Online Appendix for

Bartik Instruments: What, When, Why, and How

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift

June 2020

A Empirical example: 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*a*) (ADH).[1]

A.1 Specification

It is helpful to write the 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):

$$y_{lt} = \beta_0 + \beta x_{lt} + \epsilon_{lt}, \tag{13}$$

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:

$$x_{lt} = \gamma_0 + \gamma_1 B_{lt} + \eta_{lt}, \tag{14}$$

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 also produces a coefficient of 0.60 (see Table A3, TSLS (Bartik) row, column (2)) .

A.2 Form of endogeneity that the instrument addresses

At a high-level, it is clear that manufacturing employment can decline for many reasons, and the goal of ADH is to isolate the trade channel. Because the endogenous variable in the OLS equation and the instrument both have the Bartik form, if they were constructed using the same period to measure the shares, then in our setting there would be no way for OLS to be biased and TSLS to fix the endogeneity problem (because the choice of growth rates is not meaningful for identification in our setting).[2] Thus, the form of endogeneity that is addressed by the instrument stems from the timing of the measurement of the industry

[1] We use Census data from Ruggles et al. (2015), commuting zone data from Autor and Dorn (2013), and other replication data provided by Autor, Dorn and Hanson (2013*b*), Borusyak, Hull and Jaravel (2019*b*) and Adao, Kolesar and Morales (2019).

[2] In OLS, one can still compute Rotemberg weights. In the cross-section, $\alpha_k = \frac{g_k B_l z_{lk}}{\sum_{k'} g_k' B_l z_{lk'}}$.

shares.[3] Here, the form of endogeneity is that employment responds in anticipation of trade shocks.[4]

A.3 Rotemberg weights

As in the canonical setting, despite a very large number of instruments (397 industries) the distribution of sensitivity is skewed so that a small number of instruments get a large share of the weight. Table A1 shows that the top five instruments receive over half of the absolute weight in the estimator (0.532/1.067). These instruments are electronic computers, games and toys, household audio and video, telephone apparatus, and computer equipment. Except for games and toys, these industries are different than the ones that ADH emphasize when motivating the empirical strategy.[5] In particular, rather than being low-skill technologically stagnant industries where it is plausible that trade is the main shock hitting the industry, these are higher-skill technologically innovative industries where it is plausible that changes in technology are the main shock hitting the industry.

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, and consistent with the discussion in the previous paragraph that the growth rates provide an imperfect guide to which industries drive estimates, the g_k component explains less than twenty percent (0.430^2, see Table A1, Panel B) of the variance of the Rotemberg weights.

A.4 Discussion of the identifying assumption in terms of the shares

Why is it reasonable to interpret this paper as being about the shares? We note first that the paper does not emphasize having a large number of independent shocks (which would be necessary for the shocks interpretation to be plausible). Indeed, it is hard to conceive of a model of an "optimizing China" that would generate random patterns of exports across a wide swathe of the economy. (The random shocks assumption is more plausible in this setting if researchers control for "higher-level" fixed effects and so exploit more idiosyn-

[3] In the Borusyak, Hull and Jaravel (2019*a*) setting, it is possible to motivate why the choice of growth rates would fix an endogeneity concern, but it is harder to motivate why one would want to lag the industry shares.

[4] Autor, Dorn and Hanson (2013*a*, pg. 2129) write, "We use ten-year-lagged employment levels because, to the degree that contemporaneous employment by region is affected by anticipated China trade, the use of lagged employment to apportion predicted Chinese imports to regions will mitigate this simultaneity bias."

[5] "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 & 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).

cratic variation. When Borusyak, Hull and Jaravel (2019*a*, Table 1, column 6) control for 2 digit industries, the estimates are one-sixth the size and no longer statistically significant.) Second, the paper emphasizes particular industries and industries with particular characteristics. That is, our reading of the logic of the paper is that it emphasizes that Chinese exports were concentrated in low-skill, labor-intensive industries. This focus on particular industries is not consistent with the consistency of the estimator coming from the shocks.

Is the identification assumption necessarily implausible when viewed in terms of shares? We do not think so. Other papers in the trade literature leverage changes in trade policy and study local labor market effects of these policy changes (e.g., Topalova (2010) and Kovak (2013)). In these papers, the argument is not that the trade policy is literally random. Instead, the argument is that the change in trade policy does not coincide with shocks to locations that were highly exposed to changes in trade policy. An example of this argument in the case of ADH would be that technological changes (shifting the labor force towards automation) did not simultaneously occur to industries that were more exposed to the China trade policy. The argument does not require that the shares predict nothing in levels, but simply that the shares only predict changes through the causal channel emphasized by the paper.

In the trade policy example, there is some institutional reason to expect that there is a shock in particular industries that only operated through trade policy (because trade policy changed in these industries). By analogy, in the context of ADH this logic would suggest using institutional knowledge to pick industries where there was a large increase in exports from China because of Chinese comparative advantage (rather than technological change in the industry). Seen in this light, one motivation for ADH to look at imports to other high-income countries might be to isolate the industries where there is strong reason to think that China experienced rapid productivity gains. As we have emphasized, however, the weights that the Bartik estimator places on the just-identified IV estimates are not solely a function of the growth rates. Indeed, in this example, the growth rates explain less than twenty percent of the variation in the Rotemberg weights. As a result, weighting the shares by growth rates is an imperfect way of isolating the variation that the researcher intends. If there was further pruning of the industries, then a research design based on the shares would likely accord more closely with the goals of researchers.

A.5 Testing the plausibility of the identifying assumption

Test 1: Correlates of 1980 industry shares Table A2 shows the relationship between the covariates used in ADH and the top industries reported in Table A1. First, relative to the canonical setting, the controls explain less of the variation in shares (lower R^2 in the regressions). Second, electronic computers, computer equipment manufacturing as well as the overall measure are 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 and provide sensitivity analyses related to this industry.[6]

Test 2: Parallel pre-trends We construct our pre-trend figures as follows. We use fixed 1980 shares as our main variable, and plot the reduced form effect of each industry on manufacturing employment.[7] For our controls, we fix the controls in the same time period, and interact with time fixed effects. As in the main specification, we also control for region and time fixed effects as well. 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 combine them using growth rates from 1990 to 2000.

Figure A1 shows the plots and displays several interesting patterns. First, games and toys (Panel B), household audio and video (Panel C) and telephone apparatus (Panel D) diverge from classic pre-trends figures, which would show no trends in the pre-periods and then a sharp change at the date of the treatment. Second, the patterns in electronic computers (Panel A) and computer equipment (Panel E) are more promising in that there is a sharp drop in 2007 and no effect in 1990 and 2000; less promising, however, is that there are statistically significant effects in 1980. Note that these panels show 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 A3 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 A3 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, overidentified 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

[6] 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."

[7] We use the reduced-form effect because the endogenous variable is not available in the earlier periods. See Figure A4 for the analogous figures using fixed 1990 shares.

estimators is not reassuring,[8] and the failure of the overidentification tests points to potential misspecification.

Visualizing the overidentification tests If one wishes to interpret the failure of the overidentification tests as pointing to heterogeneity of the form outlined in Section IV rather than as evidence of misspecification, then Figure A2 shows some of the heterogeneity in treatment effects underlying the overall estimate (Figure A3 shows the relationship between the Rotemberg weights and the first-stage F-statistic). Relative to the canonical case, the patterns of heterogeneity are less concerning. In particular, visually there is less dispersion in the point estimates among the high-powered industries and the high-weight industries are clustered more closely to the overall point estimate. Finally, while there are negative Rotemberg weights, these industries are a small share of the overall weight, suggesting that there are unlikely to be negative weights on particular location-specific parameters (i.e., β_l; see also Panel E in Table A1).

[8] 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..."

Table A1: Summary of Rotemberg weights: China shock

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.067	-0.000	0.059		
Positive	1.067	0.004	0.941		
Panel B: Correlations					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	$\hat{F}_k$	Var(z_k)
$\hat{\alpha}_k$	1				
g_k	0.430	1			
$\hat{\beta}_k$	0.003	-0.320	1		
$\hat{F}_k$	0.192	0.027	0.017	1	
Var(z_k)	0.102	-0.141	0.157	0.229	1
Panel C: Variation across years in $\hat{\alpha}_k$					
	Sum	Mean			
1990	0.017	0.000			
2000	0.983	0.002			
Panel D: Top 5 Rotemberg weight industries					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Ind Share
Electronic Computers	0.183	186.231	-0.619	(-1.50,-0.20)	0.137
Games, Toys, and Children's Vehicles	0.138	243.794	-0.126	(-0.60,0.30)	0.044
Household Audio and Video Equipment	0.085	187.718	0.174	(-0.20,1.80)	0.046
Telephone and Telegraph Apparatus	0.066	92.922	-0.315	(-∞, ∞)	0.100
Computer Peripheral Equipment, NEC	0.060	34.982	-0.303	(-1.20,-0.20)	0.100
Panel E: Estimates of β_k for positive and negative weights					
	α-weighted Sum	Share of overall β	Mean		
Negative	-0.014	0.024	-0.036		
Positive	-0.582	0.976	-1.170		

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 III.C. Panel A reports the share and sum of negative Rotemberg weights. Panel B reports correlations between the weights ($\hat{\alpha}_k$), the national component of growth (g_k), the just-identified coefficient estimates ($\hat{\beta}_k$), the first-stage F-statistic of the industry share ($\hat{F}_k$), and the variation in the industry shares across locations (Var(z_k)). Panel C reports variation in the weights across years. Panel D reports the top five industries according to the Rotemberg weights. The g_k is the national industry growth rate, $\hat{\beta}_k$ is the coefficient from the just-identified regression, the 95% confidence interval is the weak instrument robust confidence interval using the method from Chernozhukhov and Hansen (2008) over a range from -10 to 10 (($-\infty$,∞) indicates that it was not possible to successfully define the confidence interval), and Ind Share is the industry share (multiplied by 100 for legibility). Panel E reports statistics about how the values of $\hat{\beta}_k$ vary with the positive and negative Rotemberg weights.

Table A2: Relationship between industry shares and characteristics: China shock

	Electronic Computers	Games, Toys, and Children's Vehicles	Household Audio and Video Equipment	Telephone and Telegraph Apparatus	Computer Peripheral Equipment, NEC	China to other
Share Empl in Manufacturing	0.016	0.002	0.006	0.002	0.009	0.099
	(0.008)	(0.001)	(0.003)	(0.003)	(0.004)	(0.011)
Share College Educated	0.016	-0.001	0.002	0.001	0.012	0.068
	(0.006)	(0.001)	(0.002)	(0.003)	(0.003)	(0.014)
Share Foreign Born	0.004	0.001	-0.001	-0.005	0.002	0.052
	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.009)
Share Empl of Women	-0.002	0.003	-0.006	-0.003	0.000	0.031
	(0.006)	(0.002)	(0.003)	(0.006)	(0.004)	(0.017)
Share Empl in Routine	-0.083	0.006	0.010	-0.010	-0.046	-0.051
	(0.041)	(0.003)	(0.007)	(0.015)	(0.018)	(0.084)
Avg Offshorability	0.410	-0.027	-0.022	0.248	0.182	-1.173
	(0.214)	(0.022)	(0.039)	(0.076)	(0.091)	(0.460)
R^2	0.18	0.02	0.01	0.04	0.12	0.22
N	1444	1444	1444	1444	1444	1444

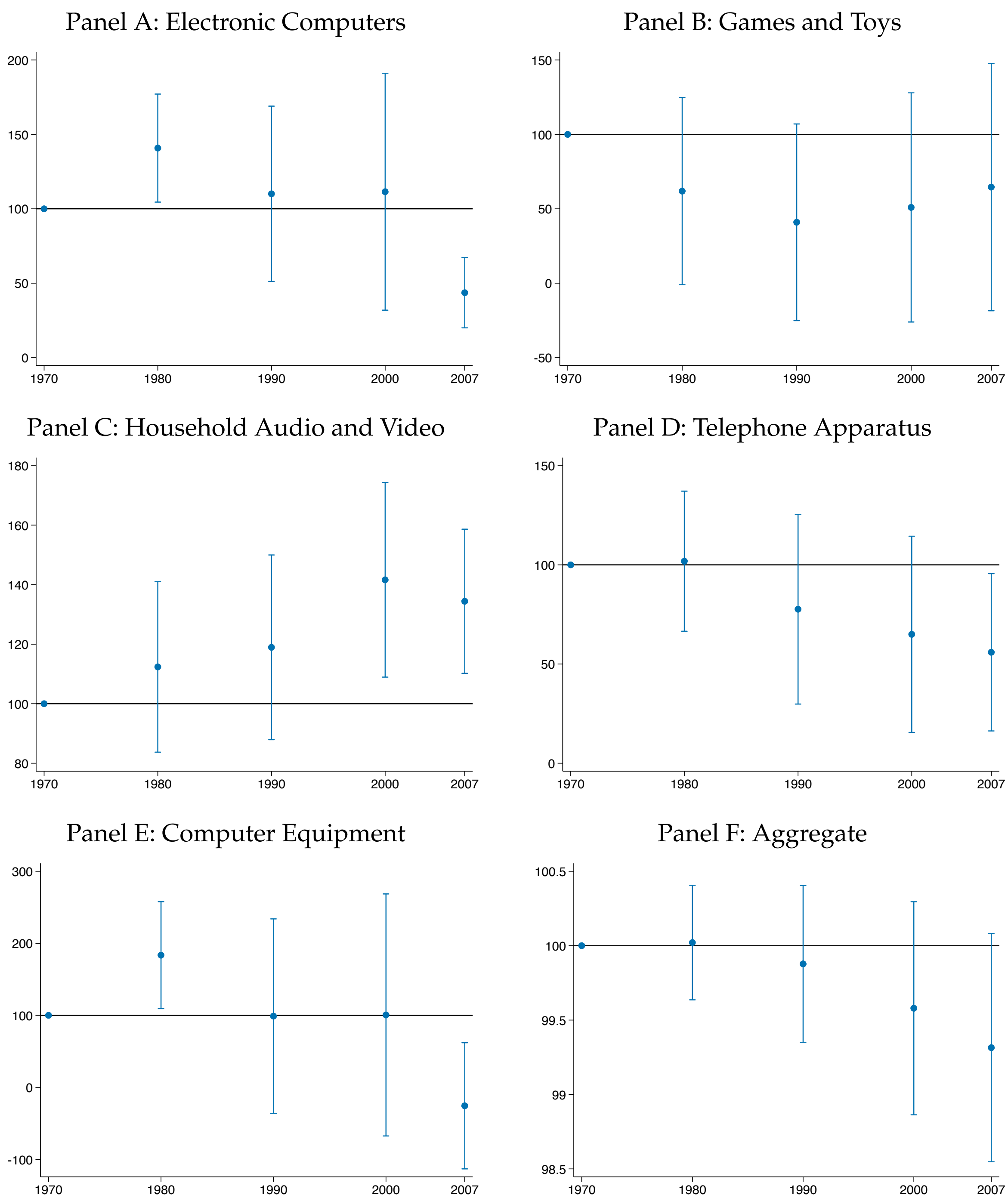
Notes: Each column reports a separate regression. The regressions are two pooled cross-sections, where one cross section is 1980 shares on 1990 characteristics, and one is 1990 shares on 2000 characteristics. The final column is constructed using 1990 to 2000 growth rates. Results are weighted by the population in the period the characteristics are measured. Standard errors in parentheses.

Table A3: OLS and IV estimates: China shock

	Δ Emp (1)	(2)	Coefficients Equal (3)	Over ID Test (4)
OLS	-0.38	-0.17	[0.00]	
	(0.07)	(0.04)		
2SLS (Bartik)	-0.73	-0.60	[0.05]	
	[0.07]	[0.10]		
2SLS	-0.45	-0.21	[0.00]	917.35
	[0.07]	[0.05]		[0.00]
MBTSLS	-0.56	-0.29	[0.00]	
	[0.07]	[0.05]		
LIML	-1.47	-1.94	[0.83]	1868.95
	(0.71)	(3.33)		[0.00]
HFUL	-1.15	-1.13	[0.47]	968.37
	(0.04)	(0.04)		[0.00]
Year and Census Division FE	Yes	Yes		
Controls	No	Yes		
Observations	1,444	1,444		

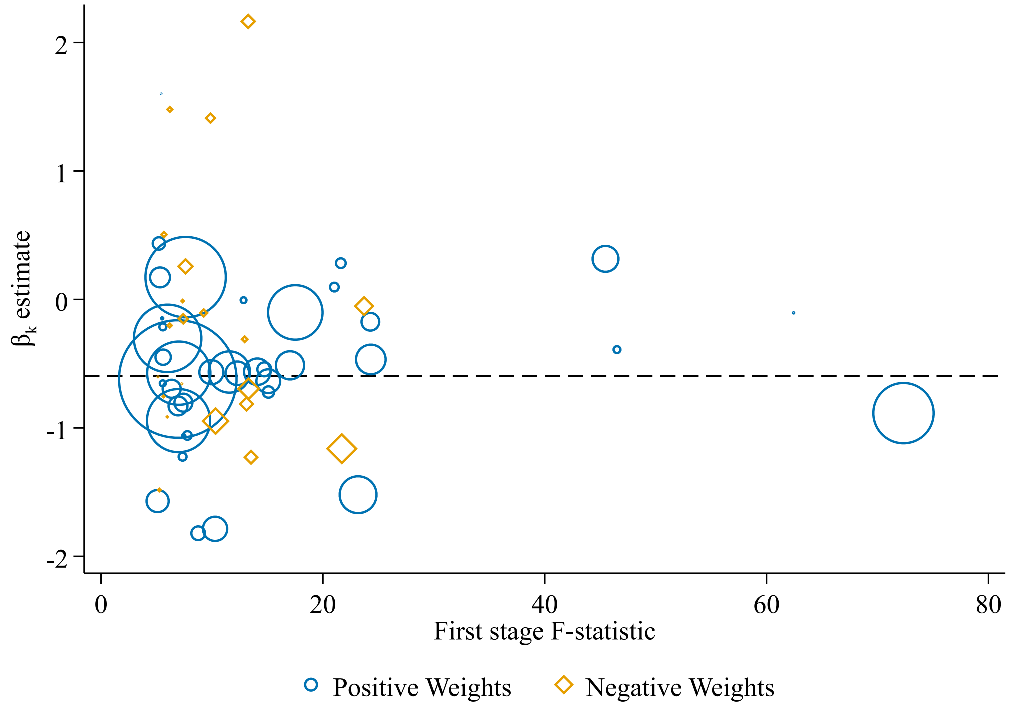
Notes: This table reports a variety of estimates of the effect of rising imports from China on US manufacturing employment. The regressions are at the CZ level and include two time periods (1990 to 2000, and 2000 to 2007). The TSLS row is our replication of Column (1) and Column (6) of Table 3 in ADH. 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) with the same set of instruments. The LIML row shows estimates using the limited information maximum likelihood estimator with the same set of instruments. Finally, the HFUL row uses the HFUL estimator of Hausman et al. (2012) with the same set of instruments. The J-statistic for HFUL comes from Chao et al. (2014). The p-value for the equality of coefficients compares the adjacent columns with and without controls. The controls are the contemporaneous characteristics displayed in Table A2. Results are weighted by start of period population. Standard errors are in parentheses and are constructed by bootstrap over commuting zones. p-values are in brackets.

Figure A1: Pre-trends for high Rotemberg weight industries: China shock



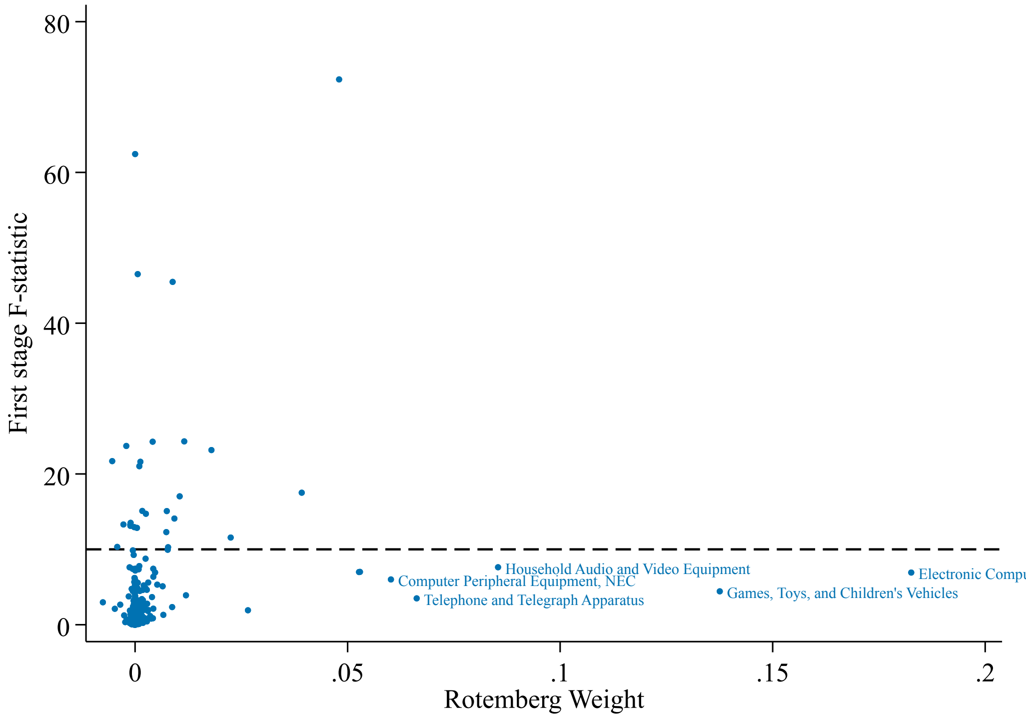
Notes: These figures report pre-trends for the overall instrument and the top-5 Rotemberg weight industries as reported in Table A1. The figures fix industry shares at the 1980 values and report the effect of these industry shares on manufacturing employment. For our controls, we fix the controls in the same time period, and interact with time fixed effects. As in the main specification, we also control for region and time fixed effects as well. 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. For the aggregate panel, we use the Bartik estimate for 1980.

Figure A2: Heterogeneity of β_k: China shock



Notes: This figure plots the relationship between each instruments' $\hat{\beta}_k$, first stage F-statistics and the Rotemberg weights. Each point is a separate instrument's estimates (industry share). The figure plots the estimated $\hat{\beta}_k$ for each instrument on the y-axis and the estimated first-stage F-statistic on the x-axis. The size of the points are scaled by the magnitude of the Rotemberg weights, with the circles denoting positive Rotemberg weights and the diamonds denoting negative weights. The horizontal dashed line is plotted at the value of the overall $\hat{\beta}$ reported in the second column in the TSLS (Bartik) row in Table A3. The figure excludes instruments with first-stage F-statistics below 5.

Figure A3: First stage versus Rotemberg weights: China shock



Notes: This figure plots each instrument's Rotemberg weight against the first stage F-statistic. Each point represents the estimates for an instrument, where instruments are aggregated across time periods following Section III.C. The labelled industries correspond to the five highest Rotemberg weight industries from Table A1. The dashed horizontal line is equal to 10.

B Instruments encompassed by our structure

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

B.1 Bank lending relationships

Greenstone, Mas and Nguyen (2020) 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 (2020) 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:

$$g_{lk} = g_l + g_k + \epsilon_{lk}, \tag{15}$$

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 (15).

B.2 Market size and demography

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

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

C Omitted proofs

Proposition 1

Proof.

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

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

Proposition 3

We use slightly more general notation than in the body of the paper. Let $\hat{W}$ be an arbitrary weight matrix and 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 k^{th} row of $\hat{W}$. We index a solution for $\hat{\beta}$ by $\hat{W}$: $\hat{\beta}(\hat{W})$. The more general version of the proposition stated in the text is:

PROPOSITION 5. *Let*

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

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. The proof is just algebra:

$$\hat{\alpha}_k(\hat{W})\hat{\beta}_k = \frac{\hat{c}_k(\hat{W})Z_k'X^\perp}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^\perp}(Z_k'X^\perp)^{-1}Z_k'Y^\perp = \frac{\hat{c}_k(\hat{W})Z_k'Y^\perp}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^\perp} \tag{16}$$

$$\sum_{k=1}^K \hat{\alpha}_k(\hat{W})\hat{\beta}_k = \frac{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'Y^\perp}{\sum_{k=1}^K \hat{c}_k(\hat{W})Z_k'X^\perp} \tag{17}$$

$$= \frac{\hat{C}(\hat{W})'Z'Y^\perp}{\hat{C}(\hat{W})'Z'X^\perp}. \tag{18}$$

□

The proposition stated in the text comes from substituting in for the Bartik definition of $\hat{W}$.

Proposition 4

Proof. For a given k,

$$\hat{\beta}_k = \frac{\sum_l z_{lk}x_l^\perp \beta_l}{\sum_l z_{lk}x_l^\perp} + \frac{\sum_l z_{lk}\epsilon_l^\perp}{\sum_l z_{lk}x_l^\perp} \tag{19}$$

$$= \frac{\sum_l z_{lk}x_l^\perp \beta_l}{\sum_l z_{lk}x_l^\perp} + o_p(1) \tag{20}$$

$$= \frac{\sum_l z_{lk}^{\perp,2}\pi_{lk}\beta_l + z_{lk}u_{lk}^\perp}{\sum_l z_{lk}^{\perp,2}\pi_{lk} + z_{lk}u_{lk}^\perp} + o_p(1). \tag{21}$$

Thus,

$$\operatorname*{plim}_{L\to\infty} \hat{\beta}_k = \mathbb{E}[\omega_{lk}\beta_l], \tag{22}$$

where $\omega_{lk} = z_{lk}^{\perp,2}\pi_{lk}/\mathbb{E}[z_{lk}^{\perp,2}\pi_{lk}]$. Since $\pi_{lk} \geq 0$ by assumption, ω_{lk} is non-negative for all l. Additionally, $\mathbb{E}[\omega_{lk}] = 1$. □

D Equivalence with K industries, L locations, and controls

The two stage least squares system of equations is:

$$y_{lt} = D_{lt}\rho + x_{lt}\beta + \epsilon_{lt} \tag{23}$$

$$x_{lt} = D_{lt}\tau + B_{lt}\gamma + \eta_{lt}, \tag{24}$$

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 $\mathbf{P_D} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ as the $n \times n$ projection matrix of $\mathbf{D}$, and $\mathbf{M_D} = \mathbf{I}_n - \mathbf{P_D}$ as the annhilator matrix. Then, because this is an exactly identified instrumental variable our estimator is

$$\hat{\beta}_{Bartik} = \frac{B'\mathbf{M_D}Y}{B'\mathbf{M_D}X}. \tag{25}$$

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

$$y_{lt} = D_{lt}\rho + x_{lt}\beta + \epsilon_{lt} \tag{26}$$

$$x_{it} = D_{lt}\tau + Z_{lt}\gamma_t + \eta_{lt}, \tag{27}$$

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

$$Y = \mathbf{D}\rho + X\beta + \epsilon \tag{28}$$

$$X = \mathbf{D}\tau + \tilde{Z}\mathbf{\Gamma} + \eta, \tag{29}$$

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

$$\mathbf{\Gamma} = [\gamma_1 \cdots \gamma_T], \tag{30}$$

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

$$\tilde{\mathbf{Z}} = \left[\begin{array}{ccc} \mathbf{Z} \odot 1_{t=1} & \cdots & \mathbf{Z} \odot 1_{t=T} \end{array} \right], \tag{31}$$

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:

$$\hat{\beta}_{GMM} = \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}\Omega\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{32}$$

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

PROPOSITION 6. *If $\Omega = GG'$, then $\hat{\beta}_{GMM} = \hat{\beta}_{Bartik}$.*

Proof. Start with the Bartik estimator,

$$\hat{\beta}_{Bartik} = \frac{B'\mathbf{M_D}Y}{B'\mathbf{M_D}X} \tag{33}$$

$$= \frac{G'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{G'\tilde{\mathbf{Z}}'\mathbf{M_D}X} \tag{34}$$

$$= \frac{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}Y}{X'\mathbf{M_D}\tilde{\mathbf{Z}}GG'\tilde{\mathbf{Z}}'\mathbf{M_D}X}, \tag{35}$$

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 $\hat{\beta}_{GMM} = \hat{\beta}_{Bartik}$. □

E Interpreting the Rotemberg weights

To interpret the Rotemberg 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 Andrews, Gentzkow and Shapiro (2017) (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 4 (Identification and Regularity). *(i) the data $\{\{x_{lt}, Z_{lt}, D_{lt}, V_{lt}, \epsilon_{lt}\}_{t=1}^{T}\}_{l=1}^{L}$ are independent and identically distributed with K and T fixed, and L going to infinity;*

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

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

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

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

LEMMA 1. *If Assumption 4 holds and* $\text{plim}_{L\to\infty} \hat{W}_L = W$ *where* W *is a positive semi-definite matrix, then*

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

Proof. Note that

$$\hat{\alpha}_k(\hat{W}) = \frac{X^{\perp\prime} Z \hat{W}_k Z_k' X^{\perp}}{X^{\perp\prime} Z \hat{W} Z' X^{\perp}} \tag{36}$$

$$= \frac{\left(\sum_{l,t} x_{lt}^{\perp} Z_{lt}\right) \hat{W}_k \left(\sum_{l,t} z_{lkt} x_{lt}^{\perp}\right)}{\left(\sum_{l,t} x_{lt}^{\perp} Z_{lt}\right) \hat{W} \left(\sum_{l,t} Z_{lt} x_{lt}^{\perp}\right)}. \tag{37}$$

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 $L \to \infty$. □

We now present results about the asymptotic behavior of our estimators with misspecification.

PROPOSITION 7. *We assume that Assumption 4 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. 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 $\mathbb{E}[\tilde{\beta}] = \sum_k \alpha_k(W)\frac{\Sigma_{ZV,k}}{\Sigma_{ZX^\perp,k}} = \sum_k \alpha_k(W)\mathbb{E}[\tilde{\beta}_k]$. □

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

COROLLARY 1. *Suppose that $\beta_0 \neq 0$. Then the percentage bias can be written in terms of the Rotemberg weights:*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \alpha_k(W)\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \quad (38)$$

COROLLARY 2. *Under the Bartik weight matrix ($W = GG'$),*

$$\frac{\mathbb{E}[\tilde{\beta}]}{\beta_0} = \sum_k \frac{g_k \Sigma_{ZX^\perp,k}}{G'\Sigma'_{ZX^\perp}}\frac{\mathbb{E}[\tilde{\beta}_k]}{\beta_0}. \quad (39)$$

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.[9] 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.

[9] 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."

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 k^{th} 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 8. *The difference in the bias from the full estimator and the estimator that leaves out the k^{th} industry is:*

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

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

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

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

$$\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right] = \sum_{k'}\alpha_{k'}(W)\mathbb{E}[\tilde{\beta}_{k'}] - \sum_{k'\neq k}\alpha_{k'}(W_{-k})\mathbb{E}[\tilde{\beta}_{k'}] \tag{40}$$

$$= \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] + \sum_{k'\neq k}(\alpha_{k'}(W) - \alpha_{k'}(W_{-k}))\mathbb{E}[\tilde{\beta}_{k'}]. \tag{41}$$

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 $\alpha_{k'}(W_{-k}) = \alpha_{k'}(W)/(1-\alpha_k(W))$.[10] This gives:

$$\mathbb{E}\left[\tilde{\beta}(\hat{W}) - \tilde{\beta}(\hat{W}_{-k})\right] = \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] + \sum_{k'\neq k}\left(\alpha_{k'}(W) - \frac{\alpha_{k'}(W)}{1-\alpha_k(W)}\right)\mathbb{E}[\tilde{\beta}_{k'}] \tag{42}$$

$$= \alpha_k(W)\mathbb{E}[\tilde{\beta}_k] - \frac{\alpha_k(W)}{1-\alpha_k(W)}\sum_{k'\neq k}(\alpha_{k'}(W))\,\mathbb{E}[\tilde{\beta}_{k'}]. \tag{43}$$

□

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

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

F Rotemberg weights and first-stage F-statistics

In this appendix, we derive the relationship between the Rotemberg weight on the k^{th} instrument, and the relative first stage F-statistic.

Let $\hat{\pi}_k \equiv \frac{Z_k^{\perp\prime} X^\perp}{Z_k^{\perp\prime} Z_k^\perp}$ and $\hat{\pi} \equiv \frac{B^{\perp\prime} X^\perp}{B^{\perp\prime} B^\perp}$ be the first stage coefficients for the k^{th} industry and the Bartik instrument. The first stage F-statistic on the k^{th} instrument can be written[11]

$$
\begin{aligned}
F_k &= \frac{\hat{\pi}_k{}^2}{\hat{\Sigma}_{\pi_k \pi_k}} && (44)\\
&= \left(\frac{Z_k' X^\perp}{Z_k^{\perp\prime} Z_k^\perp} \right)^2 \frac{1}{\hat{\Sigma}_{\pi_k \pi_k}} && (45)\\
&= \frac{1}{g_k^2} \left(\underbrace{\frac{g_k Z_k' X^\perp}{\sum_{k'} g_{k'} Z_{k'} X^\perp}}_{\hat{\alpha}_k} \right)^2 \left(\frac{(\sum_{k'} g_{k'} Z_{k'} X^\perp)}{Z_k^{\perp\prime} Z_k^\perp} \right)^2 \frac{1}{\hat{\Sigma}_{\pi_k \pi_k}} && (46)\\
&= \frac{1}{g_k^2} \hat{\alpha}_k^2 \left(\frac{(B^{\perp\prime} X^\perp}{Z_k^{\perp\prime} Z_k^\perp} \right)^2 \frac{1}{\hat{\Sigma}_{\pi_k \pi_k}} && (47)\\
&= \frac{1}{g_k^2} \frac{(B^{\perp\prime} B^\perp)^2}{(Z_k^{\perp\prime} Z_k^\perp)^2} \underbrace{\frac{(B^{\perp\prime} X^\perp)^2}{(B^{\perp\prime} B^\perp)^2}}_{\hat{\pi}^2} \hat{\alpha}_k^2 \frac{1}{\hat{\Sigma}_{\pi_k \pi_k}} && (48)\\
&= \frac{1}{g_k^2} \frac{(B^{\perp\prime} B^\perp)^2}{(Z_k^{\perp\prime} Z_k^\perp)^2} F \hat{\alpha}_k^2 \frac{\hat{\Sigma}_{\pi\pi}}{\hat{\Sigma}_{\pi_k \pi_k}}. && (49)
\end{aligned}
$$

From the first to the second line we substitute in the definition of $\hat{\pi}_k^2$, from the second to the third line we multiply by $\frac{g_k^2}{g_k^2}$ and $\left(\frac{\sum_{k'} g_{k'} Z_{k'} X^\perp}{\sum_{k'} g_{k'} Z_{k'} X^\perp}\right)^2$, from the third to the fourth line we use the definition of $\hat{\alpha}_k$ and the fact that $\sum_{k'} g_{k'} Z_{k'} X^\perp = B^{\perp\prime}$, from to the fourth to the fifth line we multiply by $\left(\frac{B^{\perp\prime}}{B^{\perp\prime}}\right)^2$, and from the fifth to the sixth line we multiply by $\frac{\hat{\Sigma_{\pi\pi}}}{\hat{\Sigma_{\pi\pi}}}$ and use the definition of F.

Hence, we have:

$$
\begin{aligned}
\frac{F_k}{F} &= \frac{(B^{\perp\prime} B^\perp)^2}{(Z_k^{\perp\prime} Z_k^\perp)^2} \frac{\hat{\Sigma}_{\pi\pi}}{\hat{\Sigma}_{\pi_k \pi_k}} \frac{1}{g_k^2} \hat{\alpha}_k^2 && (50)\\
&= \hat{\alpha}_k^2 \left(\frac{\widehat{Var}(B^\perp)}{g_k \widehat{Var}(Z_k^\perp)} \right)^2 \frac{\hat{\Sigma}_{\pi\pi}}{\hat{\Sigma}_{\pi_k \pi_k}}. && (51)
\end{aligned}
$$

[11] See, e.g., https://www.nber.org/econometrics_minicourse_2018/2018si_methods.pdf at slide 21.

G Normalization of the Rotemberg weights

This appendix presents results to understand the role of normalizations. Following Remark 1 we always "drop" industry k by subtracting off g_k from all the growth rates. Proposition 9 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 3 shows how the Rotemberg weight gets shifted across instruments depending on which instrument is dropped. Finally, corollary 4 shows that the average of the K normalizations is to set the unweighted mean of the growth rates to zero.

PROPOSITION 9. *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 $\omega_{j,k} = \frac{\Sigma_{ZX_j^{\perp}}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}}$ *and* $\mathbb{E}[\tilde{\beta}_j] = \frac{\Sigma_{ZV_j}}{\Sigma_Z X_j^{\perp}}$.

Proof. Recall from Proposition 7 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:

$$\mathbb{E}[\tilde{\beta}_k] = \frac{\Sigma_{ZV_k}}{\Sigma_{ZX_k^{\perp}}} \tag{52}$$

$$= \sum_{j \neq k} \frac{\Sigma_{ZV_j}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}} \tag{53}$$

$$= \sum_{j \neq k} \frac{\Sigma_{ZX_j^{\perp}}}{\sum_{j' \neq k} \Sigma_{ZX_{j'}^{\perp}}} \frac{\Sigma_{ZV_j}}{\Sigma_{ZX_j^{\perp}}} \tag{54}$$

$$= \sum_{j \neq k} \omega_{j,k} \mathbb{E}[\tilde{\beta}_j], \tag{55}$$

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

COROLLARY 3. *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 k^{th} instrument). Then:*

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

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

Proof. Write:

$$\alpha_{j,k}(GG') = \frac{(g_j - g_k)\Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} \tag{56}$$

$$= \frac{g_j \Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} - \frac{g_k \Sigma_{ZX_j^\perp}}{\sum_{j'}(g_{j'} - g_k)\Sigma_{ZX_{j'}^\perp}} \tag{57}$$

$$= \frac{g_j \Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'} \Sigma_{ZX_{j'}^\perp}} - \frac{g_k \Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'} \Sigma_{ZX_{j'}^\perp}}, \tag{58}$$

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

$$\alpha_{j,k}(GG') = \alpha_j(GG') - \frac{g_k \Sigma_{ZX_j^\perp}}{\sum_{j'} g_{j'} \Sigma_{ZX_{j'}^\perp}} \frac{\Sigma_{ZX_k^\perp}}{\Sigma_{ZX_k^\perp}} \tag{59}$$

$$= \alpha_j(GG') - \alpha_k(GG') \frac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}}. \tag{60}$$

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

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

□

COROLLARY 4. *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 $\omega_{j,k} = -\frac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}} = \frac{\Sigma_{ZX_j^\perp}}{\sum_{j\neq k}\Sigma_{ZX_j^\perp}}$

$$\alpha_j(GG')^{avg} = \frac{1}{K}\sum_{k=1}^{K}\alpha_{j,k}(GG') \tag{61}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\left[\alpha_j(GG') + \omega_{j,k}\alpha_k(G'G)\right] \tag{62}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\left[\alpha_j(GG') - \frac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}}\alpha_k(G'G)\right] \tag{63}$$

$$= \alpha_j(GG') - \frac{1}{K}\sum_{k=1}^{K}\left[\frac{\Sigma_{ZX_j^\perp}}{\Sigma_{ZX_k^\perp}}\frac{g_k\Sigma_{ZX_k^\perp}}{\sum_{j'=1}^{K} g_{j'}\Sigma_{ZX_{j'}^\perp}}\right] \tag{64}$$

$$= \alpha_j(GG') - \frac{\Sigma_{ZX_j^\perp}}{K}\left[\frac{\sum_{k=1}^{K} g_k}{\sum_{k=1}^{K} g_k\Sigma_{ZX_k^\perp}}\right]. \tag{65}$$

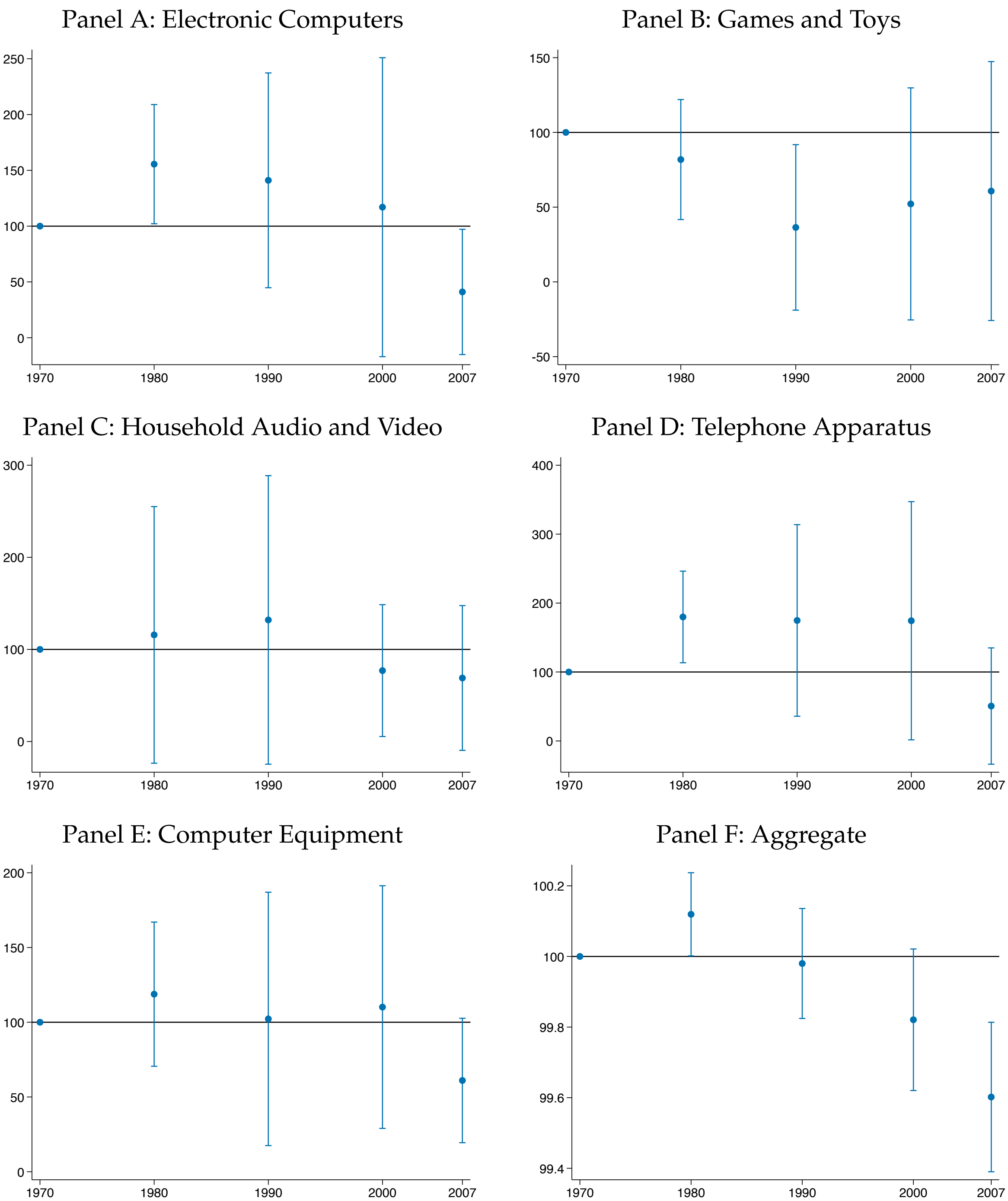
□

G.1 Empirical robustness

In the canonical Bartik setting (the only one of our three examples where the shares sum to one), we consider the impact of three ways of normalizing the growth rates on the 10 largest Rotemberg weights. Panel A of Appendix Table A4 repeats the results from Table 1 where we subtract off the simple mean of growth rates in each time period. Panel B shows what happens if we do not demean. Finally, in Panel C we demean using the mean of growth rates averaged across three time periods.

The Table shows that in this setting the Rotemberg weights are not sensitive to reasonable perturbations on the normalization. In all three cases, the industries with eight largest Rotemberg weights are the same, and in almost identical order. Though the sizes are slightly different, these differences are quite small.

Figure A4: Pre-trends for high Rotemberg weight industries (1990 shares): China shock



Notes: These figures report pre-trends for the overall instrument and the top-5 Rotemberg weight industries as reported in Table A1. The figures fix industry shares at the 1990 values and report the effect of these industry shares on manufacturing employment. For our controls, we fix the controls in the same time period, and interact with time fixed effects. As in the main specification, we also control for region and time fixed effects as well. 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. For the aggregate panel, we use the Bartik estimate for 1990.

Table A4: Robustness of Rotemberg weights: canonical setting

Panel A: Top 10 Rotemberg weight industries (baseline)		
	$\hat{\alpha}_k$	Ind Share
Oil+Gas Extraction	0.229	0.568
Motor Vehicles	0.140	1.404
Other	0.091	1.697
Guided Missiles	0.069	0.236
Blast furnaces	0.058	0.800
Construction and material handling machines	0.055	0.444
Landscaping	0.039	0.213
Electrical machinery, equipment, and supplies, n.s.	0.035	0.182
Coal mining	0.033	0.317
Petroleum refining	0.032	0.211
Panel B: Top 10 Rotemberg weight industries (no demeaning)		
	$\hat{\alpha}_k$	Ind Share
Oil+Gas Extraction	0.204	0.568
Motor Vehicles	0.167	1.404
Other	0.125	1.697
Guided Missiles	0.075	0.236
Construction and material handling machines	0.046	0.444
Blast furnaces	0.046	0.800
Landscaping	0.041	0.213
Electrical machinery, equipment, and supplies, n.s.	0.038	0.182
Computers and related equipment	0.036	0.498
National security and international affairs	0.033	0.736
Panel C: Top 10 Rotemberg weight industries (simple demeaning)		
	$\hat{\alpha}_k$	Ind Share
Oil+Gas Extraction	0.204	0.568
Motor Vehicles	0.167	1.404
Other	0.125	1.697
Guided Missiles	0.075	0.236
Construction and material handling machines	0.046	0.444
Blast furnaces	0.046	0.800
Landscaping	0.041	0.213
Electrical machinery, equipment, and supplies, n.s.	0.038	0.182
Computers and related equipment	0.036	0.498
National security and international affairs	0.033	0.736

Notes: This table reports statistics about the Rotemberg weights across alternative growth rate demeaning examples. Panel A reports the top ten industries according to the Rotemberg weights, replicating the demeaning from the main text. Panel B reports the top ten industries according to the Rotemberg weights, without demeaning. Panel C reports the top ten industries according to the Rotemberg weights, demeaning using the simple mean. The "Other" industry is the "N/A" code in the IND1990 classification system and includes full-time military personnel.

H 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 10. *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:

$$\hat{\beta}_1 = \hat{\beta}_2 \tag{66}$$

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:

$$\begin{aligned} G_j' Z'Y(G_j' Z'X)^{-1} &= \frac{\sum_k G_{j,k} Z_k' Y}{\sum_k G_{j,k} Z_k' X} && (67)\\ &= \frac{G_{j,1} Z_1' Y + \sum_{k=2}^K G_{j,k} Z_k' Y}{G_{j,1} Z_1' X + \sum_{k=2}^K G_{j,k} Z_k' X} && (68)\\ &= \frac{G_{j,1} Z_1' Y + \sum_{k=2}^K G_{j,k} 0}{G_{j,1} Z_1' X + \sum_{k=2}^K G_{j,k} 0} && (69)\\ &= \frac{Z_1' Y}{Z_1' X}, && (70) \end{aligned}$$

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:

$$
\begin{aligned}
Cov(B_1, B_2) &= \mathbb{E}[B_1 B_2] - \mathbb{E}[B_1]\mathbb{E}[B_2] &(71)\\
&= \mathbb{E}[(ZG_1)(ZG_2)] - \mathbb{E}[ZG_1]\mathbb{E}[ZG_2] &(72)\\
&= \mathbb{E}[(ZG_1)'(ZG_2)] - \mathbb{E}[ZG_1]\mathbb{E}[ZG_2] &(73)\\
&= G_1'\mathbb{E}[Z'Z]G_2 - G_1'\mathbb{E}[Z']\mathbb{E}[Z]G_2 &(74)\\
&= G_1'[\mathbb{E}[Z'Z] - \mathbb{E}[Z']\mathbb{E}[Z]]G_2 &(75)\\
&= G_1' Var(Z) G_2, &(76)
\end{aligned}
$$

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

$$TG_2 = 0. \tag{77}$$

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 $\sum_{k=1}^{K} T_k G_{2,k} = 0 \Rightarrow G_{2,K} = -\frac{\sum_{k=1}^{K-1} T_k G_{2,k}}{T_K}$.
□

I The Rotemberg weights with leave-one-out

The formulas we present in Section III 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 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:

$$B_{l,-l} = Z_l \left(\frac{L}{L-1} G_L - \frac{1}{L-1} G_l \right) \tag{78}$$

$$B_{l,-l} = \frac{L}{L-1} Z_l G - \frac{1}{L-1} Z_l G_l \tag{79}$$

$$B_{l,-l} = \frac{L}{L-1} B_l - \frac{1}{L-1} X_l, \tag{80}$$

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

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

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

Then:

$$\hat{\beta} = \frac{B'_{-l} Y}{B'_{-l} X} \tag{81}$$

$$= \frac{\left(\frac{L}{L-1} B - \frac{1}{L-1} X\right)' Y}{\left(\frac{L}{L-1} B - \frac{1}{L-1} X\right)' X} \tag{82}$$

$$= \frac{\left(\frac{L}{L-1} (ZG) - \frac{1}{L-1} X\right)' Y}{\left(\frac{L}{L-1} (ZG) - \frac{1}{L-1} X\right)' X}. \tag{83}$$

As before:

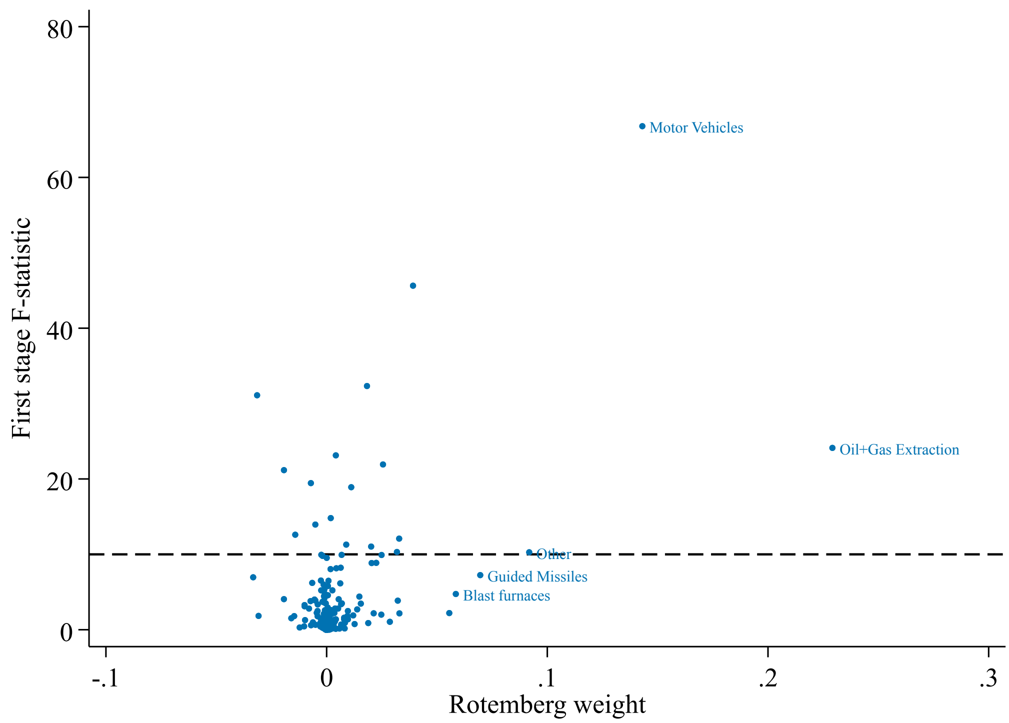
$$\beta_k = \frac{Z'_k Y}{Z'_k X}. \tag{84}$$

Then one can show:

$$\alpha_k = \frac{\frac{L}{L-1} g_k Z'_k X - \frac{1}{L-1} X' Y \beta_k^{-1}}{\sum_k \frac{L}{L-1} g_k Z'_k X - \frac{1}{L-1} X' X}. \tag{85}$$

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.

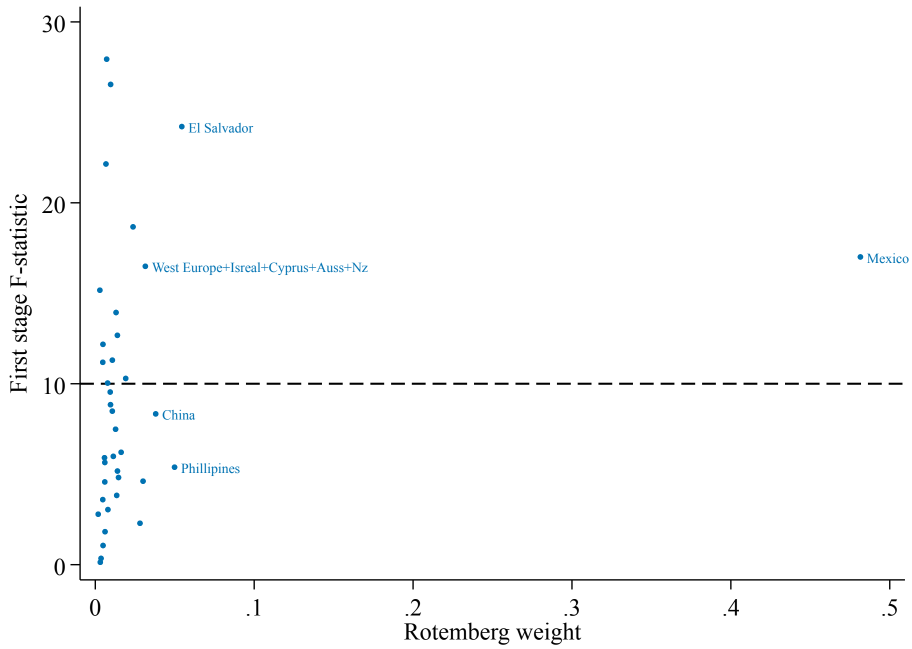
Figure A5: First stage versus Rotemberg weights: canonical setting



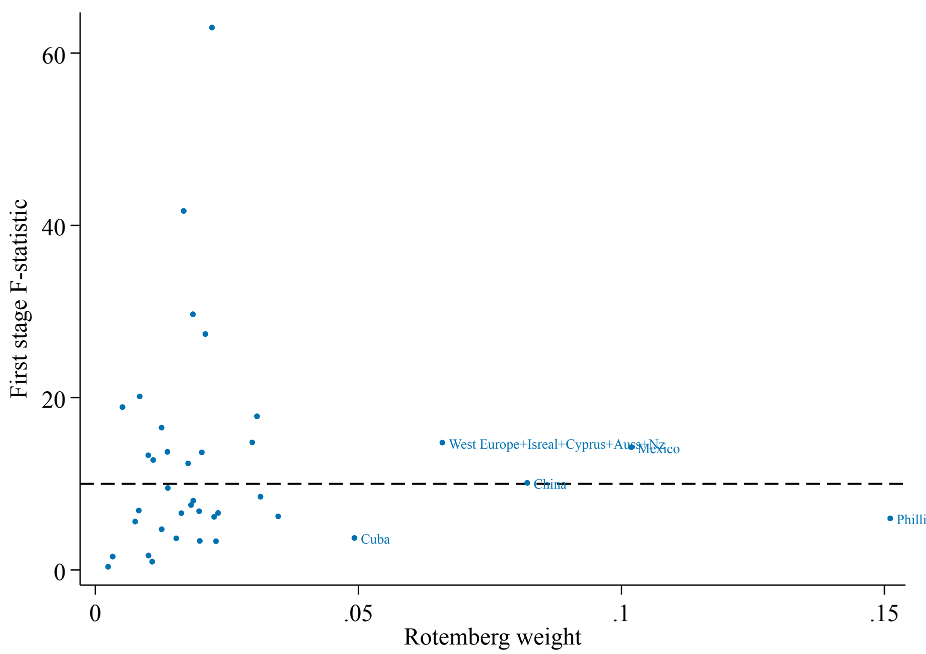
Notes: This figure plots each instrument's Rotemberg weight against the first stage F-statistic. Each point represents the estimates for an instrument, where instruments are aggregated across time periods following Section III.C. The labelled industries correspond to the five highest Rotemberg weight industries from Table 1. The dashed horizontal line is equal to 10.

Figure A6: First stage versus Rotemberg weights: immigrant enclave

Panel A: High school equivalent



Panel B: College equivalent



*Notes:*This figure plots each instrument's Rotemberg weight against the first stage F-statistic. Each point represents the estimates for an instrument, where instruments are aggregated across time periods following Section III.C. The labelled industries correspond to the five highest Rotemberg weight industries from Table 4. The dashed horizontal line is equal to 10.

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. 2019. "Replication data for: Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics [publisher]*. Unpublished data provided in personal communication by Michal Kolesar [distributor].

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

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 Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Autor, David, and David Dorn. 2013. "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market." *American Economic Review*, 103(5): 1553–1597.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013*a*. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–2168.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013*b*. "Replication data for: The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Association [publisher]*, `https://doi.org/https://doi.org/10.3886/E112670V1`. Inter-university Consortium for Political and Social Research [distributor].

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019*a*. "Quasi-experimental Shift-share Research Designs." Working Paper.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2019*b*. "Replication data for: Quasi-experimental Shift-share Research Designs." *Github repository, last accessed February 22, 2020*. `https://github.com/borusyak/shift-share`.

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.

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

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2020. "Do Credit Market Shocks affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times." *American Economic Journal: Economic Policy*, 12(1): 200–225.

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.

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.

Kovak, Brian K. 2013. "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" *American Economic Review*, 103(5): 1960–1976.

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.

Topalova, Petia. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics*, 2(4): 1–41.