Shift-Share IV

Mixtape Track

Roadmap

Who Am I?

The Groos Family Assistant Professor of Economics, Brown University

Who Am I?

The Groos Family Assistant Professor of Economics, Brown University

A big fan of instrumental variable methods:

Who Am I?

The Groos Family Assistant Professor of Economics, Brown University

A big fan of instrumental variable methods:

- Lottery- and non-lottery IVs in studies of educational quality

  (Angrist et al. 2016, 2017, 2021, 2022; Abdulkadiroğlu et al. 2016)

- Quasi-experimental evaluations of healthcare quality

  (Hull 2020; Abaluck et al. 2021, 2022)

- IV-based analyses of discrimination and bias

  (Arnold et al. 2020, 2021, 2022; Hull 2021; Bohren et al. 2022)

- Shift-share instruments (SSIV) and related designs

  (Borusyak et al. 2022; Borusyak and Hull 2021, 2022; Goldsmith-Pinkham et al. 2022)

What is This Course?

A one-day intensive on SSIV, focusing on recent practical advances

- Highlighting key points on identification, estimation, and inference
- Emphasis on *practical*: IV is meant to be used, not just studied!

What is This Course?

A one-day intensive on SSIV, focusing on recent practical advances

- Highlighting key points on identification, estimation, and inference
- Emphasis on *practical*: IV is meant to be used, not just studied!

Two 90-minute lectures

- Please ask questions in the Discord chat!

What is This Course?

A one-day intensive on SSIV, focusing on recent practical advances

- Highlighting key points on identification, estimation, and inference
- Emphasis on *practical*: IV is meant to be used, not just studied!

Two 90-minute lectures

- Please ask questions in the Discord chat!

One 45-minute coding lab

- 25 min: you seeing how far you can get on your own, or with your classmate's help (use Discord rooms!)
- 20 min: me live-coding solutions in Stata (we will also post R code)

Schedule

1:00-2:30pm	Lecture 1: Linear SSIV – Exogenous Shares and Shocks
2:30-2:35pm	*Break*
2:35-3:20pm	Coding Lab: Autor, Dorn, and Hanson (2013)
3:20-3:25pm	*Break*
3:25-4:55pm	Lecture 2: Nonlinear SSIV and Beyond – Instrument Recentering
4:55-5:00pm	Closing Remarks

What is a (Linear) SSIV?

A weighted sum of a common set of shocks, with weights reflecting heterogeneous exposure shares: $z_\ell = \sum_n s_{\ell n} g_n$

What is a (Linear) SSIV?

A weighted sum of a common set of shocks, with weights reflecting heterogeneous exposure shares: $z_\ell = \sum_n s_{\ell n} g_n$

- The shocks vary at a different "level" $n = 1, \ldots, N$ than the shares $\ell = 1, \ldots, L$, where we also observe an outcome y_ℓ & treatment x_ℓ

What is a (Linear) SSIV?

A weighted sum of a common set of shocks, with weights reflecting heterogeneous exposure shares: $z_\ell = \sum_n s_{\ell n} g_n$

- The shocks vary at a different "level" $n = 1, \ldots, N$ than the shares $\ell = 1, \ldots, L$, where we also observe an outcome y_ℓ & treatment x_ℓ

We want to use z_ℓ to estimate parameter β of the model $y_\ell = \beta x_\ell + \varepsilon_\ell$

What is a (Linear) SSIV?

A weighted sum of a common set of shocks, with weights reflecting heterogeneous exposure shares: $z_\ell = \sum_n s_{\ell n} g_n$

- The shocks vary at a different "level" $n = 1, \dots, N$ than the shares $\ell = 1, \dots, L$, where we also observe an outcome y_ℓ & treatment x_ℓ

We want to use z_ℓ to estimate parameter β of the model $y_\ell = \beta x_\ell + \varepsilon_\ell$

- Could be a "structural" equation or a potential outcomes model
- Could be misspecified, with heterogeneous treatment effects β_ℓ
- Could be a "reduced form" analysis, with $x_\ell = z_\ell$
- Could have other included controls w_ℓ

What is a (Linear) SSIV?

A weighted sum of a common set of shocks, with weights reflecting heterogeneous exposure shares: $z_\ell = \sum_n s_{\ell n} g_n$

- The shocks vary at a different "level" $n = 1, \ldots, N$ than the shares $\ell = 1, \ldots, L$, where we also observe an outcome y_ℓ & treatment x_ℓ

We want to use z_ℓ to estimate parameter β of the model $y_\ell = \beta x_\ell + \varepsilon_\ell$

- Could be a "structural" equation or a potential outcomes model
- Could be misspecified, with heterogeneous treatment effects β_ℓ
- Could be a "reduced form" analysis, with $x_\ell = z_\ell$
- Could have other included controls w_ℓ

Key question: under what assumptions does this SSIV strategy "work"?

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - $\rightarrow$ $s_{\ell n} \in [0,1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - $\rightarrow$ g_n are the *shocks* (or “shifters”)

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - $\rightarrow$ $s_{\ell n} \in [0,1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - $\rightarrow$ g_n are the *shocks* (or “shifters”)

- Bartik (1991); Blanchard and Katz (1992):
  - $\rightarrow$ β = inverse local labor supply elasticity
  - $\rightarrow$ x_ℓ and y_ℓ = employment and wage growth in region ℓ
  - $\rightarrow$ Need a labor demand shifter as an IV

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - $\rightarrow$ $s_{\ell n} \in [0,1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - $\rightarrow$ g_n are the *shocks* (or "shifters")
- Bartik (1991); Blanchard and Katz (1992):
  - $\rightarrow$ β = inverse local labor supply elasticity
  - $\rightarrow$ x_ℓ and y_ℓ = employment and wage growth in region ℓ
  - $\rightarrow$ Need a labor demand shifter as an IV
  - $\rightarrow$ g_n = national growth of industry n
  - $\rightarrow$ $s_{\ell n}$ = lagged employment shares (of industry in a region)
  - $\rightarrow$ z_ℓ = predicted employment growth due to national industry trends

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - $\rightarrow$ $s_{\ell n} \in [0, 1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - $\rightarrow$ g_n are the *shocks* (or "shifters")

- Autor, Dorn, and Hanson (2013, ADH):
  - $\rightarrow$ x_ℓ = growth of import competition in region ℓ
  - $\rightarrow$ y_ℓ = growth of manuf. employment, unemployment, etc.
  - $\rightarrow$ g_n = growth of China exports in manufacturing industry n to 8 other (i.e. non-U.S.) countries
  - $\rightarrow$ $s_{\ell n}$ = 10-year lagged employment shares (over total employment)
  - $\rightarrow$ z_ℓ = predicted growth of import competition

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - → $s_{\ell n} \in [0,1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - → g_n are the *shocks* (or "shifters")

- "Enclave instrument", e.g. Card (2009)
  - → β = inverse elasticity of substitution between native and immigrant labor of some skill level (need a relative labor supply instrument)
  - → x_ℓ and y_ℓ = relative employment and wage in region ℓ
  - → g_n = national immigration growth from origin country n
  - → $s_{\ell n}$ = lagged shares of migrants from origin n in region ℓ
  - → z_ℓ = share of migrants predicted from enclaves & recent growth

SSIV Examples

- Instrument $z_\ell = \sum_n s_{\ell n} g_n$ for model $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$
  - $\rightarrow$ $s_{\ell n} \in [0, 1]$ are the *exposure shares*; often $\sum_n s_{\ell n} = 1$
  - $\rightarrow$ g_n are the *shocks* (or "shifters")
- Hummels et al. (2014) on offshoring:
  - $\rightarrow$ x_ℓ = imports by Danish firm ℓ, y_ℓ = wages
  - $\rightarrow$ g_n = changes in transport costs by $n =$ (product, country)
  - $\rightarrow$ $s_{\ell n}$ = lagged import shares

What Do We Do With This?

Of course we can always run IV with such z_ℓ ... but what does the corresponding estimand *identify*?

What Do We Do With This?

Of course we can always run IV with such z_ℓ ... but what does the corresponding estimand *identify*?

Recall IV validity condition: $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0$ for model residual ε_ℓ

- Looks a little different than normal because we're not assuming *i.i.d.* sampling, i.e. $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = E[z_\ell \varepsilon_\ell]$ (you'll see why soon!)

What Do We Do With This?

Of course we can always run IV with such z_ℓ ... but what does the corresponding estimand *identify*?

Recall IV validity condition: $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0$ for model residual ε_ℓ

- Looks a little different than normal because we're not assuming *i.i.d.* sampling, i.e. $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = E[z_\ell \varepsilon_\ell]$ (you'll see why soon!)

What properties of shocks and shares make this condition hold?

- Is SSIV like a natural experiment? A diff-in-diff? Something new?
- Since z_ℓ combines multiple sources of variation, it can be difficult to think about it being randomly assigned across ℓ...

Roadmap

Introductions

- Me and This Course
- (Linear) SSIV

Shock Exogeneity

- Motivation
- Borusyak et al. (2022)

Share Exogeneity

- Motivation
- Goldsmith-Pinkham et al. (2020)

Choosing an Appropriate Framework

Exogenous Shocks in Industry-Level Regressions

Acemoglu-Autor-Dorn-Hanson-Price (AADHP, 2016) look at the effects of import competition with China on US manufacturing *industries*:

$$\Delta \log Emp_{nt} = \alpha + \beta \Delta IP_{nt} + \varepsilon_{nt},$$

where ΔIP_{nt} measures growth in import penetration from China in industry n, and ε_{nt} captures industry demand/productivity shocks

Exogenous Shocks in Industry-Level Regressions

Acemoglu-Autor-Dorn-Hanson-Price (AADHP, 2016) look at the effects of import competition with China on US manufacturing *industries*:

$$\Delta \log Emp_{nt} = \alpha + \beta \Delta IP_{nt} + \varepsilon_{nt},$$

where ΔIP_{nt} measures growth in import penetration from China in industry n, and ε_{nt} captures industry demand/productivity shocks

Two Key Problems with OLS estimation:

1. Endogeneity of ΔIP_{nt}: OLS is not consistent for β
2. GE spillovers: β does not capture aggregate effects

Problem 1: Endogeneity of ΔIP_{nt}

$$\Delta \log Emp_{nt} = \alpha + \beta \Delta IP_{nt} + \varepsilon_{nt}$$

- ΔIP_{nt} is driven by productivity shocks in China, but also potentially by productivity and demand shocks in the US
  - $\rightarrow$ ε_{nt} captures productivity and demand shocks in the US

Problem 1: Endogeneity of ΔIP_{nt}

$$\Delta \log Emp_{nt} = \alpha + \beta \Delta IP_{nt} + \varepsilon_{nt}$$

- ΔIP_{nt} is driven by productivity shocks in China, but also potentially by productivity and demand shocks in the US
  - $\rightarrow$ ε_{nt} captures productivity and demand shocks in the US
- AADHP instrument ΔIP_{nt} with ΔIPO_{nt}, measuring average Chinese import penetration growth in 8 non-US countries

Problem 1: Endogeneity of ΔIP_{nt}

$$\Delta \log Emp_{nt} = \alpha + \beta \Delta IP_{nt} + \varepsilon_{nt}$$

- ΔIP_{nt} is driven by productivity shocks in China, but also potentially by productivity and demand shocks in the US
  - $\rightarrow$ ε_{nt} captures productivity and demand shocks in the US
- AADHP instrument ΔIP_{nt} with ΔIPO_{nt}, measuring average Chinese import penetration growth in 8 non-US countries
  - $\rightarrow$ Relevance: both ΔIP_{nt} and ΔIPO_{nt} are driven by the same Chinese productivity shocks
  - $\rightarrow$ Validity: local productivity/demand shocks in the US are uncorrelated with those of other countries (entering ΔIPO_{nt})

Identification from a Natural Experiment

Suppose ΔIPO_{nt} is as-good-as-randomly assigned, as in a RCT:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = \mu \qquad \text{for all } n, t$$

where $\mathcal{I} = \{\varepsilon_{nt}, \text{pre-trends, balance variables}, \dots\}$

Identification from a Natural Experiment

Suppose ΔIPO_{nt} is as-good-as-randomly assigned, as in a RCT:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = \mu \qquad \text{for all } n, t$$

where $\mathcal{I} = \{\varepsilon_{nt}, \text{pre-trends, balance variables}, \dots\}$

Consistent IV estimation then follows from many observations of nt, with sufficiently independent variation in ΔIPO_{nt}

Identification from a Natural Experiment

Suppose ΔIPO_{nt} is as-good-as-randomly assigned, as in a RCT:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = \mu \qquad \text{for all } n, t$$

where $\mathcal{I} = \{\varepsilon_{nt}, \text{pre-trends, balance variables}, \dots\}$

Consistent IV estimation then follows from many observations of nt, with sufficiently independent variation in ΔIPO_{nt}

Can relax to add observables capturing systematic variation:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = q'_{nt}\mu \qquad \text{for all } n, t$$

where q_{nt} may include:

- period FE, isolating within-period variation in the shocks
- FE of 10 broad sectors, isolating within-sector variation, etc.

Identification from a Natural Experiment

Suppose ΔIPO_{nt} is as-good-as-randomly assigned, as in a RCT:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = \mu \qquad \text{for all } n, t$$

where $\mathcal{I} = \{\varepsilon_{nt}, \text{pre-trends, balance variables, } \dots\}$

Consistent IV estimation then follows from many observations of nt, with sufficiently independent variation in ΔIPO_{nt}

Can relax to add observables capturing systematic variation:

$$E[\Delta IPO_{nt} \mid \mathcal{I}] = q'_{nt}\mu \qquad \text{for all } n, t$$

where q_{nt} may include:

- period FE, isolating within-period variation in the shocks
- FE of 10 broad sectors, isolating within-sector variation, etc.

We would then just want to control for q_{nt} in the industry-level IV

Problem 2: GE Spillovers

Spillovers across different industries are likely important:

- When employment shrinks in industry n after a negative shock, aggregate employment may or may not respond

Problem 2: GE Spillovers

Spillovers across different industries are likely important:

- When employment shrinks in industry n after a negative shock, aggregate employment may or may not respond
- In a flexible labor market, comparing wages of similar workers across industries does not make sense

Problem 2: GE Spillovers

Spillovers across different industries are likely important:

- When employment shrinks in industry n after a negative shock, aggregate employment may or may not respond
- In a flexible labor market, comparing wages of similar workers across industries does not make sense

ADH Solution: specify the outcome equation for local labor markets

- Works if local economies are isolated “islands”
  (simple model in Adao-Kolesar-Morales 2019; richer structure of spatial spillovers in Adao-Arkolakis-Esposito 2020)

Problem 2: GE Spillovers

Spillovers across different industries are likely important:

- When employment shrinks in industry n after a negative shock, aggregate employment may or may not respond
- In a flexible labor market, comparing wages of similar workers across industries does not make sense

ADH Solution: specify the outcome equation for local labor markets

- Works if local economies are isolated "islands" (simple model in Adao-Kolesar-Morales 2019; richer structure of spatial spillovers in Adao-Arkolakis-Esposito 2020)

But correct specification is not the same as identification!

- Key point: the same industry-level natural experiment can be used to estimate a regional specification, via SSIV

Borusyak, Hull, and Jaravel (BHJ; 2022)

Consider the SSIV estimator of $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$ instrumented by $z_\ell = \sum_n s_{\ell n} g_n$ and, for now, $\sum_n s_{\ell n} = 1$ for all ℓ

- Reduced-form allowed: $x_\ell = z_\ell$
- Only the shift-share structure of z_ℓ matters; x_ℓ can be anything
- Note: view g_n as stochastic, so can't assume z_ℓ is iid

Borusyak, Hull, and Jaravel (BHJ; 2022)

Consider the SSIV estimator of $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$ instrumented by $z_\ell = \sum_n s_{\ell n} g_n$ and, for now, $\sum_n s_{\ell n} = 1$ for all ℓ

- Reduced-form allowed: $x_\ell = z_\ell$
- Only the shift-share structure of z_ℓ matters; x_ℓ can be anything
- Note: view g_n as stochastic, so can't assume z_ℓ is iid

E.g. $g_n = \Delta IPO_n$ aggregated w/mfg employment shares $s_{\ell n}$

- Can we leverage a natural experiment in g_n, as before?

Borusyak, Hull, and Jaravel (BHJ; 2022)

Consider the SSIV estimator of $y_\ell = \beta x_\ell + \gamma' w_\ell + \varepsilon_\ell$ instrumented by $z_\ell = \sum_n s_{\ell n} g_n$ and, for now, $\sum_n s_{\ell n} = 1$ for all ℓ

- Reduced-form allowed: $x_\ell = z_\ell$
- Only the shift-share structure of z_ℓ matters; x_ℓ can be anything
- Note: view g_n as stochastic, so can't assume z_ℓ is iid

E.g. $g_n = \Delta IPO_n$ aggregated w/mfg employment shares $s_{\ell n}$

- Can we leverage a natural experiment in g_n, as before?

First step: note that by the FWL thm. the estimator can be written

$$\hat{\beta} = \frac{\sum_\ell z_\ell y_\ell^\perp}{\sum_\ell z_\ell x_\ell^\perp} = \frac{\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp}$$

where $v_\ell^\perp$ denotes sample residuals from regressing v_ℓ on w_ℓ

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level “aggregate” of the treatment:

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level "aggregate" of the treatment:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp} =$$

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level “aggregate” of the treatment:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp} = \frac{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} y_\ell^\perp}{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} x_\ell^\perp} =$$

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level "aggregate" of the treatment:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp} = \frac{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} y_\ell^\perp}{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} x_\ell^\perp} = \frac{\sum_n s_n g_n \bar{y}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp},$$

where $s_n = \frac{1}{L}\sum_\ell s_{\ell n}$ are weights capturing the average importance of shock n, and $\bar{v}_n = \frac{\sum_\ell s_{\ell n} v_\ell}{\sum_\ell s_{\ell n}}$ is an exposure-weighted average of v_ℓ

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level "aggregate" of the treatment:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp} = \frac{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} y_\ell^\perp}{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} x_\ell^\perp} = \frac{\sum_n s_n g_n \bar{y}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp},$$

where $s_n = \frac{1}{L}\sum_\ell s_{\ell n}$ are weights capturing the average importance of shock n, and $\bar{v}_n = \frac{\sum_\ell s_{\ell n} v_\ell}{\sum_\ell s_{\ell n}}$ is an exposure-weighted average of v_ℓ

The residual $\bar{\varepsilon}_n$ of this shock-level IV procedure is the average residual of observations with a high share of n

- E.g. in ADH, the average unobserved determinants of regional employment in regions most specialized in industry n

BHJ Numerical Equivalence

BHJ show $\hat{\beta}$ can be obtained from a shock-level IV procedure that uses g_n to instrument for a shock-level "aggregate" of the treatment:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n y_\ell^\perp}{\frac{1}{L}\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp} = \frac{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} y_\ell^\perp}{\sum_n g_n \sum_\ell \frac{1}{L} s_{\ell n} x_\ell^\perp} = \frac{\sum_n s_n g_n \bar{y}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp},$$

where $s_n = \frac{1}{L}\sum_\ell s_{\ell n}$ are weights capturing the average importance of shock n, and $\bar{v}_n = \frac{\sum_\ell s_{\ell n} v_\ell}{\sum_\ell s_{\ell n}}$ is an exposure-weighted average of v_ℓ

The residual $\bar{\varepsilon}_n$ of this shock-level IV procedure is the average residual of observations with a high share of n

- E.g. in ADH, the average unobserved determinants of regional employment in regions most specialized in industry n

It follows that $\hat{\beta}$ is consistent iff this shock-level IV procedure is...

BHJ Baseline Assumptions

A1 (Quasi-random shock assignment): $E[g_n \mid \bar{\varepsilon}, s] = \mu, \forall n$

- Each shock has the same expected value, conditional on the shock-level unobservables $\bar{\varepsilon}_n$ and average exposure s_n

BHJ Baseline Assumptions

A1 (Quasi-random shock assignment): $E[g_n \mid \bar{\varepsilon}, s] = \mu, \forall n$

- Each shock has the same expected value, conditional on the shock-level unobservables $\bar{\varepsilon}_n$ and average exposure s_n
- Implies SSIV exogeneity, as $z_\ell = \mu + \sum_n s_{\ell n}(g_n - \mu) = \mu +$ "noise"

BHJ Baseline Assumptions

A1 (Quasi-random shock assignment): $E[g_n \mid \bar{\varepsilon}, s] = \mu, \forall n$

- Each shock has the same expected value, conditional on the shock-level unobservables $\bar{\varepsilon}_n$ and average exposure s_n
- Implies SSIV exogeneity, as $z_\ell = \mu + \sum_n s_{\ell n}(g_n - \mu) = \mu +$ "noise"

A2 (Many uncorrelated shocks):

- $E\left[\sum_n s_n^2\right] \to 0$: expected Herfindahl index of average shock exposure converges to zero (implies $N \to \infty$)
- $Cov(g_n, g_{n'} \mid \bar{\varepsilon}, s) = 0$ for all $n' \neq n$: shocks are mutually uncorrelated given the unobservables

BHJ Baseline Assumptions

A1 (Quasi-random shock assignment): $E[g_n \mid \bar{\varepsilon}, s] = \mu, \forall n$

- Each shock has the same expected value, conditional on the shock-level unobservables $\bar{\varepsilon}_n$ and average exposure s_n
- Implies SSIV exogeneity, as $z_\ell = \mu + \sum_n s_{\ell n}(g_n - \mu) = \mu +$ "noise"

A2 (Many uncorrelated shocks):

- $E\left[\sum_n s_n^2\right] \to 0$: expected Herfindahl index of average shock exposure converges to zero (implies $N \to \infty$)
- $Cov(g_n, g_{n'} \mid \bar{\varepsilon}, s) = 0$ for all $n' \neq n$: shocks are mutually uncorrelated given the unobservables
- Imply a shock-level law of large numbers: $\sum_n s_n g_n \bar{\varepsilon}_n \xrightarrow{p} 0$

BHJ Baseline Assumptions

A1 (Quasi-random shock assignment): $E[g_n \mid \bar{\varepsilon}, s] = \mu, \forall n$

- Each shock has the same expected value, conditional on the shock-level unobservables $\bar{\varepsilon}_n$ and average exposure s_n
- Implies SSIV exogeneity, as $z_\ell = \mu + \sum_n s_{\ell n}(g_n - \mu) = \mu +$ "noise"

A2 (Many uncorrelated shocks):

- $E\left[\sum_n s_n^2\right] \to 0$: expected Herfindahl index of average shock exposure converges to zero (implies $N \to \infty$)
- $Cov(g_n, g_{n'} \mid \bar{\varepsilon}, s) = 0$ for all $n' \neq n$: shocks are mutually uncorrelated given the unobservables
- Imply a shock-level law of large numbers: $\sum_n s_n g_n \bar{\varepsilon}_n \xrightarrow{p} 0$

Both assumptions, while novel for SSIV, would be standard for a shock-level IV regression with weights s_n and instrument g_n

BHJ Extensions

Conditional Quasi-Random Assignment: $E[g_n \mid \bar{\varepsilon}, q, s] = q_n' \mu$ for some observed shock-level variables q_n

- Consistency follows when $w_\ell = \sum_n s_{\ell n} q_n$ is controlled for in the IV

BHJ Extensions

Conditional Quasi-Random Assignment: $E[g_n \mid \bar{\varepsilon}, q, s] = q_n' \mu$ for some observed shock-level variables q_n

- Consistency follows when $w_\ell = \sum_n s_{\ell n} q_n$ is controlled for in the IV

Weakly Mutually Correlated Shocks: $g_n \mid (\bar{\varepsilon}, q, s)$ are clustered or otherwise mutually dependent

- Consistency follows when mutual correlation is not too strong

BHJ Extensions

Conditional Quasi-Random Assignment: $E[g_n \mid \bar{\varepsilon}, q, s] = q_n' \mu$ for some observed shock-level variables q_n

- Consistency follows when $w_\ell = \sum_n s_{\ell n} q_n$ is controlled for in the IV

Weakly Mutually Correlated Shocks: $g_n \mid (\bar{\varepsilon}, q, s)$ are clustered or otherwise mutually dependent

- Consistency follows when mutual correlation is not too strong

Estimated Shocks: $g_n = \sum_\ell w_{\ell n} g_{\ell n}$ proxies for an infeasible g_n^*

- Consistency may require a "leave-out" adjustment: $z_\ell = \sum_\ell s_{\ell n} \tilde{g}_{\ell n}$ for $\tilde{g}_{\ell n} = \sum_{\ell' \neq \ell} \omega_{\ell' n} g_{\ell' n}$ (akin to JIVE solution to many-IV bias)

BHJ Extensions (cont.)

Panel Data: Have $(y_{\ell t}, x_{\ell t}, s_{\ell nt}, g_{nt})$ across $\ell = 1, \dots, L$, $t = 1, \dots, T$

- Consistency can follow from either $N \to \infty$ or $T \to \infty$
- Unit fixed effects “de-mean” the shocks, if $s_{\ell nt}$ are time-invariant

BHJ Extensions (cont.)

Panel Data: Have $(y_{\ell t}, x_{\ell t}, s_{\ell n t}, g_{nt})$ across $\ell = 1, \dots, L$, $t = 1, \dots, T$

- Consistency can follow from either $N \to \infty$ or $T \to \infty$
- Unit fixed effects "de-mean" the shocks, if $s_{\ell n t}$ are time-invariant

Heterogeneous Effects: LATE theorem logic goes through

- Under a first-stage monotonicity condition, SSIV identifies a convex weighted average of heterogeneous treatment effects

Practical Consideration 1: Incomplete Sharess

So far we have assumed a constant sum-of-shares: $S_\ell \equiv \sum_n s_{\ell n} = 1$

- But in some settings, S_ℓ varies across ℓ
- E.g. in ADH, S_ℓ is region ℓ's share of non-manufacturing emp., since $s_{\ell n}$ is the share of manufacturing industry n in *total* emp.

Practical Consideration 1: Incomplete Sharess

So far we have assumed a constant sum-of-shares: $S_\ell \equiv \sum_n s_{\ell n} = 1$

- But in some settings, S_ℓ varies across ℓ
- E.g. in ADH, S_ℓ is region ℓ's share of non-manufacturing emp., since $s_{\ell n}$ is the share of manufacturing industry n in *total* emp.

BHJ show that A1/A2 are not enough for validity of z_ℓ in this case

- Now $z_\ell = \sum_n s_{\ell n} \left(\mu + (g_n - \mu)\right) = \mu S_\ell + \sum_n s_{\ell n}(g_n - \mu)$
- So z_ℓ is mechanically correlated with S_ℓ, which may be endogenous

Practical Consideration 1: Incomplete Sharess

So far we have assumed a constant sum-of-shares: $S_\ell \equiv \sum_n s_{\ell n} = 1$

- But in some settings, S_ℓ varies across ℓ
- E.g. in ADH, S_ℓ is region ℓ's share of non-manufacturing emp., since $s_{\ell n}$ is the share of manufacturing industry n in *total* emp.

BHJ show that A1/A2 are not enough for validity of z_ℓ in this case

- Now $z_\ell = \sum_n s_{\ell n} \left(\mu + (g_n - \mu)\right) = \mu S_\ell + \sum_n s_{\ell n}(g_n - \mu)$
- So z_ℓ is mechanically correlated with S_ℓ, which may be endogenous

Controlling for the sum-of-shares S_ℓ isolates clean shock variation

- Further controls are needed when A1 only holds conditional on q_n; e.g. in panels, S_ℓ should be interacted with time FE

Practical Consideration 2: Exposure Clustering

Adão, Kolesar, and Morales (2019) study a novel inference challenge when SSIV identification leverages quasi-random shocks

- Observations with similar shares $s_{\ell 1,}, \dots, s_{\ell N}$ are likely to have correlated z_ℓ, even when not "clustered" in conventional ways (e.g. by distance)
- When ε_ℓ is similarly clustered (e.g. when $\varepsilon_\ell = \sum_n s_{\ell n}\nu_n + \tilde{\varepsilon}_\ell$), large-sample distribution of $\hat{\beta}$ may not be well-approximated by standard central limit theorems (CLTs)

Practical Consideration 2: Exposure Clustering

Adão, Kolesar, and Morales (2019) study a novel inference challenge when SSIV identification leverages quasi-random shocks

- Observations with similar shares $s_{\ell 1,}, \dots, s_{\ell N}$ are likely to have correlated z_ℓ, even when not "clustered" in conventional ways (e.g. by distance)
- When ε_ℓ is similarly clustered (e.g. when $\varepsilon_\ell = \sum_n s_{\ell n}\nu_n + \tilde{\varepsilon}_\ell$), large-sample distribution of $\hat{\beta}$ may not be well-approximated by standard central limit theorems (CLTs)

They then derive a new CLT + SEs to address "exposure clustering"

- "Design-based": leverage *iid*ness of shocks, not observations

Practical Consideration 2: Exposure Clustering

BHJ show a convenient solution to exposure clustering:

Usual robust/clustered SEs can be valid when $\hat{\beta}$ is given by estimating

$$\bar{y}_n^{\perp} = \alpha + \beta \bar{x}_n^{\perp} + q_n' \tau + \bar{\varepsilon}_n^{\perp},$$

instrumenting $\bar{x}_n^{\perp}$ by g_n and weighting by s_n

Practical Consideration 2: Exposure Clustering

BHJ show a convenient solution to exposure clustering:

Usual robust/clustered SEs can be valid when $\hat{\beta}$ is given by estimating

$$\bar{y}_n^{\perp} = \alpha + \beta \bar{x}_n^{\perp} + q_n' \tau + \bar{\varepsilon}_n^{\perp},$$

instrumenting $\bar{x}_n^{\perp}$ by g_n and weighting by s_n

- Numerically identical IV estimate, when controls include $\sum_n s_{\ell n} q_n$
- Clustering logic: valid SEs are obtained when estimating the IV at the level of identifying variation (here, shocks)

Practical Consideration 2: Exposure Clustering

BHJ show a convenient solution to exposure clustering:

Usual robust/clustered SEs can be valid when $\hat{\beta}$ is given by estimating

$$\bar{y}_n^{\perp} = \alpha + \beta \bar{x}_n^{\perp} + q_n' \tau + \bar{\varepsilon}_n^{\perp},$$

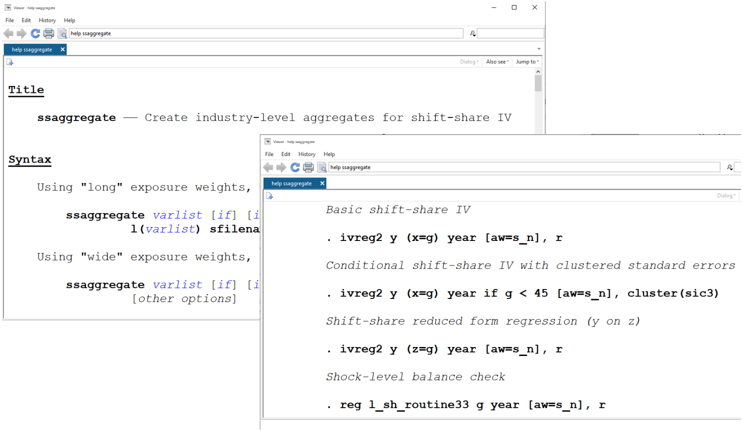
instrumenting $\bar{x}_n^{\perp}$ by g_n and weighting by s_n

- Numerically identical IV estimate, when controls include $\sum_n s_{\ell n} q_n$
- Clustering logic: valid SEs are obtained when estimating the IV at the level of identifying variation (here, shocks)

Same logic applies to performing valid balance/pre-trend tests and evaluating first-stage strength of the instrument

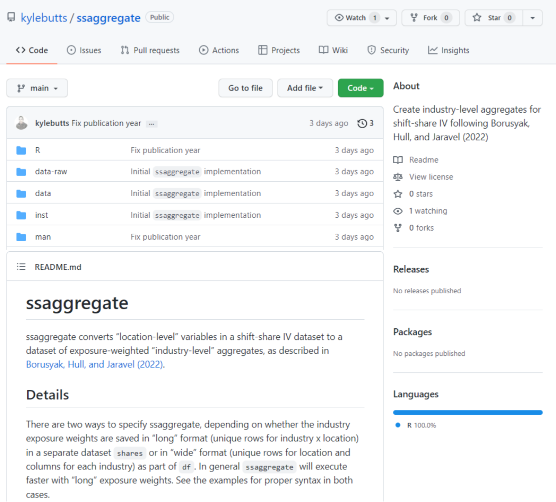
SSIV with *ssaggregate*

Stata package *ssaggregate* leverages the BHJ equivalence result: it translates data to the shock level, after which researchers can proceed with familiar estimation commands (install w/ *ssc install ssaggregate*)



SSIV with *ssaggregate*...in R!

Thanks to our own Kyle Butts, *ssaggregate* is now available in R too!



Download at `https://github.com/kylebutts/ssaggregate`

Roadmap

The Mariel Boatlift as a Basic SSIV

Card (1990) leverages a big migration “push” of low-skilled workers from Cuba to Miami, where Cubans historically settled

The Mariel Boatlift as a Basic SSIV

Card (1990) leverages a big migration "push" of low-skilled workers from Cuba to Miami, where Cubans historically settled

Imagine instrumenting immigrant inflows by the lagged share of Cuban workers $s_{\ell,\text{Cuba}}$ in a diff-in-diff setup

- Need parallel trends: regions with more/fewer Cuban workers on similar employment trends

The Mariel Boatlift as a Basic SSIV

Card (1990) leverages a big migration "push" of low-skilled workers from Cuba to Miami, where Cubans historically settled

Imagine instrumenting immigrant inflows by the lagged share of Cuban workers $s_{\ell,\text{Cuba}}$ in a diff-in-diff setup

- Need parallel trends: regions with more/fewer Cuban workers on similar employment trends

This can be viewed as a simple shift-share instrument:

$$s_{\ell,\text{Cuba}} \equiv s_{\ell,\text{Cuba}} \cdot 1 + \sum_{n \neq \text{Cuba}} s_{\ell n} \cdot 0$$

The Mariel Boatlift as a Basic SSIV

Card (1990) leverages a big migration "push" of low-skilled workers from Cuba to Miami, where Cubans historically settled

Imagine instrumenting immigrant inflows by the lagged share of Cuban workers $s_{\ell,\text{Cuba}}$ in a diff-in-diff setup

- Need parallel trends: regions with more/fewer Cuban workers on similar employment trends

This can be viewed as a simple shift-share instrument:

$$s_{\ell,\text{Cuba}} \equiv s_{\ell,\text{Cuba}} \cdot 1 + \sum_{n \neq \text{Cuba}} s_{\ell n} \cdot 0$$

If several migration origins had a push shock, we can pool them together with a more traditional SSIV...

Goldsmith-Pinkham, Sorkin, and Swift (GPSS; 2020)

GPSS view the set of n and values of g_n as fixed, so $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares

Goldsmith-Pinkham, Sorkin, and Swift (GPSS; 2020)

GPSS view the set of n and values of g_n as fixed, so $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares

They then also establish a numerical equivalence: $\hat{\beta}$ can be obtained from an overidentified IV procedure that uses N share instruments $s_{\ell n}$ and a weight matrix based on the shocks g_n

Goldsmith-Pinkham, Sorkin, and Swift (GPSS; 2020)

GPSS view the set of n and values of g_n as fixed, so $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares

They then also establish a numerical equivalence: $\hat{\beta}$ can be obtained from an overidentified IV procedure that uses N share instruments $s_{\ell n}$ and a weight matrix based on the shocks g_n

Sufficient identifying assumption: shares $s_{\ell n}$ are exogenous for each n (like parallel trends when ε_ℓ are unobserved trends)

$$E[\varepsilon_\ell \mid s_{\ell n}] = 0, \ \forall n$$

Goldsmith-Pinkham, Sorkin, and Swift (GPSS; 2020)

GPSS view the set of n and values of g_n as fixed, so $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares

They then also establish a numerical equivalence: $\hat{\beta}$ can be obtained from an overidentified IV procedure that uses N share instruments $s_{\ell n}$ and a weight matrix based on the shocks g_n

Sufficient identifying assumption: shares $s_{\ell n}$ are exogenous for each n (like parallel trends when ε_ℓ are unobserved trends)

$$E[\varepsilon_\ell \mid s_{\ell n}] = 0,\ \forall n \implies E[\textstyle\sum_\ell z_\ell \varepsilon_\ell] = \displaystyle\sum_\ell \sum_n g_n E[s_{\ell n]E[\varepsilon_\ell | s_{\ell n}}] = 0$$

Goldsmith-Pinkham, Sorkin, and Swift (GPSS; 2020)

GPSS view the set of n and values of g_n as fixed, so $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares

They then also establish a numerical equivalence: $\hat{\beta}$ can be obtained from an overidentified IV procedure that uses N share instruments $s_{\ell n}$ and a weight matrix based on the shocks g_n

Sufficient identifying assumption: shares $s_{\ell n}$ are exogenous for each n (like parallel trends when ε_ℓ are unobserved trends)

$$E[\varepsilon_\ell \mid s_{\ell n}] = 0,\ \forall n \implies E[\sum\nolimits_\ell z_\ell \varepsilon_\ell] = \sum_\ell \sum_n g_n E[s_{\ell n]E[\varepsilon_\ell | s_{\ell n}}] = 0$$

This is N moment conditions at the level of observations, e.g. 38 for Card and 397 for ADH (vs. just 1 in BHJ, at the level of industries)

Rotemberg Weights

How does SSIV pool different diff-in-diffs?

- GPSS propose “opening the black box” of overidentified IV by deriving the weights SSIV implicitly puts on each share instrument
- Builds on Rotemberg (1983), so they call these “Rotemberg weights”

$$\hat{\beta} = \sum_n \hat{\alpha}_n \hat{\beta}_n, \text{ where } \underbrace{\hat{\beta}_n = \frac{\sum_\ell s_{\ell n} y_\ell^\perp}{\sum_\ell s_{\ell n} x_\ell^\perp}}_{n\text{-specific IV estimate}} \text{ and } \underbrace{\hat{\alpha}_n = \frac{g_n \sum_\ell s_{\ell n} x_\ell^\perp}{\sum_{n'} g_{n'} \sum_\ell s_{\ell n'} x_\ell^\perp}}_{\text{Rotemberg weight}}$$

Rotemberg Weights

How does SSIV pool different diff-in-diffs?

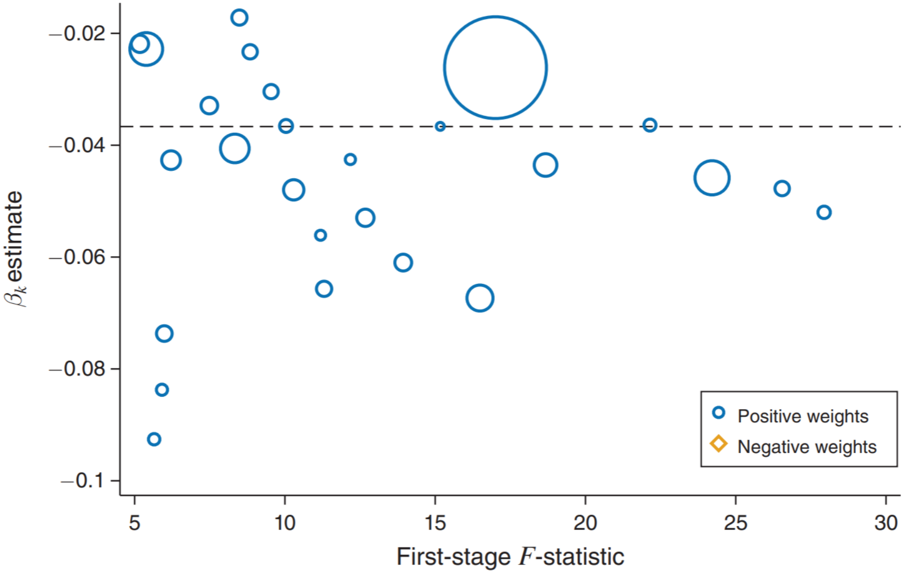
- GPSS propose "opening the black box" of overidentified IV by deriving the weights SSIV implicitly puts on each share instrument
- Builds on Rotemberg (1983), so they call these "Rotemberg weights"

$$\hat{\beta} = \sum_n \hat{\alpha}_n \hat{\beta}_n, \text{ where } \underbrace{\hat{\beta}_n = \frac{\sum_\ell s_{\ell n} y_\ell^\perp}{\sum_\ell s_{\ell n} x_\ell^\perp}}_{n\text{-specific IV estimate}} \text{ and } \underbrace{\hat{\alpha}_n = \frac{g_n \sum_\ell s_{\ell n} x_\ell^\perp}{\sum_{n'} g_{n'} \sum_\ell s_{\ell n'} x_\ell^\perp}}_{\text{Rotemberg weight}}$$

Intuitively, more weight is given to share instruments with more extreme shocks g_n and larger first stages $\sum_\ell s_{\ell n} x_\ell^\perp$

- Weights can be negative (potential issue w/heterogeneous effects)

Rotemberg Weights in Card (2009)



Is Share Exogeneity Plausible?

Share exogeneity assumption is **not** that “shares don’t causally respond to the residual” (they can’t: shares are pre-determined)

- It’s: “all unobservables are uncorrelated with anything about the local share distribution”

Is Share Exogeneity Plausible?

Share exogeneity assumption is **not** that “shares don’t causally respond to the residual” (they can’t: shares are pre-determined)

- It’s: “all unobservables are uncorrelated with anything about the local share distribution”

This sufficient condition is typically violated when there are *any* unobserved shocks ν_n that affect ε_ℓ via the same or correlated shares

- I.e. if $\varepsilon_\ell = \sum_n s_{\ell n}\nu_n + \tilde{\varepsilon}_\ell$, then $s_{\ell n}$ and ε_ℓ cannot be uncorrelated in large samples—even if ν_n are uncorelated with g_n
- E.g. in ADH, unobserved technology shocks across industries affect labor markets via lagged emp. shares, along with observed g_n
- Problem arises when shares are “generic” – predicting many things

Card and ADH Revisited

When share exogeneity is *ex ante* plauible, can test its assumptions *ex post* (focusing on high Rotemberg weight n):

- Balance/pre-trend tests
- Overidentification tests (under constant effects)
- Straightforward to implement; no different than any other IV

Card and ADH Revisited

When share exogeneity is *ex ante* plauible, can test its assumptions *ex post* (focusing on high Rotemberg weight n):

- Balance/pre-trend tests
- Overidentification tests (under constant effects)
- Straightforward to implement; no different than any other IV

GPSS find that balance/overidentification tests broadly pass for Card

... but fail badly for ADH, consistent with *ex ante* implausibility

Roadmap

Introductions

- Me and This Course
- (Linear) SSIV

Shock Exogeneity

- Motivation
- Borusyak et al. (2022)

Share Exogeneity

- Motivation
- Goldsmith-Pinkham et al. (2020)

Choosing an Appropriate Framework

A Taxonomy of SSIV Settings

BHJ distinguish between three cases of SSIVs in the literature

A Taxonomy of SSIV Settings

BHJ distinguish between three cases of SSIVs in the literature

- **Case 1**: the IV is based on a set of shocks which can be thought of as an instrument (i.e. many, plausibly quasi-randomly assigned)
  - $\rightarrow$ BHJ shows how this identifying variation can be mapped to estimate effects at a different "level" (i.e. industries $\rightarrow$ local labor markets)

A Taxonomy of SSIV Settings

BHJ distinguish between three cases of SSIVs in the literature

- **Case 1**: the IV is based on a set of shocks which can be thought of as an instrument (i.e. many, plausibly quasi-randomly assigned)
  - $\rightarrow$ BHJ shows how this identifying variation can be mapped to estimate effects at a different “level” (i.e. industries $\rightarrow$ local labor markets)
- **Case 2**: the researcher does not directly observe many quasi-random shocks, but can estimate them in-sample
  - $\rightarrow$ Canonical setting of Bartik (1991), where g_n are average industry growth rates (thought to proxy for latent demand shocks)
  - $\rightarrow$ See also Card (2009), where national immiration rates are estimated

A Taxonomy of SSIV Settings

BHJ distinguish between three cases of SSIVs in the literature

- **Case 1**: the IV is based on a set of shocks which can be thought of as an instrument (i.e. many, plausibly quasi-randomly assigned)
  - $\rightarrow$ BHJ shows how this identifying variation can be mapped to estimate effects at a different "level" (i.e. industries $\rightarrow$ local labor markets)
- **Case 2**: the researcher does not directly observe many quasi-random shocks, but can estimate them in-sample
  - $\rightarrow$ Canonical setting of Bartik (1991), where g_n are average industry growth rates (thought to proxy for latent demand shocks)
  - $\rightarrow$ See also Card (2009), where national immiration rates are estimated
- **Case 3**: the g_n cannot be naturally viewed as an instrument
  - $\rightarrow$ Either too few or implausibly exogenous, even given some q_n
  - $\rightarrow$ Identification may (or may not) instead follow from share exogeneity

Ex Ante vs. Ex Post Validity

BHJ emphasize that the decision to pursue a “shocks” vs. “shares” identification strategy must be made *ex ante*

- Undesirable to base identifying assumptions on *ex post* tests, though balance/pre-trend tests can be used to falsify assumptions
- The two identification strategies have different economic content

Ex Ante vs. Ex Post Validity

BHJ emphasize that the decision to pursue a “shocks” vs. “shares” identification strategy must be made *ex ante*

- Undesirable to base identifying assumptions on *ex post* tests, though balance/pre-trend tests can be used to falsify assumptions
- The two identification strategies have different economic content

They suggest thinking about whether shares are “tailored” to the economic question/treatment, or are “generic”

- Generic shares (e.g. ADH): unobserved ν_n are likely to enter ε_ℓ via the same or similar shares, violating share exogeneity
- Tailored shares have a diff-in-diff feel; don’t even need the shocks, except to possibly improve power or avoid many-IV bias