“Bartik Instruments: What, When, Why, and How?”

American Economic Review, 2020

by Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift

Jaime Arellano-Bover

Applied Micro in Stata Workshop

RoME, 2022

RoME Rome Masters in Economics

Motivation and research question

- *Bartik* type instrumental variables are common in labor, public, development, macro, trade, finance
- Definition:
  - Bartik-like instrument is one that uses the *inner product structure* of the endogenous variable to construct an instrument
- Examples:
  - Employment growth rates and industry shares (Bartik, 1991; Katz, 1992), Immigrant enclaves and immigrant inflows (Card, 2009) China trade shock and industry shares (Autor et al., 2013), banking relationships and credit supply (Greenstone et al., 2020)
- Until recently, even if Bartik instruments were very commonly used, they were not well understood
- Goal of the paper:
  - Open black box of Bartik instruments, formalize their structure, unpack the variation underlying the instrument, and propose tools to assess validity

RoME Rome Masters in Economics

Empirical application: Card 2009

$$y_l = \beta_0 + \beta \ln x_l + \beta_2 X_l + \varepsilon_l$$

- *l* is a city
- y_l residual log wage gap between immigrant and natives
- x_l ratio of immigrant to native total hours of work
- X_l vector of city-level controls
- β parameter of interest: inverse elasticity of substitution between immigrants and natives
- Single cross section at the city level, for 124 US cities, in 2000 (from Census)

Endogeneity concern and Bartik IV

- Endogeneity concern: positive labor demand shock that draws immigrants into a location disproportionately relative to natives
  - i.e., shock that increases both ε_l (earnings) and x_l (relative supply)

- B_l is the Bartik instrument for x_l, where:

$$B_l = \sum_k z_{lk,1980} g_k$$

  - *k* indexes a country-of-origin
  - $z_{lk,1980}$:1980 share of all immigrants in the US from country *k* who live in city *l*
  - g_k: number of immigrants from country *k* who arrived in the US between 1990-2000

RoME Rome Masters in Economics

Key results of the paper

- TSLS estimator of β using B_l as instrument, $\widehat{\beta_B}$ is numerically equivalent to GMM estimator with vector of $z_{lk,1980}$ as instruments and a weight matrix constructed from national shocks g_k
  - Bartik instrument equivalent to using local shares as instruments, so exogeneity condition should be interpreted in terms of the shares

- Bartik estimator can be decomposed into weighted sums of TSLS estimators using each individual $z_{lk,1980}$ weight as instrument:

$$\hat{\beta}_B = \sum_k \hat{\alpha}_k \hat{\beta}_k$$

(Subset of) Paper's advice for Bartik users

- Compute (Rotemberg) weights $\hat{\alpha}_k$
- Is distribution of weights very skewed?
  - If so, we are particularly worried of exogeneity assumption not holding for countries *k* who have a large weight in the Bartik estimate

- Report instruments associated with largest values of $\hat{\alpha}_k$
- Check for correlates of country origin shares $z_{lk,1980}$, especially those with high $\hat{\alpha}_k$

RoME Rome Masters in Economics