

Problem Set 6

ECON 21020 Spring, 2021

Jake Underland, Groupmates: Ian Bamford, Matthew Chen

2021-06-05

Contents

Question 1	1
Question 2	3
Question 3	8
Question 4	12

Question 1

Consider an industry of firms $i = 1, \dots, n$, where each firm i produces according to the Cobb-Douglas production function

$$Q_i = AK_i^{\beta_1} L_i^{\beta_2} M_i^{\beta_3} \cdot u_i,$$

where K_i denotes the dollar value of capital employed by firm i , L_i is the number of hours of labor, M_i is the quantity of raw materials processed and u_i is a firm specific productivity shock.

- a) Transform this production function into a linear regression model. What are the expected signs of $\beta_1, \beta_2, \beta_3$?

Solution. We transform the function into a linear function by taking the log of both sides.

$$\begin{aligned} \log Q_i &= \underbrace{\log A}_{\equiv \beta_0} + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log M_i + \underbrace{\log u_i}_{\equiv v_i} \\ &= \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log M_i + v_i \end{aligned}$$

The expected signs of the coefficients are all positive as in theory, they are all factors of production with diminishing but positive marginal returns.

- b) Researcher A believes the firms in this industry exhibit constant returns to scale.
- i) Derive a condition on $\beta_1, \beta_2, \beta_3$ which captures this assumption.

Solution.

$$\beta_1 + \beta_2 + \beta_3 = 1$$

- ii) State the null hypothesis of constant returns to scale vs. the alternative of non-constant (either increasing or decreasing) returns to scale.

Solution.

$$\begin{aligned} H_0 : \beta_1 + \beta_2 + \beta_3 &= 1 \\ H_1 : \beta_1 + \beta_2 + \beta_3 &\neq 1 \end{aligned}$$

- iii) Define a parameter θ such that your hypotheses from bii) are equivalent to $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$.

Solution.

$$\theta = 1 - (\beta_1 + \beta_2 + \beta_3)$$

Then,

$$H_0 : \theta = 1 - \underbrace{(\beta_1 + \beta_2 + \beta_3)}_1 = 0$$

$$H_1 : \theta = 1 - \underbrace{(\beta_1 + \beta_2 + \beta_3)}_{\neq 1} \neq 0$$

- c) How would you rewrite the model to get an estimate of θ , its standard error and a 95% confidence interval for θ by running a single regression in STATA?

Solution. Rearrange the equation for θ to get

$$\theta = 1 - \beta_1 - \beta_2 - \beta_3$$

$$\Rightarrow \beta_1 = 1 - \beta_2 - \beta_3 - \theta$$

Plug this into the regression equation:

$$\log Q_i = \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log M_i + v_i$$

$$= \beta_0 + (1 - \beta_2 - \beta_3 - \theta) \log K_i + \beta_2 \log L_i + \beta_3 \log M_i + v_i$$

$$\Rightarrow \log Q_i - \log K_i = \beta_0 - \theta \log K_i + \beta_2 (\log L_i - \log K_i) + \beta_3 (\log M_i - \log K_i)$$

Using this regression equation, compute the new necessary values ($\log Q_i - \log K_i$, etc.) and run the regression. The estimate of θ is the coefficient computed on $\log K_i$ times -1. The standard error and 95% confidence interval are those reported for the $\log K_i$ term.

- d) Explain how to test the hypotheses in biii) first by using a test statistic and second by using the confidence interval from part c.

Solution. Construct the test statistic by taking the estimated coefficient reported for $\log K_i$, divide it by the standard error reported for the same variable, and take the absolute value of the whole thing. It would look something like this:

$$\left| \frac{\hat{\theta} - 0}{se(\hat{\theta})} \right| = \left| \frac{\hat{\theta}}{se(\hat{\theta})} \right|$$

And then compare this with the $1-\alpha/2$ quantile of the t distribution of degree of freedom $n-3-1 = n-4$, rejecting the null hypothesis if our test statistic is greater. Using the reported confidence interval, we can check to see if $\theta = 0$ is included. If not, then we reject the null hypothesis.

- e) Researcher B believes that capital inputs do not affect output, but that firms in the industry still exhibit constant returns to scale.

- i) State Researcher B's assumptions formally as a (joint) null hypothesis.

Solution.

$$H_0 : \beta_2 + \beta_3 = 1; \beta_1 = 0$$

$$H_1 : \text{Not } H_0$$

- ii) What are the restricted and unrestricted regressions in this case?

Solution. In the restricted case, since $\beta_2 + \beta_3 = 1$, we can write $\beta_3 = 1 - \beta_2$. Plugging this and $\beta_1 = 0$ into the regression yields the restricted regression.

$$\begin{aligned} \text{Restricted : } \log Q_i &= \beta_0 + \beta_2 \log L_i + (1 - \beta_2) \log M_i + v_i \\ &\Rightarrow \log Q_i - \log M_i = \beta_0 + \beta_2 (\log L_i - \log M_i) + v_i \\ \text{Unrestricted : } \log Q_i &= \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log M_i + v_i \end{aligned}$$

- iii) Explain how to conduct an F-test of the null hypothesis that Researcher B is correct vs. the alternative that they are not at the 5% significance level. State the null and alternative hypotheses, explain how to compute the F-statistic, and give the critical value for your test.

Solution. The hypotheses are as stated earlier:

$$\begin{aligned} H_0 : \beta_2 + \beta_3 &= 1; \beta_1 = 0 \\ H_1 : \text{Not } H_0 \end{aligned}$$

In order to conduct the F-test, we need to know the residual sum of squares for both the restricted and unrestricted regression. Thus, we would run both regressions to obtain SSR_R and SSR_U . We know that the number of restrictions imposed by the null is 2, and the number of parameters in the unrestricted regression is 4. Then, we construct the F-statistic:

$$F = \frac{(SSR_R - SSR_U)/2}{SSR/(n - 4)}$$

If this F-statistic exceeds the critical value $F_{2, n-4, 0.95}$, we reject the null.

- f) Do you expect the error term to be correlated with the regressors?

Solution. It would be plausible that the error terms would be correlated with the regressors as firms producing at a similar scale in terms of capital, labor, and raw materials may have similar markets and thus may experience similar trends/shocks in their productivity.

Question 2

Download the data JTRAIN from canvas. The scrap rate for a manufacturing firm is the number of defective items out of every 100 produced. For a given number of items produced, a lower scrap rate indicates higher worker productivity. In 1988, firms became eligible to receive grants for training workers.

a.

Using the data from 1988, estimate the effect of receiving a training grant on hrsemp, the hours of training per employee. Control for log of sales and log of total employees. Is the effect statistically significant at the 5% level?

Solution.

```
df <- read.csv("JTRAIN.csv")
names(df)
```

```
## [1] "year"      "fcode"     "employ"    "sales"     "avgsal"    "scrap"
## [7] "rework"    "tothrs"    "union"     "grant"     "d89"       "d88"
## [13] "totrain"   "hrsemp"    "lscrap"    "lemploy"   "lsales"    "lrework"
## [19] "lhrsemp"   "lscrap_1"  "grant_1"   "clscrap"   "cgrant"    "clemploy"
## [25] "clsales"   "lavgsal"   "clavgsal"  "cgrant_1"  "chrsemp"   "clhrsemp"
```

```
# filtering data to 1998
df88 <- df[df["year"] == 1988,]

model1 <- lm(hrsemp ~ grant + lsales + lemploy, data = df88)

stargazer(model1, type = "latex", header = FALSE)
```

Table 1:

	<i>Dependent variable:</i>
	hrsemp
grant	26.254*** (5.592)
lsales	-0.985 (3.540)
lemploy	-6.070 (3.883)
Constant	46.665 (43.412)
Observations	105
R ²	0.237
Adjusted R ²	0.214
Residual Std. Error	24.380 (df = 101)
F Statistic	10.444*** (df = 3; 101)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Since the p-value is reported to be lower than 0.05, the effect is statistically significant at the 5% level.

b.

Using the data from 1988, estimate the effect of receiving a training grant on $\log(\text{scrap})$, where scrap is the scrap rate. Are grants estimated to make productivity better or worse? Is the effect statistically significant at the 5% level?

Solution.

```
model2 <- lm(log(scrap) ~ grant, data = df88)

stargazer(model2, type = "latex", header = FALSE, report=("vc*sp"))
```

Table 2:

	<i>Dependent variable:</i>
	log(<i>scrap</i>)
grant	0.057 (0.406) p = 0.890
Constant	0.409* (0.241) p = 0.096
Observations	54
R ²	0.0004
Adjusted R ²	-0.019
Residual Std. Error	1.423 (df = 52)
F Statistic	0.019 (df = 1; 52)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

The regression shows that the grants have an infinitesimal but positive effect on $\log(\textit{scrap})$, but the results are extremely statistically insignificant. At this level, it is hard to tell whether grants have any effect on productivity.

c.

Grants were given on a first come first served basis. Why might we not believe firms were randomly assigned to receiving grants? Is there likely to be selection bias in your estimate in b)?

Solution. Because grants were given on a first come first serve basis, it is natural to assume that the firms that would benefit most from grants applied first to receive them. This might include firms whose scrap rate is greatly reduced by training, but those that haven't spent as much time on training prior to the grants due to costs. This would create a qualitative difference in the treated and untreated groups, which amounts to selection bias.

d.

Conduct a balance test at the 5% significance level for the receipt of grants in 1988, controlling for $\log(\textit{scrap})$ from 1987, \log of sales in 1988 and \log of total employees in 1988. What do you conclude? Why is it not valid to run the balance test by checking each of the individual t-statistics for the coefficients?

Solution.

```
balance_df <- df88
balance_df$scrap <- df[df$year == 1987,]$scrap
balance_tst <- lm(grant ~ log(scrap) + lsales + lemploy, data = balance_df)

summary(balance_tst)
```

```
##
## Call:
## lm(formula = grant ~ log(scrap) + lsales + lemploy, data = balance_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5646 -0.3679 -0.2372  0.5320  0.7908
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.92096    1.62709   1.181   0.244
## log(scrap)    0.01505    0.04644   0.324   0.747
## lsales       -0.15556    0.13147  -1.183   0.243
## lemploy       0.21486    0.13007   1.652   0.105
##
## Residual standard error: 0.4731 on 46 degrees of freedom
## (107 observations deleted due to missingness)
## Multiple R-squared:  0.08248,    Adjusted R-squared:  0.02264
## F-statistic: 1.378 on 3 and 46 DF,  p-value: 0.2612
```

Since the p-value of the F-stat is greater than 0.05, the null is not rejected, meaning there does not seem to be a statistically strong relationship between the covariates studied here and grant receipt.

The balance test must be run by checking the overall F-statistic since we want to examine the total effect of covariates on selection and not the individual covariates. Individually, the covariates may not significantly influence the firm's decision to receive grants, but taken together, they may affect the firm's willingness to select into reception of grants.

e.

Repeat part b) but now control for log(scrap) from 1987. Test the null hypothesis that the effect of the grant is non-negative vs. the alternative that it is negative at the 5% significance level. State your null and alternative hypotheses, test statistic, critical value and decision.

Solution.

```
df88$scrap87 <- df[df$year == 1987,]$scrap

model3 <- lm(log(scrap) ~ grant + log(scrap87), data = df88)
```

```
stargazer(model3, type = "latex", header = FALSE, report=("

```

Table 3:

	<i>Dependent variable:</i>
	log(scrap)
grant	-0.254* (0.147) p = 0.091
log(scrap87)	0.831*** (0.044) p = 0.000
Constant	0.021 (0.089) p = 0.813
Observations	54
R ²	0.873
Adjusted R ²	0.868
Residual Std. Error	0.513 (df = 51)
F Statistic	174.941*** (df = 2; 51)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

$$H_0 : grant \geq 0$$

$$H_1 : grant < 0$$

```
# one-sided t-test
res <- summary(model3)
T1 <- (-0.254 - 0) / 0.147 # test statistic
t_critical <- qt(0.05, 54 - 3) # critical value
cat("Test statistic is: ", T1, "\nCritical value is:", t_critical,
    "\nH_0 is rejected:", T1 < t_critical) # decision: rejection if T < t_critical
```

```
## Test statistic is: -1.727891
## Critical value is: -1.675285
## H_0 is rejected: TRUE
```

f.

Test the null hypothesis that the parameter on $\log(\text{scrap})$ from 1987 equals 1 versus the 2-sided alternative at the 5% level by using the reported p -value.

Solution. p -value reported in the regression is that of the 5% 2-sided test with a null hypothesis of $\log(\text{scrap87}) = 0$. We rebuild the test below where $\theta = 1 - \beta_2$ and we test for

$$H_0 : \theta = 0$$

$$H_1 : \theta \neq 0$$

```
model_f <- lm((log(scrap) - log(scrap87)) ~ grant + log(scrap87), data = df88)
```

```
stargazer(model_f, type = "latex", header = FALSE, report=("vc*sp"))
```

Table 4:

	Dependent variable:
	(log(scrap) - log(scrap87))
grant	-0.254* (0.147) p = 0.091
log(scrap87)	-0.169*** (0.044) p = 0.0004
Constant	0.021 (0.089) p = 0.813
Observations	54
R ²	0.273
Adjusted R ²	0.244
Residual Std. Error	0.513 (df = 51)
F Statistic	9.571*** (df = 2; 51)
Note:	*p<0.1; **p<0.05; ***p<0.01

We reject the null according to the reported p -value on θ , which is the coefficient on $\log(\text{scrap87})$ reported above.

Question 3

In this problem we will decompose a naive comparison of averages into a treatment effect and selection bias. Suppose a wine seller has 1000 different wines on offer and decides to advertise half of them. We are asked to assess the effect of advertising on wine sales, and are given last month's sales figures.

- a) Explain in words why the 'naive' comparison of sales figures may not be informative of the ATE, ATT or ATU.

Solution. If we compare the sales of the wines that were advertised versus the wines that were not advertised, we cannot ensure that sales of both categories of wines respond similarly to advertisement. For example, the wines that were advertised may be wines that were already gaining popularity,

and enjoy a greater potential treatment effect than the wines that were not advertised. In order to estimate the ATT, we need to know the sales of the wines that were advertised had they not been advertised, and for ATU, we need to know the sales of wines that were not advertised had they been advertised. From here we can compute the overall ATE. Thus, the "naive" comparison, which compares the sales of the treated group and the sales of the untreated group, cannot directly inform us of the ATE, ATT or ATU unless under certain assumptions such as randomization.

- b) What do you need to assume about selection bias for the comparison of sales figures to be informative of the ATE, ATT and ATU? Also write your answer in terms of conditional expectations.

Solution.

Let $D = \{0, 1\}$ be a dummy that takes the value of 1 when the wine is part of the treated group and 0 when the wine is part of the untreated group. Let $y(i)$ be the sales of a particular group when treated ($i = 1$) or untreated ($i = 0$). We can write

$$E(Y|D) = \beta_0 + \beta_1 D$$

where

$$\begin{aligned}\beta_0 &= E(y(0)|D = 0) \\ \beta_1 &= E(y(1)|D = 1) - E(y(0)|D = 0)\end{aligned}$$

Take the naive comparison β_1 .

$$\begin{aligned}\beta_1 &= E(y(1)|D = 1) - E(y(0)|D = 0) \\ &= \underbrace{E(y(1)|D = 1) - E(y(0)|D = 1)}_{ATT} + \underbrace{E(y(0)|D = 1) - E(y(0)|D = 0)}_{Selection\ Bias}\end{aligned}$$

In this case, if *Selection Bias* in y_0 is equivalent to 0, that is, if $E(y(0)|D = 1) = E(y(0)|D = 0)$, and the expected sales of the advertised wines in the absence of advertisement are identical to the unadvertised wines, we are able to directly measure the *ATT* via the naive comparison.

$$\begin{aligned}\beta_1 &= E(y(1)|D = 1) - E(y(0)|D = 0) \\ &= \underbrace{E(y(1)|D = 0) - E(y(0)|D = 0)}_{ATU} + \underbrace{E(y(1)|D = 1) - E(y(1)|D = 0)}_{Selection\ Bias}\end{aligned}$$

In the above case, if *Selection Bias* in y_1 is equivalent to 0, that is, if $E(y(1)|D = 1) = E(y(1)|D = 0)$, and the expected sales of the advertised wines when treated are identical to the expected sale of unadvertised wines when treated, we are able to directly measure the *ATU* via the naive comparison. Furthermore, if there is no significant difference in $D = 0$ and $D = 1$, then we can take away the conditionals from our expectations and express the naive comparison as follows:

$$\begin{aligned}\beta_1 &= E(y(1)|D = 1) - E(y(0)|D = 0) \\ &= \underbrace{E(y(1)) - E(y(0))}_{ATE}\end{aligned}$$

In the above case, there is no selection bias, and the *ATE* can be directly inferred from the naive comparison.

- c) Can the naive comparison equal the ATE even with selection bias? Explain.

Solution. From the first decomposition of β_1 , we have

$$\beta_1 = \underbrace{E(y(1)|D = 1) - E(y(0)|D = 1)}_{ATT} + \underbrace{E(y(0)|D = 1) - E(y(0)|D = 0)}_{Selection\ Bias}$$

We also know the following

$$\begin{aligned}ATE &= P(D = 1)ATT + P(D = 0)ATU \\ &= 0.5ATT + 0.5ATU\end{aligned}$$

Then, we can rewrite β_1 as

$$\begin{aligned}
\beta_1 &= \underbrace{E(y(1)|D=1) - E(y(0)|D=1)}_{ATT} + \underbrace{E(y(0)|D=1) - E(y(0)|D=0)}_{Selection\ Bias} \\
&= \underbrace{0.5ATT + 0.5ATU}_{ATE} + 0.5ATT - 0.5ATU + \underbrace{E(y(0)|D=1) - E(y(0)|D=0)}_{Selection\ Bias} \\
&= ATE + Selection\ Bias + 0.5(ATT - ATU)
\end{aligned}$$

Thus, for β_1 to equal ATE with selection bias, the below must hold:

$$Selection\ Bias = 0.5(ATU - ATT)$$

	Not Advertised	Advertised	Total
Sales with advertising	\$1000	\$600	\$800
Sales without advertising	\$800	\$500	\$650
Difference	\$200	\$100	\$150

- d) Assume average sales for advertised wines were \$600, whereas non-advertised wines were \$800. Assume also that the selection bias in y_0 is $-\$300$ and the selection bias in y_1 is $-\$400$. Complete the table below and explain your answers. (Note: Half of wines are advertised).

Because

$$\text{Selection Bias}(y_0) = E(y(0)|D = 1) - \underbrace{E(y(0)|D = 0)}_{\$800} = -\$300$$

$$\text{Selection Bias}(y_1) = \underbrace{E(y(1)|D = 1)}_{\$600} - E(y(1)|D = 0) = -\$400$$

$$\Rightarrow E(y(0)|D = 1) = \$500, E(y(1)|D = 0) = \$1000$$

- e) Consider the regression $Y_i = \beta_0 + \beta_1 D_i + u_i$. What sign does β_1 have? Explain why correlation is not indicative of causation in this example by comparing β_1 with the ATE, ATT and ATU of advertising.

Solution. Because $\beta_1 = E(y(1)|D = 1) - E(y(0)|D = 0)$, and $E(y(1)|D = 1) < E(y(0)|D = 0)$ as reported in the table above, β_1 would likely have a negative sign. In this example, the ATU is given by the Difference row for the Not Advertised column, ATT is given by the Difference of the Advertised column, and ATE is given by computing $0.5ATT + 0.5ATU$. The values are below:

$$ATU = \$200$$

$$ATT = \$100$$

$$ATE = \$150$$

Thus, although the regression seems to suggest that advertising lowers sales, that is due to the strong influence of selection bias on the data. When we know the selection bias, we can ascertain that the naive comparison that leads to β_1 is indeed incorrect, and that the negative correlation does not at all imply negative causation.

- f) Suppose that the profit margin is 25% of sales. If the cost of advertising a bottle of wine is \$40 per month, should the wine merchant also advertise the 500 wines that have not yet been advertised?

Solution.

```
# Profit without advertisement
p_a = 500 * 800 * 0.25
# Profit with advertisement
p_b = 500 * 1000 * 0.25 - (40 * 500)
cat(p_b, "-", p_a, "=", p_b - p_a)
```

```
## 105000 - 1e+05 = 5000
```

therefore, since advertisement increases profits, the firm should advertise the rest of the wine bottles.

- g) If the wine merchant only knew the ATE before advertising the wines, would they pay to have all wines advertised?

Solution. The ATE is $ATE = \$150$. Thus, the firm would estimate the following profits for advertising the remaining bottles of wine:

```

# Profit without advertisement
p_a = 500 * 800 * 0.25
# Estimated profit with advertisement
p_b = 500 * (800 + 150) * 0.25 - (40 * 500)

cat(p_b, "-", p_a, "=", p_b - p_a)

## 98750 - 1e+05 = -1250

```

therefore, the firm would estimate a loss of profit from advertisement for the unadvertised stash of wines, and would not pay to have the wines advertised.

Question 4

a)

$$\begin{aligned}
 y_i &= \beta x_i^* + u_i \\
 x_i &= x_i^* + v_i \\
 y_i &= \beta(x_i - v_i) + u_i \\
 &= \beta x_i + \underbrace{u_i - \beta v_i}_{\epsilon_i} \\
 &= \beta x_i + \epsilon_i
 \end{aligned}$$

New error term is $\epsilon_i = u_i - \beta v_i$.

The true model is now

$$y = \beta x + \epsilon$$

Endogeneity problem:

$$\begin{aligned}
 Cov(x, \epsilon) &= Cov(x, u - \beta v) \\
 &= Cov(x, u) - \beta Cov(x, v) \\
 &= Cov(x^* + v, u) - \beta Cov(x^* + v, v) \\
 &= Cov(x^*, u) + Cov(v, u) - \beta Cov(x^*, v) - \beta Cov(v, v)
 \end{aligned}$$

Where

$$\begin{aligned}
 Cov(x^*, u) &= E(x^*u) - E(x^*)E(u) = 0 - 0 * E(x^*) = 0 \\
 v \perp\!\!\!\perp u &\implies Cov(v, u) = 0 \\
 v \perp\!\!\!\perp x^* &\implies Cov(x^*, v) = 0
 \end{aligned}$$

Thus,

$$\begin{aligned}
 Cov(x, \epsilon) &= Cov(x^*, u) + Cov(v, u) - \beta Cov(x^*, v) - \beta Cov(v, v) \\
 &= -\beta Var(v)
 \end{aligned}$$

If v is nondegenerate, variance is strictly greater than 0. Because $Cov(x, \epsilon) = -\beta Var(v) \neq 0$, we have an endogeneity problem.

b) Running an OLS regression on

$$y_i = \beta x_i + \epsilon_i$$

We know from Problem Set 3 that $\hat{\beta}_{OLS} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\frac{1}{n} \sum_i x_i y_i}{\frac{1}{n} \sum_i x_i^2}$. From the law of large numbers, and continuous mapping theorem,

$$\begin{aligned} \frac{1}{n} \sum_i x_i y_i &\xrightarrow{p} E(xy) \\ \frac{1}{n} \sum_i x_i^2 &\xrightarrow{p} E(x^2) \\ g(a, b) &= \frac{a}{b} \text{ is a continuous function for } b \neq 0 \\ \Rightarrow g\left(\frac{1}{n} \sum_i x_i y_i, \frac{1}{n} \sum_i x_i^2\right) &\xrightarrow{p} g(E(xy), E(x^2)) \\ \Rightarrow \frac{\frac{1}{n} \sum_i x_i y_i}{\frac{1}{n} \sum_i x_i^2} &\xrightarrow{p} \frac{E(xy)}{E(x^2)} \end{aligned}$$

Thus,

$$\begin{aligned} \hat{\beta}_{OLS} &\xrightarrow{p} \frac{E(xy)}{E(x^2)} \\ &= \frac{E(x(\beta x + \epsilon))}{E(x^2)} \\ &= \frac{E(\beta x^2 + x\epsilon)}{E(x^2)} \\ &= \beta + \frac{E(x\epsilon)}{E(x^2)} \end{aligned}$$

Where

$$\begin{aligned} E(x\epsilon) &= Cov(x, \epsilon) + E(x)E(\epsilon) \\ &= -\beta Var(v) + E(x)E(\epsilon) \\ E(\epsilon) &= E(u - \beta v) = E(u) - \beta E(v) = 0 \\ \Rightarrow E(x\epsilon) &= -\beta Var(v) \neq 0 \end{aligned}$$

Therefore,

$$\begin{aligned} \beta + \frac{E(x\epsilon)}{E(x^2)} &= \beta + \frac{-\beta Var(v)}{E(x^2)} \\ &= \beta \left(1 - \frac{Var(v)}{E(x^2)}\right) \\ &= \beta \left(1 - \frac{E(v^2)}{E(x^2)}\right) \neq \beta \end{aligned}$$

From what we have derived, the probability limit of $\hat{\beta}_{OLS}$ is inconsistent.

c)

$$\bar{\beta} = \beta \left(1 - \frac{E(v^2)}{E(x^2)}\right)$$

We know $\frac{E(v^2)}{E(x^2)} > 0$ because both numerator and denominator are positive. Then,

$$\begin{aligned} E(x^2) &= E((x^* + v)^2) \\ &= E(x^{*2}) + 2 \underbrace{E(x^*v)}_{=0} + E(v^2) \\ &= E(x^{*2}) + E(v^2) > E(v^2) \end{aligned}$$

Where the third equality follows from

$$\begin{aligned} v \perp\!\!\!\perp x^* &\Rightarrow Cov(x^*, v) = E(x^*v) - E(x^*)E(v) \\ &= E(x^*v) - E(x^*) \cdot 0 \\ &= E(x^*v) = 0 \end{aligned}$$

And the final equality is true because each element in the expectation is squared and thus nonnegative. Therefore, $\frac{E(v^2)}{E(x^2)} \in [0, 1]$, and we can determine $\bar{\beta}$ suffers from attenuation bias.

- d) Let $z = 1$. We assume validity $Cov(z, \epsilon) = 0$, which implies $E(z\epsilon) - E(z) \underbrace{E(\epsilon)}_{=E(u-\beta v)} = E(z\epsilon) - E(z) \cdot 0 = E(z\epsilon) = 0$.

$$\begin{aligned} E(zy) &= E(z(\beta x + \epsilon)) \\ &= \beta E(zx) + E(z\epsilon) \\ &= \beta E(zx) \\ \implies \beta^{IV} &= \frac{E(zy)}{E(zx)} \end{aligned}$$

The above is well defined when we consider relevance, which states the coefficient of the regression of x on $z = 1$, which is identical to the coefficient in an intercept only model, must be nonzero. $E(zx) \neq 0$, which we can derive from

$$E(zx) = E(x) \neq 0 \iff E(x^*) \neq 0$$

So, we have relevance. Since $z = 1$, $\beta^{IV} = \frac{E(zy)}{E(zx)} = \frac{E(y)}{E(x)}$. Thus,

$$\hat{\beta}^{IV} = \frac{\sum_i y_i}{\sum_i x_i}$$

- e) In order for this estimate to be consistent, we must have Validity $Cov(z, \epsilon) = 0$ which occurs if and only if

$$E(z\epsilon) - E(z) \underbrace{E(\epsilon)}_{=E(u-\beta v)} = E(z\epsilon) - E(z) \cdot 0 = E(z\epsilon) = 0$$

Then,

$$\begin{aligned} \hat{\beta}^{IV} &= \frac{\sum_i y_i}{\sum_i x_i} \\ &= \frac{\sum_i z_i y_i}{\sum_i z_i x_i} \\ &= \frac{\sum_i z_i (\beta x_i + \epsilon_i)}{\sum_i z_i x_i} \\ &= \beta + \frac{\frac{1}{n} \sum_i z_i \epsilon_i}{\frac{1}{n} \sum_i z_i x_i} \end{aligned}$$

From the law of large numbers,

$$\begin{aligned} \sum_i z_i \epsilon_i &\xrightarrow{p} E(z\epsilon) = 0 \\ \sum_i z_i x_i &\xrightarrow{p} E(zx) \end{aligned}$$

From Relevance, $g(a, b) = \frac{a}{b}$ is continuous on $(E(z\epsilon), E(zx))$. From the continuous mapping theorem,

$$\begin{aligned} g\left(\frac{1}{n} \sum_i z_i \epsilon_i, \frac{1}{n} \sum_i z_i x_i\right) &= \frac{\frac{1}{n} \sum_i z_i \epsilon_i}{\frac{1}{n} \sum_i z_i x_i} \xrightarrow{p} \frac{E(z\epsilon)}{E(zx)} = g(E(z\epsilon), E(zx)) \\ \implies \beta + \frac{\frac{1}{n} \sum_i z_i \epsilon_i}{\frac{1}{n} \sum_i z_i x_i} &\xrightarrow{p} \beta + \underbrace{\frac{E(z\epsilon)}{E(zx)}}_{=0} = \beta \end{aligned}$$

And we have shown that $\hat{\beta}^{IV}$ becomes consistent.