SOME NOTES ON THE INTUITION BEHIND POPULAR ECONOMETRIC TECHNIQUES

These notes have been put together by Dave Donaldson (EC307 class teacher 2004-05) and by Hannes Mueller (EC307 class teacher 2004-present). The notes are meant to help you understand the intuition behind some of the most common techniques used in applied development papers, from instrumental variables to fixed effects. The notes are a complement to your econometrics textbook, and they are intended to aid the transition from econometric theory to applications. Their purpose is help you navigate through the class papers and to critically assess how the various Authors deal with identification and econometric concerns. Hope you find them useful,

Oriana Bandiera
November 2005

PART I

Instrumental variable estimation: understanding a famous example[1]

Dave Donaldson, 2005

Instrumental variable estimation is something that many students find easy enough to understand in its mathematical form (which you were examined on in EC220/EC221 last year), but hard to understand in applications. Unfortunately, EC307 tests your ability to criticise IV estimates appearing in the literature, and this ability hinges on your ability to understand IV applications. This note is an attempt to explain applied IV estimation, without mathematics, by way of walking through a famous example from the literature that I think is the most intuitive example of an IV in applied work.

The famous example is Steve Levitt's (1997) attempt to answer the simple but crucial question 'If a city hires more policemen, will its crime rate fall?' – or, 'do policemen do what they're supposed to do, reduce crime?' Note that this is a *causal* question: we don't care about the correlation between policemen and crime; we care about whether adding one more policeman to a city will lower that city's crime rate.

Being an economist, Levitt (and many before him) tried regressing (using OLS), across US cities, 'number of policemen per capita' on 'crimes committed per year per capita', controlling for a bunch of other variables that are likely to affect crime rates (socioeconomic and demographic factors, and policy variables like other types of

[1] The idea of applied IV analysis is important and examinable in EC307. You could certainly be asked questions, like those on the problem sets, about whether an author's IV procedure is convincing. However, the example I teach here, along with other Further Reading below, is definitely not exam material.

social spending, welfare programmes, education spending and levels, and unemployment rates). We'd expect a negative sign for the coefficient on policemen: more policemen, all else equal, must reduce crime rates. But in his data, as in a majority of studies before his, the coefficient was positive. OLS results are typically interpreted by economists as causal parameters (this is why we use OLS – we're interested in causation). So all of these studies suggest, if you believe them, that if you hire more police, crime rates will rise. The obvious policy advice for raising crime is to fire all the policemen.

Something is clearly wrong. As we know, OLS goes wrong when there is measurement error, omitted variable bias, or endogeneity (meaning reverse causation). Here, the problem is almost definitely endogeneity, rather than ME[2] or OVB[3]. Intuitively, the problem is that as soon as crime rates go up (or are expected to go up) the city will hire more police. So you can see the endogeneity (reverse causality): policemen reduce crime (that's the causation we're looking for), and crime increases policemen (this is a sort of 'nuisance causation'). OLS results are biased in the presence of reverse causality – often very badly biased, to the extent of taking the wrong sign as we see here.

Levitt's paper is famous for suggesting a good instrument for this problem. This good instrument must be (i) correlated with policemen, and (ii) not correlated with the error in the regression equation. Can you think of anything?

His proposal was to use the timing of the electoral cycle, or simply whether or not there was a mayoral election in the city that year. This might sound crazy, because here is an economist thinking about politics[4] – but, in fact, the best instruments are often crazy. The very nature of condition (ii) is that a good IV has to be something that has nothing to do with one of the two variables we're interested in.

First, he shocked anyone who thinks US mayors are benevolent by showing that there is a positive correlation between mayoral elections and police hiring (mayors crank up the number of policemen on the beat in an election year). So condition (i) was proven. As always, condition (i) is easy to check. The IV won't even begin to work if condition (i) isn't satisfied.[5]

---

[2] Under the most classical/plausible/simple assumptions about the type of ME (such as those you saw in EC220/EC221), ME will bias the coefficient towards zero, but never 'past zero'. That is, it couldn't make the coefficient take the wrong sign.

[3] OVB is probably not a serious problem for two reasons. First, the controls look pretty exhaustive. Second, Levitt is using panel data which allows him to include 'fixed city effects' in his model. We'll see these 'fixed effects' more, later in the year. But this type of estimation method allows him to control for anything about each city that affects crime and is constant over time. So some unmeasurable/unobservable omitted variable such as 'the taste for crime', which is likely to be constant, is not really omitted at all – it turns up in the city fixed effect.

[4] Levitt probably only thought of this because he was one of the first economists to look at applying economic models and statistics to politics.

[5] You might hear people talk about the 'weak instrument' problem. This is simply the (big) problem that arises whenever the correlation between the instrument and the thing it's instrumenting for is 'low' (where 'low' is pretty dependent on the circumstances, but below 0.5 is usually problematic), and there

Now think about condition (ii). This is always the hard one to think about, and to verify in practice. In fact, it's impossible to verify using the data because you'd have to check the correlation between something observable (the IV) and something imaginary (the error term), which you obviously can't do. Condition (ii) again was: the IV is uncorrelated with the error term in the regression. The error term captures everything that affects the dependent variable (crime rate), but which hasn't already been explained by the right-hand-side variables (policemen and a bunch of socioeconomic/demographic/policy controls). So, there are two reasons why the IV might fail: (a) the IV belongs in the regression itself (or the IV is itself an omitted variable); and (b), the IV is correlated with an omitted variable. Translated into our application, these are: (a) this being a mayoral election year affects crime directly (there's something about all the election hysteria and competition that drives people to commit crimes); and (b), 'something else' happens in election years which then affects crime, and this 'something else' hasn't already been controlled for (an obvious thing that affects crime and which happens in election years is more policemen get hired but obviously this is controlled for – this is the thing we're instrumenting for). I think it's pretty safe to assume that (a) is not a problem,[6] but that (b) is more likely to be a problem for Levitt's strategy. However, he's controlled for all of the other policy variables that incumbent mayors like to crank up (like education, welfare, job-creation programmes) just before elections, so we need to think of something else. And I can't. Hence, I'm forced to conclude that the instrument must satisfy condition (ii). Feel free to disagree, if you can criticise this chain of logic.

The punchline? OLS results say the elasticity of violent crime (per capita) with respect to police officers (per capita) is 0.28 – hire 1% more policemen, and get 0.28% more crime. But, thankfully for Bobbies on the beat, IV estimates[7] of this elasticity are -1.0 – hire 1% more policemen and get 1% *less* crime.

How about the intuition for IV estimation? Why does it work? In this example, the IV estimator is finding all the times when policemen are hired at the same time as there was an election. Whenever that happens, under the assumptions that the IV satisfies conditions (i) and (ii), we know that the only reason policemen went up was because

---

aren't many observations (again, highly dependent on context, but below 500-1000 observations is usually very serious). So, in practice, condition (i) doesn't just have to hold (ie a correlation of 0.0001 will satisfy condition (i)), but it has to hold strongly. This might sound like a contradiction of the proofs (of consistency of the IV estimator) in EC220/EC221, but those proofs were of consistency and not unbiasedness. Remember, 'consistent' means 'biased, but the bias tends towards zero as the number of observations tends to infinity', whereas 'unbiased' means 'no systematic bias, even if you had only one observation'. So, if you had lots and lots of data, even weak instruments will be fine; but in the real world, where you don't have lots and lots of observations, weak instruments are a problem.

[6] But note how I have to use the word 'assume' here. As always, there's no way to check an instrument using the data you have in your study.

[7] Actually, they're called 2SLS estimates in the paper, not IV. Recall that 2SLS is just IV applied to the case where you have more than one instrument for the thing you're instrumenting for. Here, Levitt uses both mayoral elections and gubernatorial elections (those which elect state governors), and hence he has more instruments (2) than things to instrument for (1), and hence 2SLS is appropriate.

the election *caused* policemen to go up. Then the IV estimator looks at what happens to crime rates, but crucially it only looks at crime rates at those times when policemen went up because it was an election year. If crime rates go down at these times, then, condition (ii) tells us that it wasn't the election that caused crime rates to go down, so it must have been the effect of the election on policemen, and then the effect of policemen on crime, that caused crime to go down. Hence, even though there is a complicated mess of causality between crime and policing (one causing the other, in various ways that are very hard to understand), all we need is to find one tiny source of 'exogenous variation' (something that causes one of them but not the other at all) and we can infer the sign and strength of one of the directions of causation. And hence make policy prescriptions based on causality that are actually useful and believable.

Further reading:

For the original article by Levitt (AER 97), a stable JSTOR link is http://links.jstor.org/sici?sici=0002-8282%28199706%2987%3A3%3C270%3AUECIPH%3E2.0.CO%3B2-V

For a nice (non-mathematical) introduction to good instruments, including lots of examples of them, try Angrist and Krueger (JEP 2001):
http://links.jstor.org/sici?sici=0895-3309%28200123%2915%3A4%3C69%3AIVATSF%3E2.0.CO%3B2-3

The most famous (and controversial) article by Steve Levitt (joint with Donohue, QJE 2001) is his observation that crime rates in US states began to fall 18 years after the state legalised abortion (states legalised abortion at different times, so it's not a cross-US phenomenon; and states that didn't legalise abortion didn't have the 18-year-later drop in crime). One controversial interpretation is that legalising abortion cut the number of unwanted babies, the types of babies who might be more likely to commit crimes 18 years later. A link to the working paper version is: http://www.nber.org/papers/w8004

PART II

<u>Econometrics for EC 307</u> Hannes Mueller

This small guide should provide you with some very crude intuition of important tools used in the papers that are discussed in EC 307. It is neither exhaustive nor formally waterproof. This guide only serves as a help to understand some of the more difficult mechanisms it does not cover everything you will have to know for the exam.

Dummies

Dummies are variables that take two values. A "female" dummy usually takes the value one if the individual is female. In that case the omitted category is "male". For the interpretation of regressions with Dummy variables it is very important to check what the omitted category is.

This becomes clear if we think of an example in which households are categorized "very poor", "poor", "rich" and "very rich". We want to know what effect this variable has on intake of calories. One way to test the effect of these different categories is to construct three dummies: one for "very poor", one for "rich" and one for "very rich". Obviously the omitted category is "poor". That means that all the coefficients on "very poor", "rich" and "very rich" measure the difference between the intake of households in their category and that of poor households. If there is any kind of positive monotone relationship between being better off and intake of calories we expect the coefficient of "very poor" to be negative and the other two to be positive. In that case the parameters tell a story like: *on average, very poor households have a lower intake of calories than poor households. Rich and very rich households tend to have a higher intake than poor households*. How tricky the business with the omitted category is, can be seen if we think of how the parameters change with a change of the omitted category. If we omit "very rich" for example, we expect all other coefficients to be negative.

It should be noted that this view is only valid in the case of an intercept in the regression. If there is no intercept, all categories of a variable should be included and the sign of coefficients are not sufficient to decide which household has a higher intake. We would then compare the magnitude of the coefficients.

Interaction Terms

Another tricky concept that often appears in EC 307 is interaction terms. "Interaction" here means that two variables are multiplied with each other. Behind an interaction stands the hypothesis that one variable has an effect on how the other variable influences the dependant variable. What does that mean? An example:

You are interested in the effect that grades of third year undergraduates have on their income in the following year. The dataset you have, gives you information on the kind of employment (student, public sector, private sector) they follow and their average "income" (including scholarships, transfers from the family). Now your hypothesis could be something like that the effect of grades on "income" is especially high for undergrads that take up a job in the private sector after their graduation. You multiply a "student" dummy with grade and a "private sector" dummy with grade. In your regression you also include the variables grade, private sector and student without interaction. How would you interpret a positive coefficient on "private sector" * grade? Well, due to the fact that we omitted one category – namely "public sector" – but included grade, we can interpret the positive coefficient as a higher effect of grades on income for graduates in the private sector. We can conclude: *on average, the effect of grades on income is higher for graduates employed in the private sector than for graduates in public sector*.

Note that this does not necessarily tell us anything about the "income" levels themselves. That is why you should include "student" and "private sector" dummies that could have positive or negative coefficients regardless of their interaction with grade. If "private sector" has a negative coefficient but "private sector"*grade has a positive coefficient, it means that income is lower for the average graduate in the private sector than in the public sector but that income reacts more to grades in the private sector than in the public sector.

Controls

Many articles on the reading list for EC 307 mention that they "control" for something if they introduce variables on the right hand side. In order to make clear why this is done and why it is important imagine the following example.

Imagine a pop quiz in the three tuns. Only in this particular case each of the participants plays alone. You would like to find out whether male or female participants are better in this game. So as a first step you regress the scores of each player on a dummy that equals 1 if the player is male and 0 if the player is female. The coefficient for this dummy turns out to be positive and significant. So your hasty conclusion could be that male participants are better in the game. But then you realize that you also have the age of players. That means that you can control for age in your regression of scores on sex[8]. When you run the regression of scores on sex and age the coefficient on the male dummy turns out to be insignificant while age has a highly significant positive coefficient.

What does this tell you? Older participants do better on average than younger participants. Seemingly the male participants were on average older than the female players and therefore knew more. Since age was excluded in the first regression the male dummy picked up the effect of age on scores.

---

[8] Age is a classic omitted variable (see below).

Control variables are of crucial importance for social scientists since we have often no way of testing our ideas in laboratories in which controlling for everything else is relatively easy. For example, we can't run experiments on interest rate policy (even if we would find a central banker mad enough to help us) because the economy changes constantly and due to other factors than the one input we are interested in, in this case interest rate policy. Control variable are the attempt to create something similar to a controlled lab experiment. Due to the complexity of things, this task is simply impossible. The authors of articles usually try to convince the readers that they got at least close.

Omitted Variables

If we don't control for every important variable, there is likely some problems with our estimation. One example is given above. The omission of age led to an overestimation of the effect of sex. How does this work in general? Can we tell what kind of bias will be caused by omitting a variable?

These two questions can be answered if we think about the triangle between three variables: dependant variable, included explanatory variable, omitted explanatory variable. The table below shows when we will have an upwards bias or downwards bias in the regression of an explanatory variable on the dependant variable.

	Omitted variable positively affects the dependant variable	Omitted variable negatively affects the dependant variable	Omitted variable does not affect the dependant variable
Omitted variable positively correlated with included explanatory variable	a) Positive bias	b) Negative bias	No bias
Omitted variable negatively correlated with included explanatory variable	c) Negative bias	d) Positive bias	No bias
Omitted variable not correlated with included explanatory variable	No bias	No bias	No bias

The case described in the pub quiz example is case a). The omitted variable (age) was positively correlated with the dependant variable (points) and the included variable (sex). This upwards bias is especially bad because it makes us see a link where there is none. In research, it is usually important that the authors find a robust effect of one variable on another. It is therefore less worrisome if it is clear that the bias caused by an omitted variable works against the significance of the link. For example: Your idea is that for property rights are positively correlated with investment. In addition you know that you omitted a variable (e.g. informal property rights) that is negatively correlated with property rights but positively with investment. Here we are in case b. While this is definitely bad, it is less bad than a) or d) since it will cause less of a problem for your story if you can show that despite the downwards bias you still have a significant coefficient on property rights.

Fixed Effects

Imagine you are after the following question: Does sleeping more before an exam have a good or a bad effect on the grade? Some of your friends swear on nights on red bull and show you their fantastic grades achieved by that strategy. Imagine further that someone gave you a fantastic data set. A panel of all undergrads within a certain cohort with information on hours slept before the exam and the grades achieved.

A first natural approach would be to regress the grades on hours slept. But then you think again. All the friends that follow that strategy are workaholics and overachievers so there could be something else going on. You realize that there are hundreds of individual attributes that could influence the link (like ability, living in the library, tendency to consume alcohol and other drugs...). If these individual characteristics are related to both the tendency to sleep before the exam and the grade achieved they could bias your result. A wonderful way out for this problem are fixed effects. They work like giving every individual in your dataset their own dummy. What does this do: it leads to a comparison of each grade/hours of sleep pair with the individual average instead of the overall average. This means that you introduce a different benchmark for every student. Or to make it more clear: Assume you had a genius in your year who commonly drowned her sorrows over being one in heavy drinking nights before the exam and slept only four hours for 90 points. Looking at the average student in your year who slept eight hours and wrote a 64 this looks like less sleep leads to a better grade. But given that she slept only two hours for the 80s she wrote this still implies that sleep had a good effect on her performance. It should be obvious that an ordinary regression without the fixed effects might lead to radically different (and wrong) results.