Instrumental Variables

Paul Johnson

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 $X1$. What is the value of $\check{X}1_i = \widehat{X1}$? Well, obviously

$$\check{X}1_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_i$$

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

$$\hat{\beta_{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y \tag{7}$$

turns into the IV estimator

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

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

The statistical properties of (8) are pretty easy to see if we replace y with $X\beta + \epsilon$.

$$\boldsymbol{\beta}^{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T(\mathbf{X}\boldsymbol{\beta} + \epsilon) \tag{9}$$

$$= \cancel{(\mathbf{Z}^T\mathbf{X})^{-1}(\mathbf{Z}^T X)}\boldsymbol{\beta} + (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\epsilon$$

We insert $1 = \frac{N}{N}$ (N is sample size).

$$= \boldsymbol{\beta} + (\frac{1}{N}\mathbf{Z^T X})^{-1}\frac{1}{N}\mathbf{Z}^T\epsilon$$

By inserting N in this way, we can understand the large sample properties of this estimator. If, as $N \to \infty$, the correlation between the instrument and the error shrinks to 0, the second term will vanish. This presupposes that $(\frac{1}{N}\mathbf{Z^T X})$ can be "solved" (inverted), which basically requires that $\mathbf{Z}$ is related to $\mathbf{X}$, no matter how weakly.

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.

Some authors suggest that we should just pick 1 instrument Z for a predictor. Greene suggests that's a mistake, and suggests instead we should use all instrumental columns and we blend them together by 2SLS (2008, p. 318).

4 Weak and Invalid Instruments

Hahn and Hausman wrote a series of articles about problems that arise in the use of IV models. This is a review essay about their many projects.

Hahn, Jinyong, and Hausman, Jerry. 2003. "Weak Instruments: Diagnosis and Cures in Empirical Econometrics" *The American Economic Review*, Vol. 93, No. 2, pp. .

Hahn, Jinyong, and Hausman, 2005, "Estimation with Valid and Invalid Instruments" *Annales d'Économie et de Statistique*, No. 79/80, pp. 25-57.

4.1 Invalid

If $\mathbf{Z}$ does belong as a predictor of $\mathbf{y}$, but we assume it is not, then $\mathbf{Z}$ is an invalid instrument. If we use $\mathbf{Z}$ as an instrument to generate a predicted value of $\mathbf{X}$, the predictor $\breve{\mathbf{X}} = \hat{\mathbf{X}} = \mathbf{Z}\hat{\gamma}$ is not helpful. If we replace X with that, then we are inserting something that is correlated with the error term.

Jargon: "exclusion restriction": making the IV model work requires the assumption–on a theoretical level–that $\mathbf{Z}$ does not belong in the regression prediction of y.

4.2 Weak

If the instrument Z is not highly predictive of X, then the final estimates with have high standard errors. Even if X truly does influence y, the model does not help us see the effect.

You'll find some caution about IV (or 2SLS) giving biased estimates in the literature. As Hahn and Hausman point out on many occasions, the bias is not worse than OLS. In fact, as the instrument becomes weaker, the 2SLS estimate moves away from the correct value toward the value that OLS gave. As a result, there is no good reason to prefer OLS (it seems to me, anyway).

In Hahn and Hausman (2003), they consider a weak instrument with $R^2 = 0.01$, a very weak instrument, and show that the final estimates of the standard errors are distorted. "As either the number of instrument grows or the covariance between the structural and reduced-term stochastic disturbances be- comes large, the downward bias in the estimation of $\sigma_{\varepsilon\varepsilon}$will become large." (p. 121). As Hahn and Hausman observed elsewhere, "The estimated standard errors for the 2SLS estimator are downward biased, sometimes leading to the mistaken inference that the 2SLS estimates are much more precise than they actually are." (2005, p. 37)

If the instrument is invalid–even slightly correlated with the error term–the damage from weak instruments is more dramatic. "Thus, when the R^2 is small (e.g., 0.01), a large amount of bias results which does not decrease with increasing sample size" (p. 123).

5 Believable examples

This is the most difficult part of IV-based modeling. We need a predictor variable Z that

1. DOES NOT belong as a predictor in the model for y which, by implication,means that
   a) omitting it from the model for y does not make the model wrong
   b) omitting the variable, and letting its 'influence' go 'into' the error term does not have any effect. In other words, the error in the original model ϵ is uncorrelated with the instrumental variable. $\mathbf{Z}$ and ε can only be "uncorrelated" if $\mathbf{Z}$ does not belong in the model for y in the first place!
2. DOES serve as a predictor of X.

As I have searched for persuasive examples of IV, I have been mostly disappointed. I have found some comfort in the misery of others. William Greene (2008, p. 319) wrote, "An obvious question is where one is likely to find a suitable set of instrumental variables. In many time-series settings, lagged values of the variables in the model provide natural candidates. In other cases, the answer is less obvious. The asymptotic covariance matrix of the IV estimator can be rather large if $\mathbf{Z}$ is not highly correlated with $\mathbf{X}$; the elements of $(\mathbf{Z}^T\mathbf{X})^{-1}$ grow large. Unfortunately there usually is not much choice in the selection of instrumental variables. The choice of $\mathbf{Z}$ is often ad hoc. There is a bit of a dilemma in this result. It would seem to suggest that the best choices of instruments are variables that are highly correlated with $\mathbf{X}$. But the more highly correlated a variable is with the problematic columns of $\mathbf{X}$, the less defensible the claim that these same variables are *uncorrelated* with the disturbances." (2008, *Econometric Analysis 6th ed*, p. 319).

Greene's Wage example

Greene (2008, p. 320) presents a regression model for wages that depends on personal characteristics as well as weeks worked:

$$ln\,Wage_{it} = \beta_1 + \ldots + \beta_4 Wks_{it} + \beta_5 Occ_{it} + \ldots$$

He then observes that one should not estimate that model, because the variable Wks_{it} is probably jointly determined with the wage. That is to say, Wks should go up when $ln\,Wage$ goes up, and as wages go down, then people will work fewer weeks. He suggests there should be a second equation for weeks worked

$$Wks_{it} = \gamma_1 + \gamma_2 lnWage_{it} + \gamma_3 Ed_i + \gamma_4 Union_{it} + \gamma_5 Fem_i + u_{it}$$

Since Wks and $lnWage$ are jointly determined, we need instruments for $lnWage$ here, and Greene compares 2 selections of instruments. Here's a picture.

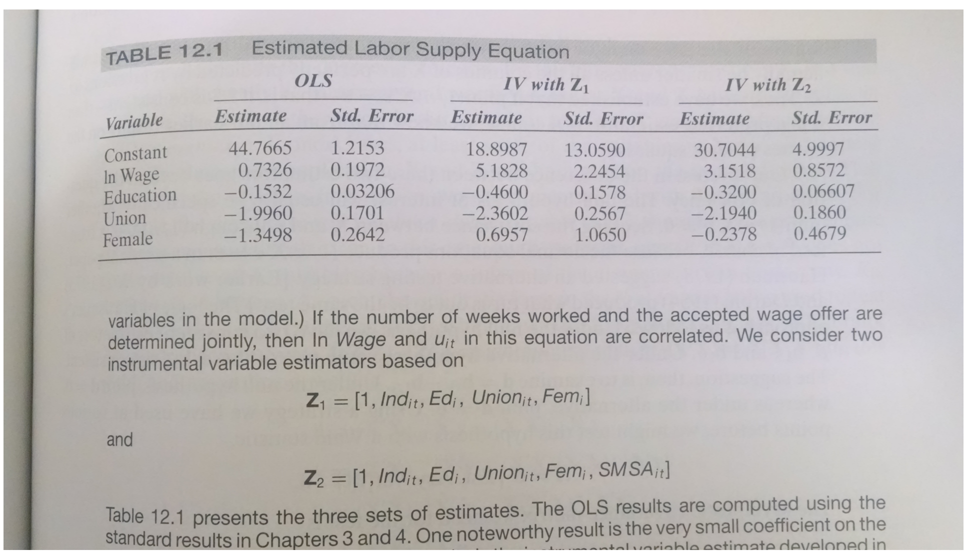

TABLE 12.1 Estimated Labor Supply Equation

	OLS		*IV with* $\mathbf{Z}_1$		*IV with* $\mathbf{Z}_2$	
Variable	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
Constant	44.7665	1.2153	18.8987	13.0590	30.7044	4.9997
ln Wage	0.7326	0.1972	5.1828	2.2454	3.1518	0.8572
Education	−0.1532	0.03206	−0.4600	0.1578	−0.3200	0.06607
Union	−1.9960	0.1701	−2.3602	0.2567	−2.1940	0.1860
Female	−1.3498	0.2642	0.6957	1.0650	−0.2378	0.4679

variables in the model.) If the number of weeks worked and the accepted wage offer are determined jointly, then ln *Wage* and u_{it} in this equation are correlated. We consider two instrumental variable estimators based on

$$\mathbf{Z}_1 = [1, Ind_{it}, Ed_i, Union_{it}, Fem_i]$$

and

$$\mathbf{Z}_2 = [1, Ind_{it}, Ed_i, Union_{it}, Fem_i, SMSA_{it}]$$

Table 12.1 presents the three sets of estimates. The OLS results are computed using the standard results in Chapters 3 and 4. One noteworthy result is the very small coefficient on the

The big point in this is that when we use OLS, and ignore the simultaneity, we find wages do not affect the number of weeks that people work, which seems, well, wrong.

Greene does not delve into the comparison of the two IV estimators, but there is an interesting difference. The first shows a larger wage effect, with a larger standard error, while the second shows a smaller effect with a much much smaller standard error.

Fish Markets

Imbens, Guido W. 2014. "Instrumental Variables: An Econometrician's Perspective", *Statistical Science* 29(3): 323–358.

Imbens describes several examples of IV. While studying impact of taxes on demand for fish, for example, the endogenous variables are the price of fish and the quantity supplied to the market. We need instruments to predict quantity that do not include the price of fish. "To identify the demand function, we look for determinants of the supply of whiting that do not affect the demand for whiting, and, similarly, to identify the supply function we look for determinants of the demand for whiting that do not affect the supply. In this specific case, Graddy (1995, 1996) assumes that weather conditions at sea on the days prior to market t, denoted by Z_t , affect supply but do not affect demand. Certainly, it appears reasonable to think that weather is a direct determinant of supply: having high waves and strong winds makes it harder to catch fish. On the other hand, there does not seem to be any reason why demand on day t, at a given price p, would be correlated with wave height or wind speed on previous days."(p. 13)

Instruments in wage/education models

Hahn and Hausman (2005) claim that the study of education as an endogenous predictor of wages was initiated by Griliches (1977). This is probably the most well known example of IV, and not a particularly persuasive example, I think, because there are so many competing, confusing claims.

Outcome variable is wage. Years of education is a predictor.

$$wage_i = \beta_0 + \beta_1 ed_i + \varepsilon_i$$

We suspect that there is an unmeasured variable, "ability" which contributes to both wages and years in education (people who are more able perhaps find school easier or more tolerable). Rather than (), we in fact are confronted by the unmeasured ability, which both affects wage and education

$$wage_i = \beta_0 + \beta_1 ed_i + \{\varepsilon_i + ability_i\}$$

As an instrument for ed_i, we need predictors that affect a person's level of education that do not also belong in the $wage_i$ equation. Card (1995) suggests distance from a person's home to nearest school as an instrument. One might use the educational achievements of the child's parents, or the IQ of the parents. In the review essay by Imbens (2014), there is a list of instruments for education that have been suggested. The key thing is that there are variables which seem not related to wages (employers don't pay more when students drive further to school, do they?).

The Hausman-Taylor model extends that IV approach by including fixed individual characteristics in repeated measures data. "Removing potentially correlated instruments has had a substantial effect: the point estimates change and their standard errors increase. All methods which control for correlation with the latent individual effects increase the schooling coefficient over those which do not; and this is certainly not the direction that many people concerned about ability bias would have expected."(1981, p. 1393).