Financial Econometrics II – Cross Section and Panel Data

Instrumental Variables

Andreas Fuster
Swiss Finance Institute @ EPFL

SFI Léman PhD program – 2025, Lecture 2

Overview

- **The problem**
- What are the assumptions when doing IV?
- How is IV implemented?
- Other issues with IV
- Examples

Motivation

- Consider the following single equation linear model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- Assume that:
  - $Cov(x_1, u) = \cdots = cov(x_{k-1}, u) = 0$
  - $Cov(x_k, u) \neq 0$

- **All of the regression coefficients will be biased**, except if x_k is uncorrelated with the other regressors (unlikely). Then only β_k will be biased.

Instrumental variables

- One way to get around this problem is to find an **instrumental variable (IV)** (call it z) for the endogenous regressor x_k.
- We can think of the variable x_k as having "good" and "bad" variation.
  - "Good" variation is **not** correlated with u
  - "Bad" variation is correlated with u

- An IV is a variable that explains variation in x_k, but does not explain variation in y.
  - It only explains the **"good" variation**
- We can use the IV to extract the "good" variation and replace x_k with only that component.

Overview

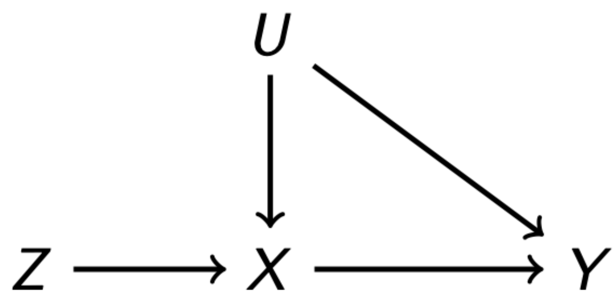
- The problem
- **What are the assumptions when doing IV?**
- How is IV implemented?
- Other issues with IV
- Examples

Instrumental variables

- Instrumental variables need to satisfy two conditions:

  1. **Relevance condition**

  2. **Exclusion restriction**



  – What are these two conditions?
  – Which is harder to satisfy?
  – Can we test whether they are true?

- Let's first think about the situation with one endogenous regressor and one instrument.

Relevance condition

- The **relevance condition** requires that the partial correlation between the **instrument z** and the endogenous variable **not be zero.**

- The coefficient γ in the regression

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + \epsilon$$

does not equal zero.

- What it means is that z is relevant in explaining the problematic regressor, x_k, **after netting out the effects of all other exogenous variables** of the original model.

Relevance condition

- This condition is empirically testable. How?

- Run the regression of x_k on all the other x's and the IV z to see if z explains x_k.
  - In this regression, the coefficient estimate of z should be statistically different from zero – ideally t-stat > 4 or so (cf. later).

- This is what people call the **"first stage"** of the IV estimation.

- Even though the relevance condition is formally testable, we should also have a good **economic argument** of why the instrumental variable is relevant in explaining the variable x_k.

Exclusion restriction

- Take the original model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- The **exclusion restriction** requires that $cov(z, u) = 0$.
- The name derives from the **exclusion** of the instrument from the original equation.

- z is uncorrelated with the error term u
  - z has no explanatory power with respect to y after conditioning on the other explanatory variables.
  - **The only role that the instrument z plays in influencing the outcome y is through its effect on the endogenous variable x_k.**

Exclusion restriction

- **Problem: the exclusion restriction cannot be tested!**
- The error term, u, is unobservable.
- How then can we check that the instrument is good?
- **We cannot formally test this condition!**
- We must find a **convincing economic argument** as to why the exclusion restriction is not violated.

Exclusion restriction

- Sometimes we see the following "support" for the exclusion restriction:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma z + u$$

  — Idea: if $\gamma = 0$, then the exclusion restriction likely holds, i.e. z does not explain y after conditioning on other x's.

- Problem?
  — If $cov(x_k, u) \neq 0$, we still get biased estimates.
  — And if we believe that the relevance condition holds, then the coefficient on z is certainly biased too.

➡ **don't run this "test"**

Good instruments?

- Good instruments often come from biological or physical events or features.
- Sometimes from institutional changes, as long as the economic question under study was not one of the reasons for the institutional change in the first place.
- Only way to find a good instrument is to understand the economics of the question at hand.
- Ask the following question: "**Does the instrumental variable affect the outcome variable y only via its effect on the endogenous regressor x_k?**"
  - If the answer to this question is no, then the instrument is likely to violate the exclusion condition, and the estimates will be inconsistent and biased.

“Good instruments should feel weird”

- Cunningham, section 7.2.2: “*Let’s say you think you have a good instrument. How might you defend it as such to someone else? A necessary but not sufficient condition for having an instrument that can satisfy the exclusion restriction is if people are confused when you tell them about the instrument’s relationship to the outcome.*”

- So it shouldn’t be obvious why z should affect y
  — until you explain the path via x_k

- Example he gives: gender of a family’s first two children affects women’s labor supply (in particular, BB/GG vs. BG)
  — why would that be?

A Nobel instrument

The Prize in Economic Sciences 2024

The Royal Swedish Academy of Sciences has decided to award the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2024 to

Daron Acemoglu
Massachusetts Institute of Technology, Cambridge, USA

Simon Johnson
Massachusetts Institute of Technology, Cambridge, USA

James A. Robinson
University of Chicago, IL, USA

- Most famous paper: "The Colonial Origins of Comparative Development: An Empirical Investigation" (AER 2001)
- Instrument z : mortality rate of the first European settlers in the countries they colonized (100+ years ago)
- Outcome y : countries' GDP "today" (1990s)
- What is the (potentially endogenous) x of interest?

Overview

- The problem
- What are the assumptions when doing IV?
- $\boxed{\textbf{How is IV implemented?}}$
- Other issues with IV
- Examples

Implementing IV estimation

- Given a good instrument, how can we consistently estimate the parameters of the main model?
- The most intuitive approach is to estimate a **two-stage least squares** (2SLS) model:
- **First stage:** regress x_k on other x's and z.
- **Second stage:** take **predicted x_k from first stage** and use it in the original model instead of x_k.

First stage

- In the first stage, estimate the following

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma z + \epsilon$$

Endogenous regressor

All other **exogenous** regressor

Instrument

- Get estimates of the α's and γ

- Calculate predicted values $\hat{x}_k$:

$$\hat{x}_k = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \cdots + \hat{\alpha}_{k-1} x_{k-1} + \hat{\gamma} z$$

Second stage

- In the second stage, use the predicted values to estimate

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k \hat{x}_k + u$$

Predicted values

- **Intuition from earlier:** extract the "good" variation in x_k
  - Predicted values represent variation in x_k that is "good" in that it is driven only by factors that are uncorrelated with u
  - Predicted values are linear function of variables that are uncorrelated with u

Reduced form

- Commonly also report "reduced form" estimates.
- The "**reduced form**" estimation is when you regress y directly onto the instrument, z, and the other exogenous x's. The endogenous variable x_k is excluded from this regression.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1} + \delta z + u$$

- It is a consistent estimate of the **effect of $\boldsymbol{z}$ on $\boldsymbol{y}$** (if $cov(z, u) = 0$) presumably through the channel of z's effect on x_k.
  - sometimes called "intent-to-treat" effect, esp. when x_k is binary

Reduced form

- It can be shown that the IV estimate for x_k, $\hat{\beta}_k^{IV}$, is given by

$$\hat{\beta}_k^{IV} = \frac{\hat{\delta}}{\hat{\gamma}}$$

Reduced form coefficient estimate for z

First stage coefficient estimate for z

- If you do not find an effect of z on y in the reduced form model, the second-stage regression will most likely not be significant, at least not "robustly" so (cf. Anderson-Rubin test below)
  — always good to check reduced form model.

Estimation in practice

- Rather than estimating the two stages manually, we usually let the **software** do it, as doing it manually may lead to wrong standard errors
  - For example, **ivreg2, ivregress, ivreghdfe** in Stata
  - Although doing it manually is fine for Anderson-Rubin test (later)
- All **the exogenous** x**'**s need to be included in the first stage (including year and firm fixed effects). Otherwise estimates are not consistent.
- Always report the **first-stage results**
  - It is a test of the relevance condition, and gives indications regarding a potential weak instrument problem.
  - In particular, report F-statistic on instrument(s) and perhaps also the partial R-sq of the instrument(s) – more below

Overview

- The problem
- What are the assumptions when doing IV?
- How is IV implemented?
- **Other issues with IV**
- Examples

More instruments

- Nothing restricts the **number of instruments** to just one.
- Any variable satisfying both relevance and exclusion condition is a **valid instrument**.
- In the case there are multiple instruments $z = (z_1, \dots, z_m)$, the **relevance condition** can be tested with an **F-test** of the joint null hypothesis that $\gamma_1 = 0, \dots, \gamma_m = 0$ against the alternative that at least one γ coefficient is non-zero in the model

$$x_k = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \gamma_1 z_1 + \cdots + \gamma_m z_m + \epsilon$$

- The **exclusion restriction** requires the correlation between **each instrument** and the error term u to be zero.

More endogenous variables

- Nothing restricts the **number of endogenous variables** to just one.
- Consider the model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + \cdots + \beta_{k+h-1} x_{k+h-1} + u$$

  — $(x_1, \ldots, x_{k-1})$ are the $k-1$ **exogenous** regressors
  — $(x_k, \ldots, x_{k+h-1})$ are h **endogenous** regressors

- We must have at least as many instruments as endogenous regressors for the coefficients to be identified.
  – If the number of IVs matches the number of problematic regressors, the model is said to be **"just identified"**.

More endogenous variables

- The **exclusion restriction is unchanged**: none of the instruments is correlated with u
- The relevance condition is similar in spirit except now there is a system of relevance conditions corresponding to the system of endogenous variables.
  - Each first stage (there will be h of them) must have at least one instrument with non-zero coefficient (based on F-test)
  - Of the m instruments, there must be at least h of them that are partially correlated with problematic regressor
- Models with more instruments (m) than endogenous variables (h) are said to be **overidentified** and there are $(m - h)$ overidentifying restrictions.

Two problematic regressors

- Consider the model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

  - $cov(x_1, u) = \cdots = cov(x_{k-2}, u) = 0$
  - $cov(x_{k-1}, u) \neq 0$
  - $cov(x_k, u) \neq 0$

- There are now **two problematic regressors** x_k and x_{k-1}.

Two problematic regressors

- We need an IV for each problematic regressor (e.g. z_1 and z_2).

- Then estimate 2SLS as before:

  – Regress x_{k-1} on all other x's (except the endogenous x_k) and **both** instruments z_1 and z_2.

  – Regress x_k on all other x's (except the endogenous x_{k-1}) and **both** instruments z_1 and z_2.

  – Get predicted values for second stage.

IVs with interactions

- Take the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

- $Cov(x_1, u) = 0$ and $Cov(x_2, u) \neq 0$
- Both x_2 and $x_1 x_2$ are problematic (endogenous).
- What if you can only find one IV, z? Can you still get consistent estimates?
  - **YES!** Use z as instrument for x_2 and $x_1 z$ as instrument for $x_1 x_2$.
  - (in practice all instruments are there for all endog. variables)

Binary variables

- If x_k is binary, do <u>not</u> run a probit (or logit) in the first-stage and then directly use the predicted $\hat{x}_k$ in the second stage
  - Angrist and Pischke call this the "forbidden regression" (because neither the conditional expectations operator nor the linear projection carry through nonlinear functions)
  - Instead, can either run linear probability model in first stage, or (potentially more efficient) run a probit and then use the predicted values as *instrument* for x_k

- If y is binary, can use IV-Probit, though 2SLS linear probability model is usually fine as well
  - for more detail, see Wooldridge 15.7.2-3

Overidentified model

- With more instruments, you can in principle extract more “good” variation from the first stage of the estimation. If #instruments > #endog. variables, can also do a test of IV validity.
- **Overidentification test** (aka Sargan test). Intuition:
  - Estimate model for all subsets of instruments that provide exact identification.
  - Are the estimates ~the same across the different subsets?
  - If the results are similar across different subsets, it suggests the IV’s are OK.
- **Limitation:** The test assumes that at least one instrument is valid, but it is left unspecified which. Also not always a very powerful test.
- And note: Finding ONE good instrument is sufficiently difficult so that it is rare to find SEVERAL good instruments.
- → **There is no test to prove an IV is valid.** We always have to motivate an IV (whether it satisfies the exclusion restriction) based on economic theory and intuition.

Weak instruments

- Weak instruments are instruments that are only **weakly correlated** with the endogenous regressor.
- This can lead to coefficient bias in finite samples and wrong t-statistics.
- **Bias from weak instruments can be severe**. Angrist and Pischke (pp. 205-8): if $y = \beta x + \eta$ is causal model of interest, and $x = \pi z + \xi$ the first stage, then bias of 2SLS:

$$E(\widehat{\beta_{2sls}} - \beta) \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}} \frac{1}{F+1}$$

where first ratio corresponds to OLS bias when π = 0 and second ratio goes to 1 as F-stat (from first stage) goes to 0.

Weak instruments

- How can we detect weak instruments?

1. **Large standard errors** in second stage of IV estimation
   — The variance of an IV estimator depends inversely on the covariance between the instrument and the endogenous variable

2. **Low F-statistic** from the first stage (on the excluded instruments – not overall first-stage regression!)
   — Indication of low correlation between instruments and endogenous regressors
   — Stock and Yogo (2002) commonly used rule of thumb: instruments "not weak" if this F-statistic is above 10. (**But**: reliance on this threshold is evolving, see below)

Which F-statistic?

- Andrews, Stock and Sun (2019) recommend that (with one endogenous regressor), researchers use “effective F-statistic” of Montiel Olea and Pflueger (2013)
  — “weakivtest” in Stata

- Equivalent to the conventional first-stage F-statistic for testing $\pi = 0$ in models with homoskedastic errors, but adds a multiplicative correction in models with non-homoskedastic errors.
  — Stock-Yogo tests (reported e.g. by Stata’s “estat firststage”) not applicable in non-homoskedastic case → most applications

Weak instruments & inference

- There has been an intense & ongoing debate in recent years about the reliability of inference/hypothesis tests in just-identified IV models. Will highlight three contributions:

- 1) Lee et al. (AER 2022): smoothly adjust t-statistic based on F-stat from first stage (for single-IV case) to get correct inference ("*tF* adjustment"). E.g. for 5% significance level:
  - F=10: multiply standard error by 1.75
  - F=30: multiply standard error by 1.20
  - only at F>=105, usual t-tests "correct"
  - Looking at 61 AER papers, "among the specifications for which *F* > 10 and |*t*| > 1.96, the use of *tF* adjustment would cause about 1/4 of the specifications to be statistically insign. at the 5% level."
  - see https://irs.princeton.edu/davidlee-supplementarytF for FAQs

Weak instruments & inference

- 2) Keane and Neal (JEconometrics, 2023) conduct straightforward simulations to show that:
  — 2SLS standard errors tend to be small exactly in samples where the 2SLS estimate is biased toward the OLS estimate
  — This leads to "power asymmetry": 2SLS t-tests will be most likely to reject null hypothesis when 2SLS is shifted toward OLS; very little power to detect a true effect of opposite sign to OLS

- They argue that **Anderson-Rubin test,** which does not have such power asymmetry, should be adopted in lieu of the t-test even when the first-stage F-statistic is well above 10.
  — See also Andrews, Stock and Sun (2019): "Anderson-Rubin confidence sets" are robust to weak identification and efficient (in a specific sense) in just-identified case

Anderson and Rubin (1949(!))

- Consider $y = \beta x + u$, with $x = \pi z + e$ where $\mathrm{cov}(e, u) \neq 0$
- **A-R test** is simply based on reduced form regression:
  $$y = \beta\pi z + (\beta e + u) \equiv \xi z + \nu$$
- Given that a valid instrument requires $\pi \neq 0$, a test of the null hypothesis $\xi = 0$ provides an alternative way to test $\beta = 0$
  - with single instrument: simple t-test of coefficient in reduced form

- To get **A-R confidence sets**: find the max and min β_0 values such that the excluded instrument is significant at exactly the 5% level in the regression of $(y - \beta_0 x)$ on all exog. variables
  - e.g. “weakiv” in Stata
  - If the instrument is weak, the A-R confidence set may be unbounded (on one or both sides)

Keane and Neal (ARE 2024)

- “A Practical Guide to Weak Instruments” – highly recommended!
- Their recommended procedure:

First, consider the case in which outcome y is regressed on the single endogenous variable x and an exogenous control variable c, and where z is the excluded instrument. We suggest reporting results from the following procedures.

1. Run and report OLS. It is important to compare 2SLS and OLS results.
   In Stata: `reg y x c, vce(`*type*`)`
2. Run the first-stage regression of x on z and c.
   In Stata: `reg x z c, vce(`*type*`)`
3. Obtain a heteroskedasticity/cluster-robust F statistic for significance of the instrument.
   In Stata: `test z = 0`
4. Compute $\hat{x}$.
   In Stata: `predict xhat, xb`
5. Run the second-stage regression of y on $\hat{x}$ and use the t-test from this regression to test $H_0{:}\beta = 0$. This t-stat is the AR test of $H_0{:}\beta = 0$ in the one instrument case.
   In Stata: `reg y xhat c, vce(`*type*`)`
6. Construct a valid confidence interval by inverting the AR test.
   In Stata: `weakiv ivregress 2sls y (x = z) c, vce(`*type*`)`

In all these commands the “vce” option determines how the variance matrix of the parameter estimates is calculated. The results are rendered robust to heteroskedasticity or clustering via the option one specifies for “*type*,” which refers to the data type—for example, vce(cluster personid) for panel data. Enter `help reg` in Stata for a list of options.

Weak instruments & inference

- 3) Lee et al. (2023) – authors of the "*tF* adjustment" above – propose a different adjustment that they argue is more efficient than the Anderson-Rubin procedure. They call it ***VtF***.
- Idea: incorporate information from $\hat{r}$, the empirical correlation between the residuals in the main equation and first stage.
- This yields shorter confidence intervals (and thus more significance) than A-R intervals
- They also show that this rule works: "if F >10+100× $\hat{r}$, use the usual '1.96' conf. intervals; otherwise use the VtF intervals"

- See https://irs.princeton.edu/davidlee-supplementVTF for FAQ
- But note: paper has not been published yet – so would stick to Anderson-Rubin for now (i.e., follow Keane-Neal guidance)

Weak instruments & inference – other considerations

- In cases with more instruments than endog. variables (over-identified case), can estimate with “LIML” or “CUE” – less biased in finite samples

- Separately, if you have many but relatively weak instruments, can combine them in an “optimal” way using machine-learning-type methods in the first stage – e.g. LASSO
  - See papers by Belloni, Chernozhukov and Hansen
  - For intro, see https://medium.com/teconomics-blog/machine-learning-meets-instrumental-variables-c8eecf5cec95

- More common advice: case with multiple weak instruments is (even) more complicated – better to use single best instrument and report just-identified case (Angrist and Pischke)

Limitations of IV

- Finding a **good** instrument is **extremely difficult**.
  — Always pay enough attention to the exclusion restriction!
  — And robustness of statistical inference

- Even with a good IV, **external validity** may be a concern:
  — **Internal validity**: estimation strategy successfully uncovers a causal effect in the empirical context's setting.
  — **External validity:** Are the estimates predictive of other outcomes in other scenarios?
  — IV estimates only tell us about subsample where the instrument is predictive (we only make use of variation in x explained by z and ignore the effect of x for observations where z does not affect x).

What is the IV strategy identifying?

- Consider a setting with a binary treatment (e.g. whether a startup was in some ‘accelerator’ scheme)

- Allow for the treatment effects to potentially differ across units i, and denote by Y_i^T i’s **potential outcome** under treatment T (0 or 1). Then the treatment effect is

$$\delta_i = Y_i^1 - Y_i^0$$

  — Allow for this to potentially be **heterogeneous.**

- We need to complicate this further b/c we now have an instrument Z_i for the treatment (e.g. whether startup was randomly sent an invitation to the accelerator scheme)

What is the IV strategy identifying?

- Won't go through all the notation and assumptions – see Cunningham section 7.6 for a good treatment – but what can be shown is that what IV identifies in such a setting is the **local average treatment effect, or LATE.**

$$\delta_{IV,LATE} = E[(Y_i^1 - Y_i^0) |\ T_i(Z_i = 1) - T_i(Z_i = 0) = 1]$$

- This is the average causal treatment effect on the subset of units whose treatment status was changed by the instrument. These are also called **"compliers".**
  - Other subgroups: "never takers", "always takers", "defiers"
  - For LATE interpretation, require "monotonicity" assumption: Z_i moves T_i in the same direction (or not at all) for all i → no defiers

What is the IV strategy identifying?

- To the extent that treatment effects are heterogeneous, the LATE may not be close to the overall ATE that we would (usually) ideally like to estimate
  - Imbens (2010): “Better LATE than nothing” – vs. Deaton/Heckman

- Also note that the LATE depends on the chosen instrument – a different IV would generally give us a different LATE

Have instrumental variables brought us closer to the truth?

- Recommended reading: Wei Jiang, RCFS 2017

A survey of 255 papers that rely on the instrumental variable (IV) approach for identifying causal effects published in the "Big Three" finance journals reveals that IV estimates are larger than their corresponding uninstrumented estimates in about 80% of the studies, regardless of whether the potential endogeneity is expected to create a positive or negative bias based on economic reasoning. The magnitude of the IV estimates is, on average, nine times of that of the uninstrumented estimates even when economic insights do not suggest a downward bias of the latter. This study provides several explanations to the "implausibly large" IV estimates in finance research, and proposes best practices for identification-conscientious researchers. (*JEL* G30, C13)

Have instrumental variables brought us closer to the truth?

- Her candidate explanations:
  - LATE > population average treatment effect
    - example from labor econ: education instrumented by proximity to college (or other instruments) – often get $\widehat{\beta_{IV}} > \widehat{\beta_{OLS}}$
  - (Relatively) weak instruments amplify any small violation of exclusion restriction
    - also see Keane-Neal point below
  - Specification search / publication bias. IV s.e. typically much larger than OLS, so to be significant, point estimate "needs" to be larger than OLS
    - S.e. increase by a factor of about $1/\sqrt{R^2}$ where R^2 is the partial R^2 in the first stage (after controlling for other covariates)
    - E.g. for 0.02 ("a respectable partial R^2"), factor of 7

Have instrumental variables brought us closer to the truth?

- What should be done?
  - — Anticipate the relative magnitude of $\widehat{\beta_{IV}}$ and $\widehat{\beta_{OLS}}$ ex ante, and reconcile/discuss ex post
  - — Be transparent regarding IV potency
    - – e.g. report partial R^2 from first stage; give a sense of how many "compliers"
  - — Reality check of economic magnitudes (not just stat. significance)

Back to Keane and Neal (ARE 2024, Section 6)

- They emphasize that a low first-stage F-statistic is not just a potential problem for inference – it also makes it more likely that 2SLS estimates will be further from the truth than OLS
  - — especially if the degree of endogeneity (ρ) is modest

Table 5 Probability of 2SLS outperforming OLS

Population F	$F_{5\%}$	Uniform prior for ρ: 0 to 1	0 to 0.45	0.35 to 0.45	0.5 to 1
1.82	8.96	45	24	41	65
2.30	10	48	26	45	69
3.84	13	56	32	55	77
5.78	16.38	62	38	64	84
10.00	23.10	70	47	77	91
29.44	50	83	65	95	99
73.75	104.70	89	76	100	100

We report the frequency of $|\hat{\beta}_{2SLS} - \beta| < |\hat{\beta}_{OLS} - \beta|$ across Monte Carlo replications, averaged across all values of ρ under a uniform prior that ρ falls in the indicated range. Abbreviations: 2SLS, two-stage least squares; OLS, ordinary least squares.

Back to Keane and Neal (ARE 2024, Section 6)

Despite the difficulty of devising a general rule of thumb, a strong case can be made that applied researchers should adopt a threshold of instrument strength of at least $\hat{F} > 50$ in the single instrument case. This makes 2SLS likely to outperform OLS at moderate levels of endogeneity, although at high levels of ρ a lower $\hat{F}$ would suffice.[20] As we have seen, this threshold renders extreme 2SLS outliers very unlikely. If such a threshold cannot be met, it is advisable to seek stronger instruments or pursue alternative strategies, such as OLS combined with a serious attempt to control for omitted variables.[21] We reiterate that robust tests (AR or CLR) should be used in lieu of 2SLS t-tests regardless of $\hat{F}$.

- Key takeaways:
  - — If we care about *testing*, AR (or "CLR" with multiple IVs) solve the main weak-instrument test size problem.
  - — If we care about *magnitudes*, 2SLS can easily be less reliable than OLS unless
    - – instruments are very strong (roughly F $\geq$ 50), and/or
    - – we have strong prior that OLS bias is large relative to the sampling noise

Overview

- The problem
- What are the assumptions when doing IV?
- How is IV implemented?
- Other issues with IV
- **Examples**

“Natural” instruments

- For corporate governance related topics, births/deaths can provide good instruments
  - Gender of first-born child for family firm succession
  - CEO deaths, etc.

- Weather events can provide good instruments as well
  - “Snow and leverage” (Giroud et al. 2012)
  - Bad weather & attendance at various events (political rallies etc.)
  - Exogenous losses depending on location of natural disasters
    - (although still need to think about endogenous sorting into locations & other possible violations of exclusion restriction)

- **But:** even completely random events don’t guarantee that excl. restriction holds. Example: 2 kids’ gender & female labor supply

Lagged variables as instruments?

- In some (rare) cases, x_{t-j} may be a valid instrument for x_t
  - usually relevant, if x is persistent
  - bigger question is if exclusion restriction is satisfied – i.e. x_{t-j} should not directly affect y_t. Often implausible, since requires that only the exogenous part (not the endog. part) of x persists.

- Not much formalization of problems with this approach; one exception is https://marcfbellemare.com/wordpress/13422

- Note: not to be confused with dynamic panel data approaches where Δy_{t-1} is among the right-hand-side var. and instrumented for by further lags of y_{t-j} (Arellano-Bond)
  - or Euler equation models – see Roberts and Whited (sect. 3.6)

Bartik / shift-share instruments

- Idea: use differential (pre-determined) exposure to common "exogenous" shocks (closely related to diff-in-diff, cf. next week)
- TWFE panel setting:

e.g. sales growth

$$y_{it} = \alpha_i + \alpha_t + \beta x_{it} + \varepsilon_{it}$$

- Decompose x_{it} into sum across segments/sectors

$$x_{it} = \sum_k w_{itk} \times e_{itk}$$

i's growth in segment *k*

i's share in segment *k* at time *t*

- Instrument based on

$$z_{it} = \sum_k w_{ik} \times e_{tk}$$

aggregate growth in segment *k*

past share in segment *k*

Bartik – examples

- Bartik (1991): estimate effects of labor demand changes on local wages, observed at the county level (*i*) over time (*t*)
- Need (quasi-) exogenous shifter in labor demand *x*. Use
  - *w:* share of past employment in county *i* in industry *k*
  - *e:* nationwide industry-specific employment growth rates

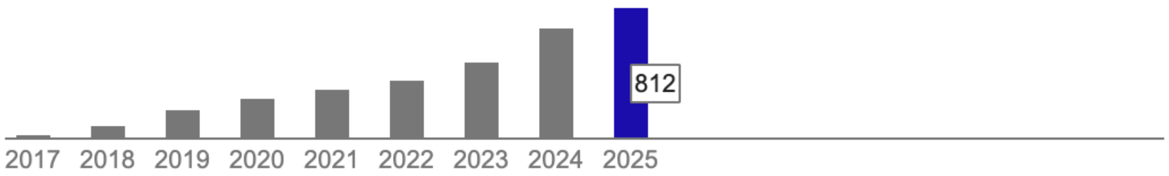
- Autor et al. (2013): “China shock”. *x* is exposure to import competition; *w* as above; *e* is imports from China in industry *k* in other high-income countries

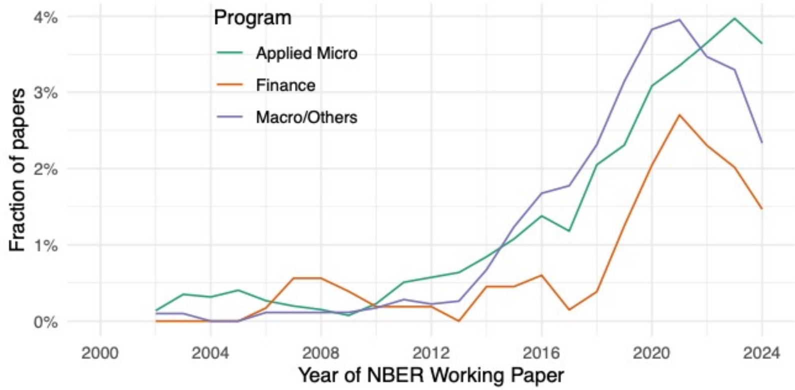
- Greenstone-Mas-Nguyen (2020): predict county-level bank lending shocks (*x*) using variation in pre-existing bank market shares (*w*) and bank supply shifts (*e*)
  - *e* itself needs to be estimated first

Very popular in recent years

Bartik instruments: What, when, why, and how

Authors	Paul Goldsmith-Pinkham, Isaac Sorkin, Henry Swift
Publication date	2020/8/1
Journal	American Economic Review
Volume	110
Issue	8
Pages	2586-2624
Publisher	American Economic Association
Description	The Bartik instrument is formed by interacting local industry shares and national industry growth rates. We show that the typical use of a Bartik instrument assumes a pooled exposure research design, where the shares measure differential exposure to common shocks, and identification is based on exogeneity of the shares. Next, we show how the Bartik instrument weights each of the exposure designs. Finally, we discuss how to assess the plausibility of the research design. We illustrate our results through two applications: estimating the elasticity of labor supply, and estimating the elasticity of substitution between immigrants and natives. (JEL *C51, F14, J15, J22, L60, R23, R32*)
Total citations	Cited by 3148





(a) Bartik and shift-share instruments

Source: Goldsmith-Pinkham (2024)

Where does identification come from?

- Recent literature has clarified assumptions that are implicit in these designs, and how to assess them
  - Goldsmith-Pinkham et al. (2020): Bartik instrument is equivalent to using local industry shares as instruments, and so the exogeneity condition should be interpreted in terms of the shares
    - are initial shares “exogenous” relative to outcome of interest?
    - method to show which shares (industries) are important
  - Borusyak et al. (2022): exogenous independent shocks to many industries allow estimation of causal effects even when local industry shares are endogenous

- See e.g. **Borusyak et al. (JEP 2025)**, Breuer (2022; https://ssrn.com/abstract=3786229), https://blogs.worldbank.org/impactevaluations/just-little-bartik-exposure, or Cunningham 7.8.3/.4 for recent summaries

Final comments

- Used to be that "a good new instrument = 1 publication"
- Probably still true, but
  - — increasingly harder to find good *new* instruments
  - — increasing scrutiny of exclusion restriction / plausibility / LATEness
  - — increasing focus on weak-IV inference
- Two strategies that *can* be promising:
  - — use well-accepted instrument for new question (and shed new light on instrument validity)
  - — when see "random" factors (= potential IV) think about what questions could be addressed with it (but can also be "dangerous")
- Bartik-type "exposure" designs very popular now – many applications but make sure to understand relevant "theory"