Instrumental Variables

Paul Johnson

KU CENTER FOR RESEARCH METHODS & DATA ANALYSIS
College of Liberal Arts & Sciences

http://crmda.ku.edu

Guide No: XXX

1 What's the problem?

Suppose you estimate a model

$$y_i = \beta_o + \beta_1 X1_i + \varepsilon_i \tag{1}$$

Unbeknownst to you, there is a devil who corrupts your data. The devil has some unobserved error and it inserts that error into both $X1$ and ε.

You thought you had (1), but really you have

$$y_i = \beta_o + \beta_1^{bad}(X1_i + v_i) + (\epsilon_i + v_i)$$

It is easy enough to run that regression, but the estimate you get, $\widehat{\beta_1^{bad}}$ is not a consistent or unbiased estimate of β_1. Why's that?

We violated a basic assumption in regression analysis, which is that the predictor is uncorrelated with the error term. It is plain as day that the correlation between $(X1_i + v_i)$ and $(\epsilon_i + v_i)$ is not zero. In the regression assumptions, we often glossed over that by asseting that $X1_i$ is a "fixed" regression (hence, it cannot be correlated with the error term).

Lets make picture on blackboard, try to visualize the bad thing that happens, how the estimate of β_1 is sent sideways.

We'd like to replace the observed $X1_i$ with an improved column of numbers, one which is not correlated with $(\epsilon_i + v_i)$. We don't clean up the error term, we instead try to protect our predictor from it.

2 Check the books

In RHS, p. 254, we find a display that explains IV. Their presentation is quite nice, I think. Their idea is that instead of a system with 1 equation, like (1), now we think of $X1_i$ as being an endogenous variable, something that is predictable by a formula like this:

$$X1_i = \gamma_o + \gamma_1 Z_i + u_i$$

you actually have a system with 2 equations:

$$\begin{aligned} y_i &= \beta_o + \beta_1 X1_i + \varepsilon_i \\ X1_i &= \gamma_o + \gamma_1 Z_i + u_i \end{aligned} \tag{2}$$

1425 Jayhawk Blvd.
470 Watson Library
Lawrence, KS 66045-7555

Phone: 785-864-3353
Website: crmda.ku.edu
Email: crmda@ku.edu

The error terms ε_i and u_i are "correlated" with each other. Because I have a difficult time saying what exactly correlated means, I shy away from that a bit. I'd rather say there is an omitted variable v_i going into both of them. But if you like the correlation idea, I don't want to argue against that.

Because $X1$ is predictable from Z, it means that the regression devil actually has a sympathetic side. It is not just v_i popping into our regression model 1. There is actually some structure.

A "reduced form" equation re-arranges a system so that the unobserved variables.

$$y_i = \beta_0 + \beta_1\{\gamma_o + \gamma_1 Z_i + u_i\} + \varepsilon_i$$

which rearranges as

$$y_i = \beta_0 + \beta_1\gamma_0 + \beta_1\gamma_1 Z_i + \{\beta_1 u_i + \epsilon_i\}$$

If we run the regression model using Z_i as a predictor, the coefficient we receive is an estimate of $\{\beta_1\gamma_1\}$, not $\beta_1$by itself.

$$\widehat{\beta_1\gamma_1} = \frac{Cov(y, Z)}{Var(Z)} \tag{3}$$

If only we could find a way to get rid of γ_1 from that formula!

Suppose we run a regression

$$X1 = \gamma_o + \gamma_1 Z_i + u_i$$

and get an estimate

$$\hat{\gamma_1} = \frac{Cov(X1, Z)}{Var(Z)} \tag{4}$$

Take the original estimator (3) and divide by (4)

$$\frac{\widehat{\beta_1\gamma_1}}{\widehat{\gamma_1}} = \frac{\frac{Cov(y,Z)}{Var(Z)}}{\frac{Cov(X1,Z)}{Var(Z)}} = \frac{Cov(y, Z)}{Cov(X1, Z)} \tag{5}$$

Can I give you an intuition?

Beyond the most basic intuition about the ratio (5) giving us an estimate of β_1, I have not much to say. Maybe there is an intuition.

The problem is that $X1$ is correlated with the residual. One can think of IV estimation as a two step process.

1. Stage 1. Create a new variable, $\check{X}1$, from which we have removed the correlation with the residual.
2. State 2. Replace $X1$ with $\check{X}1$ and then run the regression

$$y_i = \beta_o + \beta_1 \check{X}1_i + \varepsilon_i$$

The coefficient we get from that run turns out to be the same as the ratio estimate above.

In stage 1, of course, this all supposes that we have a good predictor for $X1$, commonly called Z, and that Z is uncorrelated with the residual ϵ_i. Because $E[Z, \epsilon] = 0$, we have "purged" the correlated part from the model.

To be explicit, stage 1 estimates a predicted value from

$$X1 = \gamma_o + \gamma_1 Z_i + u_i \tag{6}$$

In stage 2, we let the predicted value $\widehat{X1}$ serve as the new, improved $X1$.

Why is the predicted value have good as an instrument? Most importantly,

1. the predicted value is uncorrelated with the regression residual.
2. Because equation (6) does not have ε in it, the predicted value is, by definition, uncorrelated with ε.
   a) How do I know ε is not in the equation? Well, its is not in there, and
   b) We assume that u_i is uncorrelated with ϵ_i.

Matrices

If you work that out in matrix algebra, letting $\mathbf{X} = [1, X1]$ and $Z = [1, Z]$, the "usual" OLS regression estimator

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y \tag{7}$$

turns into the IV estimator

$$(\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T y \tag{8}$$

Of course, if $\mathbf{Z}$ is almost exactly like $\mathbf{X}$, then the IV estimator is almost just the same as the OLS estimator.

In the book *Microeconometrics Using Stata* 2ed, Cameron and Trivedi justify this by saying that the fundamental assumption of instrumental variables regression is that the new variable, Z, must be uncorrelated with the error term in the original model. The assumption is, for case i

$$E\{z_i^T(y_i - X1_i^T\beta)\} = 0$$

In matrix form, we assert that holds for all i

$$E\{Z^T(y - \mathbf{X}\beta_1)\} = \mathbf{0}.$$

If we simply start with that assertion (a moment condition),

$$Z^T(y - \mathbf{X}\beta_1)\} = \mathbf{0}$$

$$Z^T y - Z^T\mathbf{X}\beta_1 = \mathbf{0}$$

$$\beta_1 = (\mathbf{Z}^T\mathbf{X})\mathbf{Z}^{-1}y$$

3 Identification

Under identified: We don't have a Z variable that "works" as an instrument for X

Exactly identified: We have exactly one Z for each X. That means the ratio in (5) is a unique estimate.

Over identified: Suppose there are several predictors that work for X. That means we can form several different ratios like (5), one for each instrument. In that case, we'd like to either a) get rid of some instruments, or b) blend the instruments together. Two stage least squares is one way to blend these together.