Soluciones Examen Final de Econometría

Universidad Carlos III de Madrid

26 de Mayo de 2015

Conteste todas las preguntas en dos horas y media.

Pregunta 1 (33 puntos): *Un investigador está considerando las siguientes dos especificaciones alternativas para estimar la relación entre una variable X y una variable Y,*

$$\log Y = \beta_1 + \beta_2 \log X + U \tag{1}$$

$$\log \frac{Y}{X} = \alpha_1 + \alpha_2 \log X + V, \tag{2}$$

donde las letras griegas se refieren a parámetros y X e Y son dos variables para las que disponemos de una muestra aleatoria de tamaño n.

a. (4 puntos) *Determine si (2) puede expresarse como una versión restringida del modelo (1).*

(2) we puede reescribir como

$$\log Y = \alpha_1 + (\alpha_2 + 1) \log X + V$$

que es una versión reparametrizada de (1) con $\beta_1 = \alpha_1$ y $\beta_2 = \alpha_2 + 1$, y por tanto NO una versión restringida, ya que no impone ninguna restricción en los valores de β_1 o β_2.

b. (7 puntos) *Utilizando las mismas n observaciones de las variables Y y X, se ajustan los dos modelos por mínimos cuadrados ordinarios (MCO). Los ajustes son*

$$\widehat{\log Y} = \hat{\beta}_1 + \hat{\beta}_2 \log X \tag{3}$$

$$\widehat{\log \frac{Y}{X}} = \hat{\alpha}_1 + \hat{\alpha}_2 \log X, \tag{4}$$

donde las letras griegas con tilde son los valores estimados por mínimos cuadrados ordinarios. Utilizando las expresiones para los estimadores MCO, exprese $\hat{\beta}_2$ en términos de $\hat{\alpha}_2$.

Definiendo

$$\begin{aligned} y &= \log Y \\ x &= \log X \\ z &= \log \frac{Y}{X} = y - x, \end{aligned}$$

se obtiene

$$\begin{aligned} \hat{\alpha}_2 &= \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(y_i - x_i - [\bar{y} - \bar{x}])}{\sum (x_i - \bar{x})^2} \\ &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} - \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \hat{\beta}_2 - 1. \end{aligned}$$

c. (4 puntos) *Exprese $\hat{\beta}_1$ en términos de $\hat{\alpha}_1$.*

$$\begin{aligned}\hat{\alpha}_1 &= \bar{z} - \hat{\alpha}_2 \bar{x} = (\bar{y} - \bar{x}) - \left(\hat{\beta}_2 - 1\right)\bar{x} \\ &= \bar{y} - \bar{x} - \hat{\beta}_2 \bar{x} + \bar{x} = \bar{y} - \hat{\beta}_2 \bar{x} = \hat{\beta}_1\end{aligned}$$

d. (5 puntos) *Demostrar que*

$$\widehat{\log Y} - \log X = \widehat{\log \frac{Y}{X}}.$$

La expresión es equivalente a $\hat{y} - x = \hat{z}$, donde

$$\begin{aligned}\hat{z} &= \hat{\alpha}_1 + \hat{\alpha}_2 x \\ &= \hat{\beta}_1 + \left(\hat{\beta}_2 - 1\right) x \\ &= \hat{\beta}_1 + \hat{\beta}_2 x - x \\ &= \hat{y} - x.\end{aligned}$$

e. (3 puntos) *Demuestre que los residuos de (3) son idénticos a los de (4).*

$$\begin{aligned}\hat{V} &= \hat{z} - z \\ &= \hat{\alpha}_1 + \hat{\alpha}_2 x - (y - x) \\ &= \hat{\beta}_1 + \left(\hat{\beta}_2 - 1\right) x - (y - x) \\ &= \hat{\beta}_1 + \hat{\beta}_2 x - y \\ &= \hat{y} - y = \hat{U}.\end{aligned}$$

f. (3 puntos) *Demuestre que los errores estándar de $\hat{\beta}_2$ y $\hat{\alpha}_2$ son idénticos.*

$$s.e.\left(\hat{\beta}_2\right) = \sqrt{\frac{\sum \hat{U}_i^2/(n-2)}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum \hat{V}_i^2/(n-2)}{\sum (x_i - \bar{x})^2}} = s.e.\left(\hat{\alpha}_2\right)$$

g. (4 puntos) *Determine la relación entre el estadístico t utilizando $\hat{\beta}_2$ y el estadístico t utilizando $\hat{\alpha}_2$.*

El estadístico t para $H_0 : \beta_2 = 0$ es

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{s.e.\left(\hat{\beta}_2\right)} = \frac{\hat{\alpha}_2 + 1}{s.e.\left(\hat{\alpha}_2\right)} = t_{\hat{\alpha}_2 + 1}$$

que corresponde al estadístico t para la hipótesis equivalente $H_0 : \alpha_2 + 1 = 0$.

h. (3 puntos) *Explique de forma razonada si el R^2 sería el mismo en los dos ajustes.*

Los R^2 serán diferentes porque el R^2 mide la proporción de la varianza de la variable dependiente explicada por la regresión, y las variables dependientes son diferentes en cada regresión.

PREGUNTA 2 (33 puntos): *Se está llevando a cabo un estudio econométrico sobre la diferencia entre los salarios de hombres y mujeres controlando por las variables determinantes del salario. La muestra consiste en 526 empleados, de los cuales 252 son mujeres. La muestra contiene información sobre las siguientes variables.*

$$\begin{array}{lcl} W & = & \text{Salario en dólares por hora.} \\ S & = & \text{Número de años de educación formal completados.} \\ A & = & \text{Edad en años.} \\ T & = & \text{Antigüedad en la empresa en años.} \\ F & = & \text{1 si el individuo es mujer y 0 en otro caso.} \end{array}$$

Se propone el siguiente modelo

$$\begin{aligned} \ln W &= \beta_1 + \beta_2 S + \beta_3 A + \beta_4 A^2 + \beta_5 T + \beta_6 (S \cdot T) + \beta_7 F + \beta_8 (F \cdot S) \\ &\quad + \beta_9 (F \cdot A) + \beta_{10} (F \cdot A^2) + \beta_{11} (F \cdot T) + \beta_{12} (F \cdot S \cdot T) + U, \end{aligned} \tag{5}$$

donde las letras griegas denotan parámetros desconocidos, y U es el término de error. Se supone que se satisfacen todos los supuestos clásicos del modelo de regresión múltiple, incluyendo el de homocedasticidad. La muestra observada proporciona el siguiente ajuste (errores estándar de cada estimador en paréntesis),

$$\begin{aligned} \widehat{\ln W} &= \underset{(0.2385)}{-0.5667} + \underset{(0.01104)}{0.05937}\, S + \underset{(0.01216)}{0.07980}\, A - \underset{(0.00015)}{0.00093}\, A^2 - \underset{(0.01128)}{0.01057}\, T \\ &\quad + \underset{(0.00087)}{0.00227}\,(S \cdot T) + \underset{(0.3373)}{0.03593}\, F + \underset{(0.01684)}{0.01684}(F \cdot S) - \underset{(0.01715)}{0.03847}\,(F \cdot A) \\ &\quad + \underset{(0.000219)}{0.000422}(F \cdot A^2) + \underset{(0.02652)}{0.01850}(F \cdot T) - \underset{(0.002187)}{0.002107}(F \cdot S \cdot T). \end{aligned} \tag{6}$$

La correspondiente suma de los residuos al cuadrado (SR) y la suma total de cuadrados (ST) de las n=526 observaciones son

$$SR = \sum_{i=1}^{n} \hat{U}_i^2 = 80.57 \text{ y } ST = \sum_{i=1}^{n} (\ln W_i - \overline{\ln W})^2 = 148.33.$$

a. (6 puntos) *Utilice el ajuste para obtener los estimadores de mínimos cuadrados ordinarios (MCO) de los coeficientes de S, A y T cuando solamente se utilizan las 274 observaciones correspondientes a los hombres. Haga lo mismo cuando solo se utilizan las 252 observaciones correspondientes a las mujeres.*

Para hombres ($F = 0$) obtenemos

$$\begin{aligned} \widehat{\ln W} &= \underset{(0.2385)}{-0.5667} + \underset{(0.01104)}{0.05937}\, S + \underset{(0.01216)}{0.07980}\, A - \underset{(0.00015)}{0.00093}\, A^2 - \underset{(0.01128)}{0.01057}\, T \\ &\quad + \underset{(0.00087)}{0.00227}\,(S \cdot T) \end{aligned}$$

y los coeficientes son $\hat{\beta}_2 = 0.05937$, $\hat{\beta}_3 = 0.07980, \hat{\beta}_5 = -0.01057$, respectivamente, y para mujeres ($F=1$)

$$\begin{aligned}
\widehat{\ln W} &= \Big(\underset{(0.2385)}{-0.5667} + \underset{(0.3373)}{0.03593}\Big) + \Big(\underset{(0.01104)}{0.05937} + \underset{(0.01684)}{0.01684}\Big)\ S + \Big(\underset{(0.01216)}{0.07980} - \underset{(0.01715)}{0.03847}\Big)\ A \\
&+ \Big(\underset{(0.00015)}{-0.00093} + \underset{(0.000219)}{0.000422}\Big)\ A^2 + \Big(\underset{(0.01128)}{-0.01057} + \underset{(0.02652)}{0.01850}\Big)\ T \\
&+ \Big(\underset{(0.00087)}{0.00227} - \underset{(0.002187)}{0.002107}\Big)\, S \cdot T
\end{aligned}$$

por lo que los coeficientes son, respectivamente,

$$\begin{aligned}
\hat{\beta}_2 + \hat{\beta}_8 &= 0.05937 + 0.01684 = 0.07621\ , \\
\hat{\beta}_3 + \hat{\beta}_9 &= 0.07980 - 0.03847 = 0.041\,33, \\
\hat{\beta}_5 + \hat{\beta}_{11} &= -0.01057 + 0.01850 = 0.00793.
\end{aligned}$$

b. (9 puntos) *Proporcione una expresión (fórmula) del efecto parcial de A sobre $\ln W$ para los hombres, en términos de los parámetros desconocidos, implicada por la ecuación (5). Haga lo mismo con las mujeres. Interprete las dos expresiones. Proporcione una expresión para la hipótesis nula (H_0) de que los efectos parciales de A sobre $\ln W$ para hombres y mujeres son idénticos. Proporcione también la expresión para la hipótesis alternativa (H_1). Exprese la ecuación de regresión restringida implicada por H_0. La estimación por MCO de la ecuación restringida arroja un $SR = 81.7242$. Utilice esta información, junto con la que se ha proporcionado en el enunciado del problema, para calcular un estadístico del contraste, indicando cuál es su distribución aproximada bajo la hipótesis nula. Establezca una regla de decisión, utilizando un nivel de significación del 5%, y la correspondiente conclusión sobre la inferencia del contraste.*

Los efectos parciales de A son

$$\begin{aligned}
\frac{\partial E\left[\ln W | S, A, T, male\right]}{\partial A} &= \beta_3 + 2\beta_4 A \\
\frac{\partial E\left[\ln W | S, A, T, female\right]}{\partial A} &= \beta_3 + \beta_9 + 2\left(\beta_4 + \beta_{10}\right) A
\end{aligned}$$

que se interpretan de forma que $100\left(\beta_2 + 2\beta_4 A\right)\%$ es el porcentaje esperado de cambio en W cuando A cambia en una unidad, todo lo demás igual.

Las hipótesis son

$$\begin{aligned}
H_0 &: \ \beta_9 = \beta_{10} = 0 \\
H_1 &: \ H_0 \text{ es falsa (o bien } \beta_9 \text{ o } \beta_{10} \text{ o los dos son } \neq 0).
\end{aligned}$$

La regresión restringida es

$$\begin{aligned}
\ln W &= \beta_1 + \beta_2 S + \beta_3 A + \beta_4 A^2 + \beta_5 T + \beta_6\left(S \cdot T\right) + \beta_7 F + \beta_8 (F \cdot S) \\
&\quad + \beta_{11}(F \cdot T) + \beta_{12}(F \cdot S \cdot T) + U.
\end{aligned}$$

El estadístico de contraste es un estadístico de la F bajo homocedasticidad para $q = 2$ restricciones

$$\begin{aligned} F &= \frac{RSS_{res} - RSS_{nor}}{RSS_{nor}} \frac{n-k-1}{q} \\ &= \frac{81.7242 - 80.57}{80.57} \frac{526 - 11 - 1}{2} = 3.6816 \end{aligned}$$

cuya distribución asintótica bajo la nula es $\chi^2_2/2$, la cual proporciona un valor crítico al nivel de significación del 5% igual a $cv_{0.05} = 5.99/2 \approx 3$. Por tanto, como $F > cv_{0.05}$, rechazamos la hipótesis nula en favor de la conclusión de que los efectos marginales de A son diferentes para hombres y mujeres.

c. (9 puntos) *Exprese la restricción sobre los parámetros en (5) de que los efectos parciales de S sobre $\ln W$ y de T sobre $\ln W$ son iguales para mujeres y hombres. Proporcione una expresión de la ecuación (5) imponiendo estas dos restricciones. La estimación por MCO de esta ecuación restringida proporciona un $SR = 80.8747$. Utilice esta información, junto con la que se ha proporcionado en el enunciado del problema, para calcular un estadístico del contraste, indicando cuál es su distribución aproximada bajo la hipótesis nula. Establezca una regla de decisión, utilizando un nivel de significación del 5%, y la correspondiente conclusión sobre la inferencia del contraste.*

Los efectos parciales de S son

$$\begin{aligned} \frac{\partial E\left[\ln W | S, A, T, male\right]}{\partial S} &= \beta_2 + \beta_6 T \\ \frac{\partial E\left[\ln W | S, A, T, female\right]}{\partial S} &= \beta_2 + \beta_8 + \left(\beta_6 + \beta_{12}\right) T \end{aligned}$$

que se interpretan de forma que $100\left(\beta_2 + \beta_6 T\right)\%$ es el cambio porcentual esperado en W cuando S cambia en una unidad, todo lo demás igual, y los efectos parciales de T son

$$\begin{aligned} \frac{\partial E\left[\ln W | S, A, T, male\right]}{\partial T} &= \beta_5 + \beta_6 S \\ \frac{\partial E\left[\ln W | S, A, T, female\right]}{\partial T} &= \beta_5 + \beta_{11} + \left(\beta_6 + \beta_{12}\right) S \end{aligned}$$

que se interpretan de forma que $100\left(\beta_2 + \beta_6 S\right)\%$ es el cambio porcentual esperado en W cuando T cambia en una unidad, todo lo demás igual.

Para contrastar

$$\begin{aligned} \frac{\partial E\left[\ln W | S, A, T, male\right]}{\partial S} &= \frac{\partial E\left[\ln W | S, A, T, female\right]}{\partial S} \\ &\quad y \\ \frac{\partial E\left[\ln W | S, A, T, male\right]}{\partial T} &= \frac{\partial E\left[\ln W | S, A, T, female\right]}{\partial T} \end{aligned}$$

las hipótesis son

$$
\begin{aligned}
H_0 &: \beta_8 = \beta_{12} = \beta_{11} = 0 \\
H_1 &: H_0 \text{ es falsa (al menos uno de } \beta_8, \beta_{12}, \beta_{11} \text{ son } \neq 0).
\end{aligned}
$$

y la regresión restringida es

$$
\begin{aligned}
\ln W &= \beta_1 + \beta_2 S + \beta_3 A + \beta_4 A^2 + \beta_5 T + \beta_6 (S \cdot T) + \beta_7 F \\
&\quad + \beta_9 (F \cdot A) + \beta_{10} (F \cdot A^2) + U.
\end{aligned}
$$

El estadístico de contraste es un estadístico de la F bajo homocedasticidad para $q = 3$ restricciones

$$
\begin{aligned}
F &= \frac{RSS_{res} - RSS_{nor}}{RSS_{nor}} \frac{n-k-1}{q} \\
&= \frac{80.8747 - 80.57}{80.57} \frac{526 - 11 - 1}{3} = 0.648
\end{aligned}
$$

cuya distribución asintótica bajo la nula es $\chi^2_3/3$, la cual proporciona un valor crítico al nivel de significación del 5% igual a $cv_{0.05} = 7.815/3 \approx 2.6$. Por tanto, como $F < cv_{0.05}$, no podemos rechazar la hipótesis nula de que los efectos marginales son iguales para hombres y mujeres al nivel designificación del 5%.

d. (9 puntos) *Exprese la restricción sobre los parámetros en (5) bajo la cual las ecuaciones de regresión de hombres y mujeres son idénticas; esto es, una restricción bajo la cual la media del logaritmo del salario de las mujeres y la de los hombres para unos valores de S, A y T cualesquiera son idénticas. Proporcione una expresión para la hipótesis nula (H_0) de la restricción mencionada y para la alternativa (H_1). Escriba la ecuación (5) imponiendo la restricción en H_0. La estimación de esta ecuación reducida produce un $SR = 93.1805$. Utilice esta información, junto con la que se ha proporcionado en el enunciado del problema, para calcular un estadístico del contraste, indicando cuál es su distribución aproximada bajo la hipótesis nula. Establezca una regla de decisión, utilizando un nivel de significación del 5%, y la correspondiente conclusión sobre la inferencia del contraste.*

Las hipótesis son

$$
\begin{aligned}
H_0 &: \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0 \\
H_1 &: H_0 \text{ es falsa (al menos uno de } \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} \text{ es} \neq 0).
\end{aligned}
$$

y la regresión restringida es

$$
\ln W = \beta_1 + \beta_2 S + \beta_3 A + \beta_4 A^2 + \beta_5 T + \beta_6 (S \cdot T) + U.
$$

El estadístico de contraste es un estadístico de la F bajo homocedasticidad para $q = 6$ restricciones

$$
\begin{aligned}
F &= \frac{RSS_{res} - RSS_{nor}}{RSS_{nor}} \frac{n-k-1}{q} \\
&= \frac{93.1805 - 80.57}{80.57} \frac{526 - 11 - 1}{6} = 13.408,
\end{aligned}
$$

cuya distribución asintótica bajo la nula es $\chi^2_6/6$, la cual proporciona un valor crítico al nivel de significación del 5% igual a $cv_{0.05} = 12.592/6 = 2.099$. Por tanto, como $F > cv_{0.05}$, rechazamos la hipótesis nula de que las funciones de regresión son iguales para hombres y mujeres, por lo que concluimos que deben de diferir en al menos un coeficiente.

PREGUNTA 3 (34 puntos): *Considere un modelo econométrico donde dos variables X e Y se determinan conjuntamente en el modelo*

$$\begin{aligned} Y &= \beta_1 + \beta_2 X + \beta_3 Z + U \qquad (7) \\ X &= \alpha_1 + \alpha_2 Y + V, \qquad (8) \end{aligned}$$

Las letras griegas denotan parámetros desconocidos, U y V son los errores del modelo con media cero, incorrelados entre ellos, y Z es una variable exógena, independiente de los errores.

a. (4 puntos) *Explique, de forma razonada, si β_2 y α_2 están identificados bajo la información proporcionada en el enunciado.*

β_2 no está identificada porque no hay instrumentos disponibles para la primera ecuación (es decir, no hay regresores exógenos omitidos en (7) que aparezcan en (8)).

α_2 está identificada por que podemos usar Z como instrumento en la ecuación (8), ya que es independiente de V (exógeno) y correlado (relevante) con Y si $\beta_3 \neq 0$.

b. (12 puntos) *Demostrar que el estimador MCO de α_2 en la ecuación (8) es en general asintóticamente sesgado. ¿Cuál sería la conclusión si $\beta_2 = 0$? ¿Cuál sería su conclusión si $\alpha_2\beta_2 = 1$?*

Tenemos que

$$\hat{\alpha}_2^{MCO} = \frac{\sum \left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)}{\sum \left(Y_i - \bar{Y}\right)^2} = \frac{\sum \left(Y_i - \bar{Y}\right) X_i}{\sum \left(Y_i - \bar{Y}\right)^2}$$

y sustituyendo (8) encontramos que

$$\begin{aligned} \hat{\alpha}_2^{MCO} &= \frac{\sum \left(Y_i - \bar{Y}\right)(\alpha_1 + \alpha_2 Y_i + V_i)}{\sum \left(Y_i - \bar{Y}\right)^2} = \alpha_2 \frac{\sum \left(Y_i - \bar{Y}\right) Y_i}{\sum \left(Y_i - \bar{Y}\right)^2} + \frac{\sum \left(Y_i - \bar{Y}\right) V_i}{\sum \left(Y_i - \bar{Y}\right)^2} \\ &= \alpha_2 + \frac{\sum \left(Y_i - \bar{Y}\right) V_i}{\sum \left(Y_i - \bar{Y}\right)^2} \end{aligned}$$

por lo que

$$\begin{aligned} p\lim \hat{\alpha}_2^{MCO} &= \alpha_2 + p\lim \frac{\sum \left(Y_i - \bar{Y}\right) V_i}{\sum \left(Y_i - \bar{Y}\right)^2} \\ &= \alpha_2 + \frac{p\lim \frac{1}{n} \sum \left(Y_i - \bar{Y}\right)\left(V_i - \bar{V}\right)}{p\lim \frac{1}{n} \sum \left(Y_i - \bar{Y}\right)^2} \\ &= \alpha_2 + \frac{Cov(Y,V)}{Var(Y)} = \alpha_2 + \frac{\beta_2}{1 - \alpha_2\beta_2} \frac{Var(V)}{Var(Y)} \end{aligned}$$

ya que resolviendo el sistema y despejando Y encontramos

$$Y = \frac{1}{1-\alpha_2\beta_2}\left(\beta_1 + \alpha_1\beta_2 + \beta_3 Z + U + \beta_2 V\right)$$

y por tanto

$$Cov(Y,V) = \frac{\beta_2}{1-\alpha_2\beta_2} Var(V)$$

porque U, V y Z están incorrelados. Por tanto, el estimador MCO de α_2 está en general sesgado y es inconsistente.

Si $\beta_2 = 0$, entonces X no depende de Y, no hay simultaneidad y el estimador MCO es un estimador consistente (el sesgo se hace cero).

Si $\alpha_2\beta_2 = 1$ las rectas son paralelas en el plano $\{X, Y\}$ y no se cruzan. La ecuación de forma reducida no está definida

c. (5 puntos) *Explique de forma razonada si existe algún instrumento válido para estimar α_2. Si su respuesta es positiva, demuestre que el correspondiente estimador de variables instrumentales (VI) es consistente.*

α_2 está identificada por que la variable exógena Z está omitida en la ecuación (8), y por tanto se puede usar como instrumento ya que es independiente de V (es exógena). Como está presente en la ecuación (7) está correlada con Y (es relevante) si $\beta_3 \neq 0$.

$$\hat{\alpha}_2^{VI} = \frac{\sum\left(Z_i - \bar{Z}\right)\left(X_i - \bar{X}\right)}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} = \frac{\sum\left(Z_i - \bar{Z}\right)X_i}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)}$$

y sustituyendo la ecuación (8),

$$\begin{aligned}
\hat{\alpha}_2^{VI} &= \frac{\sum\left(Z_i - \bar{Z}\right)\left(\alpha_1 + \alpha_2 Y_i + V_i\right)}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} = \alpha_2\frac{\sum\left(Z_i - \bar{Z}\right)Y_i}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} + \frac{\sum\left(Z_i - \bar{Z}\right)V_i}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} \\
&= \alpha_2 + \frac{\sum\left(Z_i - \bar{Z}\right)V_i}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)}
\end{aligned}$$

por lo que

$$\begin{aligned}
p\lim \hat{\alpha}_2^{VI} &= \alpha_2 + p\lim \frac{\sum\left(Z_i - \bar{Z}\right)V_i}{\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} \\
&= \alpha_2 + \frac{p\lim \frac{1}{n}\sum\left(Z_i - \bar{Z}\right)V_i}{p\lim \frac{1}{n}\sum\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} \\
&= \alpha_2 + \frac{Cov(Z,V)}{Cov(Z,Y)} = \alpha_2
\end{aligned}$$

porque $Cov(Z,V) = 0$ y $Cov(Z,Y) \neq 0$ si $\beta_2 \neq 0$ ya que

$$Cov(Z,Y) = \frac{\beta_2}{1-\alpha_2\beta_2} Var(Z),$$

y el estimador de Variables Instrumentales es consistente.

d. (7 puntos) *Demostrar que el estimador de mínimos cuadrados en dos etapas (MC2E) de α_2 es idéntico al estimador VI en c. Recuerde que el estimador MC2E es el estimador MCO de α_2 en el modelo (8) donde Y es sustituida por sus valores predichos (ajustados) por MCO en la ecuación de forma reducida*

$$Y = \gamma_1 + \gamma_2 Z + W,$$

donde W es un término de error (las letras griegas siguen siendo parámetros).

$$\hat{\alpha}_2^{MC2E} = \frac{\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(X_i - \bar{X}\right)}{\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)^2} = \frac{\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(X_i - \bar{X}\right)}{\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(Y_i - \bar{Y}\right)}$$

porque $\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)^2 = \sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(Y_i - \bar{Y}\right)$ ya que las propiedades del ajuste MCO implican que $\sum \left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(Y_i - \hat{Y}_i\right) = 0$, donde

$$\hat{Y} = \hat{\gamma}_1 + \hat{\gamma}_2 Z$$

se ajusta por MCO. Por tanto,

$$\begin{aligned}
\hat{\alpha}_2^{MC2E} &= \frac{\sum \left([\hat{\gamma}_1 + \hat{\gamma}_2 Z_i] - \overline{[\hat{\gamma}_1 + \hat{\gamma}_2 Z_i]}\right)\left(X_i - \bar{X}\right)}{\sum \left([\hat{\gamma}_1 + \hat{\gamma}_2 Z_i] - \overline{[\hat{\gamma}_1 + \hat{\gamma}_2 Z_i]}\right)\left(Y_i - \bar{Y}\right)} \\
&= \frac{\hat{\gamma}_2 \sum \left(Z_i - \bar{Z}\right)\left(X_i - \bar{X}\right)}{\hat{\gamma}_2 \sum \left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} \\
&= \frac{\sum \left(Z_i - \bar{Z}\right)\left(X_i - \bar{X}\right)}{\sum \left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)} = \hat{\alpha}_2^{VI},
\end{aligned}$$

y concluimos que $\hat{\alpha}_2^{MC2E}$ es equivalente al estimador de VI, y por tanto, es consistente.

e. (6 puntos) *Suponga que $\alpha_1 = 0$. Determinar si $\bar{X}/\bar{Y}$ es un estimador consistente de α_2. Justifique su respuesta.*

Tenemos que

$$\frac{\bar{X}}{\bar{Y}} = \frac{\alpha_2 \bar{Y} + \bar{V}}{\bar{Y}} = \alpha_2 + \frac{\bar{V}}{\bar{Y}},$$

por lo que

$$p\lim \frac{\bar{X}}{\bar{Y}} = \alpha_2 + p\lim \frac{\bar{V}}{\bar{Y}} = \alpha_2 + \frac{p\lim \bar{V}}{p\lim \bar{Y}} = \alpha_2$$

ya que $p\lim \bar{V} = E[V] = 0$ (una media muestral de variables iid con esperanza cero), mientras que

$$\begin{aligned}
p\lim \bar{Y} &= \frac{1}{1 - \alpha_2 \beta_2}\left(\beta_1 + \alpha_1 \beta_2 + \beta_3 p\lim \bar{Z} + p\lim \bar{U} + \beta_2 p\lim \bar{V}\right) \\
&= \frac{1}{1 - \alpha_2 \beta_2}\left(\beta_1 + \alpha_1 \beta_2 + \beta_3 \mu_Z\right),
\end{aligned}$$

demuestra que $p \lim \bar{Y}$ existe y en general no será cero. Por lo tanto, el estimador es consistente.