastar si existe heteroscedasticidad condicional en el modelo lineal a partir de la información proporcionada por la Salida 1.*

$$\begin{aligned} Var\left[sport|female,\ age,\ yedu\right] &= Var\left[sport|female,\ age,\ age^2,\ yedu\right] \\ &= p\left(female,\ age,\ \ yedu\right)\left\{1 - p(female,\ age,\ yedu)\right\} \\ &= \left(\beta_0 + \beta_1 female + \beta_2 age + \beta_3 age^2 + \beta_4 yedu\right) \\ &\quad \times \left(1 - \beta_0 - \beta_1 female - \beta_2 age - \beta_3 age^2 - \beta_4 yedu\right) \end{aligned}$$

donde $p\left(female,\ age,\ yedu\right) = \Pr\left(sport = 1|female,\ age,\ \ yedu\right)$.

El contraste de homoscedasticidad sería

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

que equivale a un contraste de significación global en el modelo lineal. Comparando el estadístico el estadístico $W = qF = 4 * 173.525 = 694.1$ con una χ^2_4 de valor crítico $\chi^2_{4;0,01} = 13,28$, confirmamos que es muy significativo y H_0 se rechaza en favor de que al menos un β_j es diferente de cero y por tanto $p\left(female,\ age,\ yedu\right)$ y $Var\left[sport|female,\ age,\ yedu\right]$ no son constantes.

(c) *Usando la estimación del modelo lineal de probabilidad, indicar si la probabilidad de participar en actividad física disminuye siempre con la edad. Explicar detalladamente cómo se contrastaría que para individuos de 20 años esta probabilidad disminuye con la edad.*

Se puede comprobar que

$$\begin{aligned} \frac{\partial}{\partial age} p\left(female,\ age,\ \ yedu\right) &= \beta_2 + 2\beta_3 age \\ &= -0.020 + 2 * 0.0002 age \end{aligned}$$

y esta estimación del efecto parcial es negativa si age es menor que

$$age^* = \frac{0.020}{2 * 0.0002} = 50,$$

y positiva si es mayor. Para contrastar si para los individuos de 20 años este efecto es significativamente negativo, hay que contrastar

$$\begin{aligned} H_0 &: \beta_2 + 2\beta_3 * 20 = \beta_2 + 40\beta_3 = 0 \\ H_1 &: \beta_2 + 2\beta_3 * 20 = \beta_2 + 40\beta_3 < 0, \end{aligned}$$

y para ello compararíamos el valor del estadístico t,

$$\begin{aligned} t &= \frac{\hat{\beta}_2 + 40\hat{\beta}_3}{\widehat{Var}\left(\hat{\beta}_2 + 40\hat{\beta}_3\right)^{1/2}} = \frac{-0.020 + 40 * 0.0002}{\widehat{Var}\left(\hat{\beta}_2 + 40\hat{\beta}_3\right)^{1/2}} \\ &= \frac{-0.012}{\left\{\widehat{Var}\left(\hat{\beta}_2\right) + 40^2 * \widehat{Var}\left(\hat{\beta}_3\right) + 2 * 40 * \widehat{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right)\right\}^{1/2}} \end{aligned}$$

con una normal estándard: si $t < -1.645$, rechazaríamos H_0 en favor de H_1. También se podría estimar un modelo restringido (o reparametrizado) y comparar los ajustes.

Salida 1: Estimaciones MCO utilizando 4986 observaciones desde 1–4993
Se han quitado las observaciones ausentes o incompletas: 7
Variable dependiente: sport
Desviaciones típicas robustas ante heteroscedasticidad

Variable	*Coeficiente*	*Desv. típica*	*Estadístico t*	*valor p*
const	*0.793*	*0.141*	*5.624*	*0.000*
female	*-0.276*	*0.0127*	*-21.725*	*0.000*
age	*-0.020*	*0.0074*	*-2.701*	*0.007*
age2	*0.0002*	*1.086e-04*	*1.841*	*0.066*
yedu	*0.0148*	*0.0017*	*8.768*	*0.000*

Media de la var. dependiente	*0,359807*
D.T. de la variable dependiente	*0,479992*
Suma de cuadrados de los residuos	*1021,94*
Desviación típica de los residuos ($\hat{\sigma}$)	*0,452955*
R^2	*0,110198*
$\bar{R}^2$ *corregido*	*0,109483*
$F(4, 4981)$	*173,525*
Log-verosimilitud	*-3123,6*

PREGUNTA 3 *El siguiente es un modelo de ecuaciones simultáneas para estudiar si la apertura de la economía (open) lleva a menores tasas de inflación* (inf),

$$\begin{aligned} inf &= \delta_{10} + \gamma_{12} open + \delta_{11} \log(pcinc) + u_1 \\ open &= \delta_{20} + \gamma_{21} inf + \delta_{21} \log(pcinc) + \delta_{22} \log(land) + u_2. \end{aligned}$$

Se asume que (los logaritmos de) pcinc (renta per cápita) y land (tierra de cultivo) son exógenos en todo el ejercicio. Se han obtenido diferentes estimaciones por MCO y MC2E que se proporcionan a continuación de las cuestiones.

(a) *Obtener la forma reducida del sistema. ¿Cómo se estimarían consistentemente sus parámetros?*

Forma reducida:

$$
\begin{aligned}
open &= \frac{1}{1-\gamma_{21}\gamma_{12}}\left\{\delta_{20}+\gamma_{21}\delta_{10}+[\gamma_{21}\delta_{11}+\delta_{21}]\log(pcinc)+\delta_{22}\log(land)+\gamma_{21}u_1+u_2\right\}\\
inf &= \frac{1}{1-\gamma_{21}\gamma_{12}}\left\{\delta_{10}+\gamma_{12}\delta_{20}+[\delta_{11}+\gamma_{12}\delta_{21}]\log(pcinc)+\gamma_{12}\delta_{22}\log(land)+u_1+\gamma_{12}u_2\right\}.
\end{aligned}
$$

En cualquier forma reducida, como los regresores son exógenos, se puede aplicar MCO para obtener estimadores consistentes.

(b) *Estudiar la identificación del sistema de ecuaciones simultáneas con las estimaciones proporcionadas. ¿Hay alguna ecuación potencialmente sobreidentificada?*

La primera ecuación estructural estaría (exactamente) identificada si $\log(land)$ fuese significativa en la forma reducida de *open* : en la Salida 4 comprobamos que el estadístico t es -9,2937 con p-valor 0,0000, y por tanto es significativa, y confirmamos que la primera ecuación está identificada.

La segunda ecuación estructural no puede estar identificada en ningún caso porque no hay instrumentos disponibles para inf.

Ninguna ecuación puede estar sobreidentificada porque el único instrumento disponible es $\log(land)$ para *open* en la primera ecuación.

(c) *Contrastar si open es un regresor endógeno en la primera ecuación. En función del resultado comentar qué estimadores de los parámetros $(\gamma_{12}, \delta_{11})$ serían preferibles.*

Para contrastar si *open* es un regresor endógeno en la primera ecuación estructural usamos el contraste de Hausman en la Salida 5: $\chi^2_1 = 1{,}35333$ con valor p = 0,244697, por lo que no se rechaza la hipótesis nula de exogeneidad (y por tanto los EMCO serían consistentes y más eficientes siempre que los EMC2E).

(d) *Si se sabe que $\delta_{11} = 0$, ¿cómo cambiaría tu respuesta a la pregunta (b)?*

En ese caso la primera ecuación estructural estaría potencial sobreidentificada, ya que *open* tendría dos instrumentos, $\log(pcinc)$ y $\log(land)$ potencialmente válidos (sin embargo en la Salida 4 vemos que ambos son conjuntamente significativos, confirmando la identificación, pero $\log(pcinc)$ no lo es individualmente, descartando la sobreidentificación).

La segunda ecuación del sistema seguiría sin identificar, ya que siguen sin existir instrumentos disponibles para inf.

¿Y si alternativamente se sabe que $\delta_{21} = 0$ (pero se ignora el valor de δ_{11} o el de cualquier otro parámetro del sistema)?

Si $\delta_{21} = 0$, no afectaría a la identificación de la primera ecuación, ya que $\log(land)$ seguiría siendo un instrumento potencialmente válido para *open* (y no hay posibilidad de sobreidentificación).

En cambio ahora $\log(pcinc)$ no aparece en la segunda ecuación estructural y por tanto puede ser un instrumento válido para inf : sin embargo la Salida 3 descarta esta posibilidad, ya que $\log(pcinc)$ no es significativa en la forma reducida de inf.

Salida 1: Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	*Coeficiente*	*Desv. típica*	*Estadístico t*	*valor p*
const	*25,1040*	*15,2052*	*1,6510*	*0,1016*
open	*-0,215070*	*0,0946289*	*-2,2728*	*0,0250*
lpcinc	*0,0175673*	*1,97527*	*0,0089*	*0,9929*

Media de la var. dependiente	*17,2640*
D.T. de la variable dependiente	*23,9973*
Suma de cuadrados de los residuos	*62127,5*
Desviación típica de los residuos ($\hat{\sigma}$)	*23,6581*
R^2	*0,0452708*
$\bar{R}^2$ *corregido*	*0,0280685*
$F(2, 111)$	*2,63167*
valor p para $F()$	*0,0764453*

Salida 2: Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	116,226	15,8808	7,3187	0,0000
inf	-0,0680353	0,0715556	-0,9508	0,3438
lpcinc	0,559501	1,49395	0,3745	0,7087
lland	-7,3933	0,834814	-8,8563	0,0000

Media de la var. dependiente	37,0789
D.T. de la variable dependiente	23,7535
Suma de cuadrados de los residuos	34865,3
Desviación típica de los residuos ($\hat{\sigma}$)	17,8033
R^2	0,453162
$\bar{R}^2$ corregido	0,438249
$F(3, 110)$	30,3855
valor p para $F()$	< 0,00001

Salida 3: Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	-12,615	21,0313	-0,5998	0,5498
lpcinc	0,191394	1,98158	0,0966	0,9232
lland	2,55380	1,08049	2,3635	0,0198

Media de la var. dependiente	17,2640
D.T. de la variable dependiente	23,9973
Suma de cuadrados de los residuos	61903,2
Desviación típica de los residuos ($\hat{\sigma}$)	23,6154
R^2	0,0487174
$\bar{R}^2$ corregido	0,0315772
$F(2, 111)$	2,84229
valor p para $F()$	0,0625432

Salida 4: Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	*Coeficiente*	*Desv. típica*	*Estadístico t*	*valor p*
const	*117,085*	*15,8483*	*7,3878*	*0,0000*
lpcinc	*0,546479*	*1,49324*	*0,3660*	*0,7151*
lland	*-7,5671*	*0,814216*	*-9,2937*	*0,0000*

Media de la var. dependiente	*37,0789*
D.T. de la variable dependiente	*23,7535*
Suma de cuadrados de los residuos	*35151,8*
Desviación típica de los residuos ($\hat{\sigma}$)	*17,7956*
R^2	*0,448668*
$\bar{R}^2$ *corregido*	*0,438734*
$F(2, 111)$	*45,1654*
valor p para $F()$	*< 0,00001*

Salida 5: Estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: inf
Instrumentos: lland

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	26,8993	15,4012	1,7466	0,0807
open	-0,337487	0,144121	-2,3417	0,0192
lpcinc	0,375823	2,01508	0,1865	0,8520

Media de la var. dependiente	17,2640
D.T. de la variable dependiente	23,9973
Suma de cuadrados de los residuos	63064,2
Desviación típica de los residuos ($\hat{\sigma}$)	23,8358
$F(2,111)$	2,62498
valor p para $F()$	0,0769352

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 1{,}35333$
con valor p = 0,244697

First-stage $F(1,111) = 86{,}3734$

VALORES CRÍTICOS:

$Z_{0,025} = 1,96$	$Z_{0,05} = 1,645$	$Z_{0,01} = 2,326$	$Z_{0,005} = 2,576$	
$Z_{0,1} = 1,282$	$\chi^2_{3;0,01} = 11,34$	$\chi^2_{3;0,05} = 7,82$	$\chi^2_{5;0,05} = 11,07$	
$\chi^2_{2;0,05} = 5,99$	$\chi^2_{2;0,01} = 9,21$	$\chi^2_{6;0,05} = 12,59$	$\chi^2_{2;0,1} = 4,61$	$\chi^2_{1;0,05} = 3,84$
$\chi^2_{6;0,01} = 16,81$	$\chi^2_{4;0,05} = 9,49$	$\chi^2_{3;0,1} = 6,25$	$\chi^2_{4;0,01} = 13,28$	$\chi^2_{1;0,01} = 6,64$

Z es la normal de media cero y varianza uno y χ^2_q es la chi cuadrado con q grados de libertad, $\Pr(Z > Z_\alpha) = \alpha$; $\Pr\left(\chi^2_q > \chi^2_{q;\alpha}\right) = \alpha$.

Nótese que la distribución F se puede aproximar por la de la χ^2. Esto es, $\chi^2_q \sim q \cdot F_{q,n}$ cuando n es grande, $\Pr\left(\chi^2_q > \chi^2_{q;\alpha}\right) \simeq \Pr\left(q \cdot F_{q,n} > \chi^2_{q;\alpha}\right)$.