Econometrics II

Lecture 6: Instrumental Variables

Konrad Burchardi

Stockholm University

18th of April 2024

Literature

1. **"Mostly Harmless Econometrics"**, Angrist and Pischke
   Chapter 4.1-4.3, 4.6.1, 4.6.4

These notes draw on those books. All mistakes are mine.

Plan for Today

Introducing IV

Take standard regression framework[1]:

$$y = \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{X} = [\mathbf{X_1}, \mathbf{X_2}]$. Worried about exogeneity of $\mathbf{X_1}$.

Valid **Instrumental Variables** yield consistent estimates of β.

- ...in the presence of measurement error in $\mathbf{X_1}$;
- ...in case of endogeneity of regressors, $\mathbb{E}[\varepsilon|\mathbf{X_1}] \neq 0$.

How does this work? And what are valid instruments?

[1] Assume constant treatment effect. Later will talk about IV with treatment effect heterogeneity.

Introducing IV

We require some 'instruments' $\mathbf{Z_1}$ such that:

1. *Relevance*: $plim \frac{1}{N}(\mathbf{Z_1'X_1}) \neq 0$
2. *Exogeneity*: $plim \frac{1}{N}(\mathbf{Z_1'}\varepsilon) = 0$

Then $\hat{\beta}_{IV} = (\mathbf{Z'X})^{-1}\mathbf{Z'}y$ is consistent estimator of β, where $\mathbf{Z} = [\mathbf{Z_1}, \mathbf{X_2}]$:

$$\begin{aligned}\text{plim } \hat{\beta}_{IV} &= \text{plim } \left[\left(\frac{1}{N}\mathbf{Z'X}\right)^{-1} \frac{1}{N}\mathbf{Z'}(\mathbf{X}\beta + \varepsilon)\right] \\ &= \beta + \left[\text{plim } \left(\frac{1}{N}\mathbf{Z'X}\right)^{-1} \times \text{plim } \left(\frac{1}{N}\mathbf{Z'}\varepsilon\right)\right] = \beta,\end{aligned}$$

where we use $plim \frac{1}{N}(\mathbf{Z_1'X_1}) \neq 0$ [*Relevance*] and $plim \frac{1}{N}(\mathbf{Z_1'}\varepsilon) = 0$ [*Exogeneity*].

Generalized IV and 2SLS

1. **Generalized IV**
   The optimal choice of instruments $\mathbf{Z}$ is $\mathbf{P_Z X}$.[2]
   (Note: $\mathbf{X_2}$ is optimally instrumented with $\mathbf{X_2}$.)
   The estimator is called 'generalized IV', defined as:

$$\hat{\beta}_{GIV} = (\mathbf{X}'\mathbf{P_Z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P_Z}y.$$

2. **Two-Stage Least-Spuares (2SLS)**
   Given that P_Z is idempotent and symmetric, $\hat{\beta}_{GIV}$ is numerically equivalent to:

$$\hat{\beta}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'y,$$

   where $\hat{\mathbf{X}} = \mathbf{P_Z X}$.[3]

[2] In the least asymptotic variance sense.
[3] Proof of consistence works also for $\hat{\beta}_{GIV}$ and hence $\hat{\beta}_{2SLS}$.

Plan for Today

Anatomy of IV Formula

- The 2SLS formula shows, we can calculate IV estimator in two steps:
  1. Regress $\mathbf{X}$ on $\mathbf{Z}$ to obtain predicted values $\hat{\mathbf{X}}$
  2. Regress y on $\hat{\mathbf{X}}$.

  Intuitive interpretation: 'only exploit variation in $\mathbf{X}$ driven by the instrument'.

- Meaning *Relevance* Condition:
  1. $\mathbf{Z_1}$ needs to impact $\mathbf{X_1}$ (conditional on $\mathbf{X_2}$).
  2. At least as many instruments as endogenous variables.

  → Without it cannot estimate effect of $\mathbf{X_1}$ on y.

- Meaning *Exogeneity* Condition[4]:
  1. $\mathbf{Z_1}$ is determined 'like an experiment' (instrument is external)...
  2. and $\mathbf{Z_1}$ affects y *only* through $\mathbf{X_1}$ (instrument is excludable).

  → Without it do not solve original problem.

[4] Sometimes called "*Exclusion Restriction*" or "*Identifying Assumption*". **Fundamentally not testable!**

Intuition for IV

1. Find variables $\mathbf{Z_1}$ that...

   Relevance: "shock" $\mathbf{X_1}$, but...

   Exogeneity: ...are unrelated to y otherwise.

   Then we see how y changes when $\mathbf{X_1}$ is shocked!

2. Only exploit variation in $\mathbf{X_1}$ that we "know to be exogenous".

3. Idea much like an **experiment**:

   Shock the explanatory variable, rather than finding more controls!

First Stage, Reduced Form and Second Stage

"**First Stage**" is the (OLS) regression of each element of $\mathbf{X_1}$ on $\mathbf{Z}$.

This tells us how the instruments impact the endogenous variable.

Key to test Condition 1!

"**Reduced Form**" is the (OLS) the regression of y on $\mathbf{Z}$.

This tells us how the instrument is related to outcomes.

(Excludability not necessary.)

"**Second Stage**" is the regression of y on $\hat{\mathbf{X}}$.

This tells us how exogenous changes in $\mathbf{X_1}$ and $\mathbf{X_2}$ impact y.

IV is Reduced Form over First Stage

In case of one endogenous variable and one regressor, we can write

$$\beta_{IV} = \frac{Cov(y_i, z_i)}{Cov(x_i, z_i)} = \frac{Cov(y_i, z_i)/V(z_i)}{Cov(x_i, z_i)/V(z_i)}$$

Sample analogue is called **Indirect Least Squares** estimator.
IV estimate is ratio of the reduced form over the first stage coefficient![5]

Two-Sample IV (Angrist and Krueger, 1992):
To calculate IV estimator requires only $\frac{1}{N_A}\mathbf{Z}'\mathbf{X}$ and $\frac{1}{N_B}\mathbf{Z}'\mathbf{y}$. These might come from different samples (from the same population), so $\mathbf{X}$ and y need not be in same data set.

Split-Sample IV (Angrist and Krueger, 1995), more efficient:
Find first coefficient in sample A, $(\mathbf{Z_A'Z_A})^{-1}\mathbf{Z_A'X_A}$ and calculate IV estimate in sample B as regression y_B on $\mathbf{Z_B}(\mathbf{Z_A'Z_A})^{-1}\mathbf{Z_A'X_A}$. Adjust standard errors (Inoue and Solon, 2010).

[5] Mathematical fact, also with $\mathbf{X_2}$. If your results do not satisfy it, you did something wrong.

Simple Case: Wald Estimator

Take the case of a single dummy instrument z_i and one endogenous regressor x_i

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Using $\mathbb{E}[\varepsilon_i | z_i] = 0$, it follows that $\mathbb{E}[y_i | z_i] = \alpha + \beta \mathbb{E}[X_i | z_i]$ and:

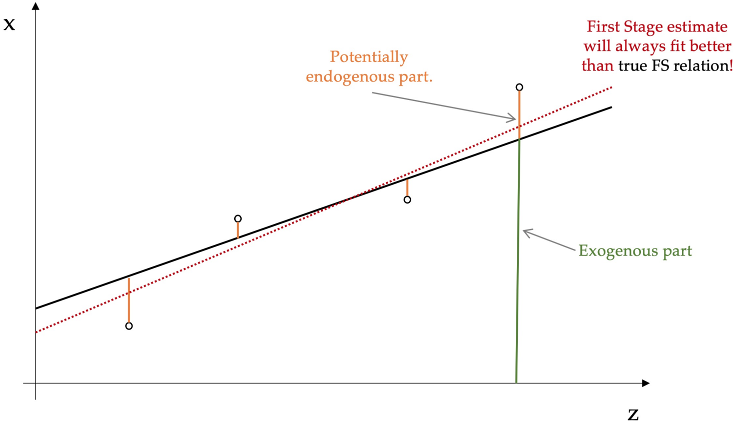
$$\beta = \frac{\mathbb{E}[y_i | z_i = 1] - \mathbb{E}[y_i | z_i = 0]}{\mathbb{E}[x_i | z_i = 1] - \mathbb{E}[x_i | z_i = 0]}$$

Population analogue of **Wald Estimator**.

Intuition?

Experiments: IV in ‘encouragement designs’, or with imperfect compliance.

Why is IV only consistent, but not unbiased?



OLS overfits the First Stage in small samples. [Problem Set 3]

But variance around true First Stage effect decreases with sample size.

Consistent but not Unbiased

What can be done about it?

1. Test how big problem (likely) is. Test *Relevance* condition!

$$\mathbb{E}[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\varepsilon}}{\sigma_\eta^2}\left[\frac{\mathbb{E}[\pi'\mathbf{Z}_1'\mathbf{Z}_1\pi]/q}{\sigma_\eta^2} + 1\right]^{-1}$$

   where $x = \mathbf{Z}_1\pi + \eta$ is the First Stage, end $\mathbb{E}[\pi'\mathbf{Z}_1'\mathbf{Z}_1\pi]/q$ is the First Stage 'population F-statistics' on the excluded instruments (not $\mathbf{X}_2$).

   - Finite sample bias of IV inversely related to "strength" of instruments; as rule of thumb: with First Stage F-statistics < 10,[6] instruments were considered 'weak' (Staiger and Stock, 1997); see also Young (2022).
   - If instruments are useless, bias as large as OLS.
   - If you add useless instruments, F-statistic falls and bias increases.
   - With multiple instruments: KP/AP test of differential variation.

2. Correct for the degree of bias: FIML estimator (less efficient with strong instr.)

[6] And then? 1. Drop weak instruments; 2. Get better instruments; 3. LIML/JIV; 4. New project.

IV and Classical Measurement Error

With classical measurement error, where $x_{1i}^* = x_{1i} + v_i$:

$$\text{plim}\ [\hat{\beta}_1] = \beta_1 \frac{Var(x_{1i})}{Var(x_{1i}) + Var(v_i)} \equiv \beta_1 \lambda$$

Now consider you have additionally another measure of x_{1i}:

$$z_i = x_{1i} + \xi_i, \text{ with } Cov(v_i, \xi_i) = 0$$

Then the reduced form and first stage identify

$$\gamma_1 = \frac{Cov(y_i, z_i)}{Var(z_i)} = \beta_1 \frac{Var(x_{1i})}{Var(x_{1i}) + Var(\xi_i)};\ \pi_1 = \frac{Cov(x_{1i}^*, z_i)}{Var(z_i)} = \frac{Var(x_{1i})}{Var(x_{1i}) + Var(\xi_i)}$$

Therefore $\beta = \frac{\gamma_1}{\pi_1}$, i.e. IV estimator identifies β_1, not $\beta_1 \lambda$.

Often if $\beta_{2SLS} > \beta_{OLS}$ in absolute value - against the readrers' expectations- authors conclude: 'IV solved measurement error'.

Plan for Today

Getting Standard Errors Right

There is a temptation to calculate the 2SLS estimator by:

1. running the First Stage as OLS regression of $\mathbf{X}$ on $\mathbf{Z}$;
2. calculate the predicted values $\hat{\mathbf{X}}$;
3. running the Second Stage as OLS regression of y on $\hat{\mathbf{X}}$.

This will provide you with the correct $\hat{\beta}_{2SLS}$ (discussed above).
However the standard errors will be wrong! Should be (without proof)

$$y - \mathbf{X}\hat{\beta}_{2SLS},$$

but in the above procedure your statistical package will calculate them as

$$y - \hat{\mathbf{X}}\hat{\beta}_{2SLS}.$$

Getting First Stage Right

Rewriting the Second Stage we get:

$$y = \hat{\mathbf{X}}_{\mathbf{1}}\beta_1 + \mathbf{X}_{\mathbf{2}}\beta_2 + (\mathbf{X}_{\mathbf{1}} - \hat{\mathbf{X}}_{\mathbf{1}})\beta_1 + \varepsilon$$

Note that:

1. $\mathbf{X}_{\mathbf{2}}$ is uncorrelated of ε (by assumption);
2. $\mathbf{X}_{\mathbf{2}}$ is uncorrelated of $\mathbf{X}_{\mathbf{1}} - \hat{\mathbf{X}}_{\mathbf{1}}$ (by construction);
3. $\hat{\mathbf{X}}_{\mathbf{1}}$ is linear combination of $[\mathbf{Z}_{\mathbf{1}}, \mathbf{X}_{\mathbf{2}}]$, asymp. uncorrelated of ε (by assumption);
4. $\hat{\mathbf{X}}_{\mathbf{1}}$ is uncorrelated of $\mathbf{X}_{\mathbf{1}} - \hat{\mathbf{X}}_{\mathbf{1}}$ (by construction).

Together these imply we can consistently estimate β.

Failure to include $\mathbf{X}_{\mathbf{2}}$ in the First Stage means (2) breaks down.
Failure to run linear First Stage means (2), (4) and (3) might break down.

Interpreting R^2 in Second Stage

The R^2 in Second Stage [when displayed] is not meaningful.

- Residuals are calculated, correctly, as $y - \mathbf{X}\hat{\beta}_{2SLS}$. The RSS might be larger than TSS, and hence $R^2 < 0$.
- The point of the Second Stage is *not* to fit y to $\mathbf{X}$, but solely to estimate $\hat{\beta}$.

What is (somewhat) meaningful is the R^2 in the Reduced Form.

Basic Mistakes in Typical IV Paper

In my (limited) experience the most common drawbacks of IV papers are:

1. Authors present an instrument that is plausibly external, but might impact y through multiple channels; authors highlight one channel.

   To save project: Is Reduced Form interesting?

2. Authors do not critically assess plausibility of exclusion restriction.

Plan for Today

Discussing Exogeneity

Relevance condition can be tested (see above).

Exogeneity condition can fundamentally not be tested.

1. Need to argue, *using understanding of the world*, that it is satisfied.
2. Might provide 'balance' tests, demonstrating that **Z** is unrelated to baseline variables that might impact y.
3. Might provide 'placebo' tests, demonstrating that **Z** has no impact on pseudo outcomes, outcomes which it should not impact.

Order of Identification

With number of instruments in $\mathbf{Z_1}$...

1. ...greater than number of variables in $\mathbf{X_1}$, model is 'over-identified'.

   Efficient to use all instruments, if they are relevant.

2. ...equal to number of variables in $\mathbf{X_1}$, model is 'exactly identified'.
3. ...less than number of variables in $\mathbf{X_1}$, model is 'under-identified'.

Overidentification Tests

In the over-identified case, can calculate Sargan-Hansen/Sargan's J test:

$$J(\hat{\beta}) = N\frac{\hat{\varepsilon}' P_Z \hat{\varepsilon}}{\hat{\varepsilon}\hat{\varepsilon}}$$

Under H_0 that *Exogeneity* is satisfied for all $\mathbf{Z}$, this is χ^2-distributed.

- Empirically straight-forward to implement as M times R^2 of regression of Second Stage residuals on all elements of $\mathbf{Z}$.
- Intuition is: in case *Exogeneity* is not satisfied for some instruments, they will be correlated with $\hat{\varepsilon}$ and H_0 is rejected.
- Omnibus test: Tells you something is wrong, not what.
- With effect heterogeneity, different instruments identify different causal effects, and test is no longer useful.

Plan for Today

Shift-Share / Bartik Instruments

$$z_l = \sum_n s_{ln} g_n,$$

a weighted sum of

- common shocks, varying at level $n = 1, \ldots, N$,
- weighted by exposure shares, varying at level of outcome $l = 1, \ldots, L$.

Shift-Share: Examples

Bartik (1991) and Blanchard and Katz (1992):
instrument region l's labor demand, where g_n is the national growth of industry n and $s_{ln} \in [0, 1]$ are lagged employment shares; and study impact on wages.

Autor, Dorn and Hanson (2013):
instrument Chinese import competition in location l, where g_n is growth in Chinese exports in manufacturing industry n to 8 non-U.S. countries, and s_{ln} are lagged employment shares; study impact on manufacturing employment / unemployment.

Shift-Share / Bartik Instruments

Recent Developments

1. Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022) “Quasi- Experimental Shift-Share Research Designs” *Review of Economic Studies*
2. Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020) “Bartik Instruments: What, When, Why and How” *American Economic Review*

Questions?

References

1. Angrist and Pischke (2008): Chapter 4.1-4.3, 4.6.1, 4.6.4.
2. Angrist, Joshua and Alan B. Krueger (1992) “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples” Journal of the American Statistical Association 87: 328–336.
3. Angrist, Joshua and Alan B. Krueger (1995) “Split-Sample Instrumental Variables Estimates of the Return to Schooling” Journal of Business and Economic Statistics 13: 225–235.
4. Autor, David, David Dorn, and Gordon Hanson (2013) “The China Syndrome: Local Labor Market Effects of Import Competition in the United States.” American Economic Review 103 (6): 2121–68.
5. Bartik, Timothy (1991 “Who Benefits from State and Local Economic Development Policies?” W.E. Upjohn Institute.
6. Blanchard, Olivier, and Lawrence Katz (1992) “Regional Evolutions” Brookings Papers on Economic Activity 23(1): 1–76.
7. Borusyak, Kirill, Peter Hull and Xavier Jaravel (2022) “Quasi-Experimental Shift-Share Research Designs” Review of Economic Studies 89(1): 181–213.

References

8 Goldsmith-Pinkham, Paul, Isaac Sorkin, Henry Swift (2020) “Bartik Instruments: What, When, Why, and How” American Economic Review 110(8): 2586–2624.

9 Inoue, Atsushi, and Gary Solon (2010) “Two-Sample Instrumental Variables Estimators” The Review of Economics and Statistics 92(3): 557–561.

10 Staiger, Douglas, and James Stock (1997) “Instrumental Variables Regressions with Weak Instruments” Econometrica 65(3): 557–586.

11 Young, Alwyn (2022) “Consistency without Inference: Instrumental Variables in Practical Application” European Economic Review 147: 104–112.