ECON306 – Quiz 5

2014 · 6 · 14

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

There are 20 questions worth 5 points each. You have 45min to solve all of them. *You must justify all your answers.* Don't forget to write your name and PSU ID (e.g. bxs5142).

1. Suppose that you suspect that some of your regressors are endogenous, but all other classical assumptions are satisfied. Can you still use OLS to estimate the model and make accurate predictions?

Yes, endogeneity does not affect the ability of OLS estimates to generate good predictions, it only affects the ability to generate reliable policy implications.

2. What is the main problem associated with endogeneity?

Bias (and inconsistency) in the estimation of $\hat{\beta}_1$, which results in policy implications of the form $\Delta y = \hat{\beta}_1 \Delta x$ to be unreliable.

3. What are the four most common sources of endogeneity?

Omitted variables, simultaneous equations (two sided causality), selection bias, and measurement error.

4. How can we test if a regressor is endogenous?

You cannot (unless you have additional exogenous instruments, but we did not discuss this in class).

5. What are the three properties that a good instrumental variable should satisfy?

It should be available (can be measured), it should be relevant (sufficiently correlated with the endogenous explanatory variable) and it should be exogenous (uncorrelated with the error term).

For questions 6–9, suppose that you are interested in the *elasticity of supply* for carbonated drinks

6. What would be the problem of just regressing sales on prices using OLS, and using $\hat{\beta}_1$ as an estimate for the elasticity of supply?

Short answer: endogeneity due to simultaneity. Explanation: prices and quantities are determined by the interaction of supply and demand. Therefore, residuals which affect the supply function are likely to be correlated with prices.

7. List three instruments that you could you use to solve this problem.

You need 3 *demand* shifters (things which affect demand but not supply). These could be: market size (population), income, preference characteristics (advertisement, fads), or prices of complements and substitutes.

8. What equation would you estimate on the first step of 2SLS?

If we are using z as an instrument, then the 1st stage equation would be:

$$P_i = \alpha_0 + \alpha_1 z_i + \eta_i$$

9. What equation would you estimate on the second step of 2SLS?

$$S_i = \beta_0 + \beta_1 \widehat{P}_i + \varepsilon_i,$$

where

$$\widehat{P}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_i$$

For questions 10–13, suppose that you want to measure the consequences of the *amount of time spend in prison* by convicts over future earnings *after release* (in the formal economy).

10. What is a potential source of endogeneity?

Omitted variable bias (or selection bias) due to the fact that people with longer sentences are more likely to have committed more serious crimes. This suggests that they also may be more likely to be involved in serious criminal activities (like organized crime/gangs) and less likely to be able to join the formal workforce.

Alternatively, even if all the people in the sample have committed similar crimes, early released are likely to be correlated with good behavior in prison, which in turn is likely to be correlated with the intention to live a normal life and hold a normal job. (Other explanations are acceptable as long as they are well justified).

11. Would the OLS estimates of such effect have a positive or a negative bias (or neither)?

Based on my answer to the previous question, I would expect a negative bias. This is because, a person who is involved in serious criminal activities is likely to expect lower future earnings in the formal economy ($\beta_2 < 0$), and to have a longer sentence ($\alpha_1 > 0$).

12. Different federal judges are known to consistently assign different sentences to similar crimes, and are assigned to cases randomly. Would the identity of the prosecuting judge be a good instrument to solve this problem?

Yes. People who get tougher judges will serve longer sentences, hence the instrument is relevant. However, since judges are assigned randomly, the identity of the judge is unlikely to be related with unobserved variables which affect earning, and thus the instrument should be exogenous. To get full credit you must have established both relevance and exogeneity.

13. Prisoners are able to participate in voluntary self-improvement programs, and those who participate may be rewarded with reduced sentences. Would the participation in such programs be a good instrument to solve this problem?

No. The instrument will most likely be relevant but also andogenous. This is because the decision to participate in this programs is likely to be correlated with unobserved characteristics (such as the intent to live a normal life) that are likely to correlate with future laboral outcomes.

For questions 14–16, suppose that you are interested in estimating the effectiveness of police presence (COPS) to reduce crime rates (CRIME).

14. What is a likely source of endogeneity?

Simultaneity, due to the fact that the government is likely to deploy more police units to regions with higher crime rates.

15. Would the OLS estimates of such effect have a positive or a negative bias (or neither)?

Positive, because things which increase crime rates will also increase police presence, and hence COPS would be positively correlated with the error term.

16. Suppose that you use OLS and obtain the estimated model

$$\widehat{\text{CRIME}}_i = \underset{(0.23)}{10.36} - \underset{(0.15)}{0.45} \cdot \text{COPS}_i.$$

Would you be able to conclude that police presence has a significant negative effect on crime rates?
[Hint: the answer to problem 15 may be helpful]

Yes, because the bias is positive. This means that I would expect the estimated coefficient to be smaller than the estimated coefficient which is already significantly negative. More formally: $0 > \mathbb{E}\left[\hat{\beta}_1\right] = \beta_1 + \text{bias} > \beta_1$.

For questions 17–20, suppose that you work at the Human Resources department of a large firm. Your firm has been offering a voluntary leadership training that takes place at the workplace after regular office hours. Your boss asks you to evaluate whether the training has a significant effect on productivity, and you have data on the following variables

$$\begin{aligned}
\text{PROD}_i &= \text{ a measure of worker productivity} \\
\text{LEAD}_i &= \text{ dummy variable equal to 1 for those who have taken the training} \\
\text{EDU}_i &= \text{the years of schooling of the employee} \\
\text{DIST}_i &= \text{distance from } i\text{'s home to the workplace} \\
\mathbf{x}_i &= \text{ other characteristics of the employees}
\end{aligned}$$

17. Suppose you use OLS to estimate the model

$$\text{PROD}_i = \beta_0 + \beta_1 \text{LEAD}_i + \beta_2 \text{EDU}_i + \beta_3 \text{DIST}_i + \boldsymbol{\gamma}\mathbf{x}_i + \varepsilon_i$$

The results of the regression are shown in table (1). Why would you think that the estimated coefficient $\hat{\beta}_1$ may be biased?

Selection bias. The program is voluntary, and I would expect more motivated people to be more likely to participate. These people would also be more productive even in the absence of the program.

18. Based on your previous answer, would you expect the bias to be positive or negative?

Positive. Because motivation (MOT_i) is likely to be positively correlated both with LEAD_i ($\alpha_1 > 0$) and with PROD_i ($\beta > 0$).

19. Suppose that you regress LEAD_i on EDU_i, and obtain the results in table (2). Do you think that EDU_i would be a valid instrument for LEAD_i?

No. The results in table (2) suggest that EDU_i is a relevant instrument. However, one could expect that EDU_i is correlated with MOT_i, and hence it is not exogenous. This hypothesis is consistent with the fact that EDU_i has a significant effect on PROD_i, as shown in table (1).

20. Suppose that you regress LEAD_i on DIST_i, and obtain the results in table (3). Do you think that DIST_i would be a valid instrument for LEAD_i?

No. In theory I would expect it to be both relevant and exogenous. Distance to the workplace is not likely to be strongly correlated with omitted variables (at least, table (1) suggests that it doesn't affect productivity). Also, I would expect that people with longer commutes are less willing to participate in after hour activities. Indeed, DIST_i has a significant effect on LEAD_i, see table (3). However, the explanatory variable of DIST_i over LEAD_i is way too small ($R^2 \approx 1.3\%$), and hence it would be a weak instrument.

```
. reg PROD LEAD EDU DIST x1 x2 x3
```

Source	SS	df	MS	Number of obs =	805
				F(6, 798) =	592.32
Model	100935.73	6	16822.6217	Prob > F =	0.0000
Residual	22664.2099	798	28.4012655	R-squared =	0.8166
				Adj R-squared =	0.8153
Total	123599.94	804	153.731269	Root MSE =	5.3293

PROD	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
LEAD	.5411501	.4883871	1.11	0.268	-.4175251	1.499825
EDU	5.188254	.1119755	46.33	0.000	4.968452	5.408055
DIST	-.0113576	.0179542	-0.63	0.527	-.0466006	.0238853
x1	2.572646	.6545689	3.93	0.000	1.287765	3.857526
x2	-.2466527	.0987435	-2.50	0.013	-.4404803	-.052825
x3	-1.039741	.3916485	-2.65	0.008	-1.808524	-.2709584
_cons	2.749862	1.914271	1.44	0.151	-1.007739	6.507463

Table (1)

```
. reg LEAD EDU
```

Source	SS	df	MS	Number of obs =	805
				F(1, 803) =	462.27
Model	71.0084359	1	71.0084359	Prob > F =	0.0000
Residual	123.346844	803	.153607526	R-squared =	0.3654
				Adj R-squared =	0.3646
Total	194.35528	804	.241735422	Root MSE =	.39193

LEAD	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
EDU	.1400715	.0065148	21.50	0.000	.1272835	.1528596
_cons	-1.624889	.0955293	-17.01	0.000	-1.812405	-1.437372

Table (2)

```
. reg LEAD DIST
```

Source	SS	df	MS	Number of obs =	805
				F(1, 803) =	10.71
Model	2.55871538	1	2.55871538	Prob > F =	0.0011
Residual	191.796564	803	.238850018	R-squared =	0.0132
				Adj R-squared =	0.0119
Total	194.35528	804	.241735422	Root MSE =	.48872

LEAD	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DIST	.0053101	.0016224	3.27	0.001	.0021255	.0084947
_cons	.3465823	.0253493	13.67	0.000	.2968236	.3963411

Table (3)