# "Bartik Shift-Share Instruments" Discussion of Goldsmith-Pinkham, Sorkin, and Swift, AER 2020

Nathan Schiff
Shanghai University of Finance and Economics

Graduate Urban Economics, Lecture 12
June 7, 2021

# Bartik or Shift-Share Instruments

Very common instrument that uses lagged stocks of a variable at a regional level combined with aggregate flows to predict current values of an endogenous variable comprised of different categories (ex: industries, immigrant groups)

The stocks are broken out into categories and transformed into shares (share of region $l$ total employment in industry $k$, share of total immigrants from country $c$ going to location $l$)

These are then multiplied by aggregate flows (total growth in industry $k$, total outflow from country $c$) and summed to create a prediction for the endogenous variable

# An Important Instrument

Name comes from 1991 book by Bartik: “Who Benefits from State and Local Economic Development Policies?” (free to read on JSTOR)

Used by Blanchard and Katz (Brookings 1992), Bound and Holzer (JoLE 2000), Card (JoLE 2001), Moretti (AER 2004), Card (AER 2009), Aizer (AER 2010), Duranton and Turner (AER 2011), Beaudry, Green, and Sand (Econometrica 2012), Autor, Dorn, Hanson (AER 2013), Diamond (AER 2016), and many others

We saw it in Baum-Snow, Brandt, Henderson, Turner, Zhang (ReStat 2017) to predict population growth (lagged provincial migrant shares with current outflows)

# Not Clear How It Works...Until Now

What is the identification assumption?

What can we do to provide support for the identification assumption?

What is the role of the national shocks and local shares?

Where does most of the identification come from; is it all group shares of just a few (ex: just a few important industries)?

# Recent Work on Bartik Instruments

1. Jaeger, Ruist, Stuhler, NBER WP, 2018
2. Adao, Kolesar, Morales, QJE 2019
3. Goldsmith-Pinkham, Sorkin, Swift, AER 2020
4. Borusyak, Hull, Jaravel, ReStud 2021
5. Broxterman, Larson, J. Regional Science 2021

We will only discuss Goldsmith-Pinkham et. al., but:

Borusyak et. al. has a different view on exogeneity requirements

Jaeger et. al. focuses specifically on predicting immigration flows

# Basic Setup: Growth Between Two Periods

We want to estimate effect of employment growth in location $l$, $x_l$, on wage growth in $l$, $y_l$:

$$y_l = \rho + \beta_0 x_l + \epsilon_l \tag{1}$$

Overall employment growth can be decomposed into industry-level growth in $l$ ($K$ industries):

$$x_l = \sum_{k=1}^{K} z_{lk} g_{lk} \tag{2}$$

Industry level growth has national and local components:

$$g_{lk} = g_k + \tilde{g}_{lk} \tag{3}$$

Bartik instrument:

$$B_l = \sum_{k=1}^{K} z_{lk} g_k \tag{4}$$

# Full Panel Setup

Full panel setup can predict multiple periods of growth, using period specific national growth rates and industry shares from a single period $t = 0$

$$y_{lt} = D_{lt}\rho + x_{lt}\beta_0 + \epsilon_{lt} \tag{5}$$

$D_{lt}$ are location-time fixed effects

$$x_{lt} = Z_{lt}G_{lt} = \sum_{k=1}^{K} z_{lkt}g_{lkt} \tag{6}$$

$$g_{lkt} = g_{kt} + \tilde{g}_{lkt} \tag{7}$$

$$B_{lt} = Z_{l0}G_t = \sum_{k=1}^{K} z_{lk0}g_{kt} \tag{8}$$

# Bartik Instrument Equivalent to GMM IV

Each industry's initial shares can be thought of as a separate instrument; multiplying by aggregate growth and summing is one way of combining multiple instruments into a single instrument

Generalized Method of Moments (GMM) is a technique to estimate coefficients by solving a set of conditions (moments)

With instruments $z_i \in Z$ and error term $\epsilon_i$, each moment is $E(z_i \epsilon_i) = 0$ (exclusion restriction)

If number of instruments equals number of coefficients, this is exactly IV estimation; if more instruments then we choose weighting matrix to minimize squared error. This is GMM IV.

GSS show that a Bartik instrument is the same as using each industry's share as an instrument and then estimating with GMM IV, where the GMM weights matrix is equal to the cross product of industry growth rates

# Identification: Shares are Exogenous

Shares are instruments: thus shares must be uncorrelated with error term, conditional on controls

This is standard conditional exogeneity assumption: $E(\epsilon Z|X) = 0$

This means shares are uncorrelated with unobservables affecting *growth* rates; note that levels are not a problem (ex: high wage level and high industry share)

What about national growth rates and exogeneity? These just change weighting matrix (efficiency); not important for identification

If shares are instruments, do we need all of them?

# Importance of each instrument: Rotemberg weights

Many applications of Bartik instruments use shares from many categories (ex: hundreds of industries)

Are all of these equally important? No.

GSS use Rotemberg weights (Rotember 1983) to provide a measure (sums to one) of how important each share is

Method is beyond scope of our class, but they provide Stata and R code to easily compute these weights: https://github.com/paulgp/bartik-weight

Weights can be very useful to see which industries are important, and thus whether exogeneity assumptions are plausible for instruments with largest weight

# Simplest case: two industries, one period growth

GSS use a simple example with just two industries and growth between two periods to build intuition

With just two industries and employment shares that add to one, we know industry mix from share of just one industry

Thus there is just a single instrument and a single coefficient, so GMM is identical to IV (generally, if there are $K$ industries whose shares add to 1, there are $K - 1$ instruments)

This makes easy to see some of their main points–we will concentrate on this simple case

# Simplest case identical to IV with shares

Want to estimate: $y_l = \rho + \beta_0 x_l + \epsilon_l$

Two industries: $z_{l1} + z_{l2} = 1$; $x_l = g_{l1} * z_{l1} + g_{l2} * z_{l2}$

$$B_l = z_{l1} g_1 + z_{l2} g_2 \qquad (9)$$

Substitute $z_{l2} = 1 - z_{l1}$ into first stage:

$$x_l = \gamma_0 + \gamma B_l + \eta_l = \gamma_0 + \gamma g_2 + \gamma(g_1 - g_2) z_{l1} + \eta_l \qquad (10)$$

What if we just used $z_{l1}$ directly as the instrument?

$$x_l = \delta_0 + \delta z_{l1} + \delta_l \qquad (11)$$

Difference is simply multiplying $z_{l1}$ by a constant $(g_1 - g_2)$, thus $\gamma = \delta/(g_1 - g_2)$ and the IV estimates of $\beta_0$ will be *identical*!

# Two industries, two periods growth

Want to estimate: $y_{lt} = \tau_t + \beta_0 x_{lt} + \epsilon_{lt}$;

Use only initial shares: $z_{l2,0} = 1 - z_{l1,0}$

$$B_{lt} = g_{1t} z_{l1,0} + g_{2t} z_{l2,0} = g_{2t} + (g_{1t} - g_{2t}) z_{l1,0} \tag{12}$$

$$x_{lt} = (\tau_t + g_{2t}\gamma) + z_{l1,0}(g_{1t} - g_{2t})\gamma + \eta_{lt} \tag{13}$$

Rewrite with time FE:

$g_{1t} - g_{2t} = 1(t=1)(g_{11} - g_{21}) + 1(t=2)(g_{12} - g_{22})$

$x_{lt} = (\tau_t + g_{2t}\gamma) + z_{l1,0} 1(t=1)(g_{11} - g_{21})\gamma + z_{l1,0} 1(t=2)(g_{12} - g_{22})\gamma + \eta_{lt}$

Now compare to first stage with shares interacted with time FE:

$$x_{lt} = \tau_t + z_{l1,0} 1(t=1)\delta_1 + z_{l1,0} 1(t=2)\delta_2 + \delta_{lt} \tag{14}$$

Identical if we constrain $\delta_1/(g_{11} - g_{21}) = \delta_2/(g_{12} - g_{22})$

# Interpretation: 2 industries, 2 periods growth

GSS point out that with two periods, this is somewhat like differences-in-differences

The industry shares can be viewed as a measure of exposure to a policy

The growth rates measure the policy size

Then we are comparing locations with more exposure to locations with less exposure, across a period where the policy is stronger and a period where the policy is weaker

My question: shouldn't the unconstrained estimator always be just as good as Bartik? In this case, why use Bartik–is it simply because the instrument has a better connection to theory?

# Worth Additional Study

Many points in the paper I did not cover, well worth additional study

Very accessible (readable) for an econometrics paper; examples are also quite interesting

Time for Stata exercise