Econometría

Universidad Carlos III de Madrid

Examen Final. SOLUCIÓN

2 de Julio de 2013

1. [**3 puntos/sobre 12**] Algunos autores sostienen que los precios de las acciones de las empresas dependen de la política de dividendos seguida por los gestores de las mismas. El modelo económico empleado para explicar el valor de mercado de las acciones de una empresa, P (precio de cotización en bolsa (en euros)) en función de VC, el valor contable (en euros) de las mismas, y DPA, el dividendo (en euros) obtenido por acción, se estima por MCO con una muestra de 22 empresas que cotizan en Bolsa. Los resultados son

```
Variable dependiente:  ln(P)
Método:  Mínimos Cuadrados Ordinarios
Muestra:  1-22
Observaciones Incluidas:  22
Variable Coeficiente Std.  Error Estadístico t P-v.
C        1.0913      0.4251     2.57           0.0190
ln(VC)   0.7781      0.0935     8.32           0.000
ln(DPA)  0.1814      0.0820     2.21           0.039
R-cuadrado 0.8669
E.E. de regresión 0.4069
Suma Cuadrados Residuos 3.1460
```

.

a.) *Interpretar los coeficientes del modelo estimado. Proponer un modelo alternativo en el que los cambios en el valor de mercado de las acciones sean proporcionales a los cambios porcentuales en el valor contable y en los dividendos.* [**1 punto/sobre 12**]

β_1 y β_2 son elasticidades: β_1 es el cambio porcentual en el precio (P) cuando VC se incrementa un 1%, β_2 es el cambio porcentual en el precio (P) cuando DPA se incrementa un 1%.

El modelo alternativo sería

$$P_i = \beta_0 + \beta_1 \ln VC_i + \beta_2 \ln DPA_i + e_i.$$

b.) *Obtener un intervalo de confianza aproximado para el efecto de los cambios del valor contable sobre el precio de cotización de las acciones, al 95% de confianza. ¿Es ese efecto significativo?.*[**1 punto/sobre 12**]

El Intervalo de confianza es $\hat{\beta}_1 \pm 1.96 se\left(\hat{\beta}_1\right) = 0.7781 \pm 1.96 * 0.0935 = [0.595, 0.961]$.

Es significativo porque el IC no incluye el valor 0.

c.) *Si ln(VC) y* $\ln(DPA)$ *están correlacionadas positivamente, ¿que ocurriría con el coeficiente de ln(VC) en una regresión que omitiese* $\ln(DPA)$*? ¿Seguiría siendo consistente, sobreestimaría o infraestimaría el valor real? Justifica tu respuesta.* **[1 punto/sobre 12]**

Escribiendo

$$\begin{aligned} \ln P_i &= \beta_0 + \beta_1 \ln VC_i + \beta_2 \ln DPA_i + u_i \\ &= \beta_0 + \beta_1 \ln VC_i + v_i, \qquad v_i = \beta_2 \ln DPA_i + u_i, \end{aligned}$$

usando que el estimador MCO de la regresión simple de $\ln P$ sobre $\ln VC$ satisface

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \frac{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right) v_i}{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right)^2} \\ &= \beta_1 + \beta_2 \frac{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right) \ln DPA_i}{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right)^2} + \frac{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right) u_i}{\sum_{i=1}^n \left(\ln VC_i - \overline{\ln VC_i}\right)^2} \\ &\to {}_p \beta_1 + \beta_2 \frac{Cov\left(\ln VC_i, \ln DPA_i\right)}{Var\left(\ln VC_i\right)}, \end{aligned}$$

y como la Cov y β_2 son positivos (porque podemos pensar que el efecto causal de DPA es positivo), tendremos un sesgo positivo en $\hat{\beta}_1$, que sobreestimará el valor verdadero de β_1.

2. **[5 puntos/sobre 12]** Se está llevando a cabo un estudio sobre los salarios/hora de mujeres trabajadoras. La muestra tiene 252 observaciones sobre las siguientes variables

$$\begin{aligned} W_i &\equiv \text{Salario hora de la trabajadora } i, \text{ medido dólares/hora.} \\ Ed_i &\equiv \text{N}^o \text{ total de años de educación formal completados por la trabajadora } i. \\ Exp_i &\equiv \text{Años de experiencia laboral acumulados por la trabajadora } i. \end{aligned}$$

Considere el siguiente modelo de regresión log-lineal,

$$\ln W_i = \beta_0 + \beta_1 Ed_i + \beta_2 Exp_i + \beta_3 Ed_i^2 + \beta_4 Exp_i^2 + \beta_5 Ed_i \cdot Exp_i + U_i, \qquad (1)$$

donde los $\beta_j's$ $(j = 0, 1, ..., 5)$ son los coeficientes de la regresión. Se supone que se satisfacen todos los supuestos clásicos del modelo de regresión; en particular, los errores son homocedásticos. Se estima este modelo por MCO y otros dos en el que se imponen ciertas restricciones sobre los coeficientes. Los resultados son los siguientes:

Regresores	(1)	(2)	(3)
Constante: $\hat{\beta}_0$	1,150	1,162	0.3348
	(0,4410)	(0,2399)	(0,1415)
Ed_i : $\hat{\beta}_1$	-0,0867	-0,0882	0,0823
	(0,0591)	(0,0382)	(0,0105)
Exp_i : $\hat{\beta}_2$	0,02102	0,02059	0,00411
	(0,0146)	(0,06428)	(0,0019)
Ed_i^2 : $\hat{\beta}_3$	0,00742	0,00746	–
	(0,0021)	(0,0016)	–
Exp_i^2 : $\hat{\beta}_4$	-0,00038	-0,00038	–
	(0,00015)	(0,00014)	–
$Ed_i \cdot Exp_i$: $\hat{\beta}_5$	-0,00003	–	–
	(??)	–	–
$RSS =$	35,4317	35,4419	39,6443
$TSS =$	49,5336	49,5336	49,5336
$n =$	252	252	252

a.) *Obtenga el efecto parcial de la educación sobre el logaritmo del salario implicado por la regresión (1). Proponga restricciones sobre los coeficientes en (1) para que este efecto parcial sea siempre igual a cero para cualquier valor de las variables explicativas. Suponga que una vez introducidas estas restricciones en el modelo, el R^2 del ajuste por MCO es de 0,0428. Exponga formalmente la hipótesis nula de que el efecto parcial es igual a cero en términos de los coeficientes originales y realice el contraste de esta hipótesis al 5% de significación.* **[1 punto/sobre 12]**

El efecto parcial es

$$\beta_1 + 2\beta_3 Ed_i + \beta_5 Exp_i,$$

que es cero siempre si se cumple

$$H_0 : \beta_1 = \beta_3 = \beta_5 = 0.$$

Para realizar el contraste usaremos un test de la F,

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} \frac{n-k-1}{q},$$

donde $R_{nr}^2 = 1 - 35.4317/49.5336 = 0.2847$ y $R_r^2 = 0.0428$ son los coeficientes de determinación de los modelos no restringido y restringido, respectivamente, $q = 3$ es el número de restricciones contrastadas y $n - k - 1 = 252 - 5 - 1 = 246$ son los grados de libertad. Por tanto

$$F = \frac{0.2847 - 0.0428}{1 - 0.2847} \frac{246}{3} = 27.731$$

donde $F \sim_{aprox.} \chi_3^2/3$ si H_0 es cierta, con valor crítico igual a $7.81/3 = 2.603$ al 5%, por lo que rechazamos H_0 en favor de la alternativa de que al menos uno de los tres coeficientes es diferente de cero.

b.) *Escriba formalmente la hipótesis de que el modelo es lineal en las variables explicativas en términos de los coeficientes del modelo. Contraste esta hipótesis al 1% de significación.* [**1 punto/sobre 12**]

La hipótesis es

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$$

que se corresponde con el modelo de la columna (3).

Para realizar el contraste usaremos un test de la F,

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q}$$

donde $R^2_{nr} = 1 - 35.4317/49.5336 = 0.2847$ y $R^2_r = 1 - 39.6443/49.5336 = 0.199\,65$ son los coeficientes de determinación de los modelos no restringido y restringido, respectivamente, $q = 3$ es el número de restricciones contrastadas y $n - k - 1 = 252 - 5 - 1 = 246$ son los grados de libertad. Por tanto

$$F = \frac{0.2847 - 0.199\,65}{1 - 0.2847} \frac{246}{3} = 9.75$$

donde $F \sim \chi^2_3/3$ si H_0 es cierta, con valor crítico igual a $7.81/3 = 2.603$ al 5%, por lo que rechazamos H_0 en favor de la alternativa de que al menos uno de los tres coeficientes es diferente de cero.

c.) *Explique en palabras que significa la restricción impuesta en la columna (2) respecto a la (1). Contraste si esta restricción no puede rechazarse con la muestra observada al 10% de significación. Basándose en este contraste, ¿qué modelo preferiría?, ¿por qué?* [**1 punto/sobre 12**]

La restricción impuesta, $H_0 : \beta_5 = 0$, implica que no hay interacción entre las variables Ed_i y Exp_i, por lo que el efecto parcial de *educación* no depende del nivel de experiencia (y viceversa). Para contrastar la hipótesis no podemos realizar un contraste de la t,

$$t = \frac{\hat{\beta}_5}{e.e.\left(\hat{\beta}_5\right)}$$

porque desconocemos el valor de $e.e.\left(\hat{\beta}_5\right)$, por lo que podemos realizar el contraste de la F equivalente (en contra de una alternativa bilateral. Ahora tenemos

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q}$$

donde $R^2_{nr} = 1 - 35.4317/49.5336 = 0.2847$ y $R^2_r = 1 - 35.4419/49.5336 = 0.2845$ son los coeficientes de determinación de los modelos no restringido y restringido, respectivamente, $q = 1$ es el número de restricciones contrastadas y $n - k - 1 = 252 - 5 - 1 = 246$ son los grados de libertad. Por tanto

$$F = \frac{0.2847 - 0.2845}{1 - 0.2847} \frac{246}{1} = 0.0688,$$

donde $F \sim \chi^2_1$ si H_0 es cierta, con valor crítico igual a $1.96^2 = 3.84$ al 5%, por lo que no rechazamos H_0 y mantenemos que dicha interacción es cero.

d.) *Utilizando los resultados en la columna (2) calcular el efecto parcial de la educación sobre el logaritmo del salario. Contrastar si dicho efecto parcial es igual a cero para todas las mujeres con 16 años de educación (*$Ed_i = 16$*). Se sabe que la covarianza estimada entre* $\hat{\beta}_1$ *y* $\hat{\beta}_3$ *es -0,00001066. Escriba la hipótesis nula y alternativa en términos de los coeficientes.* **[1 punto/sobre 12]**

El efecto parcial es

$$\beta_1 + 2\beta_3 Ed_i,$$

ya que está impuesto que $\beta_5 = 0$, por lo que la hipótesis a contrastar, con $Ed_i = 16$, es

$$H_0 : \beta_1 + 32\beta_3 = 0,$$

en contra de $H_1 : \beta_1 + 32\beta_3 \neq 0$ mediante un contraste de la t,

$$t = \frac{\hat{\beta}_1 + 32\hat{\beta}_3}{e.e.\left(\hat{\beta}_1 + 32\hat{\beta}_3\right)} = \frac{-0.0882 + 32 * 0.00746}{0.0293} = 5.137\,2$$

ya que

$$\begin{aligned} e.e.\left(\hat{\beta}_1 + 32\hat{\beta}_3\right) &= \left(e.e.\left(\hat{\beta}_1\right)^2 + 32 * e.e.\left(\hat{\beta}_3\right)^2 + 2 * 32 * \widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_3\right)\right)^{1/2} \\ &= \left(0.0382^2 + 32 * 0.0016^2 + 2 * 32 * (-0.00001066)\right)^{1/2} \\ &= 0.0293. \end{aligned}$$

Comparando con el valor crítico de una normal estándar al 5%, 1.96, se rechaza H_0 en favor de H_1.

e.) *Obtener el error estándar de* $\hat{\beta}_5$ *en columna (2) con la información de la tabla utilizando la relación entre el estadístico* F *y el* t*. ¿Es el error estándar obtenido de esta forma robusto a la heterocedasticidad?* **[1 punto/sobre 12]**

Usando que $t^2 = F$ calculado en el apartado *c.*), donde el t se calcula con la estimación del modelo no restringido, columna (1),

$$t = \frac{\hat{\beta}_5}{s.e.\left(\hat{\beta}_5\right)}$$

tenemos que

$$s.e.\left(\hat{\beta}_5\right) = \left(\frac{\hat{\beta}_5^2}{F}\right)^{1/2} = \left(\frac{(-0.00003)^2}{0.0688}\right)^{1/2} = 0.00011437.$$

Este error no es robusto a la heterocedasticidad, ya que F está calculado de la forma no robusta.

3. **[4 puntos/sobre 12]** Para estudiar los factores que determinan el peso de los niños al nacer se considera el modelo de regresión lineal

$$\ln(bwgth) = \beta_0 + \beta_1 packs + \beta_2 faminc + \beta_3 male + \beta_4 white + u$$

donde las variables se definen como:

bwgth = peso del niño al nacer, en libras;

packs = número de paquetes de tabaco que fuma la madre al día;

faminc = ingresos familiares, en miles de dólares;

male = variable binaria que toma el valor 1 si el niño es varón y 0 en caso contrario;

white = variable binaria que toma el valor 1 si el niño es de raza blanca y 0 en caso contrario;

.

Salida 1: MCO, usando las observaciones 1–1388
Variable dependiente: lbwght

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	4.69883	0.0135412	347.0026	0.0000
packs	−0.0827900	0.0170675	−4.8507	0.0000
faminc	0.000522446	0.000285778	1.8282	0.0677
male	0.0263017	0.0100504	2.6170	0.0090
white	0.0522238	0.0128152	4.0752	0.0000
	.	.	.	.

Media de la vble. dep.	4.760031	D.T. de la vble. dep.	0.190662
Suma de cuad. residuos	48.27294	D.T. de la regresión	0.186827
R^2	0.042590	R^2 corregido	0.039821
$F(4, 1383)$	15.38048	Valor p (de F)	2.59e–12

.

Salida 2: MCO, usando las observaciones 1–1388
Variable dependiente: packs

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	0.313403	0.103414	3.0306	0.0025
faminc	−0.00117833	0.000489073	−2.4093	0.0161
male	−0.0184887	0.0151293	−1.2220	0.2219
white	0.0313756	0.0212964	1.4733	0.1409
cigprice	0.000661987	0.000735415	0.9002	0.3682
fatheduc	−0.00519264	0.00372076	−1.3956	0.1631
motheduc	−0.0169061	0.00416662	−4.0575	0.0001
	.	.	.	.

Media de la vble. dep.	0.088455	D.T. de la vble. dep.	0.267189
Suma de cuad. residuos	80.12099	D.T. de la regresión	0.260134
R^2	0.056887	R^2 corregido	0.052108
$F(6, 1184)$	11.90282	Valor p (de F)	5.27e–13

.

.

Salida 3: MCO, usando las observaciones 1–1388
Variable dependiente: packs

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	0.171102	0.0208246	8.2163	0.0000
faminc	−0.00294700	0.000443055	−6.6515	0.0000
male	−0.00510169	0.0158282	−0.3223	0.7473
white	0.0273467	0.0201697	1.3558	0.1754
	.	.	.	.

Media de la vble. dep.	0.104359	D.T. de la vble. dep.	0.298634
Suma de cuad. residuos	119.8234	D.T. de la regresión	0.294241
R^2	0.031309	R^2 corregido	0.029209
$F(3, 1384)$	14.91058	Valor p (de F)	1.48e–09

.

.

Salida 4: MC2E, usando las observaciones 1–1388
Variable dependiente: lbwght
Mediante Instrumentos: packs
Instrumentos: const cigprice fatheduc motheduc faminc male white

	Coeficiente	Desv. Típica	z	Valor p
const	4.69990	0.0241955	194.2473	0.0000
packs	−0.0549186	0.117545	−0.4672	0.6403
faminc	0.000556410	0.000422354	1.3174	0.1877
male	0.0344248	0.0109986	3.1299	0.0017
white	0.0434922	0.0156143	2.7854	0.0053
	.	.	.	.

Media de la vble. dep.	4.767536	D.T. de la vble. dep.	0.188013
Suma de cuad. residuos	40.51189	D.T. de la regresión	0.184820
R^2	0.037882	R^2 corregido	0.034637
$F(4, 1186)$	6.478370	Valor p (de F)	0.000037
Estadístico J	2.414230		

.

a.) *Discuta si es razonable que packs sea exógena en la regresión anterior. ¿Qué otros factores estarían contenidos en u? ¿Qué ocurriría con los estimadores MCO de la Salida 1 si packs no es exógena?* [**1 punto/sobre 12**]

La variable *packs* no puede ser exógena, ya que está correlada con otros factores omitidos (con efecto causal sobre $\ln(bwght)$) recogidos en el término de error u, como hábitos alimenticios y sanitarios/higiénicos. En ese caso los estimadores MCO estarán sesgados y serán inconsistentes.

b.) *Se consideran tres posibles instrumentos para packs, asumiendo que el resto de variables explicativas son exógenas,*

$cigprice$ = precio medio del paquete de cigarrillos en el estado de nacimiento;

$motheduc$ = años de educación de la madre;

$fatheduc$ = años de educación del padre;

Usar la información precisa de las anteriores estimaciones para comprobar la relevancia de los instrumentos y si hay un problema de instrumentos débiles. [**1 punto/sobre 12**]

Para comprobar la relevancia debemos contrastar en la salida 2 (estimación de la forma reducida) la significación conjunta de las variables instrumentales, es decir contrastar

$$H_0 : \beta_{cigprice} = \beta_{motheduc} = \beta_{fatheduc} = 0$$

mediante un test de la F, comparando con la salida 3, que es el modelo restringido correspondiente,

$$F = \frac{R^2_{nr} - R^2_r}{1 - R^2_{nr}} \frac{n-k-1}{q},$$

donde R^2_{nr} =0.056887 y R^2_r =0.031309 son los coeficientes de determinación de los modelos no restringido y restringido, respectivamente, $q = 3$ es el número de restricciones contrastadas y $n - k - 1 = 1388 - 6 - 1 = 1381$ son los grados de libertad. Por tanto

$$F = \frac{0.056887 - 0.031309}{1 - 0.056887} \frac{1388 - 6 - 1}{3} = 12.485$$

donde $F \sim_{aprox.} \chi^2_3/3$ si H_0 es cierta, con valor crítico igual a $7.81/3 = 2.603$ al 5%, por lo que rechazamos H_0 en favor de la alternativa de que al menos uno de los tres coeficientes es diferente de cero, y por tanto los instrumentos son relevantes conjuntamente. Además como $F > 10$ concluimos que no son instrumentos débiles.

c.) *Explicar cómo se podrían obtener los estimadores MC2E de la Salida 4 mediante una regresión MCO.* [**1 punto/sobre 12**]

Habría que realizar las dos etapas, la primera estimaría por MCO la forma reducida correspondiente, usando $cigprice$, $fatheduc$, $motheduc$ como instrumentos de $packs$,

$$packs = \delta_0 + \delta_1 cigprice + \delta_2 fatheduc + \delta_3 motheduc + \delta_4 faminc + \delta_5 male + \delta_6 white + error$$

y obteniendo las predicciones $\widehat{packs}$, y la segunda etapa estimaría por MCO el modelo original, sustituyendo $packs$ por $\widehat{packs}$

$$\ln(bwgth) = \beta_0 + \beta_1 \widehat{packs} + \beta_2 faminc + \beta_3 male + \beta_4 white + error.$$

d.) *Explicar cómo se realiza el contraste de las restricciones de sobreidentificación, realizar dicho contraste con la información proporcionada si es posible e interpretar sus resultados.* **[1 punto/sobre 12]**

Como tenemos en principio más instrumentos ($m = 3$) que regresores endógenos ($k = 1$) podemos contrastar las restricciones de sobreidentificación para estudiar si los instrumentos son exógenos, que es la hipótesis nula. Para ello deberíamos regresar los residuos del modelo obtenidos por la estimación MC2E,

$$\begin{aligned} \hat{u}_i &= \ln(bwgth_i) - \ln\widehat{(bwgth_i)} \\ &= \ln(bwgth_i) - \hat{\beta}_0 - \hat{\beta}_1 packs_i + \hat{\beta}_2 faminc_i + \hat{\beta}_3 male_i + \hat{\beta}_4 white_i \end{aligned}$$

sobre todas las variables exógenas, $cigprice, fatheduc, motheduc, faminc, male$ y $white$ y contrastar la significación conjunta de los instrumentos mediante el estadístico $J = mF$ de Sargam comparado con una χ^2_2 ya que tenemos $m - k = 3 - 1 = 2$ restricciones de sobreidentificación. En este caso $J = 2.144$, que no es significativo comparado con una χ^2_2 al 5% (vc=6) y por tanto no rechazamos que los instrumentos sean exógenos.