Weak instruments: An overview and new techniques

Austin Nichols

July 21, 2006

Why Use IV?

Instrumental variables, often abbreviated IV, refers to an estimation technique used to address a variety of violations (collected under the general heading of endogeneity) of assumptions that guarantee that standard OLS estimates will be consistent, including:

- measurement error
- simultaneity (X affects y, and y affects X)
- omitted variables

Typically, the point of IV is to allow causal inference in a non-experimental setting.

Why Use IV?

Suppose one has a linear model of how y is related to X:

$$y = X\beta + \varepsilon$$

Here y is a $T \times 1$ vector of dependent variables, X is a $T \times K_1$ matrix of independent variables, β is a $K_1 \times 1$ vector of parameters to estimate, and ε is a $T \times 1$ vector of errors (capturing so-called "unexplained" variation in y). If $E[X'\varepsilon] \neq 0$ then the OLS estimate of β is biased and inconsistent.

How Use IV?

Using a $T \times K_2$ matrix of variables Z (called **the excluded instruments** or sometimes **the instrumental variables** or sometimes **the instruments**), correlated with X but not with ε, one can construct an IV estimator that will be a consistent estimator for β:

$$\hat{\beta}_{IV} = (X'P_zX)^{-1}X'P_zy$$

where P_z, the projection matrix of Z, is defined to be:

$$P_z = Z(Z'Z)^{-1}Z'.$$

- Two-stage Least Squares Two stage least squares (2SLS) is an instrumental variables estimation technique that is formally equivalent in the linear case.
  - Use OLS to regress X on Z and get $\hat{X} = Z(Z'Z)^{-1}Z'X$
  - Use OLS to regress y on $\hat{X}$ to get $\hat{\beta}_{IV}$.
- Ratio of Coefficients Another approach considers a different set of two stages, but this approach may only be used when there is one endogenous variable and one instrument.
  - Use OLS to regress y on X and get $\hat{\beta} = (X'X)^{-1}X'y$
  - Use OLS to regress y on Z and get $\hat{\pi} = (Z'Z)^{-1}Z'y$.

  Divide $\hat{\beta}$ by $\hat{\pi}$ to get $\hat{\beta}_{IV}$.
- The Control Function Approach The most useful approach considers another set of two stages.
  - Use OLS to regress X on Z and get estimated errors $\hat{\nu} = X - Z(Z'Z)^{-1}Z'X$
  - Use OLS to regress y on X and $\hat{\nu}$ to get $\hat{\beta}_{IV}$.

Forms of IV

- All of these approaches give the IV estimate of β, and each has its proponents for understanding the "intuition" of IV.
- For each, the reported standard errors for the second stage are wrong (not an issue for the one-step estimator that is used in practice).
- The advantage of the Control Function approach is that it works even outside the linear framework we are exploring here. For example, if the model is $y = \exp\left[X\beta + \varepsilon\right]$ you can still regress X on Z to get $\hat{\nu}$ and then use a Poisson regression to regress y on X and $\hat{\nu}$ to get $\hat{\beta}_{IV}$ (and then fix the standard errors). The other two approaches to constructing the linear IV estimates are not generalizable in the same way. See Wooldridge (2002).

Including Exogenous Regressors

All of the preceding assumes that the matrix of regressors X is composed entirely of potentially problematic variables (endogenous or measured with error). In fact, the usual model specification includes some variables that are not problematic, and some that are. For example, X will almost always include a constant (a vector of ones), which is neither endogenous nor measured with error. For conceptual reasons, we usually divide the set of regressors into two disjoint sets. We will refer to the potentially problematic regressors as the matrix Y and the exogenous variables as the matrix X.

The General Model for IV

The general model for IV is thus

$$y = Y\beta + X\gamma + u$$

where y is the dependent variable of interest, Y is an $N \times T$ matrix of problematic variables (or N endogenous variables), and X is a $K_1 \times T$ matrix of unproblematic variables, called the K_1 included instruments. Assume we have Z, a matrix of K_2 excluded instruments (sometimes called the "instrumental variables" when the meaning is clear), where $K_2 \geq N$, and we can write:

$$Y = Z\Pi + X\Phi + V$$

Note in particular that X and Z are identical for all endogenous variables we are instrumenting for. When the number N of endogenous variables is greater than one, there will be multiple equations to estimate in the "first stage" but we must always include the full set of exogenous variables in each equation.

Basic Assumptions

- Order Condition: When $N \leq K_2$, we say the system is identified, and when $N < K_2$ the system is overidentified.
- Rank Condition: $\mathrm{rank}(Z'Y) = N$.
- Z explains Y The regression of Y on Z produces coefficients $\Pi \neq 0$ (in population and sample).
- Z does not explain y

$$Cov(Z, u) = 0$$

  This says that Z (the set of instrumental variables) has no effect on y (the dependent variable) except through Y (the endogenous variables).

Asides

- Data mining in the "first stage"
- Estimated instrumental variables
- Polynomials in an endogenous variable: See Wooldridge (2002) on the "forbidden regression."
- Specification tests: See Baum, Schaffer, and Stillman (2003).

Why not always use IV?

- It's hard to find variables that meet the definitions of valid instruments: conceptually, most variables that have an effect on endogenous variables Y may also have a direct effect on the dependent variable y.
- The standard errors on IV estimates are likely to be larger than OLS estimates, and much larger if the excluded instrumental variables are only weakly correlated with the endogenous regressors.
- Bias. See also Kinal (1980) for related issue with small sample properties: the IV estimator may have no expected value.
- Interpretation. ATE, LATE, Random Coefficients. See Wooldridge (2002) Chapter 18.
- Weak Instruments. This set of problems is the focus of the rest of today's material.

Weak Instruments

- As pointed out by Bound, Jaeger, and Baker (1993; 1995), the "cure can be worse than the disease" when the excluded instruments are only weakly correlated with the endogenous variables.
  - IV estimates are biased in same direction as OLS, and Weak IV estimates may not be consistent.
  - With weak instruments, tests of significance have incorrect size, and confidence intervals are wrong.
- Staiger and Stock (1997) formalized the definition of "weak instruments" and most researchers seem to have concluded (incorrectly) from that work (or hearsay) that if the F-statistic on the excluded instruments in the first stage is greater than 10, one need worry no further about weak instruments.
- Stock and Yogo (2001; revised 2004) go into more detail and provide useful rules of thumb regarding the weakness of instruments based on a statistic due to Cragg and Donald (1993). Stock, Wright, and Yogo (2002) provide a summary of this work.

Weak Instruments

- Consider again the basic model

$$y = Y\beta + X\gamma + u$$

$$Y = Z\Pi + X\Phi + V$$

where y is the dependent variable of interest, Y is an $N \times T$ matrix of endogenous variables, Z is a matrix of K_2 excluded instruments and X is a matrix of K_1 included instruments.

- The concern is that the explanatory power of Z may be insufficient to allow inference on β. In this case, the first-stage statistic on the hypothesis $\Pi = 0$ may be bounded as $T \to \infty$, and various tests do not have correct size.

Weak Instruments: Diagnostics

- Historically, only informal rule-of-thumb diagnostics have been reported. These are reported by the Stata program **ivreg2.ado** (easily obtained while Stata is open by typing **ssc install ivreg2, replace** in the Command window) when the ffirst option is specified, and include the partial R^2 and first-stage F-statistics on excluded instruments.
- The newest version of **ivreg2.ado** incorporates additional code to compute eigenvalues of G_T before reporting other estimates. The minimum eigenvalue should be compared to Table 1 (to bound bias) or Table 2 (to bound size of Wald tests) in Stock and Yogo (2001; revised 2004).

Cragg-Donald statistic

- For notational compactness, let $P_W = W(W'W)^{-1}W'$ and $M_W = I - P_W$ for any matrix W, and let $W^{\perp}$ be the residuals from projection on X, so $W^{\perp} = M_X W$. Define $\underline{Z} = [XZ]$ to be the matrix of all instruments (included and excluded).
- One can construct the Cragg-Donald statistic as follows:

$$G_T = (Y'M_{\underline{Z}}Y)^{-1/2}Y^{\perp'}P_{Z^{\perp}}Y^{\perp}(Y'M_{\underline{Z}}Y)^{-1/2}\frac{(T-K_1-K_2)^2}{K_2}$$

where the matrix

$$Y^{\perp'}P_{Z^{\perp}}Y^{\perp} = (M_X Y)'M_X Z((M_X Z)'M_X Z)^{-1}(M_X Z)'(M_X Y)$$

- The minimum eigenvalue of G_T is the statistic used for testing for weak instruments.

- If $N = 1$ (there is one endogenous variable), the minimum eigenvalue of G_T is the F-statistic in the first stage regression, which Staiger and Stock (1997) suggest should be greater than 10.
- Looking at Table 1, if one used three excluded instrumental variables to instrument for a single endogenous variable (as in the returns-to-schooling regressions examined by Staiger and Stock), and one wanted to restrict the bias of the IV estimator to five percent of the OLS bias, the critical value of the first-stage F-statistic is 13.91.
- If one wanted Wald tests (of nominal size .05) of hypotheses about β to have size less than .1, the first-stage F-statistic should be greater than 22.3, according to Table 2 in Stock and Yogo (2001; revised 2004).

Tests and Confidence Sets Robust to Weak Instruments

- Anderson and Rubin (1949) propose a test of structural parameters (the AR test) that turns out to be robust to weak instruments (i.e. the test has correct size in cases where instruments are weak, and when they are not). Kleibergen (2002) proposes a Lagrange multiplier test, also called the score test, but this is now deprecated since Moreira (2003) proposes a Conditional Likelihood Ratio (CLR) test that dominates it.
- Andrews, Moreira, and Stock (2004a) and Mikusheva and Poi (2006) provide useful overviews of these alternatives.
- In theory, either the AR test or the CLR test can be inverted to produce a confidence region for the parameter β, but the AR test is much easier to work with.

AR Confidence Sets

Following Anderson and Rubin (1949), the simultaneous regression framework

$$y = Y\beta + X\gamma + u$$

$$Y = Z\Pi + X\Phi + V$$

can be rewritten (by subtracting $Y\beta_0$ from both sides) as:

$$y - Y\beta_0 = Z\theta + X\eta + \varepsilon$$

and all the assumptions of the linear regression framework are satisfied. Assuming $\Pi \neq 0$, a test of $\theta = 0$ is equivalent to a test of $\beta = \beta_0$ in this context, and seems to have correct size even in the presence of weak instruments.

AR Confidence Sets, cont.

We can construct an Anderson-Rubin (AR) confidence region for β as an N-dimensional set of values of β_0 for which we cannot reject $\theta = 0$. As discussed by Dufour and Taamouti (1999), this type of confidence region is robust to the presence of weak instruments and the accidental exclusion of relevant instruments, and allows valid inference about β. AR tests seem to have correct size under a wide variety of violations of the standard assumptions of IV regression.

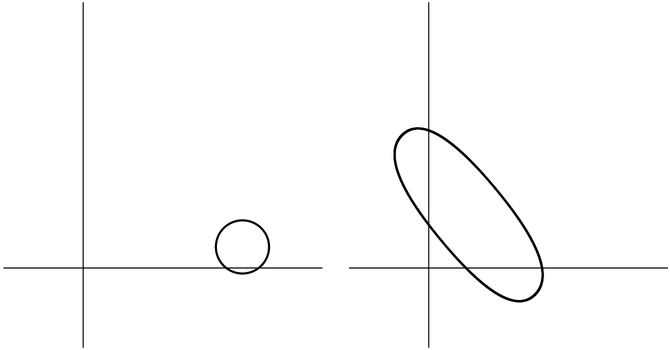
AR Confidence Sets, cont.

The AR confidence region has the unfortunate property that it need be neither bounded nor connected. In addition, constructing the region with any degree of accuracy is computationally intensive, and visual representation of the region can be quite cumbersome, whenever there is more than one endogenous variable. The AR confidence region may also be empty, or may not include the point estimate, in which cases the researcher may conclude that the model is not supported by the data. If the region is unbounded, the instrumental variables are simply too weak to conclude much about β.

AR Confidence Sets, cont.

- Assuming one has constructed the AR confidence region, any single hypothesis about β could be tested by comparing the hypothesized values of β to the region, and if the entire range of hypothesized values lay outside the confidence region, the hypothesis would be rejected.
- For example, if the AR confidence region for coefficients $\beta_{\mathbf{YRSED}}$ and $\beta_{\mathbf{IQ}}$ in an earnings equation were a disk in $\mathbf{R}^2$ whose boundary is given by the equation $(\beta_{\mathbf{YRSED}} - 8)^2 + (\beta_{\mathbf{IQ}} - 90)^2 = 100$, then the hypothesis that both coefficients are zero can be rejected, since the set does not include the origin. The hypothesis that $\beta_{\mathbf{YRSED}} = 0$ would not be rejected, however, since the set overlaps the axis where $\beta_{\mathbf{YRSED}} = 0$ (the $\beta_{\mathbf{IQ}}$ axis in this example).
- Non-spherical confidence regions are more interesting, and suggest the limitations of a projection-based method for constructing confidence intervals variable by variable.

AR Confidence Sets, cont.



AR Confidence Sets, cont.

- If you have two endogenous variables, and you want to construct an AR confidence region over a space of 100 values for each, this entails running 10,000 regressions and running 10,000 Wald tests, following the strategy outlined above. Even this is a coarse grid for graphing the confidence region. Indeed, a grid search will often not contain the confidence region, since the confidence region could be unbounded.
- But setting the test statistic equal to the critical value produces a quadric surface in β_0, and it is possible to graph "slices" or level curves of the surface, even for 3 or 4 (or more) dimensions, using Stata's graph programs. These level curves of the surface can currently be graphed in other software, but a seamless integration of the IV estimation and graphing of confidence regions is to be desired!

The Cutting Edge

- Moreira (2001) gives criteria for evaluating tests in the presence of weak instruments. Moreira (2003) proposes the Conditional Likelihood Ratio (CLR) test which Andrews, Moreira, and Stock (2004a) claim outperforms the AR test in power simulations. See also Andrews, Moreira, and Stock (2004b) and online supplements.
- Using some mathematical shortcuts proposed by Mikusheva (2005), Mikusheva and Poi (2006) provide Stata code (type **net from http://www.stata.com/users/bpoi/** and **net install condivreg** in Stata) to conduct both the CLR and AR tests, and to construct the corresponding confidence sets in the common case of **a single endogenous regressor**.
- Still plenty of work to be done here, at many levels.

Anderson, T. W. and H. Rubin (1949). "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations." *Annals of Mathematical Statistics*, 21, 570-582.

Andrews, Donald W. K.; Marcelo J. Moreira; and James H. Stock. (2004a). "Optimal Invariant Tests for Instrumental Variables Regression" NBER Technical Working Paper No. 299 with supplements at Stock's website.

Andrews, Donald W. K.; Marcelo J. Moreira; and James H. Stock. (2004b). "Performance of Conditional Wald Tests in IV Regression with Weak Instruments." *Journal of Econometrics*, forthcoming. Working paper version online with supplements at Stock's website.

Baum, Christopher F.; Mark E. Schaffer; Steven Stillman. (2003). "Instrumental variables and GMM: Estimation and testing." *Stata Journal*, StataCorp LP, vol. 3(1), 1-31. Working paper version available online.

Bound, John; David A. Jaeger; and Regina Baker. (1993). "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables." NBER Technical Working Paper No. 137.

Bound, John; David A. Jaeger; and Regina Baker. (1995). "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak." *Journal of the American Statistical Association*, 90(430), 443-450.

Cragg, J.G. and S.G. Donald. (1993). "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222-240.

Dufour, Jean-Marie (2003). "Identification, Weak Instruments, and Statistical Inference in Econometrics." *Canadian Journal of Economics*, 36, 767-808.

Dufour, Jean-Marie and Mohamed Taamouti. (1999; revised 2003) "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments." Unpublished manuscript, Department of Economics, University of Montreal.

Kinal, Terrence W. (1980). "The Existence of Moments of k-Class Estimators" *Econometrica*, 48(1), 241-250. Available via JSTOR.

Kleibergen, Frank. (2002). "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression." *Econometrica*, 70(5), 1781-1803. Available via JSTOR.

Mikusheva, Anna. (2005). "Robust Confidence Sets in the Presence of Weak Instruments." Unpublished manuscript.

Mikusheva, Anna and Brian P. Poi. (2006). "Tests and confidence sets with correct size in the simultaneous equations model with potentially weak instruments." Forthcoming in the *Stata Journal*. Working paper version available on Mikusheva's website

Moreira, Marcelo J. (2001). "Tests With Correct Size When Instruments Can Be Arbitrarily Weak" Working paper available online.

Moreira, Marcelo J. (2003). "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica*, 71 (4), 1027-1048. Working paper version available on Moreira's website.

Staiger, Douglas and James H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586. Available via JSTOR.

Stock, James H. and Motohiro Yogo (2001; revised 2004), "Testing for Weak Instruments in Linear IV Regression." Originally published as NBER Technical Working Paper No. 284; newer version available at Stock's website.

Stock, James H.; Jonathan H. Wright; and Motohiro Yogo. (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics*, 20, 518-529. Available from Yogo's website.

Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. Available from Stata.com.