Lecture 7: Instrumental Variables (Part II)

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

Recap

- IV is a powerful method to deal with bias in OLS
- IV relies on two conditions:
  - Relevance (which can be directly tested)
  - Validity (which is not directly testable)
- Under maintained IV assumptions, the IV estimator is consistent...
  - ...but still biased in finite samples
  - This bias declines with N and with the strength of the first-stage
  - When IVs are weak, even small violations of validity can cause large inconsistency

Plan for This Lecture

- We will extend discussion on the type of problems we were already discussing in the previous lecture
- The regressor of interest ("treatment") is potentially endogenous
- The regressor is assumed to have the same effect across the population
- The plan for today is to take a deeper look at further cases and diagnostics for IV estimation

Plan for today

1. Introduction
2. The Wald estimator
3. Multiple instruments
4. Multiple endogenous variables
5. Diagnostic tests for IV
   - F-statistics for weak instruments
   - Over-identification tests
   - Testing endogeneity of regressors
6. Applications of IVs for measurement error
7. Summary

Reviewing the Basic Concept of an IV

- Remember: the 2SLS coefficient is the ratio of the reduced form coefficient to the first-stage coefficient

- This becomes very simple when we have a single dummy IV

- Specifically, given the following causal model $Y_i = \alpha + \rho S_i + \eta_i$ where
  - S_i is suspected to be endogenous (i.e. $Cov(S_i, \eta) \neq 0$)
  - a dummy variable Z_i meets the requirements of an IV, then

$$\rho = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(S_i|Z_i = 1) - E(S_i|Z_i = 0)}$$

- This estimator is called the "Wald estimator" and is just a ratio of two differences-in-means
  - It shows how the IV estimate "scales up" the reduced form coefficient
  - This intuition helps a lot in applying IV methods to RCTs and RDDs

An Example of the Wald Estimator (from Angrist and Pischke 2009)

TABLE 4.1.2
Wald estimates of the returns to schooling using quarter-of-birth instruments

	(1) Born in 1st Quarter of Year	(2) Born in 4th Quarter of Year	(3) Difference (Std. Error) (1) – (2)
ln (weekly wage)	5.892	5.905	–.0135 (.0034)
Years of education	12.688	12.839	–.151 (.016)
Wald estimate of return to education			.089 (.021)
OLS estimate of return to education			.070 (.0005)

Notes: From Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930–39 birth cohorts in the 1980 census 5 percent file. The sample size is 162,515.

Multiple Instruments

- What do we do when we have more than one exogenous variable that can be excluded from the structural equation?

- Let us first look at the case with a single endogenous variable:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

  with y_2 endogenous and z_1 exogenous

- Suppose we have two exogenous variables z_2 and z_3 that can be excluded from the structural equation

- The assumptions that z_2 and z_3 are uncorrelated with u_1 and are not in the structural equation are called **exclusion restrictions**

How Do We Most Efficiently Use the Information in z_2 and z_3?

- If z_2 and z_3 are correlated with y_2 we could use either as an instrument

- However, neither IV estimator would be efficient

- Since each of z_1, z_2 and z_3 is uncorrelated with u_1, any linear combination is also uncorrelated with u_1 and is therefore a valid IV

- To find the best IV, we choose the **linear combination** that is most strongly correlated with y_2

- This is given by the reduced form equation for y_2 (which we call the first stage):

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$$

- The best IV for y_2 is the following linear combination of the z_j:

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3$$

The First-Stage Equation

- For this IV not to be perfectly correlated with z_1 we need

$$\pi_2 \neq 0 \text{ or } \pi_3 \neq 0 \text{ or both}$$

- The value of π_1 is irrelevant, but if $\pi_2 = \pi_3 = 0$ then the structural equation is not identified

- We can (and should) test $H_0 : \pi_2 = 0$ and $\pi_3 = 0$ using an F test

The First-Stage Equation

- We do not know the population parameters π_j for $j = 0, 1, 2, 3$, so we cannot simply compute y_2^*
- We can obtain estimates of π_j for $j = 0, 1, 2, 3$ by regressing y_2 on z_1, z_2 and z_3
- Using estimates $\widehat{\pi}_j$, we obtain the fitted values:

$$\widehat{y}_2 = \widehat{\pi}_0 + \widehat{\pi}_1 z_1 + \widehat{\pi}_2 z_2 + \widehat{\pi}_3 z_3$$

2SLS

- Once we have $\widehat{y}_2$, we can use it as an instrument for y_2
- In this case, the IV FOCs are

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_{i1}-\widehat{\beta}_0-\widehat{\beta}_1 y_{i2}-\widehat{\beta}_2 z_{i1}\right)=0$$

$$\frac{1}{n}\sum_{i=1}^{n}z_{i1}\left(y_{i1}-\widehat{\beta}_0-\widehat{\beta}_1 y_{i2}-\widehat{\beta}_2 z_{i1}\right)=0$$

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{i2}\left(y_{i1}-\widehat{\beta}_0-\widehat{\beta}_1 y_{i2}-\widehat{\beta}_2 z_{i1}\right)=0$$

The 2SLS Estimator in Matrix Notation

- We estimate the first-stage regression using OLS:

$$\widehat{\pi} = (Z'Z)^{-1}Z'X$$

  where Z is the matrix of instruments (excluded instruments and the exogenous variables from the structural equation) and X is the vector of endogenous X

- Form predicted values of $X = (x_1, ..., x_n)$

$$\widehat{X} = Z\widehat{\pi} = Z(Z'Z)^{-1}Z'X$$

The 2SLS Estimator in Matrix Notation

- Use the predicted $\widehat{X}$ instead of X in the structural equation to get

$$\widehat{\beta}_{2sls} = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\mathbf{y}$$

- Plugging in the value of $\widehat{X}$ from earlier and simplifying:

$$\begin{aligned}
\widehat{\beta}_{2sls} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\mathbf{y} \\
&= (Z(Z'Z)^{-1}Z'X)^{-1}Z(Z'Z)^{-1}Z'\mathbf{y} \\
&= (\widehat{X}'X)^{-1}\widehat{X}'\mathbf{y}
\end{aligned}$$

- One useful feature of the expression above is that when $X = Z$, then $\widehat{X} = X$ and we see that the 2SLS estimator is the same as the OLS estimator

Consistency of the 2SLS Estimator

$$
\begin{aligned}
\widehat{\beta}_{2sls} &= (\widehat{X}'X)^{-1}\widehat{X}'\mathbf{y} \\
&= (\widehat{X}'X)^{-1}\widehat{X}'(X\beta + \mathbf{u}) \\
&= (\widehat{X}'X)^{-1}\widehat{X}'X\beta + (\widehat{X}'X)^{-1}\widehat{X}'\mathbf{u} \\
&= \beta + \left(\frac{\widehat{X}'X}{N}\right)^{-1}\left(\frac{\widehat{X}'\mathbf{u}}{N}\right)
\end{aligned}
$$

- We will take probability limits of this expression to show consistency

Consistency of the 2SLS estimator

$$plim\widehat{\beta}_{2sls} \quad = \quad \beta + plim\left(\frac{\widehat{X}'X}{N}\right)^{-1} plim\left(\frac{\widehat{X}'\mathbf{u}}{N}\right)$$

- $\underset{n\to\infty}{plim}\left(\frac{\widehat{X}'X}{N}\right)$ exists and is invertible (non-singular), provided the instrument is relevant
- $\underset{n\to\infty}{plim}\left(\frac{\widehat{X}'u}{N}\right) = 0$ if the predicted $\widehat{X}$ and the structural error are uncorrelated (i.e. the instrument is valid)
- Thus, under the IV assumptions, $plim\beta_{2sls} = \beta$

2SLS with Multiple Endogenous Variables

- You may have more than one endogenous variable in your structural equation
- It is possible to use 2SLS to estimate the structural parameters, but you need to fulfill two conditions:
  1. The order condition
  2. The rank condition

The Order Condition

- We need as many instruments excluded from the structural equation as endogenous variables
- Specifically, given the following structure:
  - The structural equation is $y_i = x_i'\beta + u_i$ with $E(x_i u_i) \neq 0$
  - $x_i' = (1, x_{2i}, x_{3i}, ..., x_{Ki})$ is the $1 \times K$ row vector of explanatory variables
  - $z_i' = (1, z_{2i}, z_{3i}, ..., z_{Ki})$ is a $1 \times L$ row vector which contains the intercept, all exogenous variables in the structural equation and all excludable instruments
- Then the **order condition is that** $L \geqslant K$
- This is a necessary condition for identification, but not sufficient

The Rank Condition

- The sufficient condition for identification is called the **rank condition**

$$\text{The } L \times K \text{ matrix } E(z_i x_i) \text{ has full column rank}$$
$$\text{rank } E(z_i x_i) = K$$

- The rank condition can equivalently be expressed as *rank* $\Pi = K$ in the complete system of first stage equations

$$\underset{N\times K}{X} = \underset{N\times L}{Z}\ \underset{L\times K}{\Pi} + \underset{N\times K}{R}$$

- Intuitively, what the condition requires is that there is a first-stage relationship for every variable in the structural equation

- In the special case of one endogenous variable and one instrument, this is the same as the relevance condition

Weak Instruments

- In the last lecture, we discussed the problem of "weak instruments" i.e. when the correlation between the excluded instrument(s) and the endogenous variable(s) is very small
- It increases the finite sample bias in IV estimates
- Even modest violations of the exclusion restrictions can lead to substantial inconsistency
- So a key issue is: how do we identify weak instruments?
- The statistic that we use to evaluate this is the F-statistic for the joint significance of all regressors in the first-stage regression

Weak Instruments

- In a simple causal model linking outcomes (y_i) to an endogenous regressor (x_i), for which we have a vector of Q instruments ($\mathbf{z_i}$), and no covariates, it can be shown that:

$$E(\widehat{\beta}_{2SLS} - \beta) \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \left[\frac{E(\pi' Z' Z \pi)/Q}{\sigma_{\xi}^2} + 1 \right]^{-1}$$

  where η is the error term from the second stage and ξ is the error term from the first stage and π is a vector of first-stage coefficients on Z

Weak Instruments

- Define

$$F = \frac{E(\pi' Z' Z \pi)/Q}{\sigma_\xi^2}$$

- Then

$$E(\widehat{\beta}_{2SLS} - \beta) \approx \frac{\sigma_{\eta\xi}}{\sigma_\xi^2}\left[\frac{1}{F+1}\right]$$

- As the F-statistic gets bigger the finite-sample bias gets smaller

Weak Instruments

- The F-statistic is bigger, the higher the proportion of variance in x that your instruments can explain—hence it is a diagnostic to guard against weak instruments

- Note that the F-statistic is inversely proportional to the number of instruments (Q)
  - Additional instruments are only worth putting in if they add substantial more information
  - In practice, always report "just-identified" IV estimates with your strongest IVs in addition to any additional specifications which are "over-identified"

- Always report your first-stage and the F-statistic!

- How big is big? There isn't a theorem ...

- In practice, an F-statistic above 10 is taken to be sufficiently strong
  - Recent research suggests that this might be too optimistic
  - There are some options that work better for weak-IV inference

Weak Instruments (Angrist and Pischke 2009)

TABLE 4.6.2
Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	.105	.435	.089	.076	.093	.091
	(.020)	(.450)	(.016)	(.029)	(.009)	(.011)
LIML	.106	.539	.093	.081	.106	.110
	(.020)	(.627)	(.018)	(.041)	(.012)	(.015)
F-statistic (excluded instruments)	32.27	.42	4.91	1.61	2.58	1.97
Controls						
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, age squared		✓		✓		✓
Excluded instruments						
Quarter-of-birth dummies	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The age and age squared variables measure age in quarters. The OLS estimate corresponding to the models reported in columns 1–4 is .071; the OLS estimate corresponding to the models reported in columns 5 and 6 is .067. Data are from the Angrist and Krueger (1991) 1980 census sample. The sample size is 329,509. Standard errors are reported in parentheses.

First-Stage F-statistics in Stata

- We will be using the command `ivreg2` in Stata to estimate IV specifications
- The precise F-statistic to be used differs a bit depending on the specified model:
  - Single endogenous regressor: the "Angrist-Pischke F-statistic", which we just saw
  - Multiple endogenous regressors + conditional homoskedasticity: "Cragg-Donald F-statistic"
  - Multiple endogenous regressors + conditional heteroskedasticity: "Kleibergen-Paap F-statistic"
- Stata will calculate the appropriate F-statistic for you

Over-Identification Tests

- **More IV jargon**: An IV specification is said to be...
  - "just-identified" when the number of excluded instruments equals the number of endogenous regressors
  - "over-identified" when you have more excluded instruments than endogenous regressors

- In the latter case, in a constant effects model, it is possible to indirectly test for validity of the instruments

Over-Identification Tests

- Suppose we have a single endogenous variable but two informative candidate instruments
- If the IVs are valid, in a constant effects model, either of them should get you consistent estimates of the same parameter
- We can then estimate two (different) just-identified estimates of the parameter of interest and test whether they are the same
- If each just-identified estimator is consistent, the distance between them should be small relative to sampling variance and should shrink as the sample size increases
- This is called an "over-identification test" (Hansen-Sargant test)

Over-Identification Tests

- If you have over-identified specifications, you should report the over-identification test
- Usually it will be the Hansen J-statistic, since you will have heteroskedasticity
- The null of the test is that all instruments are valid (under the maintained assumption that at least one instrument is valid + we have constant causal effects)
- If you reject the null (i.e., get a very small *p*-value), you might be in trouble...
- Although with heterogeneous potential outcomes (next week), we will see interpreting this is not strightforward

Over-Identification Tests

- If you fail to reject the null, this is not proof that your IVs are all valid—remember that you “fail to reject the null”, you do not “accept”!
- Both the rejection of the null, or the failure to reject, are consistent with issues unrelated to identification
- Angrist and Pischke (2009):

  “In our experience, the “over-ID statistic” is of little value in applied work. Since IV estimates are often imprecise, we cannot take much satisfaction [...] that one estimate is within the sampling variance of another. On the other hand, [...], the fact that the overid test rejects need not point to an identification failure. Rather, this may be an effect of treatment effect heterogeneity.”

Tests for Endogeneity of the Regressor(s): Hausman Test Statistic

- The last test that we will look at is a test of whether an IV was necessary at all
- Remember, we only have a suspicion that in the second-stage, x is correlated with the error term
  - If, in fact, x was not correlated with the error term, there would be no bias in OLS
  - Specifically, while both OLS and IV are consistent, OLS will be more precise
  - With homoskedasticity, OLS is efficient (but potentially biased) while IV is consistent
- It is possible to test whether the estimates are different from each other beyond sampling variation
  - The null is that the regressor can be treated as exogenous
  - Under homoskedasticty, this is a Hausman test statistic
  - If you fail to reject it, this is perhaps a sign that bias in OLS was not very large

Tests for Endogeneity of the Regressor(s): Hausman Test Statistic

- It is possible to test whether the estimates are different from each other beyond sampling variation
- The null is that the regressor can be treated as exogenous
- Under homoskedasticty, this is a Hausman test statistic
- If you fail to reject it, perhaps a sign that bias in OLS was not very large
- Although, again, many other potential explanations
- Your IV is rubbish, maybe?

Using IVs to Correct for Measurement Error Bias

- Both the Angrist and Krueger (1991) and Acemoglu, Johnson, and Robinson (2001) examples (that we saw in the previous lecture) are applications of IV to deal with omitted variables bias
- This is perhaps the most-frequent use of IV methods in economics
- But the IV method is more general and offers neat solutions for other problems too
- One particular example is measurement error in regressors
- You should recall the bias from classical error-in-variables in Lecture 5
- IV methods provide a very neat solution to these issues

A Stylized Example

- Let the causal relationship of interest be:

$$y_i = \alpha + \beta x_i^* + A_i^{'}\gamma + \epsilon$$

- Say, y_i is test scores and x_i^* is the average time spent in a week doing homework

- A_i is some vector of controls which make true x_i^* exogenous in the regression such that $E(\epsilon | x_i^*, A_i) = E(\epsilon) = 0$

A Stylized Example

$$y_i = \alpha + \beta x_i^* + A_i^{'} \gamma + \epsilon$$

- y_i is collected through a cognitive test given to students
- x_i^* is collected by asking students about hours spent doing homework in the last week (x_{1i})
- We are concerned that students may not accurately recall how long they spent doing homework
- So $x_{1i} = x_i^* + e$ where e is some measurement error, not correlated with true x_i
- NB. For this example, we are abstracting from potential omitted variable bias issues, which would also be certain here

An IV Solution to Attenuation Bias

- We know that the following regression will be biased:

$$y_i = \alpha + \beta x_{1i} + A_i^{'}\gamma + v$$

  with the estimated $\widehat{\beta}$ being biased towards zero

- What we need is an IV for x_{1i} such that this IV is correlated with x_{1i} and not correlated with e

- Suppose we have another measure of the student's time spent doing homework (x_{2i}) collected by asking their parent

- Under certain assumptions, this can be an IV for x_{1i} in the regression above

An IV Solution to Attenuation Bias

- We can see this in the following case:

$$\begin{aligned} x_{1i} &= x_i^* + e \\ x_{2i} &= x_i^* + u \end{aligned}$$

- We can predict x_{1i} based on x_{2i} and the vector of controls and use predicted $\widehat{x_{1i}}$ in the second-stage:

$$x_{1i} = \pi_0 + \pi_1 x_{2i} + A_i^{'} \pi_3 + \xi$$
$$y_i = \alpha + \beta \widehat{x_{1i}} + A_i^{'} \gamma + \mu$$

- Assuming relevance ($\pi_1 \neq 0$) and validity ($Cov(e, u) = 0$), we will get a consistent estimator for β

Summary

- Extending the IV estimator to include additional covariates or to multiple instruments is easy; relevance and validity are key
- You can also extend the estimator to multiple endogenous variables
  - You need to satisfy the rank condition
  - Each endogenous variable needs identifying variation from the excluded instruments
- You should always report the first-stage and always report the (appropriate) first-stage F-statistics
- If the equation is over-identified, report the over-identification statistic
- You may wish to present the Hausman statistic (or the version appropriate for heteroskedasticity)
- You should always present the OLS as well

Readings

- Angrist and Pischke (2009): Chapter 4 up till Section 4.1.2 (p. 133)
- Wooldridge (2013): Chapter 15 (except Section 15.7 on time-series applications)
- Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. Journal of Experimental Criminology, 2(1), 23-44.
- Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. The Journal of Economic Perspectives, 20(4), 111-132.

Appendix

Derivation of the consistency of IV

$$
\begin{aligned}
\widehat{\beta}_{2sls} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\mathbf{y} \\
&= [(Z(Z'Z)^{-1}Z'X)'(Z(Z'Z)^{-1}Z'X)]^{-1}(Z(Z'Z)^{-1}Z'X)'y \\
&= [(X'Z(Z'Z)^{-1}Z')(Z(Z'Z)^{-1}Z'X)]^{-1}(X'Z(Z'Z)^{-1}Z')y \\
&= [X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X)]^{-1}(X'Z(Z'Z)^{-1}Z')y \\
&= [X'Z(Z'Z)^{-1}Z'X)]^{-1}(X'Z(Z'Z)^{-1}Z')y
\end{aligned}
$$

using the symmetry of $(Z'Z)^{-1}$ such that$[(Z'Z)^{-1}]' = (Z'Z)^{-1}$

Appendix

Derivation of the consistency of IV

$$
\begin{aligned}
\widehat{\beta}_{2sls} &= [X'Z(Z'Z)^{-1}Z'X)]^{-1}(X'Z(Z'Z)^{-1}Z')y \\
&= [(Z(Z'Z)^{-1}Z'X)'X]^{-1}(Z(Z'Z)^{-1}Z'X)'y \\
&= (\hat{X}X)^{-1}\hat{X}'y
\end{aligned}
$$

where the last line uses that $(Z(Z'Z)^{-1}Z'X) = \hat{X}$

The remainder of the proof follows as in the main slides