

Topic 4 Introductory Note: Instrumental Variables

These notes are provided as a supplement to the lecture slides in BSM946 and are not expected to be used as your sole understanding of this topic. The main arguments are presented in the lecture, with extra notes here and the key reading to ensure your understanding of the topic is complete.

Introduction and recap on Endogeneity

Instrumental variables (IV) estimation is something that has numerous uses in econometrics, one of which we discussed in the lecture (removing omitted variable bias). This note is an attempt to explain another famous implementation of IV, to again solve an endogeneity issue, but of a slightly different kind.

We previously discussed endogeneity as a variable that was causing the Gauss Markov assumption:

$$E[\epsilon | X] = 0$$

to break. We discussed in Topic 6 that this may happen if there is an unobserved variable that is related to our independent variable. For example, if we are measuring the returns to education we may not observe things such as work ethic and these will thus end up in our error term. If there is a relationship between these omitted variables and our independent variables then we will end up with biased results.

Thus an equation such as this:

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Gender}_i + \beta_3 \text{Age}_i + u_i$$

may not provide a reliable estimate.

There are different ways that a variable can cause endogeneity issues however. Omitted variable bias is one key way, measurement error is another (often under-considered way), and finally through reverse causality. Reverse causality will be the key premise in this note today as it is one of the largest issues that econometricians face in real world estimation.

Reverse Causality

Reverse causality refers either to a direction of cause-and-effect contrary to common belief or to a two-way causal relationship in, as it were, a loop.

Reverse causality means that X and Y are associated, but not in the way you would expect. Instead of X causing a change in Y, it is really the other way around: *Y is causing changes in X*. For example:

*"...one may be tempted to say that low social status causes schizophrenia, [but] another plausible explanation is that schizophrenia causes downward social mobility..." ~ Gerstman***

A simple example of endogeneity is the following:

$$Y = \beta_1 X + \epsilon$$

And X can be given by:

$$X = \delta_0 + \delta_1 Y + u$$

We know that we will not estimate β_1 correctly if X is correlated with ϵ . Is it?

Well another way we can write X, by substituting in the information from above, is:

$$X = \delta_0 + \delta_1(\beta_1 X + \epsilon) + u$$

Thus, X can be seen to be a function of ϵ , and of course X and ϵ are therefore correlated. This violates the OLS assumption of $E[\epsilon|X] = 0$.

According to Katz (2006), identifying reverse causality is sometimes a matter of “common sense.” A study may find that smoking and depression are linked. It could be that smoking causes depression, or it could be the other way around, that depression causes smoking. The likelihood is that they both cause each other, (smoking $\rightarrow$ depression *and* depression $\rightarrow$ smoking) perhaps because smokers feel societal pressure to quit, but may not have the willpower to do so. When this symbiotic relationship happens, it’s often called simultaneity. This will be the case that we look at from Steven Levitt’s work later in this note.

There are a few good examples of reverse causality in health economics. When lifelong smokers are told they have lung cancer, many may then quit smoking. This change of behavior after the disease develops can make it seem as if ex-smokers are actually more likely to die of emphysema or lung cancer than current smokers. This is not really the case but we are observing both factors changing at the same time due to the feedback loop between them. Smoking caused the individual to develop cancer, but the cancer has changed the individuals smoking behavior. It is also known that many alcoholics become nondrinkers after being diagnosed with liver conditions. Such changes can confuse or confound what is a causal correlation between dying of a disease and an environmental factor.

Clearly, the decisions by smokers and alcoholics to stop smoking or drinking did not cause them to get sick and die. Even so, this tendency to reduce or eliminate things that make you become sick after the fact can undermine the statistical correlation between smoking and dying of lung cancer, and heavy alcohol consumption and dying of liver disease.

Famous Example by Steven Levitt (1997)

In a famous paper Steven Levitt (the author of the popular Freakonomics series), investigates the simple question: ‘If a city hires more policemen, will its crime rate fall?’. Or, thinking of this another way, do policemen actually do what we expect and reduce crime? This is obviously a causal question. We are not interested in whether there is a correlation between these two things, we are interested in if police levels change crime rates. Because causality is key here, we should already be thinking about ensuring we do not have reverse causality.

Being an economist Levitt (as well as the many that had tried investigating this relationship prior) tried to use OLS analysis. He looked at investigating if the ‘number of policemen per capita’ impact ‘crimes committed per year per capita’ across a range of US

cities. In this regression he also controlled for numerous other independent variables that may impact crime rates such as socioeconomic and demographic factors, and policy variables like the social spending, welfare programmes, education spending, education levels and the unemployment rate.

$$\text{Crime rate per capita}_i = \beta_0 + \beta_1 \text{Policemen per capita}_i + \beta_j \text{Other Controls}_i + u_i$$

Like the many other economists that had investigated the relationship before him, Levitt found a positive value of β_1 . This would suggest that more police officers would actually cause more crime, rather than reduce crime as we would have initially expected. The policy suggestion from this would be simple, if you want to reduce crime rates then you need to fire all your police officers!

Something is clearly wrong here. We are aware that OLS can provide biased estimates if there is endogeneity, caused by measurement error, omitted variable bias or reverse causality. Here the issue is almost certainly reverse causality. Measurement error is typically assumed to bias an estimate towards zero, but not 'past' zero, so it would not explain finding a positive result if we were truly expecting to see a negative sign. We should also feel pretty happy that this is not an issue of omitted variable bias as the list of controls seems sensible and exhaustive. Secondly, in this work Levitt is actually using fixed effects. As we know (or will see in the panel data lecture if you have not seen it yet!) with fixed effects, we can control for any unobserved time-invariant factors that may affect a given city. Thus our chance of experiencing omitted variable bias is greatly reduced. Our best guess therefore is a reverse causality problem which makes sense. Intuitively, the problem is that as soon as crime rates go up, or are expected to go up, the city will hire more policemen. So whilst policemen are likely to reduce crime (the causation we are looking for), crime rates also increase policemen (a sort of nuisance causation). This will bias our OLS results, and bias them to the extent that we are getting the incorrect sign on our parameter of interest, β_1 .

Levitt's paper is famous not for finding a biased estimate of β_1 , but rather for suggesting a good instrument to use to remove this bias. As we should remember from the lecture on instrumental variables, a good instrument is one that is i) correlated with the variable it will be replacing (police levels here), and ii) not correlated with the error term in the regression equation (i.e. it does not have a direct impact on the dependent variable crime rate).

The proposed instrument in this paper was the timing of the electoral cycle, or simply, whether or not there was a mayoral election in the city that year. This might sound like a strange instrument but the best instruments are often things that we might not initially think of. The very nature of the second condition, ii), is that the variable does not have a relationship with crime and thus we need to think slightly outside the box to find a variable that fits this assumption.

Levitt's research showed that there is a positive relationship between the mayoral elections in a city and police hiring. Mayors can be seen to increase the number of policemen on the streets in an election year in an effort to lower crime and have a higher chance of being re-elected. This proves condition i). This is the only assumption that we can actually check when selecting an instrument and we often end up with what is called the 'weak instrument problem' where the correlation between the instrument and the variable it is instrumenting is low. Therefore in reality we need to find a significant relationship for assumption i) and a strongly significant relationship.

In terms of the second assumption, ii), this is often difficult to consider and almost impossible to verify in practice. As we demonstrated in the lecture it is impossible to test as you have to verify if the correlation between something observable (the instrument) and something imaginary (the error term). There are two common reasons why this second assumption might fail:

1. The instrument should actually be in the regression itself and by not including it we have omitted an important variable.
2. The instrument is correlated with an omitted variable.

Translated into the application we are discussing here would mean that 1. There is something about it being a mayoral election year that causes crime rates to rise or 2. 'something else' happens in election years which then affects crime, and this 'something else' hasn't already been controlled for in our regression. Most economists assume that 1 is not a problem here, but this is still an assumption. It is more likely that 2 is a problem in this paper. As Levitt controls for a lot of other factors, especially other policy variables that mayoral candidates might typically change in an election year, it is difficult to think of any important omitted variables so we will have to accept assumption 2. We therefore believe that a mayoral election satisfies both i) and ii) and thus can be used effectively. If you can think of something that is important and missing feel free to disagree, and critique the chain of logic.

What impact does this instrument have on the results? The OLS result for the elasticity of violent crime (per capita) with respect to police officers (per capita) is 0.28, i.e. if you hire 1% more policemen you get 0.28% more crime. Using a two stage least squares application of IV as we saw in the lectures, the estimates are -1, i.e. if you hire 1% more policemen you get 1% less crime.

We have seen how IV works in the lectures but can we explain intuitively what is happening in this instance? In this example the issue is that the number of police officers increase when crime increases so that we are not finding the true impact of police officers on crime as the two variables are moving together. In this application the IV estimator is finding all of the times when policemen are hired at the same time as there was an election. Given that assumptions i) and ii) hold, whenever we observe this then we know the only reason that police officers were increased was due to the election, not due to changing crime rates. This thus gives the *causal* relationship as only the election *caused* police officer numbers to increase. So the IV estimator looks at what happens to crime rates, but crucially it only looks at crime rates in those times when policemen went because it was an election year. If crime rates go down, it must have been the effect of the election on policemen, and then the effect of policemen on crime that caused the falling crime rates. Therefore, even though there is a complicated mess of causality between crime and policing, all we need to find is one source of exogenous variation (something that causes one of the factors to change but not the other) and we can infer the sign and strength of one of the directions of causality. This leads us to being able to make useful policy recommendations here, in terms of causal estimates that are believable, and to tell politicians that if they want 5% lower crime they need 5% more police officers.

Further Reading

There is some really good further reading on this topic. One of the best things to read is the original article by Levitt (AER 1997) which I have added to the folder. There is also a very nice (non-mathematical) introduction to good instruments, including lots of

examples of them in Angrist and Krueger (JEP 2001). Finally, the most famous and definitely the most controversial article by Steven Levitt (Levitt and Donohue, JQE 2001), is his observation that crime rates in US states began to fall 18 years after the states legalised abortion (states legalised abortion at different times and thus differences across states that legalised and did not could be investigated). One of the controversial interpretations of this is that the legalisation of abortion cut the number of unwanted babies, the types of babies that might be more likely to commit crimes 18 years later. Whilst very controversial it is an interesting piece to read.

That's all from me this week!

David Morris