BARTIK INSTRUMENTS:
WHAT, WHEN, WHY, AND HOW

Paul Goldsmith-Pinkham
Isaac Sorkin
Henry Swift

Bartik Instruments: What, When, Why, and How
Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift
NBER Working Paper No. 24408
March 2018, Revised July 2018
JEL No. C1,C18,C2,J0,J2

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 in a GMM estimator and discuss how different asymptotics imply different identifying assumptions. We argue that in most applications the identifying assumption is in terms of industry shares. Finally, we show how to decompose the Bartik instrument into the weighted sum of the just-identified instrumental variables estimators. These weights measure how sensitive the parameter estimate is to each instrument. We illustrate our results through four applications: estimating the inverse elasticity of labor supply, estimating local labor market effects of Chinese imports, estimating the fiscal multiplier using defense spending shocks, and using simulated instruments to study the effects of Medicaid expansions.

Paul Goldsmith-Pinkham
Federal Reserve Bank of New York
33 Liberty St
New York, NY 10045
paulgp@gmail.com

Isaac Sorkin
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
sorkin@stanford.edu

Henry Swift
Massachusetts Institute of Technology,
henry.swift@gmail.com

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

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

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

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

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

We first show that the Bartik instrument is numerically equivalent to a generalized method of moments (GMM) estimator with the local industry shares as instruments and

---

[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.

a weight matrix constructed from the national 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. Under the natural asymptotic assumption of fixed number of time periods and industries, and locations growing to infinity, the GMM estimator suggests that the identifying assumption for the Bartik instrument is best stated in terms of local industry composition.

Using the industry shares as instruments does, however, rely on a stronger identifying assumption than using the Bartik instrument directly. The Bartik instrument imposes a restriction on a weighted sum of the industry shares and the error terms ($\mathbb{E}[\sum_k z_{lk} g_k \epsilon_l] = 0$), whereas the identification condition for using the industry shares as instruments imposes an orthogonality condition on each of the industry shares ($\mathbb{E}[z_{lk} \epsilon_l] = 0, \forall k$). Under the assumption of fixed number of industries and time periods, however, we view the difference between these two assumptions as hard to motivate.

We consider alternative asymptotics that allow for other identifying assumptions. In our first departure, we consider settings where the number of time periods go to infinity, but we fix the number of locations. This setting is a relabelling of our benchmark case. Here, the instrument is the national growth rates (interacted with local industry composition). So the identifying assumption is best stated in terms of the national growth rates. As an example of this setting, which we discuss more below, is Nakamura and Steinsson (2014), where the instrument is national defense spending shocks.

In our second departure, we allow the number of industries to go to infinity. This set-up provides a robust justification for an important distinction between an orthogonality condition in terms of the Bartik instrument, and in terms of the industry shares as instruments. We construct an argument formally identical to that in Kolesar et al. (2015) of "many invalid instruments". The set-up allows the industry shares to enter the error term directly, making each industry share an invalid instrument. However, if there are many industries and the direct effect of each industry is independent of its first stage coefficient, then as the number of industries goes to infinity the misspecification "averages out." Borusyak, Hull, and Jaravel (2018) and Adao, Kolesar, and Morales (2018) develop additional results about this asymptotic setting.

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 and sum to 1. 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 instrument. The weights depend on the covariance between the kth instrument's fitted value of the endogenous variable and the endogenous variable itself.

We use these Rotemberg weights to assess the plausibility of the "many invalid instruments" view of the Bartik instrument. The key observation is that the many invalid instruments asymptotics imply that in the limit no single instrument "matters" very much. In contrast, in our empirical examples the Rotemberg weights tend to be quite skewed, with the top five industries accounting for about half of the positive weight. We show analytically in a special case that the many invalid instrument asymptotics imply that the Rotemberg weight on any given industry goes to zero. We conduct Monte Carlo simulations designed to mimic U.S. industry composition and growth rates. Even when there are over 200 instruments, if the instruments with the five largest Rotemberg weights are misspecified in a way that is correlated with the first-stage coefficient, then the Bartik estimates are biased.

We suggest researchers perform three relatively standard tests of the identifying assumption. First, researchers should test how balanced instruments are across potential confounders. Researchers should conduct these tests both in terms of the overall instrument as well as the high Rotemberg weight industries. 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, in settings where there is a pre-period, researchers can test for parallel pretrends. We suggest that researchers examine parallel pre-trends in terms of the overall Bartik instrument, as well as in terms of the industries that receive a large Rotemberg weight. Failure of parallel pre-trends in terms of any given instrument suggests misspecification in that instrument, and so raises questions about the broader research design.

Third, researchers can consider alternative estimators and also perform overidentification tests. Under the null of constant effects, alternative estimators should deliver similar point estimates, and the divergence of point estimates across estimators is typically interpreted as reason to worry.[2] An alternative interpretation of such divergence emphasized by Kolesar et al. (2015) is that it is evidence of misspecification. A direct way of assessing the importance of misspecification is to perform overidentification tests.[3]

---

[2]Angrist and Pischke (2008, pg. 213) write: "Check overidentified 2SLS estimates with LIML. LIML is less precise than 2SLS but also less biased. If the results come out similar, be happy. If not, worry..."

[3]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 overiden-

We note two limitations to our analysis. First, we assume locations are independent and so ignore the possibility of spatial spillovers or correlation.[4] Second, we assume that the data consist of a series of steady states.[5]

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 under plausible asymptotics, 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.

Applications: We illustrate our results through four 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, the weights are skewed, with about a third of the weight on the top five industries. 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. Third, industry shares, including oil and gas extraction, are correlated with many observables, including the immigrant share, which is thought to predict innovations in labor supply. Fourth, an overidentification test rejects the null of exogeneity, and alternative estimators deliver substantively different point estimates.

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, the two highest weight industries are games and toys and electronic computers. 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. Third, the industries that get the most weight tend to be in more educated areas. Fourth, we examine pre-trends among the industries with high weights and find that none of the comparisons

---

tifying restrictions and thus explicitly embrace the constant effects interpretation (e.g., Beaudry, Green, and Sand (2012) and Hornbeck and Moretti (2018)).

[4]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.

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

implied by the industries (i.e., places with more and less of the industry) exhibits flat pre-trends and effects that are dramatically larger in the 2000s (when the China shock was largest). Fifth, alternative estimators deliver substantively different point estimates and overidentification tests reject the null of exogeneity.

In our third application, we examine the use of defense spending shocks to estimate the fiscal multiplier following Nakamura and Steinsson (2014). This setting is a relabelling of the canonical Bartik setting. To map to the Bartik setting, the equivalent of local industry shares times time is the national industry growth rates times location, and the equivalent of the national industry rates is location industry shares. We find that the weights are quite skewed, with the most weight placed on Northeastern states.

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 receive the largest weight.

Besides these four examples, a much broader set of instruments is Bartik-like. We define a Bartik-like instrument as one that uses the inner product structure of the endogenous variable to construct an instrument. This encompasses at least three 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.

5

Literature: A vast literature of papers uses Bartik-like instruments, and many of these discuss the identifying assumptions in ways that are close to the benchmark results in this paper. For example, Baum-Snow and Ferreira (2015, pg. 50) survey the literature and state that the "validity [of the Bartik instrument]...relies on the assertion that neither industry composition nor unobserved variables correlated with it directly predict the outcomes of interest conditional on controls." Similarly, Beaudry, Green, and Sand (2012) provide a careful discussion of identifying assumptions in the context of an economic model. We only intend to claim novelty for the formalism along this dimension.

Beyond the vast literature of papers using Bartik-like instruments, this paper is also related to a growing literature that comments on specific papers (or literatures) that use Bartik-like instruments. This literature includes at least three papers: Christian and Barrett (2017), which comments on Nunn and Qian (2014), Jaeger, Joyce, and Kaestner (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.

1 Equivalence between Bartik IV and GMM with industry shares

We first show that the Bartik instrument is numerically equivalent to using industry shares as instruments. 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 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. Section 2 discusses asymptotics.

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:

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

In the canonical setting, l indexes a location, t a time period, y_{lt} is wage growth, D_{lt} is a vector of Q controls which could include location and time fixed effects, x_{lt} is employment

growth and ϵ_{lt} is a structural error term. The estimand of interest is β_0. We assume that the ordinary least squares (OLS) estimator for β_0 is biased and we need an instrument to estimate β_0.

In this section, we focus on numerical equivalence of different estimators, and so leave the sampling process unspecified. When we next study the consistency of these estimators, we consider three asymptotic sampling frames: one where we fix K and T, and let $L \to \infty$; second, where we fix T and let both $L, K \to \infty$; and in a third we fix K and L, and let $T \to \infty$. In all cases, we will assume that the units are drawn in an independent and identically distributed manner, but allow for within-unit correlation.

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, and G_{lt} is a $K \times 1$ vector of industry-location-time period growth rates where the kth entry is g_{lkt}. We decompose the industry-location-period growth rate into industry-period, and idiosyncratic industry-location-period components:

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

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:[6]

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

where G_t is a $K \times 1$ vector of the industry growth rates in period t (the kth entry is g_{kt}), and Z_{l0} is the $1 \times K$ vector of industry shares in location l.

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

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

---

[6] 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.

1.2 Equivalence in three special cases

We build up to the result that the Bartik instrument is numerically equivalent to using industry shares as instruments through three special cases.

Two industries and one time period

With two industries whose shares sum to one and one time period, the Bartik instrument is identical to using one of the industry shares as an instrument. To see this, expand the Bartik instrument:

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

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

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

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

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

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

Two industries and two time periods

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

varies over time:

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

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

To see the relationship between the cross-sectional and panel estimating equations, 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 $\Delta g_1 = g_{12} - g_{11}$, $\Delta g_2 = g_{22} - g_{21}$, and $\mathbb{1}$ is the indicator function. Then, rewrite the first stage as

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

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

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

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

In this case, again $\tilde{\gamma} = \gamma/(\Delta g_1 - \Delta g_2)$. 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, $\Delta g_{1t} - \Delta g_{2t}$.

Viewing the growth rates as a measure of policy size and the industry shares as measures of exposure emphasizes a useful connection to difference-in-differences. In some settings, there are more than two time periods, and there is a pre-period before a policy takes effect. In this case, we can write the reduced form in terms of the Bartik instrument, or, alternatively, in terms of industry shares interacted with time:

$$y_{lt} = \tau_l + \tau_t + \gamma\beta B_{lt} + \underbrace{\tilde{\epsilon}}_{\eta_{lt}\beta + \epsilon_{lt}}$$

$$y_{lt} = \tau_l + \tau_t + \sum_{s \neq s_0} \mathbb{1}(s = t) \underbrace{\tilde{\gamma}_s}_{\gamma \Delta_{gs}} \beta z_{l10} + \underbrace{\tilde{\epsilon}}_{\eta_{lt}\beta + \epsilon_{lt}}.$$

In this case, the testable implication of parallel pre-trends is that $\gamma_t\beta = 0$ for $t < s_0$, where s_0 demarcates the pre-period. Relative to a setting in which there is no pre-period, settings with pre-periods provide additional ways of testing the design using the standard tools of applied microeconomics and potentially add credibility to the use of a Bartik instrument.

K industries and one time period

Finally, we show that with K industries as instruments in a generalized method of moments (GMM) estimator set-up with a specific weight matrix, the Bartik estimator 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 when Bartik and GMM are numerically equivalent.

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.

REMARK 1.1. *When $\sum_{k=1}^{K} z_{lk} = 1$, there are $K - 1$ instruments and not K instruments. In practice, any of the K industries can be dropped by subtracting off that industry's growth rate from the G vector, and the Bartik instrument will maintain its numerical equivalence from Proposition 1.1. To*

see the intuition behind this, suppose that $\sum_k z_{lk} = 1 \forall l$. Consider the first stage regression:

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

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

$$x_l = \underbrace{\gamma_0 + \gamma_1 \sum_k z_{lk} g_j}_{\gamma_0 + \gamma_1 g_j} + \gamma_1 \underbrace{\sum_k z_{lk}(g_k - g_j)}_{B_l - g_j} + \eta_l. \tag{1.3}$$

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

1.3   Summary

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

2   Asymptotic consistency and identifying assumptions

The previous section established finite sample equivalence between the GMM estimator, using industry shares as instruments and industry growth weights in the weight matrix, and the TSLS IV estimator using the Bartik instrument. We now discuss the consistency of these estimators under different asymptotic regimes and the implied identification conditions.

To fix ideas, the difference between estimator and estimand is:

$$\hat{\beta} - \beta = \frac{B' \epsilon^\perp}{B' X^\perp} = \frac{X^{\perp\prime} Z W Z' \epsilon^\perp}{X^{\perp\prime} Z W Z' X^\perp}, \tag{2.1}$$

where $W = GG'$. In the subsections below, we outline what assumptions are necessary for this difference to converge in probability to zero. This also requires that the denominator converges to a constant in the limit, which we assume. Hence, we focus on the necessary identification conditions for the numerator to converge to zero.

Since the goal of asymptotics is to approximate the distributional behavior of estimators in finite samples, the appropriate asymptotic assumption will vary by application. We

encourage researchers to be explicit about which asymptotic framework they are using (and why). Section 4.4 provides some data-driven guidance.

2.1 Case 1: Fix K, fix T and let $L \to \infty$

The first case is an application of standard theorems for GMM. In this case, we fix the number of industries and time periods, and hence view the g_{kt} as non-stochastic. As a result, the identification condition when we treat the instruments as industry shares (times time period) is:

$$\mathbb{E}\left[z_{lk0}\mathbb{1}(s=t)\epsilon_{lt}|\mathbf{D}_l\right] = 0, \forall k, s, t.$$

This condition is stronger than the condition implied by using the Bartik instrument directly where it is only the weighted sum of industry shares that is zero. Formally:

$$\sum_k \sum_s g_{ks}\mathbb{E}\left[z_{lk0}\mathbb{1}(s=t)\epsilon_{lt}|\mathbf{D}_l\right] = 0, \forall t.$$

It is typically hard, however, to motivate why the g_{kt} would have the special property that would cause the misspecification in each instrument to exactly cancel out. Hence, given these asymptotics, we view the condition that each instrument (initial industry share times time period) is orthogonal to the error term as the reasonable identification condition.

2.2 Case 2: Fix K, fix L, and let $T \to \infty$.

Second, consider the case where the number of locations and industries is fixed, and instead the number of time periods grows. This case switches the positions of the G_t vector of national industry growth rates to form the moment condition, and the initial industry shares, Z_l, to form the GMM weight matrix.[7] Here, the equivalence between Bartik and GMM implies viewing the instrument as national industry growth rates times location, and the weight matrix is the outer product of the local industry shares. The Bartik version imposes that the first-stage coefficients are proportional to industry-location shares.

Following the arguments of case 1, the natural identification condition is then

$$\mathbb{E}\left[g_{kt}\mathbb{1}(s=l)\epsilon_{lt}|\mathbf{D}_l\right] = 0, \forall k, s, l.$$

---

[7]Formally, we redefine the matrix notation that constructs the $LT \times 1$ vector of the Bartik instrument, $\mathbf{B}$. In slight abuse of notation, let $\mathbf{G}$ be an $LT \times LK$ matrix, where each industry's growth rate in a given period is multiplied by a location fixed effect. Then, let $\mathbf{Z}$ be an $LK \times 1$ vector of the $K \times 1$ vectors of industry-location shares. Then we can write, $\mathbf{B} = \mathbf{GZ}$, and similar to the results in Section 1, using $\mathbf{B}$ in a just-identified TSLS setting is identical to GMM with a weight matrix defined as $W = \mathbf{ZZ}'$.

This identification assumption would fail if places that respond more to industry k's shock also have endogenous innovations that occur at the same time.

2.3 Case 3: Let $K \to \infty$, fix $T = 1$, and let $L \to \infty$

Finally, we consider a setting where the number of industries goes to infinity along with the number of locations. Relative to the knife-edge condition in Case 1, these asymptotics generate a set of restrictions on the properties of the g_{kt} such that an alternative identification condition holds. Any given industry share might be an invalid instrument. The key idea is that the invalid effects of the instruments averages out because there are a large number of national industry growth rates (g_{kt}) that are randomly assigned. Hence, the consistency argument relies on the number of instruments, K, growing to infinity along with the number of locations.

For simplicity, we assume that $T = 1$, but little would change with any fixed T. Our results in this section closely follow Kolesar et al. (2015), who term this case the many invalid instrument case.

Let $\epsilon_l = Z_l \lambda + u_l$, and then define the instrumental variables setup as

$$Y_l = X_l \beta_0 + D_l \rho + \underbrace{Z_l \lambda + u_l}_{\epsilon_l}$$

$$X_l = Z_l G \pi_1 + D_l \pi_2 + v_l.$$

If $\lambda = 0$, then the same standard assumptions from Case 1 above hold, and the vector of Z_l is a valid instrument. However, if $\lambda \neq 0$, then the industry shares Z_l have a direct effect on the outcome of interest. A key simplification in our analysis is that we treat the G as known datapoints, and not estimated. Borusyak, Hull, and Jaravel (2018) provide a more general treatment of these asymptotics that, among other things, allows for estimation of the G.[8]

We make the same standard assumptions as Kolesar et al. (2015), and highlight two assumptions.[9] First, the number of instruments grows with the sample size:

ASSUMPTION 1. *For some* $0 \leq c_K < 1$

$$K_L / L = c_K + o(L^{-1/2}).$$

Second, the first-stage is non-zero, and the direct effect and the growth rates are uncorrelated:

---

[8] Adao, Kolesar, and Morales (2018) discuss inference in a version of this setting.
[9] We list the full set of assumptions in Appendix D.

ASSUMPTION 2. $\pi_1 \neq 0$ *and* $\plim_{K,L \to \infty} \lambda' \mathbf{Z}^{\perp\prime} \mathbf{Z}^\perp G / L = 0$.

Under these assumptions, the TSLS estimate using $B_l = Z_l G_l$ is a consistent estimator of β.[10]

To see the logic of this case most clearly, consider a simpler set-up where each instrument is binary, there are no controls, and each location only has one industry.[11] In this case, Assumption 2 simplifies to:

$$\lambda' \mathbf{Z}' \mathbf{Z} G / L = \sum_k \omega_{L,k} \lambda_k g_k$$

where $\omega_{L,k} = \sum_l Z_{lk}/L$ is the share of locations that have industry k. The economic interpretation of this condition is that the direct effect of industry composition on wage growth is independent of the direct effect of industry composition on employment growth, and the share of locations that have industry k goes to zero. For example, it allows for the possibility that places with industry k systematically have higher wage growth, but says that this endogeneity is independent of the growth rate (when averaged across k). If the g_k are random shocks, then this condition holds.

3   Opening the black box of the Bartik estimator

The previous sections showed that under standard panel asymptotics, the Bartik instrument is equivalent to using industry shares (interacted with time fixed effects) as instruments. Thus, the Bartik estimator combines many instruments using a specific weight matrix.

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

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

---

[10]See Appendix D for a more formal statement. Note that in this setting, in contrast to Kolesar et al. (2015), it is not necessary to adjust for the many instruments in choice of estimator, since the instrument B_l is a scalar. This relies on the fact that the choice of G is known *ex ante*, and still gives first-stage power (e.g. $\pi_1 \neq 0$).

[11]This example is nearly identical to Kolesar et al. (2015, Section 2).

instruments are more sensitive to misspecification, and hence are the instruments that are most important for researchers to defend.

3.1 Decomposing the Bartik estimator

We first present a finite sample decomposition of the linear overidentified GMM estimator due to Rotemberg (1983).[12] For expositional simplicity, we use a single cross-section, though it is straightforward to extend results to the fixed T panel.

Consider the minimization 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 $\hat{W}$:

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

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 3.1. *Let*
$$\hat{C}(\hat{W}) = \hat{W} Z' X^\perp \text{ and } \hat{c}_k(\hat{W}) = \hat{W}_k Z' X^\perp,$$

where $\hat{W}_k$ is the kth row of $\hat{W}$.

We index a solution for $\hat{\beta}$ by $\hat{W}$: $\hat{\beta}(\hat{W})$. The following result (which is a special case of Rotemberg (1983, Proposition 1)) decomposes $\hat{\beta}(\hat{W})$ into the contribution from each of the K just-identified regressions.

PROPOSITION 3.1. *Let*

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

Then:

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

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

Proof. See appendix B. □

Proposition 3.1 has three implications. First, mirroring our results from Section 2.1, the validity of each just-identified $\hat{\beta}_k$ depends on the exogeneity of a given Z_k, and hence for

---

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

fixed K, the overall estimate's consistency 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, and thus the overidentified IV estimate may lie outside of the range of the just-identified estimates.

We now look at the Rotemberg weights for the Bartik estimator.

REMARK 3.1. *The Rotemberg weights for the Bartik instrument are given by:*

$$\hat{\alpha}_k(GG') = \frac{g_k Z'_k X^\perp}{\sum_{k=1}^K g_k Z'_k X^\perp} = \frac{\hat{\gamma} g_k Z'_k X^\perp}{\hat{\gamma} B' X^\perp} = \frac{\hat{X}_k^{Bartik\prime} X^\perp}{\hat{X}^{Bartik\prime} X^\perp},\tag{3.1}$$

where g_k is the kth 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 kth industry.

Contrast Bartik's Rotemberg weights with the weights that arise from TSLS using the industry shares as instruments:

REMARK 3.2. *The Rotemberg weights from TSLS are given by:*

$$\hat{\alpha}_k((Z^{\perp\prime}Z^\perp)^{-1}) = \frac{\hat{\pi}_k Z'_k X^\perp}{\sum_{k=1}^K \hat{\pi}_k Z'_k X^\perp} = \frac{\hat{X}_k^{TSLS\prime} X^\perp}{\hat{X}^{TSLS\prime} X^\perp},\tag{3.2}$$

where $\hat{\pi}_k$ is the kth 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 kth 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).[13] Second, the weights reflect the covariance between the kth 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 kth 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.[14] If we relax these two assumptions, then negative weights are possible.[15]

---

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

[14]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.

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

3.2 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 $\hat{W}$ 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 AGS. As such, the results in this section are largely special cases of AGS.

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 exogenous, however, the population version of the Rotemberg weights measures how much the overidentified estimate of β_0 is affected by this misspecification. To allow for this interpretation, we modify our estimating equation:

$$y_{lt} = D_{lt}\rho + x_{lt}\beta_0 + V_{lt}\kappa + \epsilon_{lt},$$

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 3 (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 kth 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 3.1. *If Assumption 3 holds and* $\operatorname{plim}_{L\to\infty}\hat{W}_L = W$ *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 misspecification.

PROPOSITION 3.2. *We assume that Assumption 3 holds and* $\text{plim}_{L\to\infty}\hat{W}_L = W$ *where W is a positive semi-definite matrix.*

If $\kappa = L^{-1/2}$, then

(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. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

This proposition shows that in the presence of misspecification, the estimator is asymptotically biased. Two useful corollaries follow:

COROLLARY 3.1. *Suppose that $\beta_0 \neq 0$. Then the percentage bias can be written in terms of the Rotemberg weights:*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \alpha_k(W)\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \tag{3.3}$$

COROLLARY 3.2. *Under the Bartik weight matrix ($W = GG'$),*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \frac{g_k\Sigma_{ZX^\perp,k}}{G'\Sigma'_{ZX^\perp}}\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \tag{3.4}$$

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 AGS sensitivity parameter to be unit-invariant, and hence is comparable across instruments.[16] 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.

An alternative approach to measuring sensitivity 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 kth instrument and define the bias term for $\hat{\beta}(\hat{W}_{-k})$ as $\tilde{\beta}(\hat{W}_{-k}) = \hat{\beta}(\hat{W}_{-k}) - \beta$.

PROPOSITION 3.3. *The difference in the bias from the full estimator and the estimator that leaves*

---

[16] 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."

out the kth 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 misspecified the instrument is relative to the remaining instruments.

To summarize, 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 a more concrete way to describe the empirical strategy. Second, to the extent that the researcher is concerned about misspecification, these are the instruments that are most worth probing.

3.3   Normalization

When the industry shares sum to one, the instruments are linearly dependent and so we can write each instrument as a function of the remaining $K - 1$ instruments. This fact has a couple implications. First, following Remark 1.1, we can drop any industry through normalization by subtracting off g_j from all the growth rates, and leave our point estimates unchanged. Second, the fact that we can drop any one industry means that the Rotemberg weights are not invariant to the choice of which industry to drop. To take an extreme example, suppose industry j has the largest weight. Then, by dropping industry j through normalization, a researcher could make industry j have a weight of zero, but the Bartik estimate would remain the same.

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

To understand the intuition for why the normalization matters, return to the two industry example: suppose we think that Bartik is biased in this case. Does the bias arise from the fact that the industry 1 share is correlated with the error term, or that the industry 2

share is correlated with the error term? Conceptually, it is not meaningful to distinguish between these two possibilities, because industry 1 and 2 shares are exactly negatively correlated. Hence, saying the bias is correlated with industry 1 is the same as saying the bias is correlated with industry 2. In this case, our normalization assigns weight 0.5 to each industry.

3.4 Aggregation

Below, we consider applications with panel data and multiple time periods. As a result, the underlying instruments are industry shares interacted with time fixed effects. Rather than reporting results at the level of $\alpha_{k,t}$, we report $\alpha_k = \sum_t \alpha_{k,t}$, or, sometimes, $\alpha_t = \sum_k \alpha_{k,t}$. It is typically easier to think about the variation coming from a cross-sectional difference, rather than the variation coming from a cross-sectional difference in a particular time period. When aggregating to the kth industry, we report $\hat{\beta}_k$, which comes from using $B_{lkt} = z_{lk0}g_{kt}$, the Bartik instrument built from just the kth industry, as the instrument.

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

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

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

4 Testing the plausibility of the identifying assumptions

The identifying assumptions necessary for consistency are typically not directly testable. However, it is possible to partially assess their plausibility. We first focus on the assumptions from Section 2.1; in the context of the canonical setting of estimating the inverse elasticity of labor supply, the identifying assumption is that industry composition (Z_{l0}) does not predict innovations to labor supply (ϵ_{lt}).[17] For applications where this exclusion restriction may not hold, we discuss a test of the plausibility of "averaging out" asymptotics from Section 2.3.

4.1 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

---

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

provides an empirical description of what the variation is correlated with, and the types of mechanisms that may be problematic for the exclusion restriction.

Since we argued in footnote 6 that it is typically desirable to fix industry shares to an initial time period (Z_{l0}), we recommend considering the correlation with initial period characteristics, as this reflects the instruments' cross-sectional variation. If Z_{l0} is correlated with potential confounding factors, this can imply that there are omitted variables biasing estimation. Naturally, it is always possible to control for observable confounders, but following the logic of Altonji, Elder, and Taber (2005) and Oster (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.

4.2 Empirical Test 2: Pre-trends

In some applications, there is a policy change in period s_0. As we discussed in Section 1.2, a researcher can use this sharp policy change to implement difference-in-differences research design. The analogy to difference-in-differences is most straightforward when the shares are fixed over time (emphasizing the point in footnote 6). In this case, the industry shares measure the exposure to the policy change, while the national growth rates proxy for the size of the policy change.[18] In these settings, it is natural to test for pre-trends. We recommend looking at pre-trends in terms of the instruments with the largest Rotemberg weights, as well as looking at pre-trends in terms of the overall Bartik instrument. We suspect that researchers will be more comfortable with the plausibility of their empirical design if parallel pre-trends are satisfied for the instruments to which their estimates are most sensitive to misspecification. For more details on pre-trends tests, see DiNardo and Lee (2011).

4.3 Empirical Test 3: Alternative estimators and overidentification tests

Under the assumptions in Section 2.1, there are many moment conditions that must hold, and it is possible to use the full set of industry shares (Z_{l0}) with a more flexible weight matrix, instead of the Bartik estimator. The simplest approach would be to use the TSLS estimator; however, in finite samples, the overidentified TSLS estimator is biased. As a result, we encourage researchers to use three alternative estimators which have better properties with many instruments: the Modified Bias-corrected TSLS (MBTSLS) estimator from Anatolyev (2013) and Kolesar et al. (2015), the Limited Information Maximum Likelihood (LIML) estimator and the HFUL estimator from Hausman et al. (2012).

---

[18]Some examples of this include Autor, Dorn, and Hanson (2013) and Lucca, Nadauld, and Chen (Forthcoming).

These estimators may not give the same estimates, as their underlying assumptions are different. We follow Kolesar et al. (2015, pg. 481-2) and interpret differences between HFUL and LIML on the one hand, and MBTSLS and TSLS on the other, as pointing in the direction of potential misspecfication. The reason is that LIML and HFUL are maximum likelihood estimators and so exploit cross-equation restrictions while both MBTSLS and TSLS are two-step estimators and so do not exploit these cross-equation restrictions. Comparing these estimates, along with the Bartik TSLS estimate, provides a useful first pass diagnostic for misspecification concerns. If these estimators agree, then researchers can be more confident in their identifying assumption.

Overidentification tests provide more formal tests for misspecification. These estimators permit test statistics under different assumptions. For the HFUL estimator, we suggest the overidentification test from Chao et al. (2014), and for LIML estimator, we use the Cragg and Donald (1993) statistic, as suggested by Kolesar et al. (2015).[19] Conceptually, the overidentification test asks whether the instruments are correlated with the error term beyond what would be expected by chance, and relies on the validity of at least one of the instruments.

When overidentification tests reject, and when HFUL and LIML differ from MBTSLS and Bartik TSLS, these findings point to misspecification. In this case, researchers should be especially interested in the high Rotemberg weight industries, as these are the industries where the estimates are most sensitive to misspecification.

One reaction to the divergence between various estimators and failure of the overidentification tests is that this tends to happen in any heavily overidentified setting. Kolesar et al. (2015, Table 3 and 4) present results from these estimators and tests in the Angrist and Krueger (1991) data, and find that overidentification tests do *not* reject the null of common effects, and the LIML and MBTSLS estimates are numerically very similar (though TSLS and LIML estimates diverge due to many instruments bias). Hence, we view failure of overidentification tests and differences between estimators as informative about problems with the underlying estimates.

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

We recommend that researchers use the Rotemberg weights to quantify what varia-

---

[19]Code to implement both overid tests are available on request and will be posted on Github.

tion each Bartik instrument is using, and whether the two Bartik instruments use different sources of variation. Specifically, researchers can report the top-5 Rotemberg weights across the two instruments and also their rank correlation. If these statistics are low, then the two Bartik instruments are likely using different sources of variation and the conclusion discussed above is warranted.[20]

4.4 Empirical Test 4: Using the Rotemberg weights to assess the plausibility of many invalid instrument asymptotics

Under the assumption of many invalid instrument asymptotics (as in Section 2.3), we show that in a special case the Rotemberg weight on any given instrument goes to zero. Hence, to the extent that there are a few large Rotemberg weights, then the many invalid instrument asymptotics is not a good approximation to the finite sample behavior of the estimator.

We first present an analytical result in the special case we discussed in Section 2.3 where instruments are binary.

PROPOSITION 4.1. *Suppose that g_k and $\tilde{g}_{lk}$ are mean zero. Also suppose that locations are randomly assigned a single industry so that z_{lk} is equal to 1 for one k and zero for all other $k' \neq k$. If $\lim_{L\to\infty} \frac{K}{L} = constant$ and for any L $\bar{z}_{L,k} = \sum_l z_{lk}$ is bounded by a constant, then*

$$\lim_{L\to\infty} \alpha_k(GG') = 0.$$

Proof. See Appendix B. □

This proposition says that under the many invalid instrument asymptotics the weight on each industry goes to zero. In contrast, in our applications, we typically find that a few industries get large weight, implying that the estimator could still be biased through the direct effects of those industries.

We now present Monte Carlo evidence for the more general version of this point. Our simulations are broadly designed to mimic properties of industry growth rates and industry shares in the U.S. See Appendix F for details.

The first four rows of Table 1 show how the misspecification can average out. The first row shows that when $\lambda = 0$, OLS is biased, and (infeasible) Bartik is unbiased. Rows 2

---

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

through 4 show the effect of allowing the λ term to be non-zero: in particular, we draw g_k and λ_k independently and scale the variance of the λ_k terms relative to the variance of the g_k term. In these simulations, infeasible Bartik remains unbiased.

The last two rows of Table 1 show that even with many instruments, the Bartik estimate can be biased when there are a few industries with high Rotemberg weights. Both rows deviate from row 3 by replacing the λ_k with the realized draw of g_k for five industries. In Row 5, the five industries with the **smallest** (in absolute value) Rotemberg weights are replaced, and the Bartik estimator is unbiased, similar to rows 1-4. However, in Row 6, we replace the five industries with the **largest** (in absolute value) Rotemberg weights, and the Bartik estimator is no longer approximately unbiased, even though all other industries' g_k and λ_k are uncorrelated. We view this as a consequence of the large share—50%—of the Rotemberg weights for the top 5 industries. This degree of sensitivity to misspecification in a few instruments is empirically realistic: in our empirical applications, the top 5 industries tend to also have shares of the positive weights that are around 0.5.

Finding a few high Rotemberg weight industries raises concerns about the extent to which bias in a given industry may affect the overall bias of the estimator. Diffuse Rotemberg weights, however, are a necessary but not sufficient condition for the many invalid instrument asymptotics and do not, by themselves, show that these asymptotics provide a good guide to the finite sample properties of the estimator.

5 Empirical example I: Canonical Setting

We now present empirical examples to make our theoretical ideas concrete, specifically focusing on using our empirical tests from Section 4. 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.

5.1 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.[21] 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

---

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

and 2000 to 2010 leave-one-out growth rates. To construct the industry growth rates, we weight by employment. We weight all regressions by 1980 population.

We use the leave-one-out means to construct the national growth rates to address the finite sample bias that comes from using own-observation information. Specifically, using own-observation information allows the first-stage to load on the idiosyncratic industry-location component of the growth rate, $\tilde{g}_{lk}$, which is endogenous. This finite sample bias is generic to overidentified instrumental variable estimators and is the motivation for jack-knife instrument variable estimators (e.g. Angrist, Imbens, and Krueger (1999)). In practice, because we have 722 locations, using leave-one-out to estimate the national growth rates matters little in point estimates.[22]

5.2   Rotemberg weights

We compute the Rotemberg weights of the Bartik estimator with controls, aggregated across time periods. The distribution of sensitivity is skewed, so that a small number of instruments have a large share of the weight: Table 2 shows that the top five instruments account for almost a third ($0.592/1.968$) of the positive weight in the estimator. These top five instruments are: oil and gas extraction, motor vehicles, other,[23] guided missiles, and blast furnaces.

These weights give a way of describing the research design that reflects the variation in the data that the estimator is using, and hence makes concrete for the reader what types of deviations from the identifying assumption are likely to be important. In this canonical setting, one of the important comparisons is across places with greater and smaller shares of oil and gas extraction. Hence, the estimate is very sensitive to deviations from the identifying assumption related to geographic variation in employment share in oil and gas extraction. Interestingly, a common short-hand to talk about Bartik is to discuss the fate of the automobile industry (e.g. Bound and Holzer (2000, pg. 24)), and this analysis confirms that the motor vehicle industry plays a large role in the Bartik instrument.

Finally, Panel B shows that the national growth rates are weakly correlated with the sensitivity to misspecification elasticities. In contrast, the elasticities are quite related to the variation in the industry shares across locations—$\text{Var}(z_k)$. This observation explains why the industries with high weight tend to be tradables: almost by definition, tradables have industry shares that vary across locations, while non-tradables do not.[24]

---

[22]In Appendix J, 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.

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

[24]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.

5.3 Testing the plausibility of the identifying assumption

Test 1: Correlates of 1980 industry shares   Table 3 shows the relationship between 1980 characteristics of commuting zones and the share of the top 5 industries in Table 2, as well the overall Bartik instrument using 1980 to 1990 growth rates. First, the R^2 in these regressions are quite high: for example, we can explain 43% of the variation in share of the "other" industry via our covariates. Second, "other," oil and gas extraction, blast furnaces, and the overall Bartik instrument are statistically significantly correlated with the share of native-born workers. In the immigrant enclave literature, the share of native born (immigrant share) is thought to predict labor supply shocks.

Test 3: Alternative estimators and overidentification tests   Rows 1 and 2 of Table 4 report the OLS and IV estimates, with and without for the 1980 covariates as controls (we discuss these covariates below) and makes two main points. First, the IV estimates are bigger than the OLS estimates. Second, the Bartik results are sensitive to the inclusion of controls, though these are not statistically distinguishable.

Rows 3-6 of Table 4 report alternative estimators as well as overidentification tests. We focus on column (2), where we control for covariates. TSLS with the Bartik instrument and LIML are quite similar. This finding is typically viewed as reassuring. In contrast, TSLS and MBTSLS are similar, while HFUL is substantially larger. The different point estimates suggest the presence of misspecification. In column (4), we see that the overidentification tests reject the null that all instruments are exogenous, which also points to misspecification.

Test 4: Plausibility of many invalid instrument asymptotics   As previously discussed, Table 2 shows that the top five instruments account for almost a third ($^{0.592}/_{1.968}$) of the positive weight in the estimator. This finding suggests that the many invalid instrument asymptotics do not provide a good approximation to the finite sample performance of the estimator.

6   Empirical example II: China shock

We estimate the effect of Chinese imports on manufacturing employment in the United States using the China shock approach of Autor, Dorn, and Hanson (2013) (ADH).

6.1   Specification

It is helpful to write main regression specification of ADH in our notation. The paper is interested in a regression (where we omit covariates for simplicity, but include them in the

regressions):

$$y_{lt} = \beta_0 + \beta x_{lt} + \epsilon_{lt}, \tag{6.1}$$

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

$$x_{lt} = \gamma_0 + \gamma_1 B_{lt} + \eta_{lt}, \tag{6.2}$$

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

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.

6.2   Rotemberg weights

As in the canonical setting, despite a very large number of instruments (397 industries) 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 receive about half of the absolute weight in the estimator ($0.628/1.379$). These instruments are games and toys, electronic computers, household audio and video, computer equipment and telephone apparatus. Except for games and toys, these industries are different than the ones that ADH emphasize when motivating the empirical strategy.[25]

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 (0.581^2) of the variance of the Rotemberg weights.

6.3   Testing the plausibility of the identifying assumption

Test 1: Correlates of 1980 industry shares   Table 6 shows the relationship between the covariates used in ADH and the top industries reported in Table 5. First, relative to the

---

[25]"The main source of variation in exposure is within-manufacturing specialization in industries subject to different degrees of import competition...there is differentiation according to local labor market reliance on labor-intensive industries...By 2007, China accounted for over 40 percent of US imports in four four-digit SIC industries (luggage, rubber and plastic footwear, games and toys, and die-cut paperboard) and over 30 percent in 28 other industries, including apparel, textiles, furniture, leather goods, electrical appliances, and jewelry" (pg. 2123).

canonical setting, the controls explain less of the variation in shares. Second, electronic computers, computer equipment manufacturing as well as the overall measure are both concentrated in more college educated areas; in contrast, games and toys is concentrated in places with fewer college educated workers. This pattern emphasizes that researchers should be concerned about other trends potentially affecting manufacturing employment in more educated areas. Interestingly, the identifying assumption related to the computer industry is precisely one that ADH worry about.[26]

Test 2: Parallel pre-trends We construct our pre-trend figures as follows. We use fixed 1980 shares as the instruments, and plot the reduced form effect of each industry on manufacturing employment.[27] We then convert the growth rates to levels and we index the levels in 1970 to 100. Standard errors are constructed using the delta method. For the aggregate Bartik, we use the industry shares fixed in 1980, and aggregate them using growth rates from 1990 to 2000.

Figure 1 shows the plots and displays several interesting patterns. First, all of the panels diverge from classic pre-trends figures, which show no trends in the pre-periods and then a sharp change at the date of the treatment. Second, as was true in the covariates in Table 6, the patterns in electronic computers (Panel B) and computer equipment (Panel D) are similar to the aggregate, with the decline in manufacturing from 1990 to 2007 undoing growth from 1970 to 1990. Note that Panel B and D shows comparisons of places with more and less of these particular industries in 1980, while the outcome is employment *for all manufacturing industries*.

Test 3: Alternative estimators and overidentification tests Rows 1 and 2 of Table 7 report the OLS and IV estimates using Bartik, with and without for the 1980 covariates as controls, though these are not statistically distinguishable for the IV estimates. Rows 3-6 of Table 7 shows alternative estimators as well as overidentification tests. We focus on column (2), where we control for covariates. The estimates range from half the size of the baseline Bartik TSLS estimate (MBTSLS), to several times the size (LIML). The divergence between the two-step estimators (TSLS with Bartik, TSLS and MBTSLS) and the maximum likelihood estimators (LIML and HFUL) is evidence of misspecification. Similarly, the overidentification tests reject. Combined, the movement in the estimates across estimators is not reassuring[28],

---

[26] ADH (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."

[27] In unreported results, we find that this specification delivers the same point estimate as the time-varying (but lagged) shares that ADH use, albeit with larger standard errors. See Figure A1 for the analogous figures using fixed 1990 shares. And Figure A2 shows analogous figures for the time-varying shares.

[28] Angrist and Pischke (2008, pg. 213) write: "Check overidentified 2SLS estimates with LIML. LIML is less precise than 2SLS but also less biased. If the results come out similar, be happy. If not, worry..."

and the failure of the overidentification tests points to potential misspecification.

Test 4: Plausibility of many invalid instrument asymptotics   As previously discussed, Table 5 shows that the top five instruments account for almost half ($0.628/1.379$) of the positive weight in the Bartik estimator. This finding suggests that the many invalid instrument asymptotics do not provide a good approximation to the finite sample performance of the estimator.

7   Empirical example III: defense spending shocks

In our third application, we examine the use of defense spending shocks to estimate the fiscal multiplier following Nakamura and Steinsson (2014). To map to the Bartik setting, the equivalent of local industry shares times time is the national industry growth rates times location, and the equivalent of the national industry rates is initial location industry share (the average level of military spending in that state relative to state output in the first five years of the sample). Notably, Nakamura and Steinsson (2014) implement both TSLS using the initial location industry share, and TSLS using the Bartik measure.

In Table 8, we report the Rotemberg weights associated with their benchmark specification (Table 2, column 2 of their paper). Consistent with the previous examples, the weights are very skewed: the top region receives half of the positive weight ($1.001/1.933$).

8   Empirical example IV: simulated instruments

Finally, we look at a simulated instrument (Currie and Gruber (1996a) and Currie and Gruber (1996b)) application. Specifically, we look at Cohodes et al. (2016), which studies the effect of Medicaid expansion on educational attainment. We begin by discussing how to map the simulated instrument example to our setting. We then compute the sensitivity elasticities to illustrate how the weights can be useful for researchers to clarify their identifying variation.

8.1   A simulated instrument as a Bartik instrument

A simplified version of the estimating equation in Cohodes et al. (2016) is

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

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 variation in 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: $x_l = \sum_k z_{lk} g_{lk}$.

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. Thus, under plausible asymptotics the instrument is the cross-state differences in eligibility for each eligibility type.

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. 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, and 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. The large number of instruments explains why Currie and Gruber (1996b, pg. 446) describe a simulated instrument as "a convenient parameterization of legislative differences."

8.2   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.[29] The second dimension is the characteristics of households

---

[29]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

affected by the different instruments.[30]

Figure 2 summarizes the state-year eligibility changes that drive estimates, and hence the state-years a researcher should be looking for potential confounds. There are several things to note. First, the empirical strategy leverages variation that is spread throughout time. Second, there are a few large spikes. The notable spike is in 1990, which generated changes in eligibility in all 51 states (and DC).[31] Third, the figure highlights the six most important state-law changes. 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.

Figure 3 summarizes which types of households get more weight in estimates. 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 robustness analysis in Cohodes et al. (2016, Table 5 and 6).

9   Summary

The central contribution of this paper revolves around understanding identification and the Bartik instrument. Our first set of formal results relate to identification in the sense typically used by econometricians. We show that Bartik is numerically equivalent to a GMM estimator with the industry shares as instruments. We then argue that under plausible asymptotics 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

---

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.

[30]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.

[31]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 the federal poverty line (whether or not they were receiving AFDC assistance). 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.

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 results. Our results clarify the set of reasonable concerns a consumer of the Bartik literature should have. We hope that researchers will use the results and tools in this paper to be clearer about how identification works in their papers: both in the econometric sense of stating the identifying assumption, and in the practical sense of showing what variation drives estimates.

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.

Adao, Rodrigo, Michal Kolesar, and Eduardo Morales. 2018. "Shift-Share Designs: Theory and Inference." Working paper.

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.

Anatolyev, Stanislav. 2013. "Instrumental variables estimation and inference in the presence of many exogenous regressors." *The Econometrics Journal* 16:27–72.

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.

Angrist, Joshua D. and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106 (4):979–1014.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

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.

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

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, Peter Hull, and Xavier Jaravel. 2018. "Quasi-experimental Shift-share 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.

Cragg, John G and Stephen G Donald. 1993. "Testing identifiability and specification in instrumental variable models." *Econometric Theory* 9 (2):222–240.

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.

DiNardo, John and David S Lee. 2011. "Program evaluation and research designs." In *Handbook of labor economics*, vol. 4. Elsevier, 463–536.

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

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.

Lucca, David O, Taylor Nadauld, and Karen Chen. Forthcoming. "Credit supply and the rise in college tuition: Evidence from the expansion in federal student aid programs." *Review of Financial Studies* .

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

Nakamura, Emi and Jon Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3):753–792.

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: Monte Carlo evidence on many invalid instruments

	OLS	Feasible Bartik			Infeasible Bartik			Top 5 α_k share	
	$\hat{\mathbb{E}}\left[\hat{\beta}\right]$ (1)	$\hat{\mathbb{E}}\left[\hat{\beta}\right]$ (2)	Med[$\hat{\beta}$] (3)	$\hat{\mathbb{E}}\left[F\right]$ (4)	$\hat{\mathbb{E}}\left[\hat{\beta}\right]$ (5)	Med[$\hat{\beta}$] (6)	$\hat{\mathbb{E}}\left[F\right]$ (7)	$\hat{\mathbb{E}}$ (8)	Med (9)
(1) Standard	2.73	1.95	1.98	33.63	1.98	2.01	35.59	0.50	0.50
(2) $\sigma^2_{\lambda_k} = 0.2\sigma^2_{g_k}$	2.73	1.94	1.98	32.18	1.97	2.01	34.07	0.49	0.49
(3) $\sigma^2_{\lambda_k} = 1.0\sigma^2_{g_k}$	2.73	1.94	1.98	34.40	1.97	2.00	36.36	0.50	0.51
(4) $\sigma^2_{\lambda_k} = 5.0\sigma^2_{g_k}$	2.72	1.93	1.94	35.05	1.96	1.97	36.97	0.50	0.51
(5) $\sigma^2_{\lambda_k} = 1.0\sigma^2_{g_k}$, $\lambda_k = g_k$ (smallest 5 α_k)	2.73	1.94	1.98	33.66	1.97	2.00	35.62	0.50	0.51
(6) $\sigma^2_{\lambda_k} = 1.0\sigma^2_{g_k}$, $\lambda_k = g_k$ (largest 5 α_k)	2.75	2.55	2.57	34.00	2.56	2.58	35.93	0.50	0.51

Notes: This table reports Monte Carlo simulations of the performance of the Bartik estimator. The true value of β is 2. Column (1) reports OLS estimates. Column (2) through (4) report feasible Bartik estimates, where the growth rates are estimated using a leave-one-out estimator. Column (2) is the mean of the $\hat{\beta}$, column (3) is the median, and column (4) reports the mean of the first stage F-statistics. Columns (5) to (7) repeat the same exercise, except that it reports infeasible Bartik where we use the true value of g_k to construct the Bartik instrument. Columns (8) and (9) report the mean and median share of the positive weight for the top-5 Rotemberg weight industries. Row (1) reports a simulation where the Bartik instrument is valid. Rows (2) through (4) add a component of the error term $\sum_k z_{lk}\lambda_k$ where λ_k is drawn independently of g_k. Rows (5) and (6) replace 5 draws of the λ_k with the g_k for the kth industry where the five industries are selected on the basis of the Rotemberg weights: Row (5) picks the five industries with the smallest weights (in absolute value) and row (6) repeats the exercise with the industries with the five largest weights. In these simulations there are 800 locations and 228 industries, where the industry shares are drawn from the empirical distribution of 3 digit IND1990 industries in 1980. All random variables are normally distributed and the variances are as follows: $\sigma^2_{g_k} = 0.0046$, $\sigma^2_{g_{lk}} = 0.01$, $\sigma^2_{g_l} = 0.001$, and $\sigma^2_{\epsilon} = 0.0245$. The Table reports the mean and median values of $\hat{\beta}$ across 2000 simulation runs. See Appendix F for additional details.

Table 2: Summary of Rotemberg weights: canonical setting

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.968	-0.003	0.444
Positive	1.968	0.006	0.501

Panel B: Correlations of Industry Aggregates

	α_k	g_k	β_k	$\text{Var}(z_k)$
α_k	1			
g_k	-0.049	1		
β_k	-0.085	0.047	1	
$\text{Var}(z_k)$	0.576	-0.164	-0.019	1

Panel C: Variation across years in α_k

	Sum	Mean
1980	0.461	0.002
1990	0.177	0.001
2000	0.362	0.002

Panel D: Top 5 Rotemberg weight industries

	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Ind Share
Oil+Gas Extraction	0.231	-0.002	1.778	(1.25,2.85)	1.601
Motor Vehicles	0.139	-0.012	1.381	(1.2,1.65)	5.006
Other	0.093	-0.019	1.344	(-10,10)	4.840
Guided Missiles	0.070	0.029	0.392	(-4.1,.75)	1.074
Blast furnaces	0.059	-0.034	1.176	(.8,2.6)	1.831

Panel E: Summary of $\hat{\beta}_k$

	Mean	Median	25th P	75th P	Share Negative
β_k	1.637	0.775	2.181	-0.609	0.349

Notes: This table reports statistics about the Rotemberg weights. In all cases, we report statistics about the aggregated weights, where we aggregate a given industry across years as discussed in Section 3.4. 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 ($\text{Var}(z_k)$). Panel C reports variation in the weights across years. Panel D reports the top five industries according to the Rotemberg weights. The g_k is the national industry growth rate, 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

	Oil+gas	Motor Vehicles	Other	Guided Missiles	Blast Furnaces	Bartik (1980 shares)
Male	40.94	-25.86	98.73	36.14	14.14	-0.64
	(8.39)	(7.61)	(14.91)	(18.38)	(7.21)	(0.05)
White	1.07	-23.61	-34.41	14.85	-32.32	-0.08
	(3.09)	(28.37)	(7.66)	(14.84)	(13.66)	(0.06)
Native Born	10.96	-3.92	21.25	-42.83	-14.99	-0.19
	(2.94)	(5.38)	(6.95)	(41.64)	(6.85)	(0.04)
12th Grade Only	-32.51	64.36	26.75	-68.41	16.09	0.35
	(7.53)	(17.72)	(8.88)	(27.39)	(8.25)	(0.06)
Some College	-9.51	23.24	24.12	28.93	-40.43	0.63
	(4.63)	(22.71)	(7.12)	(25.74)	(11.00)	(0.07)
Veteran	-10.57	19.09	-142.65	86.93	111.08	0.55
	(7.05)	(41.04)	(22.65)	(43.69)	(30.84)	(0.12)
# of Children	-2.75	45.35	-57.66	11.33	5.53	0.27
	(4.62)	(23.48)	(13.46)	(21.27)	(9.77)	(0.06)
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	722	722	722	722	722	722
R^2	0.22	0.11	0.43	0.25	0.21	0.58

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: canonical setting

| | Δ Emp | | Coefficient Equal | Over ID test |
	(1)	(2)	(3)	(4)
OLS	0.71	0.63	[0.69]	
	(0.06)	(0.07)		
TSLS (Bartik)	1.75	1.28	[0.14]	
	(0.34)	(0.15)		
TSLS	0.74	0.67	[0.08]	1014.05
	(0.06)	(0.07)		[0.00]
MBTSLS	0.76	0.69	[0.17]	
	(0.04)	(0.06)		
LIML	1.60	1.42	[0.86]	2820.96
	(0.06)	(0.06)		[0.00]
HFUL	2.85	2.69	[0.00]	804.19
	(0.15)	(0.13)		[0.00]
Year and CZone FE	Yes	Yes		
Controls	No	Yes		
1980 Population Weighted	Yes	Yes		
Observations	2,166	2,166		

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

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

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.379	-0.001	0.477
Positive	1.379	0.003	0.523

Panel B: Correlations of Industry Aggregates

	α_k	g_k	β_k	$Var(z_k)$
α_k	1			
g_k	0.581	1		
β_k	-0.005	-0.041	1	
$Var(z_k)$	0.154	-0.038	0.054	1

Panel C: Variation across years in α_k

	Sum	Mean
1990	0.329	0.001
2000	0.671	0.002

Panel D: Top 5 Rotemberg weight industries

	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Ind Share
Games and Toys	0.182	174.841	-0.151	(-0.40,0.20)	0.270
Electronic Computers	0.182	85.017	-0.620	(-1.55,-0.05)	1.091
Household Audio and Video	0.130	118.879	0.287	(-0.10,5.70)	0.378
Computer Equipment	0.076	28.110	-0.315	(-1.60,0.20)	0.519
Telephone Apparatus	0.058	37.454	-0.305	(-10.00,10.00)	0.920

Panel E: Summary of $\hat{\beta}_k$

	Mean	Median	25th P	75th P	Share Negative
β_k	-0.909	-0.514	0.734	-1.687	0.633

Notes: This table reports statistics about the Rotemberg weights. In all cases, we report statistics about the aggregated weights, where we aggregate a given industry across years as discussed in Section 3.4. 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 -10 to 10, and Ind Share is the industry share (multiplied by 100 for legibility). Panel E reports statistics about the dispersion in the $\hat{\beta}_k$.

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

	Games and toys	Electronic computers	Household audio and video	Computer equipment	Telephone apparatus	China to other
Share Empl in Manufacturing	0.01	0.21	0.08	0.21	-0.07	0.57
	(0.03)	(0.18)	(0.08)	(0.15)	(0.06)	(0.07)
Share College Educated	-0.08	0.20	0.01	0.22	-0.07	0.30
	(0.03)	(0.11)	(0.04)	(0.10)	(0.06)	(0.06)
Share Foreign Born	0.01	-0.01	-0.02	-0.01	-0.08	0.15
	(0.01)	(0.04)	(0.01)	(0.04)	(0.03)	(0.03)
Share Empl of Women	0.05	-0.04	-0.08	-0.02	-0.02	0.10
	(0.03)	(0.12)	(0.05)	(0.12)	(0.07)	(0.06)
Share Empl in Routine	0.04	-0.37	0.06	-0.36	-0.01	-0.08
	(0.03)	(0.14)	(0.05)	(0.12)	(0.07)	(0.13)
Avg Offshorability	0.02	0.33	0.00	0.29	0.23	-0.24
	(0.02)	(0.10)	(0.05)	(0.08)	(0.04)	(0.09)
1980 Population Weighted	Yes	Yes	Yes	Yes	Yes	Yes
N	1,444	1,444	1,444	1,444	1,444	1,444
R^2	0.02	0.08	0.01	0.08	0.05	0.22

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. The final column is constructed using 1990 to 2000 growth rates. Results are weighted by the population in the period the characteristics are measured. Standard errors in parentheses.

Table 7: OLS and IV estimates: Autor, Dorn, and Hanson (2013)

| | Δ Emp | | Coefficients Equal | Over ID Test |
	(1)	(2)	(3)	(4)
OLS	-0.38	-0.17	[0.00]	
	(0.07)	(0.04)		
TSLS (Bartik)	-0.68	-0.62	[0.33]	
	(0.09)	(0.11)		
TSLS	-0.46	-0.22	[0.00]	872.69
	(0.07)	(0.06)		[0.00]
MBTSLS	-0.61	-0.33	[0.00]	
	(0.07)	(0.05)		
LIML	-1.57	-2.07	[0.82]	1348.50
	(0.82)	(3.52)		[0.00]
HFUL	-1.14	-1.13	[0.64]	1141.08
	(0.04)	(0.04)		[0.00]
Year and Census Division FE	Yes	Yes		
Controls	No	Yes		
Observations	1,444	1,444		

Notes: This table reports a variety of estimates of the effect of rising imports from China on US manufacturing employment. The regressions are at the CZ level and include two time periods (1990 to 2000, and 2000 to 2007). The TSLS row is our replication of Column (1) and Column (6) of Table 3 in ADH. Column (1) does not contain controls, while column (2) does. The TSLS (Bartik) row uses the Bartik instrument. The TSLS row uses each industry share (times time period) separately as instruments. The MBTSLS row uses the estimator of Anatolyev (2013) and Kolesar et al. (2015) using each industry share (times time period) separately as instruments. The LIML row shows estimates using the limited information maximum likelihood estimator. Finally, the HFUL row uses the HFUL estimator of Hausman et al. (2012). The J-statistic comes from Chao et al. (2014). The p-value for the equality of coefficients compares the adjacent columns with and without controls. The controls are the contemporaneous characteristics displayed in Table 6. Results are weighted by start of period population. Standard errors are in parentheses.

Table 8: Summary of Rotemberg weights: Nakamura and Steinsson (2014)

Panel A: Negative and positive weights

	Sum	Mean	Share
Negative	-0.933	-0.155	0.600
Positive	1.933	0.483	0.400

Panel B: Correlations of Aggregates

	α_k	g_k	β_k
α_k	1		
g_k	0.877	1	
β_k	0.081	0.111	1

Panel C: Top 5 Rotemberg weight (region times growth rates)

	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI
CT MA ME NH RI VT	1.001	1.825	1.739	(1.6,1.9)
TX OK LA AR	0.400	1.202	1.185	(.95,1.4)
CA WA OR AK HI	0.295	1.614	5.727	(5.25,6.25)
MO KS IA NE MN SD ND	0.237	0.964	-1.841	(-2.25,-1.45)
NC SC GA FL	-0.133	0.771	5.647	(5.05,6.3)

Panel D: Summary of $\hat{\beta}_k$

	Mean	Median	25th P	75th P	Share Negative
β_k	1.514	1.846	5.040	-2.016	0.400

Notes: This table reports statistics about the Rotemberg weights. In all cases, we report statistics about the aggregated weights, where we aggregate a national growth rate time region as discussed in Section 3.4. 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) and the just-identified coefficient estimates. Panel C reports the top five regions times national growth rates according to the Rotemberg weights. The g_k is the share of defense spending in the region, 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). Panel D reports statistics about the dispersion in the $\hat{\beta}_k$.

Figure 1: Pre-trends for high Rotemberg weight industries: Autor, Dorn, and Hanson (2013)
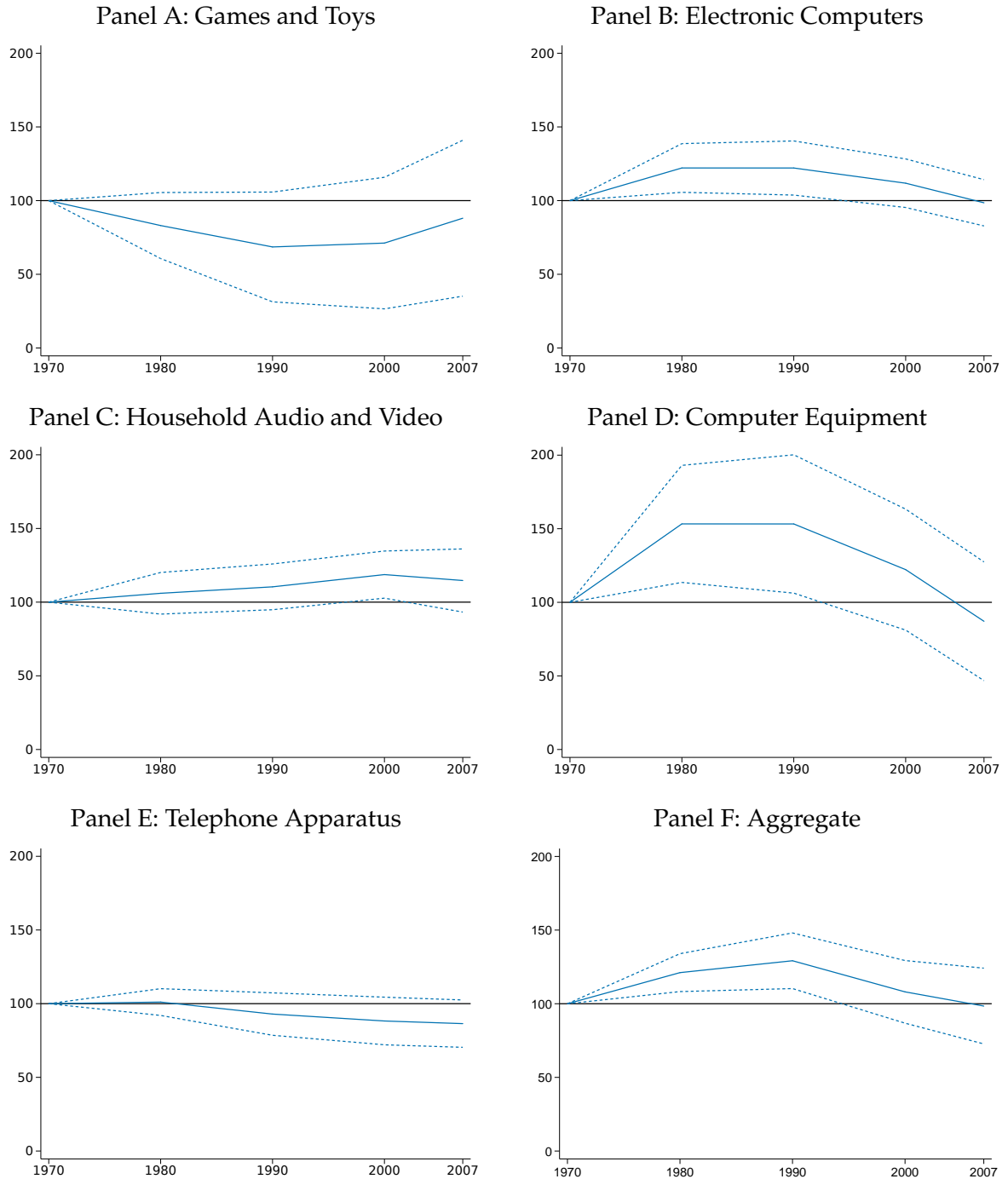
Panel A: Games and Toys

Panel B: Electronic Computers

Panel C: Household Audio and Video

Panel D: Computer Equipment

Panel E: Telephone Apparatus

Panel F: Aggregate

Notes: These figures report pre-trends for the overall instrument and the top-5 Rotemberg weight industries as reported in Table 5. The Figures fix industry shares at the 1980 values and report the effect of these industry shares on manufacturing employment. We run regressions in growth rates and then convert to levels. We normalize 1970 to 100, and compute the standard errors using the delta method.

Figure 2: Rotemberg weights by state-year policy changes and income

Panel A: Overall
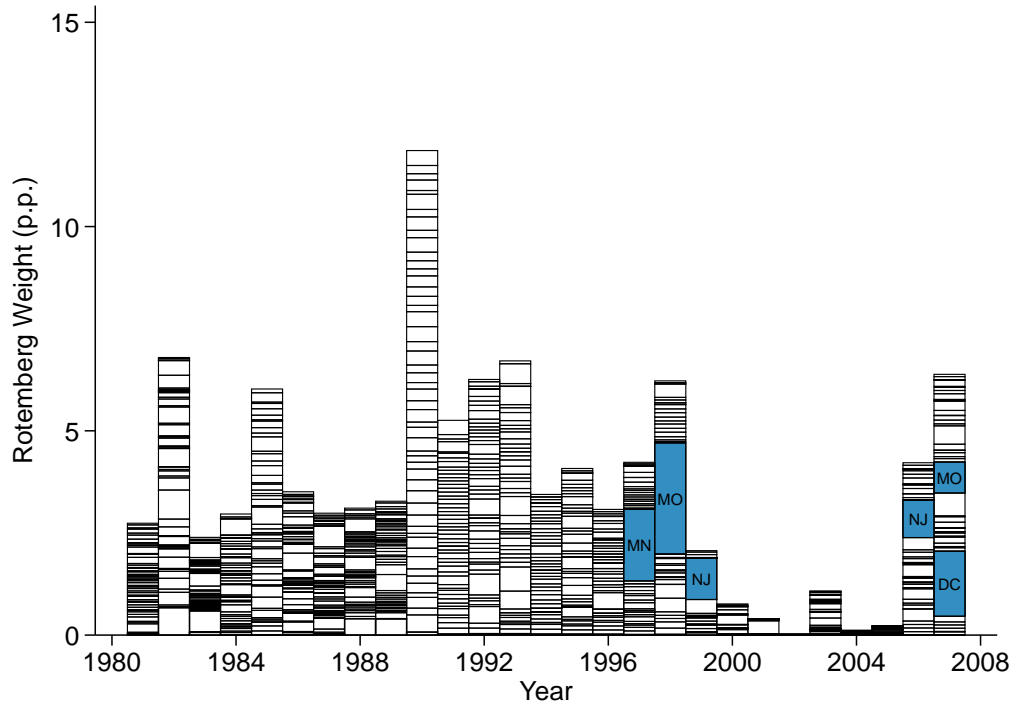
Panel B: Low-income (< $10,000 in 1986 USD)   Panel C: High-income (≥ $10,000 in 1986 USD)
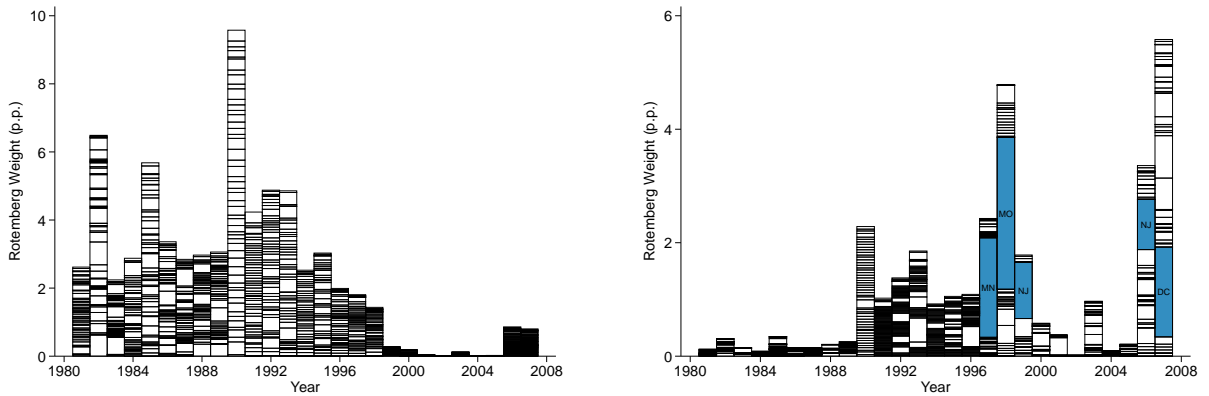
Notes: These figures report the state-year policy changes to which the estimates are most sensitive to misspecification. The figure is constructed by weighting the state-year Medicaid eligibility *changes* that go into the definition of eligibility types. The state-year Medicaid eligibility changes are 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 eligibility types by their mean income and so the combination of Panels (b) and (c) sum to Panel (a).

Figure 3: Rotemberg weights by charactestics

Panel A: Number of Children
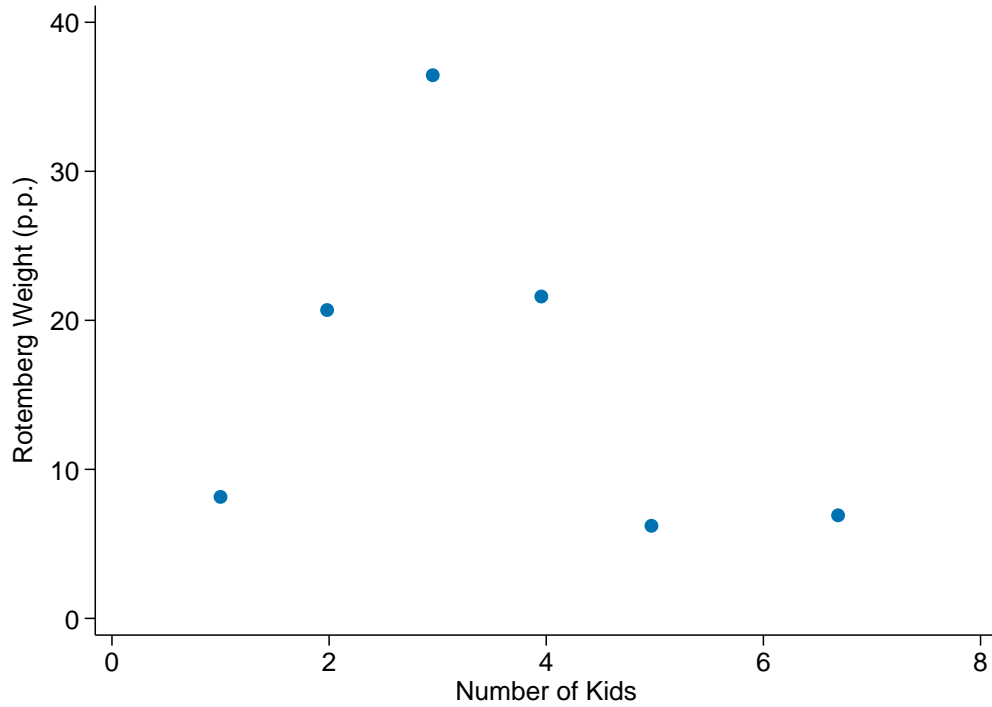
Panel B: Family Income
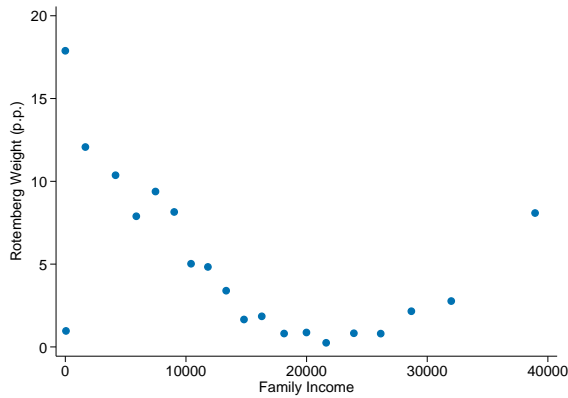
Panel C: Child Age
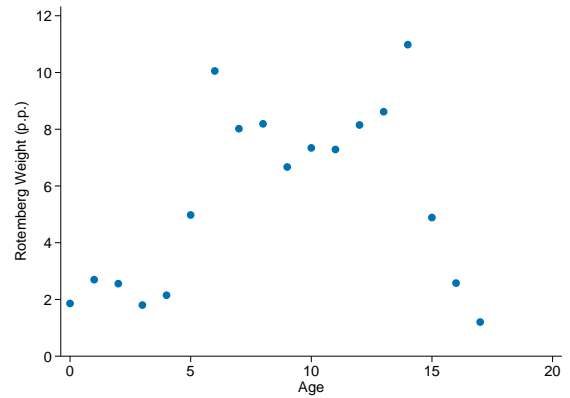
Notes: These figures report the characteristics of the eligibility types to which the estimates are most sensitive to misspecification. Each of the three panels reports computing the average characteristics of each eligibility type, sorting the eligibility types according to the characteristic, and then 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. Natu-

rally, 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.

$$
\begin{aligned}
\hat{\beta}_{GMM} &= \frac{X^{\perp'}ZGG'Z'Y^{\perp}}{X^{\perp'}ZGG'Z'X^{\perp}} \\
&= \frac{X^{\perp'}BB'Y^{\perp}}{X^{\perp'}BB'X^{\perp}} \\
&= \hat{\beta}_{Bartik},
\end{aligned}
$$

where $X^{\perp'}B$ is a scalar and so cancels. □

Proposition 3.1

Proof. The proof is algebra:

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

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

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

□

Proof of Lemma 3.1.

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 $L \to \infty$. $\qquad\square$

Proof of Proposition 3.2

Proof. First, note that

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

$$\hat{\beta}_k - \beta_0 = L^{-1/2} \frac{\sum_{l,t} z_{lkt} V_{lt}}{\sum_{l,t} z_{lkt} x_{lt}} + \frac{\sum_{l,t} z_{lkt} \epsilon_{lt}}{\sum_{l,t} z_{lkt} x_{lt}}.$$

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 $\mathbb{E}[\tilde{\beta}] = \sum_k \alpha_k(W) \frac{\Sigma_{ZV,k}}{\Sigma_{ZX^{\perp},k}} = \sum_k \alpha_k(W) \mathbb{E}[\tilde{\beta}_k]$. $\qquad\square$

Proposition 3.3

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 $\alpha_{k'}(W_{-k}) = \alpha_{k'}(W)/(1 - \alpha_k(W))$.[32] This gives:

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

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

$\square$

Proposition 4.1

Proof.

$$\alpha_k = \frac{g_k Z'_k X}{\sum_k g_k Z'_k X}$$

$$\alpha_k = \frac{g_k \bar{z}_{L,k} g_k + g_k \bar{z}_{L,k} \sum_{l,st.z_{lk}=1} \tilde{g}_{lk}}{\sum_k \left(g_k \bar{z}_{L,k} g_k + g_k \bar{z}_{L,k} \sum_{l,st.z_{lk}=1} \tilde{g}_{lk} \right)}$$

$$\alpha_k = \frac{g_k^2 \bar{z}_{L,k} + g_k \bar{z}_{L,k} \sum_{l,st.z_{lk}=1} \tilde{g}_{lk}}{\sum_k \left(g_k^2 \bar{z}_{L,k} + g_k \bar{z}_{L,k} \sum_{l,st.z_{lk}=1} \tilde{g}_{lk} \right)}$$

Note that as $L \to \infty$, the numerator remains bounded for any k, but the denominator grows. Hence, $\alpha_k \to 0$. $\square$

C Equivalence with K industries, L locations, and controls

The two stage least squares system of equations is:

$$y_{lt} = D_{lt}\rho + x_{lt}\beta + \epsilon_{lt} \tag{A1}$$

$$x_{lt} = D_{lt}\tau + B_{lt}\gamma + \eta_{lt}, \tag{A2}$$

---

[32]Note that with TSLS, 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 $1 \times S$ 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 $\mathbf{P_D} = \mathbf{D}(\mathbf{D'D})^{-1}\mathbf{D'}$ 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

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

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

$$y_{lt} = D_{lt}\rho + x_{lt}\beta + \epsilon_{lt} \tag{A4}$$
$$x_{it} = D_{lt}\tau + Z_{lt}\gamma_t + \eta_{lt}, \tag{A5}$$

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 Section 1.2, the t subscript allows the effect of a given industry share to be time-varying. In matrix notation, we write

$$Y = \mathbf{D}\rho + X\beta + \epsilon \tag{A6}$$
$$X = \mathbf{D}\tau + \tilde{Z}\mathbf{\Gamma} + \eta, \tag{A7}$$

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. Then, using the $\tilde{\mathbf{Z}}$ as instruments, the GMM estimator is:

$$\hat{\beta}_{GMM} = \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{A10}$$

where Ω is a $KT \times KT$ weight matrix.

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

Proof. Start with the Bartik estimator,

$$\hat{\beta}_{Bartik} = \frac{B'\mathbf{M_D}Y}{B'\mathbf{M_D}X} \tag{A11}$$

$$= \frac{G'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{G'\tilde{\mathbf{Z}}'\mathbf{M_D}X} \tag{A12}$$

$$= \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{A13}$$

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

D   Consistency under Many Invalid Instruments

D.1   Setup

For simplicities' sake, we will fix the number of controls, and focus on the many instruments version of this proof. These assumptions follow almost directly from Kolesar et al. (2015). Our general instrumental variables model follows as:

$$Y_l = X_l\beta_0 + D_l\rho + Z_l\lambda + \epsilon_l \tag{A1}$$
$$X_l = Z_lG\pi_1 + D_l\pi_2 + \nu_l. \tag{A2}$$

Let $\mathbf{Y}$ be the L-component vector with Y_l as its lth element, $\mathbf{X}$ the L-component vector with lth element X_l, $\boldsymbol{\epsilon}$ the L-component vector with lth element ϵ_l, the L-component vector with lth element X_l, $\mathbf{D}$ be the $L \times S$-matrix with lth row equal to D_l, $\mathbf{Z}$ be the $L \times K_L$-matrix with lth row equal to Z_l and $\mathbf{B} = \mathbf{Z}'G$ be the L-component vector with the lth element $B_l = Z_l'G$. Let $\bar{\mathbf{Z}} = (\mathbf{Z}, \mathbf{D})$ be the full matrix of exogeneous variables. If $\sum_k z_{lk} = 1$, we drop one of the $\mathbf{Z}$ instruments such that the full rank condition holds below. Finally, let $\mathbf{P_D} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ and $\mathbf{M_D} = \mathbf{I} - \mathbf{P_D}$. Recall we write $\mathbf{M_D}X = X^{\perp}$.

These parameters can be used to define the following matrix Λ_L:

$$\Lambda_L = \begin{pmatrix} \Lambda_{L,11} & \Lambda_{L,12} \\ \Lambda_{L,12} & \Lambda_{L,22} \end{pmatrix} = \begin{pmatrix} \lambda & G\pi_1 \end{pmatrix}' \mathbf{Z}^{\perp\prime}\mathbf{Z}^{\perp} \begin{pmatrix} \lambda & G\pi_1 \end{pmatrix}. \tag{A3}$$

ASSUMPTION 4 (Instruments and exogeneous variables).   *(i) $Z_l \in R^{K_L}, G \in R^{K_L}, D_l \in R^S, \epsilon_l \in R, \nu_l \in R$, for $l = 1, \dots, L$, $L = 1, \dots$ are triangular arrays of random variables with $(G, Z_l, W_l, \epsilon_l, \eta_l), l = 1, \dots, L$ exchangeable.*

(ii) $\bar{\mathbf{Z}}$ is full column rank with probability one.

ASSUMPTION 5 (Model). *$(\epsilon_l, \nu_l)'|\mathbf{Z}, \mathbf{D}, G$ are iid with mean zero, positive definite covariance matrix Σ and finite fourth moments.*

ASSUMPTION 6 (Number of instruments). *For some $0 \le c_K < 1$,*

$$K_L/L = c_K + o(L^{-1/2}). \tag{A4}$$

ASSUMPTION 7 (Concentration parameter). *For some positive semidefinite 2×2 matrix Λ with $\Lambda_{22} > 0$,*

$$\Lambda_L / L \to^p \Lambda \qquad \text{and} \qquad \mathbb{E}[\Lambda_L / L] \to \Lambda. \tag{A5}$$

ASSUMPTION 8 (Zero correlation). $\Lambda_{12} = 0$.

DEFINITION D.1. *Let*

$$\hat{\beta}_{bartik} = \left(\mathbf{B}'\mathbf{X}^\perp \right)^{-1} \left(\mathbf{B}'\mathbf{Y}^\perp \right). \tag{A6}$$

THEOREM D.1. *Assume Assumptions 4, 5, 6, 7,8 hold. Then,*

$$\beta_{bartik} - \beta_0 = 0. \tag{A7}$$

Proof. Note the reduced-form for β_{bartik}:

$$(Y_l \ X_l) = Z_l'(\psi_1 \ G\pi_1) + D_l'(\psi_2 \ \pi_2) + V_l', \tag{A8}$$

where $\psi_1 = \lambda + G\pi_1\beta_0$, $\psi_2 = \delta + \pi_2\beta_0$ and $V_l = (\epsilon_l + v_l\beta, v_l)'$, and let $\mathbf{V}$ be the $L \times 2$ matrix with lth row equal to V_l'. Let $\Pi = (\psi_1, \pi_1)$, and $\overline{\mathbf{Y}}^\perp = \mathbf{M_D}(\mathbf{Y} \ \mathbf{X})$. Finally, let $\Omega = E[V_l V_l']$. Then, let

$$\Gamma = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix}. \tag{A9}$$

Thus,

$$\Sigma = \Gamma^{-1\prime}\Sigma\Gamma^{-1} = \begin{pmatrix} \Sigma_{11} + 2\Sigma_{12}\beta_0 + \Sigma_{22}\beta_0^2 & \Sigma_{12} + \Sigma_{22}\beta_0 \\ \Sigma_{21} + \Sigma_{22}\beta_0 & \Sigma_{22} \end{pmatrix} \tag{A10}$$

First, note by Lemma A.3 of Kolesar et al. (2015) that

$$\overline{\mathbf{Y}}^{\perp\prime}\overline{\mathbf{Y}}^\perp / L \to^p \Psi + \Omega \tag{A11}$$

$$\overline{\mathbf{Y}}^{\perp\prime}\mathbf{P_{Z^\perp}}\overline{\mathbf{Y}}^\perp / L \to^p \Psi + \alpha_K\Omega, \tag{A12}$$

where

$$\Psi = \begin{pmatrix} \Lambda_{11} + 2\Lambda_{12}\beta_0 + \Lambda_{22}\beta_0^2 & \Lambda_{12} + \Lambda_{22}\beta_0 \\ \Lambda_{12} + \Lambda_{22}\beta_0 & \Lambda_{22} \end{pmatrix}. \tag{A13}$$

Note that under the Assumption 8, this simplifies to

$$\Psi = \begin{pmatrix} \Lambda_{11} + \Lambda_{22}\beta_0^2 & \Lambda_{22}\beta_0 \\ \Lambda_{22}\beta_0 & \Lambda_{22} \end{pmatrix}. \tag{A14}$$

Then, under a slight modification, we now show that

$$\overline{\mathbf{Y}}^{\perp\prime}\mathbf{P_{B^\perp}}\overline{\mathbf{Y}}^\perp / L \to^p \Psi + \alpha_K\Omega, \tag{A15}$$

First, note that

$$\mathbb{E}[\mathbf{V}'\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L|\mathbf{Z}^\perp, G_L] = (1/N)\Sigma \tag{A16}$$

since B_l is scalar. Fix $a \in \mathbb{R}^2$. By an identical argument to Kolesar et al. (2015) for Lemma A.3,

$$Var(\mathbf{V}'\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L) = O(L^{-2}), \tag{A17}$$

with a slight difference from Kolesar et al. (2015) because the rank of $\mathbf{B} = 1$. Hence

$$\mathbf{V}'\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L \to^p 0. \tag{A18}$$

This is standard in a just-identified setting.

Next, note by the same argument from Kolesar et al. (2015) in Lemma A.3,

$$\Pi\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L \to^p 0. \tag{A19}$$

This follows from Assumption 5, since $\mathbb{E}[\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L] = 0$, and from the fact that

$$Var(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L) = \mathbb{E}\left[Var(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}|\mathbf{Z}^\perp, \Pi, G)\right] \tag{A20}$$

$$= (a'\Sigma a)\mathbb{E}\left[\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi/N^2\right] \tag{A21}$$

$$= (a'\Sigma a)\Gamma^{-1'}\mathbb{E}\left[\Lambda_L/N^2\right]\Gamma^{-1} = O(1/L). \tag{A22}$$

where

$$\Gamma = \begin{pmatrix} 1 & 0 \\ -\beta & 1 \end{pmatrix}. \tag{A23}$$

This comes from the fact that the entries of $\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi$ can be written as

$$(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi)_{(11)} = \lambda\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G(G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G)^{-1})^{-1}G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp\lambda + 2\beta_0\pi_1 G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp\lambda + \pi_1 G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G\pi_1\beta_0^2$$
$$\tag{A24}$$

$$(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi)_{(12)} = \lambda\mathbf{Z}^{\perp'}\mathbf{Z}G\pi_1 + \pi_1 G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G\pi_1\beta_0 \tag{A25}$$

$$(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi)_{(21)} = \lambda\mathbf{Z}^{\perp'}\mathbf{Z}G\pi_1 + \pi_1 G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G\pi_1\beta_0 \tag{A26}$$

$$(\Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi)_{(22)} = \pi_1 G'\mathbf{Z}^{\perp'}\mathbf{Z}^\perp G\pi_1\beta_0^2. \tag{A27}$$

Then,

$$\overline{\mathbf{Y}}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\overline{\mathbf{Y}}^\perp/L = \Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi/L + \Pi'\mathbf{Z}^{\perp'}\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L + \mathbf{V}'\mathbf{P}_{\mathbf{B}^\perp}\mathbf{Z}^\perp\Pi/L + \mathbf{V}'\mathbf{P}_{\mathbf{B}^\perp}\mathbf{V}/L \tag{A28}$$
$$\to^p \Psi. \tag{A29}$$

Finally, it follows that

$$\hat{\beta}_{bartik} \to^p \frac{\Psi_{12}}{\Psi_{22}} = \beta_0. \tag{A30}$$

$\square$

E   Normalization

This appendix presents results to understand the role of normalizations. Following Remark 1.1 we always "drop" industry k by subtracting off g_k from all the growth rates. Proposition E.1 shows that the bias coming each instrument can be written as a weighted average of the bias coming from the remaining $K-1$ instruments. Corollary E.1 shows how the Rotemberg weight gets shifted across instruments depending on which instrument is dropped. Finally, corollary E.2 shows that the average of the K normalizations is to set the unweighted mean of the growth rates to zero.

PROPOSITION E.1. *If the $\sum_{k=1}^{K} z_{lk} = 1 \forall l$, then we can write*

$$\mathbb{E}[\tilde{\beta}_k] = \sum_{j \neq k} \omega_{j,k} \mathbb{E}[\tilde{\beta}_j]$$

where $\omega_{j,k} = \dfrac{\Sigma_{ZX_j^{\perp}}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}}$ and $\mathbb{E}[\tilde{\beta}_j] = \dfrac{\Sigma_{ZV_j}}{\Sigma_Z X_j^{\perp}}$.

Proof. Recall that

$$\mathbb{E}[\tilde{\beta}_k] = \frac{\Sigma_{ZV_k}}{\Sigma_Z X_k^{\perp}}.$$

When $\sum_{k=1}^{K} z_{lk} = 1$, then $\sum_{k=1}^{K} \Sigma_{ZX_k^{\perp}} = 0$ and $\sum_{k=1}^{K} \Sigma_{ZV_k} = 0$. Then we can write

$$\Sigma_{ZV_k} = -\sum_{j \neq k} \Sigma_{ZV_j}$$

and

$$\Sigma_{ZX_k^{\perp}} = -\sum_{j \neq k} \Sigma_{ZX_j^{\perp}}.$$

Then:

$$\mathbb{E}[\tilde{\beta}_k] = \frac{\Sigma_{ZV_k}}{\Sigma_{ZX_k^{\perp}}} \tag{A1}$$

$$= \sum_{j \neq k} \frac{\Sigma_{ZV_j}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}} \tag{A2}$$

$$= \sum_{j \neq k} \frac{\Sigma_{ZX_j^{\perp}}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}} \frac{\Sigma_{ZV_j}}{\Sigma_{ZX_j^{\perp}}} \tag{A3}$$

$$= \sum_{j \neq k} \omega_{j,k} \mathbb{E}[\tilde{\beta}_j], \tag{A4}$$

where $\omega_{j,k} = \dfrac{\Sigma_{ZX_j^{\perp}}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}}$ and $\mathbb{E}[\tilde{\beta}_j] = \dfrac{\Sigma_{ZV_j}}{\Sigma_Z X_j^{\perp}}$. $\square$

COROLLARY E.1. *Let $\sum_{k=1}^{K} z_{lk} = 1 \forall l$. Let $\{\alpha_k(GG')\}_{k=1}^{K}$ be the set of sensitivity-to-misspecification*

elasticities given a weight matrix formed by a set of growth rates G. Now renormalize the growth rates by subtracting off g_k. Define $\alpha_{j,k}(GG') = \alpha_j((G - g_k)(G - g_k)')$ to be the resulting sensitivity to misspecification elasticities (which imply that we have "zeroed out" the kth instrument). Then:

$$\alpha_{j,k}(GG') = \alpha_j(GG') + \omega_{j,k}\alpha_k(GG'),$$

where $\omega_{j,k} = \dfrac{\Sigma_{ZX_j^\perp}}{\sum_{j' \neq k}\Sigma_{ZX_{j'}^\perp}}.$

Proof. Write:

$$\alpha_{j,k}(GG') = \frac{(g_j - g_k)\Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} \tag{A5}$$

$$= \frac{g_j\Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} - \frac{g_k\Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} \tag{A6}$$

$$= \frac{g_j\Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'}\Sigma_{ZX_{j'}^\perp}} - \frac{g_k\Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'}\Sigma_{ZX_{j'}^\perp}}, \tag{A7}$$

because $g_k\sum_{j'}\Sigma_{ZX_{j'}^\perp} = 0$. Then:

$$\alpha_{j,k}(GG') = \alpha_j(GG') - \frac{g_k\Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'}\Sigma_{ZX_{j'}^\perp}}\frac{\Sigma_{ZX_k^\perp}}{\Sigma_{ZX_k^\perp}} \tag{A8}$$

$$= \alpha_j(GG') - \alpha_k(GG')\frac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}}. \tag{A9}$$

Recall that $\Sigma_{ZX_k^\perp} = -\sum_{j \neq k}\Sigma_{ZX_j^\perp}$. So that: $-\dfrac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}} = \dfrac{\Sigma_{ZX_j^\perp}}{\sum_{j \neq k}\Sigma_{ZX_j^\perp}} = \omega_{j,k}$. Hence:

$$\alpha_{j,k}(GG') = \alpha_j(GG') + \omega_{j,k}\alpha_k(GG').$$

$\square$

COROLLARY E.2. *The average of the K normalizations is:*

$$\alpha_j(GG')^{avg} = \alpha_j(GG') - \frac{\Sigma_{ZX_j^\perp}}{K}\left[\frac{\sum_{k=1}^K g_k}{\sum_{k=1}^K g_k\Sigma_{ZX_k^\perp}}\right].$$

If $\sum_{k=1}^K g_k = 0$, then $\alpha_j(GG')^{avg} = \alpha_j(GG')$.

Proof. Note that we have two expressions for $\omega_{j,k} = -\dfrac{\Sigma_{ZX_j^{\perp}}}{\Sigma_{ZX_k^{\perp}}} = \dfrac{\Sigma_{ZX_j^{\perp}}}{\Sigma_{j \neq k} \Sigma_{ZX_j^{\perp}}}$

$$\alpha_j(GG')^{avg} = \frac{1}{K} \sum_{k=1}^{K} \alpha_{j,k}(GG') \tag{A10}$$

$$= \frac{1}{K} \sum_{k=1}^{K} \left[\alpha_j(GG') + \omega_{j,k} \alpha_k(G'G) \right] \tag{A11}$$

$$= \frac{1}{K} \sum_{k=1}^{K} \left[\alpha_j(GG') - \frac{\Sigma_{ZX_j^{\perp}}}{\Sigma_{ZX_k^{\perp}}} \alpha_k(G'G) \right] \tag{A12}$$

$$= \alpha_j(GG') - \frac{1}{K} \sum_{k=1}^{K} \left[\frac{\Sigma_{ZX_j^{\perp}}}{\Sigma_{ZX_k^{\perp}}} \frac{g_k \Sigma_{ZX_k^{\perp}}}{\sum_{j'=1}^{K} g_{j'} \Sigma_{ZX_{j'}^{\perp}}} \right] \tag{A13}$$

$$= \alpha_j(GG') - \frac{\Sigma_{ZX_j^{\perp}}}{K} \left[\frac{\sum_{k=1}^{K} g_k}{\sum_{k=1}^{K} g_k \Sigma_{ZX_k^{\perp}}} \right] . \tag{A14}$$

$\square$

F   Appendix: Simulation details

We report simulations with 228 industries, where this reflects the number of non-missing IND1990 3 digit industries in 1980. Similarly, we use 800 locations to correspond (loosely) to PUMAs. We assume (possibly unlike in the data) that the locations are independent, and simulate from the empirical distribution of industry shares. Finally, we anchor the properties of the industry-location growth rates to the U.S. data.

The top panel of Table A1 reports the empirical variances of industry employment growth with PUMAs and 3 digit industries, while the bottom panel reports the overall variances of industry-location growth rates. In all simulations, we begin by randomly drawing industry shares for each location (with replacement) from the empirical distribution of industry shares, as well as g_k, $\tilde{g}_{lk}$ and g_l terms, which we assume to be normally distributed. We use these terms to construct an x_l, and then draw a random error term ϵ_l for each location. To create endogeneity in OLS, we add the g_l term to ϵ_l.

In all simulations, the true value of β is assumed to be 2. We also consider a many invalid instrument simulation where we add an additional component to the error term $\tilde{\epsilon}_l = \sum_k z_{lk} \lambda_k$, where λ_k is drawn independently of g_k, and vary the size of the variance of λ_k. Finally, to show how the Rotemberg weights matter, in two simulations, we set $\lambda_k = g_k$ for five industries; in the first case, the five industries with the small Rotemberg weights (in absolute value), and in the second, the five industries with the largest Rotemberg weights.

G   When the first-stage coefficient is one

PROPOSITION G.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 $\tilde{g}_{lk}$. 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 $Cov(G, G_l - G) = 0$. $\qquad\square$

H 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:[33]

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

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.[34] Then the infeasible Bartik instrument that isolates the industry component of the innovations to demand shocks is $B_{lt} = \sum_k z_{lk0}d\alpha_{kt}$.

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.

H.1 The model's empirical analogue

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

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

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

[33]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 $d\alpha_{kt}$.

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

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.

I   Using growth rates to test overidentification restrictions

We consider a setting where only one instrument has first stage power. We consider a researcher choosing two sets of weights. We show that given one set of weights, denoted by G_1, and all but one entry 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. In this case, however, both Bartik instruments use the same identifying variation and so finding that they are uncorrelated does not imply that they leverage different sources of variation.

PROPOSITION I.1. *Suppose that $Z'Z$ is full rank. Suppose that only the first entry in $Z'X$ (a $K \times 1$ vector) is non-zero. Since we assume that the Z constitute a valid instrument, then only the first entry in $Z'Y$ is non-zero. Suppose that we are given two sets of weights, G_1 and G_2, with $G_{1,1} \neq 0$ and $G_{2,1} \neq 0$. Suppose we leave the last entry of the second vector unknown ($G_{2,K}$). 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. Then it is always possible to find $G_{2,K}$ such that:*

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

2. *The two Bartik instruments are uncorrelated.*

The proof shows that the first constraint is always satisfied, and derives an expression for the second constraint.

Proof. The first constraint is that:

$$\hat{\beta}_1 = \hat{\beta}_2 \tag{A1}$$

where for $j \in \{1, 2\}$ $\hat{\beta}_j = G'_j Z'Y (G'_j Z'X)^{-1}$. Since only the first entries in $Z'X$ and $Z'Y$ are nonzero, we have:

$$G'_j Z'Y (G'_j Z'X)^{-1} = \frac{\sum_k G_{j,k} Z'_k Y}{\sum_k G_{j,k} Z'_k X} \tag{A2}$$

$$= \frac{G_{j,1} Z'_1 Y + \sum_{k=2}^{K} G_{j,k} Z'_k Y}{G_{j,1} Z'_1 X + \sum_{k=2}^{K} G_{j,k} Z'_k X} \tag{A3}$$

$$= \frac{G_{j,1} Z'_1 Y + \sum_{k=2}^{K} G_{j,k} 0}{G_{j,1} Z'_1 X + \sum_{k=2}^{K} G_{j,k} 0} \tag{A4}$$

$$= \frac{Z'_1 Y}{Z'_1 X}, \tag{A5}$$

where this derivation uses the fact that only the first entry in $Z'X$ (and $Z'Y$) is nonzero. Hence, if $G_{1,1} \neq 0$ and $G_{2,1} \neq 0$, $\hat{\beta}_1 = \hat{\beta}_2$, which is true by assumption. Hence, the first constraint always holds.

The second constraint is that the covariance between the two Bartik instruments is zero:

$$Cov(B_1, B_2) = \mathbb{E}[B_1 B_2] - \mathbb{E}[B_1] \mathbb{E}[B_2] \tag{A6}$$

$$= \mathbb{E}[(ZG_1)(ZG_2)] - \mathbb{E}[ZG_1] \mathbb{E}[ZG_2] \tag{A7}$$

$$= \mathbb{E}[(ZG_1)'(ZG_2)] - \mathbb{E}[ZG_1] \mathbb{E}[ZG_2] \tag{A8}$$

$$= G'_1 \mathbb{E}[Z'Z] G_2 - G'_1 \mathbb{E}[Z'] \mathbb{E}[Z] G_2 \tag{A9}$$

$$= G'_1 [\mathbb{E}[Z'Z] - \mathbb{E}[Z'] \mathbb{E}[Z]] G_2 \tag{A10}$$

$$= G'_1 Var(Z) G_2, \tag{A11}$$

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

$$TG_2 = 0. \tag{A12}$$

Note that T is $1 \times K$. By assumption, the last entry in T are nonzero. We now construct an expression for this entry. To make $TG_2 = 0$, we need $\sum_{k=1}^{K} T_k G_{2,k} = 0 \Rightarrow G_{2,K} = -\frac{\sum_{k=1}^{K-1} T_k G_{2,k}}{T_K}$. $\qquad \square$

J The Rotemberg weights with leave-one-out

The formulas we present in Section 3 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 3.1 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:

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

$$B_{l,-l} = \frac{L}{L-1}Z_lG - \frac{1}{L-1}Z_lG_l \tag{A2}$$

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

where the observation is that $Z_lG_l = X_l$. 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:

$$\beta_k = \frac{Z'_kY}{Z'_kX}. \tag{A7}$$

Then one can show:

$$\alpha_k = \frac{\frac{L}{L-1}g_kZ'_kX - \frac{1}{L-1}X'Y\beta_k^{-1}}{\sum_k \frac{L}{L-1}g_kZ'_kX - \frac{1}{L-1}X'X}. \tag{A8}$$
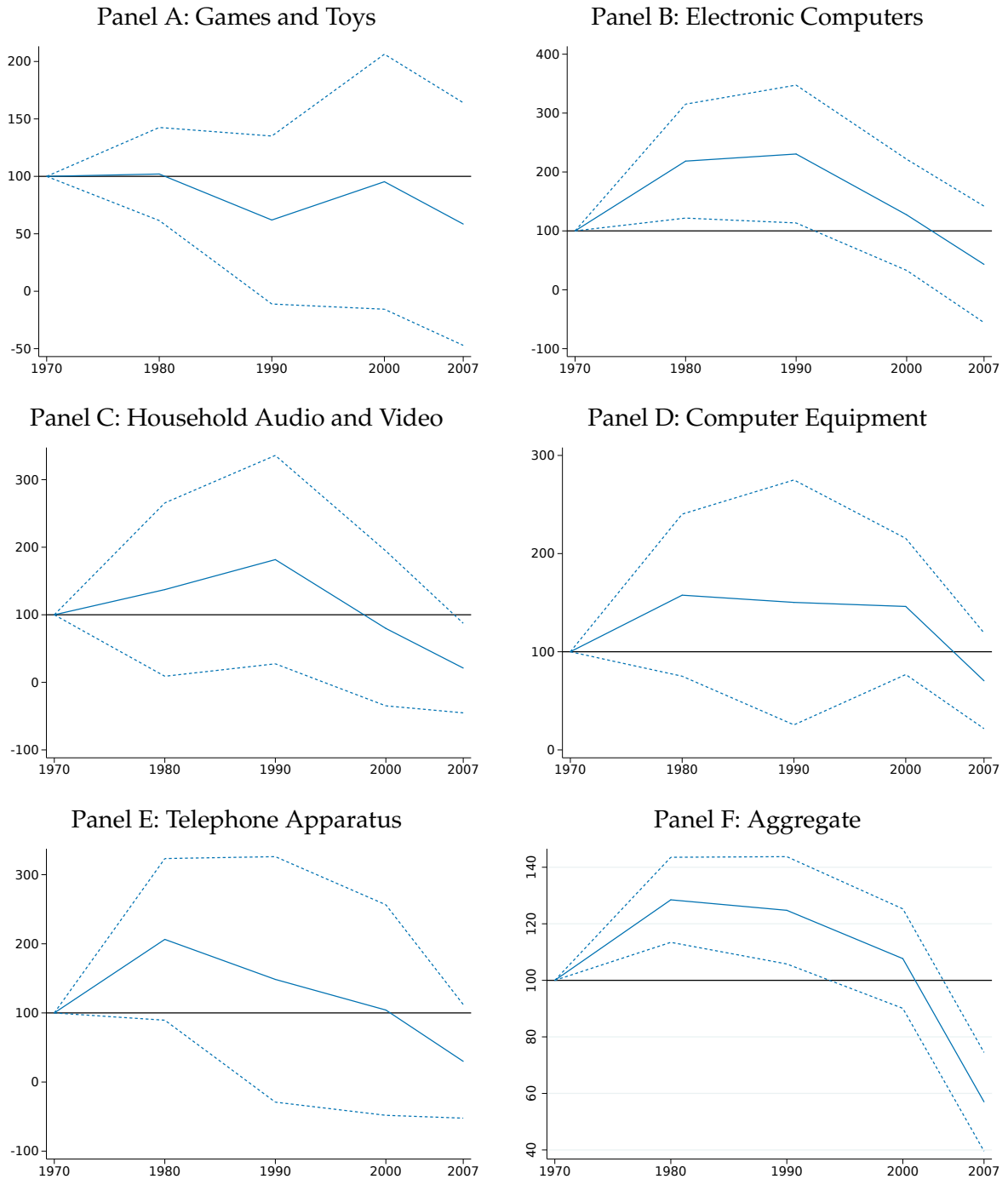
By inspection, $\sum_k \alpha_k \neq 1$. However, as $L \to \infty$ the sum converges to 1 as the leave-one-out terms drop out.

Table A1: Growth summary statistics

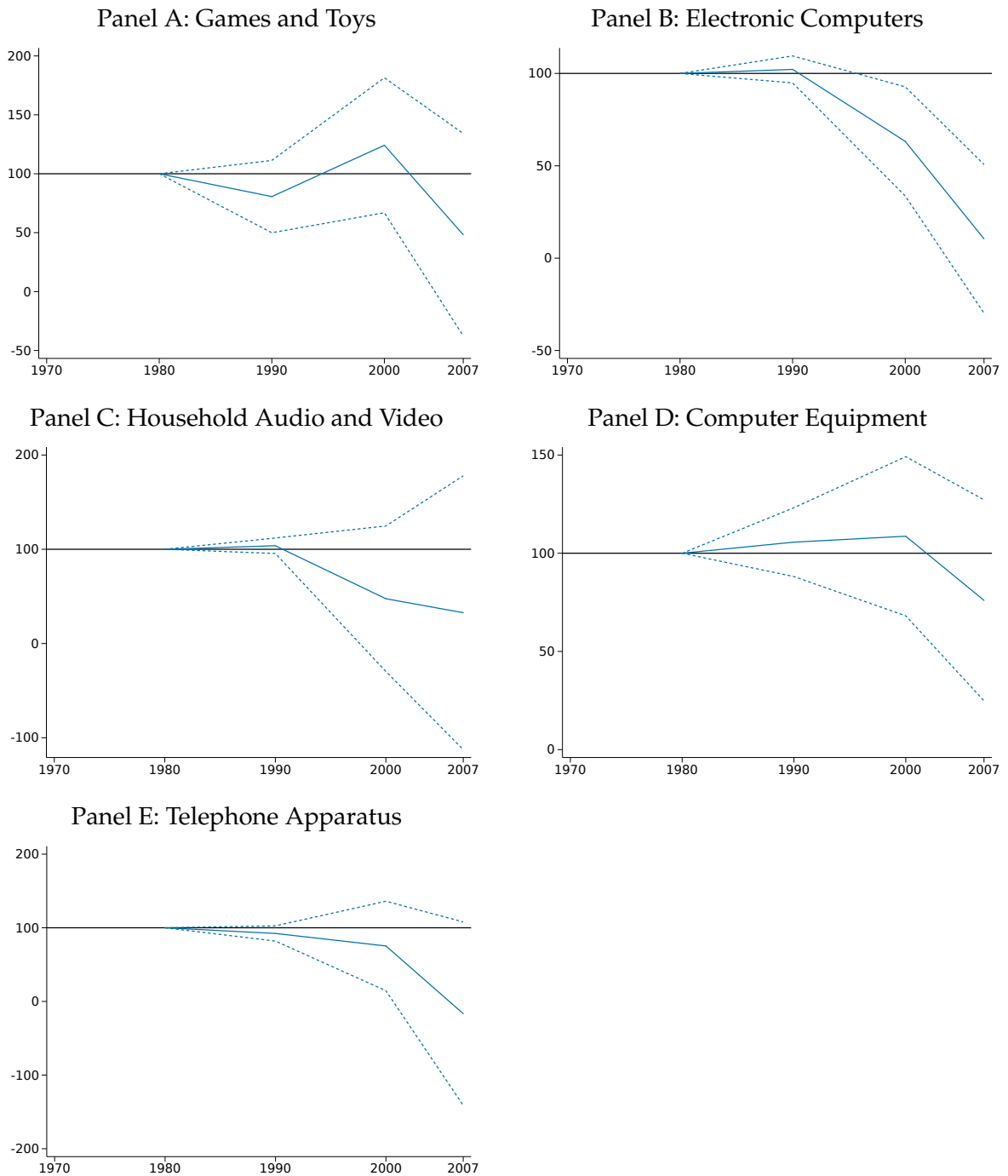|	Mean	Variance
Wage Growth		0.0006
Panel A: Industry:		
Puma 3 Digit Emp. Growth	0.0443	0.0044
Panel B: Pooled:		
Puma 3 Digit Emp. Growth	0.0472	0.0364

Notes: This table reports a variance decomposition of industry and industry-location growth rates. Panel A reports the means and variances of industry growth rates. Panel B reports the means and variance of industry-location growth rates.

Figure A1: Pre-trends for high Rotemberg weight industries (1990 shares): Autor, Dorn, and Hanson (2013)

Panel A: Games and Toys

Panel B: Electronic Computers

Panel C: Household Audio and Video

Panel D: Computer Equipment

Panel E: Telephone Apparatus

Panel F: Aggregate

Notes: These figures report pre-trends for the overall instrument and the top-5 Rotemberg weight industries as reported in Table 5. The Figures fix industry shares at the 1990 values and report the effect of these industry shares on manufacturing employment. We run regressions in growth rates and then convert to levels. We normalize 1970 to 100, and compute the standard errors using the delta method.

Figure A2: Pre-trends for high Rotemberg weight industries (time-varying shares): Autor, Dorn, and Hanson (2013)

Panel A: Games and Toys

Panel B: Electronic Computers

Panel C: Household Audio and Video

Panel D: Computer Equipment

Panel E: Telephone Apparatus

Notes: These figures report pre-trends for the overall instrument and the top-5 Rotemberg weight industries as reported in Table 5. The Figures update industry shares as in the benchmark ADH regressions and report the effect of these industry shares on manufacturing employment. We run regressions in growth rates and then convert to levels. Because 1970 shares are not available in the data, we normalize 1980 to 100, and compute the standard errors using the delta method.