Canonical Research Designs V: Bartik and Simulated Instruments

Paul Goldsmith-Pinkham

April 6, 2021

Roadmap for Today

- In some cases, the source of exogeneous variation (either in an IV setting, or just OLS) is straightforward
  - There is a single policy or source of variation
- However, in other settings, there are more complicated sources of variation exploited to identify effects. Today we'll focus on two:
  - Bartik (shift-share) instruments: three recent papers on commonly used identification approach
  - Simulated instruments: reframe an older literature in a new light
- Key historical feature of these approaches is that they had an "intuitive" feature of identification, but formal properties were not established for several decades
  - Analagous to staggered DinD lit!

Bartik instruments are used everywhere

Bartik

Labor supply + immigration

Trade

Finance, innovation, others...

- Thread that links all Bartik applications:
  - local markets composed of many "categories"
  - need for identification
- Approach has been used since the early 90:
  - sometimes called "shift-share" or "industry mix" instruments

Examples of Bartik instruments in many subfields

Immigration: Altonji and Card (1991), Card (2001)

Bank Lending: Amiti and Weinstein (2018), Greenstone, Mas and Nguyen (2015)

Market Size + Demography: Acemoglu and Linn (2004), Jaravel (2018)

Labor Supply Elasticity: Blanchard and Katz (1992), **Bartik** (1991)

Fiscal Multipliers: Nakamura and Steinsson (2014)

Trade + Labor: Autor, Dorn, and Hanson (2013), Autor, Dorn and Hanson (2018), etc.

Foreign Aid: Nunn and Qian (2014)

Portfolio Allocation: Calvet, Campbell, and Sodini (2009)

Trade + Prices: Piveteau and Smagghue (2017), de Roux et al. (2017)

Automation: Acemoglu and Restrepo (2017)

Many paths lead to Bartik

- Diverse literature leads to many motivations and justifications for Bartik approach
- Two distinct approaches in the literature:
  1. Applied micro statistical approach: interested in a reduced form causal relationship; need an instrument that is uncorrelated with error term; make argument that Bartik instrument is defensible
  2. Structural approach: interested in particular parameters from model; assumptions of model motivate certain estimating equations
- So what is the Bartik approach anyway?

Motivation: local labor market approaches + reduced form

Consider a local labor market regression like the following:

$$y_l = \beta_0 + \beta x_l + \epsilon_l$$

- $\mathbb{E}[x_l \epsilon_l] \neq 0 \Rightarrow$ need an instrument to estimate β
- E.g. Autor, Dorn and Hanson (2013) setting:
  - l: location (commuting zone)
  - y_l: manufacturing employment *growth*
  - x_l: import exposure to China *growth*
  - β: effect of rise of China on manufacturing employment
  - an instrument for location-level exposure to trade with China

The Bartik instrument

Accounting identity #1:

$$x_l = \sum_{k=1}^{K} z_{lk} g_{lk}$$

- z_{lk}: location-industry shares (Z_l)
- g_{lk}: location-industry growth (in imports) rates (G_l)

Accounting identity #2:

$$\underbrace{g_{lk}}_{\text{location-industry}} = \underbrace{g_k}_{\text{industry}} + \underbrace{\tilde{g}_{lk}}_{\substack{\text{idiosyncratic} \\ \text{location- industry}}}$$

Infeasible Bartik:

$$B_l = \sum_{k=1}^{K} z_{lk} g_k$$

This gives us a simple 2SLS structure

$$y_l = \beta_0 + \beta x_l + \epsilon_l$$
$$x_l = \pi_0 + \pi_1 B_l + u_l$$
$$B_l = \sum_{k=1}^{K} z_{lk} g_k$$
$$g_{lk} = g_k + \tilde{g}_{lk}$$

China shock: e.g., Autor, Dorn and Hanson (2013)

- z_{lk}: location (l) industry (k) composition
- g_{lk}: location (l) industry (k) growth in imports from China
- g_k: industry k growth of imports (from China)

Other instruments have this structure

$$
\begin{aligned}
y_l &= \beta_0 + \beta x_l + \epsilon_l \\
x_l &= \pi_0 + \pi_1 B_l + u_l \\
B_l &= \sum_{k=1}^{K} z_{lk} g_k \\
g_{lk} &= g_k + \tilde{g}_{lk}
\end{aligned}
$$

Immigrant enclave: e.g., Altonji and Card (1991)

- z_{lk}: share of people from foreign k living in l (in a base period)
- g_{lk}: growth in number of people from k to l
- g_k: growth in people from k nationally

Other instruments have this structure

$$y_l = \beta_0 + \beta x_l + \epsilon_l$$
$$x_l = \pi_0 + \pi_1 B_l + u_l$$
$$B_l = \sum_{k=1}^{K} z_{lk} g_k$$
$$g_{lk} = g_k + \tilde{g}_{lk}$$

Bank-lending relationships: e.g., Greenstone, Mas and Nguyen (2015)

- z_{lk}: location (l) share of loan origination from bank k
- g_{lk}: loan growth in location l by bank k
- g_k: part of loan growth due to bank supply shock

Other instruments have this structure

$$
\begin{aligned}
y_l &= \beta_0 + \beta x_l + \epsilon_l \\
x_l &= \pi_0 + \pi_1 B_l + u_l \\
B_l &= \sum_{k=1}^{K} z_{lk} g_k \\
g_{lk} &= g_k + \tilde{g}_{lk}
\end{aligned}
$$

Market size and demography: e.g., Acemoglu and Linn (2004)

- z_{lk}: spending share on drug l from age group k
- g_{lk}: growth in spending of group k on drug l
- g_k: growth in spending of group k (due to population aging)

What's necessary for consistency?

$$
\begin{aligned}
y_l &= \beta_0 + \beta x_l + \epsilon_l \\
x_l &= \pi_0 + \pi_1 B_l + u_l \\
B_l &= \sum_{k=1}^{K} z_{lk} g_k \\
g_{lk} &= g_k + \tilde{g}_{lk}
\end{aligned}
$$

- We need B_l to be a valid instrument
- Requires two conditions with constant effects:
  1. Relevance: $\pi_1 \neq 0$, e.g. $Cov(B_l, x_l) \neq 0$
  2. Exclusion: $E(B_l \epsilon_l) = 0$
- Key flaw in this literature until recently: economic + statistical content of exclusion has been vague and sometimes confused

Key thing to remember from today

- Assuming independence or exogeneity on the basis of a model does not necessarily make it true
  - E.g. Hausman instruments in IO models – model may assume that exclusion restriction is satisfied, but not necessarily true in reality
- Assuming that two things are independent because they don't seem "related" doesn't make it true
  - Bartik literature many times argues that national nature of shocks "decouples" the instrument from local market conditions. However, it still exploits local characteristics. Need to make very specific arguments to validate claim (will come to this).
- When evaluating an identification strategy, you should be able to describe counterfactual claims using the measure. This is typically not concrete in Bartik – try to make it concrete! What is exactly changing in China? Why is it random?

More general econometric set-up

$$
\begin{aligned}
y_{lt} &= \mathbf{D}_{lt}\beta_0 + x_{lt}\beta + \epsilon_{lt}, \\
x_{lt} &= \mathbf{D}_{lt}\tau + B_{lt}\gamma + \eta_{lt} \\
\mathbf{D}_{lt} &= \text{controls, f.e.} \\
g_{lkt} &= g_{kt} + \tilde{g}_{lkt} \\
B_{lt} &= \sum_{k=1}^{K} z_{lk0} g_{kt}, \\
&\{\{x_{lt}, \mathbf{D}_{lt}, \epsilon_{lt}\}_{t=1}^{T}\}_{l=1}^{L}, \text{ iid, } L \to \infty
\end{aligned}
$$

Assumptions for IV in terms of B_{lt}:

- Exogeneity: $\mathbb{E}\left[B_{lt}\epsilon_{lt} | \mathbf{D}_{lt}\right] = 0$

- Relevance: $\text{Cov}\left[B_{lt}, x_{lt} | \mathbf{D}_{lt}\right] \neq 0$

Question:

- What do these statements about B_{lt} imply about z_{lk0} and g_{kt}?

Recent Literature on this topic

- Three papers addressed this question, and can be split into two grouping
- The division between papers can be split based on focus on z_{lk0} vs. g_{kt}
  1. Goldsmith-Pinkham, Sorkin and Swift (2020) focus on z_{lk0} and make an analogy to difference-in-differences
  2. Adao, Kolesar and Morales (2019) and Borusyak, Hull and Jaravel (2020) focus on g_{kt}, and make a strong connection to the design based approach (e.g. these are as-if random shocks)
- I will focus on GPSS, but the important fact is to pick a concrete explanation for identification
  - Key problem, historically, in this literature, was the lack of a coherent defense of the identifying variation
  - Ideally, this choice should be made ex ante! I.e. don't run the tests and then choose...

Bartik Instruments: What, When, Why and How
Goldsmith-Pinkham, Sorkin and Swift

Outline:

- **Understanding the identifying assumption**
- Opening the black box
- Testing the plausibility of the identifying assumption

Three special cases

1. One time period, two industries
2. T time periods, two industries
3. One time period, K industries

Special case #1: One time period, two industries

- $z_{l2} = 1 - z_{l1}$
- Bartik:

$$
\begin{aligned}
B_l &= z_{l1} g_1 + z_{l2} g_2 = z_{l1} g_1 + (1 - z_{l1}) g_2 \\
&= g_2 + (g_1 - g_2) z_{l1}
\end{aligned}
$$

First-stage:

$$
\begin{aligned}
x_l &= \gamma_0 + \gamma B_l + \eta_l \\
x_l &= \underbrace{\gamma_0 + \gamma g_2}_{\text{constant}} + \underbrace{\gamma (g_1 - g_2)}_{\text{coefficient}} z_{l1} + \eta_l
\end{aligned}
$$

The instrument is z_{l1}, while g_k affects relevance

▸ Why OLS is biased

Special case #2: T time periods, two industries

Panel Bartik:

$$B_{lt} = z_{l10} g_{1t} + z_{l20} g_{2t} = g_{2t} + \underbrace{\Delta_{gt}}_{g_{1t} - g_{2t}} z_{l10}$$

First stage:

$$x_{lt} = \tau_l + \tau_t + \gamma B_{lt} + \eta_{lt}$$
$$x_{lt} = \tau_l + \underbrace{(\tau_t + \gamma g_{2t})}_{\tilde{\tau}_t} + \underbrace{\gamma \Delta_{gt}}_{\tilde{\gamma}_t} z_{l10} + \eta_{lt}$$

- Industry shares times time period is the instrument
- (Updated industry shares: similar)

Special case #2: T time periods, two industries

- Analogy to continuous difference-in-differences
  - Δ_{gt} is size of policy
  - z_{l10} is exposure to policy
- Sometimes a "pre-period" before policy: test for parallel pre-trends
  - E.g., in ADH, what happens from 1970 to 1990?

Special case #3: One time period, K industries

- G: $K \times 1$ vector of g_k
- Z: $L \times K$, matrix of Z_l
- $Y^\perp$, $X^\perp$, $B = (ZG)$: $L \times 1$, vectors of $y_l^\perp$, $x_l^\perp$ and B_l
- Ω: $K \times K$

$$\hat{\beta}_{Bartik} = \frac{B'Y^\perp}{B'X^\perp}$$
$$\hat{\beta}_{GMM} = \frac{(X^{\perp\prime}Z)\Omega(Z'Y^\perp)}{(X^{\perp\prime}Z)\Omega(Z'X^\perp)}$$

If $\Omega = (GG')$, then $\hat{\beta}_{Bartik} = \hat{\beta}_{GMM}$

Full general result with T time periods and K industries

Two estimators are numerically identical:

- TSLS with Bartik instrument
- GMM with industry shares $\times$ time period as instruments and a particular weight matrix

$$\hat{\beta}_{Bartik} = \frac{\mathbf{B}'\widetilde{\mathbf{Y}}^{\perp}}{\mathbf{B}'\widetilde{\mathbf{X}}^{\perp}}$$

$$\hat{\beta}_{GMM} = \frac{(\mathbf{X}^{\perp\prime}\widetilde{\mathbf{Z}})\Omega(\widetilde{\mathbf{Z}}'\mathbf{Y}^{\perp})}{(\mathbf{X}^{\perp\prime}\widetilde{\mathbf{Z}})\Omega(\widetilde{\mathbf{Z}}'\mathbf{X}^{\perp})}$$

$\Omega = (\mathbf{GG}')$, and $\widetilde{\mathbf{Z}}$ is an $LT \times KT$ stacked vector of Z_0 interacted with time fixed effects and $\mathbf{G}$ is a $KT \times 1$ stacked vector of growth rates g_{kt}.

When is the estimator consistent for the estimand of interest?

What is the identification condition?

$$\hat{\beta}_{Bartik} = \frac{\sum_{l=1}^{L} \sum_{t=1}^{T} \sum_{k=1}^{K} z_{lkt} g_{kt} y_{lt}^{\perp}}{\sum_{l=1}^{L} \sum_{t=1}^{T} \sum_{k=1}^{K} z_{lkt} g_{kt} x_{lt}^{\perp}}$$

Two ideas:

- "Shares" (this paper): talk about properties of z_{lkt}
    - Conditional exogeneity
- "Shocks" (Borusyak, Hull and Jaravel (2018)): talk about properties of g_{kt}
    - Random, and a large number (equivalent industry-level regression)

When are these views plausible? What do they mean?

Shares

Conditional exogeneity:

- Typically: exogenous to changes in error term, not levels of outcome
- Standard in diff-in-diff (exclusion): in a period, exposure to an industry matters for outcome only through x

Shocks

- Large number of industries (shares are misspecified, need it to average out)
- Random shocks across industries

Is ADH about shocks or shares?

Shocks:

- Explains why $g_{kt}^{high-income}$ rather than g_{kt}^{US} (hard to rationalize under shares)
- Natural in a trade model: why would imports from China rise (in a trade model)? Independent industry-specific shocks

Shares:

- Explains why z_{lkt-1} rather than z_{lkt} (hard to rationalize under shocks)
- Explains why it is important for identification to study local labor markets (as opposed to parameter of interest where we want to think about spillovers)

Bottom line: a little hard to tell what exactly ADH are assuming; ADH approach does not appear to satisfy testable assumptions under GPSS, but do appear to under BHJ.

Decomposing Bartik (GPSS 2020)

(Special case of Rotemberg (1983), proposition 1)

$$\hat{\beta}_{Bartik} = \sum_k \hat{\alpha}_k \hat{\beta}_k, \quad \sum_k \hat{\alpha}_k = 1$$

IV estimate using only the k^{th} instrument:

$$\hat{\beta}_k = (Z_k' X)^{-1} Z_k' Y$$

"Rotemberg" weight:

$$\hat{\alpha}_k = \frac{g_k Z_k' X}{\sum_{k=1}^{K} g_k Z_k' X}$$

▸ Proof

Interpretation: sensitivity to misspecification elasticity

Conley, Hansen and Rossi (2012); Andrews, Gentzkow and Shapiro (2017)
Local misspecification: $\epsilon_{lt} = L^{-1/2} V_{lt} + \tilde{\varepsilon}_{lt}, \quad Cov(V_{lt}, Z_{lt}) \neq 0,$

- $\sqrt{L}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} \tilde{\beta}, \qquad \mathbb{E}\left[\tilde{\beta}\right] =$ bias (misspecification) of Bartik instrument
- $\sqrt{L}\left(\hat{\beta}_k - \beta_0\right) \xrightarrow{d} \tilde{\beta}_k, \qquad \mathbb{E}\left[\tilde{\beta}_k\right] =$ bias (misspecification) of kth instrument

Suppose $\beta_0 \neq 0$. Percentage bias:

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \alpha_k \frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}$$

Industry with high α_k:

- an industry where it matters whether it is misspecified (endogenous)
  - because it is “important” in the estimate

Top five industries (out of 397)

	$\hat{\alpha}_k$	$g_k^{\text{high-income}}$	$\hat{\beta}_k$
Games and Toys	0.182	174.841	-0.151
Electronic Computers	0.182	85.017	-0.620
Household Audio and Video	0.130	118.879	0.287
Computer Equipment	0.076	28.110	-0.315
Telephone Apparatus	0.058	37.454	-0.305
	0.628/1.379		-0.230

*The **main source of variation in exposure is** within-manufacturing specialization in industries subject to different degrees of import competition...there is differentiation according to **local labor market reliance on labor-intensive industries**...By 2007, China accounted for over 40 percent of US imports in four four-digit SIC industries (**luggage, rubber and plastic footwear, games and toys, and die-cut paperboard**) and over 30 percent in 28 other industries, including **apparel, textiles, furniture, leather goods, electrical appliances, and jewelry**.*

— Autor, Dorn and Hanson (2013) , pg. 2123

▸ Correlation matrix

Outline

- Understanding the identifying assumption
  - *If fixed K and T, and $L \to \infty$, then in terms of industry composition*
- Opening the black box
  - *Which industries are "important" in estimates*
- **Testing the plausibility of the identifying assumption**

Three tests of the identifying condition (under GPSS (2020)

1. Confounds (or correlates)
2. Pre-trends
3. Alternative estimators and overidentification

- There are *also* tests for BHJ – similar to assuming strict ignorability, you can test for balance on observables (like the confounds above) of industries and locations

Test #1: Correlates of initial industry composition

- How are initial characteristics (D_{l0} less F.E.) related to Z_{l0}?
- Look at high-Rotemberg weight industries (and aggregate)

Not definitive:

- Shows source of variation
- Can address by controlling for observables ($D_{l0} \times$ time)

Test #1: Correlates

	Games and toys	Electronic computers	Household audio and video	Computer equipment	Telephone apparatus	China to other
Share Empl in Manufacturing	0.01	0.21	0.08	0.21	-0.07	0.57
	(0.03)	(0.18)	(0.08)	(0.15)	(0.06)	(0.07)
Share College Educated	-0.08	0.20	0.01	0.22	-0.07	0.30
	(0.03)	(0.11)	(0.04)	(0.10)	(0.06)	(0.06)
Share Foreign Born	0.01	-0.01	-0.02	-0.01	-0.08	0.15
	(0.01)	(0.04)	(0.01)	(0.04)	(0.03)	(0.03)
Share Empl of Women	0.05	-0.04	-0.08	-0.02	-0.02	0.10
	(0.03)	(0.12)	(0.05)	(0.12)	(0.07)	(0.06)
Share Empl in Routine	0.04	-0.37	0.06	-0.36	-0.01	-0.08
	(0.03)	(0.14)	(0.05)	(0.12)	(0.07)	(0.13)
Avg Offshorability	0.02	0.33	0.00	0.29	0.23	-0.24
	(0.02)	(0.10)	(0.05)	(0.08)	(0.04)	(0.09)
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	1,444	1,444	1,444	1,444	1,444	1,444
R^2	0.02	0.08	0.01	0.08	0.05	0.22

Test #1: Correlates

	Games and toys	Electronic computers	Household audio and video	Computer equipment	Telephone apparatus	China to other
Share Empl in Manufacturing	0.01	0.21	0.08	0.21	-0.07	0.57
	(0.03)	(0.18)	(0.08)	(0.15)	(0.06)	(0.07)
Share College Educated	-0.08	0.20	0.01	0.22	-0.07	0.30
	(0.03)	(0.11)	(0.04)	(0.10)	(0.06)	(0.06)
Share Foreign Born	0.01	-0.01	-0.02	-0.01	-0.08	0.15
	(0.01)	(0.04)	(0.01)	(0.04)	(0.03)	(0.03)
Share Empl of Women	0.05	-0.04	-0.08	-0.02	-0.02	0.10
	(0.03)	(0.12)	(0.05)	(0.12)	(0.07)	(0.06)
Share Empl in Routine	0.04	-0.37	0.06	-0.36	-0.01	-0.08
	(0.03)	(0.14)	(0.05)	(0.12)	(0.07)	(0.13)
Avg Offshorability	0.02	0.33	0.00	0.29	0.23	-0.24
	(0.02)	(0.10)	(0.05)	(0.08)	(0.04)	(0.09)
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	1,444	1,444	1,444	1,444	1,444	1,444
R^2	0.02	0.08	0.01	0.08	0.05	0.22

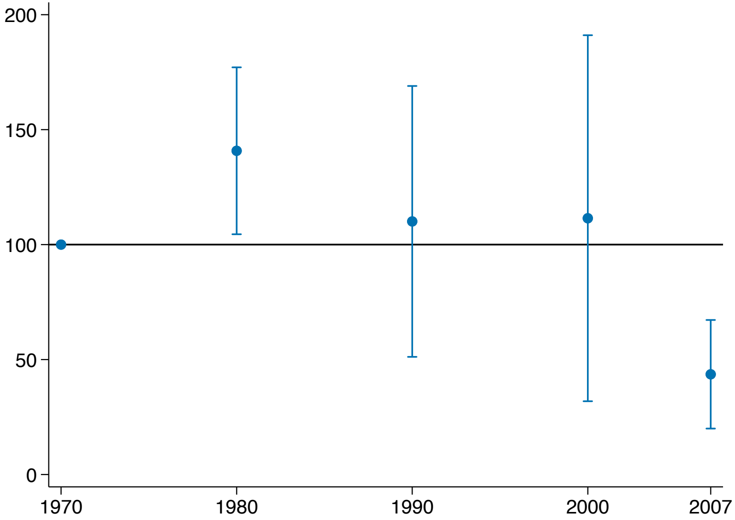
Test #2: Pre-trends

$$\Delta \text{Manufacturing Emp}_{lt} = \alpha + \sum_s \mathbb{1}(s = t)\gamma_{k,s} z_{lk,1980} + \epsilon_{lt}$$

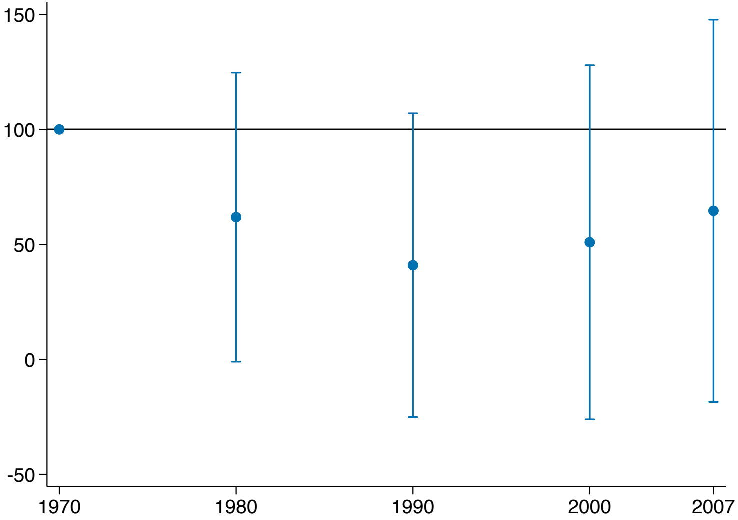
- Four time periods: 1970-1980, 1980-1990, 1990-2000, 2000-2007
- Convert $\hat{\gamma}_{k,t}$ to levels (1970 = 100)
- k is top five Rotemberg weight industries in 1980, and "aggregate"
    - Aggregate: 1980 shares, aggregated using $g_{k,1990-2000}^{\text{high-income}}$

"Pre-period" prior to 1990

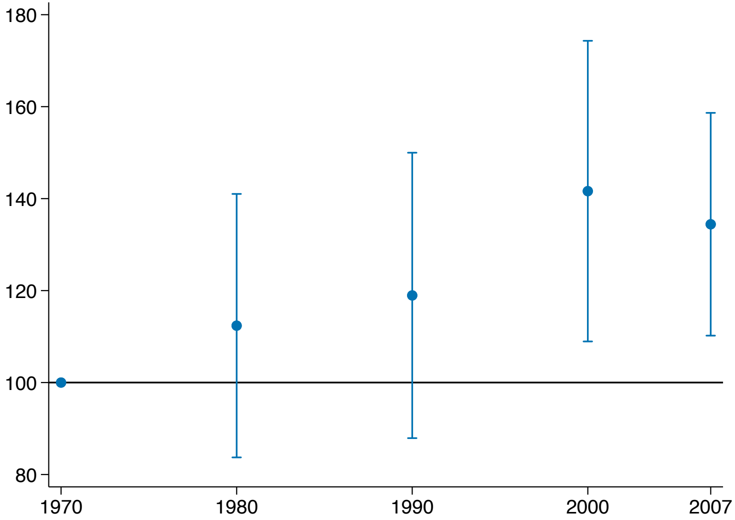
Electronic computers, fixed 1980 industry shares, Rotemberg weight 0.183



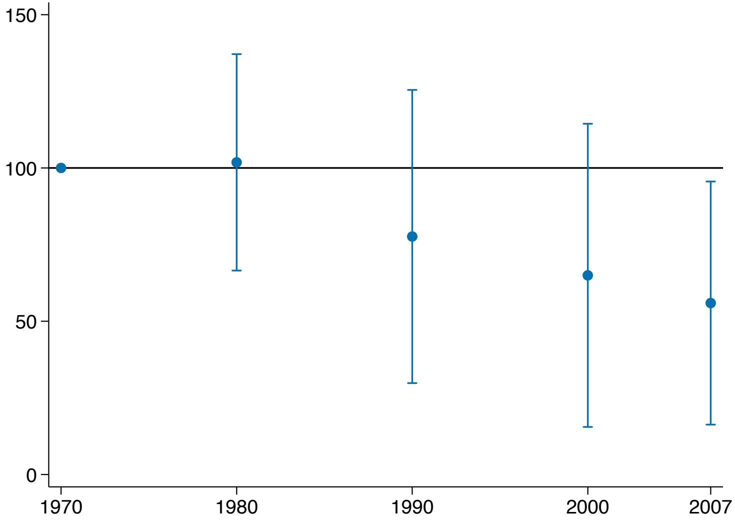
Games and toys, fixed 1980 industry shares, Rotemberg weight 0.138



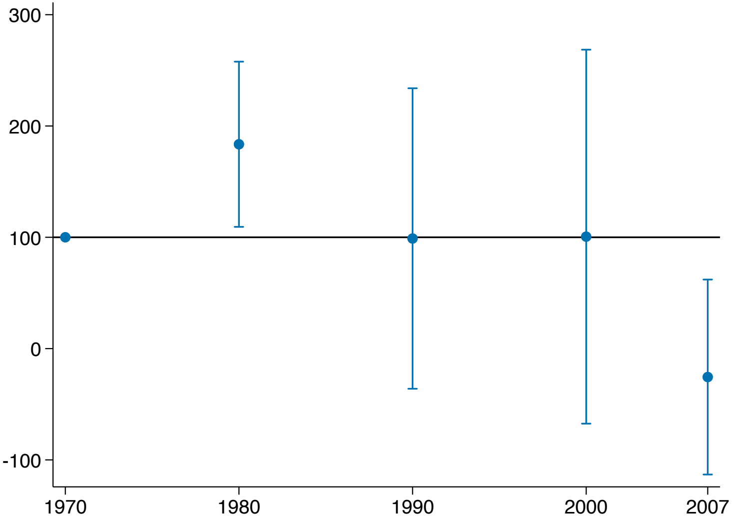
Household audio and video, fixed 1980 industry shares, Rotemberg weight 0.085



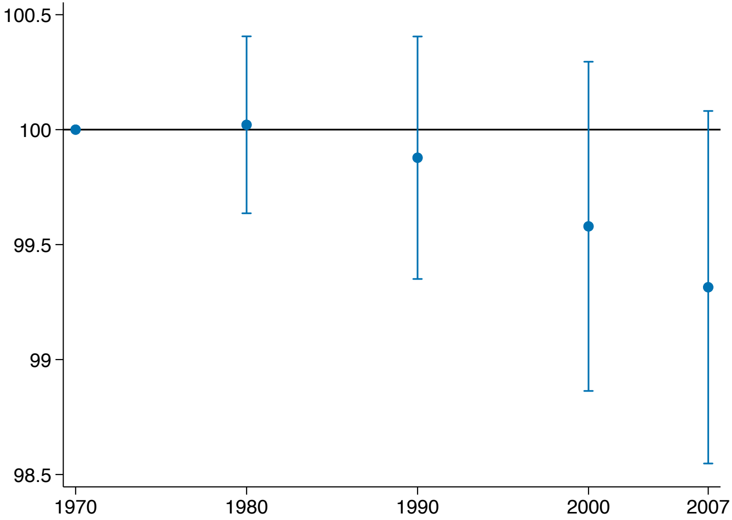
Telephone apparatus, fixed 1980 industry shares, Rotemberg weight 0.066



Computer equipment, fixed 1980 industry shares, Rotemberg weight 0.060



Aggregate, fixed 1980 industry shares



Test #3: Alternative estimators and overidentification tests

Basic insight: many instruments

- Estimators (maximum likelihood): LIML, Hausman, Newey, Woutersen, Chao and Swanson (2012) HFUL (heteroskedasticity-Fuller (1977))
- Estimators (two-step): TSLS (problematic), Bartik TSLS, MBTSLS (Anatolyev (2013), and Kolesar et al (2015))

Interpretation:

- Gap between maximum likelihood and two-step estimators is evidence of misspecification

Also, overidentification tests, which provides evidence of misspecification

Test #3: Alternative estimators and overidentification

	Δ Emp	Over ID Test
OLS	-0.17	
	(0.04)	
TSLS (Bartik)	-0.62	
	(0.11)	
TSLS	-0.22	872.69
	(0.06)	[0.00]
MBTSLS	-0.33	
	(0.05)	
LIML	-2.07	1348.50
	(3.52)	[0.00]
HFUL	-1.13	1141.08
	(0.04)	[0.00]
Year and Census Division FE	Yes	
Controls	Yes	
Observations	1,444	

Simulated Instruments

- Pivoting now to simulated instruments
- Running example – effect of Medicaid on Y (e.g. child mortality)
- Eligibility depends on:
  - State-year (legislative changes)
  - Individual characteristics: Income less disregards (child care and work expenses), adults (if married, unemployed husband can maintain eligibility), number of children, whether first-time mother, etc.

  Actual eligibility is clearly endogenous:
  - Poorer people are more likely to be eligible, "worse" potential outcomes

Build an instrument

- Idea: isolate (changes in) eligibility coming from policy changes
  - Take a fixed population (say, 1986 CPS)
  - Compute "share eligible" in each state-year
- Parameterizes (in one-dimension) how generous the policy is
- Benefits:
  - Intuitive: isolate variation that you think is exogenous
  - "Split-sample": Can compute simulated eligibility in one sample, and outcomes in another sample

Set up

- Y_i: outcome (e.g. utilization of health care, whether baby is low birthweight)
- D_i: treatment (e.g. Medicaid eligibility)
  - Today: $D_i \in \{0, 1\}$
- $D_i \equiv g(S_i, X_i)$: *known* eligibility function $g(\cdot, \cdot)$
  S_i: location (e.g. state)
  X_i: characteristics driving eligibility (e.g. income, age, # of children)
  $\mathcal{X}$: simplifying case - discrete support of X_i
- $g(s, x)$ *can be highly non-linear in x*

Aronow, Goldsmith-Pinkham, & Sorkin: treatment effect heterogeneity

It is natural to imagine that we are not in a constant effects world:

- In observables: differential effects by income
- In unobservables: differential effects by latent health and etc.

Two questions with TE heterogeneity:

- How plausible is it that monotonicity holds?
- What is the LATE? (And is the LATE of economic interest?)

Fundamental problem of causal inference

- $Y_i(D_i)$ denote potential outcomes for individual i
- $Y_i(1) - Y_i(0)$: individual treatment effect of eligibility
- Observe $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$

Solution:

- Assume state *policies* are as good as random...
  - Concern that X_i is correlated with the potential outcomes;
  - Poor individual more likely to be eligible, but may have worse outcomes ($Y_i(D_i)$);
- How to get state policy variation without endogeneity from composition?

Quote - Currie & Gruber (1996)

“Our IV strategy ... treats the legislative environment ... as a source of exogenous identification.”

The simulated instrument

$$Z_i = \sum_{x \in \mathcal{X}} g(S_i, x) \underbrace{Pr(X_i = x)}_{\text{non-stoch func}}$$

- $Pr(X_i = x)$ is, e.g., distribution of characteristics in 1986 CPS

TSLS:

$$Y_i = \alpha + \beta D_i + \epsilon_i$$
$$D_i = \pi_0 + \pi_1 Z_i + \eta_i.$$

Standard constant effects assumptions:

$$\underbrace{Cov(Z_i, \epsilon_i)}_{\text{exclusion}} = 0, \qquad \underbrace{Cov(Z_i, D_i)}_{\text{relevance}} \neq 0$$

TSLS Estimator and Heterogeneity

- TSLS estimator:

$$\hat{\beta}_{sim} = \frac{\widehat{Cov}(Z_i, Y_i)}{\widehat{Cov}(Z_i, D_i)}.$$

- What average TE is $\hat{\beta}_{sim}$? Most researchers appeal to monotonicity for LATE:

Assumption

Let $\mathcal{Z}$ be the set of values of Z_i. For all $z, z' \in \mathcal{Z}$, either $D_i(z) \geq D_i(z')$ or $D_i(z) \leq D_i(z')$.

- How plausible is this?

Idea #1: monotonicity is very likely to be violated

Recall why we need an instrument:

- Eligibility depends on "high-dimensional" observables
- States change many dimensions of eligibility at once
- If it were just income, we'd do an RD in income

Hence:

- When a state expands eligibility **on average**, it can make some groups ineligible
    - Law changes represent "menu of treatments"
- Violate monotonicity in **observables**; can document empirical prevalence (!)

Why you probably shouldn't use simulated instruments (in this setting): potentially uninterpretable estimates in presence of TE heterogeneity

Idea #2: we have a selection on observables design

- Basic assumption in the simulated instrument design: state legislative policy changes don't depend on characteristics of the population
- **Strengthen (?) this assumption:** characteristics don't **respond** to policy changes (i.e., become poorer to become eligible for Medicaid)
  - This form of response affects interpretation of IV estimate
- If you buy this (which you might not in some settings!):
  - You can compare outcomes of "types" of people (vectors of characteristics) who are eligible in some state-years and not in others

Monotonicity assumption is strong, and testable

- Since $g(\cdot, \cdot)$ is known for all x and s, can look for $x, x' \in \mathcal{X}$ where $Z_i(s) > Z_{i'}(s')$ for x and x' but $D_i(z) > D_i(z')$ for x and $D_i(z) < D_i(z')$
- Simple example: states where eligibility is not ordered
  - $\mathcal{X} = \{1, 2, 3, 4, 5\}$, $Pr(X_i = x) = 1/5$
  - Types 1,2,5 are eligible in s, and 1,3,4,5 are eligible in s'.
  - Then, $\forall i$, $Z_i(s') > Z_i(s)$.
  - But, for $X_i = 3$, $D_i(s) < D_i(s')$, while for $X_i = 2$, $D_i(s') < D_i(s)$
- Implication: motonoicity is violated, and $\hat{\beta}_{sim}$ combines positive and negative eligibility effects.

How important are these violations?

- Cohodes, Grossman, Kleiner and Lovenheim (2016) study long-term impact of Medicaid expansions on childrens' educational outcomes
- Information on eligibility for fixed distribution of 1986 CPS
- Define an eligibility "type":
  - A unique pattern of state-year (from 1980-2007) eligibility for a respondent in the 1986 CPS
  - 48,036 children in the 1986 CPS, 18, 881 eligibility types

Quantifying violations of monotonicity

- Take state-years where the value of the instrument changes (i.e., state policy becomes more/less generous)
- Ask for each eligibility type whether its eligibility moved in the direction of the instrument
- Weight by the eligibility type share of the national population
  - In the CPS, small samples at state level

Idea #2.5: What you can sometimes do instead

With infinite data and outcomes observed in same dataset as eligibility (no split sample):

- Compute type specific treatment effects
  - Simple comparison of means, no IV!
- Report TE for populations of interest where eligibility varies across state-years
  - I.e., if we expanded the income threshold nationally, what would happen...
- Because monotonicity violation is in observables, not hard to fix (!)

In practice: use the propensity score....

Caveats

This is about Currie and Gruber (1996a, b) and not Gruber and Saez (2002):

- Gruber and Saez is an individual-level instrument so the monotonicity point doesn't apply
- Gruber and Saez is all about characteristics (income) responding endogenously

Not clear how useful/implementable idea #2.5 is:

- Need outcomes and eligibility in same dataset

In progress:

- A complete empirical example
- Even if you can't do #2.5, can bound TE heterogeneity so that still have interpretable estimates
- Formalism is a bit sloppy: currently just for cross-section, and not panel with two-way F.E.

Borusyak and Hull: formalizing this point even further

- Borusyak and Hull (2021) bring this point even further, emphasizing this propensity score setup in a wider range of settings
- They allow for potential *interference* across designs as well (due to, for example, spatial correlation in things like railroads)
- BH approach builds on the BHJ/AKM notion of random shocks being combined using non-random exposure
    - Key p-score result is the same, but the build out a more general (and more efficient) set of structures
    - Additional, in cases with interference (like railroad upgrades), BH show how to "recenter" the pscore to account for interference