Shift-Share IV

Mixtape Track

MIXTAPE SESSIONS

Roadmap

- Shift-Share IV
  - Approach
  - Cautions
  - Recentered IV

Approach

A shift-share instrument takes the form $Z_i = \sum_n s_{in} g_n$ for a set of shocks g_n and a set of exposure shares $s_{in} \geq 0$ (for each i)

Approach

A shift-share instrument takes the form $Z_i = \sum_n s_{in} g_n$ for a set of shocks g_n and a set of exposure shares $s_{in} \geq 0$ (for each i)

- Bartik (1991): national industry employment growth g_n, local industry employment shares s_{in} for regions i
- Autor et al. (2013): increase in (non-U.S.) Chinese import growth across manufacturing industries g_n, local employment shares s_{in}
- Card (2009): growth of immigrant inflows across origin countries g_n, local immigrant shares s_{in}

Approach

A shift-share instrument takes the form $Z_i = \sum_n s_{in} g_n$ for a set of shocks g_n and a set of exposure shares $s_{in} \geq 0$ (for each i)

- Bartik (1991): national industry employment growth g_n, local industry employment shares s_{in} for regions i
- Autor et al. (2013): increase in (non-U.S.) Chinese import growth across manufacturing industries g_n, local employment shares s_{in}
- Card (2009): growth of immigrant inflows across origin countries g_n, local immigrant shares s_{in}

The literature has taken two econometric approaches to such Z_i...

Exogenous Shares

Goldsmith-Pinkham et al. (2020) consider the shocks g_n as fixed numbers and consider the “exogeneity” of the shares: $E[s_{in}\varepsilon_i] = 0$

- Often regressions are run in first-differences, so this is like DD-IV
- The twist here is we have many instruments: In Autor et al. (2013) there are 398 industries n (and $1,444$ regional observations!)

Exogenous Shares

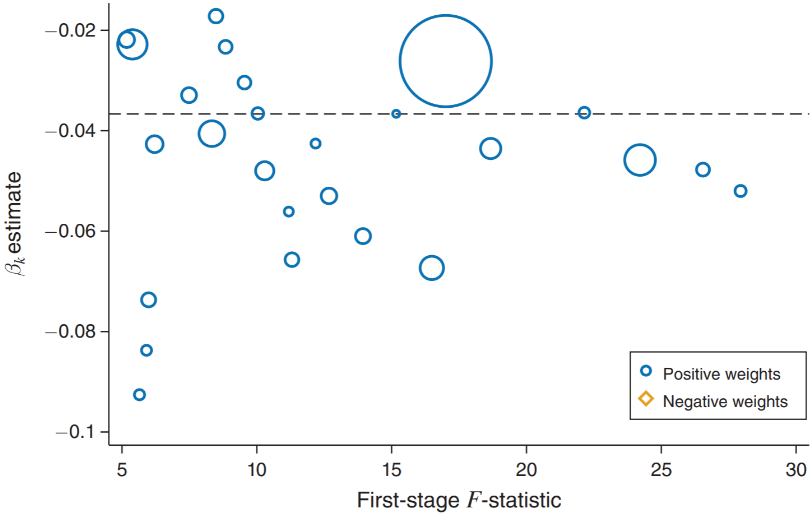
Goldsmith-Pinkham et al. (2020) consider the shocks g_n as fixed numbers and consider the “exogeneity” of the shares: $E[s_{in}\varepsilon_i] = 0$

- Often regressions are run in first-differences, so this is like DD-IV
- The twist here is we have many instruments: In Autor et al. (2013) there are 398 industries n (and $1,444$ regional observations!)

They propose tools to measure the “importance” of different share IVs (“Rotemberg weights”) and discuss other subtlties in estimation

- Kind of like judge IV, except with known “leniency” g_n
- Can check (many) overidentifying restrictions, pre-trends, etc

Rotemberg Weights for Card (2009) Exposure Shares



Source: Goldsmith-Pinkham et al. (2020)

Exogenous Shocks

Borusyak et al. (2022) consider the shocks g_n as exogenous, (quasi-randomly assigned + excludable), conditional on the shares

- E.g. different industries saw higher/lower import growth from China for reasons unrelated to local U.S. employment trends
- Need a “shock-level law of large numbers” (i.e. many shocks)

Exogenous Shocks

Borusyak et al. (2022) consider the shocks g_n as exogenous, (quasi-randomly assigned + excludable), conditional on the shares

- E.g. different industries saw higher/lower import growth from China for reasons unrelated to local U.S. employment trends
- Need a “shock-level law of large numbers” (i.e. many shocks)

They propose tools to test for shock exogeneity (e.g. balance/ pre-trend checks) and quantify the extent of identifying variation

- No overidentifying restrictions: a single instrument g_n, as if we were running an “industry-level” IV regression
- Also show how to relax exogeneity to hold conditional on some observed shock-level confounders

Caution 1: Incomplete Shares

In some shift-share applications exposure weight sum $S_i = \sum_n s_{in}$ varies across observations i

- E.g. in Autor et al. (2013), the total manufacturing share S_i varies

Caution 1: Incomplete Shares

In some shift-share applications exposure weight sum $S_i = \sum_n s_{in}$ varies across observations i

- E.g. in Autor et al. (2013), the total manufacturing share S_i varies

Borusyak et al. (2022) show this can be a problem if you only want to leverage variation in the shocks and not also in S_i

- Intuitively, if $E[g_n|s] = \mu$ then $E[Z_i|s] = E\left[\sum_n s_{in} g_n | s\right] = \mu S_i$, so the "expected instrument" varies non-randomly across observations
- If S_i is correlated with ε_i, this non-random variation can create bias

Addressing Incomplete Shares

An easy fix to incomplete shares is to control for $S_i = \sum_n s_{in}$

- Alternatively, construct shares such that $S_i = 1$ for everyone
- The former may be more powerful if $X_i = \sum_n s_{in}\tilde{g}_{in}$ for $S_i \neq 1$

Addressing Incomplete Shares

An easy fix to incomplete shares is to control for $S_i = \sum_n s_{in}$

- Alternatively, construct shares such that $S_i = 1$ for everyone
- The former may be more powerful if $X_i = \sum_n s_{in}\tilde{g}_{in}$ for $S_i \neq 1$

If other controls are needed to make the shocks as-good-as- random (e.g. time dummies, to isolate within-period variation) then S_i needs to be added as an *interaction* with them

- In Autor et al. (2013), this means interacting the manufacturing sum-of-shares with period FE...

Sum-of-Share Controls in Autor et al. (2013)

Table 4: Shift-Share IV Estimates of the Effect of Chinese Imports on Manufacturing Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Coefficient	-0.596	-0.489	-0.267	-0.314	-0.310	-0.290	-0.432
	(0.114)	(0.100)	(0.099)	(0.107)	(0.134)	(0.129)	(0.205)
Regional controls							
Autor et al. (2013) controls	✓	✓	✓		✓	✓	✓
Start-of-period mfg. share	✓						
Lagged mfg. share		✓	✓	✓	✓	✓	✓
Period-specific lagged mfg. share			✓	✓	✓	✓	✓
Lagged 10-sector shares					✓		✓
Local Acemoglu et al. (2016) controls						✓	
Lagged industry shares							✓
SSIV first stage F-stat.	185.6	166.7	123.6	272.4	64.6	63.3	27.6
# of region-periods	1,444	1,444	1,444	1,444	1,444	1,444	1,444
# of industry-periods	796	794	794	794	794	794	794

Source: Borusyak et al. (2022)

Caution 2: Exposure Clustering

Adáo et al. (2019) show another problem with exogenous shocks: conventional robust/clustered SEs may be wrong

- Intuitively, the structure of $Z_i = \sum_n s_{in} g_n$ may make observations with similar $s_{i1} \dots s_{in}$ correlated, even when otherwise "far apart"
- They derive non-standard central limit theorems to account for such "exposure clustering" (with R/Stata code)

Caution 2: Exposure Clustering

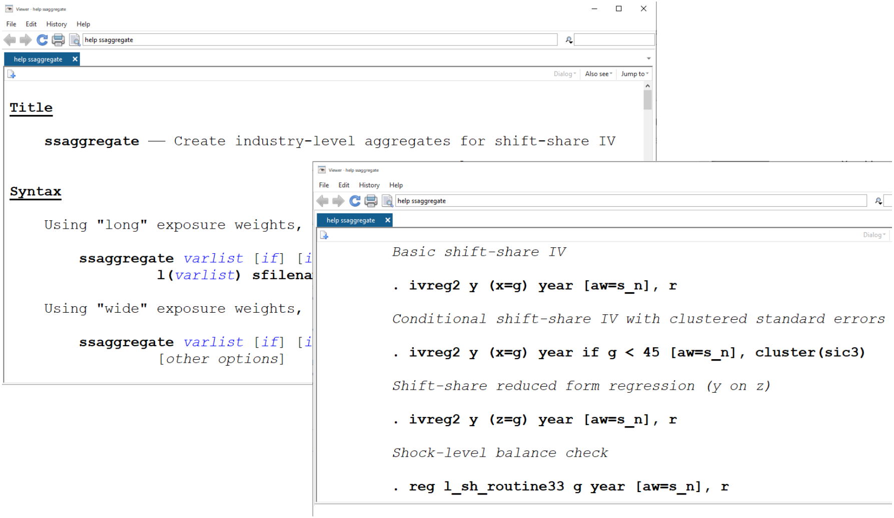
Adáo et al. (2019) show another problem with exogenous shocks: conventional robust/clustered SEs may be wrong

- Intuitively, the structure of $Z_i = \sum_n s_{in} g_n$ may make observations with similar $s_{i1} \dots s_{in}$ correlated, even when otherwise "far apart"
- They derive non-standard central limit theorems to account for such "exposure clustering" (with R/Stata code)

Borusyak et al. (2022) build on this theory to propose an alternative approach: estimate the IV at the level of identifying variation (shocks)

- Derive an equivalent regression where the g_n are used directly as the instrument for shock-level outcomes and treatments
- Standard robust SEs address the exposure clustering problem

Estimating Shock-Level SSIV Regressions



Install in Stata: `ssc install ssaggregate`

Recentered IV

Remember the “expected instrument” in shift-share IV? It turns out the incomplete shares problem may generalize to related settings

- Network spillover IVs (e.g. Miguel and Kremer 2004)
- Transportation upgrade IVs (e.g. Donaldson and Hornbeck 2016)
- Simulated instruments (e.g. Currie and Gruber 1996)
- Nonlinear shift-share (e.g. Chodorow-Reich and Wieland 2020)

Recentered IV

Remember the “expected instrument” in shift-share IV? It turns out the incomplete shares problem may generalize to related settings

- Network spillover IVs (e.g. Miguel and Kremer 2004)
- Transportation upgrade IVs (e.g. Donaldson and Hornbeck 2016)
- Simulated instruments (e.g. Currie and Gruber 1996)
- Nonlinear shift-share (e.g. Chodorow-Reich and Wieland 2020)

Borusyak and Hull (2021) develop a general identification framework for IVs combining multiple sources of variation, w/only some random

- Propose “recentering” to avoid bias from non-random “exposure”

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

μ_i is measured by taking a stand on the shock *assignment process*

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

μ_i is measured by taking a stand on the shock *assignment process*

1. Specify *counterfactual* shocks $\tilde{g}^{(1)}, \ldots, \tilde{g}^{(K)}$ which were as likely to have occured (by, e.g., permuting the rows of g)

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

μ_i is measured by taking a stand on the shock *assignment process*

1. Specify *counterfactual* shocks $\tilde{g}^{(1)}, \ldots, \tilde{g}^{(K)}$ which were as likely to have occured (by, e.g., permuting the rows of g)
2. Recompute $Z_i^{(1)}, \ldots, Z_i^{(K)}$ for each observation i: $Z_i^{(k)} = f_i(\tilde{g}^{(k)}; s)$

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

μ_i is measured by taking a stand on the shock *assignment process*

1. Specify *counterfactual* shocks $\tilde{g}^{(1)}, \ldots, \tilde{g}^{(K)}$ which were as likely to have occured (by, e.g., permuting the rows of g)
2. Recompute $Z_i^{(1)}, \ldots, Z_i^{(K)}$ for each observation i: $Z_i^{(k)} = f_i(\tilde{g}^{(k)}; s)$
3. Average the counterfactual instruments for each i: $\mu_i = \frac{1}{K} \sum_k Z_i^{(k)}$

The Borusyak and Hull (2021) Proposal

Consider a instrument $Z_i = f_i(g; s)$ for some known mapping $f_i(\cdot)$ of exogenous shocks g and non-random exposure s

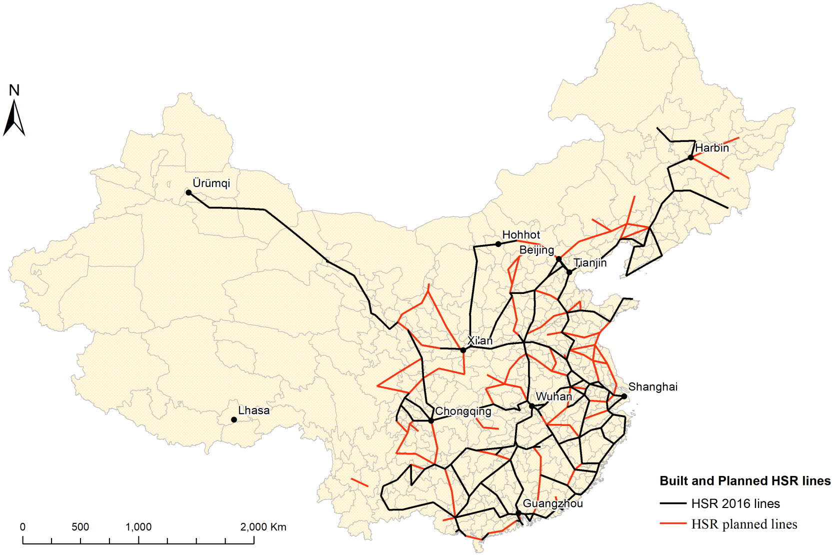
- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

μ_i is measured by taking a stand on the shock *assignment process*

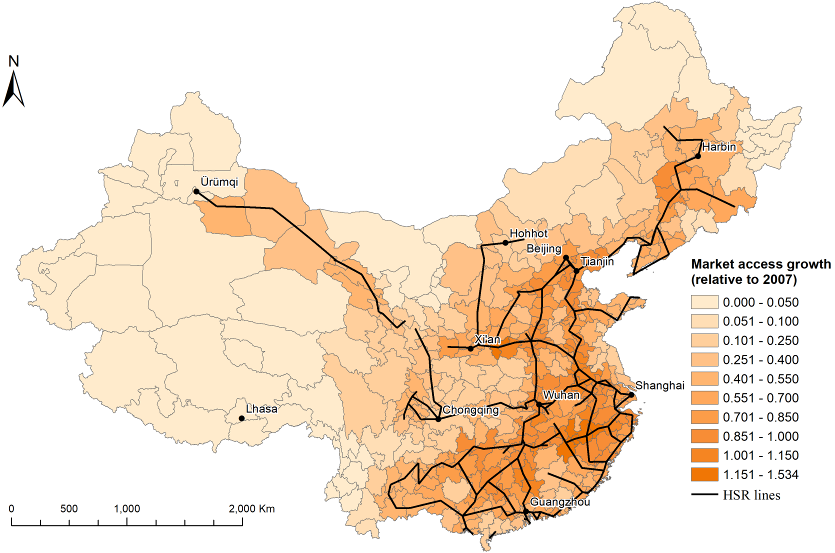
1. Specify *counterfactual* shocks $\tilde{g}^{(1)}, \dots, \tilde{g}^{(K)}$ which were as likely to have occured (by, e.g., permuting the rows of g)
2. Recompute $Z_i^{(1)}, \dots, Z_i^{(K)}$ for each observation i: $Z_i^{(k)} = f_i(\tilde{g}^{(k)}; s)$
3. Average the counterfactual instruments for each i: $\mu_i = \frac{1}{K} \sum_k Z_i^{(k)}$

Besides recentering, μ_i can also be controlled for with the original Z_i

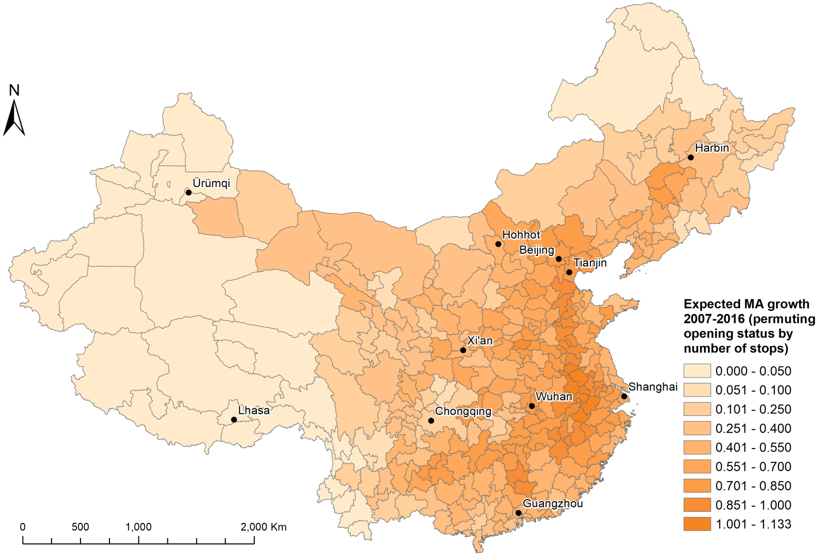
Illustration: High-Speed Rail in China, 2007-2016



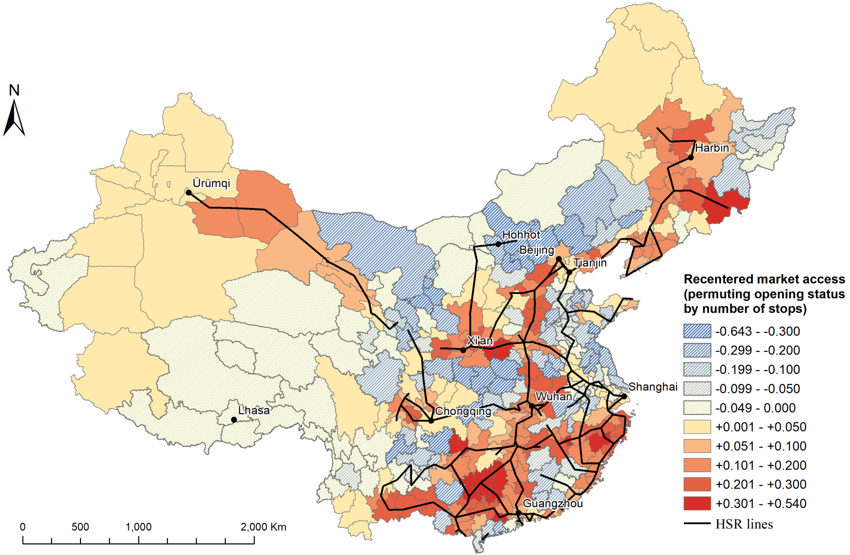
Market Access Growth, Computed from Rail Growth



Expected MA Growth, Assuming Random Rail Timing



Recentered Market Access Growth = Actual - Expected



Recentering Can Matter a Lot Empirically!

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
Panel A. No Controls			
Market Access Growth	0.232	0.081	0.069
	(0.075)	(0.098)	(0.094)
		[-0.315, 0.328]	[-0.209, 0.331]
Expected Market Access Growth			0.318
			(0.095)
Panel B. With Geography Controls			
Market Access Growth	0.132	0.055	0.045
	(0.064)	(0.089)	(0.092)
		[-0.144, 0.278]	[-0.154, 0.281]
Expected Market Access Growth			0.213
			(0.073)
Recentered	No	Yes	Yes
Prefectures	274	274	274

Source: Borusyak and Hull (2021)