Bartik Instruments: What, When, Why, and How*

Paul Goldsmith-Pinkham Isaac Sorkin Henry Swift

March 2018

Abstract

The Bartik instrument is formed by interacting local industry shares and national industry growth rates. We show that the Bartik instrument is numerically equivalent to using local industry shares as instruments. Hence, the identifying assumption is best stated in terms of these shares, with the national industry growth rates only affecting instrument relevance. We then show how to decompose the Bartik instrument into the weighted sum of the just-identified instrumental variables estimators, where the weights sum to one, can be negative and are easy to compute. These weights measure how sensitive the parameter estimate is to each instrument. We illustrate our results through three applications: estimating the inverse elasticity of labor supply, estimating local labor market effects of Chinese imports, and using simulated instruments to study the effects of Medicaid expansions.

*Goldsmith-Pinkham: Federal Reserve Bank of New York. Email: paulgp@gmail.com. Sorkin: Stanford University and NBER. Email: sorkin@stanford.edu. Swift: Unaffiliated. Email: henryswift@gmail.com. Thanks to Isaiah Andrews, David Autor, Kirill Borusyak, Jediphi Cabal, Gabriel Chodorow-Reich, Richard Crump, Rebecca Diamond, Mark Duggan, Matt Gentzkow, Gordon Hanson, Caroline Hoxby, Peter Hull, Guido Imbens, Xavier Jaravel, Pat Kline, Magne Mogstad, Maxim Pinkovskiy, Luigi Pistaferri, Giovanni Righi, Juan Carlos Suarez Serrato, Jan Stuhler, Kenneth West, Wilbert van der Klaauw, Eric Zwick and seminar participants at ASU, Chicago, Maryland, NBER SI: Labor Studies, New York Fed, St. Louis Fed./Wash. U., SoLE, Stanford, and Syracuse for helpful comments. Thanks to Maya Bidanda and Jacob Conway for research assistance. Thanks to Sarah Cohodes, Daniel Grossman, Samuel Kleiner and Michael Lovenheim for sharing code and data. 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.

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 variants have since been used across many fields in economics, including labor, public, development, macroeconomics, international trade, and finance. Indeed, as we discuss at the end of the introduction, numerous instruments have the same formal structure, including simulated instruments (Currie and Gruber (1996a) and Currie and Gruber (1996b)).

Our goal is to open the black box of the Bartik instrument by formalizing its structure and unpacking the variation that the instrument uses. 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 are used. For simplicity, consider the cross-sectional structural equation linking wage growth to employment growth

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

where y_l is wage growth in location l, x_l is the employment growth rate, and ϵ_l is a structural error term that is correlated with x_l. Our estimand 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

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

where g_k is the industry growth rate, g_l is the location 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, $B_l = \sum_k z_{lk} g_k$.

We first show that using the Bartik instrument is equivalent to using local industry

[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.

shares as instruments, with variation in the common industry component of growth only contributing to instrument relevance. More precisely, using the Bartik instrument in just-identified two-stage least squares is numerically equivalent to a generalized methods of moments (GMM) estimator with the local industry shares as instruments and a weight matrix constructed from the national industry growth rates. The intuition is that the variation in outcomes is at the location (l) level, and the only component of the instrument that varies at the local level is the industry shares. This result suggests that the identifying assumption for the Bartik instrument is best stated in terms of local industry composition. Moreover, it means that the Bartik *instrument* is not an instrument; instead, it is an overidentified GMM *estimator*.

While the industry shares are the instruments, the Bartik functional form is not arbitrary because the national growth rates are important for relevance. Economically, the Bartik combination of instruments can be derived from models in which there are industry-level shocks (e.g., Galle, Rodriguez-Clare, and Yi (2017) and Beaudry, Green, and Sand (Forthcoming)). Econometrically, the Bartik functional form is a set of restrictions on the first-stage coefficients in a two stage least squares (TSLS) estimator with the industry shares as instruments: specifically, it imposes that the first-stage coefficients are proportional to the national growth rates. While replacing the national growth rates with a random vector would not affect the validity of the exclusion restriction, it would be unlikely to generate a relevant instrument.

We next show how to measure the relative importance of each industry share in determining parameter estimates. We build on Rotemberg (1983) and decompose the Bartik estimator into a weighted sum of the just-identified instrumental variable (IV) estimators using each industry share (z_{lk}) as a separate instrument. The weights, which we refer to as Rotemberg weights, are simple to compute, sum to 1, and can be negative. They 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 given instrument. The weights depend on the covariance between the k^{th} instrument's fitted value of the endogenous variable and the endogenous variable itself. Empirically, the national growth rate plays a small role in determining this covariance.

We suggest researchers perform two relatively standard tests of the identifying assumptions. First, while there are many industries, the identifying variation is typically driven by a smaller number of industries. Reseachers should look at these high Rotemberg weight industries directly, and test how balanced the high-weight instruments are across potential confounders. This dimension-reduction focuses researchers' argument for their identifying assumption to instruments that matter in their estimate. Naturally, it is always possible to control for observable confounders, but following the logic of Altonji, Elder, and Taber (2005) and Oster (Forthcoming), movements in point estimates when conditioning on observable confounders suggest the potential importance of unobserved confounders.

Second, researchers can also perform overidentification tests. Under the null of constant effects, a rejection implies that some of the instruments are endogenous. While a rejection of the null of exogeneity is sometimes interpreted as evidence of heterogeneous treatment effects, without additional assumptions that are typically context specific, overidentified linear IV with unordered treatments has no obvious local average treatment effect (LATE) interpretation (e.g., Kirkeboen, Leuven, and Mogstad (2016) and Hull (2018)). Moreover, some papers using Bartik instruments test overidentifying restrictions and thus explicitly embrace the constant effects interpretation (e.g., Beaudry, Green, and Sand (2012) and Hornbeck and Moretti (2018)).

We illustrate our results through three 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 one percent of the variance of the Rotemberg weights. Second, we show that the weights are skewed, with about a third of the weight on the top five industry-time periods (out of 678 total industry-time periods). The top instrument is oil and gas extraction with a weight of 0.16. Hence, a concrete example of the comparisons being made by the estimator is comparing changes in employment growth and wage growth in places with more and less oil and gas extraction. Moreover, the empirical strategy is sensitive to any innovations in labor supply that are correlated with a location's oil and gas extraction employment share. Third, industry shares, including oil and gas extraction, are correlated with many observables, including the immigrant share. Controlling for these observables attenuates estimates (though not in a statistically significant way). Fourth, an overidentification test rejects the null of exogeneity.

In our second application, we estimate the effect of Chinese imports on manufacturing employment in the United States (using the China shock of Autor, Dorn, and Hanson (2013)). We first show that the growth rates of imports from China to other high-income countries explain about 30% of the variance in the Rotemberg weights. Second, we show that the highest-weight instrument is electronic computers in the 2000s with a weight 0.14. Hence, a concrete example of the comparisons being made by the estimator is comparing outcomes in locations with high and low shares of the electronic computers industry. Interestingly, Autor, Dorn, and Hanson (2013, pg. 2138) discuss that one might be worried that computer share is correlated with demand shocks and so would not be a valid instrument, and report sensitivity to dropping the instrument. Our weights allow the reader to do this calculation for themselves. Third, the industries that get the most weight tend to be in more educated areas, and controlling for observables attenuates estimates (though not in a statistically significant way). Fourth, an overidentification test rejects the null of exogeneity.

In our final application, we extend our set-up to the simulated instruments framework of Currie and Gruber (1996a) and Currie and Gruber (1996b). The key idea is that the variation is at the level of the eligibility type, where an eligibility type is a unique pattern of Medicaid eligibility across state-years among the households in the fixed population used to build the simulated instrument (in this case, the 1986 Current Population Survey). To map to the Bartik setting, the equivalent of the industry shares is state-year indicators of Medicaid eligibility for different eligibility types, and the equivalent of the national growth rates are the national population shares of each eligibility type.

We consider estimating the effect of Medicaid eligibility on schooling attainment as in (Cohodes et al., 2016). We use the Rotemberg weights to show which state-year Medicaid eligibility changes—and which household characteristics—drive the estimates. We find that the expansions between 1980 and 1997 had the largest effect for low-income households (less than 10,000 dollars in 1986 USD), while post-1996 expansions had the largest effect for higher-income households (greater than 10,000 dollars in 1986 USD). These policy changes were concentrated in Missouri, Minnesota, New Jersey and Washington, DC. In terms of household characteristics, we find that most weight is on households below twelve thousand dollars in household income, and in households with between two to four kids. Moreover, changes during school age receives the largest weight.

Besides the three examples discussed in the paper, 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. This encompasses at least three instruments, which are not always labelled as Bartik instruments. First, the "immigrant enclave" instrument introduced by Altonji and Card (1991) interacts initial immigrant composition of a place with immigration flows from origin countries. Second, researchers, such as Greenstone, Mas, and Nguyen (2015), interact pre-existing bank lending shares with changes in bank lending volumes to instrument for credit supply. Third, Acemoglu and Linn (2004) interact age-group spending patterns with demographic changes to instrument for market size. We discuss these examples in greater detail in Appendix A.

We note three limitations to our analysis. First, we assume locations are independent and so ignore the possibility of spatial spillovers or correlation.[2] Second, we assume that the data consist of a series of steady states.[3] Third, we assume that the number of locations

[2]Monte, Redding, and Rossi-Hansberg (2017) document the presence and economic importance of spatial spillovers through changes in commuting patterns in response to local labor demand shocks.

[3]See Jaeger, Ruist, and Stuhler (2018) for discussion of out-of-steady-state dynamics in the context of immi-

grows large, but fix the number of instruments (locations and industries).[4] Fixing the number of instruments means that we do not consider asymptotics in the spirit of Kolesar et al. (2015) where each instrument is potentially misspecified, but the misspecification averages out in the limit. In such a setting, the results of this paper are still important in highlighting that the researcher is using industry shares as (invalid) instruments.

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 (2017), which comments on Kearney and Levine (2015), and Jaeger, Ruist, and Stuhler (2018), 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.

To summarize, we view our contribution as explaining identification in the context of Bartik instruments in two senses. First, our GMM result shows that Bartik is numerically equivalent to using industry shares as instruments. Hence, we argue that the identifying assumption is best stated in terms of industry shares. 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.

Roadmap: Section 1 shows that the Bartik instrument is equivalent to using industry shares as instruments. Section 2 shows how to open the black box of the Bartik instrument and show which estimation moments the parameter estimate is most sensitive to. Section 3 discusses the implementation of the balance and overidentification tests. Section 4 presents an application to the canonical setting of estimating the inverse elasticity of labor supply. Section 5 presents our application to the China Shock. Section 6 discusses the simulated instrument application.

1 Equivalence between Bartik and industry shares

We first show that the Bartik instrument is equivalent to using industry shares as instruments. This result suggests that the identifying assumption is best stated 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 that Bartik is (numerically) equivalent to using local industry shares

gration.

[4]See Borusyak and Jaravel (2017) for discussion of the setting where the number of industries grows.

as instruments. To focus on identification issues, we discuss infeasible Bartik, where we assume that we know the common national component of industry growth rates.

1.1 Full panel setup

We begin by setting up the general panel data case with K industries and T time periods. This set-up 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:

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

We consider $\{\{x_{lt}, D_{lt}, \epsilon_{lt}\}_{t=1}^{T}\}_{l=1}^{L}$ to be independent and identically distributed with T fixed and L growing to infinity. 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. We allow for $\mathbb{E}[x_{lt}\epsilon_{lt}|D_{lt}] \neq 0$. The estimand of interest is β_0, typically assumed to be the inverse elasticity of labor supply. However, 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 $1 \times K$ row vector of industry-location-time period shares, $\sum_{k=1}^{K} z_{lkt} = 1$, and G_{lt} is a $K \times 1$ vector of industry-location-time period growth rates where the k^{th} entry is g_{lkt}. We decompose the industry-location-period growth rate into industry-period, location-period and idiosyncratic industry-location-period components:

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

Let the location-period and idiosyncratic industry-location-period components (g_{lt} and $\tilde{g}_{lkt}$) be mean zero random variables. Since K and T are fixed, g_{kt} is assumed to be non-stochastic and fixed. We fix industry shares to an initial time period, so that the Bartik instrument is the inner product of the initial industry-location shares and the industry-period growth rates:[5]

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

where G_t is a $K \times 1$ vector of the industry growth rates in period t (the k^{th} entry is g_{kt}), and Z_{l0} is the $1 \times K$ vector of industry shares in location l. Define Z_0 to be the $L \times K$ matrix of industry-location shares (we suppress L subscripts for notational simplicity, but the dimension of these matrices grows with the number of locations).

Hence, we have a standard two-stage least squares set-up where the first-stage is a regression of employment growth on the set of controls and the Bartik instrument:

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

The two assumptions for the validity of the TSLS estimator are that the Bartik instrument is exogenous, $\mathbb{E}\left(B_{lt}\epsilon_{lt} | D_{lt}\right) = 0$, and that it is relevant: $\gamma \neq 0$.

1.2 Equivalence in three special cases

We build up to the result that the Bartik instrument is equivalent to using industry shares as instruments through three special cases which each cleanly illustrate one aspect of the general result.

Two industries and one time period

With two industries and one time period, it is clear that the Bartik instrument is identical to industry shares as instruments. To see this, write the Bartik instrument explicitly:

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

where g_1 and g_2 are the industry components of growth. Since the shares sum to one, with only two industries, 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:

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

[5] If ϵ_{lt} are correlated with growth rates, and the ϵ_{lt} are serially correlated, then future shares will be endogenous. This potential for serial correlation motivates fixing industry shares to some initial period. Beaudry, Green, and Sand (Forthcoming, pg. 18-19) discuss Bartik instruments and advocate updating the shares under the assumption that the error term is not serially correlated.

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.

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 using industry shares as instruments. To see this result, we 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 the panel estimating equations, restrict our panel setup to have the vector of controls consist solely of location and time fixed effects. Then the first-stage is

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

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

$$x_{lt} = \tau_l + \underbrace{(\tau_t + g_{2t}\gamma)}_{\equiv \tilde{\tau}_t} + z_{l10} \underbrace{(g_{1t} - g_{2t})\gamma}_{\equiv \tilde{\gamma}_t} + \eta_{lt}. \tag{1.1}$$

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 $g_{1t} - g_{2t} = (g_{11} - g_{21}) + (\Delta g_1 - \Delta g_2)\mathbb{1}(t = 2)$, where $\Delta g_1 = g_{12} - g_{11}$ and $\Delta g_2 = g_{22} - g_{21}$. Then, rewrite the first stage as

$$x_{lt} = \underbrace{\tau_l + z_{l10}(g_{11} - g_{21})\gamma}_{\equiv \tilde{\tau}_l} + \underbrace{(\tau_t + g_{2t}\gamma)}_{\equiv \tilde{\tau}_t} + z_{l10}\underbrace{\mathbb{1}(t = 2)(\Delta g_1 - \Delta g_2)\gamma}_{\equiv \tilde{\gamma}_t} + \eta_{lt}. \quad (1.2)$$

We can now see the equivalence between Bartik and using the shares as instruments:

$$x_{lt} = \tilde{\tau}_l + \tilde{\tau}_t + z_{l10}\mathbb{1}(t = 2)(\Delta g_1 - \Delta g_2)\gamma + \eta_{lt} \quad \text{(Bartik)}$$

$$x_{lt} = \tilde{\tau}_t + \tilde{\tau}_t + z_{l10}\mathbb{1}(t = 2)\tilde{\gamma} + \eta_{lt}. \quad \text{(Industry Shares)}$$

In this case, $\tilde{\gamma} = \gamma/(\Delta g_1 - \Delta g_2)$. If we view z_{l10} as the effect of exposure to a policy, then $\tilde{\gamma}$ 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 dispersion in national industry growth rates, $\Delta g_{1t} - \Delta g_{2t}$.

K industries and one time period

Finally, we show that with K industries as instruments in a generalized method of moments (GMM) set-up with a specific weight matrix, Bartik is identical to using the set of industry shares as instruments.

To prove this result, we introduce some additional notation. Let G be the $K \times 1$ vector of industry growth rates, let Z be the $L \times K$ matrix of industry shares, let Y be the $L \times 1$ vector of outcomes, let X be the $L \times 1$ vector of endogenous variables, let $B = ZG$ be the $L \times 1$ vector of Bartik instruments, and let W be an arbitrary $K \times K$ matrix. Finally, let $M_D = I_L - D(D'D)^{-1}D$ denote the annhilator matrix for D, the $L \times Q$ matrix of controls. We denote $X^{\perp} = M_D X$ and $Y^{\perp} = M_D Y$ to be the residualized X and Y. We define the Bartik and the GMM estimator using industry shares as instruments:

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

The following proposition says that Bartik and GMM are equivalent for a particular choice of weight matrix.

PROPOSITION 1.1. *If $W = GG'$, then $\hat{\beta}_{GMM} = \hat{\beta}_{Bartik}$.*

Proof. See appendix B. □

Hence, the Bartik instrument and industry shares as instruments are numerically equivalent for a particular choice of weight matrix. The decision of what growth rates to use is about how to weight a set of instruments, and not about the identification assumption.

1.3 Equivalence with K industries and T time periods

We now have the ideas necessary to show the equivalence between Bartik and industry shares as instruments in the general set-up of Section 1.1, which includes arbitrary controls including location and time fixed effects.[6] With K industries and T time periods, the 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 Appendix C.

2 Opening the black box of the Bartik estimator

The previous section showed that Bartik is equivalent to using industry shares (interacted with time) as instruments. Thus, the Bartik estimator combines many instruments using a particular weight matrix.

While empirical work using a single instrument is quite transparent because there is a clear set (and a small number) of variances and covariances that enter the estimator, with many instruments it is not clear how the estimator combines the different instruments. Hence, empirical work using Bartik instruments often feels like a black box, in that it is hard to explain what variation in the data drives estimates.

In this section, we show how to open the black box of the Bartik estimator. We first decompose the Bartik estimator into a weighted combination of just-identified estimates. We view this decomposition as increasing the transparency of the estimator because the weights highlight the industries whose variation in the data drives the estimate. Building on Andrews, Gentzkow, and Shapiro (2017), we show that in the limit, these weights can be interpreted as sensitivity-to-misspecification elasticities. High-weight instruments are more sensitive to misspecification, and hence are the instruments that are most important for researchers to defend.

2.1 Decomposing the Bartik estimator

We first present an exact 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 the panel by multiplying the dimensions

[6] The identification condition strengthens when moving from the Bartik instrument to industry shares as instruments. Specifically it moves from $\sum_k g_{kt}\mathbb{E}[z_{lkt}\epsilon_{lt}|D_{lt}] = 0$ to $\mathbb{E}[z_{lkt}\epsilon_{lt}|D_{lt}] = 0, \forall k, t$. It is logically possible for the first to hold but not the second if the g_{kt} are "special." We view this case as a knife-edge case.

[7] Andrews (2017, Section 3.1) reports this decomposition for constant-effect linear instrumental variables.

of vectors by T.[8]

Consider the GMM problem of estimating a scalar β_0, using K empirical moment conditions $\hat{g}(\beta) = Z'(Y^{\perp} - X^{\perp}\beta)$, where we have residualized for a matrix of control variables D, and a $K \times K$ weight matrix $\hat{W}$:

$$\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)'\hat{W}\hat{g}(\beta). \tag{2.1}$$

Define a $K \times 1$ vector, $\hat{C}(\hat{W})$, which also depends on the instrument set, Z, and the endogenous variable $X^{\perp}$:

DEFINITION 2.1. *Let*

$$\hat{C}(\hat{W}) = \hat{W}Z'X^{\perp} \text{ and } \hat{c}_k(\hat{W}) = \hat{W}_k Z'X^{\perp}, \tag{2.2}$$

where $\hat{W}_k$ is the k^{th} row of $\hat{W}$.

We index a solution for $\hat{\beta}$ by $\hat{W}$:

$$\hat{\beta}(\hat{W}) = \frac{\hat{C}(\hat{W})'Z'Y^{\perp}}{\hat{C}(\hat{W})'Z'X^{\perp}}. \tag{2.3}$$

The following result (which is a special case of Rotemberg (1983, Proposition 1)) shows how to decompose $\hat{\beta}(\hat{W})$ into the contribution from each of the K just-identified regressions.

PROPOSITION 2.1. *Let*

$$\hat{\beta}(\hat{W}) = \frac{\hat{C}(\hat{W})'Z'Y^{\perp}}{\hat{C}(\hat{W})'Z'X^{\perp}},\ \hat{\alpha}_k(\hat{W}) = \frac{\hat{c}_k(\hat{W})Z_k'X^{\perp}}{\sum_{k'} \hat{c}_{k'}(\hat{W})Z_k'X^{\perp}},\text{ and } \hat{\beta}_k = (Z_k'X^{\perp})^{-1}Z_k'Y^{\perp}. \tag{2.4}$$

Then:

$$\hat{\beta}(\hat{W}) = \sum_{k=1}^{K} \hat{\alpha}_k(\hat{W})\hat{\beta}_k, \tag{2.5}$$

where $\sum_{k=1}^{K} \hat{\alpha}_k(\hat{W}) = 1$.

Proof. See appendix B. □

Proposition 2.1 has three important implications. First, mirroring our results from Section 1, the validity of each just-identified $\hat{\beta}_k$ estimate depends on the exogeneity of a given Z_k, and is not related to the choice of $\hat{W}$. Second, the $\hat{\alpha}_k(\hat{W})$ weights sum to 1, and differ depending on the choice of $\hat{W}$. Finally, for some k, $\hat{\alpha}_k(\hat{W})$ can be negative, which means that the overidentified IV estimate can lie outside of the range of the just-identified estimates.

[8] For example, Z becomes $LT \times 1$, and X becomes $LT \times Q$.

We now look at Bartik's Rotemberg weights.

REMARK 2.1. *The Rotemberg weights for the Bartik instrument are given by:*

$$\hat{\alpha}_k(GG') = \frac{g_k Z_k' X^{\perp}}{\sum_{k=1}^{K} g_k Z_k' X^{\perp}} = \frac{\hat{\gamma} g_k Z_k' X^{\perp}}{\hat{\gamma} B' X^{\perp}} = \frac{\hat{X}_k^{Bartik\prime} X^{\perp}}{\hat{X}^{Bartik\prime} X^{\perp}}, \tag{2.6}$$

where g_k is the k^{th} entry in G, and $\hat{X}_k^{Bartik}$ is the $L \times 1$ vector of the fitted values from the first-stage regression using the full Bartik instrument, but applying the coefficient to the k^{th} industry.

To see the specific restrictions imposed by Bartik, we contrast Bartik's Rotemberg weights with the weights that arise from TSLS using the industry shares as instruments:

REMARK 2.2. *The Rotemberg weights from TSLS are given by:*

$$\hat{\alpha}_k((Z^{\perp\prime} Z^{\perp})^{-1}) = \frac{\hat{\pi}_k Z_k' X^{\perp}}{\sum_{k=1}^{K} \hat{\pi}_k Z_k' X^{\perp}} = \frac{\hat{X}_k^{TSLS\prime} X^{\perp}}{\hat{X}^{TSLS\prime} X^{\perp}}, \tag{2.7}$$

where $\hat{\pi}_k$ is the k^{th} entry in $(Z^{\perp\prime} Z^{\perp})^{-1} Z' X^{\perp}$, which is the first stage regression when using all K industries as instruments, and $\hat{\pi}_k Z_k = \hat{X}_k^{TSLS}$ is the $L \times 1$ vector of fitted values based on the k^{th} industry.

This comparison lets us see two points. First, the Bartik and TSLS estimators are identical when the TSLS first-stage coefficients are proportional to g_k (the national growth rates).[9] Second, the weights reflect the covariance between the k^{th} instrument's fitted value of the endogenous variable and the endogenous variable itself. To understand this covariance, let $\hat{X}_k$ be a first stage fitted value using the k^{th} instrument (e.g., $g_k Z_k$ or $\hat{\pi}_k Z_k$) so that $Cov(\hat{X}_k, X) = Var(\hat{X}_k) + \sum_{j \neq k} Cov(\hat{X}_k, \hat{X}_j) + Cov(\hat{X}_k, \hat{\epsilon})$, where $\hat{\epsilon} = X - \sum_{k=1}^{K} \hat{X}_k$. If the instruments are mutually orthogonal, then all the covariance terms are zero. If, in addition, the coefficients on the instruments come from a regression (i.e., in TSLS), then the covariance with the error term is also zero. Under these two assumptions, the weights measure the share of first-stage partial R^2 that is attributable to each instrument and all weights are positive.[10] If we relax these two assumptions, then negative weights are possible.[11]

2.2 Interpreting the weights

To interpret these weights, we move from finite samples to population limits. We first state the standard assumptions such that GMM estimators are consistent for all sequences of

[9] In the limit, this occurs when $\mathbb{E}[G|Z] = \mathbb{E}[G]$. This assumption also implies that the first stage coefficient is 1. See Appendix D for details.

[10] Angrist and Imbens (1995, Theorem 2) present a related result where the instruments are mutually orthogonal and they study TSLS so the weights are all positive.

[11] To generate mutually orthogonal instruments, we could take the PCA components of the industry shares.

$\hat{W}$ matrices. We then consider local-to-zero asymptotics (e.g., Conley, Hansen, and Rossi (2012)) to interpret the Rotemberg weights in terms of sensitivity-to-misspecification as discussed in Andrews, Gentzkow, and Shapiro (2017) (AGS). As such, the results in this section largely follow as special cases of AGS's results.

The Rotemberg weights depend on the choice of weight matrix, $\hat{W}$. Given standard assumptions, the choice of weight matrix does not affect consistency or bias of the estimates, and only affects the asymptotic variance of the estimator (there is a rich literature studying how to optimize this choice).

When some of the instruments are not exogeneous, however, the population version of the Rotemberg weights measures how much the estimate of β_0 is affected by this moment misspecification. To allow for this misspecification interpretation, we modify our estimating equation:

$$
\begin{aligned}
y_{lt} &= D_{lt}\rho + x_{lt}\beta_0 + \tilde{\epsilon}_{lt} \\
\tilde{\epsilon}_{lt} &= V_{lt}\kappa + \epsilon_{lt},
\end{aligned}
$$

where we assume that for some k, $\mathbb{E}[Z_{lkt}V_{lt}|D_{lt}] \neq 0$. We follow Conley, Hansen, and Rossi (2012, Section III.C) and AGS (pg. 1569) and allow κ to be proportional to $L^{-1/2}$ such that we have local misspecification. We make the following standard regularity assumptions:

ASSUMPTION 1 (Identification and Regularity). *(i) the data $\{\{x_{lt}, Z_{lt}, D_{lt}, V_{lt}, \epsilon_{lt}\}_{t=1}^{T}\}_{l=1}^{L}$ are independent and identically distributed with K and T fixed, and L going to infinity;*

(ii) $\mathbb{E}[\epsilon_{lt}] = 0$, $\mathbb{E}[V_{lt}] = 0$ and $Var(\tilde{\epsilon}) < \infty$;

(iii) $\mathbb{E}[z_{lkt}\epsilon_{lt}|D_{lt}] = 0$ for all values of k; $\mathbb{E}[z_{lt}V_{lt}] = \Sigma_{ZV}$, where Σ_{ZV} is a $1 \times K$ covariance vector with at least one non-zero entry; and $\mathbb{E}[Z_{lt}x_{lt}^{\perp}] = \Sigma_{ZX^{\perp}}$ is a $1 \times K$ covariance vector with all non-zero entries (x_{lt} is a scalar), and $\Sigma_{ZX^{\perp},k}$ is the k^{th} entry; and

(iv) $Var(z_{lkt}\epsilon_{lt}) < \infty$, $Var(z_{lkt}V_{lt}) < \infty$ and $Var(z_{lkt}x_{lt}^{\perp}) < \infty$ for all values of k.

We first establish the population version of $\hat{\alpha}_k(\hat{W})$:

LEMMA 2.1. *If Assumption 1 holds and $\text{plim}_{L\to\infty} \hat{W}_L = W$ where W is a positive semi-definite matrix, then*

$$
\operatorname*{plim}_{L\to\infty} \hat{\alpha}_k(\hat{W}) = \alpha_k(W) = \frac{\Sigma_{ZX^{\perp}} W_k \Sigma_{ZX^{\perp},k}}{\Sigma_{ZX^{\perp}} W \Sigma'_{ZX^{\perp}}}.
$$

Proof. See appendix B. □

We now present results about the asymptotic behavior of our estimators with and without misspecification.

PROPOSITION 2.2. *We assume that Assumption 1 holds and* $\operatorname{plim}_{L\to\infty} \hat{W}_L = W$ *where* W *is a positive semi-definite matrix.*

(i) If $\kappa = 0$*, then standard GMM results apply*

$$\operatorname*{plim}_{L\to\infty} \sum_{k=1}^{K} \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \sum_{k=1}^{K} \alpha_k(W)\beta_0 = \beta_0 \tag{2.8}$$

and $\sqrt{L}(\hat{\beta} - \beta_0)$ *converges in distribution to a random variable* $\tilde{\beta}$*, with* $\mathbb{E}[\tilde{\beta}] = 0$*.*

(ii) If $\kappa = L^{-1/2}$*, then*

$$\operatorname*{plim}_{L\to\infty} \sum_{k=1}^{K} \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \sum_{k=1}^{K} \alpha_k(W)\beta_0 = \beta_0 \tag{2.9}$$

and

(a) $\sqrt{L}(\hat{\beta}_k - \beta_0)$ *converges in distribution to a random variable* $\tilde{\beta}_k$*, with* $\mathbb{E}[\tilde{\beta}_k] = \frac{\Sigma_{ZV,k}}{\Sigma_{ZX^\perp,k}}$ *and*

(b) $\sqrt{L}(\hat{\beta} - \beta_0)$ *converges in distribution to a random variable* $\tilde{\beta}$*, with* $\mathbb{E}[\tilde{\beta}] = \sum_{k=1}^{K} \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] = \sum_{k=1}^{K} \alpha_k(W)\frac{\Sigma_{ZV,k}}{\Sigma_{ZX^\perp,k}}$*.*

Proof. See appendix B. □

The first part of the proposition is the standard result that in the absence of misspecification the estimator is consistent and asymptotically unbiased. The second part shows that in the presence of misspecification while the estimator is consistent (because the misspecification disappears in the limit), it is asymptotically biased.

Two useful corollaries follow:

COROLLARY 2.1. *Suppose that* $\beta_0 \neq 0$*. Then the percentage bias can be written in terms of the Rotemberg weights:*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \alpha_k(W)\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \tag{2.10}$$

COROLLARY 2.2. *Under the Bartik weight matrix (*$W = GG'$*),*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \frac{g_k\Sigma_{ZX^\perp,k}}{G'\Sigma'_{ZX^\perp}}\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \tag{2.11}$$

The first corollary interprets the $\alpha_k(W)$ as a sensitivity-to-misspecification *elasticity*. Because of the linear nature of the estimator, it rescales the Andrews, Gentzkow, and Shapiro

(2017) sensitivity parameter to be unit-invariant, and is hence comparable across instruments.[12] Specifically, $\alpha_k(W)$ is the percentage point shift in the bias of the over-identified estimator given a percentage point change in the bias from a single industry. The second corollary gives the population version of Bartik's Rotemberg weights.

We can also use these weights to understand an alternative approach to measuring sensitivity which is to drop an instrument and then re-estimate the model. Let $\hat{\beta}(\hat{W}_{-k})$ be the same estimator as $\hat{\beta}(\hat{W})$, except excluding the k^{th} instrument. Then, define the bias term for $\hat{\beta}(\hat{W}_{-k})$ as $\tilde{\beta}(\hat{W}_{-k}) = \hat{\beta}(\hat{W}_{-k}) - \beta$.

PROPOSITION 2.3. *The difference in the bias from the full estimator and the estimator that leaves out the k^{th} industry is:*

$$\frac{\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right]}{\beta} = \alpha_k(W)\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta} - \frac{\alpha_k(W)}{1-\alpha_k(W)}\sum_{k'\neq k}\alpha_{k'}(W)\frac{\mathbb{E}[\tilde{\beta}_{k'}]}{\beta}.$$

If $\mathbb{E}[\tilde{\beta}_{k'}] = 0$ for $k' \neq k$, then we get a simpler expression:

$$\frac{\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right]}{\beta} = \alpha_k(W)\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta}.$$

Proof. See appendix B. □

As emphasized by AGS (Appendix A.1), dropping an instrument and seeing how estimates change does not directly measure sensitivity. Instead, this measure combines two forces: the sensitivity of the instrument to misspecification, and how misspecificed the instrument is relative to the remaining instruments.

The implications for practice is that researchers should report the instruments associated with the largest values of α_k. The reason is twofold: first, reporting the instruments with the largest α_k provides the researcher a way to describe their empirical strategy in a way that reflects what variation they are using. Second, to the extent that the researcher is concerned about misspecification, these are the instruments that are most worth probing.

3 Testing the plausibility of the identifying assumption

In the context of the canonical setting of using Bartik to estimate the inverse elasticity of labor supply, the identifying assumption is that industry composition does not predict in-

[12] AGS (pg. 1558) write: "The second limitation is that the units of [our sensitivity vector] are contingent on the units of [the moment condition]. Changing the measurement of an element [j of the moment condition] from, say, dollars to euros, changes the corresponding elements of [the sensitivity vector]. This does not affect the bias a reader would estimate for specific alternative assumptions, but it does matter for qualitative conclusions about the relative importance of different moments."

novations to labor supply.[13] While this exclusion restriction is not directly testable, we discuss two tests of its plausibility.[14]

3.1 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 descriptive evidence both of what the variation is correlated with, and the types of mechanisms that would be problematic for the exclusion restriction.

Since we argued in footnote 5 that it is typically desirable to fix industry shares to an initial time period (Z_{l0}), it is useful to consider the correlation with initial period characteristics, as this reflects the cross-sectional variation that the instrument uses. If a researcher finds that Z_{l0} is correlated with potential confounding factors, this can imply that there are omitted variables influencing estimation. Naturally, it is always possible to control for observable confounders, but following the logic of Altonji, Elder, and Taber (2005) and Oster (Forthcoming), 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 because the estimator is most sensitive to misspecification of these instruments.

3.2 Empirical Test 2: Overidentification tests

Under the assumption of constant effects in β_0, we can use an overidentification test to test the null hypothesis that all of the industry shares are uncorrelated with innovations to supply shocks. Conceptually, the test asks whether the instruments are correlated with the error term beyond what would be expected by chance.

To implement the overidentification test, we run into the problem that there are many instruments, and standard overidentification tests perform poorly (in Monte Carlo experiments, they overreject). To address this problem, we use estimators and tests that are designed to have good finite sample performance in the presence of many instruments and heteroskedasticity. In particular, we use the HFUL estimator from Hausman et al. (2012), which is a heteroskedasticity robust version of the Fuller (1977) estimator (hence, H-FUL). This estimator is a jackknife estimator. See Hausman et al. (2012) for further discussion

[13] In Appendix E we write down an economic model which allows us to derive this statement more precisely.

[14] In general, it is not possible to test for pre-trends without additional assumptions because the instrument is industry times time and so there is no pre-period. An example of an additional assumption would be that a given industry has no predictive power for employment growth in some pre-period.

of the properties of this estimator.[15] We implement the overidentification test from Chao et al. (2014). Under the assumption of constant effects, this tests for whether all instruments are exogenous. While the constant effects assumption might seem unappealing, as we discussed in the introduction there are not currently results in the literature that define a LATE in our setting. Hence, a weaker interpretation of the rejection of the null is simply that it suggests a lack of clarity in what the estimates mean.

These overidentification tests test whether each instrument is estimating the same parameter. Beaudry, Green, and Sand (2012) and others test overidentifying restrictions in Bartik settings by using different vectors of national growth rates to weight the industry shares and testing whether these different weighted combinations of instruments estimate the same parameter. We show in Appendix F that it is possible to fail to reject the null using this test, even if a more traditional overidentification test would reject.

4 Empirical example I: Canonical Setting

So far we have emphasized that Bartik is equivalent to using industry shares (interacted with time) as instruments and hence that the identifying assumption is best viewed in terms of the industry shares. We have also shown how to open the black box of the Bartik estimator and show which industries drive the results.

We now present empirical examples to make these ideas concrete. Our first example is the canonical setting of estimating the inverse elasticity of labor supply. We begin by reporting the main estimates and then report the industries with the highest Rotemberg weight. We then probe the plausibility of the identifying assumption for these instruments.

4.1 Dataset

We use the 5% sample of IPUMS of U.S. 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 and 3-digit IND1990 industries.[16] 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.

[15] Arguably, researchers use the Bartik instrument to address the problem of having many weak instruments. Under this argument, researchers should instead be drawing on the large econometrics literature that addresses this issue and using an estimator like HFUL.

[16] There are 228 non-missing 3-digit IND1990 industries in 1980. There are 722 continental US commuting zones.

Why do we use leave-one-out means to construct the national growth rates? The reason is that a leave-one-out estimator addresses the finite sample bias that comes from using own-observation information to fit the first stage. Specifically, using own-observation information allows the first-stage to load on the idiosyncratic industry-location component of the growth rate, which is why OLS is biased. 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)). Hence, the desirability of using a leave-one-out estimator of the national growth rate should not be interpreted as evidence that identification "comes from" the national growth rates. In practice, because we have 722 locations, using leave-one-out to estimate the national growth rates matters little in point estimates.[17]

4.2 Parameter estimates

Table 1 reports OLS and IV estimates with and without controlling for the 1980 covariates (we discuss these covariates below) and makes two main points. First, the IV estimates in columns (5) and (6) are bigger than the OLS estimates in columns (1) and (2). Second, the Bartik results are sensitive to the inclusion of controls. Adding controls moves the parameter estimate from 1.75 to 1.28, though these are not statistically distinguishable.

4.3 Rotemberg weights

We compute the Rotemberg weights on the IV estimates with controls. Despite the very large number of instruments ($684 = 228 \times 3$), the distribution of sensitivity is skewed so that in practice a small number of instruments get a large share of the weight: Table 2 shows that the top five instruments account for almost a third ($^{0.507}/_{1.877}$) of the positive weight in the estimator. These top five instruments are: oil and gas extraction in the 2000s, other in the 1990s,[18] motor vehicles in the 2000s, guided missiles in the 1980s, and motor vehicles in the 1990s. (Recall that we measure industry shares in 1980, so the dates refer to the time span of the outcomes).

These weights give a way of describing the research design in a way that reflects the variation in the data that the estimator is using, and hence makes more 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 of outcomes in places from 2000 to 2010 with greater and smaller shares of oil and gas extraction. This single industry

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

[18] The "Other" industry is the "N/A" code in the IND1990 classification system.

has a sensitivity elasticity of 0.16. So 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, pg. 24)), and this analysis confirms that the motor vehicle industry plays a large role in the Bartik instrument. Relative to the conventional story where the instrument uses the common decline in the U.S. auto industry in the 1980s, this analysis shows that it is driven more by events in these places in the 1990s and 2000s. More generally, however, the estimator does place more weight on the 1980s: Panel C sums the weights across years and shows that the Bartik estimator is most sensitive to misspecification in the 1980s.

Negative weights are quantitatively important. Panel A shows that that there are substantial negative weights—they sum to -0.8.

Finally, Panel B shows the determinants of the sensitivity to misspecification elasticities. Because the national growth rates—the g_k—enter the formula for the elasticities (see Remark 2.1) one might think that these would be the primary determinants of the weights. Instead, the Table shows that the national growth rates are weakly correlated with the elasticities. In contrast, the elasticities are quite related to the variation in the industry shares across locations—Var(z_k). Mechanically, efficient estimators place more weight on the instruments that vary more, and Var(z_k) measures how much the instrument varies across locations. This observation also explains why the industries with high weight tend to be tradables: almost by definition, tradables have industry shares that vary across locations, while non-tradables do not.[19]

4.4 Testing the plausibility of the identifying assumption

Test #1: Correlates of 1980 industry shares Table 3 shows the relationship between 1980 characteristics and the 4 industries that appear among the top 5 instruments in Table 2 (recall that motor vehicles appear in two time periods).[20] It also shows the relationship between these characteristics and the Bartik instrument. The first point to notice is that the R^2 in these regressions are quite high: for example, we can explain 43% of the variation in share of the "other" industry via male, white, native-born share, education composition, veteran status, and number of children. For the remaining three industries the R^2 is also high: 11% for motor vehicles, 22% for oil and gas extraction and 25% for guided missiles. Using the 1980-1990 growth rates, the Bartik instrument is even more strongly explained

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

[20] We compute the Rotemberg weights in a regression that includes the controls.

by these correlates. The second thing to note is that "other," oil and gas extraction, and the overall Bartik instrument are statistically significantly correlated with share native born. In the immigrant enclave literature, share native born (immigrant share) is thought to predict labor supply shocks.

Test #2: Overidentification tests Columns (7) and (8) of Table 1 shows the point estimates from the Hausman et al. (2012) HFUL estimator and the Chao et al. (2014) overidentification tests. Relative to the TSLS estimates using the Bartik instrument in columns (5) and (6), the HFUL estimator produces substantively larger point estimates. The key punchline of the overidentification tests (which can be approximately anticipated from looking at the dispersion in the $\hat{\beta}_k$ in Panels D and E of Table 2) is that the tests rejects the null that all instruments are exogenous under the assumption of constant treatment effects.

5 Empirical example II: China shock

For our second empirical example, we estimate the effect of Chinese imports on manufacturing employment in the United States using the China shock of Autor, Dorn, and Hanson (2013) (ADH).

5.1 Specification

It is helpful to write ADH's main regression specification in our notation. ADH are interested in a regression (where we omit covariates for simplicity, but include them in the regressions):

$$y_{lt} = \beta_0 + \beta X_{lt} + \epsilon_{lt}, \tag{5.1}$$

where y_{lt} is the percentage point *change* in manufacturing employment rate, and $X_{lt} = \sum_k z_{lkt} g_t^{US}$ is import exposure, where z_{lkt} is contemporaneous start-of-period industry-location shares, and g_t^{US} is a normalized measure of the *growth* of imports from China to the US. The first stage is:

$$X_{lt} = \gamma_0 + \gamma_1 B_{lt} + \eta_{lt}, \tag{5.2}$$

where $B_{lt} = \sum_k z_{lkt-1} g_t^{high-income}$, the z are lagged, and $g_t^{high-income}$ is a normalized measure of the *growth* of imports from China to other high-income countries (mainly in Europe).

The fact that B_{lt} and X_{lt} differ primarily in the g makes it tempting to explain the empirical strategy as instrumenting "for growth in Chinese imports to the United States using the contemporaneous composition and growth of Chinese imports in eight other developed countries" (ADH (pg, 2128)). Our analytical results in Section 1 show that the $g^{high-income}$ are a particular way of weighting the moment conditions implied by the instruments $\{z_{lkt-1}\}_{k=1}^{K}$.

We focus on the TSLS estimate in column (6) of Table 3 of ADH, which reports that a $1,000 increase in import exposure per worker led to a decline in manufacturing employment of 0.60 percentage points. Our replication produces a coefficient of 0.62.[21]

5.2 Rotemberg weights

As in the canonical setting, despite a very large number of instruments ($794 = 397 \times 2$) the distribution of sensitivity is skewed so that in practice a small number of instruments get a large share of the weight: Table 5 shows that the top five instruments get just over a third of the absolute weight in the estimator ($^{0.391}/_{1.116}$). These instruments are electronic computers, games and toys, household audio and video, telephone apparatus and computer equipment (all in the 2000s) (recall that ADH update the shares but use shares that are lagged by one period so that the 2000 to 2007 changes are instrumented using industry shares measured in 1990).

The Table shows that one of the important comparisons in the estimator is differences in outcomes from 2000 to 2007 of places with high and low shares of electronic computer manufacturing (measured in 1990). So the estimate is sensitive to deviations from the identifying assumption in terms of the electronic computer industry (the sensitivity elasticity is 0.14).

Interestingly, the identifying assumption related to the computer industry is precisely one that ADH worry about. They write (pg. 2138) "Computers are another sector in which demand shocks may be correlated [across countries], owing to common innovations in the use of information technology," and so movements in manufacturing employment from 2000 to 2007 that are predicted by the 1990 employment share in electronic computers might not reflect the economic rise of China. To address this concern, the paper reports that dropping the industry entirely has minimal effects on their estimates. From our table, we can see that dropping the instrument "electronic computers in the 2000s" would have minimal effects on the point estimates because the $\hat{\beta}_k$ for the instrument (-0.62) is

[21] Column (1) of Table A1 reproduces ADH Table 3 column (6), and column (2) reports our replication. ADH update their industry shares over time which we argued in footnote 5 is undesirable. Column (3) shows that when we fix industry shares to the base period but let the controls be time-varying that the coefficient estimate dramatically attenuates and becomes a statistically insignificant -0.16. Column (4) also fixes the controls to 1980, but interacts the fixed controls with time. This specification produces a coefficient of -0.60, which is substantively nearly identical to our replication in column (2), but becomes statistically insignificant. The specification in column (4) is the one we use in Section 4 (and we used this specification in the July 2017 draft of this paper before we replicated ADH).

identical to the overall $\hat{\beta}$ (-0.62). This observation highlights the distinction, discussed formally in Proposition 2.3, between the sensitivity elasticity, which measures how responsive point estimates are to *changes* in misspecification, and the dropping instrument measure of sensitivity, which also depends on how far the $\hat{\beta}_k$ is from the overall $\hat{\beta}$.

More generally, our sensitivity elasticities allow us to formalize which industries "drive" the estimates. ADH (pg. 2138) write, "one may worry that the results are being driven by a handful of consumer goods industries in which China has assumed a commanding role. Dropping apparel, footwear, and textiles..." Our table shows precisely which industries—and in which time periods—are most important in driving the results.

Relative to the canonical setting, negative weights are less prominent and the variation in the national growth rates (or, imports from China to other high-income countries) explains more of the variation in the sensitivity elasticities. Even so, the g_k component only explains about thirty percent (0.545^2) of the variance of the Rotemberg weights.

5.3 Testing the plausibility of the identifying assumption

Test #1: Correlates of 1980 industry shares Table 6 shows the relationship between the covariates (used by ADH) and the top industries reported in Table 5. First, relative to the canonical setting and those controls, the controls explain less of the variation in shares. Second, the share of employment in manufacturing is never significant. By controlling for overall manufacturing share, the most "typical" manufacturing industries (i.e., those that covary the most with overall manufacturing share), get less weight. Third, electronic computers and computer equipment manufacturing are both concentrated in more educated areas with less routine employment (note that Table 4 shows that the inclusion of controls does not generate a statistically significant attenuation of coefficients).

Test #2: Overidentification tests Columns (7) and (8) of Table 4 show the point estimates from the Hausman et al. (2012) HFUL estimator and the Chao et al. (2014) overidentification tests. The results are similar to the canonical setting. Relative to the TSLS estimates using the Bartik instrument in columns (5) and (6), the HFUL estimator produces substantively larger point estimates. The key punchline of the overidentification tests (which can be approximately anticipated from looking at the dispersion in the $\hat{\beta}_k$ in Panels D and E of Table 2) is that the tests rejects the null that all instruments are exogenous under the assumption of constant treatment effects.

6 Empirical example III: simulated instruments

We now look at a simulated instrument (Currie and Gruber (1996a) and Currie and Gruber (1996b)) application. These papers are interested in understanding the effects of Medicaid expansions on various outcomes. We begin by discussing how to map the simulated instrument example to our setting. We then compute the sensitivity elasticities in the context of Cohodes et al. (2016), where the outcome is educational attainment.

6.1 A simulated instrument as a Bartik instrument

To make the discussion concrete, we begin by outlining the empirical approach in Cohodes et al. (2016), and discuss how we map their empirical specification into our notation.

To simplify their empirical strategy, they are interested in:

$$y_l = \beta x_l + \epsilon_l$$

where y_l is the educational attainment of people living in location l, and x_l is the average Medicaid eligibility of people living in location l. They are concerned that x_l is endogenous to the population's other characteristics (e.g., a poor state will have a high share of its population eligible for Medicaid), and so want to instrument for eligibility using just variation in the laws across states and time.

To write the simulated instrument in our notation, we first write the endogenous variable, x_l, as an inner product. Specifically, suppose that that there are K eligibility "types" indexed by k, z_{lk} is an indicator for whether eligibility type k is eligible in location l, and g_{lk} is the share of people living in location l who are of eligibility type k. Then $\{z_{lk}\}_{k=1}^{K}$ is a description of the legal environment in location l and we can write the endogenous variable in inner product form: $x_l = \sum_k z_{lk} g_{lk}$.[22]

The instrument is built by replacing the g_{lk}—the location shares of eligibility types—with g_k—the national shares of the eligibility types. The simulated instrument is then

$$S_l = \sum_k z_{lk} g_k,$$

and measures the share of people in location l who would be eligible if each location had the national distribution of eligibility types. The instrument only varies across locations based on laws, and not on the population characteristics. Thus, the results in Section 1 imply that the instrument is the cross-state differences in eligibility for each eligibility type.

[22]Relative to the examples so far there are a couple differences. First, $\sum_k z_{lk} \neq 1$, whereas previously the z_{lk} have summed to one. Second, $\sum_k g_{lk} = 1$, whereas previously the g_{lk} have not summed to one.

How do we determine the set of eligibility types? The laws for Medicaid eligibility across states take into account numerous factors including family income and household structure. We take a top-down approach to defining eligibility types. An eligibility type is a unique combination of eligibility across all 51 states (and 28 years used in the estimates) among the individuals in the 1986 CPS. Among the 48,036 children in the 1986 CPS, we find 18,881 distinct eligibility types, which vastly exceeds the number of observations. This explains why Currie and Gruber (1996b, pg. 446) describe a simulated instrument as "a convenient parameterization of legislative differences."

Cohodes et al. (2016) use 11 birth cohorts (born from 1980 to 1990) and the average eligibility of each cohort from age 0 to 17. They also separate whites and non-whites. So the instrument is defined at the level of eligibility type $\times$ cohort $\times$ age $\times$ white. Hence, there are 7,476,876 distinct instruments.

6.2 Rotemberg weights

Whereas in the canonical and ADH settings the top five instruments receive a quantitatively large share of the weight, here, because they account for a small share of the instruments, the top five instruments receive a small share of the weight. As a result, to understand what variation matters in point estimates, we project the instruments into lower-dimensional space along two dimensions. The first dimension is the state-year variation in laws that underlies the instruments.[23] The second dimension is the characteristics of households affected by the different instruments.[24]

Figure 1 summarizes the state-year eligibility changes that drive estimates, and hence the state-years a researcher should be looking for potential confounds. Because we are in changes, the first bar in the figure is for 1981. There are several things to note in the figure. First, the empirical strategy leverages variation that is spread throughout time: there is variation in eligibility that matters in the instrument throughout the 1980s and 1990s. Second, there are a few large spikes. The notable spike is in 1990, which generated changes in eligibility in all 51 states (and DC). This law change is the 1990 Federal budget which mandated coverage of children ages 6 through 18 in families with income at or below 100% of

[23]First, we sum the weights across birth-cohort $\times$ age $\times$ race, to have the weight for each eligibility type. Second, for each eligibility type we code the state-years in which Medicaid eligibility changed. (There are 1,372 distinct state-years (out of a total of $51 \times 27 = 1377$ possible state-years) where some eligibility type has a change in eligibility.) Third, so that the weights continue to sum to one, we divide the Rotemberg weight for the eligibility type by the number of state-year policy changes experienced by that eligibility type. Finally, we sum up the these normalized weights at the state-year level.

[24]First, we sum the weights across birth-cohort $\times$ age $\times$ race, to have the weight for each eligibility type. Second, we compute the average of various characteristics of the households in the 1986 CPS that make up each eligibility type. Finally, we sort eligibility types based on the characteristics and compute the sum of the Rotemberg weights within bins defined by the characteristics.

the federal poverty line (whether or not they were receiving AFDC assistance).[25] Third, the figure highlights the six largest state-law changes in terms of sensitivity to misspecification elasticity. These occur later in the sample. The bottom panels decompose the weights into those applying to lower- and "higher"-income families (where the cutoff is $10,000 1986 dollars). This decomposition shows that later law changes affect higher income households. A researcher should be interested if there are other things affecting higher-income families in Minnesota in 1997, Missouri in 1998, or New Jersey in 1999 as the estimates are sensitive to any confounding changes in these state-years.

Figure 2 summarizes which types of households get more weight in estimates. Hence, researchers should be concerned if there are other policies or trends in this period that differentially affect households with the characteristics that get weight. Panel A shows that the estimator places the most weight on families with 3 children. Panel B shows that the estimator generally weights lower income households more, except that there is some non-monotonicity: the poorest households are always eligible and so get no weight in estimates, and there are some higher-income households where variation matters more. Panel C shows the estimator places the most weight on variation that occurs at schooling ages (e.g. 5-16), with less weight in early childhood. This analysis complements the analysis in Cohodes et al. (2016, Table 5 and 6). The paper splits the sample based on age. It also complements the paper's controls for a natural confounder that would have a similar temporal pattern: the paper shows the robustness of their main estimates to controlling for school spending and the EITC.

7 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 formally equivalent to a GMM estimator with the industry shares as instruments. Hence, we argue that the identifying assumption is best stated in terms of the industry shares—the national growth rates are simply a weight matrix. Our second set of formal results relate 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 pursued a number of applications to illustrate what can be learned from our

[25] See `https://kaiserfamilyfoundation.files.wordpress.com/2008/04/5-02-13-medicaid-timeline.pdf` for a description of Medicaid law changes. Last accessed January 26, 2018.

results. While none of our applications produces a smoking gun that the particular use of a Bartik instrument is wrong, our results do 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.

References

Acemoglu, Daron and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119 (3):1049–1090.

Altonji, Joseph G. and David Card. 1991. "The Effects of Immigration on the Labor market Outcomes of Less-skilled Natives." In *Immigration, Trade and the Labor Market*, edited by John M. Abowd and Richard B. Freeman. University of Chicago Press, 201–234.

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):151–184.

Andrews, Isaiah. 2017. "On the Structure of IV Estimands." Working paper, MIT.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments." *Quarterly Journal of Economics* 132 (4):1553–1592.

Angrist, Joshua D. and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430):431–442.

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 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):2121–2168.

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

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.

———. Forthcoming. "In Search of Labor Demand." *American Economic Review* .

Blanchard, Olivier Jean and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 1992 (1):1–75.

Borusyak, Kirill and Xavier Jaravel. 2017. "Consistency and Inference in Bartik Research Designs." Working paper.

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.

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* 178:15–21.

Chernozhukhov, Victor and Christian Hansen. 2008. "The reduced form: A simple approach to inference with weak instruments." *Economics Letters* 100:68–71.

Christian, Paul and Christopher B. Barrett. 2017. "Revisiting the Effect of Food Aid on Conflict: A Methodological Caution." Working paper.

Cohodes, Sarah R., Daniel S. Grossman, Samuel A. Kleiner, and Michael F. Lovenheim. 2016. "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions." *Journal of Human Resources* 51 (3):727–759.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1):260–272.

Currie, Janet and Jonathan Gruber. 1996a. "Health Insurance Eligibility, Utilization, Medical Care and Child Health." *Quarterly Journal of Economics* 111 (2):431–466.

———. 1996b. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy* 104 (6):1263–1296.

Freeman, Richard B. 1980. "An Empirical Analysis of the Fixed Coefficient "Manpower Requirement" Mode, 1960-1970." *Journal of Human Resources* 15 (2):176–199.

Fuller, Wayne A. 1977. "Some Properties of a Modification of the Limited Information Estimator." *Econometrica* 45 (4):939–953.

Galle, Simon, Andres Rodriguez-Clare, and Moises Yi. 2017. "Slicing the pie: Quantifying the aggregate and distributional effects of trade." Working Paper 23737, NBER.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2015. "Do Credit Market Shocks affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times." Working paper.

Hall, Alistair R. 2005. *Generalized Method of Moments*. Oxford: Oxford University Press.

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* 3:211–255.

Hornbeck, Richard and Enrico Moretti. 2018. "Who Benefits From Productivity Growth? n The Direct and Indirect Effects of Local TFP Shocks." Working paper.

Hull, Peter. 2018. "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons." Working paper.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner. 2017. "Did Reality TV Really Cause a Decline in Teenage Childbearing? A Cautionary Tale of Evaluating Identifying Assumptions." Working paper.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact 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 Tade Forum* 2005:75–133.

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.

Kolesar, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and inference with many invalid instruments." *Journal of Business and Economic Statistics* 33 (4):474–484.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2017. "Commuting, Migration and Local Employment Elasticities." Working paper.

Nunn, Nathan and Nancy Qian. 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104 (6):1630–1666.

Oster, Emily. Forthcoming. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* :1–18.

Perloff, Harvey S. 1957. "Interrelations of State Income and Industrial Structure." *Review of Economics and Statistics* 39 (2):162–171.

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

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.

Table 1: OLS and IV estimates: canonical setting

	OLS		First Stage (Bartik)		Second Stage (Bartik)		Second Stage (HFUL)	
	Δ Wage (1)	Δ Wage (2)	Δ Emp (3)	Δ Emp (4)	Δ Wage (5)	Δ Wage (6)	Δ Wage (7)	Δ Wage (8)
Δ Emp	0.71	0.63			1.75	1.28	2.85	2.63
	(0.06)	(0.07)			(0.34)	(0.15)	(0.12)	(0.14)
Bartik (1980 shares)			0.90	0.94				
			(0.15)	(0.12)				
Year and CZone FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,166	2,166	2,166	2,166	2,166	2,166	2,166	2,166
R-squared	0.91	0.94	0.40	0.59	0.80	0.91		
Coefficient Equal p-value	0.09		0.80		0.14			
J-Statistic							697.16	861.44
P-value of J-Statistic							0.00	0.00

Notes: This table reports OLS and TSLS estimates of the inverse elasticity of labor supply. The regression are at the CZ level and the instruments are 3-digit industry-time periods (1980-1990, 1990-2000, and 2000-2010). The odd columns do not contain controls, while the even-numbered columns contain controls. Column (1) and (2) show the OLS estimates. Columns (3) and (4) show the first stage using the Bartik instruments. Columns (5) and (6) show the TSLS estimates using the Bartik instruments. Column (7) and (8) report estimates using the HFUL estimator of Hausman et al. (2012). The J-statistic comes from Chao et al. (2014). The R^2 is partial after absorbing location and time fixed effects. 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 3. Results are weighted by 1980 population. Standard errors are in parentheses.

Table 2: Summary of Rotemberg weights: canonical setting

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.874	-0.003	0.412		
Positive	1.874	0.005	0.588		
Panel B: Correlations					
	α_k	g_k	$\hat{\beta}_k$	Var(z_k)	
$\hat{\alpha}_k$	1				
g_k	0.022	1			
β_k	-0.002	0.043	1		
Var(z_k)	0.418	-0.100	-0.012	1	
Panel C: Variation across years in α_k					
	Sum	Mean			
1980	0.457	0.002			
1990	0.182	0.001			
2000	0.360	0.002			
Panel D: Top five Rotemberg weight industries					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	β_k 95% CI	Ind Share
Oil + Gas Extraction, 2000	0.156	0.080	1.169	(0.75, 2.20)	1.590
Other, 1990	0.098	-0.033	0.752	(0.40, 1.55)	4.808
Motor Vehicles, 2000	0.093	-0.030	1.351	(1.20, 1.55)	4.973
Guided Missiles, 1980	0.081	0.100	0.174	(-0.30, 0.55)	1.067
Motor Vehicles, 1990	0.077	0.031	1.607	(1.35,2.20)	4.973
Panel E: Summary of $\hat{\beta}_k$					
	Mean	Median	25th perc.	75th perc.	Share Negative
β_k	1.637	0.775	2.181	-0.609	0.349

Notes: This table reports statistics about the Rotemberg weights. Panel A reports the share and sum of negative weights. Panel B reports correlations between the weights, as well as the national component of growth (g_k), the just-identified coefficient estimates, and the variation in the industry shares across locations (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, β_k is the coefficient from the just-identified regression, the 95% 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 the dispersion in the $\hat{\beta}_k$. The "Other" industry is the "N/A" code in the IND1990 classification system.

Table 3: Relationship between industry shares and characteristics: canonical setting

	Other	Oil + gas Extraction	Motor Vehicles	Guided Missiles	Bartik (1980-1990)
Male	98.73	40.94	-25.86	36.14	-0.64
	(14.91)	(8.39)	(7.61)	(18.38)	(0.05)
White	-34.41	1.07	-23.61	14.85	-0.08
	(7.66)	(3.09)	(28.37)	(14.84)	(0.06)
Native Born	21.25	10.96	-3.92	-42.83	-0.19
	(6.95)	(2.94)	(5.38)	(41.64)	(0.04)
12th Grade Only	26.75	-32.51	64.36	-68.41	0.35
	(8.88)	(7.53)	(17.72)	(27.39)	(0.06)
Some College	24.12	-9.51	23.24	28.93	0.63
	(7.12)	(4.63)	(22.71)	(25.74)	(0.07)
Veteran	-142.65	-10.57	19.09	86.93	0.55
	(22.65)	(7.05)	(41.04)	(43.69)	(0.12)
# of Children	-57.66	-2.75	45.35	11.33	0.27
	(13.46)	(4.62)	(23.48)	(21.27)	(0.06)
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes
N	722	722	722	722	722
R^2	0.43	0.22	0.11	0.25	0.58
F	8.16	5.40	8.50	2.93	71.00
p	0.00	0.00	0.00	0.00	0.00

Notes: Each column reports results of a single regression of a 1980 industry share on 1980 characteristics. Each characteristic is standardized to have unit standard deviation. 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.

Table 4: OLS and IV estimates: Autor, Dorn, and Hanson (2013)

	OLS		First Stage (Bartik)		Second Stage (Bartik)		Second Stage (HFUL)	
	Δ Mfg (1)	Δ Mfg (2)	China to US (3)	China to US (4)	Δ Mfg (5)	Δ Mfg (6)	Δ Mfg (7)	Δ Mfg (8)
China to US	-0.38	-0.08			-0.68	-0.62	-1.14	-1.12
	(0.06)	(0.03)			(0.10)	(0.13)	(0.05)	(0.04)
China to Other			0.49	0.39				
			(0.05)	(0.04)				
Year and Census Division FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,444	1,444	1,444	1,444	1,444	1,444	1,444	1,444
R-squared	0.25	0.55	0.46	0.55	0.18	0.33		
Coefficient Equal p-value	0.00		0.00		0.26			
J-Statistic							967.82	969.34
P-value of J-Statistic							0.00	0.00

Notes: This table reports OLS and TSLS estimates of the inverse elasticity of labor supply. The regression are at the CZ level. The odd-numbered columns show estimates without controls and the even-numbered columns show estimates with controls Column (1) and (2) show the OLS estimates. Columns (3) and (4) show the first stage using the Bartik instrument. Columns (5) and (6) show the TSLS estimates using the Bartik instrument. Column (7) and (8) report estimates using the HFUL estimator of Hausman et al. (2012). The J-statistic comes from Chao et al. (2014). The R^2 is partial after absorbing location and time fixed effects. 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 6. Results are weighted by start of period population. Standard errors are in parentheses.

Table 5: Summary of Rotemberg weights: Autor, Dorn, and Hanson (2013)

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.116	-0.000	0.381		
Positive	1.116	0.002	0.619		
Panel B: Correlations					
	α_k	g_k	β_k	Var(z_k)	
α_k	1				
g_k	0.545	1			
β_k	0.002	0.007	1		
Var(z_k)	-0.023	-0.073	0.013	1	
Panel C: Variation across years in α_k					
	Sum	Mean			
1990	0.239	0.001			
2000	0.761	0.002			
Panel D: Top five Rotemberg weight industries					
	α_k	g_k	β_k	β_k 95% CI	Ind Share
Electronic Computers, 2000	0.140	189.12	-0.620	(-1.59, -0.02)	1.091
Games and Toys, 2000	0.098	320.64	-0.179	(-4.00, 3.00)	0.270
Household Audio and Video, 2000	0.055	218.22	-0.147	(-0.68, 0.13)	0.378
Telephone Apparatus, 2000	0.051	94.58	-0.308	(-4.00, 3.00)	0.920
Computer Equipment, 2000	0.047	41.68	-0.232	(-4.00, 3.00)	0.519
Panel E: Summary of $\hat{\beta}_k$					
	Mean	Median	25th perc.	75th perc.	Share Negative
β_k	-0.909	-0.514	0.734	-1.687	0.633

Notes: This table reports statistics about the Rotemberg weights. Panel A reports the share and sum of negative Rotemberg weights. Panel B reports correlations between the weights, as well as the national component of growth (g_k), the just-identified coefficient estimates, and the variation in the industry shares across locations (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, β_k is the coefficient from the just-identified regression, the 95% confidence interval is the weak instrument robust confidence interval using the method from Chernozhukhov and Hansen (2008) over a range from -4 to 3, and Ind Share is the industry share (multiplied by 100 for legibility). Panel E reports statistics about the dispersion in the $\hat{\beta}_k$.

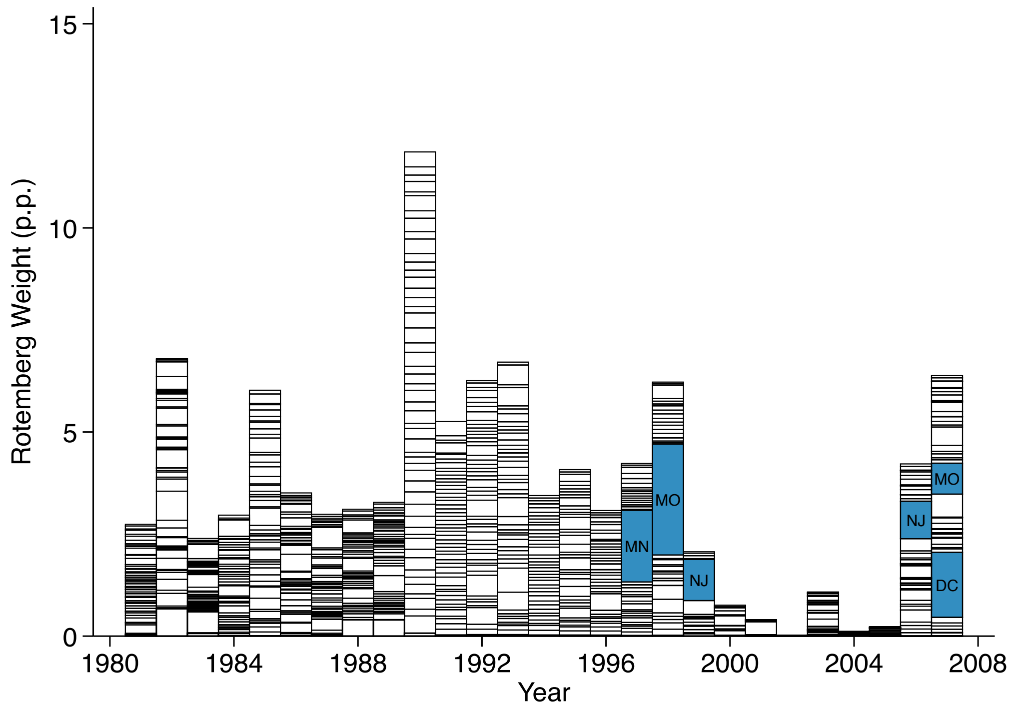
Table 6: Relationship between industry shares and characteristics: Autor, Dorn, and Hanson (2013)

	Electronic Computers	Computer Equipment	Household Audio and Video	Telephone Apparatus	Games and Toys
Share Empl in Manufacturing	0.21	0.21	0.08	-0.07	0.01
	(0.18)	(0.15)	(0.08)	(0.06)	(0.03)
Share College Educated	0.20	0.22	0.01	-0.07	-0.08
	(0.11)	(0.10)	(0.04)	(0.06)	(0.03)
Share Foreign Born	-0.01	-0.01	-0.02	-0.08	0.01
	(0.04)	(0.04)	(0.01)	(0.03)	(0.01)
Share Empl of Women	-0.04	-0.02	-0.08	-0.02	0.05
	(0.12)	(0.12)	(0.05)	(0.07)	(0.03)
Share Empl in Routine	-0.37	-0.36	0.06	-0.01	0.04
	(0.14)	(0.12)	(0.05)	(0.07)	(0.03)
Avg Offshorability	0.33	0.29	0.00	0.23	0.02
	(0.10)	(0.08)	(0.05)	(0.04)	(0.02)
N	1,444	1,444	1,444	1,444	1,444
R^2	0.08	0.08	0.01	0.05	0.02
F	3.36	3.57	0.99	5.80	3.53
p	0.01	0.01	0.44	0.00	0.01

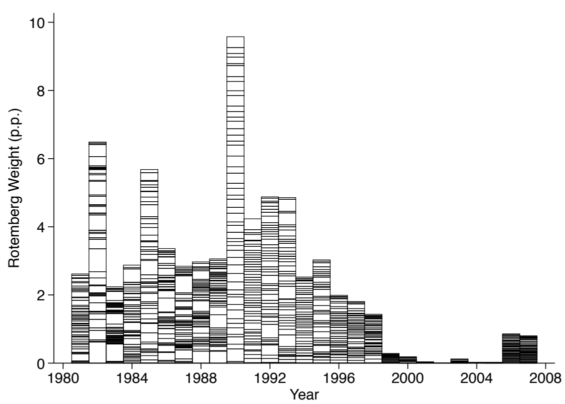
Notes: Each column reports a separate regression. The regressions are two pooled cross-sections, where one cross section is 1980 shares on 1990 characteristics, and one is 1990 shares on 2000 characteristics. Each characteristic is standardized to have unit standard deviation. Results are weighted by the population in the period the characteristics are measured. Standard errors in parentheses.

Figure 1: Rotemberg weights by state-year policy changes and income

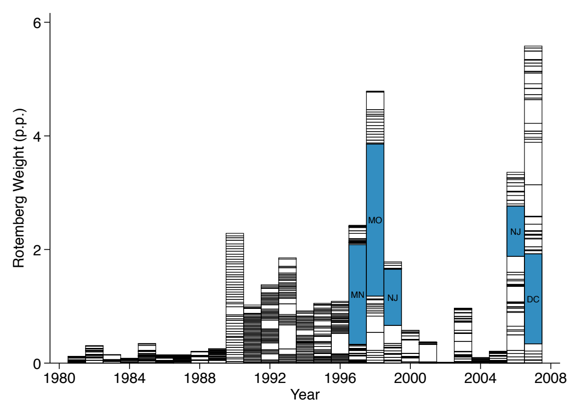
Panel A: Overall



Panel B: Low-income ($<$ \$10,000 in 1986 USD)



Panel C: High-income ($\geq$ \$10,000 in 1986 USD)



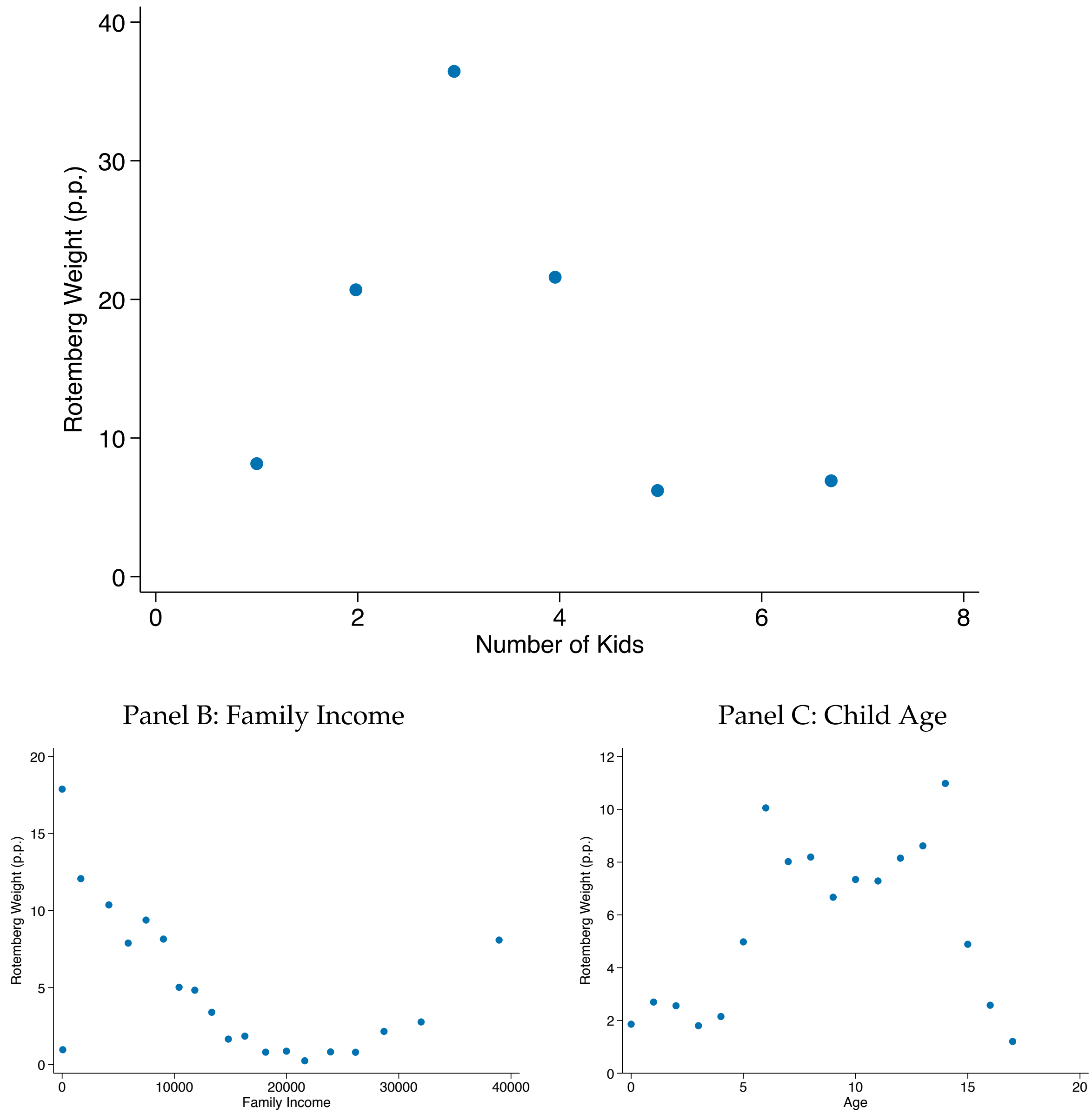
Notes: These figures report the state-year policy changes to which the estimates are most sensitive to misspecification. The figure is constructed by weighting the state-year Medicaid eligibility *changes* that go into the definition of eligibility types. The state-year Medicaid eligibility changes are weighted by the number of such changes experienced by each eligibility type so that the bars in Panel (a) sum to one (if an eligibility type experiences more state-year eligibility changes, then each state-year eligibility change it experiences receives less weight in the figure). Because the figures represent policy *changes* the first year is 1981. Within each year, each of the rectangles indicates a single state. In Panel (a), the bars sum to one. Panels (b) and (c) split the eligibilility types by their mean income and so the combination of Panels (b) and (c) sum to Panel (a).

Figure 2: Rotemberg weights by charactestics

Panel A: Number of Children

Panel B: Family Income

Panel C: Child Age



Notes: These figures report the characteristics of the eligibility types to which the estimates are most sensitive to misspecification. Each of the three panels reports computing the average characteristics of each eligibility type, sorting the eligibility types according to the characteristic, and then reporting bin sums of the Rotemberg weights. Thus, in each figure the weights sum to one.

A Instruments encompassed by our structure

We now discuss three other instruments that our encompassed by our framework. This list cannot be exhaustive, but illustrates the widespread applicability of our results.

A.1 Immigrant enclave instrument

Altonji and Card (1991) are interested in the effects of immigration on native wages, but are concerned that the correlation between immigrant inflows and local economic conditions may confound their estimates. To fit our notation, let x_l denote the number of newly arriving immigrants in location l in a given interval. Let k denote one of K countries of origin and let z_{lk} denote the share of people arriving from origin country k living in location l. Hence, $\sum_{l=1}^{L} z_{lk} = 1, \forall k$. In contrast, in the industry-location setting it is the sum over k that sums to one. Let g_k denote the number of people arriving from origin k. The instrument comes from lagging the z_{lk}. Once we lag z, say z_{lk0} in some initial period, then let i_{lk} be the number of immigrants from origin country k arriving in destination l. Then define $g_{lk} = \frac{i_{lk}}{z_{lk0}}$ to be the hypothetical flow of immigrants from k that would have to have occurred to have generated the extent of flows; this allows us to write $x_l = \sum_k z_{lk0} g_{lk}$. Then rather than using the g_{lk} that makes this an identity, the researcher uses $g_k = \sum_l i_{lk} = \sum_l g_{lk} z_{lk0}$. (This is analogous to in the industry-location setting weighting the g_{lk} by the z_{lk} to compute the g_k, rather than equal-weighting across locations).

A.2 Bank lending relationships

Greenstone, Mas, and Nguyen (2015) are interested in the effects of changes in bank lending on economic activity during the Great Recession. They observe county-level outcomes and loan origination by bank to each county. In our notation, let x_l be credit growth in a county, let z_{lk} be the share of loan origination in county l from bank k in some initial period, and let g_{lk} be the growth in loan origination in county l by bank k over some period. Then $x_l = \sum_k z_{lk} g_{lk}$.

The most straightforward Bartik estimator would compute $\hat{g}_{-l,k} = \frac{1}{L-1} \sum_{l' \neq l} g_{l'k}$. However, Greenstone, Mas, and Nguyen (2015) are concerned that there is spatial correlation in the economic shocks and so leave-one-out is not enough to remove mechanical correlations. One approach would be to instead leave out *regions*. Instead, they pursue a generalization of this approach and regress:

$$g_{lk} = g_l + g_k + \epsilon_{lk}, \tag{A1}$$

where the g_l and g_k are indicator variables for location and bank. Then the $\hat{g}_l$ captures the change in bank lending that is common to a county, while $\hat{g}_k$ captures the change in bank lending that is common to a bank. To construct their instrument, they use $\hat{B}_l = \sum_k z_{lk} \hat{g}_k$, where the $\hat{g}_k$ comes from equation (A1).

A.3 Market size and demography

Acemoglu and Linn (2004) are interested in the effects of market size on innovation. Naturally, the concern is that the size of the market reflects both supply and demand factors: a good drug will increase consumption of that drug. To construct an instrument, their basic observation is that there is an age structure to demand for different types of pharmaceuticals and there are large shifts in the age structure in the U.S. in any sample. They use this observation to construct an instrument for the change in market size.

In our notation, z_{lk} is the share of spending on drug category l that comes from age group k. Hence, $\sum_k z_{lk} = 1$. Then g_{lk} is the growth in spending of age group k on drug category l. Hence, $x_l = \sum_k z_{lk} g_{lk}$. To construct an instrument, they use the fact that there are large shifts in the age distribution. Hence, they estimate $\hat{g}_k$ as the increase in the *number of people* in age group k, and sometimes as the *total income* (people times incomes) in age group k. This instrument is similar to the "China shock" setting where for both conceptual and data limitation issues g_{lk} is fundamentally unobserved and so the researcher constructs $\hat{g}_k$ using other information.

B Omitted proofs

Proposition 1.1

Proof. Let a bar over a vector denote the mean of the vector times the all ones vector of the appropriate dimension.[26]

$$\begin{aligned}\hat{\beta}_{GMM} &= \frac{X^{\perp\prime}(Z-\bar{Z})(G-\bar{G})(G-\bar{G})'(Z-\bar{Z})'Y^{\perp}}{X^{\perp\prime}(Z-\bar{Z})(G-\bar{G})(G-\bar{G})'(Z-\bar{Z})'X^{\perp}} \\ &= \frac{X^{\perp\prime}(B-\bar{B})(B-\bar{B})'Y}{X^{\perp\prime}(B-\bar{B})(B-\bar{B})'X^{\perp}} \\ &= \hat{\beta}_{Bartik},\end{aligned}$$

where $B-\bar{B} = (Z-\bar{Z})(G-\bar{G})$ because $(Z-\bar{Z})G = ZG - \bar{Z}G = B - \bar{B}$, $(Z-\bar{Z})\bar{G} = \mathbf{0}$ (an $L \times 1$ vector of zeros), and $X^{\perp\prime}(B-\bar{B})$ is a scalar and so cancels. □

Proposition 2.1

Proof. The proof is algebra:

$$\hat{\alpha}_k(\hat{W})\hat{\beta}_k = \frac{\hat{c}_k(\hat{W})Z_k'X^{\perp}}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^{\perp}}(Z_k'X^{\perp})^{-1}Z_k'Y^{\perp} = \frac{\hat{c}_k(\hat{W})Z_k'Y^{\perp}}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^{\perp}} \tag{A1}$$

$$\sum_{k=1}^K \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \frac{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'Y^{\perp}}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^{\perp}} \tag{A2}$$

$$= \frac{\hat{C}(\hat{W})'Z'Y^{\perp}}{\hat{C}(\hat{W})'Z'X^{\perp}}. \tag{A3}$$

□

[26] In the case of $\bar{Z}$, which is a matrix, these are the column means (i.e., the mean share of industry k across locations).

Proof of Lemma 2.1.

Proof. Note that

$$\hat{\alpha}_k(\hat{W}) = \frac{X^{\perp\prime} Z \hat{W}_k Z_k' X^{\perp}}{X^{\perp\prime} Z \hat{W} Z' X^{\perp}} \tag{A4}$$

$$= \frac{\left(\sum_{l,t} x_{lt}^{\perp} Z_{lt}\right) \hat{W}_k \left(\sum_{l,t} z_{lkt} x_{lt}^{\perp}\right)}{\left(\sum_{l,t} x_{lt}^{\perp} Z_{lt}\right) \hat{W} \left(\sum_{l,t} Z_{lt} x_{lt}^{\perp}\right)}. \tag{A5}$$

Since our data is i.i.d. and the variance of $x_{lt}^{\perp} Z_{lt}$ is bounded, the law of large numbers holds as $L \to \infty$. □

Proof of Proposition 2.2

Proof. (i) Note that $\sum_{k=1}^{K} \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \hat{\beta}$ is the standard GMM estimator using weight matrix $\hat{W}$. Since the standard GMM estimator is a consistent estimator for any positive semi-definite matrix given Assumption 1 and $\kappa = 0$, $\sum_{k=1}^{K} \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \beta_0$.

Next, note that

$$\begin{aligned}
\beta_0 &= \frac{\Sigma_{ZX^{\perp}} W \mathbb{E}[Z_{lt} Y_{lt}^{\perp}]}{\Sigma_{ZX^{\perp}} W \Sigma_{ZX^{\perp}}'} \\
&= \sum_k \frac{\Sigma_{ZX^{\perp}} W_k \mathbb{E}[z_{lkt} Y_{lt}^{\perp}]}{\Sigma_{ZX^{\perp}} W \Sigma_{ZX^{\perp}}'} \\
&= \sum_k \frac{\Sigma_{ZX^{\perp}} W_k \Sigma_{ZX^{\perp},k}}{\Sigma_{ZX^{\perp}} W \Sigma_{ZX^{\perp}}'} \frac{\mathbb{E}[z_{lkt} Y_{lt}^{\perp}]}{\Sigma_{ZX^{\perp},k}} \\
&= \sum_{k=1}^{K} \alpha_k(W)\beta_0.
\end{aligned}$$

Finally, since the standard necessary assumptions for asymptotic normality of the GMM estimator are satisfied (e.g., Hall (2005, Theorem 3.2, pg. 71)), $\tilde{\beta}$ is normally distributed with mean zero. Hence, $\mathbb{E}[\tilde{\beta}] = 0$.

(ii) First, note that

$$\begin{aligned}
\hat{\beta}_k &= \frac{\sum_{l,t} z_{lkt} y_{lt}^{\perp}}{\sum_{l,t} z_{lkt} x_{lt}^{\perp}} = \beta_0 + \frac{\sum_{l,t} z_{lkt}(L^{-1/2} V_{lt} + \epsilon_{lt})}{\sum_{l,t} z_{lkt} x_{lt}} \\
\hat{\beta}_k - \beta_0 &= L^{-1/2} \frac{\sum_{l,t} z_{lkt} V_{lt}}{\sum_{l,t} z_{lkt} x_{lt}} + \frac{\sum_{l,t} z_{lkt} \epsilon_{lt}}{\sum_{l,t} z_{lkt} x_{lt}}.
\end{aligned}$$

The second term goes to zero because $\mathbb{E}[z_{lkt}\epsilon_{lt}] = 0$. The first term goes to zero as $L \to \infty$. Finally, since our summand terms have bounded variance, the law of large numbers holds. A similar argument holds for the broader summand.

The asymptotic bias of $\tilde{\beta}_k$ follows from Proposition 3 of AGS. A sketch of the proof for this case follows:

$$\sqrt{L}(\hat{\beta}_k - \beta_0) = \frac{\sum_{l,t} z_{lkt} V_{lt}}{\sum_{l,t} z_{lkt} x_{lt}} + \sqrt{L}\frac{\sum_{l,t} z_{lkt}\epsilon_{lt}}{\sum_{l,t} z_{lkt} x_{lt}}$$
$$\sqrt{L}(\hat{\beta}_k - \beta_0) - \frac{\sum_{l,t} z_{lkt} V_{lt}}{\sum_{l,t} z_{lkt} x_{lt}} = \sqrt{L}\frac{\sum_{l,t} z_{lkt}\epsilon_{lt}}{\sum_{l,t} z_{lkt} x_{lt}}.$$

Since $\frac{\sum_{l,t} z_{lkt} V_{lt}}{\sum_{l,t} z_{lkt} x_{lt}}$ converges to $\frac{\Sigma_{ZV,k}}{\Sigma_{ZX^{\perp},k}}$, this implies that $\sqrt{L}(\hat{\beta}_k - \beta_0)$ converges in distribution to a normally distributed random variable $\tilde{\beta}_k$ with $\mathbb{E}[\tilde{\beta}_k] = \frac{\Sigma_{ZV,k}}{\Sigma_{ZX^{\perp},k}}$. Finally, since $\hat{\alpha}_k(\hat{W})$ converges in probability to $\alpha_k(W)$, by a similar argument this implies that $\sqrt{L}(\hat{\beta} - \beta_0)$ converges in distribution to a normally distributed random variable $\tilde{\beta}$ with $\mathbb{E}[\tilde{\beta}] = \sum_k \alpha_k(W)\frac{\Sigma_{ZV,k}}{\Sigma_{ZX^{\perp},k}} = \sum_k \alpha_k(W)\mathbb{E}[\tilde{\beta}_k]$.

□

Proposition 2.3

Proof. Consider the difference in the bias for the two estimators:

$$\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right] = \sum_{k'} \alpha_{k'}(W)\mathbb{E}[\tilde{\beta}_{k'}] - \sum_{k' \neq k} \alpha_{k'}(W_{-k})\mathbb{E}[\tilde{\beta}_{k'}] \tag{A6}$$
$$= \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] + \sum_{k' \neq k} (\alpha_{k'}(W) - \alpha_{k'}(W_{-k}))\mathbb{E}[\tilde{\beta}_{k'}]. \tag{A7}$$

Now, consider $\alpha_{k'}(W) - \alpha_{k'}(W_{-k})$. If $W = GG'$, then $C(W) = GB'X^{\perp}$ and $\alpha_{k'}(W) = \frac{g_{k'} Z_{k'} X^{\perp}}{\sum_{k'} g_{k'} Z_{k'} X^{\perp}}$. If $W_{-k} = G_{-k}G'_{-k}$, then $\alpha_{k'}(W_{-k}) = \frac{g_{k'} Z_{k'} X^{\perp}}{\sum_{k' \neq k} g_{k'} Z_{k'} X^{\perp}}$, or $\alpha_{k'}(W_{-k}) = \alpha_{k'}(W)/(1 - \alpha_k(W))$.[27] This gives:

$$\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right] = \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] + \sum_{k' \neq k} \left(\alpha_{k'}(W) - \frac{\alpha_{k'}(W)}{1 - \alpha_k(W)}\right)\mathbb{E}[\tilde{\beta}_{k'}] \tag{A8}$$
$$= \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] - \frac{\alpha_k(W)}{1 - \alpha_k(W)} \sum_{k' \neq k} (\alpha_{k'}(W))\,\mathbb{E}[\tilde{\beta}_{k'}]. \tag{A9}$$

□

C Equivalence with K industries, T locations, and controls

The two stage least squares system of equations is:

$$y_{lt} = D_{lt}\alpha + x_{lt}\beta + \epsilon_{lt} \tag{A1}$$
$$x_{lt} = D_{lt}\tau + B_{lt}\gamma + \eta_{lt}, \tag{A2}$$

[27] Note that with 2SLS, these results would not hold, as the estimates for the first stage parameters after dropping an industry would be different.

where D_{lt} is a $1 \times L$ vector of controls. Typically in a panel context, D_{lt} will include location and year fixed effects, while in the cross-sectional regression, this will simply include a constant. It may also include a variety of other variables. Let $n = L \times T$, the number of location-years. For simplicity, let Y denote the $n \times 1$ stacked vector of y_{lt}, $\mathbf{D}$ denote the $n \times L$ stacked vector of D_{lt} controls, X denote the $n \times 1$ stacked vector of x_{lt}, G the stacked $K \times T$ vector of the g_{kt}, and B denote the stacked vector of B_{lt}. Denote $\mathbf{P_D} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ as the $n \times n$ projection matrix of $\mathbf{D}$, and $\mathbf{M_D} = \mathbf{I}_n - \mathbf{P_D}$ as the annhilator matrix. Then, because this is an exactly identified instrumental variable our estimator is

$$\hat{\beta}_{Bartik} = \frac{(\mathbf{M_D}B)'Y}{(\mathbf{M_D}B)'X}. \tag{A3}$$

We now consider the alternative approach of using industry shares as instruments. The two-equation system is:

$$y_{lt} = D_{lt}\alpha + x_{lt}\beta + \epsilon_{lt} \tag{A4}$$

$$x_{it} = D_{lt}\tau + Z_{lt}\gamma_t + \eta_{lt}, \tag{A5}$$

where Z_{lt} is a $1 \times K$ row vector of industry shares, and γ_t is a $K \times 1$ vector, and, reflecting the lessons of previous section, the t subscript allows the effect of a given industry share to be time-varying. In matrix notation, we write

$$Y = \mathbf{D}\alpha + X\beta + \epsilon \tag{A6}$$

$$X = \mathbf{D}\tau + \tilde{Z}\mathbf{\Gamma} + \eta, \tag{A7}$$

where $\mathbf{\Gamma}$ is a stacked $1 \times (T \times K)$ row vector such that

$$\mathbf{\Gamma} = [\gamma_1 \cdots \gamma_T], \tag{A8}$$

and $\tilde{\mathbf{Z}}$ is a stacked $n \times (T \times K)$ matrix such that

$$\tilde{\mathbf{Z}} = \left[\begin{array}{ccc} \mathbf{Z} \odot 1_{t=1} & \cdots & \mathbf{Z} \odot 1_{t=T} \end{array} \right], \tag{A9}$$

where $1_{t=t'}$ is an $n \times K$ indicator matrix equal to one if the nth observation is in period t', and zero otherwise. $\odot$ indicates the Hadamard product, or pointwise product of the two matrices. Let $\mathbf{Z}^{\perp} = \mathbf{M_D}\tilde{\mathbf{Z}}$ and $\mathbf{P}_{Z^{\perp}} = \tilde{\mathbf{Z}}^{\perp}(\tilde{\mathbf{Z}}^{\perp\prime}\tilde{\mathbf{Z}}^{\perp})^{-1}\tilde{\mathbf{Z}}^{\perp\prime}$. Then, the TSLS estimator is

$$\hat{\beta}_{TSLS} = \frac{X'\mathbf{P}_{Z^{\perp}}Y}{X'\mathbf{P}_{Z^{\perp}}X}. \tag{A10}$$

Alternatively, using the $\mathbf{Z}$ as instruments, the GMM estimator is:

$$\hat{\beta}_{GMM} = \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{A11}$$

where Ω is a $(K \times T) \times (K \times T)$ weight matrix.

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

Proof. Start with the Bartik estimator,

$$\hat{\beta}_{Bartik} = \frac{(\mathbf{M_D}B)'Y}{(\mathbf{M_D}B)'X} \tag{A12}$$

$$= \frac{B'\mathbf{M_D}Y}{B'\mathbf{M_D}X} \tag{A13}$$

$$= \frac{G'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{G'\tilde{\mathbf{Z}}'\mathbf{M_D}X} \tag{A14}$$

$$= \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{A15}$$

where the second equality is algebra, the third equality follows from the definition of B, and the fourth equality follows because $X'\mathbf{M_D}\tilde{\mathbf{Z}}G$ is a scalar. By inspection, if $\Omega = GG'$, then $\hat{\beta}_{GMM} = \hat{\beta}_{Bartik}$. □

D When the first-stage coefficient is one

PROPOSITION D.1. *Let G_l be the $K \times 1$ vector of industry-location growth rates in location l and let Z_l be the $1 \times K$ row vector of industry shares in location l. Suppose that G_l and Z_l are independent. Then $\mathbb{E}[G_l|Z_l] = \mathbb{E}[G_l]$ and the expectation of the first stage coefficient from using the Bartik instrument is 1. (For notational simplicity we suppress notation that residualizes for controls.)*

Proof. Note that we can write $G_l = G + \tilde{G}_l$ where G is the vector of national growth rates and $\tilde{G}_l$ is a $K \times 1$ vector made up of $\tilde{g}_{lk}$ (we suppress the location term for simplicity). Similarly, $B_l = Z_l G$. Hence, the population expression is:

$$Var(X_l) = Var(Z_l G_l) = Var(Z_l G + Z_l \tilde{G}_l) \tag{A1}$$

$$= Var(Z_l G) + 2Cov(Z_l G, Z_l \tilde{G}) + Var(Z_l \tilde{G}), \tag{A2}$$

The probability limit of the first-stage coefficient is then:

$$plim_{L\to\infty}\gamma = \frac{Cov(B_l, X_l)}{Var(B_l)} = 1 + \frac{Cov(Z_l G, Z_l \tilde{G})}{Var(Z_l G)}. \tag{A3}$$

Hence, whether the first stage coefficient is 1 depends on the properties of $Cov(Z_l G, Z_l \tilde{G}_l)$. We now show that a sufficient condition for $Cov(Z_l G, Z_l \tilde{G}_l) = 0$ is that $\mathbb{E}[G_l|Z_l] = \mathbb{E}[G_l]$.

$$Cov(Z_l G, Z_l \tilde{G}_l) = \mathbb{E}[Z_l G Z_l \tilde{G}_l] - \mathbb{E}[Z_l G]\mathbb{E}[Z_l \tilde{G}_l] \tag{A4}$$

$$= \mathbb{E}[Z_l G Z_l (G_l - G)] - \mathbb{E}[Z_l G]\mathbb{E}[Z_l (G_l - G)] \tag{A5}$$

$$= \mathbb{E}[Z_l G (G_l - G)' Z_l] - \mathbb{E}[Z_l G]\mathbb{E}[Z_l (G_l - G)] \tag{A6}$$

$$= \mathbb{E}[Z_l \mathbb{E}[G(G_l - G)'|Z_l] Z_l] - \mathbb{E}[Z_l \mathbb{E}[G|Z_l]]\mathbb{E}[Z_l \mathbb{E}[(G_l - G)]|Z_l] \tag{A7}$$

$$= \mathbb{E}[Z_l \mathbb{E}[G(G_l - G)'] Z_l] - \mathbb{E}[Z_l]\mathbb{E}[G]\mathbb{E}[Z_l]\mathbb{E}[(G_l - G)]] \tag{A8}$$

$$= 0. \tag{A9}$$

The first line is the definition of covariance, the second line is the definition of $\tilde{G}_l$, the third line takes the transpose of a scalar, the fourth line is the law of iterated expectations, the fifth line is the assumption that G and Z are independent, and the sixth follows from the fact that $\mathbb{E}[G_l - G] = 0$ and $Cov(G, G_l - G) = 0$. □

E An economic model

We consider L independent locations indexed by l. Labor is homogeneous so that the wage in location l in period t is w_{lt}. The labor supply curve in location l in period t is:

$$\ln N_{lt}^S = \sigma_{lt} + \theta \ln w_{lt}. \tag{A1}$$

Here, N_{lt}^S is the quantity of labor supplied and σ_{lt} is a location-period-specific shifter of the level of labor supply. The local labor supply elasticity, θ, is the parameter of interest and is common across industries and locations.

The demand curve for industry k in location l at time t is given by

$$\ln N_{lkt}^D = T_{lk}\alpha_{lkt} - \phi \ln w_{lt}. \tag{A2}$$

Here, N_{lkt}^D is the quantity of labor demanded, T_{lk} is a fixed factor that generates persistent differences in industry composition, α_{lkt} is the time-varying industry-location level of labor demand, and ϕ is the common elasticity of local labor demand. Letting $\alpha_{lt} = \ln\left(\sum_k \exp\{T_{lk}\alpha_{lkt}\}\right)$ be the aggregated location-specific shifter of labor demand, the location-level demand curve is:

$$\ln N_{lt}^D = \alpha_{lt} - \phi \ln w_{lt}. \tag{A3}$$

The equilibrium condition in market l in period t is a labor market clearing condition: $N_{lt} = N_{lt}^S = \sum_k N_{lkt}^D = N_{lt}^D$. We let $\tilde{x}_t = \ln x_t$ and dx_t be the per-period change in x_t.

To construct the infeasible Bartik instrument, write the change in log employment in an industry-location, and then label the components of this decomposition in the same notation as the previous section:[28]

$$d\tilde{N}_{lkt} = \underbrace{d\alpha_{kt}}_{g_{kt}} - \underbrace{\left(\frac{\phi}{\theta+\phi}d\alpha_{lt} - \frac{\phi}{\theta+\phi}d\sigma_{lt}\right)}_{g_{lt}} + \underbrace{T_{lk}d\alpha_{lkt} - d\alpha_{kt}}_{\tilde{g}_{lkt}}.$$

Define $z_{lk0} \equiv \frac{exp(T_{lk}\alpha_{lk0})}{\sum_{k'} exp(T_{lk'}\alpha_{lk'0})}$ to be the industry shares in period 0.[29] Then the infeasible Bartik instrument that isolates the industry component of the innovations to demand shocks is $B_{lt} = \sum_k z_{lk0} d\alpha_{kt}$.

[28]Combine equation (A1) and (A3) to have the following equilibrium wage equation: $\ln w_{lt} = \frac{1}{\theta+\phi}\alpha_{lt} - \frac{1}{\theta+\phi}\sigma_{lt}$. Then substitute in to equation (A2) for the equilibrium wage, take differences, and add and subtract a $d\alpha_{kt}$.

[29]Note that $\frac{N_{lkt}^D}{N_{lt}^D} = \frac{exp(T_{lk}\alpha_{lkt} - \phi \ln w_{lt})}{exp(\alpha_{lt} - \phi \ln w_{lt})} = \frac{exp(T_{lk}\alpha_{lkt})}{exp(\alpha_{lt})} = \frac{exp(T_{lk}\alpha_{lkt})}{exp(\ln(\sum_k \exp\{T_{lk}\alpha_{lkt}\}))} = \frac{exp(T_{lk}\alpha_{lkt})}{\sum_k \exp\{T_{lk}\alpha_{lkt}\}}$.

In differences and with only two time periods, the equation we are interested in estimating is:

$$(d\tilde{w}_{lt+1} - d\tilde{w}_{lt}) = (\tau_{t+1} - \tau_t) + \beta(d\tilde{N}_{lt+1} - d\tilde{N}_{lt}) + (\epsilon_{lt+1} - \epsilon_{lt}) \tag{A4}$$

where we have differenced out a location fixed effect, ϵ_{lt} is an additive error term and the goal is to recover the inverse labor supply elasticity $\beta = \frac{1}{\theta}$. Traditional OLS estimation of equation (A4) is subject to concerns of endogeneity and hence the Bartik instrument may provide a way to estimate β consistently.

E.1 The model's empirical analogue

It is instructive to compare the population expressions for $\hat{\beta}_{OLS}$ and $\hat{\beta}_{Bartik}$:

$$\hat{\beta}_{OLS} = \frac{1}{\theta} \frac{\frac{\theta}{(\theta+\phi)^2} Var(d\alpha_{lt+1} - d\alpha_{lt}) - \frac{\phi}{(\theta+\phi)^2} Var(d\sigma_{lt+1} - d\sigma_{lt}) + \frac{\phi-\theta}{\phi+\theta} Cov(d\alpha_{lt+1} - d\alpha_{lt}, d\sigma_{lt+1} - d\sigma_{lt})}{\frac{\theta}{(\theta+\phi)^2} \underbrace{Var(d\alpha_{lt+1} - d\alpha_{lt})}_{\text{demand}} + \frac{\phi}{\theta} \frac{\phi}{(\theta+\phi)^2} \underbrace{Var(d\sigma_{lt+1} - d\sigma_{lt})}_{\text{supply}} + \frac{\phi}{(\theta+\phi)^2} \underbrace{Cov(d\alpha_{lt+1} - d\alpha_{lt}, d\sigma_{lt+1} - d\sigma_{lt})}_{\text{covariance}}}$$

$$\hat{\beta}_{Bartik} = \frac{1}{\theta} \frac{Cov[d\alpha_{lt+1} - d\alpha_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})] - Cov[d\sigma_{lt+1} - d\sigma_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})]}{Cov[d\alpha_{lt+1} - d\alpha_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})] + Cov[d\sigma_{lt+1} - d\sigma_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})]}.$$

We see that for $\hat{\beta}_{OLS}$ to be consistent, an important sufficient condition is that there are no changes in supply shocks, or $Var(d\sigma_{lt+1} - d\sigma_{lt}) = 0$. In contrast, for $\hat{\beta}_{Bartik}$ to be consistent, industry composition must not be related to innovations in supply shocks, or $Cov[d\sigma_{lt+1} - d\sigma_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})] = 0$. Bartik is invalid if the innovations in the supply shocks are predicted by industry composition. For example, Bartik would not be valid if $d\sigma_{lt+1} - d\sigma_{lt} = d\tilde{\sigma}_{lt+1} - d\tilde{\sigma}_{lt} + \sum_k z_{lk0}(d\sigma_{kt+1} - d\sigma_{kt})$. The relevance condition is that $Cov[d\alpha_{lt+1} - d\alpha_{lt}, \sum_k z_{lk0}(d\alpha_{kt+1} - d\alpha_{kt})] \neq 0$. A necessary condition for instrument relevance is that there is variation in the innovations to demand shocks between at least two industries.

The condition for Bartik to be consistent is weaker than for OLS, since the variance of the innovations to the supply shocks enters into the location-level component of growth (g_{lt}) and Bartik removes these (but not their correlation with demand shocks). The observation that the Bartik estimator does not include the variance of the innovations to the supply shocks helps explain why Bartik tends to produce results that "look like" a demand shock.

In this model, any given industry share would be a valid instrument. The exclusion restriction is that the industry share does not predict innovations to supply shocks: $Cov(d\sigma_{lt+1} - d\sigma_{lt}, z_{lk0}) = 0$. The relevance condition is that $Cov[d\alpha_{lt+1} - d\alpha_{lt}, z_{lk0}] \neq 0$, which says that the industry share is correlated with the innovations in the demand shocks.

F Using growth rates to test overidentification restrictions

Beaudry, Green, and Sand (2012, pg. 1084) test for overidentification of Bartik instruments by looking at whether different choices of weights lead to different parameter estimates. One point they emphasize is that their two choices of weights lead to instruments that are only weakly correlated.[30] Their assertion is that if two different choices of weights lead to instruments that are weakly correlated and these instruments generate similar point estimates, then this implies that the instrument is valid, in the sense that under the null of constant treatment effects neither instrument is correlated with the error term.

We now show that this logic is incomplete because it tests whether different ways of weighting the instruments give rise to different point estimates, rather than whether different instruments give rise to different point estimates. So it is possible to pass the overidentification test proposed by Beaudry, Green, and Sand (2012), while failing an overidentification test in terms of industry shares. (Of course, the converse is not true.)

In particular, we consider a researcher choosing two sets of weights (we denote those weights by $G = \hat{C}(GG'))$. We show that given one set of weights, denoted by G_1, and all but two entries in a second vector G_2, it is possible to generate two instruments that have a covariance of 0 and lead to identical parameter estimates and thus would pass this proposed overidentification test. Importantly, however, it would still be possible to fail the overidentification tests that are in terms of whether different industry shares generate similar parameter estimates.

PROPOSITION F.1. *Suppose that we are given two sets of weights, G_1 and G_2, with the last two entries of the second vector unknown ($G_{2,K-1}$ and $G_{2,K}$), and a set of industry shares Z. Use these two sets of weights to construct two Bartik instruments: $B_1 = ZG_1$ and $B_2 = ZG_2$. Assume further that all the entries in $G_1' Var(Z)$ are non-zero and that all the entries in $(((G_1'Z'Y(G_1'Z'X)^{-1}X')' - Y')Z)$ are non-zero. Then it is always possible to find $G_{2,K-1}$ and $G_{2,K}$ such that:*

1. *The two Bartik instruments are uncorrelated.*
2. *The two Bartik instruments lead to identical parameter estimates.*

The proof shows that the desiderata of the proposition implies two linear equations in two unknowns.

Proof. The first constraint is that the covariance between the two Bartik instruments is zero:

$$
\begin{aligned}
Cov(B_1, B_2) &= \mathbb{E}[B_1 B_2] - \mathbb{E}[B_1]\mathbb{E}[B_2] && \text{(A1)}\\
&= \mathbb{E}[(ZG_1)(ZG_2)] - \mathbb{E}[ZG_1]\mathbb{E}[ZG_2] && \text{(A2)}\\
&= \mathbb{E}[(ZG_1)'(ZG_2)] - \mathbb{E}[ZG_1]\mathbb{E}[ZG_2] && \text{(A3)}\\
&= G_1'\mathbb{E}[Z'Z]G_2 - G_1'\mathbb{E}[Z']\mathbb{E}[Z]G_2 && \text{(A4)}\\
&= G_1'[\mathbb{E}[Z'Z] - \mathbb{E}[Z']\mathbb{E}[Z]]G_2 && \text{(A5)}\\
&= G_1' Var(Z) G_2, && \text{(A6)}
\end{aligned}
$$

where this exploits the fact that $B_{1,l}$ is a scalar so we can take the transpose, and G_1 and G_2 are non-stochastic so that we can pull them out of the expectation. Let $T = G_1'\Sigma_Z$, where

[30] "We believe that the fact that the two IV approaches, which focus on very different data variation, give very similar results provides considerable support [for the instrument]."

$\Sigma_Z = Var(Z)$. So we can write this first constraint as:

$$TG_2 = 0. \tag{A7}$$

The second constraint is that:

$$\hat{\beta}_1 = \hat{\beta}_2 \tag{A8}$$

where for $j \in \{1, 2\}$ $\hat{\beta}_j = G_j' Z' Y (G_j' Z' X)^{-1}$. Equating:

$$\hat{\beta}_1 = \hat{\beta}_2 \tag{A9}$$
$$G_1' Z' Y (G_1' Z' X)^{-1} = G_2' Z' Y (G_2' Z' X)^{-1} \tag{A10}$$
$$G_1' Z' Y (G_1' Z' X)^{-1} X' Z G_2 = Z Y' G_2. \tag{A11}$$

Note that $G_1' Z' Y (G_1' Z' X)^{-1} X' = \hat{Y}_1$, or the predicted value of Y using the instrument built from G_1. So we have:

$$((\hat{Y}_1' - Y')Z)G_2 = 0. \tag{A12}$$

Let $S = ((\hat{Y}_1' - Y')Z)$, where we pick S because this expression is the basis of the Sargan test. Then we have that the second constraint is:

$$SG_2 = 0 \tag{A13}$$

Note that T and S are both $1 \times K$. By assumption, the last two entries in both S and T are nonzero. We now construct expressions for these two entries. To make $TG_2 = 0$, we need $\sum_{k=1}^{K} T_k G_{2,k} = 0 \Rightarrow G_{2,K} = -\frac{\sum_{k=1}^{K-1} T_k G_{2,k}}{T_K} = -\frac{\sum_{k=1}^{K-2} T_k G_{2,k} + T_{K-1} G_{2,K-1}}{T_K}$. To make $SG_2 = 0$ we have $G_{2,K-1} = -\frac{\sum_{k=1}^{K-2} S_k G_{2,k} + S_K G_{2,K}}{S_{K-1}}$. This gives us two equations in two unknowns.

□

G The Rotemberg weights with leave-one-out

The formulas we present in Section 2 apply to the case where the weights are common to all locations (i.e., we compute the national industry growth rates using a weighted average that included all locations). Here we present the formulas for the α_k that obtain when we use leave-one-out growth rates to construct the Bartik estimator. We note a few things. First, the numerical equivalence between GMM and Bartik obtains in the limit as the number of locations goes to infinity when we use a leave-one-out estimator. Second, when we use a leave-one-out estimator, the weights sum to one in the limit as the number of locations goes to infinity. (For notational simplicity we suppress notation that residualizes for controls.)

First, we derive how the leave-location-l-out estimator of G, which we denote by G_{-l}, relates to the overall average, G and the location-specific G_l (L is the number of locations):

$$G = \frac{L-1}{L} G_{-l} + \frac{1}{L} G_l \Rightarrow G_{-l} = \frac{L}{L-1} G - \frac{1}{L-1} G_l.$$

Second, we derive a version of Proposition 2.1 with the leave-one-out estimator of G. Note that the instrument constructed using leave-l-out growth rates in location l is: $B_{l,-l} = Z_l \left(\frac{L}{L-1}G - \frac{1}{L-1}G_l\right)$ where G and G_l are $K \times 1$ vectors and Z_l is a $1 \times K$ vector (and Z will be the $L \times K$ stacked matrix). Then:

$$B_{l,-l} = Z_l \left(\frac{L}{L-1} G_L - \frac{1}{L-1} G_l \right) \tag{A1}$$

$$B_{l,-l} = \frac{L}{L-1} Z_l G - \frac{1}{L-1} Z_l G_l \tag{A2}$$

$$B_{l,-l} = \frac{L}{L-1} B_l - \frac{1}{L-1} X_l, \tag{A3}$$

where the observation is that $Z_l G_l = X_l$. Then the stacked version is:

$$B_{-l} = \frac{L}{L-1} B - \frac{1}{L-1} X,$$

where B is the vector of B_l and B_{-l} is the vector of $B_{l,-l}$.

Then:

$$\hat{\beta} = \frac{B'_{-l} Y}{B'_{-l} X} \tag{A4}$$

$$= \frac{\left(\frac{L}{L-1} B - \frac{1}{L-1} X\right)' Y}{\left(\frac{L}{L-1} B - \frac{1}{L-1} X\right)' X} \tag{A5}$$

$$= \frac{\left(\frac{L}{L-1} (ZG) - \frac{1}{L-1} X\right)' Y}{\left(\frac{L}{L-1} (ZG) - \frac{1}{L-1} X\right)' X}. \tag{A6}$$

As before:

$$\beta_k = \frac{Z'_k Y}{Z'_k X}. \tag{A7}$$

Then one can show:

$$\alpha_k = \frac{\frac{L}{L-1} g_k Z'_k X - \frac{1}{L-1} X' Y \beta_k^{-1}}{\sum_k \frac{L}{L-1} g_k Z'_k X - \frac{1}{L-1} X' X}. \tag{A8}$$

By inspection, $\sum_k \alpha_k \neq 1$. However, as $L \to \infty$ the sum converges to 1 as the leave-one-out terms drop out.

Table A1: Replication of ADH Table 3, Column (6) Results

	ADH	Replicated	Fixed Shares	Fixed All
Δ imports from China to US	-0.60	-0.62	-0.16	-0.60
	(0.10)	(0.13)	(1.41)	(0.78)
Share Empl in Manufacturing	-0.04	-0.04	-0.09	
	(0.01)	(0.02)	(0.16)	
Share College Educated	0.01	0.01	0.01	
	(0.01)	(0.01)	(0.03)	
Share Foreign Born	0.03	0.03	0.02	
	(0.01)	(0.01)	(0.03)	
Share Empl of Women	-0.01	-0.01	-0.01	
	(0.02)	(0.02)	(0.03)	
Share Empl in Routine	-0.24	-0.25	-0.20	
	(0.06)	(0.06)	(0.13)	
Avg Offshorability	-0.06	-0.05	-0.23	
	(0.24)	(0.23)	(0.56)	
1990 $\times$ Share Empl in Manufacturing (1980)				-0.03
				(0.06)
2000 $\times$ Share Empl in Manufacturing (1980)				-0.06
				(0.12)
1990 $\times$ Share College Educated (1980)				0.01
				(0.02)
2000 $\times$ Share College Educated (1980)				0.02
				(0.03)
1990 $\times$ Share Foreign Born (1980)				-0.00
				(0.02)
2000 $\times$ Share Foreign Born (1980)				0.07
				(0.03)
1990 $\times$ Share Empl of Women (1980)				0.04
				(0.04)
2000 $\times$ Share Empl of Women (1980)				0.04
				(0.04)
1990 $\times$ Share Empl in Routine (1980)				-0.14
				(0.08)
2000 $\times$ Share Empl in Routine (1980)				-0.27
				(0.15)
1990 $\times$ Avg Offshorability (1980)				-1.05
				(0.44)
2000 $\times$ Avg Offshorability (1980)				0.11
				(0.47)
Year and Census Division FE	Yes	Yes	Yes	Yes
N	1,444	1,444	1,444	1,444

Notes: This table reports our replication of ADH, as well as the results of small variations on the specification. Column (1) of Table A1 reproduces ADH Table 3 column (6), and column (2) reports our replication. Column (3) fixes industry shares to the 1980 values, but allows the controls to vary over time as in column (2). Like column (3), column (4) fixes the industry shares to their 1980 values, but also fixes the controls to their 1980 values and interacts these controls with time.