Instrumental Variable Estimation II

Lingguo Cheng

Nanjing University

December 1, 2025

Outline

1. Specification Tests for IV Estimation

2. Bartik Instrument

Specification Test I: Weak IV

- We refer to a variable as a ***weak instrument*** in the sense that one or more of the first-stage regressions have a poor fit.
- When IV is weak, the asymptotic theory can provide a poor guide to actual finite-sample distribution, even if the sample has thousands of observations.
- That is, even when IV estimators are consistent, they are biased in finite samples.

What is a weak instrumental variable

- Give a population regress function

$$y = \beta_0 + \beta_1 x_1 + u$$

- Use z_1 as an instrument for x_1, that is, z_1 satisfies:

$$\mathrm{Cov}(z_1, u) = 0 \quad \text{and} \quad \mathrm{Cov}(z_1, x_1) \neq 0$$

- The plim of the IV estimator is easily shown to be:

$$\operatorname{plim} \hat{\beta}_1 = \beta_1 + \mathrm{Cov}(z_1, u)/\,\mathrm{Cov}(z_1, x_1)$$

Proof: Remember that (ILS Estimator)

$$\beta_1 = \frac{\operatorname{Cov}(y, z_1)}{\operatorname{Var}(z_1)} \Big/ \frac{\operatorname{Cov}(x_1, z_1)}{\operatorname{Var}(z_1)} = \frac{\operatorname{Cov}(y, z_1)}{\operatorname{Cov}(x_1, z_1)}$$

Plugging in $y = \beta_0 + \beta_1 x_1 + u$, we have

$$\beta_1 = \frac{\operatorname{Cov}(y, z_1)}{\operatorname{Cov}(x_1, z_1)} = \beta_1 + \operatorname{Cov}(z_1, u) / \operatorname{Cov}(z_1, x_1)$$

Therefore, probability limit of IV estimator:

$$\operatorname{plim} \hat{\beta}_1 = \beta_1 + \frac{\operatorname{Cov}(z_1, u)}{\operatorname{Cov}(z_1, x_1)} = \beta_1 + \frac{\sigma_u}{\sigma_{x_1}} \cdot \frac{\operatorname{Corr}(z_1, u)}{\operatorname{Corr}(z_1, x_1)}$$

Where $\operatorname{corr}(\cdot, \cdot)$ denotes correlation.

Consequences of Weak IV Estimation

1. If z_1 and u are correlated, the inconsistency in the IV estimator gets arbitrarily large if $\mathrm{Corr}(z_1, x_1)$ gets close to zero.
   - Thus, if z_1 is only weakly correlated with x_1, seemingly small correlations between z_1 and u can cause severe inconsistency—and therefore severe finite sample bias.
2. In fact, it may be better to just use OLS. Let $\tilde{\beta}_1$ denote the OLS estimator, then:

$$\operatorname{plim} \tilde{\beta}_1 = \beta_1 + \frac{\mathrm{Cov}(x_1, u)}{\mathrm{Var}(x_1)} = \beta_1 + \frac{\sigma_u}{\sigma_{x_1}} \mathrm{Corr}(x_1, u)$$

Bound et al. (1995)

- What was the cost of adding instruments without predictive power?
- Adding more weak instruments causes the first-stage F statistic to approach zero and increase the bias of 2SLS.
- Bound et al. (1995) studied this empirically, replicating Angrist and Krueger (1991), and using simulations.

Effect of completed schooling on men's log weekly wages.

Independent variable	**OLS**	**2SLS**	**OLS**	**2SLS**	**OLS**	**2SLS**
Years of schooling	0.063 (0.000)	0.142 (0.033)	0.063 (0.000)	0.081 (0.016)	0.063 (0.000)	0.060 (0.029)
First stage *F*		13.5		4.8		1.6
Excluded instruments						
Quarter of birth		Yes		Yes		Yes
Quarter of birth × year of birth		No		Yes		Yes
Number of excluded instruments		3		30		28

Note: Standard errors in parentheses. First stage is quarter of birth dummies.

How to test a weak IV

- Consider the model

$$y_{1i} = \mathbf{y}_{2i}\beta_1 + \mathbf{x}_{1i}\beta_2 + u_i, \;\; i = 1, \ldots, N$$

- $\mathbf{y}_2$ is a vector of m endogenous variables. $\mathbf{x}_2$ is a vector of K_1 exogenous vectors. Assume the existence of m IV $\mathbf{x}_2$ for $\mathbf{y}_2$ that satisfies the assumption that E$(u_i|\mathbf{x}_2)$=0. Assume each component y_{2j} of $\mathbf{y}_2$ satisfies the first-stage equation

excluded effect.

$$y_{2ji} = \mathbf{x}_{1i}\pi_{1j} + \mathbf{x}_{2i}\pi_{2j} + v_{ji}, \;\; j = 1, \ldots, m$$

- The F statistics for joint significance of the instruments $\mathbf{x}_2$ in first-stage regression of the endogenous regressor $\mathbf{y}_2$ on $\mathbf{x}_1$ and $\mathbf{x}_2$. This is a test that $\pi_2 = 0$.

Case 1: One Endogenous Variable

- H_0: the instruments are weak versus H_1: the instruments are strong
- **Case 1: one endogenous variable, no less than one instrumental variable**
- Staiger and Stock (1997) suggested a rule of thumb that an F statistic of less than 10 indicates weak instruments. The rule of thumb is ad hoc.

Case 2: More than one endogenous variable

- **Stock and Yogo (2005)**
- The test statistic is the minimum eigenvalue of a matrix analog of the F statistic (equals the F value statistic if there is just one endogenous regressor). It is proposed originally by **Cragg and Donald (1993)**.
- A low minimum eigenvalue indicates weak instruments.
- **To determine the critical values:**
  - First choose b, the largest relative bias of the 2SLS estimator relative to OLS, that is acceptable
  - The test critical value varies with:
    - Number of endogenous regressors (m)
    - Number of exclusion restrictions (K)
  - Critical value: $cv(b, m, K)$; Critical values are only available when the model has at least two identifying restrictions
- Stata syntax:

```
estat firststage [, forcenonrobust all]
```

Specification Test II:Testing for Endogeneity: Durbin-Wu-Hausman Test

- **Durbin (1954), Wu (1973) and Hausman (1978)**
- **H0**: all elements of $\mathbf{x}$ are exogenous
- The test is based on the difference $\hat{\beta}_{\text{2SLS}} - \hat{\beta}_{\text{OLS}}$. If all elements of $\mathbf{x}$ are exogenous (and $\mathbf{z}$ is also exogenous - a maintained assumption), then 2SLS and OLS should differ only due to sampling error. To determine whether this is so, we need to estimate the asymptotic variance of $\sqrt{N}(\hat{\beta}_{\text{2SLS}} - \hat{\beta}_{\text{OLS}})$
- Under the null hypothesis $E(\mathbf{x}'u) = 0$ and the appropriate **homoskedasticity assumption**:

$$\text{Avar}[\sqrt{N}(\hat{\beta}_{\text{2SLS}} - \hat{\beta}_{\text{OLS}})] = \sigma^2[E(\mathbf{x}^*\mathbf{x}^*)]^{-1} - \sigma^2[E(\mathbf{x}'\mathbf{x})]^{-1}$$

- The **DWH statistics**:

$$(\hat{\beta}_{\text{2SLS}} - \hat{\beta}_{\text{OLS}})' \left[(\hat{X}\hat{X})^{-1} - (XX)^{-1}\right]^{-1} (\hat{\beta}_{\text{2SLS}} - \hat{\beta}_{\text{OLS}})/\hat{\sigma}^2_{\text{OLS}}$$

Comments on DWH Statistic

- A more serious drawback about DWH statistics is that it is not robust to heteroskedasticity. A robust variance matrix estimator for $\text{Avar}[\sqrt{N}(\hat{\beta}_{\text{SLS}} - \hat{\beta}_{\text{OLS}})]$ can be and should be obtained.
- Infrequent use of robust DWH statistic may be partially due to misunderstandings of when the principle of comparing estimators applies. It is commonly thought that one estimator should be asymptotically efficient under the null hypothesis, but it is not necessary—indeed it is essentially irrelevant.

Specification Test III: Overidentification test

- When we have more instruments than we need to identify an equation, we can test whether the additional instruments are valid in the sense that they are uncorrelated with u_1.
- **Hansen-Sargan test**
- H_0: all instruments are valid
- H_1: at least one of the instruments is not valid

Tests of overidentifying restrictions

After obtaining $\hat{\beta}_{OGMM}$, then

$$Q(\hat{\beta}) = \left\{ \frac{1}{N}(\mathbf{y} - X\hat{\beta})'Z \right\} \hat{S}^{-1} \left\{ \frac{1}{N} Z'(\mathbf{y} - X\hat{\beta}) \right\}$$

If the population moment conditions

$$E\left\{ Z'(\mathbf{y} - X\beta) \right\} = 0$$

are correct, then $Z'(\mathbf{y} - X\hat{\beta}) \approx 0$, so $Q(\hat{\beta})$ should be close to zero. Under the null hypothesis that all instruments are valid, it can be shown that $Q(\hat{\beta})$ has an asymptotic chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions.
Large values of $Q(\hat{\beta})$ lead to rejection of $H_0 : E\left\{ Z'(\mathbf{y} - X\beta) \right\} = 0$

Overidentification Test: Comments

- **Comments:**

  一般至多一个内生性变量和一个IV.

  - Rejection means at least one of the validity of the overidentifying instruments is not valid.
  - It is possible that rejection of H0 indicates that the model for the conditional mean is misspecified.
- **Stata syntax:**

```
estat overid
```

Outline

1. Specification Tests for IV Estimation

2. Bartik Instrument

Popular IV Design

- The lottery design– The randomized trial case.
- Bartik instruments
- Peer effect

Bartik Instrument

- Also known as **shift-share instruments**, were named after Timothy Bartik, popularized in Blanchard and Katz (1992)
- The idea behind a Bartik instrument is to measure the change in a region's labor demand due to changes in the national demand for different industries' products.
- Consider the model:

$$y_{l,t} = \alpha + \delta I_{l,t} + \rho X_{l,t} + \varepsilon_{l,t}$$

Where:

- $y_{l,t}$ is log wages in location l (e.g., Detroit) in time period t (e.g., 2000) among native workers
- $I_{l,t}$ are immigration flows in region l at time period t; $X_{l,t}$ are controls that include region and time fixed effects, among other things
- The parameter δ is some average treatment effect of the immigration flows' effect on native wages

ATE

Endogeneity Problem

- The problem: it is almost certainly the case that immigration flows are highly correlated with the disturbance term such as the time-varying characteristics of location l (e.g., changing amenities) [Sharpe, 2019]
- Reverse causality
- The Bartik instrument is created by interacting initial "shares" of geographic regions, prior to the contemporaneous immigration flow, with national growth rates.
- The deviations of a region's growth from the US national average are explained by deviations in the growth prediction variable from the US national average. And deviations of the growth prediction variables from the US national average are due to the shares because the national growth effect for any particular period is the same for all regions.

Bartik Instrument

- We can define the Bartik instrument as follows:

$$B_{l,t} = \sum_{k=1}^{K} \underbrace{\overset{\text{share}}{z_{l,k,t_0}}}_{\text{外生的初始状态}} \overset{\text{shift}}{m_{k,t}}$$

  where:
  - z_{l,k,t_0} are the "initial" t_0 share of immigrants from source country k (e.g., Mexico) in location l (e.g., Detroit)
  - $m_{k,t}$ is the change in immigration from country k (e.g., Mexico) into the US as a whole
  - The first term is the ***share variable*** and the second is the ***shift variable***
- The predicted flow of immigrants, B, into destination l (e.g., Detroit) is then just a weighted average of the national inflow rates from each country in which weights depend on the initial distribution of immigrants.

Shifts vs Shares

- There are two perspectives as to what is needed to leverage a Bartik design to identify a causal effect and they separately address the roles of the exogeneity of the shares versus the shifts. Which perspective you take will depend on the ex-ante plausibility of certain assumptions. They will also depend on different tools.
- Goldsmith-Pinkham et al. (2020) explain the share perspective. They show that while the shifts affect the strength of the first stage, it is the initial shares that provide the exogenous variation.
- They write that "the Bartik instrument is 'equivalent' to using local industry shares as instruments, and so the exogeneity condition should be interpreted in terms of the shares."

Shift Perspective

- Borusyak et al. (2019) explain the shift perspective: Temporal shocks may provide exogenous sources of variation.
- They show that exogenous independent shocks to many industries allow a Bartik design to identify causal effects regardless of whether the shares are exogenous so long as the shocks are uncorrelated with the bias of the shares.
- Otherwise, it may be the shock itself that is creating exogenous variation, in which case the focus on excludability moves away from the initial shares and more towards the national shocks themselves.