ECON306 – Quiz 4

2014· 6 · 7

Name: Bruno Salcedo (Answer Key) PSU ID: bxs5142

There are 20 questions worth 2 points each. You have 30min to solve all of them. Don't forget to write your name and PSU ID (e.g. bxs5142).

1. Which of the following would result in biased estimates of the slope coefficient? (mark all that apply)

(a) Omission of an important variable
(b) Inclusion of an irrelevant variable
(c) Imperfect multicollinearity
(d) Pure serial correlation
(e) Heterosckedasticity

2. Is it true that we should always include every available variable in order to obtain unbiased estimators?

No, including irrelevant variables will increase the variance of the OLS estimates.

3. What are the consequences for OLS of having pure serial correlation?

(a) The usual formula for standard errors is no longer valid
(b) OLS is no longer the most efficient unbiased estimator

4. According to the following estimated model:

$$\hat{y}_i = 1 + x_i + 0.5x_i^2,$$

what would be the average effect on y_i of a change Δx for an individual with $x_i = 1$?

$$\Delta y_i = (1 + 2 \times 0.5\, x_i)\Delta x_i = 2\Delta x_i$$

5. According to the following model:

$$\log(y_i) = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i,$$

by what percentage would y_i change on average, if x_i changed by 8%?

$$\text{percentual change} = \frac{\Delta y_i}{y_i} = \beta_1 \frac{\Delta x_i}{x_i} = \beta_1 \times 8\%$$

6. Consider the following estimated model:

$$\log(\hat{y}_i) = 20 - 2.5x_{1i} + 3.45x_{2i}.$$

Also consider an individual with $y_i = 10$, $x_{2i} = 1.15$ and $x_{2i} = 0.75$. According to the estimated model, what would be the average change in y_i if x_{1i} increased by $\Delta x_{1i} = 4$?

$$\begin{aligned} \frac{\Delta y_i}{y_i} &= \hat{\beta}_1 \Delta x_i = -2.5 \times 4 = -10 \\ \Rightarrow \quad \Delta y_i &= -10 y_i = -100 \end{aligned}$$

For problems 7–8, consider the following model

$$\text{ACC}_i = \beta_0 + \beta_1 \text{DRINK}_i + \beta_2 \text{SPEED}_i + \beta_3 \text{DRINK}_i \times \text{SPEED}_i + \varepsilon_i,$$

$$\begin{aligned} \text{ACC}_i =&\ \text{the probability of being involved in a car accident} \\ \text{SPEED}_i =&\ \text{the average driving speed} \\ \text{DRINK}_i =&\ \text{the blood alcohol content} \end{aligned}$$

7. What sign would you expect the coefficient β_3 to have? [Justify your answer]

Positive, because drinking increases the probability of being involved in an accident, and this effect is more pronounced the faster you are driving.

8. What would be the average effect of driving 10mph faster on the probability of having a car accident?

$$\Delta\text{ACC}_i = (\beta_2 + \beta_3 \text{DRINK}_i)\Delta\text{SPEED}_i = 10 \times (\beta_2 + \beta_3 \text{DRINK}_i)$$

For problems 9 and 10, suppose that you are interested on evaluating the effectiveness of certain given policy to foment economic growth in small rural villages. You have data for three variables

$$\begin{aligned} \text{GROWTH}_i =&\ \text{the real per capita GDP growth in the village} \\ \text{POL}_i =&\ \text{dummy variable indicating whether the policy was enforced in the } i\text{th village} \\ \text{POP}_i =&\ \text{the population of the village} \end{aligned}$$

9. According to the estimated model:

$$\widehat{\text{GROWTH}}_i = 0.1 - 0.005\,\text{POP}_i + 0.012\,\text{POL}_i,$$

what is the average effect of the policy on growth?

$$\hat{\beta}_2 = 0.012$$

10. Suppose that you suspect that the effectiveness of the policy depends on the size (population) of the village. What model could you use to capture this effect?

$$\text{GROWTH}_i = \beta_0 + \beta_1 \text{POP}_i + \beta_2 \text{POL}_i + \beta_3 \text{POP}_i \times \text{POL}_i + \varepsilon_i$$

11. Consider the data depicted in figure (1). What model could generate a good fit for this data? What can you tell about the coefficients of this model from looking at the data?

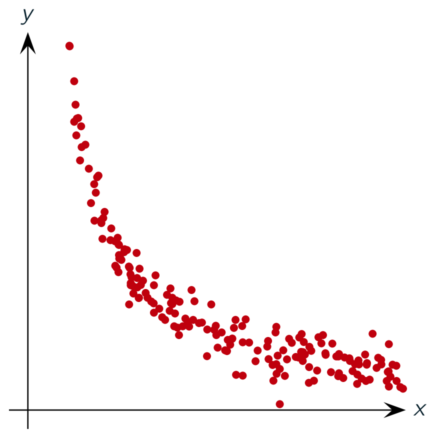

Figure (1)

$$\log(y_i) = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i, \qquad -1 < \beta_1 < 0$$
$$\text{or} \quad y_i = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i, \qquad \beta_1 < 0$$

12. Consider the data depicted in figure (2). What model could generate a good fit for this data? What can you tell about the coefficients of this model from looking at the data?

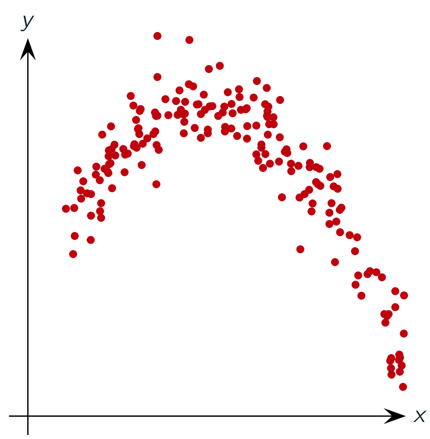

Figure (2)

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \qquad \beta_0 > 0,\ \beta_1 > 0,\ \beta_2 < 0$$

For problems 13–18, consider the following model for crime, which omits the relevant variable INC_i:

$$\text{CRIME}_i = \beta_0 + \beta_1 \text{EDU}_i + \varepsilon_i$$

$$\begin{aligned} \text{CRIME}_i &= \text{ the probability that individual } i \text{ commits a crime} \\ \text{EDU}_i &= \text{ the number of schooling years of the } i^{\text{th}} \text{ individual} \\ \text{INC}_i &= \text{ the yearly income of the } i^{\text{th}} \text{ individual} \end{aligned}$$

13. What sign would you expect for β_1?

$\beta_1 < 0$, because education may improve the legal alternatives of income and the opportunity cost of criminal records, and may have an effect on moral values.

14. What sign would you expect for the relationship between EDU_i and INC_i?

Wealthier people have greater access to education, and a lesser need for income in the short run. Hence I would expect a positive relation.

15. What sign would you expect for the relationship between INC_i and CRIME_i?

Wealthier people probably have less reasons to commit crimes, hence I would expect a negative relation.

16. What sign would you expect for the bias resulting from the omission of INC_i?

The bias is equal to the relationship between EDU_i and INC_i (positive) times the relationship between CRIME_i and INC_i (negative). Hence I would expect a negative bias, i.e. $\mathbb{E}\left[\hat{\beta}_1\right] < \beta_1$.

For problems 18 and 19, suppose that the output of your linear regression is given by:

```
---------------------------------------------------------------------------
 CRIME |      Coef.   Std. Err.      t      P>|t|     [95% Conf. Interval]
-------+-------------------------------------------------------------------
   EDU |    -0.0261   0.0053     -4.9657   0.000      [-0.0364 , -0.0158]
 _cons |     0.5355
---------------------------------------------------------------------------
```

17. Can you conclude with 95% confidence that each additional year of education reduces (on average) the probability of committing a crime by at least 1.58%? Briefly justify your answer (2-4 sentences).

No. The regression table allows to say with 95% confidence that $\mathbb{E}\left[\hat{\beta}_1\right] \in (-0.0364, -0.0158)$, and thus $\mathbb{E}\left[\hat{\beta}_1\right] < -0.0158$. However, we expect a negative bias, i.e. $\mathbb{E}\left[\hat{\beta}_1\right] < \beta_1$, and thus it could be the case that $\beta_1 > -0.0158$.

18. Can you conclude with 95% confidence that each additional year of education reduces (on average) the probability of committing a crime by no more than 3.64%?
Briefly justify your answer (2-4 sentences).

Yes. The regression table allows to say with 95% confidence that $\mathbb{E}\left[\hat{\beta}_1\right] \in (-0.0364, -0.0158)$, and thus $\mathbb{E}\left[\hat{\beta}_1\right] > -0.0364$. The estimation may be biased, but the bias is negative and hence $\beta_1 > \mathbb{E}\left[\hat{\beta}_1\right] > -0.0364$.

For problems 19 and 20, suppose that you are interested on measuring the effect of x_1 over y, and you have data for x_1, x_2 and y. Suppose that you estimate the following regressions:

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \\
y_i &= \beta_0 + \beta_1 x_{1i} + \varepsilon_i, \\
x_{2i} &= \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i,
\end{aligned}
$$

and you obtain the following results:

Y	Coef.	Std. Err.	t	P>|t|
X1	-2.0229	0.0638	-42.15	0.000
X2	-5.0504	0.0482	-104.76	0.000
_cons	1.0638	0.0638	16.66	0.000

Y	Coef.	Std. Err.	t	P>|t|
X1	1.7246	0.2153	8.01	0.000
_cons	1.6776	0.4276	3.92	0.000

X2	Coef.	Std. Err.	t	P>|t|
X1	-0.74201	0.04215	-17.60	0.000
_cons	-0.12154	0.08372	-1.45	0.148

19. How can you explain the difference between the estimates for β_1 in the first and second regressions?

The first table suggests that x_2 has a significant negative effect on y ($\beta_2 < 0$). The third table suggests that x_2 has a significant negative relation with x_1 ($\alpha_1 < 0$). Hence, when omitting x_2 in the second equation, we would expect a significant positive bias $\mathbb{E}\left[\hat{\beta}_1\right] = \beta_1 + \alpha_1\beta_2 > \beta_1$

20. Which of the two estimates do you think is more reliable? [justify your answer]

The previous answer suggests that x_2 is a relevant variable, and excluding it would generate significant bias on the estimation of β_1. Hence the first regression appears to be more reliable.