ECO375 Tutorial 6
Wooldridge: Chapter 15

Matt Tudball

University of Toronto St. George

October 27, 2017

Today's coverage:

- Chapter 15, #2 (in slides)
- Chapter 15, #5 (in slides)
- Chapter 15, #7 (in slides)
- Chapter 15, #9 (in slides)

Chapter 15, #2

Suppose that you wish to estimate the effect of class attendance on student performance. A basic model is

$$stndfl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u$$

where *stndfl* is standardised outcome on a final exam and *atndrte* is the percentage of classes attended.

i) Let *dist* be the distance from the students' living quarters to the lecture hall. Do you think *dist* is uncorrelated with u?

Suppose that you wish to estimate the effect of class attendance on student performance. A basic model is

$$stndfl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u$$

where *stndfl* is standardised outcome on a final exam and *atndrte* is the percentage of classes attended.

i) Let *dist* be the distance from the students' living quarters to the lecture hall. Do you think *dist* is uncorrelated with *u*?

It is unlikely that living quarters are assigned with the convenience of particular students in mind, so we would think that *u* and *dist* are uncorrelated. It could be, however, that living quarters which are located closer to the lecture halls are more expensive and so students from richer families tend to live there. If these students are also more likely to have, for example, a better diet or access to private tutors then *u* may be correlated with *dist*.

Chapter 15, #2

ii) Assuming that *dist* and *u* are uncorrelated, what other assumption must *dist* satisfy to be a valid IV for *atndrte*?

ii) Assuming that *dist* and *u* are uncorrelated, what other assumption must *dist* satisfy to be a valid IV for *atndrte*?

***dist* must be correlated with *atndrte*. To put this more precisely, in the reduced form we must have**

$$\boldsymbol{atbdrte} = \pi_0 + \pi_1 \boldsymbol{priGPA} + \pi_2 \boldsymbol{ACT} + \pi_3 \boldsymbol{dist} + \boldsymbol{v}$$

we must have $\pi_3 \neq 0$. Given a sample of data, we can test $H_0 : \pi_3 = 0$ against $H_1 : \pi_3 \neq 0$ using a t test.

Chapter 15, #2

iii) Suppose we add the interaction term $priGPA \cdot atndrte$

$$stndfl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \cdot atndrte + u$$

If $atndrte$ is correlated with u then, in general, so is $priGPA \cdot atndrte$. What might be a good IV for $priGPA \cdot atndrte$?

iii) Suppose we add the interaction term $priGPA \cdot atndrte$

$$stndfl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \cdot atndrte + u$$

If $atndrte$ is correlated with u then, in general, so is $priGPA \cdot atndrte$. What might be a good IV for $priGPA \cdot atndrte$?

In this model we only assumed that $\boldsymbol{atndrte}$ was endogenous (meaning that it is correlated with the error term). We can assume that $\mathbb{E}(\boldsymbol{u}|\boldsymbol{priGPA}, \boldsymbol{ACT}, \boldsymbol{dist}) = 0$. It follows from the properties of independence that any function of the exogenous variables will also be independent, so that $\boldsymbol{priGPA} \cdot \boldsymbol{dist}$ is independent of u. Since $\boldsymbol{priGPA} \cdot \boldsymbol{dist}$ is also going to be correlated with $\boldsymbol{priGPA} \cdot \boldsymbol{atndrte}$ it is a viable IV and can be estimated via 2SLS.

Chapter 15, #5

Refer to equations (15.19) and (15.20). Assume that $\sigma_u = \sigma_x$ so that the population variance in the error term is the same as the population variance in x. Suppose that the instrumental variable z is slightly correlated with u, Corr$(z,u) = 0.1$. Suppose also that z and x have a somewhat stronger correlation, Corr$(z,x) = 0.2$.

i) What is the asymptotic bias in the IV estimator?

(15.19): plim$\hat{\beta}_{1,IV} = \beta_1 + \frac{\textbf{Corr}(z,u)}{\textbf{Corr}(z,x)}\frac{\sigma_u}{\sigma_x} = \beta_1 + \frac{1}{2}$
So the asymptotic bias for the estimator for β_1 **is** $\frac{1}{2}$**.**

ii) How much correlation would have to exist between x and u before OLS has more asymptotic bias than 2SLS?

(15.20): plim$\hat{\beta}_{1,OLS} = \beta_1 + \text{Corr}(x,u)\frac{\sigma_u}{\sigma_x} = \beta_1 + \text{Corr}(x,u)$
This will occur when Corr$(x,u) > \frac{1}{2}$. This is a simple illustration of how a seemingly small correlation (.1 in this case) between the IV (z) and error (u) can still result in IV being more biased than OLS if the correlation between z and x is weak (.2).

The following is a simple model to measure the effect of a school choice programme on standardised test performance (see Rouse (1998) for motivation):

$$score = \beta_0 + \beta_1 choice + \beta_2 faminc + u_1$$

where *score* is the score on a state-wide test, *choice* is a binary variable indicating whether a student attended a choice school in the last year, and *faminc* is family income. The IV for *choice* is *grant*, the dollar amount granted to students to use for tuition at choice schools. The grant amount differed by family income level, which is why we control for *faminc* in the equation.

i) Even with *faminc* in the equation, why might *choice* be correlated with u_1?

Chapter 15, #7

i) Even with *faminc* in the equation, why might *choice* be correlated with u_1?

Even at a given income level, some students are more motivated and more able than others, and their families are more supportive (say, in terms of providing transportation) and enthusiastic about education. Therefore, there is likely to be a self-selection problem: students that would do better anyway are also more likely to attend a choice school.

ii) If within each income class, the grant amounts were assigned randomly, is *grant* uncorrelated with u_1?

ii) If within each income class, the grant amounts were assigned randomly, is *grant* uncorrelated with u_1?

Assuming we have the functional form for *faminc* correct, the answer is yes. Since u_1 does not contain income, random assignment of grants within income class means that grant designation is not correlated with unobservables such as student ability, motivation, and family support.

iii) Write the reduced form equation for *choice*. What is needed for *grant* to be partially correlated with *choice*?

The reduced form is

$$\boldsymbol{choice} = \pi_0 + \pi_1 \boldsymbol{faminc} + \pi_2 \boldsymbol{grant} + \boldsymbol{v}_2$$

and we need $\pi_2 \neq 0$. In other words, after accounting for *income*, the grant amount must have some effect on choice. This seems reasonable, provided the grant amounts differ within each income class.

iv) Write the reduced form equation for *score*. Explain why this is useful. (*Hint* : How do you interpret the coefficient on *grant*?

We obtain the reduced form for *score* by substituting the reduced form for *choice* into the structural equation *score*.

$$\begin{aligned} \boldsymbol{score} &= \beta_0 + \beta_1 \left(\pi_0 + \pi_1 \boldsymbol{faminc} + \pi_2 \boldsymbol{grant} + \boldsymbol{v}_2 \right) + \beta_2 \boldsymbol{faminc} + \boldsymbol{u}_1 \\ &= \left(\beta_0 + \beta_1 \pi_0 \right) + \left(\beta_2 + \beta_1 \pi_1 \right) \boldsymbol{faminc} + \beta_1 \pi_2 \boldsymbol{grant} + \left(\boldsymbol{u}_1 + \pi_1 \boldsymbol{v}_2 \right) \\ &= \alpha_0 + \alpha_1 \boldsymbol{faminc} + \alpha_2 \boldsymbol{grant} + \boldsymbol{v}_1 \end{aligned}$$

This allows us to estimate the direct effect of increasing the grant amount on school scores. This may be of some policy interest.

Consider a model estimating the effect of skipping classes on final exam score. In a simple regression framework, we have

$$score = \beta_0 + \beta_1 skipped + u$$

Suppose we do not have a good instrumental variable for *skipped*. But we have two other pieces of information on students: combined SAT score and cumulative GPA prior to the semester. What could we do instead of IV estimation?

Just use OLS on an expanded equation, where *SAT* and *cumGPA* are added as proxy variables for student ability and motivation.