Recent Advances in Shift-Share IV

Peter Hull
U Chicago and NBER

November 2020

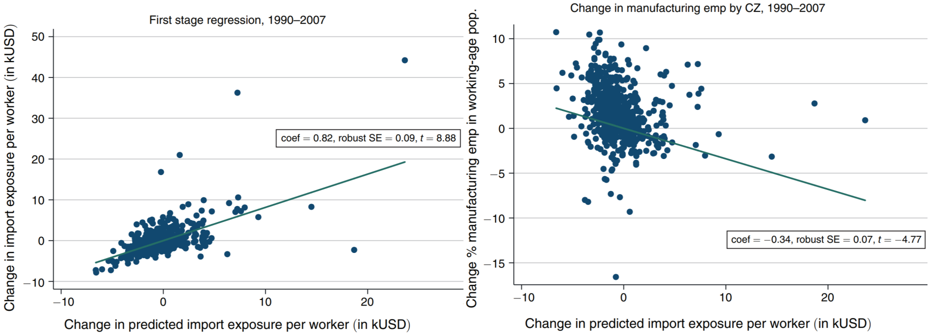
Introduction

- Many canonical instrumental variables (IVs) leverage the quasi-random assignment of some z_ℓ across observations ℓ
  - Angrist (1990): randomly assigned draft lottery number z_ℓ as an instrument for individual ℓ's service in the Vietnam War
- But some z_ℓ are more complicated, combining variation across both observations and some other common dimension n
  - Bartik (1991): predicted employment growth $z_\ell = \sum_n s_{\ell n} g_n$ as an IV for region ℓ's employment growth, where g_n is the national growth of industry n and the $s_{\ell n} \in [0, 1]$ are lagged employment shares
  - Similar z_ℓ: Blanchard & Katz (1992), Card (2009), Autor et al. (2013)
- A recent methodological literature studies when/how such "shift-share" IVs (SSIVs) can be used for causal inference
  - Formalizes two paths to identification: via "shocks" g_n or "shares" $s_{\ell n}$
  - Raises new practical considerations for SSIV estimation and inference

Autor, Dorn, & Hanson (ADH; 2013): "The China Shock"

- ADH study the effects of rising Chinese import competition on US local labor markets, 1990-2007
  - Share of US spending on Chinese goods: 0.6%→4.6%
  - Share of working-age pop employed in manufacturing: 12.6%→8.4%
  - Reverse causality concern: weak markets more likely to import
- To address endogeneity challenge, they use a SSIV $z_\ell = \sum_n s_{\ell n} g_n$
  - g_n: industry n's growth of Chinese imports in eight non-U.S. economies
  - $s_{\ell n}$: lagged share of mfg. industry n in total employment of location ℓ
  - Treatment x_ℓ: local growth of Chinese imports ($1,000/worker)
  - Main outcome y_ℓ: local change in manufacturing employment share
- ADH derive this instrument from a simple trade model:
  - *"Our IV strategy will identify the Chinese productivity and trade-shock component of United States import growth if the common within-industry component of rising Chinese imports to the United States and other high-income countries stems from China's rising comparative advantage and (or) falling trade costs in these sectors."* (p. 2129)

ADH First Stage and Reduced Form



- Main IV estimate (state-clustered SE), RF/FS: -0.596 (0.099)
- If causal (and setting aside general equilibrium effects, etc.), would explain 33% of the fall in manufacturing employment
- How might one assess causality with this IV?
    - Hard to think of changes in predicted import exposure like a lottery #...

Card (2009) Immigration "Enclave" IV

- Card studies the effect of local immigration on local wages
  - Outcome $y_{\ell j}$: log wage gap between immigrant and native men in skill group j and region ℓ
  - Treatment $x_{\ell j}$: log ratio of immigrant to native hours in (ℓ, j)
  - Seek to estimate immigrant-native inverse elasticity of substitution

- He constructs a SSIV $z_{\ell j}$ by combining lagged shares $s_{\ell n}$ of immigrants from countries n in region ℓ & national immigration rates g_{jn}
  - Intended to address endogeneity from local labor demand shocks
  - *"To the extent that initial immigrant shares are correlated with other unobserved features that affect relative wage differentials in a city, an enclave-based identification strategy may be less attractive..."* (p. 15)
  - Again, hard to think of predicted immigration inflows like a lottery #

Card Reduced Form

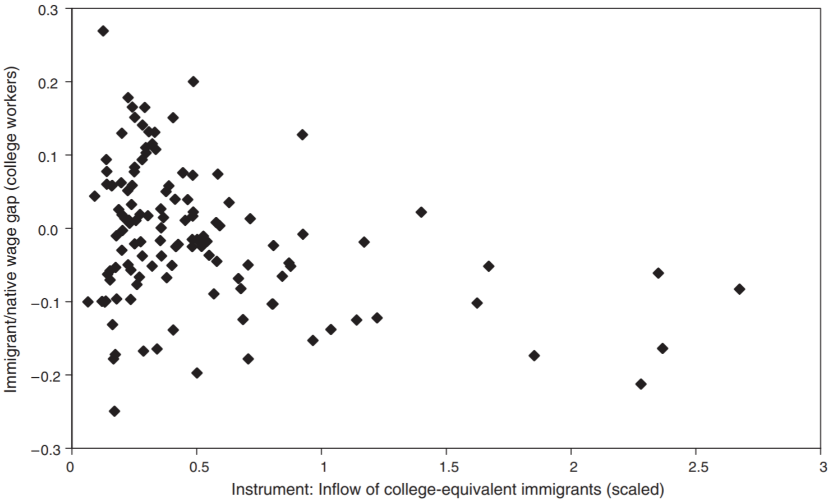


FIGURE 4. REDUCED FORM: INFLOW IV AND IMMIGRANT WAGE GAP (*College*)

Outline

1. The SSIV Setting
2. Identification from Shares: Goldsmith-Pinkham et al. (2020)
3. Identification from Shocks: Borusyak et al. (2020)
4. New Inference Challenges: Adão et al. (2019)
5. Further Settings: Borusyak and Hull (2020)

The SSIV Setting

- Suppose we are interested in estimating some parameter β of a linear causal or structural model $y_\ell = \beta x_\ell + e_\ell$
  - Straightforward to generalize to heterogeneous treatment effects

- Residualize e_ℓ on a vector of observed controls w_ℓ to get second stage:

$$y_\ell = \beta x_\ell + w_\ell' \gamma + \varepsilon_\ell,$$

  where w_ℓ and ε_ℓ are orthogonal by construction: $E\left[\sum_\ell w_\ell \varepsilon_\ell\right] = 0$

- We instrument x_ℓ with $z_\ell = \sum_n s_{\ell n} g_n$, where $\sum_n s_{\ell n} = 1$ (for now)
  - Call $s_{\ell n}$ the "exposure shares" and g_n the "shocks"
  - Share vary across observations, shocks do not

- IV is valid if $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0$; identification follows from a first stage
  - Note no *iid* assumption, $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] \neq E[z_\ell \varepsilon_\ell]$; will be important later

The SSIV Estimator

- SSIV divides the regression of y_ℓ on z_ℓ, controlling for w_ℓ, ("reduced form") by the regression of x_ℓ on z_ℓ, controlling for w_ℓ ("first stage")

- By the Frisch-Waugh-Lovell theorem, this 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

- Plugging in the model $y_\ell = \beta x_\ell + w_\ell' \gamma + \varepsilon_\ell$ gives

$$\hat{\beta} = \beta + \frac{\sum_\ell \sum_n s_{\ell n} g_n \varepsilon_\ell^\perp}{\sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp},$$

- Consistency: $\hat{\beta} \xrightarrow{p} \beta$ if $\frac{1}{L} \sum_\ell \sum_n s_{\ell n} g_n \varepsilon_\ell^\perp \xrightarrow{p} 0$, $\frac{1}{L} \sum_\ell \sum_n s_{\ell n} g_n x_\ell^\perp \xrightarrow{p} \pi \neq 0$
- Asymptotic inference: find a σ_L such that $(\hat{\beta} - \beta)/\sigma_L \Rightarrow N(0,1)$

Outline

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

- GPSS are interested in understanding when/how SSIV can be seen as leveraging quasi-experimental variation across observations
  - Viewing the g_n as fixed, $z_\ell = \sum_n s_{\ell n} g_n$ is a linear combination of shares
  - It follows that z_ℓ is a valid instrument when the shares are exogenous
- Formally, GPSS 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 condition for identification: quasi-experimental shock exposure across observations

$$E[\varepsilon_\ell \mid s_{\ell n}] = 0,\ \forall n \implies E\left[\frac{1}{L}\sum\nolimits_\ell z_\ell \varepsilon_\ell\right] = \frac{1}{L}\sum_\ell \sum_n g_n E\left[s_{\ell n} E[\varepsilon_\ell \mid s_{\ell n}]\right] = 0$$

- Diff-in-diff logic: when ε_ℓ are unobserved outcome *trends* (as in ADH) $E[\varepsilon_\ell \mid s_{\ell n}] = 0$ is akin to a "parallel trends" assumption
  - Consistency/inference follow from standard conditions (e.g. *iid* data)

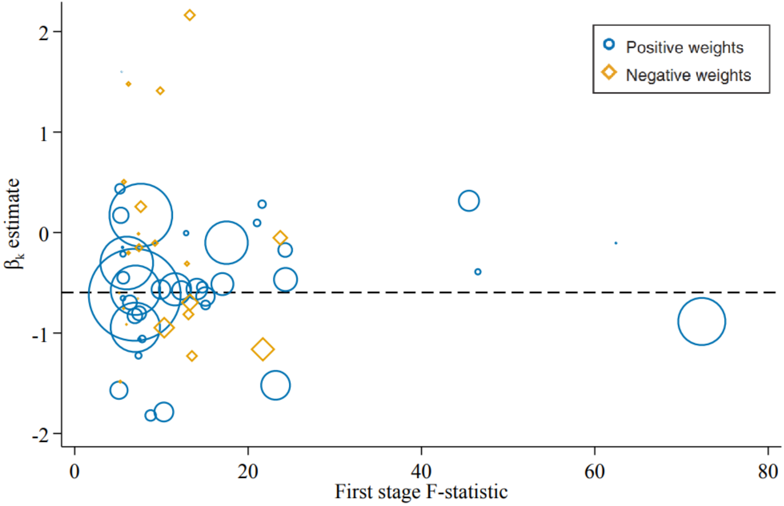
Rotemberg Weights

- The GPSS view of SSIV is one of many share instruments
  - In Card, $N = 38$. In ADH, $N = 397$ (!)
- They 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"

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

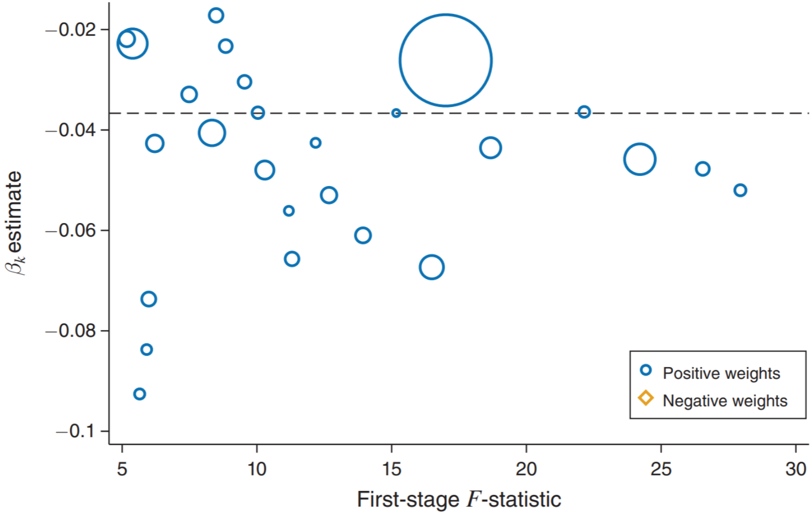
- Intuitively, more weight is given to share instruments with higher shocks g_n and larger first stages $\sum_\ell s_{\ell n} x_\ell^\perp$
  - Weights can be negative (potential problem with heterogeneous effects)
  - Under constant effects, show a formal link to Conley et al. (2012)'s and Andrews et al. (2017)'s measures of sensitivity-to-misspecification

Rotemberg Weights in ADH (via GPSS)



- Negative weights, large heterogeneity in individual $\hat{\beta}_n$ estimates

Rotemberg Weights in Card (via GPSS)



- No negative weights, low heterogeneity in individual $\hat{\beta}_n$ estimates

Is Share Exogeneity a Plausible Identifying Assumption?

- Several ways to probe the plausibility of exogenous $s_{\ell n}$ *ex post*:
  - Balance/pre-trend tests, overidentification tests (w/constant effects)
  - Straightforward to implement; no different than any other IV
  - GPSS find these tests broadly pass for Card, but fail badly for ADH
- In some settings, share exogeneity is *ex ante* implausible
  - Suppose unobserved shocks ν_n affect ε_ℓ via the same/correlated shares
  - E.g. in ADH, unobserved technology shocks across industries n can affect labor markets via employment shares, along with observed g_n
  - Then share exogeneity cannot hold: the shares drive outcomes through both observed and unobserved channels
  - Formally, 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 they are randomly assigned to observations
- Likely share endogeneity calls for a new approach to SSIV...

Outline

Borusyak, Hull, and Jaravel (BHJ; 2020)

- BHJ are interested in understanding when/how SSIV can be seen as leveraging quasi-random variation in the shocks
- Like GPSS, they establish a *numerical equivalence*:
  - $\hat{\beta}$ can be obtained from a just-identified shock-level IV procedure that uses g_n to instrument for a shock-level "aggregate" of the treatment:

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

    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
  - It follows that $\hat{\beta}$ is consistent iff this shock-level IV procedure is
- They then derive new conditions for SSIV identification + consistency
  - Want to view g_n as random shocks, so can't assume $z_\ell = \sum_n s_{\ell n} g_n$ is *iid*

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 validity: $E\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = E[\sum_n s_n g_n \bar{\varepsilon}_n] = 0$
- **A2** (Many uncorrelated shocks): $E[\sum_n s_n^2] \to 0$ and $\forall (n, n')$ with $n' \neq n$, $Cov(g_n, g_{n'} \mid \bar{\varepsilon}, s) = 0$
  - First part: expected Herfindahl index of average shock exposure converges to zero (implies $N \to \infty$)
  - Second part: shocks are mutually uncorrelated given the unobservables
  - Imply a shock-level law of large numbers: $\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell = \sum_n s_n g_n \bar{\varepsilon}_n \xrightarrow{p} 0$
- Both assumptions, while novel for SSIV, would be standard for a shock-level IV regression with weights s_n and instrument g_n
- Identification of β follows given a first stage: $\frac{1}{L}\sum_\ell z_\ell x_\ell^\perp \xrightarrow{p} \pi \neq 0$
  - Sufficient condition: most observations are mostly exposed to a small number of shocks affecting treatment

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
- **Panel Data**: Have $(y_{\ell t}, x_{\ell t}, s_{\ell n t}, g_{nt})$ across $\ell = 1, \ldots, L$, $t = 1, \ldots, T$
  - 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
  - Also see Jaeger et al. (2019) for dynamic biases in panel SSIVs
- **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: $z_\ell = \sum_\ell w_{\ell n} \tilde{g}_{\ell n}$ for $\tilde{g}_{\ell n} = \sum_{\ell' \neq \ell} w_{\ell' n} g_{\ell' n}$ (akin to JIVE solution to many-IV bias)
- **Multiple shocks**: Propose new overidentified SSIV procedures

The "Incomplete Shares" Issue

- 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 observations
  - E.g. in ADH, S_ℓ is region ℓ's share of non-manufacturing employment since $s_{\ell n}$ is the share of manufacturing industry n in *total* employment
- BHJ show that A1/A2 are not enough for identification in this case
  - The IV implicitly uses variation across S_ℓ, which may be endogenous
- Controlling for the sum-of-shares S_ℓ isolates clean shock variation
  - Can be seen as a special case of conditional quasi-random assignment: "dummying out" the non-manufacturing sector, in ADH
  - Further controls needed when A1 holds conditional on q_n; e.g. isolating within-period variation in panels requires interacting S_ℓ with period FE

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 itself be thought of as an instrument (i.e. many, plausibly quasi-randomly assigned)
  - E.g. Acemoglu et al. (2016) use the ADH shocks to conduct an industry-level IV analysis
  - 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
  - Canonical setting of Bartik (1991), where g_n are average industry growth rates (thought to proxy for latent demand shocks)
  - See also Card (2009), where national immiration rates are estimated
- Case 3: the g_n cannot be naturally viewed as an instrument
  - Either too few (small N) or implausibly exogenous, even given some q_n
  - 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 should 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 may 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 (e.g. Mohnen 2019) have a DD feel; don’t even need the shocks, except to possibly improve power / avoid many-IV bias

ADH Revisited

- BHJ show how ADH can be seen as leveraging quasi-random shocks
  - *Ex ante* plausible (unlike exogenous shares): imagine random industry productivity shocks in China affecting imports in U.S. & elsewhere
  - Many shocks (industries), plausibly weakly mutually correlated
- Evaluate A1 by regional and industry-level balance tests
  - Industry shocks are uncorrelated with five observables considered by Acemoglu et al. (2016) (e.g. lagged capital to value-added ratios)
- Evaluate A2 by studying variation across industries
  - Effective sample size (1/HHI of s_n weights): 58-192
  - Shocks appear mutually uncorrelated across sectors (SIC3)
- Check sensitivity to adjusting for potential industry-level confounders
  - Control for $w_\ell = \sum_n s_{\ell n} q_n$, where q_n include the Acemoglu et al. (2016) observables, sector FE, industry pre-trends ...

BHJ do ADH

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

- Robust coefficient of ≈ -0.3, after accounting for incomplete shares

Outline

1. The SSIV Setting
2. Identification from Shares: Goldsmith-Pinkham et al. (2020)
3. Identification from Shocks: Borusyak et al. (2020)
4. New Inference Challenges: Adão et al. (2019)
5. Further Settings: Borusyak and Hull (2020)

Adão, Kolesar, and Morales (AKM, 2019)

- AKM study a novel inference challenge when SSIV identification leverages quasi-random shocks
  - Observations with similar shares $s_{\ell 1,}, \ldots, s_{\ell N}$ 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. $\varepsilon_\ell = \sum_n s_{\ell n} \nu_n + \tilde{\varepsilon}_\ell$), the large-sample distribution of $\hat{\beta}$ may not be well-approximated by standard CLTs
- They show by simulation that this can lead to large size distortions
  - Tests with nominal 5% rejection rates can reject true nulls in 55% of placebo shock realizations (ADH-based Monte Carlo)
  - Reminiscent of Bertrand et al. (2004) study of robust SEs in diff-in-diff
- They then derive a new CLT + SEs to address "exposure clustering"
  - "Design-based:" leverage *iid*ness of shocks, not observations, building on BHJ identification framework
  - Also develop null-imposed (AKM0) CIs, which help in finite samples

ADH Monte Carlos (Robust/Clustered)

Table 1: Standard errors and rejection rate of the hypothesis $H_0: \beta = 0$ at 5% significance level

	Estimate		Median std. error		Rejection rate	
	Mean (1)	Std. dev. (2)	Robust (3)	Cluster (4)	Robust (5)	Cluster (6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	0.73	0.92	48.5%	38.1%
Employed in manufacturing	−0.01	1.88	0.60	0.76	55.7%	44.8%
Employed in non-manufacturing	0.00	0.94	0.58	0.67	23.2%	17.6%
Panel B: Change in average log weekly wage						
Employed	−0.03	2.66	1.01	1.33	47.3%	34.2%
Employed in manufacturing	−0.03	2.92	1.68	2.11	26.7%	16.8%
Employed in non-manufacturing	−0.02	2.64	1.05	1.33	45.4%	33.7%

ADH Monte Carlos (AKM/AKM0)

Table 2: Median standard errors and rejection rates for $H_0: \beta = 0$ at 5% significance level

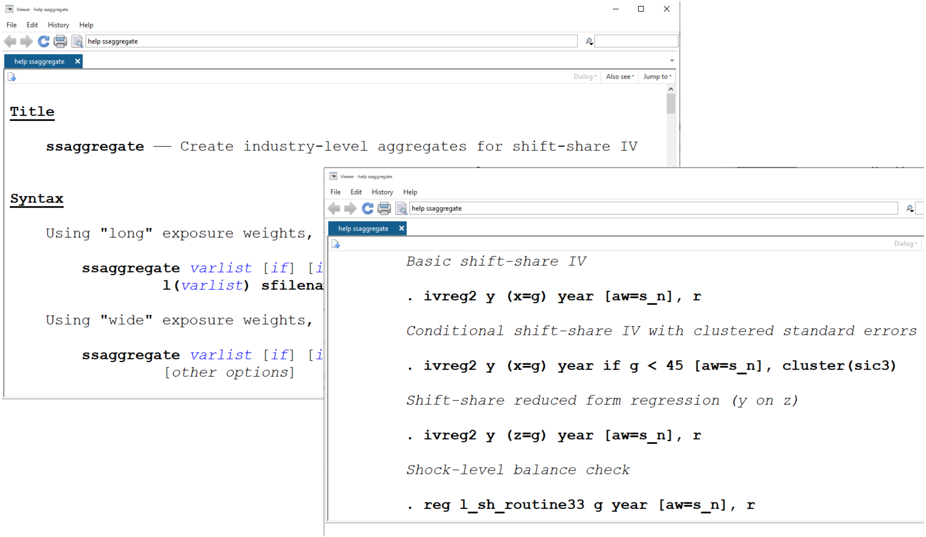
	Estimate		Median eff. s.e.		Rejection rate	
	Mean	Std. dev	AKM	AKM0	AKM	AKM0
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	1.90	2.21	7.8%	4.5%
Employed in manufacturing	−0.01	1.88	1.77	2.06	8.0%	4.3%
Employed in non-manufacturing	0.00	0.94	0.89	1.04	8.2%	4.5%
Panel B: Change in average log weekly wage						
Employed	−0.03	2.66	2.57	2.99	7.5%	4.3%
Employed in manufacturing	−0.03	2.92	2.74	3.18	9.1%	4.5%
Employed in non-manufacturing	−0.02	2.64	2.55	2.96	7.8%	4.5%

Exposure-Robust SEs

- BHJ show a convenient solution to exposure clustering, via their equivalent shock-level IV regression
- 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 $\sum_n s_{\ell n} q_n$
  - Null-imposed CIs similarly straightforward at the shock level
- Clustering logic: can get valid SEs by 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
  - New Stata package *ssaggregate* helps translate data to the shock level, after which researchers can proceed with familiar estimation commands



Install with *ssc install ssaggregate*; please send us comments to improve!

Outline

1. The SSIV Setting
2. Identification from Shares: Goldsmith-Pinkham et al. (2020)
3. Identification from Shocks: Borusyak et al. (2020)
4. New Inference Challenges: Adão et al. (2019)
5. Further Settings: Borusyak and Hull (2020)

Borusyak and Hull (BH; 2020)

- Many instruments may be seen as being SSIV-like, combining a set of exogenous shocks and measures of non-random shock exposure
  - **Nonlinear SSIVs**: $z_\ell = f(g_1, \ldots, g_N, s_{\ell 1}, \ldots, s_{\ell N})$ for nonlinear $f(\cdot)$
  - **Network treatments/instruments**: z_ℓ combines shocks to other nodes with observed network linkages
  - **Transportation instruments**: z_ℓ combines transportation infrastructure upgrades with geography and nearby market sizes
  - **Simulated eligibility instruments**: z_ℓ combines variation in state policies with individual demographics / income / etc.

- BH develop a general framework for such settings, building on BHJ
  - **Identification** generally requires an adjustment for non-random exposure, akin to the adjustment for "incomplete shares" in linear SSIV
  - **Inference** leverages "design" of the shocks to account for non-random "exposure clustering" (randomization inference)

- BH illustrate this framework by addressing bias in "market access" regressions & boosting power in a simulated instrument setting

Summary

- We've learned a lot about shift-share IV over the past few years
  - Identification can come from exogenous shock exposure (akin to a DD)
  - But as-good-as-random shocks may be a more plausible identifying assumption; then consistency/inference is non-standard
  - Many new tools to solve practical issues in either case
- General advice for researchers hoping to use a SSIV:
  - Decide in advance whether exogenous shares or exogenous shocks is a plausible assumption (BHJ taxonomy may be a helpful guide)
  - Apply appropriate tests to probe your *a priori* claims (i.e. GPSS / BHJ)
  - If exogenous shocks, address exposure clustering and be careful with the "incomplete shares" issue, especially in panels