SOLUCIÓN Examen de Introducción a la Econometría
Universidad Carlos III de Madrid

1ª Convocatoria
 Curso 2006/2007

Conteste las preguntas siguientes en cuadernillos separados en 2 horas

1. [3 puntos] *Suponga que $\log(\text{wage})$ mide el logaritmo del salario mensual, educ el número de años de educación, y abil el cociente de inteligencia (IQ) y que el modelo*

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{abil} + u$$

satisface el supuesto $E(u|\text{educ}, \text{abil}) = 0$.

- (a) *Interprete el coeficiente β_1 . Debido a que la muestra de la que inicialmente se dispone no contiene información sobre el cociente de inteligencia de los trabajadores, se estima el modelo de regresión simple*

$$\begin{aligned} \widehat{\log(\text{wage})} &= 5.97 + \underset{(.006)}{.06} \text{educ} \\ n &= 935, R^2 = 0.097, SSR = 149.52 \end{aligned}$$

Explique bajo qué condiciones la estimación OLS del parámetro de la variable educ es un estimador insesgado de β_1 . Proporcione un intervalo de confianza al 95% para la pendiente del modelo estimado.

β_1 refleja la variación relativa esperada en el salario al aumentar en un año la educación, manteniendo fijo el nivel de inteligencia. Multiplicando por 100 obtenemos la variación porcentual.

El estimador es un estimador insesgado de $\frac{\text{Cov}(\log(\text{wage}), \text{educ})}{\text{Var}(\text{educ})}$, que verifica

$$\frac{\text{Cov}(\log(\text{wage}), \text{educ})}{\text{Var}(\text{educ})} = \beta_1 + \beta_2 \frac{\text{Cov}(\text{educ}, \text{abil})}{\text{Var}(\text{educ})}$$

Por consiguiente, será un estimador insesgado de β_1 si $\beta_2 = 0$ y/o $\text{Cov}(\text{educ}, \text{abil}) = 0$.

Para obtener un intervalo de confianza necesitamos estimar el valor mínimo y el valor máximo del intervalo. El intervalo de confianza del 95% con una muestra tan grande puede obtenerse con los valores críticos de la normal estándar:

$$\begin{aligned} \hat{l} &= 0.06 - 1.96 * 0.006 \simeq 0.048 \\ \hat{u} &= 0.06 + 1.96 * 0.006 \simeq 0.072 \end{aligned}$$

- (b) *Suponga ahora que se consigue información sobre el cociente de inteligencia de los trabajadores de la muestra, obteniéndose la siguiente estimación,*

$$\begin{aligned} \widehat{\log(\text{wage})} &= 5.66 + \underset{(.007)}{.04} \text{educ} + \underset{se(\hat{\beta}_2)}{.0059} \text{abil} \\ n &= 935, R^2 = 0.130 \end{aligned}$$

Contraste al nivel de significatividad del 5% que la inteligencia no afecta el salario.

La hipótesis nula es: $H_0 : \beta_2 = 0$ mientras que la alternativa es $H_1 : \beta_2 \neq 0$. El resultado de la estimación del apartado anterior corresponde con el resultado bajo la hipótesis nula mientras que el presentado en este apartado corresponde con el modelo no restringido. Además, no hay cambios en la definición de la variable dependiente. Por ello, podemos utilizar el estadístico F expresado con los R^2 correspondientes a ambos modelos:

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} * \frac{n - 3}{1} \simeq 35.35$$

que es mayor que el valor crítico al nivel de significatividad pedido, tanto utilizando la distribución chi-cuadrado como la F .

- (c) *Obtenga la covarianza muestral entre el nivel educativo y el cociente de inteligencia. A la vista de los resultados, interprete el parámetro asociado a la educación en las dos regresiones.*

Para obtener la covarianza muestral, basta con considerar la expresión para la muestra

$$\frac{Cov(\widehat{\log(wage)}, educ)}{Var(educ)} = \widehat{\beta}_1 + \widehat{\beta}_2 \frac{Cov(\widehat{educ}, abil)}{Var(\widehat{educ})}$$

sustituyendo por los datos directamente dados en el enunciado nos queda:

$$0.06 = 0.04 + 0.0059 \frac{Cov(\widehat{educ}, abil)}{Var(\widehat{educ})}$$

Para obtener la covarianza muestral, necesitamos la varianza en la muestra de *educ*, que se puede obtener a través del valor de la desviación típica del parámetro de la pendiente del modelo “corto”:

$$0.006^2 = \frac{SSR/(n - k - 1)}{Var(educ) * n}$$

Por tanto $Var(\widehat{educ}) = \frac{(149.52/933)}{0.006^2 * 935} \simeq 4.76$ y $Cov(\widehat{educ}, abil) = (1/0.0059) * 4.76 * (0.06 - 0.04) \simeq 16.13$.

En la ecuación larga obtenemos un estimador insesgado de β_1 , cuya interpretación ya hemos dado en el primer apartado. El parámetro estimado en la ecuación corta refleja la variación relativa esperada en el salario al aumentar en un año la educación y el nivel de inteligencia, éste por la cuantía esperada al aumentar la educación en un año. Multiplicando por 100 obtenemos la variación porcentual.

2. [4 puntos] *Para explicar el salario de un director general, salary, se ha estimado la siguiente ecuación:*

$$\begin{aligned} \log(\widehat{salary}) &= 4.362 + 0.275 \log(sales) + 0.0179roe \\ &\quad (0.294) \quad (0.033) \quad (0.0039) \\ n &= 209, \quad \widehat{R}^2 = 0.282 \end{aligned}$$

donde *sales* son las ventas anuales y *roe* es el rendimiento sobre el valor nominal de una acción (*return on equity*).

- (a) *Interpretar el coeficiente de $\log(sales)$ y contrastar si es significativamente positivo.*

Si las ventas (*sales*) aumentan en un 1%, entonces el salario (*salary*) aumenta en un 0.275%.

Para el contraste

$$\begin{aligned} H_0 &: \beta_{\log(sales)} = 0 \\ H_0 &: \beta_{\log(sales)} > 0 \end{aligned}$$

usamos el estadístico *t*,

$$t_{\log(sales)} = \frac{\widehat{\beta}_{\log(sales)}}{se(\widehat{\beta}_{\log(sales)})} = \frac{0.275}{0.033} = 8.33$$

que es positivo y muy significativo: el coeficiente es signif. positivo.

- (b) *Se decide incluir una variable dummy, rosneg, que es igual a 1 cuando ros es negativa y 0 si ros es cero o positivo, donde ros es el rendimiento sobre el valor real de la acción,*

$$\begin{aligned} \log(\widehat{salary}) &= 4.074 + 0.314 \log(sales) + 0.017roe \\ &\quad (0.307) \quad (0.035) \quad (0.004) \\ &\quad + 2.094rosneg - 0.258 \log(sales) * rosneg - 0.00353roe * rosneg \\ &\quad (1.009) \quad (0.112) \quad (0.178) \\ n &= 209, \quad \widehat{R}^2 = 0.315, \end{aligned}$$

obteniéndose las siguientes varianzas estimadas para los coeficientes de $\log(\text{sales})$, roe , rosneg , $\log(\text{sales}) * \text{rosneg}$ y $\text{roe} * \text{rosneg}$, respectivamente,

$$\widehat{\text{Var}} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{pmatrix} = \begin{pmatrix} 0.001236 & 1.47E-05 & 0.010414 & -0.001236 & -1.47E-05 \\ 1.47E-05 & 1.64E-05 & 0.000410 & -1.47E-05 & -1.64E-05 \\ 0.010414 & 0.000410 & 1.018975 & -0.109247 & -0.003193 \\ -0.001236 & -1.47E-05 & -0.109247 & 0.012544 & -0.000116 \\ -1.47E-05 & -1.64E-05 & -0.003193 & -0.000116 & 0.000318 \end{pmatrix} ..$$

Contraste si es necesario distinguir en el modelo las empresas en función del signo de ros . Debemos contrastar si los coeficientes de las variables que usan la variable ficticia son conjuntamente zero o no,

$$H_0 : \beta_{\text{rosneg}} = \beta_{\log(\text{sales}) * \text{rosneg}} = \beta_{\text{roe} * \text{rosneg}} = 0$$

en contra de que es falsa. El modelo restringido es el del apartado (a) por lo que

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} \frac{n - k - 1}{q} = \frac{0.315 - 0.282}{1 - 0.315} \frac{209 - 5 - 1}{3} = 3.26$$

que es significativo comparado con una $F_{3,203}$ y con una χ_3^2 al 5% (pero no al 1%) por lo si hay cierta evidencia en favor de la necesidad de distinguir en función del ros .

- (c) *Contraste si para las empresas que tienen ros negativo, un aumento de las ventas implica necesariamente un aumento del salario del director general, todos los demás factores constantes.*

Si $\text{rosneg} = 1$, entonces

$$\Delta E [\log(\text{salary}) | \text{rosneg} = 1] / \Delta \log(\text{sales}) = \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}}$$

y por tanto

$$\Delta \log(\widehat{\text{salary}}) / \Delta \log(\text{sales}) = 0.314 - 0.258 = 0.056$$

y para contrastar si ese efecto es positivo,

$$\begin{aligned} H_0 & : \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}} = 0 \\ H_1 & : \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}} > 0 \end{aligned}$$

usamos

$$t_{\beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}}} = \frac{\hat{\beta}_{\log(\text{sales})} + \hat{\beta}_{\log(\text{sales}) * \text{rosneg}}}{\text{se}(\hat{\beta}_{\log(\text{sales})} + \hat{\beta}_{\log(\text{sales}) * \text{rosneg}})} = \frac{0.056}{0.10632} = 0.52671$$

donde

$$\begin{aligned} & \text{se}(\hat{\beta}_{\log(\text{sales})} + \hat{\beta}_{\log(\text{sales}) * \text{rosneg}}) \\ &= \left(\widehat{\text{Var}}(\hat{\beta}_{\log(\text{sales})}) + \widehat{\text{Var}}(\hat{\beta}_{\log(\text{sales}) * \text{rosneg}}) + 2\widehat{\text{Cov}}(\hat{\beta}_{\log(\text{sales})}, \hat{\beta}_{\log(\text{sales}) * \text{rosneg}}) \right)^{1/2} \\ &= (0.001236 + 0.01254 + 2 * -0.001236)^{1/2} \\ &= 0.10632. \end{aligned}$$

El estadístico t no es significativo a los niveles habituales, por lo que no se puede concluir que un aumento de las ventas necesariamente conlleve un aumento del salario en esas circunstancias.

- (d) *Explique cómo contrastaría la hipótesis de que para un director general, de una empresa con ros negativo, $\log(\text{sales}) = 10$ y $\text{roe} = 20$, es igual de interesante (a) aumentar las ventas en un 1%, o (b) conseguir que el ros pase a ser positivo (en ambos casos sin cambiar las otras variables).*

Tenemos que

$$\begin{aligned} (a) & \rightarrow \Delta E[\log(\text{salary}) | \text{rosneg} = 1] = \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}} \\ (b) & \rightarrow E[\log(\text{salary}) | \text{rosneg} = 0, \log(\text{sales}) = 10, \text{roe} = 20] - E[\log(\text{salary}) | \text{rosneg} = 1, \log(\text{sales}) = 10, \text{roe} = 20] \\ & = -\left(\beta_{\text{rosneg}} + 10 * \beta_{\log(\text{sales}) * \text{rosneg}} + 20 * \beta_{\text{roe} * \text{rosneg}}\right) \end{aligned}$$

y la hipótesis a contrastar sería

$$H_0 : \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}} = -\left(\beta_{\text{rosneg}} + 10 * \beta_{\log(\text{sales}) * \text{rosneg}} + 20 * \beta_{\text{roe} * \text{rosneg}}\right)$$

Habría que utilizar el modelo restringido imponiendo esa restricción y comparando su R^2 o SCE con el modelo sin restringir mediante un contraste de la F con $q = 1$. Alternativamente podemos definir

$$\begin{aligned} \theta & = \beta_{\log(\text{sales})} + \beta_{\log(\text{sales}) * \text{rosneg}} + \beta_{\text{rosneg}} + 10 * \beta_{\log(\text{sales}) * \text{rosneg}} + 20 * \beta_{\text{roe} * \text{rosneg}} \\ & = \beta_{\log(\text{sales})} + \beta_{\text{rosneg}} + 11\beta_{\log(\text{sales}) * \text{rosneg}} + 20 * \beta_{\text{roe} * \text{rosneg}} \end{aligned}$$

y estimarlo directamente y calcular su estadístico para contrastar

$$H_0 : \theta = 0.$$

3. [3 puntos] *Las siguientes ecuaciones de salarios han sido estimadas utilizando datos correspondientes a trabajadores en Bangladesh:*

$$\log(\widehat{\text{salario}}) = \underset{(0,35)}{1,25} + \underset{(0,03)}{0,15\text{hombre}} + \underset{(0,004)}{0,02\text{ experiencia}} \quad (1)$$

$$\log(\widehat{\text{salario}}) = \underset{(0,48)}{1,55} + \underset{(0,05)}{0,10\text{hombre}} + \underset{(0,005)}{0,015\text{ experiencia}} - \underset{(0,002)}{0,005\text{ hombre} * \text{ experiencia}}, \quad (2)$$

donde el salario está medido en US \$, y las variables explicativas son una ficticia para los hombres y los años de experiencia. Los números en paréntesis son los errores estándar.

- (a) *¿Cuál es la diferencia media estimada entre el salario de un hombre con 5 años de experiencia y una mujer con 10 años de experiencia utilizando la ecuación (1)?*

Diferencia:

$$\begin{aligned} & \log(\widehat{\text{salario}} | h = 1, \text{exp} = 5) - \log(\widehat{\text{salario}} | h = 0, \text{exp} = 10) \\ & = 0,15 + 0,02 * 5 - (0,02 * 10) \\ & = 0,05, \end{aligned}$$

es decir un hombre de 5 años de experiencia tendrá un salario aproximadamente un 5% superior en media que una mujer con 10 años de experiencia, ceteris paribus.

Diferencia exacta:

$$(\exp(0,15 + 0,02 * 5 - (0,02 * 10)) - 1) 100\% = (\exp(0,05) - 1) 100 = 5.13\%$$

- (b) *¿Cuál es la diferencia media estimada entre el salario de un hombre con 5 años de experiencia y una mujer con 10 años de experiencia utilizando la ecuación (2)?*

Diferencia:

$$\begin{aligned} & \log(\widehat{\text{salario}} | h = 1, \text{exp} = 5) - \log(\widehat{\text{salario}} | h = 0, \text{exp} = 10) \\ & = 0,10 + 0,015 * 5 - 0,005 * 5 - (0,015 * 10) \\ & = 0 \end{aligned}$$

es decir un hombre de 5 años de experiencia tendrá un salario medio igual que una mujer con 10 años de experiencia, ceteris paribus.

(c) *Contrasta que la diferencia salarial entre hombres y mujeres no depende de la experiencia.*

Test

$$H_0 : \beta_{\text{hom bre*experiencia}} = 0$$

$$H_1 : \beta_{\text{hom bre*experiencia}} \neq 0$$

donde usamos que

$$t_{\text{hom bre*experiencia}} = \frac{\hat{\beta}_{\text{hom bre*experiencia}}}{se(\hat{\beta}_{\text{hom bre*experiencia}})} = \frac{0,005}{0,002} = 2,5$$

que suponiendo que n es grande y aproximando la distribución del estadístico t por una $N(0, 1) = t_\infty$ es significativo al 5 (y al 2%) pero no al 1%.