Econometría
Universidad Carlos III de Madrid
Examen Final: SOLUCION
22 de Mayo de 2013

1. [5 puntos/sobre 15] Se está llevando a cabo una investigación sobre los salarios anuales de los CEO's (Chief Executive Officers) de compañías privadas con alto nivel de negocio. La muestra consiste en observaciones de 177 compañías sobre las siguientes variables:

$Salary_i \equiv$ Salario anual de un CEO de la compañía i, medido en miles de dólares.
$TR_i \equiv$ Volumen total de ventas de la empresa i, medido en millones de dólares.
$Value_i \equiv$ Valor de mercado de la empresa i, medido en millones de dólares.
$Ten_i \equiv$ N^o de años que el CEO de la compañía i ha sido empleado por su empresa.
$Age_i \equiv$ La edad del CEO de la compañía i en años.

Se propone el siguiente modelo de regresión para los salarios de los CEO:

$$\ln Salary_i = \beta_0+\beta_1 \ln TR_i+\beta_2 \ln Value_i+\beta_3 Ten_i+\beta_4 Ten_i^2+\beta_5 Age_i+\beta_6 Age_i^2+U_i, \quad (1)$$

donde los $\beta's$ son los coeficientes de regresión y se supone que se satisfacen todos los supuestos del modelo de regresión. Se sabe que los errores no sufren de heterocedasticidad. Se obtiene las siguientes estimaciones por MCO:

Regresores	(1)	(2)	(3)
Constante $\hat{\beta}_0$	5,572	4,369	6,060
	(1,144)	(0,2587)	(1,152)
$\ln TR_i$ $\hat{\beta}_1$	0,1820	0,1646	0,1800
	(0,0412)	(0,0386)	(0,0423)
$\ln Value_i$ $\hat{\beta}_2$	0,1023	0,1085	-
	(0,0493)	(0,0488)	-
Ten_i $\hat{\beta}_3$	0,04593	0,0451	-
	(0,0142)	(0,0141)	-
Ten_i^2 $\hat{\beta}_4$	-0,00122	-0.00121	-
	(0.0412)	(0.00047)	-
Age_i $\hat{\beta}_5$	-0.04201	-	-0.0533
	(0.0412)	-	(0.0415)
Age_i^2 $\hat{\beta}_6$	0.000332	-	0.00046
	(0.00036)	-	(0.00036)
$SCR =$	42,060	42,474	44,879
$SCT =$	64,646	64,646	64,646
$n =$	177	177	177

a.) *Utilice la columna (1) en esta pregunta. Interprete el coeficiente β_1. Contraste al 5% de significación que la edad (Age) entra de forma lineal en el modelo.* **[1 punto/sobre 15]**

(1/2) Interpretación de β_1 : si el volumen total de ventas aumenta un 1%, entonces el salario aumenta en promedio un 0,18% (ambas variables en logaritmos). (1/2) *Age* entra en forma lineal si el coeficiente de Age^2 no es significativo en el contraste

$$H_0 : \beta_6 = 0$$

en contra de $H_0 : \beta_6 \neq 0$. Calculando

$$t_{\beta_6} = \frac{0.00032}{0.00036} = 0.888\,89$$

vemos que no es significativo a los niveles habituales y por tanto el término cuadrático no es significativo: *Age* entra en el modelo en forma lineal.

b.) *Utilice la columna (1) para realizar un contraste de significación conjunta al 1% sobre todas la variables que entran en el modelo.* **[1 punto/sobre 15]**

(1/4) Este es un contraste de significatividad global (sin incluir la constante), con la hipótesis nula

$$H_0 : \beta_1 = \cdots = \beta_6 = 0$$

en contra de la alternativa de que algún coeficiente es diferente de cero. (1/2) Sólo podemos hacer la versión no robusta, que es válida porque sabemos que los errores no son heterocedásticos:

$$\begin{aligned} F &= \frac{R^2}{1-R^2}\frac{n-k-1}{k}, \\ &= \frac{0.349}{1-0.349}\frac{177-6-1}{6} = 15.189, \end{aligned}$$

donde

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{42.060}{64.646} = 0.349$$

(1/4) El valor del estadístico F hay que compararlo con el valor crítico de una $\chi^2_k/k = \chi^2_6/6$ que al 1% es igual a $16.81/6 = 2.802$ y por tanto es muy significativo: rechazamos H_0 y concluimos que las variables son significativas conjuntamente.

c.) *Se ha realizado la siguiente estimación*

$$\begin{aligned} \ln\widehat{Salary}_i &= \underset{(1{,}144)}{5{,}572} - \underset{(0{,}0841)}{0.0797}\ln Value_i + \underset{(0{,}0493)}{0{,}1023}X_i + \underset{(??)}{??}\,Ten_i - \underset{(0{,}0412)}{0{,}00122}\,Ten_i^2 \\ &\quad - \underset{(0{,}0412)}{0{,}04201}Age_i + \underset{(0{,}00036)}{0{,}000332}Age_i^2, \quad SCR = 42{,}060,\ n = 177, \end{aligned}$$

donde $X_i = \ln TR_i + \ln Value_i$. *Contraste que los efectos sobre el salario del ejecutivo de un incremento del 1% en el volumen total de ventas y en el valor de mercado de la empresa son idénticos al 1% de significación. Considere una alternativa bilateral. ¿Cuál sería el coeficiente estimado de* Ten_i *y su error estándar en la regresión anterior?* **[1 punto/sobre 15]**

(1/4) La hipótesis nula es

$$H_0 : \beta_1 = \beta_2, \text{ o igualmente, } H_0 : \theta \equiv \beta_2 - \beta_1 = 0,$$

en contra de que $H_1 : \theta \equiv \beta_2 - \beta_1 \neq 0$, donde θ es el coeficiente de $\ln Value_i$ en el modelo reparametrizado. (3/4) Empleamos el estadístico t,

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_1}{e.e.\left(\hat{\beta}_2 - \hat{\beta}_1\right)} \equiv \frac{\hat{\theta}}{e.e.\left(\hat{\theta}\right)}$$

por lo que

$$t = \frac{0.1023 - 0.1820}{0.0841} = \frac{-0.0797}{0.0841} = -0.9477,$$

que no es significativo comparado el valor crítico al 1% de una $N(0,1)$, 2.576, y no rechazamos la hipótesis de que los efectos son iguales al 1%.

El coeficiente estimado de Ten_i y su error estándar son los mismos que en la regresión original, columna (1).

d.) *Contraste al 10% de significación que la edad del ejecutivo (Age) no explica nada del salario.* **[1 punto/sobre 15]**

(1/4) Este es un contraste conjunto de

$$H_0 : \beta_5 = \beta_6 = 0,$$

en contra de que alguno de los coeficientes es diferente de cero. (1/2) Como es una hipótesis conjunta de $q = 2$ restricciones debemos calcular el estadístico F correspondiente: la versión no robusta se calcula como

$$\begin{aligned} F &= \frac{RSS_r - RSS_{nr}}{RSS_{nr}} \frac{n-k-1}{q}, \\ &= \frac{42.474 - 42.060}{42.060} \frac{177 - 6 - 1}{2} = 0.8367. \end{aligned}$$

donde RSS_{nr} y RSS_r corresponden a las regresiones (1) y (2), respectivamente. (1/4) El estadístico F se compara con el valor crítico de una $\chi^2_2/2$, que al nivel del 10% es $4,61/2 = 2,3$ aproximadamente, por lo que no es significativo y no se rechaza la hipótesis nula.

e.) *Calcule el efecto sobre el salario de un cambio de los años que un ejecutivo ha estado trabajando en la empresa (Ten) de 10 a 20, usando la columna (1). ¿Es el efecto de Ten sobre el salario siempre positivo?* **[1 punto/sobre 15]**

(1/2) En este caso tenemos que la diferencia entre las funciones de regresión estimadas y evaluadas en $Ten = 20$ y $Ten = 10$ es

$$\begin{aligned} & \hat{E}\left[\ln Salary | Ten = 20\right] - \hat{E}\left[\ln Salary | Ten = 10\right] \\ = \;& \hat{\beta}_3 (20 - 10) + \hat{\beta}_4 \left(20^2 - 10^2\right) \\ = \;& 0.04593 * (20 - 10) - 0.00122 * \left(20^2 - 10^2\right) = 0.0933 \end{aligned}$$

lo que se interpreta como que el salario promedio sube un $0.0933 * 100\% = 9.33\%$. Si el efecto se calcula como

$$\begin{aligned} & \frac{\partial}{\partial Ten} \hat{E}\left[\ln Salary | Ten_0 = 10\right] \Delta Ten \\ = \;& \left(\hat{\beta}_3 + 2\hat{\beta}_4 * Ten_0\right) \Delta Ten \\ = \;& (0.04593 - 2 * 0.00122 * 10) * 10 \\ = \;& 0.2153 \end{aligned}$$

se obtienen números bastante diferentes, ya que ΔTen es relativamente grande.

(1/2) El efecto parcial de Ten será negativo para valores muy grandes de Ten, mayores que

$$Ten^* = \frac{\hat{\beta}_3}{-2\hat{\beta}_4} = \frac{0.04593}{2 * 0.00122} = 18.824.$$

.

2. [**7 puntos/sobre 15**] Nuestro objetivo es evaluar los efectos de la televisión divulgativa sobre la formación y las actitudes del público. En el experimento, los participantes se dividieron aleatoriamente en dos grupos: el grupo A y grupo B. El grupo A se le informó sobre el próximo especial del PBS (Public Broadcasting Service) sobre la adicción a las drogas. Al Grupo B no se le dió esta información, pero se le comunicó que podría informarse sobre el programa de PBS por su cuenta. Los datos disponibles son:

$groupA$:	variable ficticia igual a 1 para los participantes en el grupo A.
$viewer$:	variable ficticia igual a 1 para los participantes que han visto el programa.
age:	edad del participante.
edu:	educación en años.
$gender$:	igual a 1 si mujer, 0 en otro caso
$income$:	renta en miles de dólares
$test\_after$:	Puntuación obtenida en la prueba de "conocimiento de adicciones", realizada después de emitir el programa medido en respuestas correctas (0-8)
$test\_before$:	Puntuación obtenida en la prueba de "conocimiento de adicciones", realizada antes de emitir el programa medido en respuestas correctas (0-8)

Supongamos que queremos estimar el efecto de ver el programa de PBS (variable $viewer$) en el conocimiento acerca de la adicción a las drogas (variable $test\_after$).

a.) *Explique por qué la regresión MCO de $test\_after$ en $viewer$ da una estimación inconsistente de los efectos del programa de PBS.* [**1 punto/sobre 15**]

La variable $viewer$ es endógena porque existen variables omitidas que están correlacionadas con ella y por tanto no se cumple el supuesto $E[u|x] = 0$. Una variable omitida obvia es el interés del individuo sobre el problema de la adicción a las drogas, que está obviamente correlacionado con $viewer$ y también con el conocimiento previo acerca del problema de la adicción y, por tanto, con $test\_after$. Este conocimiento previo y el interés sobre el problema que tiene un efecto causal sobre $test\_after$ puede estar relacionado con diferentes variables sociales y culturales.

b.) *Supongamos que queremos deshacernos de la endogeneidad de $viewer$ en la regresión de la pregunta a) mediante la inclusión de más variables explicativas. ¿Qué variables (entre las que se enumeran arriba) podrían utilizarse como variables de control para corregir la endogeneidad en a)? Esto es, considere un modelo del tipo*

$$test\_after = \beta_0 + \beta_1 viewer + \beta_2 W_1 + \cdots + \beta_{1+r} W_r + U, \tag{2}$$

donde $W_1, ..., W_r$ corresponden a las variables de control.

Dada la respuesta en a.), podríamos deshacernos de la posible endogeneidad incluyendo $test\_before$ como variable de control, que mediría el nivel de conocimientos e interés previos. También podríamos controlar por características personales que puedan afectar a que la persona vea el programa PBS (que dependeran de la curiosidad, disponibilidad de tiempo etc.), así como la puntuación de los tests; por ejemplo, $income$, $edu, gender$ y age son buenos candidatos. En principio $groupA$ es asignado aleatoriamente y podemos pensar que no tiene efecto causal sobre $test\_after$ ya que sólo la afectaría a través de $viewer$, la decisión de ver el programa, por lo que no lo incluiríamos.

c.) *Supongamos que el anterior método, incluyendo variables de control, no evita que el término de error esté correlacionado con viewer. Explique por qué razones groupA puede ser un buen instrumento de viewer. Explicar cómo se estimará el modelo (2) por variables instrumentales con el instrumento groupA. Describir las dos etapas del estimador MC2E.* **[1 punto/sobre 15]**

(1/2) Como la asignación a los grupos A y B se ha hecho **aleatoriamente**, la variable $groupA$ no está correlacionada con el término de error, esto es, es exógena. Como los individuos en el grupo A han sido informados del programa PBS, se incrementarán las posibilidades de que vea el programa; por lo tanto $groupA$ está correlacionado positivamente con $viewer$ y es un instrumento relevante. Por lo tanto, $groupA$ es un instrumento válido. (1/2) En la primera etapa regresaría por MCO $viewer$ sobre la VI $groupA$ y las variables de control. Las predicciones $\widehat{viewer}$ de dicha regresión se usarían en la segunda etapa para estimar por MCO la ecuación de interés reemplazando $viewer$ por $\widehat{viewer}$.

d.) *¿Por qué no se debe permitir que los participantes elijan pertenecer a los grupos A o B? Explíquese en términos de las condiciones necesarias para la consistencia del estimador de VI en c.).* **[1 punto/sobre 15]**

Los participantes pueden decidir si reciben o no reciben la información en base a sus restricciones de tiempo, que están correlacionadas con el hecho de ver el programa PBS. En este caso $groupA$ dejará de ser exógena porque afectará a la variable dependiente directamente. $groupA$ estará correlacionada con los factores contenidos en el término de error del modelo relacionados con motivación, interés y conocimiento sobre el tema. Sin embargo, es muy probable que siga correlada con $viewer$.

e.) *¿Qué debería concluirse sobre la relevancia del instrumento groupA teniendo en cuenta los siguientes datos?* **[1 punto/sobre 15]**

Número de personas en grupo A:	510
Número de personas en grupo A que vieron el programa:	244
Número de personas en grupo B:	579
Número de personas en grupo B que vieron el programa:	74

(3/4) El $100 * 244/510 = 47.84\%$ de las personas del grupo A vieron el programa, mientras que sólo el $100 * 74/579 = 12.78\%$ de las personas en el grupo B no lo vieron, por lo que podemos concluir (dado el tamaño de las muestras) que sí existe una correlación (positiva) entre pertener al grupo A ($groupA$) y ver el programa ($viewer$), y por tanto el instrumento es relevante.

(1/4) Para ello podemos plantearnos el error estándard de cada proporción muestral, por ejemplo para el grupo A sería $(0.4784 * (1 - 0.4784) / 510)^{1/2} = 2.212\,0 \times 10^{-2}$, es decir $\hat{p}_A = 0.4784\,(0.022)$, y aun menor para el grupo B, y un contraste formal de $H_0 : p_A = p_B$, que se rechazaría claramente.

f.) *Supongamos que tenemos la siguiente información:*

	Personas en el grupo A	Personas en el Grupo B
Edad Media	55	54
Porcentaje de blancos	76	74
Porcentaje de mujeres	54	52

¿Qué evidencia nos ofrece esta comparación sobre la consistencia del estimador de VI en c.)? **[1 punto/sobre 15]**

Si la asignación a los grupos A y B ha sido aleatoria, debemos observar aproximadamente las mismas características en los dos grupos. Esto es lo que se desprende de la tabla, lo que garantiza que $groupA$ es exógena (incorrelada con cualquier otro factor que determina $test\_after$ aparte de $viewer$), y por tanto un instrumento válido.

i.) *Se obtiene el siguiente estimador de VI*

$$\widehat{test\_after} = 4.09 - \underset{(0.164)}{0.160} \cdot viewer + \cdots$$

Interprete la estimación de la pendiente de viewer. Comente sobre su significatividad. ¿Qué concluiría sobre la eficacia del programa? **[1 punto/sobre 15]**

El coeficiente $-0,160$ quiere decir que la estimación del efecto causal de ver el programa es un descenso del nivel de conocimientos de $0,160$ puntos, ceteris paribus. Sin embargo, contrastando $H_0 : \beta_{viewer} = 0$, en contra de $H_1 : \beta_{viewer} \neq 0$ vemos que $t = -0.160/0.164 \approx -1$, que no es significativo a ningún nivel habitual- por lo que concluiríamos que el programa no es efectivo (en términos de mejorar el nivel general de conocimientos).

.

3. **[3 puntos/sobre 15]** Utilizando una muestra de individuos con estudios secundarios de la Encuesta sobre Estructura Salarial en España para el año 1995, se ha estimado el siguiente modelo para la oferta de trabajo,

$$h = \beta_0 + \beta_1 male + \beta_2 \ln(w) + \beta_3 (male * \ln(w)) + u,$$

donde h es el número de horas trabajadas durante el año, $male$ es una variable binaria que que toma el valor 1 si el individuo es un hombre y 0 si es mujer, y w es el salario-hora en euros. Utilizando una muestra aleatoria se ha ajustado el modelo por MCO con el siguiente resultado (errores estándar robustos en paréntesis):

$$\widehat{h} = \underset{(64,63)}{1429,66} + \underset{(20,68)}{224,9} male + \underset{(6,14)}{91,67} \ln(w) - \underset{(7,04)}{66,41} (male * \ln(w)),$$

a.) *¿Cuál es el efecto medio estimado sobre las horas trabajadas provocado por un aumento del 1% en el salario para la población de hombres y para la población de mujeres respectivamente? Contraste si ambos efectos son diferentes.* **[1 punto/sobre 15]**

(1/2) El efecto medio de un aumento del 1% del salario es $(91,67 - 66,41)/100 = 0,2526$ horas trabajadas para los hombres y $91,67/100 = 0,9167$ para las mujeres. (1/2) Para contrastar si son diferentes es necesario contrastar si el coeficiente de $male * \ln(w)$ es significativo,

$$H_0 : \beta_3 = 0.$$

Usando el estadístico t, $t = -66.41/7.04 = -9.4332$ vemos que es muy significativo y por tanto los efectos son estadísticamente diferentes.

b.) *Teniendo en cuenta que la media muestral de* $\ln(w)$ *es* $\overline{\ln(w)} = 3{,}38$ *para hombres y* $\overline{\ln(w)} = 2{,}48$ *para mujeres, proporcione los valores estimados de las horas medias trabajadas para hombres y para mujeres a partir del ajuste reportado.* **[1 punto/sobre 15]**

Para hombres tendríamos

$$\widehat{h} = 1429.66 + 224.9 + (91.67 - 66.41) * 3.38 = 1739$$

mientras que para mujeres

$$\widehat{h} = 1429.66 + 91.67 * 2.48 = 1657.$$

c.) *Explicar cómo contrastaría que las funciones de regresión son iguales para hombres y para mujeres. Razonar porque en este caso puede ser necesario usar errores estandar robustos a la heterocedasticidad.* **[1 punto/sobre 15]**

(1/2) Hay que contrastar la significatividad conjunta de todas las variables que incluyen la variable binaria $male$, ($male$ y $male * \ln(w)$),

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

en contra de la alternativa de que algún coeficiente es diferente de zero, mediante un contraste robusto de la F. (1/2) En este caso podemos pensar que la variabilidad de la oferta de horas de trabajo es diferente para hombres y mujeres, y por tanto $Var(u/x)$ depende de los regresores x.