Econometrics Final Exam Solutions

Universidad Carlos III de Madrid

May 26th, 2015

Answer all questions in two hours and a half.

QUESTION 1 (33 marks): *A researcher is considering two regression specifications to estimate the relationship between a variable X and a variable Y,*

$$\log Y = \beta_1 + \beta_2 \log X + U \quad (1)$$
$$\log \frac{Y}{X} = \alpha_1 + \alpha_2 \log X + V, \quad (2)$$

where the Greek letters refer to parameters and X and Y are two random variables for which we have a random sample of size n.

a. (4 marks) *Determine whether (2) can be expressed as a restricted version of (1).*

(2) can be rewritten as

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

so this is a reparameterized version of (1) with $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2 + 1$, but NOT a restricted version since it does not impose any restriction on the values of β_1 or β_2.

b. (7 marks) *Using the same n observations of the variables Y and X, the researcher fits the two specifications using ordinary least squares (OLS). The fits are*

$$\widehat{\log Y} = \hat{\beta}_1 + \hat{\beta}_2 \log X \quad (3)$$
$$\widehat{\log \frac{Y}{X}} = \hat{\alpha}_1 + \hat{\alpha}_2 \log X, \quad (4)$$

where the Greek letters with tilde are the estimated values by OLS. Using the expressions for the estimates, write $\hat{\beta}_2$ in terms of $\hat{\alpha}_2$.

Write

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

so that

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

c. (4 marks) *Write $\hat{\beta}_1$ in terms of $\hat{\alpha}_1$.*

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

d. (5 marks) *Demonstrate that*

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

This is $\hat{y} - x = \hat{z}$, where

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

e. (3 marks) *Demonstrate that the residuals of (3) are identical to those of (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 marks) *Demonstrate that the standard errors of $\hat{\beta}_2$ and $\hat{\alpha}_2$ are identical.*

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

g. (4 marks) *Determine the relationship between the t statistic using $\hat{\beta}_2$ and the t statistic using $\hat{\alpha}_2$.*

The t statistic for $H_0 : \beta_2 = 0$ is

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

which corresponds to the t statistic for the equivalent hypothesis $H_0 : \alpha_2 + 1 = 0$.

h. (3 marks) *Explain with detailed arguments whether R^2 would be the same in the two regressions.*

R^2 will be different because it measures the proportion of the variance of the dependent variable explained by the regression and the dependent variables are different in each regression.

QUESTION 2 (33 marks): *You are conducting an econometric investigation into the hourly wage rates of male and female employees. Your particular interest is in comparing the determinants of wage rates for female and male workers. The sample data consist of a random sample of observations on 526 paid employees, 252 of whom are females and 274 of whom are males. The sample data provide observations on the following variables:*

W = hourly wage rate, measured in dollars per hour
S = number of years of formal education completed, in years
A = age, in years;
T = firm tenure of employee, in years;.
F = 1 if employee is female and =0 if employee is male

The following model is proposed

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

where the Greek letters denote unknown parameters and U is an error term. It is assumed that all the classic assumptions of the multiple regression model hold, including homoskedasticity. The observed sample provides the following estimates (standard errors of each estimate in parenthesis),

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

The corresponding residual sum of squares (RSS) and the total sum of squares (SST) of the $n = 526$ observations are

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

a. (6 marks) *Use the estimation results to compute the OLS estimates of the slope coefficients of the regressors S, A and T when only the 274 observations corresponding to male employees are used. Repeat the exercise when only the 252 observations corresponding to female employees are used.*

For males ($F = 0$) we have

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

so the coefficients are $\hat{\beta}_2 = 0.05937$, $\hat{\beta}_3 = 0.07980, \hat{\beta}_5 = -0.01057$ and for females ($F = 1$)

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

so the coefficients are, respectively,

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

b. (9 marks) *Write the expression (or formula) for the marginal effect of A over $\ln W$ for male employees in terms of the unknown parameters, implied by regression equation (5). Repeat the exercise for female employees. Provide the null hypothesis that the marginal effect of A on $\ln W$ for male employees is equal to the marginal effect of A on $\ln W$ for female employees. Provide also the expression for the alternative hypothesis. Give the equation for the restricted regression implied by H_0. The restricted OLS estimation provides $RSS = 81.7242$. Use this information, together with the information provided in the problem formulation, to calculate the test statistic, indicating which is its approximate distribution under the null hypothesis. Establish a decision rule, using a significance level of 5%, and the corresponding conclusion on the inference provided by the test.*

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

which are to be interpreted as $100\left(\beta_2 + 2\beta_4 A\right)\%$ being the expected percentage change in W when A changes in one unit, everything else fixed.
The hypotheses are

$$\begin{aligned}H_0 &: \beta_9 = \beta_{10} = 0 \\ H_1 &: H_0 \text{ is false (either } \beta_9 \text{ or } \beta_{10} \text{ or both are } \neq 0).\end{aligned}$$

The restricted regression is

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

Then the test statistic is an F statistic under homoskedasticity for $q = 2$ restrictions

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

whose asymptotic distribution under the null is $\chi^2_2/2$, which provides a 5% critical value equal to $cv_{0.05} = 5.99/2 \approx 3$. Therefore, since $F > cv_{0.05}$, we reject the null hypothesis in favour of the conclusion that the marginal effect is different for males and females.

c. (9 marks) *Write the restriction over the parameters in equation (5) that the marginal effects of S over $\ln W$ and of T over $\ln W$ are equal for male and female employees. Provide an expression of equation (5) imposing these two restrictions. The restricted OLS estimation provides $RSS = 80.8747$. Use this information, together with the information provided in the problem formulation, to calculate the test statistic, indicating which is its approximate distribution under the null hypothesis. Establish a decision rule, using a significance level of 5%, and the corresponding conclusion on the inference provided by the test.*

The marginal effects of S are

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

which are to be interpreted as $100(\beta_2 + \beta_6 T)\%$ being the expected percentage change in W when S changes in one unit, everything else fixed, and the marginal effects of T are

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

which are to be interpreted as $100(\beta_2 + \beta_6 S)\%$ being the expected percentage change in W when T changes in one unit, everything else fixed.

To test

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

the hypotheses are

$$\begin{aligned} H_0 &: \beta_8 = \beta_{12} = \beta_{11} = 0 \\ H_1 &: H_0 \text{ is false (at leat one of } \beta_8, \beta_{12}, \beta_{11} \text{ is } \neq 0). \end{aligned}$$

and the restricted regression is

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

Then the test statistic is an F statistic under homoskedasticity for $q = 3$ restrictions

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

whose asymptotic distribution under the null is $\chi^2_3/3$, which provides a 5% critical value equal to $cv_{0.05} = 7.815/3 \approx 12.6$. Therefore, since $F < cv_{0.05}$, we can not reject the null hypothesis that the marginal effect is equal for males and females at the 5% significance level.

d. (9 marks) *Write the restriction over the parameters in equation (5) that the regression equations are identical for male and female employees, that is, the mean log-wage of female employees with any given values of S, A and T equals the mean log-wage of male employees with the same values of S, A and T. Provide an expression for the null hypothesis (H_0) of the mentioned restriction and for the alternative (H_1). The restricted OLS estimation provides $RSS = 93.1805$. Use this information, together with the information provided in the problem formulation, to calculate the test statistic, indicating which is its approximate distribution under the null hypothesis. Establish a decision rule, using a significance level of 5%, and the corresponding conclusion on the inference provided by the test.*

The hypotheses are

$$\begin{aligned} H_0 &: \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0 \\ H_1 &: H_0 \text{ is false (at least one of } \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12} \text{ is} \neq 0). \end{aligned}$$

and the restricted regression is

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

Then the test statistic is an F statistic under homoskedasticity for $q = 6$ restrictions

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

whose asymptotic distribution under the null is $\chi^2_6/6$, which provides a 5% critical value equal to $cv_{0.05} = 12.592/6 = 2.0987$. Therefore, since $F > cv_{0.05}$, we reject the null hypothesis that the regression functions are equal for males and females, so we conclude that they differ in at least one coefficient.

QUESTION 3 (34 marks): *Consider an econometric model where two variables Y and X are jointly determined by the following equations*

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

The Greek letters denote unknown parameters, U and V are model errors, mutually uncorrelated, with zero mean, and Z is an exogenous variable, independent of the errors.

a. (4 marks) *Explain with detailed arguments, whether β_2 are α_2 identified under the information provided in the problem presentation.*

β_2 is not identified since there are no instruments available for the first equation (i.e. there are not omitted exogenous regressors in (7) which are present in (8)).

α_2 is identified because we can use Z as instrument in equation (8), since it is independent of V (exogenous) and correlated with Y (relevant) if $\beta_3 \neq 0$.

b. (12 marks) *Demonstrate that the OLS estimate of α_2 in the equation (8) is asymptotically biased in general. Which would be your conclusion if $\beta_2 = 0$? Which would be your conclusion if $\alpha_2\beta_2 = 1$?*

We have that

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

and substituting (8) we find that

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

so that

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

because solving the system we find that

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

and that

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

since U, V and Z are uncorrelated. Then the OLS estimate of α_2 is in general biased and inconsistent.

If $\beta_2=0$, then X is not influenced by Y, there is not simultaneity and OLS will be a consistent estimate (the bias cancels out).

If $\alpha_2\beta_2=1$ the lines are parallel in $\{X,Y\}$ dimensions and they do not intersect. The reduced form relationship is undefined.

c. (5 marks) *Explain with detailed arguments whether there is any valid instrument to estimate α_2. If your answer is positive, demonstrate that the corresponding instrumental variable (IV) estimate is consistent.*

α_2 is identified because the exogenous variable Z is omitted in equation (8), so it can be used as instrument as it is independent of V. Since it is present in equation (7) it is correlated with Y (relevant) if $\beta_3\neq 0$.

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

and substituting (8) we find that

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

so that

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

because $Cov(Z,V)=0$ and $Cov(Z,Y)\neq 0$ if $\beta_2\neq 0$ since

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

and the IV estimate is consistent.

d. (7 marks) *Demonstrate that the two stage least squares (2SLS) estimate of α_2 is identical to the IV estimate in c. Recall that the 2SLS estimate is the OLS estimate of α_2 in model (8) when Y is replaced by its predicted (fitted) values by OLS in the reduced form equation*

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

where W is an error term (the Greek letters continue to be parameters).

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

because $\sum\left(\hat{Y}_i - \overline{\hat{Y}}\right)^2 = \sum\left(\hat{Y}_i - \overline{\hat{Y}}\right)(Y_i - \bar{Y})$ since by the properties of OLS estimation $\sum\left(\hat{Y}_i - \overline{\hat{Y}}\right)\left(Y_i - \hat{Y}_i\right) = 0$, where

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

is fitted by OLS. Then

$$\begin{aligned}
\hat{\alpha}_2^{2SLS} &= \frac{\sum\left([\hat{\gamma}_1 + \hat{\gamma}_2 Z_i] - \overline{[\hat{\gamma}_1 + \hat{\gamma}_2 Z_i]}\right)(X_i - \bar{X})}{\sum\left([\hat{\gamma}_1 + \hat{\gamma}_2 Z_i] - \overline{[\hat{\gamma}_1 + \hat{\gamma}_2 Z_i]}\right)(Y_i - \bar{Y})} \\
&= \frac{\hat{\gamma}_2 \sum (Z_i - \bar{Z})(X_i - \bar{X})}{\hat{\gamma}_2 \sum (Z_i - \bar{Z})(Y_i - \bar{Y})} \\
&= \frac{\sum (Z_i - \bar{Z})(X_i - \bar{X})}{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})} = \hat{\alpha}_2^{IV},
\end{aligned}$$

hence $\hat{\alpha}_2^{2SLS}$ is equivalent to the IV estimate, and therefore consistent.

e. (6 marks) *Suppose that $\alpha_1 = 0$. Determine whether $\bar{X}/\bar{Y}$ is a consistent estimate of α_2. Justify your answer.*

We have that

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

so then

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

since $p\lim \bar{V} = E[V] = 0$ (a sample average of zero mean iid variables), while

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

so that $p\lim \bar{Y}$ exists and in general will not be zero. Therefore, the estimate is consistent.