Instrumental Variables

IV Frontiers

MIXTAPE SESSIONS

Roadmap

Judge/Examiner IV

- Approach
- Cautions

Shift-Share IV

- Approach

Other Frontiers

- Diff-in-Diff IV
- Recentered IV

Approach

A judge (or examiner) IV design leverages the idiosyncratic assignment of individuals to a set of decision-makers

- Kling (2006): sentencing judges
- Doyle (2007): foster care investigators
- Maestas et al. (2013): SSDI benefit examiners
- Doyle et al. (2015): ambulance companies

Approach

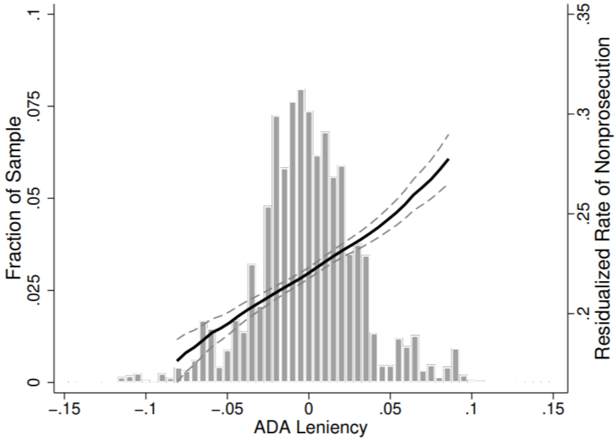
A judge (or examiner) IV design leverages the idiosyncratic assignment of individuals to a set of decision-makers

- Kling (2006): sentencing judges
- Doyle (2007): foster care investigators
- Maestas et al. (2013): SSDI benefit examiners
- Doyle et al. (2015): ambulance companies

The typical approach is to IV a treatment D_i with a measure of the “leniency” $E[D_i \mid Z_i]$ of one’s assigned judge $Z_i \in \{1, \dots, J\}$

- E.g. a leave-one-out average, $\hat{L}_i = \frac{1}{|i' \neq i, Z_{i'} = Z_i|} \sum_{i' \neq i, Z_{i'} = Z_i} D_i$

Agan et al. (2021) “Misdemeanor Prosecution”



Note: This figure shows the distribution of our leave-out mean measure of ADA “leniency,” residualized by court-by-month and court-by-day-of-week. More lenient ADAs have higher rates of not prosecuting nonviolent misdemeanor cases. The solid line is a local linear regression of nonprosecution on ADA leniency, along with the 95% confidence interval, estimated from the 1st to 99th percentiles of ADA leniency—a local linear version of our first stage. A case assigned to a more lenient ADA (computed using all cases except the current case and other cases with the same defendant) has a higher likelihood of being not prosecuted.

Agan et al. (2021) “Misdemeanor Prosecution”

	(1) Nonprosecution	(2) ADA Leniency
Number Counts	-0.019***	-0.000
	(0.003)	(0.000)
Number Misdemeanor Counts	0.018***	0.000
	(0.004)	(0.001)
Number of Serious Misdemeanor Counts	-0.102***	-0.000
	(0.006)	(0.000)
Misd Conviction within Past Year	-0.068***	-0.001
	(0.005)	(0.000)
Felony Conviction within Past Year	-0.053***	-0.001
	(0.006)	(0.001)
Citizen	0.042***	-0.000
	(0.004)	(0.000)
Disorderly/Theft	-0.014*	-0.001
	(0.008)	(0.001)
Motor Vehicle	0.105***	-0.000
	(0.009)	(0.000)
Drug	-0.094***	-0.001
	(0.009)	(0.001)
Constant	0.224***	0.001
	(0.009)	(0.002)
Observations	67553	67553
Joint F-Test p-value	0	0.234

Note: This table reports regressions testing the random assignment of cases to arraigning ADAs. ADA leniency is estimated using data from other nonviolent misdemeanor cases assigned to an arraigning ADA following the procedure described in the text. Column (1) reports estimates from an OLS regression of nonprosecution on the variables listed and court-by-time fixed effects. Column (2) reports estimates from an OLS regression of ADA leniency on the variables listed and court-by-time fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. The p-value reported at the bottom of Columns (1) and (2) is for an F-test of the joint significance of the variables listed with standard errors two-way clustered at the individual and ADA level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Agan et al. (2021) “Misdemeanor Prosecution”

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Criminal Complaint Within 2 Years				
Not Prosecuted	-0.14***	-0.10***	-0.34***	-0.33***
	(0.01)	(0.01)	(0.10)	(0.11)
			[-0.55, -0.13]	[-0.54, -0.10]
Mean Dep Var Prosecuted	0.37			
Mean Dep Var Prosecuted Compliers	0.57			
Panel B: Misdemeanor Complaint Within 2 Years				
Not Prosecuted	-0.08***	-0.06***	-0.24***	-0.24***
	(0.00)	(0.00)	(0.09)	(0.09)
			[-0.42, -0.06]	[-0.43, -0.05]
Mean Dep Var Prosecuted	0.24			
Mean Dep Var Prosecuted Compliers	0.40			
Panel C: Felony Complaint Within 2 Years				
Not Prosecuted	-0.06***	-0.04***	-0.10*	-0.08
	(0.00)	(0.00)	(0.06)	(0.07)
			[-0.22, 0.03]	[-0.21, 0.06]
Mean Dep Var Prosecuted	0.13			
Mean Dep Var Prosecuted Compliers	0.17			
Observations	67553	67553	67553	67553
Court x Time FE	Yes	Yes	Yes	Yes
Case/Def Covariates	No	Yes	No	Yes

Note: This table reports OLS and two-stage least squares estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within two years. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variables are identified in the panel headings. Each panel reports the mean of the dependent variable for all prosecuted defendants, and for prosecuted defendants within the set of compliers. See Appendix C.3 for details on the calculation of mean outcomes among prosecuted compliers. Two-stage least squares models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses in Columns (1)-(4). For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Caution 1: Monotonicity

"Strict" first-stage monotonicity requires judges to have a common ordering of individuals for treatment

- E.g. no differences in "skill" at identifying appropriate cases

Caution 1: Monotonicity

"Strict" first-stage monotonicity requires judges to have a common ordering of individuals for treatment

- E.g. no differences in "skill" at identifying appropriate cases

Imbens and Angrist (& Ridder) saw this coming in 1994!

EXAMPLE 2 (Administrative Screening):[5] Suppose applicants for a social program are screened by two officials. The two officials are likely to have different admission rates, even if the stated admission criteria are identical. Since the identity of the official is probably immaterial to the response, it seems plausible that Condition 1 is satisfied. The instrument is binary so Condition 3 is trivially satisfied. However, Condition 2 requires that if official A accepts applicants with probability $P(0)$, and official B accepts people with probability $P(1) > P(0)$, official B must accept *any* applicant who would have been accepted by official A. This is unlikely to hold if admission is based on a number of criteria. Therefore, in this example we *cannot* use Theorem 1 to identify a local average treatment effect nonparametrically despite the presence of an instrument satisfying Condition 1.

[5] This example was suggested to us by Geert Ridder.

Monotonicity Solutions

Frandsen et al. (2019) formalize a weaker “average monotonicity” condition: intuitively, that skill differences are uncorrelated with TEs

- Similar to de Chaisemartin (2017) “tolerating defiance”
- Also propose non-parametric tests of monotonicity + exclusion (similar to Kitagawa (2015), but with multiple IVs + controls)

Monotonicity Solutions

Frandsen et al. (2019) formalize a weaker “average monotonicity” condition: intuitively, that skill differences are uncorrelated with TEs

- Similar to de Chaisemartin (2017) “tolerating defiance”
- Also propose non-parametric tests of monotonicity + exclusion (similar to Kitagawa (2015), but with multiple IVs + controls)

Other tests include checking whether leniency has the same first stage in different subgroups (Norris, 2021)

- Another solution is to parameterize variation in judge skill and estimate it jointly with TEs (Chan et al. 2021; Arnold et al. 2021)

Caution 2: Exclusion

“Strict” exclusion requires judges to only affect the outcome through one treatment channel

- E.g. a judge more likely to sentence a defendant to jail does not differentially change sentence conditions

Caution 2: Exclusion

“Strict” exclusion requires judges to only affect the outcome through one treatment channel

- E.g. a judge more likely to sentence a defendant to jail does not differentially change sentence conditions

Like monotonicity, this can be weakened to an “on average” condition

- Kolesár et al. (2015): exclusion restriction violations are uncorrelated with leniency variation (see also Angrist et al. 2021)
- Need many judges for a “judge-level law of large numbers” to kick in

Adding Treatment Channels

Of course if multiple potential treatment channels are observed they can be included + instrumented by judges

- See Autor/Maestas/Mullen/Strand (2017), which adds a decision-time treatment to Maestas et al. (2013)
- Two instruments: examiner leniency and (leave-out) average examiner decision time

Adding Treatment Channels

Of course if multiple potential treatment channels are observed they can be included + instrumented by judges

- See Autor/Maestas/Mullen/Strand (2017), which adds a decision-time treatment to Maestas et al. (2013)
- Two instruments: examiner leniency and (leave-out) average examiner decision time

Careful though: IV with multiple treatments can be difficult to interpret in a LATE framework (maybe OK as a robustness check)

- See e.g. Kirkeboen et al. (2016) and Kline and Walters (2016)

Caution 3: Leniency Estimation

One concern that doesn't get enough attention (IMO) is the fact that judge leniency is estimated: after all, we don't know $E[D_i \mid Z_i]$!

Caution 3: Leniency Estimation

One concern that doesn't get enough attention (IMO) is the fact that judge leniency is estimated: after all, we don't know $E[D_i \mid Z_i]$!

- Estimating leniency as (non-leave-out) sample averages == using 2SLS with judge dummies (recall the "2S" in "2SLS'!')

Caution 3: Leniency Estimation

One concern that doesn't get enough attention (IMO) is the fact that judge leniency is estimated: after all, we don't know $E[D_i \mid Z_i]$!

- Estimating leniency as (non-leave-out) sample averages == using 2SLS with judge dummies (recall the "2S" in "2SLS'!')
- For leave-out averages, the equivalent regression uses Jacknife Instrumental Variables Estimation (JIVE; Angrist et al. 1999)

Caution 3: Leniency Estimation

One concern that doesn't get enough attention (IMO) is the fact that judge leniency is estimated: after all, we don't know $E[D_i \mid Z_i]$!

- Estimating leniency as (non-leave-out) sample averages == using 2SLS with judge dummies (recall the "2S" in "2SLS'!')
- For leave-out averages, the equivalent regression uses Jacknife Instrumental Variables Estimation (JIVE; Angrist et al. 1999)

JIVE may be better with many judge instruments (as it can avoid 2SLS many-weak bias), but it is not bulletproof

- Kolesár (2013) shows many-weak bias can creep back in with many covariates (e.g. court-by-time FE, needed to make judges random)

State-of-the-Art: UJIVE

Kolesár (2013) also derives a solution to many-IV/many-control bias

- "Unbiased" Jackknife Instrumental Variables Estimation (UJIVE) adjusts the leave-out means for controls by (basically) leave-out-Frisch-Waugh-Lovell residualization

State-of-the-Art: UJIVE

Kolesár (2013) also derives a solution to many-IV/many-control bias

- "Unbiased" Jackknife Instrumental Variables Estimation (UJIVE) adjusts the leave-out means for controls by (basically) leave-out-Frisch-Waugh-Lovell residualization

Michal Kolesár, Paul Goldsmith-Pinkham, and I are currently working on a Stata package to implement UJIVE

- We hope to publish it and an accompanying R package soon! In the meantime you'll be one of the first to beta-test our current code...

UJIVE Repo (In Progress)

`version 0.5.0 14Nov2021` | Installation | Usage | Examples | Compiling

Installation

From the command line:

```
git clone git@github.com:mcaceresb/stata-manyiv
```

(or download the code manually and unzip). From Stata:

```
cap noi net uninstall manyiv
net install manyiv, from(`c(pwd)'/stata-manyiv)
```

(Change `stata-manyiv` if you download the package to a different folder; e.g. `stata-manyiv-main`.) Note if the repo were public, this could be installed directly from Stata:

```
local github "https://raw.githubusercontent.com"
net install manyiv, from(`github'/mcaceresb/stata-manyiv/master/)
```

Usage

```
manyiv depvar (endogenous = instrument) [exogenous], options
help manyiv
```

Roadmap

Judge/Examiner IV

- Approach
- Cautions

Shift-Share IV

- Approach

Other Frontiers

- Diff-in-Diff IV
- 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

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

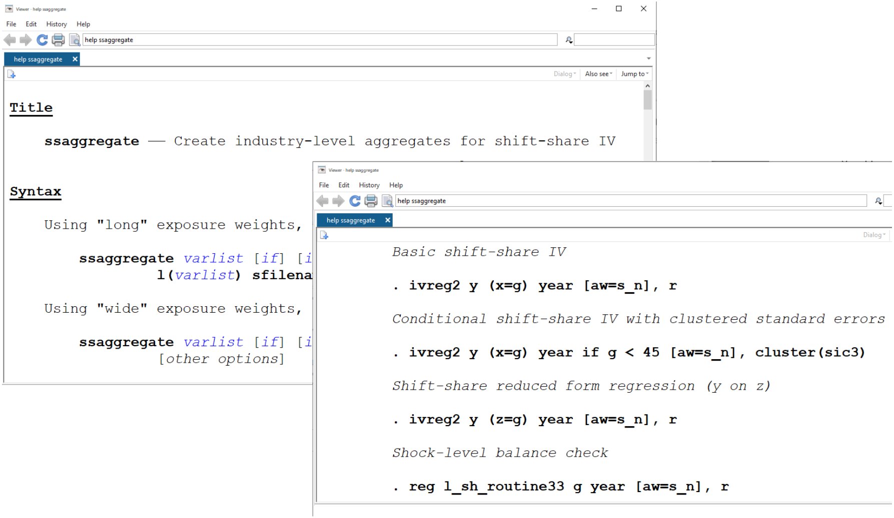
Borusyak et al. (2022) consider the shocks g_n as exogenous, (quasi-randomly assigned + excludable), conditional on the shares

- 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)

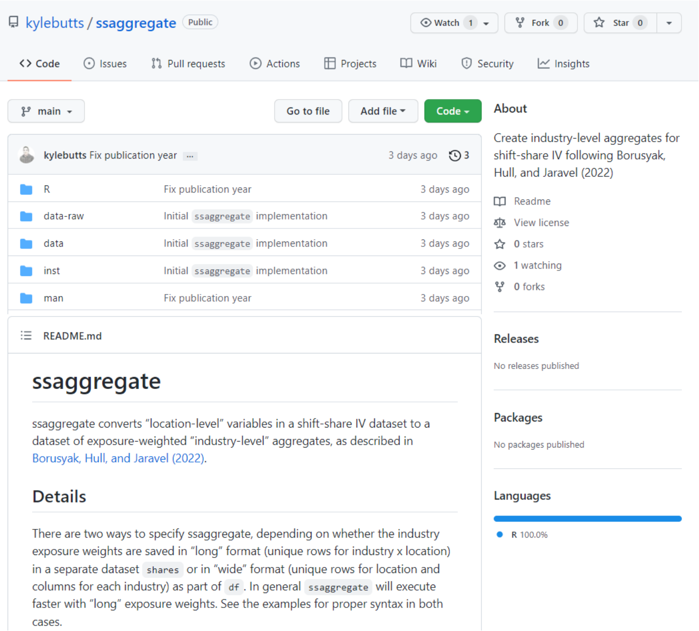
They propose tools to test for shock exogeneity (e.g. balance/ pre-trend checks) and quantify the extent of identifying variation

- 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

Estimating Exogenous-Shock SSIV Regressions in Stata



Estimating Exogenous-Shock | SSIV Regressions in R



For More on SSIV and Related Methods ...

Shift-Share IV

Mixtape Track

Shift-Share Instrumental Variables (SSIV) are used to address endogeneity and selection challenges in many economic settings. This half-day workshop will introduce the basics of SSIV and cover the recent literature on its econometric foundations. Special focus will be paid on the different assumptions underlying the "exogenous shares" and "exogenous shocks" approaches to SSIV identification, and their practical implications. We will also cover a more general class of instrumental variable strategies combining exogenous shocks and non-random exposure. Group programming exercises will be used to illustrate various theoretical concepts in real-world applications.

Learn more about the course

Instructor

Prof. Peter Hull

Dates

Starting May 21st

Your Own Workshop

Get in touch!

We can host a custom workshop for your company in person!

Roadmap

Judge/Examiner IV

- Approach
- Cautions

Shift-Share IV

- Approach

Other Frontiers

- Diff-in-Diff IV
- Recentered IV

Diff-in-Diff IV

Remember panel data IVs? We haven't talked about them in a heterogeneous-effects setup but Hudson et al. (2017) do just that

- Intuitively, a LATE interpretation requires parallel trends in both the outcome and the treatment and a subtle exclusion restriction: the IV can only affect outcomes in one period
- This note actually grew out of my Abdulkadiroglu et al. (2016) work

Diff-in-Diff IV

Remember panel data IVs? We haven't talked about them in a heterogeneous-effects setup but Hudson et al. (2017) do just that

- Intuitively, a LATE interpretation requires parallel trends in both the outcome and the treatment and a subtle exclusion restriction: the IV can only affect outcomes in one period
- This note actually grew out of my Abdulkadiroglu et al. (2016) work

De Chaisemartin and D'Haultfoeuille propose an alternative "fuzzy difference-in-differences" approach which makes other assumptions

- Key question is whether you think the RF and FS diff-in-diffs are causal or not (if so, keep calm and `ivreg2` on!)

The Recent Diff-in-Diff Literature

You may have noticed there's been, uh, a lot going on with DD recently

- Goodman-Bacon, Sun and Abraham, Callaway and Sant'Anna, Borusyak/Jaravel/Spiess, de Chaisemartin and D'Haultfoeuille ...
- As far as I can tell most/all of this analysis is about "reduced form" difference-in-differences

The Recent Diff-in-Diff Literature

You may have noticed there's been, uh, a lot going on with DD recently

- Goodman-Bacon, Sun and Abraham, Callaway and Sant'Anna, Borusyak/Jaravel/Spiess, de Chaisemartin and D'Haultfoeuille ...
- As far as I can tell most/all of this analysis is about "reduced form" difference-in-differences

My guess is these problems only get worse with IV (work to be done!)

- But presumably if you can use any of these approaches to estimate the RF & FS, LATE goes through á la Hudson et al. (2017)
- I don't really have anything smarter to say about that for now...

Recentered IV

Often we're interested in using instruments that combine multiple sources of variation, only some of which is random

- 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

Often we're interested in using instruments that combine multiple sources of variation, only some of which is random

- 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

- 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

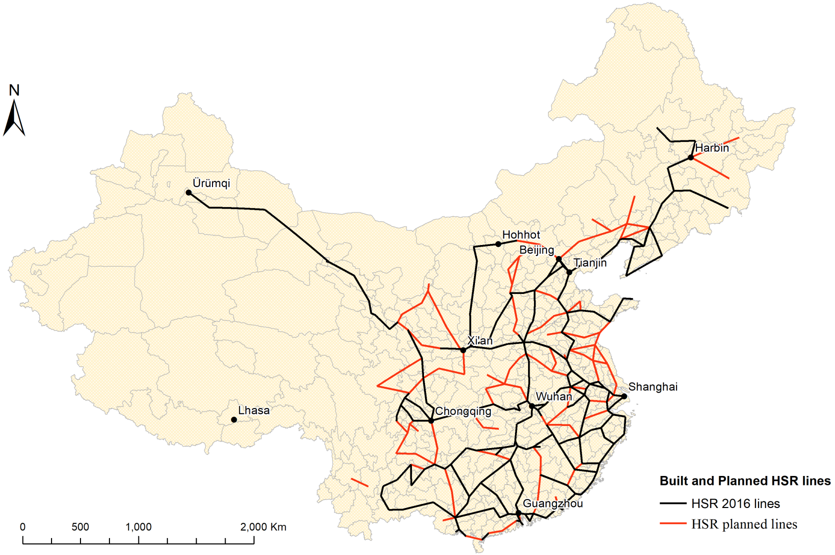
- BH show that the *expected instrument* $\mu_i = E[f_i(g; s) \mid s]$ is the sole source of bias and the *recentered instrument* $Z_i - \mu_i$ is free of bias

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

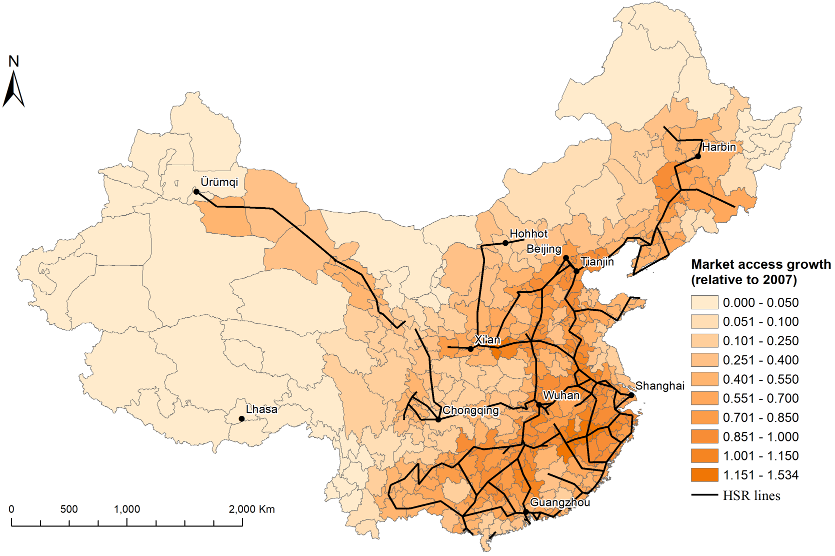
1. Specify *counterfactual* shocks $\tilde{g}^{(1)}, \ldots, \tilde{g}^{(K)}$ which were as likely to have occured (by, e.g., permuting the rows of g)
2. Recompute $Z_i^{(1)}, \ldots, Z_i^{(K)}$ for each observation i: $Z_i^{(k)} = f_i(\tilde{g}^{(k)}; s)$
3. Average the counterfactual instruments for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

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

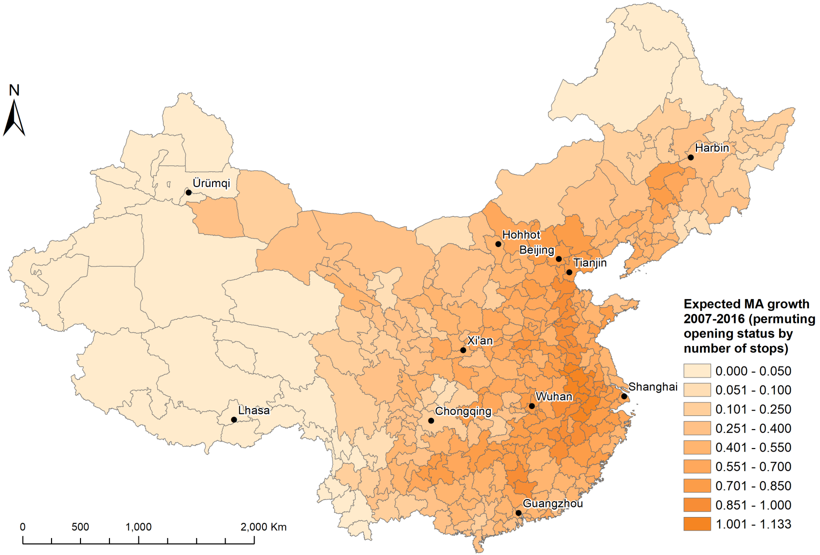
Illustration: High-Speed Rail in China, 2007-2016



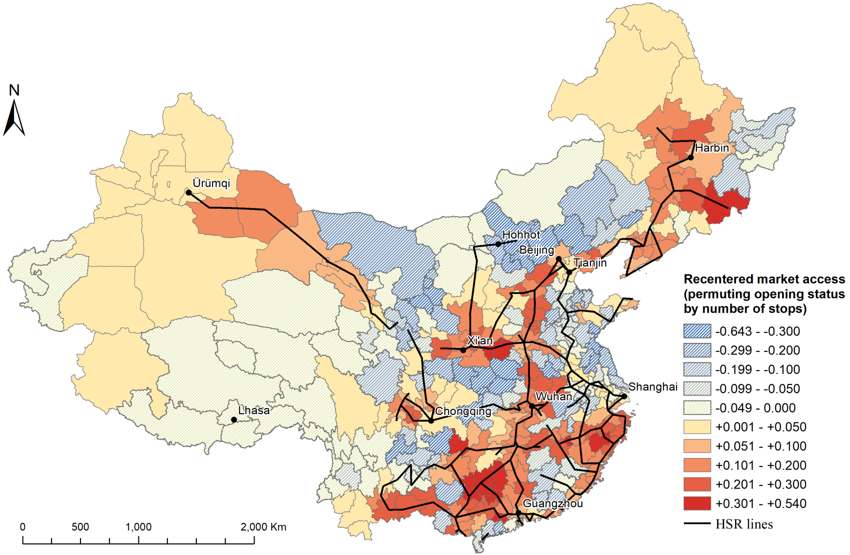
Market Access Growth, Computed from Rail Growth



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)