American Economic Review 2020, 110(8): 2586–2624
https://doi.org/10.1257/aer.20181047

Bartik Instruments: What, When, Why, and How[†]

By Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift*

The Bartik instrument is formed by interacting local industry shares and national industry growth rates. We show that the typical use of a Bartik instrument assumes a pooled exposure research design, where the shares measure differential exposure to common shocks, and identification is based on exogeneity of the shares. Next, we show how the Bartik instrument weights each of the exposure designs. Finally, we discuss how to assess the plausibility of the research design. We illustrate our results through two applications: estimating the elasticity of labor supply, and estimating the elasticity of substitution between immigrants and natives. (*JEL* C51, F14, J15, J22, L60, R23, R32)

The Bartik instrument is named after Bartik (1991), and popularized in Blanchard and Katz (1992).[1] These papers define the instrument as the local employment growth rate predicted by interacting local industry employment shares with national industry employment growth rates. The Bartik approach and its formally identical variants have since been used across many fields in economics, including labor, public, development, macroeconomics, international trade, and finance.

In our exposition, we focus on the canonical setting of estimating the labor supply elasticity, but our results apply more broadly wherever Bartik-like instruments

* Goldsmith-Pinkham: Yale School of Management (email: paul.goldsmith-pinkham@yale.edu); Sorkin: Stanford University and NBER (email: sorkin@stanford.edu); Swift: Unaffiliated (email: henryswift@gmail.com). Thomas Lemieux was the coeditor for this article. Thanks to the anonymous referees, Rodrigo Adao, Isaiah Andrews, David Autor, Tim Bartik, Paul Beaudry, Kirill Borusyak, Jediphi Cabal, Arun Chandrasekhar, Gabriel Chodorow-Reich, Damon Clark, Richard Crump, Rebecca Diamond, Mark Duggan, Matt Gentzkow, Andrew Goodman-Bacon, David Green, Gordon Hanson, Caroline Hoxby, Peter Hull, Guido Imbens, Xavier Jaravel, Pat Kline, Magne Mogstad, Maxim Pinkovskiy, Luigi Pistaferri, Giovanni Righi, Ben Sand, Pedro Sant'Anna, Juan Carlos Suárez Serrato, Jan Stuhler, Melanie Wallskog, Kenneth West, Wilbert van der Klaauw, Eric Zwick, and numerous seminar participants for helpful comments. Thanks to Maya Bidanda, Jacob Conway, and Victoria de Quadros for research assistance. Thanks to David Card for sharing code, and Rodrigo Adao, Kirill Borusyak, Peter Hull, Xavier Jaravel, Michal Kolesár, and Eduardo Morales for sharing data. Swift was supported by the National Science Foundation Graduate Research Fellowship. Part of the work on this paper was completed while Goldsmith-Pinkham was employed by the Federal Reserve Bank of New York. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve Board. All errors are our own, please let us know about them.

[†] Go to https://doi.org/10.1257/aer.20181047 to visit the article page for additional materials and author disclosure statements.

[1] The intellectual history of the Bartik instrument is complicated. The earliest use of a shift-share type decomposition we have found is Perloff (1957, Table 6), which shows that industrial structure predicts the *level* of income. Freeman (1980) is one of the earliest uses of a shift-share decomposition interpreted as an instrument: it uses the change in industry composition (rather than differential growth rates of industries) as an instrument for labor demand. What is distinctive about Bartik (1991) is that the book not only treats it as an instrument, but also, in the Appendix, explicitly discusses the logic in terms of the national component of the growth rates.

are used. For simplicity, consider the cross-sectional structural equation linking wage growth to employment growth,

$$y_l = \rho + \beta_0 x_l + \epsilon_l,$$

where y_l is wage growth in location l between two time periods, x_l is the employment growth rate, ρ is a constant, and ϵ_l is a structural error term that is correlated with x_l. Our parameter of interest is β_0, the inverse elasticity of labor supply. We use the Bartik instrument to estimate β_0.

The Bartik instrument combines two accounting identities. The first is that employment growth is the inner product of industry shares and local industry growth rates:

$$x_l = \sum_k z_{lk} g_{lk},$$

where z_{lk} is the share of location l's employment in industry k, and g_{lk} is the growth rate of industry k in location l. The second is that we can decompose the industry-growth rates as

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

where g_k is the industry growth rate and $\tilde{g}_{lk}$ is the idiosyncratic industry-location growth rate. The Bartik instrument is the inner product of the industry-location shares and the industry component of the growth rates; formally, $B_l = \sum_k z_{lk} g_k$.

Because the Bartik instrument combines two accounting identities, it is always possible to construct it. It is not plausible, however, that the Bartik instrument always provides a valid identification strategy. In this paper, we open the black box of the Bartik instrument by formalizing its structure and unpacking the variation that the instrument uses. Our goal is to enable researchers to use familiar tools to distinguish between situations where the Bartik instrument would and would not be valid.

In this paper, we discuss the Bartik instruments' identification as coming from the shares. The basis of this view is a numerical equivalence result: we show that the two-stage least squares (TSLS) estimator with the Bartik instrument (the Bartik estimator) is numerically equivalent to a generalized method of moments (GMM) estimator with the local industry shares as instruments and a weight matrix constructed from the national growth rates. We interpret this result as saying that using the Bartik instrument is "equivalent" to using local industry shares as instruments, and so the exogeneity condition should be interpreted in terms of the shares. In contrast, Borusyak, Hull, and Jaravel (2019) emphasizes that under some assumptions the consistency of the estimator can also come from the shocks,[2] and they also provide a motivating numerical equivalence result. How can researchers tell which quasi-experimental design they are using? We argue that a researcher is likely using a research design based on the shares assumption if they (i) describe their research design as reflecting differential exogenous exposure to common

[2] Adao, Kolesár, and Morales (2019) discusses inferential issues in this setup.

shocks, (ii) emphasize a two-industry example, and/or (iii) emphasize shocks to specific industries as central to their research design.

Once we think about the shares as the instruments, the implied empirical strategy is an exposure research design, where the industry shares measure the differential exogenous exposure to the common shock. In settings where the researcher has a pre-period, this empirical strategy is just difference-in-differences. Because the shares are typically equilibrium objects and likely codetermined with the level of the outcome of interest, it can be hard to assume that the shares are uncorrelated with the levels of the outcome. But this assumption is not necessary for the empirical strategy to be valid. Instead, the strategy asks whether differential exposure to common shocks leads to differential *changes* in the outcome. For example, in the canonical setting, the outcome is wage *growth*, in the China shock setting the outcome is *change* in manufacturing employment, and in the immigrant enclave setting it is *changes* in the residual log wage gap between immigrants and natives. Hence, the empirical strategy can be valid even if the shares are correlated with the levels of the outcomes.

How does one build the credibility of such an exposure design? The central identification worry is that the industry shares predict outcomes through channels other than those posited by the researcher. One way to assess this possibility is to look at correlates of the shares. If these correlates suggest other channels through which the shares affect outcomes in the relevant period, then we might be skeptical of the identifying assumption. Second, in some settings there is a pre-period, as in a standard difference-in-differences design. In this case, we can test for parallel pretrends. Given that the design exploits level differences in the shares, by exploring trends in changes we can assess the plausibility of the assumption that the common shock caused the change in the changes, or whether there were preexisting differences in the changes.

There is a third way to explore the validity of the research design, based on the observation that the Bartik instrument is a particular way of combining many instruments. Under the null of constant effects, a researcher can consider alternative estimators which combine multiple instruments or run overidentification tests. One interpretation of the divergence between estimators and the failure of overidentification tests is that the null of constant effects is unreasonable, and to instead interpret these tests as pointing to the presence of treatment effect heterogeneity, rather than failure of exogeneity. We follow Borusyak, Hull, and Jaravel (2019) and Adao, Kolesár, and Morales (2019) and consider a restricted form of linear heterogeneity where there are constant effects within each location. We highlight that even if each instrument separately places convex weights on each location's parameter, it is possible that the Bartik estimator would not have a local average treatment effect-like interpretation as a weighted average of treatment effects. To the extent that researchers wish to embrace a treatment effect heterogeneity interpretation of the Bartik instrument, they should be comfortable with the patterns of underlying heterogeneity. We develop a visual diagnostic to aid researchers in this task.

How does the Bartik instrument combine the exposure designs? We build on Rotemberg (1983) and decompose the Bartik estimator into a weighted sum of the just-identified instrumental variable estimators that use each industry share (z_{lk}) as a separate instrument. The weights, which we refer to as Rotemberg weights, are simple to compute and sum to 1. They depend on the covariance between the *k*th instrument's fitted value of the endogenous variable and the endogenous variable itself. The weights are a scaled version of the Andrews, Gentzkow, and Shapiro (2017) sensitivity-to-misspecification parameter, and tell us how sensitive the overidentified estimate of β_0 is to misspecification (i.e., endogeneity) in any instrument. Heuristically, they also tell us which exposure design gets more weight in the overall estimate, and thus which of these identifying assumptions is most worth testing. If the high-weight designs, where it is concrete what comparisons the researcher is doing, pass basic specification tests, then researchers should feel reassured about the overall empirical strategy.

In many contexts where researchers use Bartik instruments, they are used in the reduced form, whereas in our analysis we discuss the instrumental variables setting. We note that the insights of this paper still apply when Bartik is used in the reduced form. Specifically, the relevant moment condition (exclusion restriction) is still the same. Moreover, it is still possible to compute the Rotemberg weights.

We note two limitations to our analysis. First, we assume locations are independent and so ignore the possibility of spatial spillovers or correlation.[3] Second, we assume that the data consist of a series of steady states.[4]

To summarize, we view our contribution as explaining identification in the context of Bartik instruments in two ways. First, our GMM result shows that Bartik is numerically equivalent to using industry shares as instruments. Hence, we argue that the typical identifying assumption is best interpreted in terms of industry shares, rather than growth rates. Second, we build on Andrews, Gentzkow, and Shapiro (2017) to provide tools to measure the "identifying variation," and formalize how to use Rotemberg weights to highlight the subset of instruments to which the estimated parameter is most sensitive to endogeneity.

Applications.—We illustrate our results through two applications. In our first application, we look at the canonical example of estimating the inverse elasticity of labor supply in US Census data using decadal differences from 1980–2010 and instrumenting for labor demand with the Bartik instrument. We first show that the national growth rates explain less than 1 percent of the variance of the Rotemberg weights. Hence, the growth rates are a poor guide to understanding what variation in the data is driving estimates. Second, the weights are skewed, with over 40 percent of the weight on the top five industries. In the particular, the oil and gas extraction industry receives the largest weight. Hence, a concrete example of the comparisons being made by the estimator is between changes in employment growth and wage growth in places with more and less oil and gas extraction. Third, industry shares, including oil and gas extraction, are correlated with many observables, including the immigrant share, which potentially predicts innovations in labor supply. Fourth, alternative estimators deliver substantively different point estimates and overidentification tests reject the null of exogeneity. Fifth,

[3] This force is standard in spatial equilibrium models: see Redding and Rossi-Hansberg (2017) for a recent survey. Monte, Redding, and Rossi-Hansberg (2018) presents evidence for the presence and economic importance of spatial spillovers through changes in commuting patterns in response to local labor demand shocks.

[4] See Jaeger, Ruist, and Stuhler (2019) for discussion of out-of-steady-state dynamics in the context of immigration.

consistent with the overidentification tests rejecting, we find substantial visual dispersion in the estimates from each individual instrument. Moreover, some of the outlying point estimates receive negative Rotemberg weights, which suggests that, under the treatment effect heterogeneity interpretation, some of the underlying effects receive negative weight so that there is unlikely to be a LATE-like interpretation of the parameter estimate.

In our second application, we estimate the inverse elasticity of substitution between immigrants and natives in 2000 (following the empirical strategy of Card 2009). Here, the relevant shares are the share of migrants from an origin country who live in a particular location in the base year, and the shocks are the immigrant inflows. First, we find that for high school equivalent workers, the Rotemberg weights are almost completely explained by the immigrant inflows. For the college equivalent workers, the explanatory power of the inflows is higher than in our other two examples. Hence, the growth rates (the shocks) are a good guide to the variation in the data that drives estimates. Second, for high school equivalent workers, the share of Mexican immigrants in a city in 1980 gets almost one-half of the weight in the estimator, a possibility that Card (2009) acknowledges. Hence, for high school equivalent workers, a concrete example of the comparison the estimator is making is between places with more and fewer Mexican immigrants in 1980. For college equivalent workers, the highest weight instrument is the Philippines, and so the comparison is between places with higher and lower Philippines share. Third, among the covariates used by Card (2009), we do not find any systematic patterns of correlations with the immigrant shares. Fourth, unlike in our other examples, most overidentification tests fail to reject and we do not find differences among estimators. Fifth, we find limited evidence of statistically significant pretrends for the high school equivalent workers. In contrast, we find statistically significant pretrends for the estimates involving the college equivalent workers, consistent with the concerns emphasized by Jaeger, Ruist, and Stuhler (2019).

Besides these two examples, a much broader set of instruments is Bartik-like. We define a Bartik-like instrument as one that uses the inner product structure of the endogenous variable to construct an instrument. In online Appendix Section A, we discuss the China shock of Autor, Dorn, and Hanson (2013) and present our complete set of diagnostics in this application. In online Appendix Section B, we discuss two additional examples. First, researchers, such as Greenstone, Mas, and Nguyen (2020), interact preexisting bank lending shares with changes in bank lending volumes to instrument for credit supply. Second, Acemoglu and Linn (2004) interacts age-group spending patterns with demographic changes to instrument for market size.

Literature.—A vast literature uses Bartik-like instruments, and many of these papers discuss the identifying assumptions in ways that are close to the benchmark results here. For example, Baum-Snow and Ferreira (2015, p. 50) surveys the literature and states that the "validity [of the Bartik instrument] … relies on the assertion that neither industry composition nor unobserved variables correlated with it directly predict the outcomes of interest conditional on controls." Similarly, Beaudry, Green, and Sand (2012) provides a careful discussion of identifying assumptions in the context of an economic model. Given the vast diversity of ways in which Bartik instruments are discussed and understood in the literature, we can only claim novelty for the formalism along this dimension.

Beyond the vast literature of papers using Bartik-like instruments, this paper is also related to a growing literature that comments on specific papers (or literatures) that use Bartik-like instruments. This literature includes at least three papers: Christian and Barrett (2017), which comments on Nunn and Qian (2014); Jaeger, Joyce, and Kaestner (2020), which comments on Kearney and Levine (2015); and Jaeger, Ruist, and Stuhler (2019), which comments on the use of the immigrant enclave instrument. Relative to this literature, our goal is to develop a formal econometric understanding of the Bartik instrument and provide methods to increase transparency in its use.

I. Equivalence between Bartik IV and GMM with Industry Shares

We first show that the Bartik instrument is numerically equivalent to using industry shares as instruments, which we use to argue that the identification condition is best interpreted in terms of industry shares. We begin this section by setting up the most general case: panel data with K industries, T time periods, and controls. Through a series of special cases, we then build up to the main result. To focus on identification issues, we discuss infeasible Bartik, where we assume that we know the common national component of industry growth rates. Section II discusses consistency and identifying assumptions.

A. *Full Panel Setup*

We begin by setting up the general panel data case with K industries and T time periods. This setup most closely matches that used in empirical work. It allows for the inclusion of both location and time fixed effects as well as other controls.

We are interested in the following structural equation:

$$y_{lt} = D_{lt}\rho + x_{lt}\beta_0 + \epsilon_{lt}. \tag{1}$$

In the canonical setting, l indexes a location, t a time period, y_{lt} is wage growth, D_{lt} is a vector of Q controls which could include location and time fixed effects, x_{lt} is employment growth, and ϵ_{lt} is a structural error term. The parameter of interest is β_0. We assume that the ordinary least squares (OLS) estimator for β_0 is biased and we need an instrument to estimate β_0.

The Bartik instrument exploits the inner product structure of employment growth. Specifically, employment growth is the inner product of industry shares and industry-location growth rates

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

where Z_{lt} is a $1 \times K$ vector of industry-location-time period shares, and G_{lt} is a $K \times 1$ vector of industry-location-time period growth rates where the kth entry is g_{lkt}.

We decompose the industry-location-period growth rate into an industry-period and an idiosyncratic industry-location-period component:

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

where we now define $\tilde{G}_{lt}$ as the $K \times 1$ vector of the $\tilde{g}_{lkt}$. In some applications it is natural to use the sample mean (or a leave-one-out sample mean) of g_{lkt} as an estimator for g_{kt}, but none of our results are specific to this choice. We fix industry shares to an initial time period. We do this for two reasons. First, this choice follows convention. Second, this choice makes the analogy to difference-in-differences clearer: by fixing the shares to an initial time period prior to the shock, there is a single cross-sectional exposure difference that the design is exploiting. Then the Bartik instrument is the inner product of the initial industry-location shares and the industry-period growth rates:

$$B_{lt} = Z_{l0} G_t = \sum_k z_{lk0} g_{kt},$$

where G_t is a $K \times 1$ vector of the industry growth rates in period t (the kth entry is g_{kt}), and Z_{l0} is the $1 \times K$ vector of industry shares in location l. Hence, we have a standard two-stage least squares setup where the first stage is a regression of employment growth on the controls and the Bartik instrument:

$$x_{lt} = D_{lt}\tau + B_{lt}\gamma + \eta_{lt},$$

and the structural equation is given by (1).

Let $y_l = (y_{l1}, \ldots, y_{lT})$, $x_l = (x_{l1}, \ldots, x_{lT})$, $Z_l = (Z_{l1}, \ldots, Z_{lT})$, $\tilde{G}_l = (\tilde{G}_{l1}, \ldots, \tilde{G}_{lT})$, $G_K = (G_1, \ldots, G_T)$, $D_l = (D_{l1}, \ldots, D_{lT})$, and $\epsilon_l = (\epsilon_{l1}, \ldots, \epsilon_{lT})$. We assume that the data

$$\left\{y_l, D_l, \tilde{G}_l, Z_l, Z_{l0}\right\}_{l=1}^{L}$$

are drawn i.i.d. across l, and view G_K as fixed.[5]

We assume that D_{lt} is strictly exogenous, and focus on estimating β_0 using residual regression. Define $Y_L = (y_1, \ldots, y_L)$, $X_L = (x_1, \ldots, x_L)$, $D_L = (D_1, \ldots, D_L)$, and $\boldsymbol{\epsilon}_L = (\epsilon_1, \ldots, \epsilon_L)$. Let $M_D = I_L - D_L(D_L' D_L)^{-1} D_L'$ denote the annihilator matrix for D, the $L \times Q$ matrix of controls, where I_L is the $L \times L$ identity matrix. We define $X_L^{\perp} \equiv M_D X_L$ and $Y_L^{\perp} \equiv M_D Y_L$ to be the residualized X_L and Y_L such that $M_D(Y_L - X_L\beta_0) = M_D(D_L\rho + \epsilon_L) = M_D\epsilon_L$, since $M_D D_L = 0$. Finally, define $\epsilon_L^{\perp} \equiv M_D\epsilon_L$.

B. *Equivalence in Three Special Cases*

We build up to the general result that the Bartik instrument is numerically equivalent to using industry shares as instruments for a particular weight matrix

[5] This assumption allows for dependence within l: the data are *not* i.i.d. within l.

in GMM through three special cases. Each of these special cases also illustrates a research design implicit in using a Bartik instrument and suggests a specification test.

Two Industries and One Time Period.—With two industries whose shares sum to 1 within each location and one time period, the Bartik instrument is identical to using one of the industry shares as an instrument. To see this, expand the Bartik instrument:

$$B_l \;=\; z_{l1} g_1 + z_{l2} g_2,$$

where g_1 and g_2 are the industry components of growth. Since the shares sum to 1, we can write the second industry share in terms of the first, $z_{l2} \;=\; 1 - z_{l1}$, and simplify the Bartik instrument to depend only on the first industry share:

$$B_l \;=\; g_2 + (g_1 - g_2) z_{l1}.$$

Because the only term on the right-hand side with a location subscript is the first industry share, the cross-sectional variation in the instrument comes from the first industry share. Substitute into the first stage

$$x_l \;=\; \gamma_0 + \gamma B_l + \eta_l \;=\; \underbrace{\gamma_0 + \gamma g_2}_{\text{constant}} + \underbrace{\gamma (g_1 - g_2)}_{\text{coefficient}} z_{l1} + \eta_l.$$

This equation shows that the difference between using the first industry share and Bartik as the instrument is to rescale the first-stage coefficients by the difference in the growth rates between the two industries $(1/(g_1 - g_2))$. But whether we use the Bartik instrument or the first industry share as an instrument, the predicted employment growth (and hence the estimate of the inverse elasticity of labor supply) would be the same. Hence, with two industries, using the Bartik instrument in TSLS is numerically identical to using z_{l1} (or z_{l2}) as an instrument.

What is the research design inherent in this special case? Here, z_{l1} measures exposure to the policy that affects industry 1, and $g_1 - g_2$ is the size of the policy. The outcome is y_l, which is the *change* in outcomes between two periods. Hence, in this special case the empirical strategy asks about the effects of levels of z_{l1} on changes in y_l. The identification concern is whether z_{l1} is correlated with *changes* in the outcome, and not *levels* of the outcome. As we discuss more in Test 1 in Section V, studying correlates of z_{l1} is helpful in making clear the types of concerns one might have. Concretely, while z_{l1} might be correlated with many variables that predict the level of the outcome, this correlation is not necessarily a problem for the research design. Instead, the central question a researcher should have in mind is whether these correlates predict changes in the outcome in the relevant period.

Why would OLS be biased but Bartik be a valid instrument? The form of endogeneity that Bartik can address is correlation between ϵ_l and the location-specific portions of the growth rates: $\tilde{g}_{l1}$ and $\tilde{g}_{l2}$. For example, if there are amenity shocks in an area, then these shocks show up as local industries growing faster than the national average. But these amenity shocks also directly affect wage growth and so generate endogeneity.

Two Industries and Two Time Periods.—In a panel with two time periods, if we interact the time-invariant industry shares with time, then Bartik is equivalent to a special case of using industry shares as instruments. To see this result, we again specialize to two industries, and define the Bartik instrument so that it varies over time:

$$B_{lt} = g_{1t} z_{l10} + g_{2t} z_{l20} = g_{2t} + (g_{1t} - g_{2t}) z_{l10},$$

where g_{1t} and g_{2t} are the industry-by-time growth rate for industry 1 and 2. Because we fix the shares to an initial time-period, denoted by z_{lk0}, the time variation in B_{lt} comes from the difference between g_{1t} and g_{2t}.

To see the relationship between the cross-sectional and panel estimating equations, we restrict our panel setup to have the vector of controls consist solely of time fixed effects. Then the first stage is

$$x_{lt} = \tau_t + B_{lt}\gamma + \eta_{lt}.$$

Now substitute in the Bartik instrument and rearrange the first stage:

$$x_{lt} = \underbrace{(\tau_t + g_{2t}\gamma)}_{\equiv \tilde{\tau}_t} + z_{l10}(g_{1t} - g_{2t})\gamma + \eta_{lt}. \tag{2}$$

This first stage is more complicated than in the cross-sectional case because there is a time-varying growth rate multiplying the time-invariant industry share.

To recover the equivalence between Bartik and using shares as instruments in the panel setting, write $g_{1t} - g_{2t} = \mathbf{1}(t = 1)(g_{11} - g_{21}) + \mathbf{1}(t = 2)(g_{12} - g_{22})$, where $\mathbf{1}(\cdot)$ is the indicator function. Then, rewrite the first stage as

$$x_{lt} = \underbrace{(\tau_t + g_{2t}\gamma)}_{\equiv \tilde{\tau}_t} + z_{l10}\mathbf{1}(t = 1)(g_{11} - g_{21})\gamma + z_{l10}\mathbf{1}(t = 2)(g_{12} - g_{22})\gamma + \eta_{lt}. \tag{3}$$

We can now see the equivalence between Bartik and using the shares as instruments. Since we fix the industry shares in the initial time period, to create time variation in our industry shares regression, consider the regression with initial industry shares interacted with time fixed effects:

$$\text{(Bartik)} \quad x_{lt} = \tilde{\tau}_t + z_{l10}\underbrace{(g_{11} - g_{21})\mathbf{1}(t = 1)\gamma}_{\equiv \tilde{\gamma}_1} + z_{l10}\underbrace{(g_{12} - g_{22})\mathbf{1}(t = 2)\gamma}_{\equiv \tilde{\gamma}_2} + \eta_{lt},$$

$$\text{(Industry Shares)} \quad x_{lt} = \tau_t + z_{l10}\mathbf{1}(t = 1)\tilde{\gamma}_1 + z_{l10}\mathbf{1}(t = 2)\tilde{\gamma}_2 + \tilde{\eta}_{lt}.$$

In this case, the panel regression (with the industry share) gives us two parameters, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$. When will they be identical to γ (the parameter from the Bartik equation)? If we restrict $\tilde{\gamma}_1 = \gamma/(g_{11} - g_{21})$, and $\tilde{\gamma}_2 = \gamma/(g_{12} - g_{22})$, then both parameters will be proportional to the underlying Bartik parameter, γ. If we view z_{l10} as the effect of exposure to a policy, then each $\tilde{\gamma}$ captures the "unscaled" effect on x_{lt}, while γ is rescaled by the size of the policy, where the size of the policy is the difference in national industry growth rates, $g_{1t} - g_{2t}$.

What is the research design inherent in this special case? Viewing the growth rates as a measure of policy size and the industry shares as measures of exposure emphasizes a useful connection to difference-in-differences. In the equations above, a researcher is already considering outcomes and regressors in *changes*, which allows for the possibility of level differences across locations. By using initial industry shares as the right-hand-side regressor in the panel regression, the researcher is asking whether locations with high shares of a particular industry experience differential changes in outcomes following shocks whose effect depends on the size of that industry.

With the two time periods, we can consider period 1 to be a pre-period before a policy takes effect. That is, $g_{11} - g_{21} = 0$. In this case, a researcher can test whether $\tilde{\gamma}_1$ is zero (a test of the parallel pretrends assumption). Intuitively, a researcher is asking whether in the pre-period, the level of z_{l1} predicts changes in the outcome. Failing to find a pretrend gives credence to a research design where the researcher assumes that z_{l1} is relevant for predicting the change in period 2. We return to this point in Test 2 in Section V.

K Industries and One Time Period.—Finally, we show that with K industries as instruments in a generalized method of moments (GMM) estimator setup with a specific weight matrix, the Bartik estimator is identical to using the set of industry shares as instruments.

To show this result, recall that G is the $K \times 1$ vector of industry growth rates, Z is the $L \times K$ matrix of industry shares, Y is the $L \times 1$ vector of outcomes, X is the $L \times 1$ vector of endogenous variables, and $B = ZG$ is the $L \times 1$ vector of Bartik instruments. Let W be an arbitrary $K \times K$ matrix.

We define the Bartik and the GMM estimator using industry shares as instruments:

$$\hat{\beta}_{Bartik} = \frac{B'Y^{\perp}}{B'X^{\perp}}; \quad \text{and} \quad \hat{\beta}_{GMM} = \frac{X^{\perp\prime}ZWZ'Y^{\perp}}{X^{\perp\prime}ZWZ'X^{\perp}}.$$

PROPOSITION 1: *If* $W = GG'$, *then* $\hat{\beta}_{GMM} = \hat{\beta}_{Bartik}$.

PROOF:

See online Appendix Section C.

Proposition 1 says the Bartik instrument and industry shares as instruments are numerically equivalent for a particular choice of weight matrix.

What is the research design inherent in this special case? Under the shares interpretation that we discuss further below, if there is a shock in a single period, then this research design pools many different exposure designs. In Section III, we show the way that Bartik pools these designs. The tools for building the credibility of any given share are the same as in the single instrument case. Moreover, the many instruments provide the researcher with the opportunity to test whether the parameter estimates from all of these instruments are the same using overidentification tests. Alternatively, if these parameters are not similar, the researcher might be interested in trying to characterize this heterogeneity. In Test 3 in Section V, we discuss overidentification tests. In Section IV, we discuss heterogeneity.

Remark 1: When $\sum_{k=1}^{K} z_{lk} = 1$, there are $K - 1$ instruments and not K instruments. In practice, any of the K industries can be dropped by subtracting off that industry's growth rate from the G vector, and the Bartik instrument will maintain its numerical equivalence from Proposition 1. To see the intuition behind this, suppose that $\sum_k z_{lk} = \mathbf{1}\ \forall l$. Consider the first-stage regression:

$$x_l = \gamma_0 + \gamma_1 B_l + \eta_l.$$

Now add and subtract $\gamma_1 \sum_k z_{lk} g_j$ from the right-hand side:

$$x_l = \underbrace{\gamma_0 + \gamma_1 \sum_k z_{lk} g_j}_{\gamma_0 + \gamma_1 g_j} + \gamma_1 \underbrace{\sum_k z_{lk}\left(g_k - g_j\right)}_{B_l - g_j} + \eta_l. \tag{4}$$

This expression generalizes our result from the two industry and one time period example. It says that normalizing the growth rates by a constant g_j changes the first-stage intercept and does not affect the slope estimate. Hence, the first-stage prediction is unaffected.

C. *Summary*

With K industries and T time periods, the numerical equivalence involves creating $K \times T$ instruments (industry shares interacted with time periods). Then, an identical GMM result holds as we proved in the cross section with K industries. Extending the result is notationally cumbersome so we leave the formal details to online Appendix Section D. We now turn to discussing how these finite sample results map into identification conditions.

II. Asymptotic Consistency and Identifying Assumptions

We now consider consistency of the TSLS estimator that uses the Bartik instrument. In the previous section, we established a finite sample equivalence result between the TSLS estimator using the Bartik instrument, and the GMM estimator using industry shares as instruments and a weight matrix defined by the industry growth rates. Here, we use this equivalence to show that a sufficient condition for consistency is strict exogeneity of the shares.

To fix ideas, consider the difference between the TSLS estimator and the parameter of interest:

$$\hat{\beta} - \beta_0 = \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} g_{kt} \sum_{l=1}^{L} z_{lk0}\, \epsilon_{lt}^{\perp}}{\sum_{t=1}^{T} \sum_{k=1}^{K} g_{kt} \sum_{l=1}^{L} z_{lk0}\, x_{lt}^{\perp}}. \tag{5}$$

Broadly, conditions for the consistency of $\hat{\beta}$ can be stated either in terms of the *shares*, the z_{lk0}, or the *shocks*, the g_{kt}. In this paper, we consider a setting where we observe increasingly larger samples of locations, but a fixed number of time periods and industries (fixed T and K). As we show below, in this setting it is natural to state conditions for consistency in terms of the shares.

A natural extension of this setup studied by Borusyak, Hull, and Jaravel (2019) considers a setting where we not only observe increasingly larger samples of locations, but also of industries. They show that while a sufficient condition for consistency of $\hat{\beta}$ is exogeneity of the shares, it is not necessary. With many industries, it is possible to use the exogeneity of the shocks, e.g., g_{kt}, instead.

In this section, we first state the sufficient conditions in our setting, highlighting the relevance and exogeneity assumptions. We then discuss when these exogenous shares assumptions are reasonable, and how they contrast to the exogenous shocks assumptions.

A. *Identifying Assumptions*

Two assumptions must hold for consistency. First, the denominator must converge to a nonzero term. Intuitively, for this assumption to hold, there must be an industry and time period when the industry share has predictive power for x_{lt}, conditional on the controls, and the growth rates g_{kt} cannot weight the covariances in such a way that they exactly cancel. This first condition holds under the following low-level assumption.

ASSUMPTION 1 (Relevance): *For all* $k \in \{1,\ldots,K\}$ *and* $s \in \{1,\ldots,T\}$,

$$x_{lt} = D_{lt}\tau + z_{lk0}\mathbf{1}(t = s)C_{k,s} + \eta_{lt},$$

where $E[\eta_{lt}|z_{lk0},D_{lt}] = 0$, $C_{k,s}$ *is finite for all* k *and* s, *and* $\sum_s\sum_k g_{ks}C_{ks} \neq 0$.

The second necessary assumption for consistency is that the numerator must converge to zero. This assumption is the exclusion restriction, and to hold generically, the industry share must be uncorrelated with the structural error term, *after controlling for* D_{lt}, for industries that have nonzero growth rates. The following identifying assumption ensures that the numerator converges to 0.

ASSUMPTION 2 (Strict Exogeneity): $E[\epsilon_{lt}z_{lk0}|D_{lt}] = 0$ *for all* k *where* $g_k \neq 0$.

This assumption is standard in empirical settings that use exposure designs. For example, this assumption is made in difference-in-differences designs that use location fixed effects.[6]

[6]Even if $E[\epsilon_l z_{lk0}|D_{lt}] \neq 0$, then the numerator can still converge to zero nongenerically if the g_{kt} are such that these biases cancel out exactly. For fixed K and T, this case is unlikely to hold in practice. When K increases, Borusyak, Hull, and Jaravel (2019) shows that this can hold generically. We discuss this point below.

It is now straightforward to show consistency.

PROPOSITION 2: *Given Assumptions 1 and 2*,

$$\underset{L\to\infty}{\text{plim}}\,\hat{\beta} - \beta_0 \;=\; \underset{L\to\infty}{\text{plim}} \frac{\sum_{t=1}^{T}\sum_{k=1}^{K} g_{kt} L^{-1} \sum_{l=1}^{L} z_{lk0}\,\epsilon_{lt}^{\perp}}{\sum_{t=1}^{T}\sum_{k=1}^{K} g_{kt} L^{-1} \sum_{l=1}^{L} z_{lk0}\,x_{lt}^{\perp}} \;=\; 0.$$

PROOF:

This is a standard GMM consistency result (e.g., Wooldridge 2002, Theorem 8.1).

As a result, the Bartik TSLS IV estimator is consistent.

These results have two implications. First, under our sampling process, strict exogeneity of the industry shares is necessary for the Bartik estimator to be generically consistent. This assumption is standard in many difference-in-differences settings. Second, it highlights that the Bartik estimator uses a particular weighting of these exogeneity conditions; other weightings would imply other estimators.

B. *When Are These Assumptions Plausible?*

The exogenous shares assumption discussed in the last section might seem implausible because shares are equilibrium objects likely codetermined with the level of the outcome of interest. But this reasoning does not reflect the assumption that is typically being made. Instead, the assumption is about exogeneity *conditional on observables*, which typically controls for level differences either by focusing on changes (as in our baseline setup where we define y_{lt} and x_{lt} to be in changes), or else by operating in levels but including unit fixed effects. Hence, in typical specifications, the assumption is that the shares are exogenous to *changes* in the error term (i.e., changes in the outcome variable), rather than *levels* of the outcome variable.

The plausibility of the substantive restrictions implied by this identifying assumption might be more intuitive in a setting with two industries and a differential exposure design, which we discussed in Section I. In this setting, the identifying assumption is that the differential effect of higher exposure of one industry (compared to another) only affects the change in the outcome (y_{lt}) through the endogenous variable of interest, and not through any potential confounding channel. This assumption is standard in difference-in-differences. In the shares view, the identifying assumption underlying the Bartik setting is simply this differential exposure design applied to each industry separately.

This type of identification assumption is natural to make when the shares create differential exposure to a common economic or policy shock (or sets of shocks). In these cases, the most natural description of the identification comes from highlighting a few key industries which best illustrate the exposure design. In Section III, we show how to do this. While natural to make, this type of assumption may not always be satisfied. For example, areas with high versus low exposure may have other features that predict change in the outcome through channels other than the endogenous variable, violating the exclusion restriction.

In cases when the assumption of exogenous shares is not plausible, consistency of the estimator can instead come from many exogenous shocks. As proved in Borusyak, Hull, and Jaravel (2019), exogenous independent shocks to many industries lead the Bartik estimator to be consistent, even when the shares are not exogenous. The core intuition to this result can be seen in equation (5). In cases when the shares are not exogenous, the numerator does not converge to zero. As a result, the weighted sum of the industry shocks and the shares are nonzero. With many exogenous and independent shocks, however, Borusyak, Hull, and Jaravel (2019) shows that the estimator is still consistent. The reason is that the random shocks are uncorrelated with the bias from the shares ($E[\epsilon_{lt} z_{lk0} | D_{lt}]$), and the presence of many shocks causes this bias to average out to zero (see also Kolesár et al. 2015).

How can researchers tell which quasi-experimental design they are using? When a researcher explains her research design (and hence, implicitly, her estimator) using a two-industry example, she is emphasizing differential exogenous exposure, which underlies a research design based on the shares assumption. The reason is that under the shocks view the Bartik estimator is only consistent as the number of industries grows. Hence, the logic of how consistency in this research design works is not captured by the two-industry example. Similarly, if a researcher emphasizes the performance of a particular industry (or a small handful of industries), then this reasoning also suggests that she is appealing to a research design based on the shares. In contrast, when having a large number of industries is central to how the researcher thinks about consistency (and identification), then it is likely that she is building a research design based on the shocks assumption.

While a best case scenario for a researcher using a Bartik instrument is for both the exogenous shares and shocks assumptions to hold, in practice, this coincidence seems unlikely. Typically, a researcher will only have one identification strategy at their disposal. We encourage researchers to pick one or the other, be clear about why, and then defend the relevant assumptions in their setting.

III. Opening the Black Box of the Bartik Estimator

The previous sections showed that under standard panel asymptotics, the Bartik instrument is equivalent to using industry shares as instruments. Thus, the Bartik *estimator* combines many instruments using a specific weight matrix.

Empirical work using a single instrument is transparent because there is a small number of covariances that enter the estimator. With many instruments, it is less intuitive how the estimator combines the different instruments. This lack of intuition underlies much of the empirical work using Bartik instruments, where it is hard to explain what variation in the data drives estimates, and can often feel like a black box.

In this section, we show how to open the black box of the Bartik estimator. First, we decompose the Bartik estimator into a weighted combination of just-identified estimates based on each instrument. This decomposition increases the transparency of the estimator because the weights highlight the industries whose variation in the data drives the overall Bartik estimate. Building on Andrews, Gentzkow, and Shapiro (2017), we show that these weights can be interpreted as sensitivity-to-misspecification elasticities. The Bartik estimate is most sensitive to misspecification in high-weight instruments, and hence these are the instruments that are most important for researchers to defend.

A. *Decomposing the Bartik Estimator*

We first present a finite sample decomposition of the linear overidentified GMM estimator due to Rotemberg (1983).[7] For expositional simplicity, we use a single cross section, though it is straightforward to extend results to a panel with T time periods.

PROPOSITION 3: *We can write*

$$\hat{\beta}_{Bartik} = \sum_k \hat{\alpha}_k \hat{\beta}_k,$$

where

$$\hat{\beta}_k = \left(Z_k' X^{\perp}\right)^{-1} Z_k' Y^{\perp} \quad and \quad \hat{\alpha}_k = \frac{g_k Z_k' X^{\perp}}{\sum_{k'} g_{k'} Z_{k'}' X^{\perp}},$$

so that $\sum_k \hat{\alpha}_k = 1$.

PROOF:

See online Appendix Section C.

Proposition 3 has two implications. First, mirroring our results from Section II, the validity of each just-identified $\hat{\beta}_k$ depends on the exogeneity of a given Z_k. Second, for some k, $\hat{\alpha}_k$ can be negative. Under the constant effects assumption we have maintained so far, these negative weights do not pose a conceptual problem. In Section IV, we introduce a restricted form of treatment effect heterogeneity and revisit the implications of the negative Rotemberg weights.

In online Appendix Section E, we discuss how to interpret the Rotemberg weights in terms of sensitivity-to-misspecification following work by Conley, Hansen, and Rossi (2012) and Andrews, Gentzkow, and Shapiro (2017). The basic intuition is that if any particular instrument is misspecified, then α_k tells us how much that misspecification translates into the overall bias of the estimator. For example, if α_k is small, then bias in the kth instrument does not affect the overall bias in the estimator very much. We also show that this measure is different than simply dropping instruments and seeing how estimates change, since dropping an instrument combines sensitivity-to-misspecification (i.e., α_k) as well as the relative misspecification of different instruments (i.e., how far $\hat{\beta}_k$ diverges from $\hat{\beta}$).

We recommend researchers report the instruments associated with the largest values of α_k for two reasons. First, reporting the instruments with the largest α_k provides a more concrete way to describe the empirical strategy. Second, to the extent that the researcher is concerned about misspecification, these are the instruments that are most worth probing.

[7] Andrews (2019, Section 3.1) reports this decomposition for constant-effect linear instrumental variables.

In our applications, we report the share of the variance in the Rotemberg weights that can be explained by the g_k. The primary reason is that the α_k are nonlinear functions of g_k, x_l, and z_{lk} and so there is not a simple decomposition which shows why the α_k end up with the particular patterns that they do. The share of the variance of the α_k that can be explained by the g_k quantifies the extent to which it is explained by g_k. For similar reasons, we also report the correlation between $\text{var}(z_{lk})$ (across l) and α_k. A secondary reason to focus on the g_k is that there is a common intuition that the variation in the g_k explains the "sources of variation" in the empirical design. Given that the α_k is a formal way of quantifying the "sources of variation," we find it helpful to contrast this intuition with our formal measure.

Similarly, we also relate the Rotemberg weight to the first stage F-statistic. In online Appendix Section F we show that the first-stage F-statistic on the kth instrument is related to Rotemberg weight by the following formula:

$$\frac{\hat{F}_k}{\hat{F}} = \hat{\alpha}_k^2 \left(\frac{\widehat{\text{var}}(B^{\perp})}{g_k \widehat{\text{var}}(Z_k^{\perp})} \right)^2 \frac{\hat{\Sigma}_{\pi\pi}}{\hat{\Sigma}_{\pi_k \pi_k}}, \tag{6}$$

where $\hat{\Sigma}_{\pi_k\pi_k}$ is the estimated sampling variance around the first-stage coefficient on the kth instrument, $\hat{\Sigma}_{\pi\pi}$ is the estimated sampling variance around the first-stage coefficient on the Bartik instrument, and $\hat{F}$ is the first-stage F-statistic when using the Bartik instrument. This equation helps explain when and how the Rotemberg weight differs from the (relative) first-stage F-statistic. If the precision of the first-stage coefficient (third term) is proportional to the variance of the instrument (second term), then the product of the last two terms will be constant across instruments and hence the relative F-statistic will be proportional to the Rotemberg weight. In contrast, when the estimation noise does not vary proportionally with the variance of the instrument (perhaps because the instrument varies, but is not correlated with the endogenous variable), then the Rotemberg weight and relative F-statistic will be less strongly related.

B. *Normalization*

When the industry shares sum to 1 within a location, the instruments are linearly dependent and so we can write each instrument as a function of the remaining $K - 1$ instruments. This fact has a couple implications. First, following Remark 1, we can drop any industry through normalization by subtracting off g_j from all the growth rates, and leave our point estimates unchanged. Second, the fact that we can drop any one industry means that the Rotemberg weights are not invariant to the choice of which industry to drop. To take an extreme example, suppose industry j has the largest weight. Then, by dropping industry j through normalization, a researcher could make industry j have a weight of zero, but the Bartik estimate would remain the same.

To address this issue, in applications where the industry shares sum to 1, we report Rotemberg weights that come from demeaning the (unweighted) industry growth rates. In online Appendix Section G, we show that this normalization is the average of the K possible normalizations of dropping each of the industries.[8]

C. *Aggregation*

Below, we consider applications with panel data and multiple time periods. As a result, the underlying instruments are industry shares interacted with time fixed effects. Rather than reporting results at the level of k, t, we aggregate to the k level. The reason is that it is typically easier to think about the variation coming from a cross-sectional difference, rather than the variation coming from a cross-sectional difference in a particular time period. Formally, we define

$$\alpha_k \equiv \sum_t \alpha_{k,t},$$

and

$$\beta_k \equiv \sum_t \frac{\alpha_{k,t}}{\alpha_k} \beta_{k,t},$$

where the empirical estimator versions are defined analogously.[9] Note that we could analogously aggregate to the time level and define $\alpha_t \equiv \sum_k \alpha_{k,t}$ and $\beta_t \equiv \sum_k \left(\alpha_{k,t}/\alpha_t\right)\beta_{k,t}$.

To interpret such an aggregated α in terms of the underlying misspecification, suppose that $\tilde{\beta}_{kt} = \tilde{\beta}_k$ for all t. Then,

$$\tilde{\beta} = \sum_k \alpha_k \sum_t \frac{\alpha_{kt}}{\alpha_k} \tilde{\beta}_{kt} = \sum_k \alpha_k \tilde{\beta}_k \sum_t \frac{\alpha_{kt}}{\alpha_k} = \sum_k \alpha_k \tilde{\beta}_k.$$

These equations say that the α_k measures the sensitivity-to-misspecification where we assume that the endogeneity associated with the kth industry is constant across time.

IV. Heterogeneous Effects

In previous sections, we showed that the Bartik estimator combines many instruments with a specific weight matrix. A key assumption was that of constant effects. In many contexts, a researcher might prefer to think that there are heterogeneous effects that vary across locations or time. For example, in the canonical labor supply elasticity application that we discuss below, some locations might have more elastic labor supply than others.

In this section, we discuss a heterogeneous effects interpretation of the Bartik instrument. Because the Bartik instrument combines multiple unordered instruments, it is difficult to allow unrestricted heterogeneity of the form discussed in Imbens

[8] In cases when the shares sum to 1, if a researcher suspects that one instrument is invalid, then simply dropping that instrument does not fix the problem. Instead, the researcher would need to drop that instrument and then renormalize the shares to sum to 1.

[9] A numerically identical way of arriving at $\hat{\beta}_k$ is to use $B_{lkt} = z_{lk0} g_{kt}$, the Bartik instrument built from just the kth industry, as the instrument.

and Angrist (1994) and ensure interpretable estimates.[10] Specifically, assuming monotonicity as in Imbens and Angrist (1994) is not sufficient to ensure estimates reflect nonnegative weights on the underlying heterogeneity. For further lucid discussion of these issues, see Kirkeboen, Leuven, and Mogstad (2016), among others. Instead, we impose a restricted form of linear heterogeneity and then state assumptions to ensure interpretable just-identified estimates. We also emphasize that even if each just-identified IV estimate produces a convex combination of heterogeneous effects, the overall Bartik instrument can produce negative weights if there are negative Rotemberg weights.

A. *Setup with Restricted Heterogeneity*

We follow Borusyak, Hull, and Jaravel (2019) and expand our model to include location specific coefficients.[11] Formally, consider the structural model:

$$y_l = D_l \rho + x_l \beta_l + \epsilon_l, \tag{7}$$

where now β_l replaces β_0.[12] For the purposes of the results below, we focus on discrete saturated controls (i.e., dummies) for D_l in order to ignore differences in specification error when residualizing. We also assume the following linear relationship between z_{lk} and x_l:

$$x_l = D_l \tau + z_{lk} \pi_{lk} + u_{lk}, \tag{8}$$

where π_{lk} is the location-industry specific first-stage coefficient and u_{lk} is the location-industry specific error. We assume that β_l is a random variable with well-defined moments.

Relative to Imbens and Angrist (1994), this setup is restricted because it assumes constant linear effects within a location over the whole support of $x_l^{\perp}$. One substantive restriction it imposes is that identically sized shocks have identical effects regardless of the level of employment in the location.

We now impose assumptions which are sufficient to ensure that in this linear model the weights on the β_l are all weakly positive. In this sense, they are analogous to monotonicity assumptions in nonparametric models.

ASSUMPTION 3:

(*i*) *For each k*, π_{lk} *is* (*weakly*) *the same sign for all l.*

(*ii*) $E[z_{lk} u_{lk} \beta_l | D_l] = 0.$

[10] To see why industry shares are unordered instruments, note that increasing the share of an industry can increase the predicted growth rates in some locations and decrease it in others depending on which industry share decreases to offset.

[11] Adao, Kolesár, and Morales (2019) includes location-industry coefficients. For simplicity, we maintain location specific coefficients.

[12] We focus on a single time period, but these points generalize.

We now state the result that the just-identified IV estimates represents a convex combination of the β_l.

PROPOSITION 4: *Suppose that equations* (7) *and* (8) *are true*, *and Assumption* 3 *holds*, *then we can write*

$$\underset{L\to\infty}{plim}\,\hat{\beta}_k \;=\; E[\omega_{lk}\beta_l] \tag{9}$$

where $\omega_{lk} \;=\; \left(z_{lk} - E[z_{lk}|D_l]\right)^2 \pi_{lk}/E\Big[\left(z_{lk} - E(z_{lk}|D_l)\right)^2 \pi_{lk}\Big] \;\geq\; 0$ *and* $E[\omega_{lk}] \;=\; 1$.

PROOF:

See online Appendix Section C.

This result explains why in the presence of heterogeneity using different instruments (i.e., z_{lk}) would generate different point estimates (i.e., $\hat{\beta}_k$) even without misspecification. Each instrument estimates a parameter that is a different weighted combination of location-specific parameters. Because these parameters differ (i.e., there is heterogeneity), different instruments generate different estimates.

B. *The Bartik Estimator with Heterogeneity*

In this heterogeneous effects interpretation of Bartik, we can combine the Rotemberg weights and the ω_{lk} to write the Bartik estimate in terms of the location-specific coefficients:

$$\hat{\beta}_{Bartik} \;=\; \sum_l \beta_l \sum_k \alpha_k \omega_{lk} + o_p(1). \tag{10}$$

When $\sum_k \alpha_k \omega_{lk}$ is nonnegative for all l, the Bartik estimator thus reflects a convex combination of the β_l. When are these weights nonnegative? In the previous section, we discussed assumptions such that the ω_{lk} are nonnegative. These assumptions, however, do not imply that the α_k are all positive. Thus, negative α_k are possible, which raises the possibility (but does not necessarily imply) nonconvex weights on the β_l, in which case the overall Bartik estimate does not have a LATE-like interpretation as a weighted average of treatment effects.

When are negative weights on the β_l likely to arise? We note first that we cannot estimate the ω_{lk} and hence we cannot directly compute the weights on the β_l. We can, however, estimate the α_k and the β_k, and use information in these two estimates to gauge the possibility of negative weights on the β_l.

If the $\hat{\beta}_k$ are all similar, then the negative weights on the k are unlikely to generate negative weights on the β_l. The reason is that the similarity of the β_k suggests that the ω_{lk} are similar across k, so that each instrument is likely estimating a similar weighted combination of effects. Hence, the negative α_k are likely just subtracting off the same β_l, with the overall weight on each β_l remaining positive.

In contrast, if the β_k are very different, then the ω_{lk} are different across k and each instrument is estimating a different weighted combination of effects. It is then more likely that there are negative weights on the β_l, as the negative α_k place weight on β_l that do not receive positive weight from other instruments. A way to assess the quantitative importance of these negative weights is to split the instruments into those with positive and negative α_k and compare their weighted sums; i.e., to compare $\sum_{k|\alpha_k>0}\hat{\alpha}_k\hat{\beta}_k$ and $\sum_{k|\alpha_k<0}\hat{\alpha}_k\hat{\beta}_k$. If the weighted sum of the instruments with the negative α_k is relatively large, then it is more likely that there are negative weights on the β_l that are important in the overall estimate.

V. Testing the Plausibility of the Identifying Assumptions

The identifying assumptions necessary for consistency are typically not directly testable. However, it is possible to partially assess their plausibility. We focus on the assumptions from Section II; in the context of the canonical setting of estimating the inverse elasticity of labor supply, the identifying assumption is that initial industry composition (Z_{l0}) does not predict innovations to labor supply (ϵ_{lt}).

A. *Empirical Test 1: Correlates of Industry Composition*

It is helpful to explore the relationship between industry composition and location characteristics that may be correlated with innovations to supply shocks. This relationship provides an empirical description of the variation and the types of mechanisms that may be problematic for the exclusion restriction. In particular, the key question researchers should have in mind is whether the correlates of the levels of the shares predict *changes* in the outcome. For the empirical strategy to be valid, it is fine if the level of the correlates are related to the level of the outcome.

Since convention suggests fixing industry shares to an initial time period (Z_{l0}), we recommend considering the correlation with initial period characteristics, as this reflects the instruments' cross-sectional variation. This exercise is instructive for two reasons. First, the correlation in levels helps describe the cross-sectional variation the researcher is using, and so makes the variation more concrete. Second, if Z_{l0} is correlated with factors that predict changes (and not just levels), then this finding hints at omitted variables biasing estimation. Naturally, it is always possible to control for observable confounders, but following the logic of Altonji, Elder, and Taber (2005) and Oster (2019), movements in point estimates when conditioning on observable confounders suggest the potential importance of unobserved confounders. Looking at industries with the largest Rotemberg weights focuses attention on the instruments where confounding variables are most problematic.

One set of controls worth considering is shares at coarser levels. Intuitively, if the industry shares are at the 4-digit level, then it might be that places with different 2-digit compositions are on different trends (i.e., places with more and less manufacturing) and so the shares would not look like valid instruments. The variation in composition within each 2-digit sector (i.e., types of manufacturing) might generate comparisons of places that look more similar in trends.

B. *Empirical Test 2: Pretrends*

In some applications, there is a policy change in period s_0. As we discussed in Section IB, a researcher can use this sharp policy change to implement a difference-in-differences research design. The analogy to difference-in-differences is most straightforward when the shares are fixed over time. In this case, the industry shares measure the exposure to the policy change, while the national growth rates proxy for the size of the policy change.[13] In these settings, it is natural to test for pretrends. We recommend looking at pretrends in terms of the instruments with the largest Rotemberg weights, as well as looking at pretrends in terms of the overall Bartik instrument. We suspect that researchers will be more comfortable with the plausibility of their empirical design if parallel pretrends are satisfied for the instruments to which their estimates are most sensitive to misspecification.

This test would typically use a measure of industry shares that is fixed in time, prior to the policy change. Analogous to industry shares, it also makes sense to measure controls in the same time period as the industry shares, and interact these time-invariant controls with time fixed effects. The reason to fix controls is that using controls measured after the policy change can biasing estimates by controlling for an intermediate outcome affected by the policy change. For more details on pretrends tests, see DiNardo and Lee (2011). We additionally present examples below.

C. *Empirical Test 3: Alternative Estimators*, *Overidentification Tests*, *and Patterns of Heterogeneity*

So far, we have emphasized that the Bartik estimator combines many moment conditions with a particular weight matrix. In this section, we discuss how researchers can use these moment conditions. Broadly speaking, there are two directions that a researcher can go. Under homogeneous effects, researchers can consider alternative estimators that combine the moment conditions in potentially more efficient ways. Additionally, researchers can use overidentification tests. If alternative estimators yield different estimates and overidentification tests reject, then these findings point to misspecification. In contrast, under heterogeneous effects, each instrument will converge to a different estimate (say, β_k) as discussed in Section IV. Under this assumption, it is important that the patterns of heterogeneity make sense, and we discuss some ways of assessing this.

Homogeneous Effects.—We begin in a world of homogeneous effects. Because the overidentified TSLS estimator (i.e., the one using each industry share as a separate instrument) is biased in finite samples, we encourage researchers to use three alternative estimators which have better properties with many instruments: the Modified Bias-corrected TSLS (MBTSLS) estimator from Anatolyev (2013) and Kolesár et al. (2015), the Limited Information Maximum Likelihood (LIML) estimator, and the HFUL estimator from Hausman et al. (2012). These estimators may not give the same estimates, as their underlying assumptions are different.[14]

[13] Some examples of this include Autor, Dorn, and Hanson (2013) and Lucca, Nadauld, and Shen (2019).

[14] The LIML estimator, as discussed in Hausman et al. (2012), is inconsistent under heteroskedasticity and many instruments. The HFUL estimator is consistent under both heteroskedasticity and many instrument asymptotics, while the literature on MBTSLS has not developed yet under heteroskedasticity. Inference under clustering asymptotics has not, to our knowledge, been worked out for any of these estimators under many instrument asymptotic settings.

Comparing these estimates, along with the Bartik TSLS estimate, provides a useful first pass diagnostic for misspecification concerns. If these estimators agree, then researchers can be more confident in their identifying assumption. In our applications, we follow Kolesár et al. (2015) and interpret differences between HFUL and LIML on the one hand, and MBTSLS and overidentified TSLS on the other, as pointing in the direction of potential misspecification. The reason is that LIML and HFUL are maximum likelihood estimators and so exploit cross-equation restrictions while both MBTSLS and overidentified TSLS are two-step estimators and so do not exploit these cross-equation restrictions.

Overidentification tests provide more formal tests for misspecification. These estimators permit test statistics under different assumptions. For the HFUL estimator, we suggest the overidentification test from Chao et al. (2014); for LIML estimator, we use the Anderson and Rubin (1950) χ^2-test; and for overidentified TSLS, we use the Sargan (1958) χ^2-test. Again, these tests may not give the same results, as their underlying assumptions are different.[15] Conceptually, the overidentification test asks whether the instruments are correlated with the error term beyond what would be expected by chance, and relies on the validity of at least one of the instruments.

Heterogeneous Effects.—When overidentification tests reject, and when HFUL and LIML differ from MBTSLS and Bartik TSLS, under homogeneous effects these findings point to misspecification. An alternative interpretation of these results is that they point to heterogeneous effects of the form we outlined in Section IV. Under these assumptions, researchers may wish to probe the patterns of heterogeneity and see if there is a reasonable interpretation.

We now outline a visual diagnostic to help researchers assess the pattern of heterogeneity. The fundamental feature of the data that illustrates the heterogeneity is to consider the distribution of the just identified IV estimates (i.e., the $\hat{\beta}_k$). In order to visualize this dispersion, we advocate a particular figure. Here we describe the figure and discuss our reasoning, and below we present examples of it (see Figures 1 and 4 and online Appendix Figure A2).[16] Briefly, the x-axis is the first-stage F-statistic and the y-axis is the $\hat{\beta}_k$ associated with each instrument. So as to not visually overstate dispersion, the figure only includes instruments with reasonable first-stage power (in our applications, we plot instruments with first-stage F-statistics greater than 5). To show how the $\hat{\beta}_k$ compare to the Bartik estimate, the figure includes a horizontal line that reflects the overall Bartik estimate. Because first-stage power does not perfectly explain the Rotemberg weights, we weight the individual points of β_k by the absolute size of the α_k from the Bartik Rotemberg weights. Finally, to illustrate the role of negative Rotemberg weights, we shade the points differently depending on the sign of the Rotemberg weights.

[15] Specifically, the Anderson-Rubin and Sargan tests are only valid under homoskedasticity, which is likely not satisfied in this setting. The HFUL overidentification test does require the assumption of homoskedasticity, but is not solved for the general clustering setting. Code to implement the HFUL overid test is available on request and is posted at https://github.com/paulgp/gpss_replication.

[16] Code to create this figure is included in the package that computes the Rotemberg weights and is posted on GitHub and the replication archive.

Researchers can use this figure to think about three questions. First, why do the overidentification tests reject, and what industries drive the rejection? Intuitively, a researcher might be less concerned by a rejection where the β_k are less rather than more dispersed around the Bartik estimate. Similarly, the figure helps isolate which industries are driving the failure of overidentification tests. Researchers should feel comfortable with why the comparisons implied by some instruments are outliers relative to the comparisons implied by other instruments. Second, why does the Bartik estimate end up where it does relative to the underlying β_k? The relative Rotemberg weights help explain why the Bartik estimate lies where it does relative to the underlying distribution. As we emphasized in Section III, a researcher should feel comfortable that the largest Rotemberg weight industries make sense with the causal mechanism in the paper. Third, how plausible is it that there are negative weights on some β_l? Visualizing the industries with the negative Rotemberg weights helps to highlight which industries would potentially generate negative weights on β_l, as we discussed further in Section IV. Naturally, whether the patterns of heterogeneity make sense will rely on application-specific knowledge, and so we view this figure as providing a useful starting point for an application-specific investigation, rather than an ending point.

A Comment on Alternative Approaches to Overidentifying Tests.—An alternative approach to overidentification tests (e.g., by Beaudry, Green, and Sand 2012 and others) is to construct multiple Bartik instruments using different vectors of national growth rates, and then to test whether these different weighted combinations of instruments estimate the same parameter. Often, the correlation between the Bartik instruments constructed with different growth rates is quite low. This fact is interpreted as reassuring because it suggests that exploiting “different sources of variation” gives the same answer.

We recommend instead that researchers use the Rotemberg weights to quantify what variation each Bartik instrument is using, and whether the two Bartik instruments use different sources of variation. Specifically, researchers can report the top-5 Rotemberg weights across the two instruments and also their rank correlation. If these statistics are low, then the two Bartik instruments are likely using different sources of variation and the conclusion discussed above is warranted.[17]

VI. Empirical Example I: Canonical Setting

We now present two empirical examples to make our theoretical ideas concrete, focusing on our empirical tests from Section V (online Appendix Section A presents a third empirical example). Our first example is the canonical setting of estimating the inverse elasticity of labor supply. We begin by reporting the main estimates and

[17] To illustrate the theoretical distinction between looking at correlations between Bartik instruments and comparing Rotemberg weights implied by the two instruments, in online Appendix Section H we produce an example where only one industry has identifying power, but the two instruments are uncorrelated and find the same $\hat{\beta}$. While this example might seem like a theoretical curiosity, in our empirical settings we typically find that a small number of industries provide most of the identifying variation and the variation in the growth rates explains little of the variation in the Rotemberg weights. Hence, there is typically scope for different national growth rates that produce weakly correlated Bartik instruments to rely on the same “identifying variation” (that is, have similar Rotemberg weights).

then report the industries with the highest Rotemberg weight. We then probe the plausibility of the identifying assumption for these instruments.

A. *Dataset and Specification*

We use the 5 percent sample of IPUMS of US Census Data (Ruggles et al. 2015) for 1980, 1990, and 2000 and we pool the 2009–2011 ACSs for 2010. We look at continental US commuting zones (Autor and Dorn 2013) and 3-digit IND1990 industries.[18] In the notation given above, our y variable is earnings growth, and x is employment growth. We use people aged 18 and older who report usually working at least 30 hours per week in the previous year. We fix industry shares at the 1980 values, and then construct the Bartik instrument using 1980 to 1990, 1990 to 2000, and 2000 to 2010 leave-one-out growth rates. To construct the industry growth rates, we weight by employment. We weight all regressions by 1980 population.

We use the leave-one-out means to construct the national growth rates to address the finite sample bias that comes from using own-observation information. Specifically, using own-observation information allows the first stage to load on the idiosyncratic industry-location component of the growth rate, $\tilde{g}_{lk}$, which is endogenous. This finite sample bias is generic to overidentified instrumental variable estimators and is the motivation for jackknife instrument variable estimators (e.g., Angrist, Imbens, and Krueger 1999). In practice, because we have 722 locations, using leave-one-out to estimate the national growth rates matters little in point estimates (compare rows 2 and 3 in Table 3).[19]

B. *Form of Endogeneity that the Instrument Addresses*

OLS is biased but the Bartik instrument is valid when the idiosyncratic industry-location components of growth are correlated with the error term. In this setting, an amenity shock is an example because it would jointly draw people into a location (increasing employment growth in each industry beyond the national average) and directly affect wage growth (i.e., it appears in the error term in the wage equation).

C. *Rotemberg Weights*

We compute the Rotemberg weights of the Bartik estimator with controls, aggregated across time periods. The distribution of sensitivity is skewed, so that a small number of instruments have a large share of the weight. Table 1 shows that the top five instruments account for over 40 percent (0.587/1.368) of the positive weight

[18] There are 228 nonmissing 3-digit IND1990 industries in 1980. There are 722 continental US commuting zones.

[19] In online Appendix Section I, we show that with a leave-one-out estimator of the g_k component, the Rotemberg weights do not sum to 1. In our applications below, when we compute the Rotemberg weights we use simple averages so that the weights sum to 1.

Table 1—Summary of Rotemberg Weights: Canonical Setting

	Sum	Mean	Share		
Panel A. Negative and positive weights					
Negative	−0.368	−0.004	0.212		
Positive	1.368	0.010	0.788		
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	$\hat{F}_k$	$\text{var}(z_k)$
Panel B. Correlations					
$\hat{\alpha}_k$	1				
g_k	−0.015	1			
$\hat{\beta}_k$	0.017	−0.495	1		
$\hat{F}_k$	0.476	−0.032	0.016	1	
$\text{var}(z_k)$	0.549	−0.036	−0.003	0.316	1
	Sum	Mean			
Panel C. Variation across years in $\hat{\alpha}_k$					
1980	0.458	0.002			
1990	0.182	0.001			
2000	0.360	0.002			
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 percent CI	Ind. Share
Panel D. Top five Rotemberg weight industries					
Oil + gas extraction	0.229	0.034	1.170	(0.80, 1.90)	0.568
Motor vehicles	0.140	−0.017	1.525	(1.30, 1.90)	1.404
Other	0.091	−0.062	0.759	(0.10, 1.70)	1.697
Guided missiles	0.069	0.047	0.115	(−2.20, 0.70)	0.236
Blast furnaces	0.058	−0.078	1.084	(0.60, 5.10)	0.800
	α-weighted sum	Share of overall β	Mean		
Panel E. Estimates of β_k *for positive and negative weights*					
Negative	−0.074	−0.061	1.622		
Positive	1.290	1.061	−0.584		

Notes: This table reports statistics about the Rotemberg weights. In all cases, we report statistics about the aggregated weights with normalized growth rates, where we aggregate a given industry across years as discussed in Section IIIC and normalize growth rates to the per-period average as discussed in Section IIIB. Panel A reports the share and sum of negative weights. Panel B reports correlations between the weights $(\hat{\alpha}_k)$, the national component of growth (g_k), the just-identified coefficient estimates $(\hat{\beta}_k)$, the first-stage F-statistic of the industry share $(\hat{F}_k)$, and the variation in the industry shares across locations $(\text{var}(z_k))$. Panel C reports variation in the weights across years. Panel D reports the top five industries according to the Rotemberg weights. The g_k is the national industry growth rate, $\hat{\beta}_k$ is the coefficient from the just-identified regression, the 95 percent confidence interval is the weak instrument robust confidence interval using the method from Chernozhukhov and Hansen (2008) over a range from −10 to 10, and *Ind. Share* is the industry share (multiplied by 100 for legibility). Panel E reports statistics about how the values of $\hat{\beta}_k$ vary with the positive and negative Rotemberg weights. The Other industry is the N/A code in the IND1990 classification system and includes full-time military personnel.

in the estimator.[20] These top five instruments are: oil and gas extraction, motor vehicles, other,[21] guided missiles, and blast furnaces.

These weights give a way of describing the research design that reflects the variation in the data that the estimator is using, and hence makes concrete for the reader what types of deviations from the identifying assumption are likely to be important. In this canonical setting, one of the important comparisons is across places with greater and smaller shares of oil and gas extraction. Hence, the estimate is very sensitive to deviations from the identifying assumption related to geographic variation in employment share in oil and gas extraction. Interestingly, a common short-hand to talk about Bartik is to discuss the fate of the automobile industry (e.g., Bound and Holzer 2000), and this analysis confirms that the motor vehicle industry plays a large role in the Bartik instrument.

Finally, panel B shows that the national growth rates are weakly correlated with the sensitivity-to-misspecification elasticities. Hence, the growth rates provide a poor guide to understanding what variation in the data drives estimates. In contrast, the elasticities are quite related to the variation in the industry shares across locations $(\text{var}(z_{lk}))$. This observation explains why the industries with high weight tend to be tradables: almost by definition, tradables have industry shares that vary across locations, while nontradables do not.[22]

D. *Discussion of the Identifying Assumption in Terms of the Shares*

As we discussed in Section II, a heuristic for figuring out which identifying assumption researchers have in mind is whether they mention particular industries. It is common in the canonical setting to discuss particular industries (e.g., as mentioned above, Bound and Holzer 2000 discusses the automobile industry). Hence, we think that in many settings researchers have in mind this differential exposure design.

E. *Testing the Plausibility of the Identifying Assumption*

Test 1: Correlates of 1980 Industry Shares.—Table 2 shows the relationship between 1980 characteristics of commuting zones and the share of the top 5 industries in Table 1, as well the overall Bartik instrument using 1980 to 1990 growth rates. First, the R^2 in these regressions are quite high: for example, we can explain 46 percent of the variation in share of the Other industry via our covariates. Second, Other, oil and gas extraction, blast furnaces, and the overall Bartik instrument are statistically significantly correlated with the share of native-born workers. The complement of the native-born share is the immigrant share. In the immigrant enclave literature, and under the shares interpretation, the

[20] The calculation is to sum the five $\hat{\alpha}_k$: $0.229 + 0.14 + 0.091 + 0.069 + 0.058 = 0.587$ and divide by the total positive weight 1.368.

[21] The Other industry is the N/A code in the IND1990 classification system. Our understanding is that in 1980 the Other code includes full-time military personnel. Hence, in 1990 and 2000, we place full-time military personnel in the Other category to compute growth rates.

[22] This logic is the basis of Jensen and Kletzer's (2005) measure of the offshorability of services; as Jensen and Kletzer (2005) recognizes, there are other reasons for concentration besides tradability.

Table 2—Relationship between Industry Shares and Characteristics: Canonical Setting

	Oil and gas extraction	Motor vehicles	Other	Guided missiles	Blast furnaces	Bartik (1980 shares)
Male	1.319	−0.501	4.076	0.126	0.344	−0.178
	(0.242)	(0.160)	(0.600)	(0.063)	(0.159)	(0.035)
White	0.043	−0.714	−1.310	0.057	−0.681	−0.088
	(0.102)	(0.653)	(0.281)	(0.043)	(0.256)	(0.029)
Native born	0.364	−0.129	0.824	−0.157	−0.312	−0.172
	(0.092)	(0.110)	(0.281)	(0.133)	(0.129)	(0.019)
12th grade only	−1.096	1.283	1.040	−0.193	0.202	0.036
	(0.218)	(0.392)	(0.356)	(0.091)	(0.150)	(0.030)
Some college	−0.311	0.687	1.060	0.033	−0.808	0.376
	(0.143)	(0.520)	(0.288)	(0.072)	(0.254)	(0.042)
Veteran	−0.295	0.895	−5.793	0.202	2.526	0.000
	(0.227)	(0.917)	(0.879)	(0.126)	(0.714)	(0.072)
Number of children	−0.043	0.954	−2.409	−0.006	0.003	−0.070
	(0.142)	(0.538)	(0.558)	(0.047)	(0.223)	(0.034)
R^2	0.24	0.08	0.46	0.27	0.23	0.77
Observations	722	722	722	722	722	722

Notes: Each column reports results of a single regression of a 1980 industry share on 1980 characteristics. The final column is the Bartik instrument constructed using the growth rates from 1980 to 1990. Results are weighted by 1980 population. Standard errors in parentheses. The Other industry is the N/A code in the IND1990 classification system and includes full-time military personnel.

immigrant share is used to predict inflows of immigrants, which are interpreted as labor supply shocks. Hence, an industry share which is interpreted as predicting labor demand shocks is correlated with something that also predicts labor supply shocks.

Test 2: Parallel Pretrends.—We note that in this setting there is no pre-period and so it is not possible to test for parallel pretrends without further assumptions.

Test 3: Alternative Estimators and Overidentification Tests.—Rows 1, 2, and 3 of Table 3 report the OLS and IV estimates (row 2 leaves out the own-CZ growth rate to construct the instrument, while row 3 uses all CZs to construct the growth rates), with and without for the 1980 covariates as controls and makes two main points. First, the IV estimates are bigger than the OLS estimates. Second, the Bartik results are sensitive to the inclusion of controls, though these are not statistically distinguishable.

Rows 4–7 of Table 3 report alternative estimators as well as overidentification tests. We focus on column 2, where we control for covariates. TSLS with the Bartik instrument and LIML are quite similar. This finding is typically viewed as reassuring. In contrast, overidentified TSLS and MBTSLS are similar, while HFUL is substantially larger. The different point estimates suggest the presence of misspecification. In column 4, we see that the overidentification tests reject the null that all instruments are exogenous, which also points to misspecification.

Table 3—OLS and IV Estimates: Canonical Setting

	Δ Emp		Coefficient equal	Over ID test
	(1)	(2)	(3)	(4)
OLS	0.71	0.63	[0.04]	
	(0.06)	(0.07)		
TSLS (leave-out Bartik)	1.76	1.28	[0.23]	
	(0.33)	(0.42)		
TSLS (Bartik)	1.65	1.22	[0.19]	
	(0.34)	(0.15)		
TSLS	0.74	0.67	[0.10]	1,014.05
	(0.05)	(0.07)		[0.00]
MBTSLS	0.76	0.69	[0.13]	
	(0.06)	(0.07)		
LIML	1.60	1.42	[0.76]	2,820.96
	(0.00)	(0.57)		[0.00]
HFUL	2.85	2.69	[0.00]	804.19
	(0.14)	(0.13)		[0.00]
Year and CZone FE	Yes	Yes		
Controls	No	Yes		
Observations	2,166	2,166		

Notes: This table reports a variety of estimates of the inverse elasticity of labor supply. The regressions are at the commuting zone level and the instruments are 3-digit industry-time periods (1980–1990, 1990–2000, and 2000–2010). Column 1 does not contain controls, while column 2 does. The *TSLS* (*Bartik*) row uses the Bartik instrument. The *TSLS* row uses each industry share (times time period) separately as instruments. The *MBTSLS* row uses the estimator of Anatolyev (2013) and Kolesár et al. (2015) with the same set of instruments. The *LIML* row shows estimates using the limited information maximum likelihood estimator with the same set of instruments. Finally, the *HFUL* row uses the HFUL estimator of Hausman et al. (2012) with the same set of instruments. The *J*-statistic for HFUL comes from Chao et al. (2014). The *p*-value for the equality of coefficients compares the adjacent columns with and without controls. The controls are the 1980 characteristics (interacted with time) displayed in Table 2. Results are weighted by 1980 population. Standard errors are in parentheses and are constructed by bootstrap over commuting zones. *p*-values are in brackets.

Visualizing the Overidentification Tests.—If one wishes to interpret the failure of the overidentification tests as pointing to heterogeneity of the form outlined in Section IV rather than as evidence of misspecification, then Figure 1 shows some of the heterogeneity in treatment effects underlying the overall Bartik estimate (online Appendix Figure A5 shows the relationship between the Rotemberg weights and the first-stage F-statistic). First, the figure shows that among the "high-powered" (i.e., those with a first-stage F-statistic above 5) industries, there is substantial dispersion around the Bartik $\hat{\beta}$. Second, the largest weight industries do tend to be closest to the overall Bartik $\hat{\beta}$. Third, if a researcher wishes to adopt a heterogeneous effects interpretation of the rejection of the null in the overidentification tests, then the patterns of heterogeneity suggest that there are likely to be negative weights on some of the underlying location-specific coefficients. In particular, there is substantial dispersion in the $\hat{\beta}_k$ and some of the outlier $\hat{\beta}_k$ have negative weights. Thus, the underlying location-specific effects (the β_l) that lead to a negative coefficient likely receive negative weights so that the overall Bartik estimate does not reflect convex weights. To see this more generally, panel E of Table 1 shows that the mean of

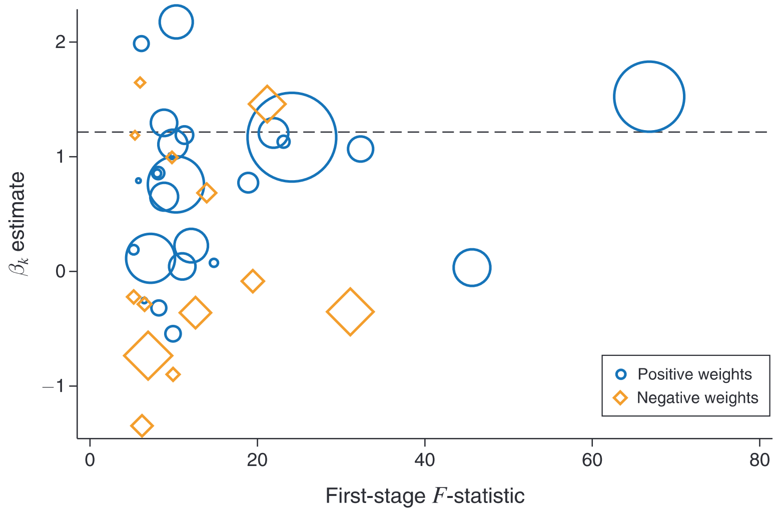

Figure 1. Heterogeneity of β_k: Canonical Setting

Notes: This figure plots the relationship between each instruments' $\hat{\beta}_k$, first-stage F-statistics, and the Rotemberg weights. Each point is a separate instrument's estimates (industry share). The figure plots the estimated $\hat{\beta}_k$ for each instrument on the y-axis and the estimated first-stage F-statistic on the x-axis. The size of the points are scaled by the magnitude of the Rotemberg weights, with the circles denoting positive Rotemberg weights and the diamonds denoting negative weights. The horizontal dashed line is plotted at the value of the overall $\hat{\beta}$ reported in the second column in the *TSLS* (*Bartik*) row in Table 3. The figure excludes instruments with first-stage F-statistics below 5.

the β_k among the negative weight industries is very different than the mean of the β_k among the industries with positive weights.

VII. Empirical Example II: Immigrant Enclave

For our second empirical example, we estimate the (negative) inverse elasticity of substitution between immigrants and natives following Card (2009). In particular, we focus on the results in Table 6 of that paper (in particular columns 3 and 7), which provides two sets of results: one for high-school equivalent workers, and one for college equivalent workers.

A. *Dataset and Specification*

We use the 5 percent sample of US Census data for 1980, 1990, and 2000 and following Card (2009) use the ICPSR version (US Census Bureau 2000).[23] It is helpful to convert Card's (2009) specification into our notation. The paper is interested in a regression:

$$y_{lj} = \beta_0 + \beta \ln x_{lj} + \beta_2 \mathbf{X}_l + \epsilon_{lj}, \tag{11}$$

[23] To build the dataset, we use code provided by Card (2019).

where l is a location (a city) and j is a skill group (either high school or college equivalent). Here, y_{lj} is the residual log wage gap between immigrant and native men in skill group j, x_{lj} is the ratio of immigrant to native hours in skill group j (of both men and women), and $\mathbf{X}_l$ is a vector of city-level controls. Hence, β is the (negative) inverse elasticity of substitution between immigrants and natives in the relevant skill group. Unlike other examples, the controls do not include place and time fixed effects because the paper considers a single cross section of outcomes in 2000 in 124 cities. The paper does, however, explore robustness to including the lagged dependent variable.

The first stage is

$$\ln x_{lj} \;=\; \gamma_0 + \gamma_1 B_{lj} + \gamma_2 \mathbf{X}_l + \eta_l\,, \tag{12}$$

where $B_{lj} = \sum_k z_{lk,1980}\, g_{kj}$. Here, $z_{lk,1980} = \left(N_{lk,1980}/N_{k,1980}\right) \times \left(1/P_{l,2000}\right)$, where $N_{k,1980}$ is the number of immigrants from 1 of 38 country (groups) k in the United States in 1980, $N_{lk,1980}$ is the number of immigrants from country (group) k in location l in 1980, and $P_{l,2000}$ is the population of location l in 2000. Here, g_{kj} is the number of people arriving in the United States from 1990 to 2000 from country (group) k and skill group j. Notice that the shares, the immigrant enclave, are not skill-specific, while the shocks, the immigrant inflows, are skill-specific. Relative to our other examples, the shares do not sum to 1 within a location.

B. *Form of Endogeneity that the Instrument Addresses*

A form of endogeneity that the instrument addresses is a positive labor demand shock that draws immigrants into a location disproportionately relative to natives: a positive labor demand shock to immigrants will increase ϵ_{lj} (relative earnings) as well as x_{lk} (relative supply).

C. *Rotemberg Weights*

In this setting, there are 38 country groups. For high school equivalent workers, panel A of Table 4 shows that the top country is Mexico, which by itself receives almost one-half of the weight, and the top five countries (in order: Mexico, El Salvador, Philippines, China, and country group of West Europe, Israel, Cyprus, Australia, and New Zealand) get almost two-thirds of the overall weight. The large weight on Mexico is perhaps unsurprising. Indeed, Card (2009) emphasizes that one might be concerned that for high-school equivalent workers the instrument is largely just initial Mexican immigrant shares. Unlike in the other examples, all the weights are positive. One reason the weights accord so closely with intuition is that for this instrument the weights are almost perfectly explained by the shocks, the immigrant inflows. Panel AII shows that the correlation between the weights and the g_k is 0.991, which is dramatically higher than in the other examples.

For college equivalent workers, panel B of Table 4 shows that the top five sending countries receive almost one-half (45 percent) of the weight and all the weights are

Table 4—Summary of Rotemberg Weights: Immigrant Enclave

Panel A. High school equivalent	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	$\hat{F}_k$	$\text{var}(z_k)$
I. Correlations					
$\hat{\alpha}_k$	1				
g_k	0.991	1			
$\hat{\beta}_k$	0.169	0.164	1		
$\hat{F}_k$	0.203	0.173	0.181	1	
$\text{var}(z_k)$	0.043	−0.032	−0.106	−0.260	1
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 percent CI	
II. Top five Rotemberg weight origin countries					
Mexico	0.482	4.95e+06	−0.026	(−0.040, 0.000)	
El Salvador	0.054	4.65e+05	−0.046	(−0.070, −0.030)	
Philippines	0.050	5.31e+05	−0.023	(−0.040, 0.130)	
China	0.038	4.28e+05	−0.041	(−0.070, −0.010)	
West Europe and others	0.031	6.41e+05	−0.067	(−0.110, −0.050)	

Panel B. College equivalent	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	$\hat{F}_k$	$\text{var}(z_k)$
I. Correlations					
$\hat{\alpha}_k$	1				
g_k	0.766	1			
$\hat{\beta}_k$	0.293	0.255	1		
$\hat{F}_k$	−0.028	−0.055	0.230	1	
$\text{var}(z_k)$	0.033	−0.381	−0.075	−0.225	1
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 percent CI	
II. Top five Rotemberg weight origin countries					
Philippines	0.151	6.32e+05	−0.065	(−0.125, −0.040)	
Mexico	0.102	5.44e+05	−0.062	(−0.095, 0.000)	
China	0.082	3.74e+05	−0.084	(−0.125, −0.060)	
West Europe and others	0.066	5.31e+05	−0.090	(−0.145, −0.065)	
Cuba	0.049	1.86e+05	−0.008	(−0.045, 0.500)	

Notes: This table reports statistics about the Rotemberg weights, which are all positive in this application. Panels AI and BI report correlations between the weights ($\hat{\alpha}_k$), the national component of growth (g_k), the just-identified coefficient estimates ($\hat{\beta}_k$), the first-stage F-statistics ($\hat{F}_k$), and the variation in the origin country shares across locations ($\text{var}(z_k)$). Panels AII and BII report the top five origin countries according to the Rotemberg weights. The *Others* are Australia, Cyprus, Israel, and New Zealand. The g_k is the number of immigrants from 1990 to 2000, $\hat{\beta}_k$ is the coefficient from the just-identified regression, the 95 percent confidence interval is the weak instrument robust confidence interval using the method from Chernozhukhov and Hansen (2008) over a range from −10 to 10.

positive. The top five countries are similar to the high-school equivalent workers, with El Salvador replaced by Cuba. The top country is the Philippines, with 15 percent of the weight. Relative to our other examples, the shocks have much more explanatory power for the weights (the shocks explain about 60 percent ($= 0.766^2$) of the weights), though this explanatory power is lower than for the high-school equivalent workers.

D. *Discussion of the Identifying Assumption in Terms of the Shares*

We think that it is typically reasonable to interpret the immigrant enclave setting as having an identifying assumption in terms of the shares. The Card (2009) setting considers a single cross section but emphasizes the analogy to difference-in-differences by showing robustness to controlling for the lagged dependent variable so that the effect of the instrument is similar to changes. More broadly, a natural way to think of the immigrant enclave instrument is that in any period there are immigrants arriving from different countries and this then naturally affects places differently. For example, even though in Card (1990) the boatlift was not caused by trends in Miami, the shock only hits Miami because of the strong "pull" factor of the immigrant enclave and the discussion of identification is thus about whether Miami would counterfactually have evolved similarly to places without an existing stock of Cuban immigrants. We view it is as reasonable to interpret the immigrant enclave instrument, especially when applied to a particular time period, as pooling this logic. Hence, a researcher should explain and defend why places with different initial stocks of immigrants would have counterfactually evolved in a similar way.

If a researcher does not feel comfortable embracing the shares view, then it is important to understand what the shocks view means in this setting. To embrace the shocks view of identification in the immigrant enclave setting requires not only that there are random "push" factors, but also that there are enough independent push factors that the endogeneity of the shares averages out. Making this case typically requires a large number of independent "push" factors.

E. *Testing the Plausibility of the Identifying Assumption*

Test 1: Correlates of 1980 Origin Country Shares.—Table 5 shows the relationship between the 1980 covariates used in Card (2009) and the top origin countries reported in Table 4. First, similar to the canonical setting, the characteristics explain a fair amount of the cross-sectional variation in the shares, especially for the overall instrument. Second, and related to the canonical setting, we tend not to find a significant relationship between manufacturing share and any of the individual country shares or the aggregate instruments (the only exception is *West Europe and others*).

Test 2: Parallel Pretrends.—We construct our pretrend figures using the reduced-form regression of equations (11) and (12) with their 1980, 1990, and 2000 values (that is, we include all the controls in Card 2009 in Table 6, columns 3 and 7 and re-estimate year-by-year).[24] Hence, the 2000 coefficient corresponds to the reduced-form coefficient estimated in Table 6.

Figure 2 shows that for the high school equivalent native-immigrant wage gap, the variation in 1980 shares of Mexican immigrants did not predict statistically or economically larger wage gaps in 1980 or 1990. That is, conditional on controls, the figures suggest that there was a shock in the 1990s that led to a widening gap

[24] Because of the structure of the data and the specification in Card (2009), it is not feasible to fix controls in each time period as we discuss in Section VB.

Table 5—Relationship between Origin Country Shares and Characteristics: Immigrant Enclave

	Mexico	Philippines	El Salvador	China	Cuba	West Europe and others	Bartik high school	Bartik college
City size	0.054	0.026	0.106	0.057	0.049	0.039	0.059	0.023
	(0.018)	(0.021)	(0.027)	(0.019)	(0.060)	(0.009)	(0.009)	(0.004)
College share	−0.545	0.559	0.692	1.318	−0.828	0.530	−0.021	0.157
	(0.370)	(0.416)	(0.554)	(0.389)	(1.206)	(0.175)	(0.189)	(0.072)
Mean wage residuals for all natives	0.601	−0.428	0.595	−0.212	−0.199	0.052	0.267	0.041
	(0.388)	(0.437)	(0.582)	(0.408)	(1.266)	(0.184)	(0.199)	(0.076)
Mean wage residuals for all immigrants	−0.652	0.596	−0.856	0.152	−0.079	−0.209	−0.385	−0.061
	(0.361)	(0.406)	(0.540)	(0.379)	(1.175)	(0.170)	(0.185)	(0.070)
Mfg. share	0.059	−0.379	0.268	−0.192	−0.653	0.230	−0.006	−0.010
	(0.202)	(0.228)	(0.303)	(0.213)	(0.660)	(0.096)	(0.104)	(0.039)
Observations	124	124	124	124	124	124	124	124
R^2	0.150	0.095	0.216	0.246	0.020	0.294	0.371	0.430

Notes: Each column reports results of a single regression of a 1980 origin country share on 1980 characteristics. Results are weighted by 1990 population. Standard errors in parentheses. For legibility, coefficients and standard errors of the first six columns are multiplied by 10,000,000. Coefficients and standard errors of the last two columns are not scaled. The *Others* are Australia, Cyprus, Israel, and New Zealand.

in 2000. Given the large weight on Mexico, it is not surprising that the aggregate instrument looks like Mexico. Perhaps more surprisingly, all the other countries look similar to Mexico.

Figure 3 shows less reassuring patterns for the college equivalent regressions. To take the Philippines (the highest weight instrument) as an example, the 1980 variation in the share of people from the Philippines implies as large an effect of the native-immigrant ratio on the native-immigrant wage gap in 1980 and 1990 as in 2000. That is, there is no evidence of change in 2000. Similarly, for other countries there are statistically significant pretrends. This evidence is consistent with the argument in Jaeger, Ruist, and Stuhler (2019) that the immigrant inflows are typically serially correlated and so the immigrant enclave instrument does not generate a well-defined shock to the supply of immigrants.

Test 3: Alternative Estimators and Overidentification Tests.—Panel A of Table 6 shows the results of alternative estimators and some overidentification tests for high school equivalent workers. Unlike in our other examples, the results are quite stable across estimators, with Bartik, overidentified TSLS, LIML, and MBTSLS all giving the same point estimate (HFUL, in contrast, is quite different). Similarly, the overidentification tests on the overidentified TSLS estimator fail to reject (though on LIML it does). This result can be approximately anticipated from Table 4 where the β_k on each individual instrument are quite similar.

Panel B of Table 6 shows that the results are broadly similar for college equivalent workers. Namely, the results are quite stable across estimators and the overidentification test fails to reject for both TSLS and LIML. Again, this result can be approximately anticipated from Table 4.

Table 6—OLS and IV Estimates: Immigrant Enclave

	Δ Emp		Coefficients equal	Over ID test
	(1)	(2)	(3)	(4)
Panel A. High school equivalent				
OLS	−0.02	−0.03	[0.00]	
	(0.01)	(0.01)		
TSLS (Bartik)	−0.02	−0.04	[0.07]	
	(0.01)	(0.01)		
TSLS	−0.02	−0.04	[0.02]	43.30
	(0.01)	(0.01)		[0.22]
MBTSLS	−0.03	−0.04	[0.08]	
	(0.01)	(0.01)		
LIML	−0.03	−0.04	[0.06]	73.16
	(0.01)	(0.01)		[0.00]
HFUL	0.03	0.02	[0.26]	82.45
	(0.01)	(0.00)		[0.00]
Panel B. College equivalent				
OLS	−0.06	−0.06	[0.65]	
	(0.01)	(0.01)		
TSLS (Bartik)	−0.08	−0.08	[0.93]	
	(0.01)	(0.01)		
TSLS	−0.06	−0.06	[0.71]	35.54
	(0.01)	(0.01)		[0.54]
MBTSLS	−0.06	−0.07	[0.71]	
	(0.01)	(0.01)		
LIML	−0.06	−0.06	[0.72]	33.67
	(0.01)	(0.01)		[0.63]
HFUL	0.04	0.04	[0.23]	67.95
	(0.01)	(0.00)		[0.00]
Controls	No	Yes		
Observations	124	124		

Notes: This table reports a variety of estimates of the negative of the inverse elasticity of substitution between immigrants and natives. The regressions are at the city level and include a single time period (2000). The *TSLS* row is our replication of column 3 and column 7 of Table 6 in Card (2009). Column 1 does not contain controls, while column 2 does. The *TSLS* (*Bartik*) row uses the Bartik instrument. The *TSLS* row uses each origin country share separately as instruments. The *MBTSLS* row uses the estimator of Anatolyev (2013) and Kolesár et al. (2015) with the same set of instruments. The *LIML* row shows estimates using the limited information maximum likelihood estimator with the same set of instruments. Finally, the *HFUL* row uses the HFUL estimator of Hausman et al. (2012) with the same set of instruments. The *J*-statistic for HFUL comes from Chao et al. (2014). The *p*-value for the equality of coefficients compares the adjacent columns with and without controls. The controls are the contemporaneous characteristics displayed in Table 5. Results are weighted by 1990 population. Standard errors are in parentheses and are constructed by bootstrap over commuting zones. *p*-values are in brackets.

Visualizing the Overidentification Tests.—Given that for several of the estimators the overidentification tests fail to reject, it is not surprising that visually there is not a great deal of dispersion in the point estimates across instruments. Figure 4 shows the heterogeneity in the $\hat{\beta}_k$ and the relationship to the first stage F-statistic. To compare to our other examples, note that the y-axis is dramatically compressed. Moreover, the high-weight industries are all very close to the overall estimate.

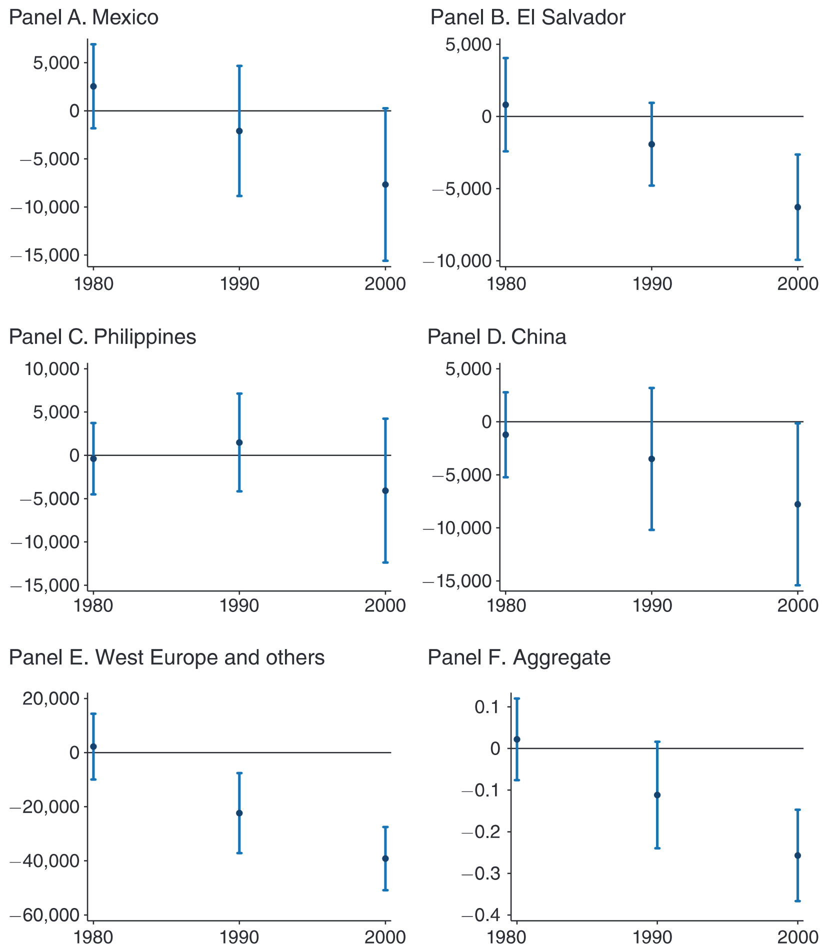

Figure 2. Pretrends for High Rotemberg Weight Origin Countries: Immigrant Enclave, High School Equivalent

Notes: These figures report pretrends for the overall instrument and the top-5 Rotemberg weight origin countries as reported in panel B of Table 4. The coefficients are estimated using the reduced-form regression of equations (11) and (12) with their 1980, 1990, and 2000 values (that is, we include all the controls in Card 2009 in Table 6, columns 3 and 7, and re-estimate year-by-year). Hence, the 2000 coefficient corresponds to the reduced-form coefficient estimated in Table 6. The *Others* are Cyprus, New Zealand, Israel, and Australia.

VIII. Summary

The central contribution of this paper revolves around understanding identification and the Bartik instrument. Our first set of formal results relate to identification in the sense typically used by econometricians. We show that Bartik is numerically equivalent to a GMM estimator with the industry shares as instruments. We use this equivalence to argue that in many settings the way to interpret the research design implicit in a Bartik instrument is a pooled exposure design. The shares measure the differential exposure to common shocks (the national growth rates), and so the relevant identification assumption, familiar from difference-in-differences, is that there are no other shocks correlated with this differential exposure.

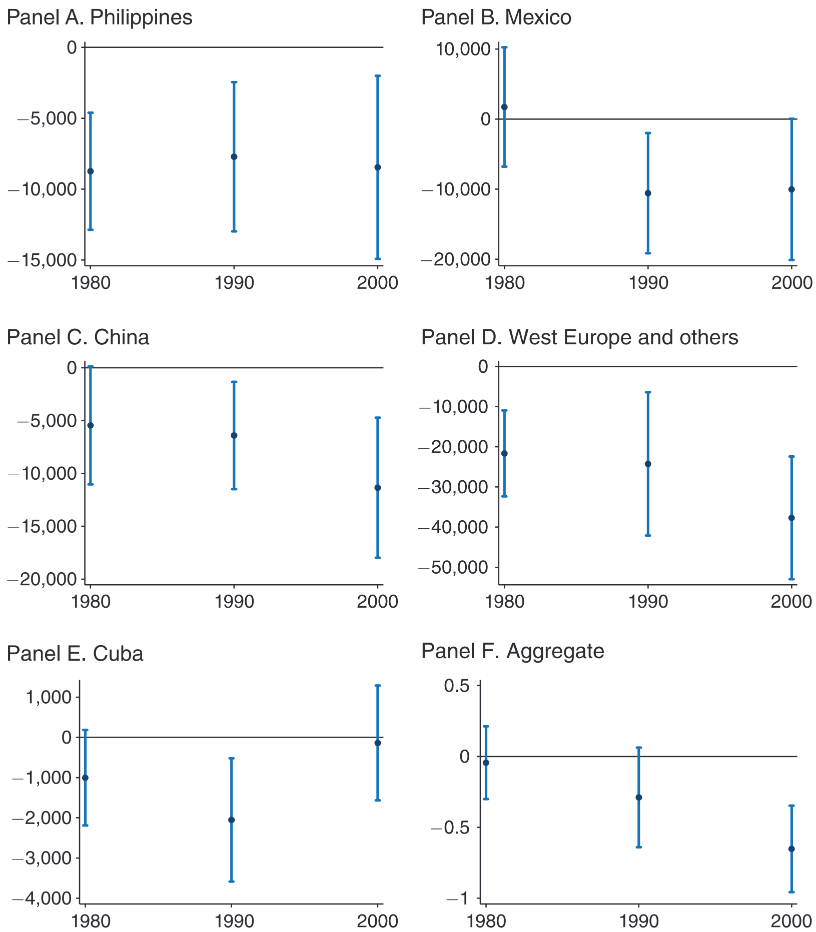

FIGURE 3. PRETRENDS FOR HIGH ROTEMBERG WEIGHT ORIGIN COUNTRIES: IMMIGRANT ENCLAVE, COLLEGE EQUIVALENT

Notes: These figures report pretrends for the overall instrument and the top-5 Rotemberg weight origin countries as reported in panel B of Table 4. The coefficients are estimated using the reduced-form regression of equations (11) and (12) with their 1980, 1990, and 2000 values (that is, we include all the controls in Card 2009 in Table 6, columns 3 and 7, and re-estimate year-by-year). Hence, the 2000 coefficient corresponds to the reduced-form coefficient estimated in Table 6. The *Others* are Cyprus, New Zealand, Israel, and Australia.

Our second set of formal results relates to identification in the sense often used by practitioners: we show how to compute which of the many instruments "drive" the estimates. Building on Andrews, Gentzkow, and Shapiro (2017) we show that these weights can be interpreted as sensitivity-to-misspecification elasticities and so highlight which identifying assumptions are most worth discussing and probing.

We then elaborated on a number of specification tests that researchers can carry out, and illustrated these tests through a number of applications. Our results clarify the set of reasonable concerns a consumer of the Bartik literature should have. We hope that researchers will use the results and tools in this paper to be clearer about how identification works in their papers, both in the econometric sense of stating the identifying assumption and in the practical sense of showing what variation drives estimates.

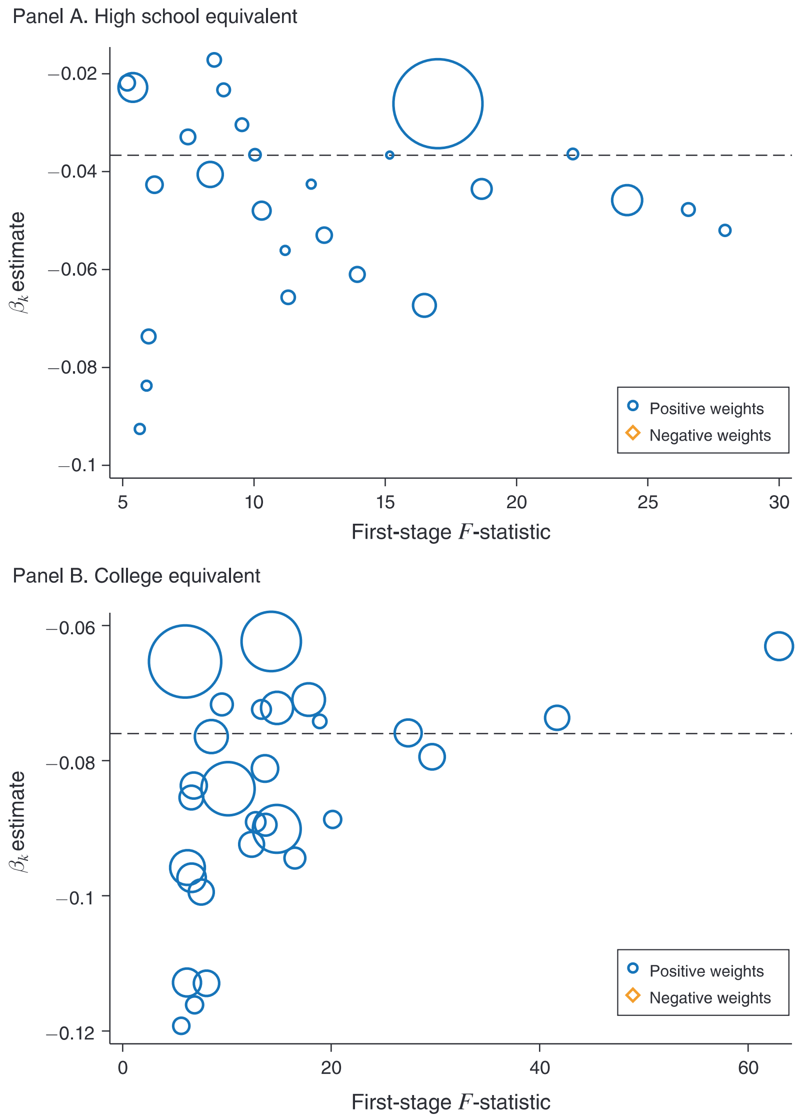

Figure 4. Heterogeneity of β_k: Immigrant Enclave

Notes: This figure plots the relationship between each instruments' $\hat{\beta}_k$, first-stage F-statistics and the Rotemberg weights. Each point is a separate instrument's (country of origin) estimates. The figure plots the estimated $\hat{\beta}_k$ for each instrument on the y-axis and the estimated first-stage F-statistic on the x-axis. The size of the points are scaled by the magnitude of the Rotemberg weights, with the circles denoting positive Rotemberg weights and the diamonds denoting negative weights. The horizontal dashed line is plotted at the value of the overall $\hat{\beta}$ reported in the second column in the *TSLS* (*Bartik*) row in Table 6. The figure excludes instruments with first-stage F-statistics below 5.

REFERENCES

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119 (3): 1049–90.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134 (4): 1949–2010.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

Anatolyev, Stanislav. 2013. "Instrumental Variables Estimation and Inference in the Presence of Many Exogenous Regressors." *Econometrics Journal* 16 (1): 27–72.

Anderson, T. W., and Herman Rubin. 1950. "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 21 (4): 570–82.

Andrews, Isaiah. 2019. "On the Structure of IV Estimands." *Journal of Econometrics* 211 (1): 294–307.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments." *Quarterly Journal of Economics* 132 (4): 1553–92.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14 (1): 57–67.

Autor, David, and David Dorn. 2013. "Replication Data for: The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market." American Economic Association [publisher]. https://www.ddorn.net/data.htm. Last accessed February 22, 2020 [distributor].

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute.

Baum-Snow, Nathaniel, and Fernando Ferreira. 2015. "Causal Inference in Urban and Regional Economics." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 3–68. Amsterdam: Elsevier.

Beaudry, Paul, David A. Green, and Benjamin Sand. 2012. "Does Industrial Composition Matter for Wages? A Test of Search and Bargaining Theory." *Econometrica* 80 (3): 1063–1104.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* (1).

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019. "Quasi-Experimental Shift-Share Research Designs." Unpublished.

Bound, John, and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics* 18 (1): 20–54.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43 (2): 245–57.

Card, David. 2009. "Immigration and Inequality." *American Economic Review* 99 (2): 1–21.

Card, David. 2019. "Code to Build Datasets Underlying Card (2009)." Unpublished code.

Chao, John C., Jerry A. Hausman, Whitney K. Newey, Norman R. Swanson, and Tiemen Woutersen. 2014. "Testing Overidentifying Restrictions with Many Instruments and Heteroskedasticity." *Journal of Econometrics* 178 (1): 15–21.

Chernozhukhov, Victor, and Christian Hansen. 2008. "The Reduced Form: A Simple Approach to Inference with Weak Instruments." *Economics Letters* 100 (1): 68–71.

Christian, Paul, and Christopher B. Barrett. 2017. "Revisiting the Effect of Food Aid on Conflict: A Methodological Caution." World Bank Policy Research Working Paper 8171.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1): 260–72.

DiNardo, John, and David S. Lee. 2011. "Program Evaluation and Research Designs." In *Handbook of Labor Economics*, Vol. 4A, edited by Orley Ashenfelter and David Card, 463–536. Amsterdam: Elsevier.

Freeman, Richard B. 1980. "An Empirical Analysis of the Fixed Coefficient 'Manpower Requirements' Model, 1960–1970." *Journal of Human Resources* 15 (2): 176–99.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Replication Data for: Bartik Instruments: What, When, Why, and How." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E117405V1.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2020. "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times." *American Economic Journal: Economic Policy* 12 (1): 200–25.

Hausman, Jerry A., Whitney K. Newey, Tiemen Woutersen, John C. Chao, and Norman R. Swanson. 2012. "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments." *Quantitative Economics* 3 (2): 211–55.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner. 2020. "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?" *Journal of Business & Economic Statistics* 38 (2): 317–26.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2019. "Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration." NBER Working Paper 24285.

Jensen, J. Bradford, and Lori G. Kletzer. 2005. "Tradable Services: Understanding the Scope and Impact of Services Offshoring." *Brookings Trade Forum: Offshoring White-Collar Work*: 75–133.

Kearney, Melissa S., and Phillip B. Levine. 2015. "Media Influences on Social Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing." *American Economic Review* 105 (12): 3597–3632.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics* 131 (3): 1057–1111.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and Inference with Many Invalid Instruments." *Journal of Business & Economic Statistics* 33 (4): 474–84.

Lucca, David O., Taylor Nadauld, and Karen Shen. 2019. "Credit Supply and the Rise in College Tuition: Evidence from the Expansion in Federal Student Aid Programs." *Review of Financial Studies* 32 (2): 423–66.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review* 108 (12): 3855–90.

Nunn, Nathan, and Nancy Qian. 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104 (6): 1630–66.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

Perloff, Harvey S. 1957. "Interrelations of State Income and Industrial Structure." *Review of Economics and Statistics* 39 (2): 162–71.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics* 9: 21–58.

Rotemberg, Julio J. 1983. "Instrumental Variable Estimation of Misspecified Models." MIT Sloan Working Paper 1508-83.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Sargan, J. D. 1958. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica* 26 (3): 393–415.

US Census Bureau. 2000. "Census of Population: Public-Use Microdata Sample, 1980–2000." United States Department of Commerce (Publisher). ICPSR (Distributor).

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.