ECON 3310 - Topics in Urban Economics & Migration

Shift-Share Instrumental Variables

Andy Ferrara
University of Pittsburgh

Fall 2022

The history of shift-share IVs

Bartel (1989) and Lalonde and Topel (1991) noticed that immigrants

- systematically sort into arrival locations
- based on whether there are pre-existing clusters of people from their home countries

$\rightarrow$ this is called **chain migration**

Altonji and Card (1991) first exploited this as an instrument

- the **past-settlement** instrument
- what's the problem?

The history of shift-share IVs

Bartel (1989) and Lalonde and Topel (1991) noticed that immigrants

- systematically sort into arrival locations
- based on whether there are pre-existing clusters of people from their home countries

$\rightarrow$ this is called **chain migration**

Altonji and Card (1991) first exploited this as an instrument

- the **past-settlement** instrument
- what's the problem?

Card (2001) refined this instrument by combining it with the aggregate flows of immigrants from each country.

- shift-share IVs originally attributed to Bartik (1991), hence sometimes referred to as **Bartik** IV

Shift-share IVs have become very popular

- China shock and U.S. employment (Autor, Dorn, and Hanson, 2016)
- global trade (Xu, 2019)
- foreign aid and conflict (Nunn and Qian, 2014)
- immigration (Card, 2001)
- local public spending (Nakamura and Steinsson, 2012)
- historical settings (Fouka, Mazumder, Tabellini, 2022)
- portfolio allocation (Calvet, Campbell, and Sodini, 2009)
- judge leniency (Kling, 2006)
- work automation (Acemoglu and Restrepo, 2017)
- ...

Especially in the migration literature, this has become a work horse model.

Setup

Imagine you want to regress,

$$\% \text{ Republican}_{c,2016} = \alpha_s + \beta \% \text{ white Southerners}_{c,1940} + X_c' \gamma + \epsilon_c$$

where

- c are non-Southern counties in 2016
- α_s are state fixed effects
- % white Southerners$_{c,1940}$ is the share of white Southern-born individuals living in c in 1940

Setup

Imagine you want to regress,

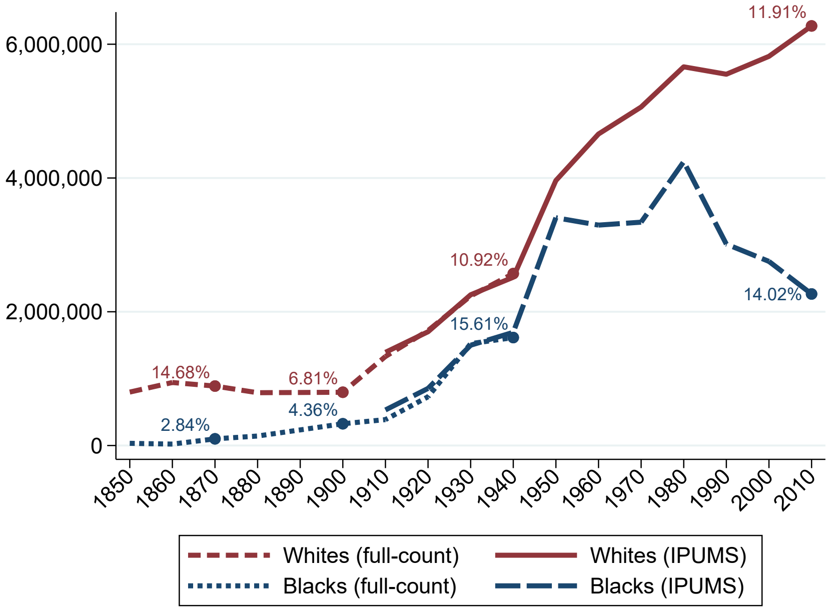
$$\% \text{ Republican}_{c,2016} = \alpha_s + \beta \% \text{ white Southerners}_{c,1940} + X_c' \gamma + \epsilon_c$$

where

- c are non-Southern counties in 2016
- α_s are state fixed effects
- $\%$ white Southerners$_{c,1940}$ is the share of white Southern-born individuals living in c in 1940

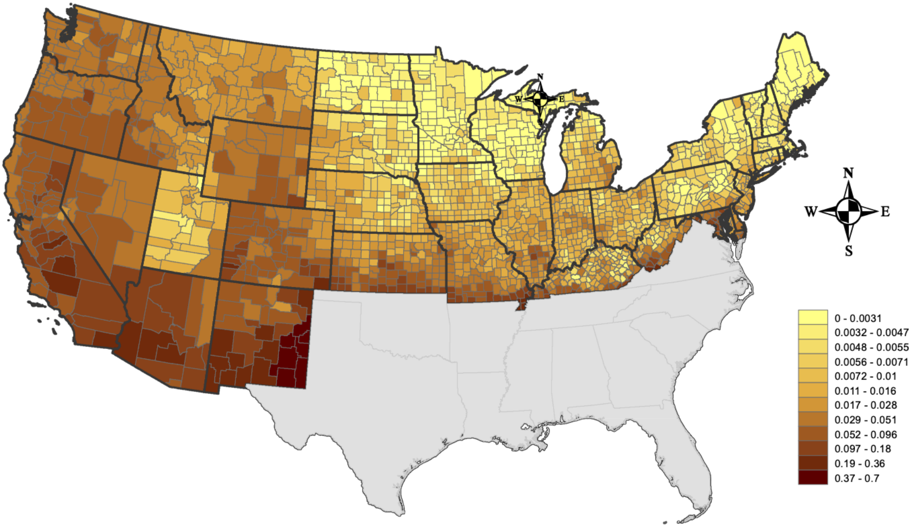
That's our setup in Bazzi et al. (2021) "The Other Great Migration: Southern Whites and the New Right", NBER working paper no. 29506

Background

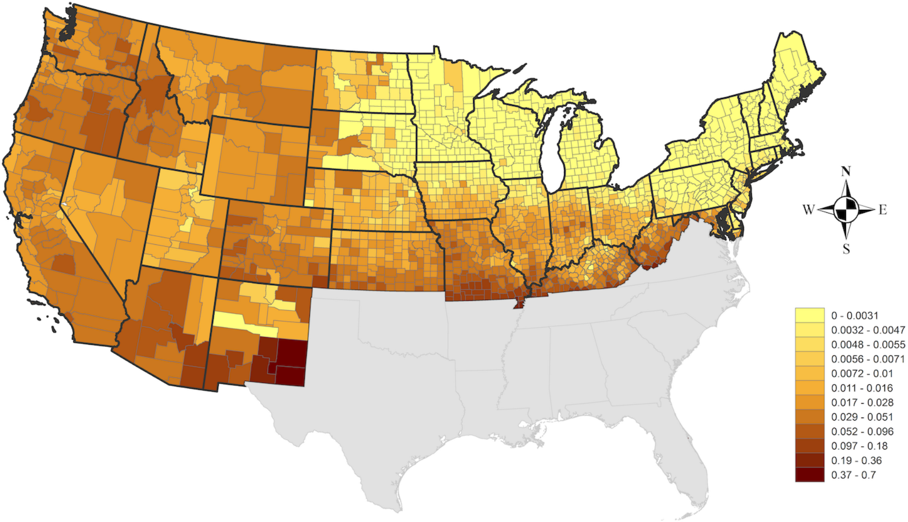


6,000,000
4,000,000
2,000,000
0
11.91%
10.92%
15.61%
14.02%
14.68%
6.81%
4.36%
2.84%
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Whites (full-count)
Whites (IPUMS)
Blacks (full-count)
Blacks (IPUMS)

Spatial distribution in 1940 (as % of pop.)



Spatial distribution in 1900 (as % of pop.)



Empirical issues

The problem: white Southerners do not sort randomly in space.

We could use % white Southerners$_{c,1900}$ as past settlement instrument for the 1940 share. But does this solve the problem?

Empirical issues

The problem: white Southerners do not sort randomly in space.

We could use % white Southerners$_{c,1900}$ as past settlement instrument for the 1940 share. But does this solve the problem?

No. If white Southerners went to places that were already more "conservative"

- correlated chain migrations
- correlate with future political outcomes

So the idea is to introduce another more exogenous element to the instrument, i.e. the shift component.

Shift-share setup

Predict the share of white Southerners in 1940 based on,

$$Z_{c,1940} = \sum_{j=1}^{J} \pi_{c,1900} \Delta M_{j,1900-1940}$$

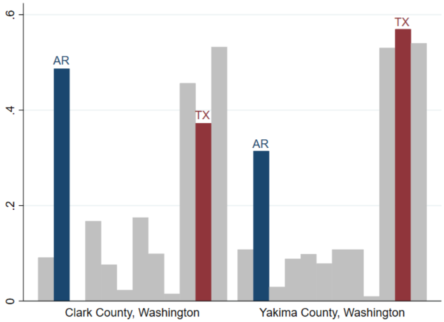
where

- j indexes Southern sending states
- $\pi_{c,1900}$ is the baseline share of white Southerners from state j
    - why did we choose 1900?
- $M_{j,1900-1940}$ is the change in the total number of white Southerners living outside the South from 1900 to 1940

then divide the sum by 1940 population in c and use this as instrument for the observed % white Southerners$_{c,1940}$.

Intuition

Where is the variation coming from?

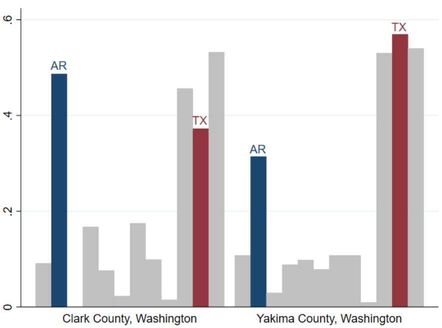


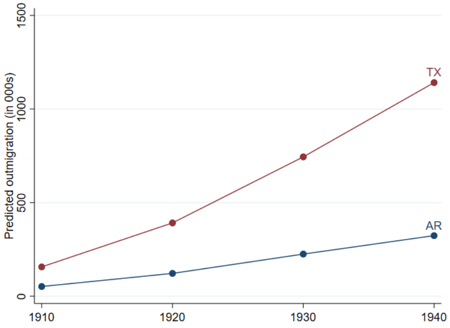
Consider Clark and Yakima County in WA

- same share of white Southerners in 1900 (2.5% and 2.6%, resp.)
- but a different "mix" of white Southerners (AR vs TX)

Intuition

Where is the variation coming from?





Consider Clark and Yakima County in WA

- same share of white Southerners in 1900 (2.5% and 2.6%, resp.)
- but a different "mix" of white Southerners (AR vs TX)

Since relatively more ppl migrated from TX than AR betw. 1900-40,

- Yakima's share incr. to 4.8% in 1940 (3.8% in Clark)

What's the gain of having the shift?

The big concern with the past settlement IV

- unobserved local characteristics that drive
  1. location choice of migrants
  2. changes in the outcome

The shift component is useful because

- aggregate flows of migrants arguably unrelated to conditions in c
- additional variation in the instrument (more on this soon!)

What's the gain of having the shift?

The big concern with the past settlement IV

- unobserved local characteristics that drive
  1. location choice of migrants
  2. changes in the outcome

The shift component is useful because

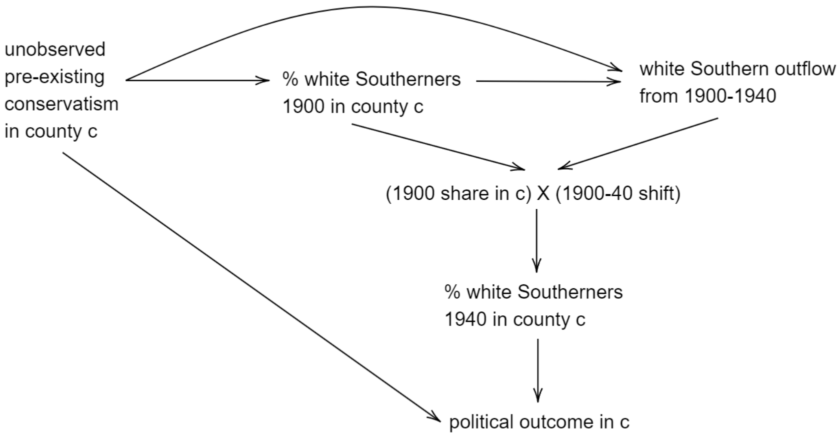
- aggregate flows of migrants arguably unrelated to conditions in c
- additional variation in the instrument (more on this soon!)

However, it's harder to think about the exclusion restriction $Cov(Z_c, \epsilon_c)$,

- the assumption now is that shift **and** share must be endogenous

SSIV have an extremely complicated **exclusion restriction**. Have you done your homework?

A possible exclusion restriction violation



What the heck!

Exclusion restriction difficulties

Since thinking about the exclusion restriction is hard, referees oftentimes get this wrong

- "I don't believe that the initial shares are exogenous"
- Yes, duh. That's why we use a shift-share design (not the answer you want to give in your response letter)

Exclusion restriction difficulties

Since thinking about the exclusion restriction is hard, referees oftentimes get this wrong

- "I don't believe that the initial shares are exogenous"
- Yes, duh. That's why we use a shift-share design (not the answer you want to give in your response letter)

In your response letter, you say

- thank you for this excellent point
- we have done X, Y, and Z to address it
- in the process of doing so, we found paper P which shows that the required assumptions for SSIV are actually on the shift **and** the share
- following their approach, we get the same result but thanks again for your amazing comment

The importance of controls

Convincing these kinds of referees is important

- their point is not entirely moot in our setting
- we don't have county FE α_c because X-sectional setting (more on this in a bit)

Remember that the excl. restriction is *conditional* on covariates. Incl. baseline controls in 1900

- enlistment and mortality rate of Union Army soldiers
- Breckindridge vote share in 1860
- soil suitability for cotton
- other resources for extractive industries (mining, oil)
- log population, population density, etc.

Based on contextual and theoretical considerations.

The importance of robustness checks

Controls rarely convince anyone. We also

- have an alternative identification strategy (the railroad IV)
- use counties w pop density <2 in 1860 that then receive white Southerners in 1870
- diff. definitions of treatment/instrument timing
- match counties based on 1900 characteristics and then have a pairwise comparison of
  - counties w similar observables in 1900
  - counties w same % Republican share in 1900
- check that this is true for counties w same 1870 white Southerner shares

Lots of hoops to jump through to defend a hard to define exclusion restriction.

Quick caveat: interpretation

Not only the exclusion restriction is hard to think about but also *what do we estimate*?

Typical LATE argument hard to articulate,

- effect of change in treatment status of compliers
- i.e. those who changed treatment status because of the instrument

That's hard to motivate with

- multivalued instrument
- and, making it even harder, multivalued treatment

Quick caveat: multivalued instruments

Assume we have a binary treatment D, multivalued valid instrument Z with finite $\mathcal{S}(Z) = 0, 1, ..., J$. Further, assume

- multivariate first stage:
  $P(D = 1|Z = j) > P(D = 1|Z = j - 1), \forall j \geq 1$

Quick caveat: multivalued instruments

Assume we have a binary treatment D, multivalued valid instrument Z with finite $\mathcal{S}(Z) = 0, 1, ..., J$. Further, assume

- multivariate first stage:
  $P(D = 1|Z = j) > P(D = 1|Z = j - 1), \forall j \geq 1$
- multivariate monotonicity:
  for $0 \leq j \leq j' \leq J$, we have that $D(j') \geq D(j)$

Quick caveat: multivalued instruments

Assume we have a binary treatment D, multivalued valid instrument Z with finite $\mathcal{S}(Z) = 0, 1, ..., J$. Further, assume

- multivariate first stage:
  $P(D = 1|Z = j) > P(D = 1|Z = j - 1), \forall j \geq 1$
- multivariate monotonicity:
  for $0 \leq j \leq j' \leq J$, we have that $D(j') \geq D(j)$
- random instrument:
  $\left[(Y(j, d))_{0 \leq j \leq J, d \in \{0,1\}}, (D(j))_{0 \leq j \leq J}\right] \perp\!\!\!\perp Z$
- exclusion restriction:
  for every $0 \leq j \neq j' \leq J$ and $d \in \{0, 1\}$, we have that $Y(j, d) = Y(j', d) = Y(d)$

where d is the treatment status under a given potential outcome.

Quick caveat: multivalued instruments

For two instrument values $j < j'$, the typical 2SLS estimand here is

$$\text{wald}_{j,k} = \frac{E(Y|Z=j') - E(Y|Z=j)}{P(D=1|Z=j') - P(D=1|Z=j)}$$

i.e. reduced form divided by the first stage coefficient.

Quick caveat: multivalued instruments

For two instrument values $j < j'$, the typical 2SLS estimand here is

$$\text{wald}_{j,k} = \frac{E(Y|Z = j') - E(Y|Z = j)}{P(D = 1|Z = j') - P(D = 1|Z = j)}$$

i.e. reduced form divided by the first stage coefficient.

However, with instrument values $j = (1, ..., J)$, the 2SLS estimand becomes

$$\sum_{j=1}^{J} \text{wald}_{j,-1}\mu_j$$

Quick caveat: multivalued instruments

For two instrument values $j < j'$, the typical 2SLS estimand here is

$$\text{wald}_{j,k} = \frac{E(Y|Z=j') - E(Y|Z=j)}{P(D=1|Z=j') - P(D=1|Z=j)}$$

i.e. reduced form divided by the first stage coefficient.

However, with instrument values $j = (1, ..., J)$, the 2SLS estimand becomes

$$\sum_{j=1}^{J} \text{wald}_{j,-1} \mu_j$$

where the weight μ_j (where $\sum_j \mu_j = 1$) is

$$\mu_j = \frac{P(D=1|Z=j) - P(D=1|Z=j-1)}{\sum_{j=1}^{J} P(D=1|Z=j) - P(D=1|Z=j-1)}$$

Weights

These weights are called **Rotemberg weights** in the shift-share context.

Recent advances in the SSIV literature focus on these weights.

Goldsmith-Pinkham, Sorkin, and Swift (2020, AER)

- SSIV is a weighted series of exposure designs
- i.e. shares measure diff. exposure to shocks
- shares required to be exogenous, shocks can be endogenous
- weights can be **negative** $\rightarrow$ under TE heterogeneity, no LATE-like interpretation

They argue that SSIV is equivalent to using the shares as instrument, hence these must be exogenous.

SSIV in the Goldsmith-Pinkham et al world

They think of it in diff-in-diff analogies

- shares measure exposure to a policy shock
- shifts are the size of the policy change (policy maker's choice)

SSIV in the Goldsmith-Pinkham et al world

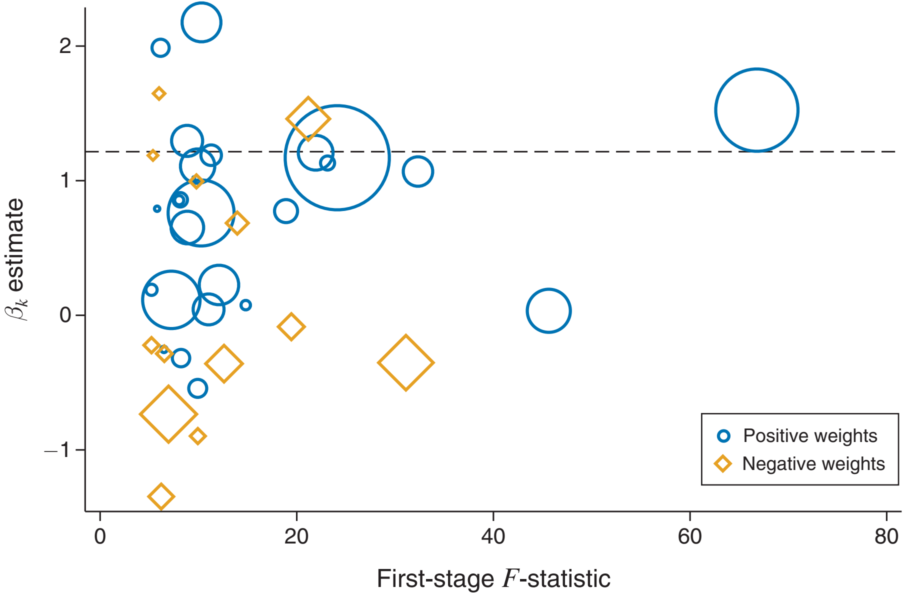
They think of it in diff-in-diff analogies

- shares measure exposure to a policy shock
- shifts are the size of the policy change (policy maker's choice)

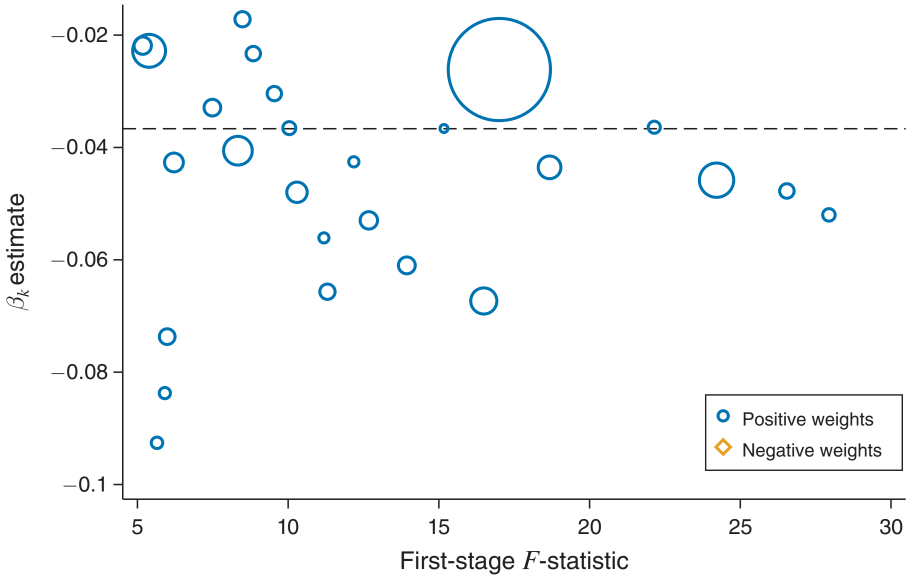
Can this exogeneity condition be tested? Yes.

- balance test: regress initial shares on pre-shock location characteristics
- pay special attention to the observations/locations/industries with the largest Rotemberg weights
  - these are the most sensitive in terms of bias
- with a time dimension, check for common trends in Z overall and for Z among obs with largest Rotemberg weights

Rotemberg weights and treatment effect heterogeneity



Rotemberg weights and treatment effect heterogeneity



Rotemberg weights and treatment effect heterogeneity

Goldsmith-Pinkham et al.

- $\widehat{\beta}$ can be identified from an overidentified IV estimation using J share instruments and a weight matrix based on the shocks g_j

Borusyak, Hull, and Jaravel (2022, REStud)

- $\widehat{\beta}$ can be identified from an just-identified IV estimation using the g_j shock-level aggregate of the treatment

This implies: shocks must be exogenous, initial shares can be endogenous.

So how do you know which of these two worlds you are in?

GPSS vs BHJ

According to GPSS, you are in the share-based identification world if,

1. research design reflects differential exposure to common shocks
2. you have very few shocks
3. you emphasize specific shocks as critical to your research design

Their exogeneity tests

- hold for Card (2001); migration from many different countries to the U.S.
- fail for Autor, Dorn, and Hanson (2013); China shock and U.S. employment (main driver is manufacturing)

GPSS vs BHJ

Being in the BHJ world is easier to deal with.

You need

- "many" uncorrelated shocks
- quasi-random shock assignment

Usually easier to find exogenous variation in the shifts than in the shares

- we link individuals across census years
- observe individuals in Southern location in $t-1$ and in non-Southern location in t
- regress migration decision on exogenous shocks (weather shocks) $\rightarrow$ many, uncorrelated shocks
- then aggregate up to state shares

SSIV dynamics

Jaeger, Ruist, and Stuhler (2018)

- standard SSIV approaches mix short- and long-run responses
- if migration is not a one-off shock, then
  - past migrations (long-run) may trigger GE effects
  - spill over into the short-run effects

Short-run $\neq$ long-run response, e.g. in the Solow model

- negative effect of immigration on wages in SR
- adjustment to zero effect in LR

even if your data is a modern cross section, immigration from the past will affect it.

Problem is particularly acute if shift and share are close to each other in time.

SSIV dynamics

They consider a wage equation based on

- a Cobb-Douglas production function
- sluggish adjustment of capital

which gives them the following estimating equation,

$$\Delta \ln w_{l,t} = \beta_0 + \beta_1 m_{l,t} + \left[\Delta \ln \theta_{l,t} + \beta_1 \gamma (\ln k^*_{l,t-1} - \ln k_{l,t-1}) + \epsilon_{l,t}\right]$$

where

- θ is local factor productivity (endog.)
- $m_{l,t}$ is migration to location l in time t
  - mix of past settlement and labor market pull factors
- k is the capital-labor ratio (adjusts at rate γ)

SSIV dynamics

They propose a solution by controlling for the lagged immigration shares,

$$\Delta \ln w_{l,t} = \beta_0 + \beta_1 m_{l,t} + \beta_1 m_{l,t-1} + \eta_{l,t}$$

where they construct and instrument for each of these based on

- the baseline migrant shares in t^0
- the overall shift in migrants from
  - $t-1$ to t for the IV of $m_{l,t}$
  - $t-2$ to t for the IV of $m_{l,t-1}$

The length of the lags depends on the application, data, and speed of adjustment γ.

- This works mainly in panels.
- Prohibitively strong data demands on cross sectional data.

Measurement in the treatment

Our initial equation was (simplified),

$$\% \text{ Republican}_{c,2016} = \alpha_s + \beta \% \text{ white Southerners}_{c,1940} + \epsilon_c$$

now we control for

$$\begin{aligned} \% \text{ Republican}_{c,2016} = \alpha_s &+ \beta_1 \% \text{ white Southerners}_{c,1940} \\ &+ \beta_2 \% \text{ white Southerners}_{c,1900} + \epsilon_c \end{aligned}$$

but we only have one instrument.

Measurement in the treatment

Our initial equation was (simplified),

$$\% \text{ Republican}_{c,2016} = \alpha_s + \beta\% \text{ white Southerners}_{c,1940} + \epsilon_c$$

now we control for

$$\begin{aligned}\% \text{ Republican}_{c,2016} = \alpha_s &+ \beta_1\% \text{ white Southerners}_{c,1940} \\ &+ \beta_2\% \text{ white Southerners}_{c,1900} + \epsilon_c\end{aligned}$$

but we only have one instrument. Another specification you often see in the literature is,

$$\% \text{ Republican}_{c,2016} = \alpha_s + \beta\Delta\% \text{ white Southerners}_{c,1900-40} + \epsilon_c$$

What's the difference to the second equation?

Measurement in the treatment

None.

$$\beta\Delta\% \text{ white Southerners}_{c,1900-40} = \beta\% \text{ white Southerners}_{c,1940} - \beta\% \text{ white Southerners}_{c,1900}$$

However, this makes the strong assumption that

- β is the same in 1940 and 1900
- we can somehow identify these two terms with only one instrument (i.e. we are underidentified)

Not clear whether/why this should work but

- these flaws are less obvious to most referees
- a lot of papers in the literature do this

No uniformly agreed specification in the literature at this point.

Inference

Adão, Kolesar, and Morals (2019) study inference in SSIV settings with quasi-random shocks (BHJ setting).

Main take-away from there paper:

- Observations with similar shares are likely to have correlated shocks
- If the errors are similarly clustered, the large sample distribution of $\widehat{\beta}$ is not well approximated by standard CLT
- Clustering does not help much here (even if clustered by distance, for instance)

Inference

Adão, Kolesar, and Morals (2019) study inference in SSIV settings with quasi-random shocks (BHJ setting).

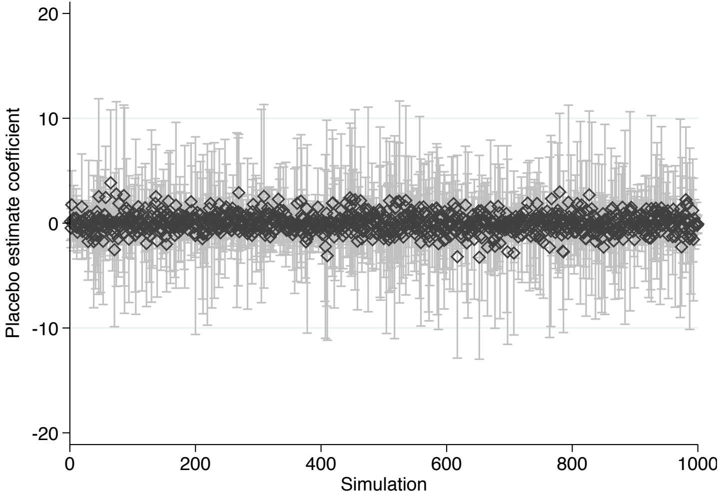
Main take-away from there paper:

- Observations with similar shares are likely to have correlated shocks
- If the errors are similarly clustered, the large sample distribution of $\widehat{\beta}$ is not well approximated by standard CLT
- Clustering does not help much here (even if clustered by distance, for instance)

This can lead to large size distortions in significance testing,

- tests with nominal 5% rejection rate can reject true nulls in up to 55% of placebo shock realizations
- i.e. SSIV tend to falsely give significant results

Inference



Generate data with 1,000 random draw of

- simulated standard normal $N(0, 5)$, use this as shift in your SSIV
- run your SSIV regression with the placebo shifts 1,000 times

Inference

Solution 1:

- Estimate IV regression using
  - the shocks as instruments (i.e. level of IV variation in BHJ)
  - weighting by the shares

Stata package `ssaggregate` can help translate your data to the shock-level.

Solution 2:

- AKM derive a new CLT and s.e. estimation method for "exposure clustering"
- Stata command `ivreg_ss`

Still need more work to know what happens in the GPSS world.