Appendix to "Quasi-Experimental Shift-Share Research Designs"

Kirill Borusyak, UCL
Peter Hull, University of Chicago
Xavier Jaravel, London School of Economics

December 2020

Contents

A Appendix Results

A.1 Heterogeneous Treatment Effects

In this appendix we consider what a linear SSIV identifies when the structural relationship between y_ℓ and x_ℓ is nonlinear. We show that under a first-stage monotonicity condition the large-sample SSIV coefficient estimates a convexly weighted average of heterogeneous treatment effects. This holds even when the instrument has different effects on the outcome depending on the underlying realization of shocks, for example when $y_\ell = \sum_n s_{\ell n}\tilde{\beta}_{\ell n} x_{\ell n} + \varepsilon_\ell$ with $\tilde{\beta}_{\ell n}$ capturing the effects of (possibly unobserved) observation- and shock-specific treatments $x_{\ell n}$ making up the observed $x_\ell = \sum_n s_{\ell n} x_{\ell n}$.

Consider a general structural outcome model of

$$y_\ell = y(x_{\ell 1}, \dots, x_{\ell R}, \varepsilon_\ell), \tag{A1}$$

where the R treatments are given by $x_{\ell r} = x_r(g, \eta_{\ell r})$ with g collecting the vector of shocks g_n and with $\eta_\ell = (\eta_{\ell 1}, \dots, \eta_{\ell R})$ capturing first-stage heterogeneity. We consider an IV regression of y_ℓ on some aggregated treatment $x_\ell = \sum_r \alpha_{\ell r} x_{\ell r}$ with $\alpha_{\ell r} \geq 0$. Note that this nests the case of a single aggregate treatment ($R = 1$ and $\alpha_{\ell 1} = 1$) with arbitrary effect heterogeneity, as well as the special case above ($R = N$ and $\alpha_{\ell r} = s_{\ell n}$). We abstract away from controls w_ℓ and assume each shock is as-good-as-randomly assigned (mean-zero and mutually independent) conditional on the vector of second-stage unobservables ε_ℓ and the matrices of first-stage unobservables $\eta_{\ell r}$, exposure shares $s_{\ell n}$, importance weights e_ℓ, and aggregation weights $\alpha_{\ell r}$, collected in $\mathcal{I} = \{\varepsilon_\ell, e_\ell, \{\eta_{\ell r}, \alpha_{\ell r}\}_r, \{s_{\ell n}\}_n\}_\ell$. This assumption is stronger than Assumption 3 but generally necessary in a non-linear setting while still allowing for the endogeneity of exposure shares. For further notational simplicity we assume that $y(\cdot, \varepsilon_\ell)$ and each $x_r(\cdot, \eta_{\ell r})$ are almost surely continuously differentiable, such that $\beta_{\ell r}(\cdot) = \frac{\partial}{\partial x_r} y(\cdot, \varepsilon_\ell)$ captures the effect, for observation ℓ, of marginally increasing treatment r on the outcome and $\pi_{\ell n r}(\cdot) = \frac{\partial}{\partial g_n} x_r(\cdot, \eta_{\ell r})$ captures the effect of marginally increasing the nth shock on the rth treatment at ℓ.

Under an appropriate law of large numbers, the shift-share IV estimator approximates the IV estimand:

$$\hat{\beta} = \frac{\mathbb{E}\left[\sum_\ell e_\ell z_\ell y_\ell\right]}{\mathbb{E}\left[\sum_\ell e_\ell z_\ell x_\ell\right]} + o_p(1) = \frac{\sum_\ell \sum_n \mathbb{E}\left[s_{\ell n} e_\ell g_n y_\ell\right]}{\sum_\ell \sum_n \sum_r \mathbb{E}\left[s_{\ell n} e_\ell g_n \alpha_{\ell r} x_{\ell r}\right]} + o_p(1). \tag{A2}$$

Given this, we have the following result:

Proposition A1 When $\pi_{\ell n r}([\breve{g}; g_{-n}]) \geq 0$ almost surely for all $\breve{g} \in \mathbb{R}$, equation (A2) can be written

$$\hat{\beta} = \frac{\sum_\ell \sum_n \sum_r \mathbb{E}\left[\int_{-\infty}^{\infty} \tilde{\beta}_{\ell n r}(\breve{g}) \omega_{\ell n r}(\breve{g})\right] d\gamma}{\sum_\ell \sum_n \sum_r \mathbb{E}\left[\int_{-\infty}^{\infty} \omega_{\ell n r}(\breve{g})\right] d\gamma} + o_p(1), \tag{A3}$$

where $\omega_{\ell nr}(\breve{g}) \geq 0$ almost surely and

$$\tilde{\beta}_{\ell nr}(\breve{g}) = \frac{\beta_{\ell r}(x_1([\breve{g}; g_{-n}], \eta_{\ell 1}), \dots x_R([\breve{g}; g_{-n}], \eta_{\ell R}))}{\alpha_{\ell r}} \tag{A4}$$

is a rescaled treatment effect, evaluated at $(x_1([\breve{g}; g_{-n}], \eta_{\ell 1}), \dots x_R([\breve{g}; g_{-n}], \eta_{\ell R})$ for $[\breve{g}; g_{-n}] = (g_1, \dots, g_{n-1}, \breve{g}, g_{n+1}, \dots g_N)'$.

Proof See Appendix B.3.

This shows that in large samples $\hat{\beta}$ estimates a convex average of rescaled treatment effects, $\tilde{\beta}_{\ell nr}(\breve{g})$, when the first stage is monotone in each shock. Appendix B.3 shows that the weights $\omega_{\ell nr}(\breve{g})$ are proportional to the first-stage effects $\pi_{\ell nr}([\breve{g}; g_{-n}])$, exposure shares $s_{\ell n}$, regression weights e_ℓ, treatment aggregation weights $\alpha_{\ell r}$, and a function of the shock distribution. In the case without aggregation, i.e. $R = \alpha_{\ell r} = 1$, there is no rescaling in the $\tilde{\beta}_{\ell nr}(\breve{g})$. Equation (A3) then can be seen as generalizing the result of Angrist et al. (2000), on the identification of heterogeneous effects of continuous treatments, to the continuous shift-share instrument case. Intuition for the $\omega_{\ell nr}(\breve{g})$ weights follows similarly from this connection. With aggregation—that is, when the realization of shocks may have heterogeneous effects on y_ℓ holding the aggregated x_ℓ fixed—equation (A3) shows that SSIV captures a convex average of treatment effects per aggregated unit. Thus in the leading example of $y_\ell = \sum_n s_{\ell n}\tilde{\beta}_{\ell n} x_{\ell n} + \varepsilon_\ell$ and $x_\ell = \sum_n s_{\ell n} x_{\ell n}$, this result establishes identification of a convex average of the $\tilde{\beta}_{\ell n}$. In this way the result generalizes Adão et al. (2019), who establish the identification of convex averages of rescaled treatment effects in reduced form shift-share regressions.

A.2 Unobserved n-level Shocks Violate Share Exogeneity

In this appendix, we show that the assumption of SSIV share exogeneity from Goldsmith-Pinkham et al. (2020) is violated when there are unobserved shocks ν_n that affect outcomes via the exposure shares $s_{\ell n}$, i.e. when the residual has the structure

$$\varepsilon_\ell = \sum_n s_{\ell n}\nu_n + \breve{\varepsilon}_\ell. \tag{A5}$$

We consider large-sample violations share exogeneity in terms of the asymptotic non-ignorability of the $\bar{\varepsilon}_n$ terms in the equivalent moment condition (5). It is intuitive that the cross-sectional dependence between $s_{\ell n}$ and ε_ℓ will not asymptotically vanish when N is fixed (as in Goldsmith-Pinkham et al. (2020)) and each ν_n shock contributes significantly to the residual, causing $\bar{\varepsilon}_n \overset{p}{\not\to} 0$ for some or all n. We next prove this result and show that it generalizes to the case of increasing N, where the contribution of each ν_n to the variation in ε_ℓ becomes small. The intuition here is that the SSIV relevance condition generally requires individual observations to be sufficiently concentrated in a small number of shocks (see Section 3.1), and under this condition the share exogeneity violations remain asymptotically non-ignorable even as $N \to \infty$.

We define share endogeneity as non-vanishing $\text{Var}\left[\bar{\varepsilon}_n\right]$ at least for some n. This will tend to make the SSIV estimator inconsistent, unless shocks are as-good-as-randomly assigned (Assumption 1), even if the importance weights of individual shocks, s_n, converge to zero (Assumption 2). Here we treat e_ℓ and $s_{\ell n}$ as non-stochastic to show this result with simple notation.

Proposition A2 Suppose condition (A5) holds with the ν_n mean-zero and uncorrelated with the $\check{\varepsilon}_\ell$ and with each other, and with $\text{Var}\left[\nu_n\right] = \sigma_n^2 \geq \sigma_\nu^2$ for a fixed $\sigma_\nu^2 > 0$. Also assume $H_L = \sum_\ell e_\ell \sum_n s_{\ell n}^2 \to \bar{H} > 0$ such that first-stage relevance can be satisfied. Then there exists a constant $\delta > 0$ such that $\max_n \text{Var}\left[\bar{\varepsilon}_n\right] > \delta$ for sufficiently large L.

Proof See Appendix B.4.

A.3 Comparing SSIV and Native Shock-Level Regression Estimands

In this appendix we illustrate economic differences between the estimands of two regressions that researchers may consider: SSIV using outcome and treatment observations y_ℓ and x_ℓ (which we show in Proposition 1 are equivalent to certain shock-level IV regressions), and more conventional shock-level IV regressions using "native" y_n and x_n. These outcomes and treatments capture the same economic concepts as the original y_ℓ and x_ℓ, in contrast to the constructed $\bar{y}_n$ and $\bar{x}_n$ discussed in Section 2.3. In line with the labor supply and other key SSIV examples, we will for concreteness refer to the ℓ and n as indexing regions and industries, respectively. We consider the case where both the outcome and treatment can be naturally defined at the level of region-by-industry cells (henceforth, cells)—$y_{\ell n}$ and $x_{\ell n}$, respectively—and thus suitable for aggregation across either dimension with some weights $E_{\ell n}$ (e.g., cell employment growth rates aggregated with lagged cell employment weights): $y_\ell = \sum_n s_{\ell n} y_{\ell n}$ for $s_{\ell n} = \frac{E_{\ell n}}{\sum_{n'} E_{\ell n'}}$ and $y_n = \sum_\ell \omega_{\ell n} y_{\ell n}$ for $\omega_{\ell n} = \frac{E_{\ell n}}{\sum_{\ell'} E_{\ell' n}}$, with analogous expressions for x_ℓ and x_n. We further define $E_\ell = \sum_n E_{\ell n}$ and $E_n = \sum_\ell E_{\ell n}$.[1]

We consider the estimands of two regression specifications: β from the regional level model (2), instrumented by z_ℓ and weighted by $e_\ell = E_\ell / E$ for $E = \sum_\ell E_\ell$, and β_{ind} from a simpler industry-level IV regression of

$$y_n = \beta_{\text{ind}} x_n + \varepsilon_n, \tag{A6}$$

instrumented by the industry shock g_n and weighted by $s_n = E_n / E$. For simplicity we do not include any controls in either specification and implicitly condition on $\{E_{\ell n}\}_{\ell, n}$ (and some other variables as

[1] This formulation nests reduced-form shift-share regressions when $x_{\ell n} = g_n$ for each ℓ. The labor supply example of Section 2.1 fits only partially in this formal setup because the industry or regional wage growth y_n is not equal to a weighted average of wage growth across cells: reallocation of employment affects the average wage growth even in the absence of wage changes in any given cell.

described below), viewing them as non-stochastic.[2]

We show that β and β_{ind} generally differ when there are within-region spillover effects or when treatment effects are heterogeneous. We study these cases in turn, maintaining several assumptions: (i) a first stage relationship analogous to the one considered in Section 3.1:

$$x_{\ell n} = \pi_{\ell n} g_n + \eta_{\ell n}, \tag{A7}$$

for non-stochastic $\pi_{\ell n} \geq \bar{\pi} > 0$, (ii) a stronger version of our Assumption 1 that imposes $\mathbb{E}[g_n] = \mathbb{E}[g_n \varepsilon_{\ell n'}] = \mathbb{E}[g_n \eta_{\ell n'}] = 0$ for all ℓ, n, and n', with $\varepsilon_{\ell n'}$ denoting the unobserved cell-level residual of each model, (iii) the assumption that g_n is uncorrelated with $g_{n'}$ for all n and n', and (iv) that all appropriate laws of large numbers hold.

Within-Region Spillover Effects Suppose the structural model at the cell level is given by

$$y_{\ell n} = \beta_0 x_{\ell n} - \beta_1 \sum_{n'} s_{\ell n'} x_{\ell n'} + \varepsilon_{\ell n}. \tag{A8}$$

Here β_0 captures the direct effect of the shock on the cell outcome, and β_1 captures a within-region spillover effect. The local employment effects of industry demand shocks from the model in Appendix A.7 fit in this framework, see equation (A31).[3] The following proposition shows that the SSIV estimand β captures the effect of treatment net of spillovers (i.e. $\beta_0 - \beta_1$), whereas β_{ind} subtracts the spillover only partially; this is intuitive since the spillover effect is fully contained within regions but not within industries.

Proposition A3 Suppose equation (A8) holds and the average local concentration index $H_L = \sum_{\ell,n} e_\ell s_{\ell n}^2$ is bounded from below by a constant $\bar{H}_L > 0$. Further assume $\pi_{\ell n} = \bar{\pi}$ and $\text{Var}[g_n] = \sigma_g^2$ for all ℓ and n. Then the SSIV estimator satisfies

$$\hat{\beta} = \beta_0 - \beta_1 + o_p(1) \tag{A9}$$

while the native industry-level IV estimator satisfies

$$\hat{\beta}_{\text{ind}} = \beta_0 - \beta_1 H_L + o_p(1), \tag{A10}$$

If $\beta_1 \neq 0$ (i.e. in presence of within-region spillovers), $\hat{\beta}$ and $\hat{\beta}_{\text{ind}}$ asymptotically coincide if and only if $H_L \xrightarrow{p} 1$, which corresponds to the case where the average region is asymptotically concentrated in one industry.

[2] Note that we thereby condition on the shares $s_{\ell n}$ and importance weights e_ℓ. Yet we still allow for share endogeneity by not restricting $\mathbb{E}[\varepsilon_{\ell n}]$ to be zero.

[3] In the labor supply example from the main text $y_{\ell n}$ is the cell wage, which is equalized within the region, and $x_{\ell n}$ is cell employment. Equation (A8) therefore holds for $\beta_0 = 0$ and $-\beta_1$ being the inverse labor supply elasticity.

PROOF See Appendix B.5.

Treatment Effect Heterogeneity Now consider a different structural model which allows for heterogeneity in treatment effects:

$$y_{\ell n} = \beta_{\ell n} x_{\ell n} + \varepsilon_{\ell n}. \tag{A11}$$

We also allow the first-stage coefficients $\pi_{\ell n}$ and shock variance σ_n^2 to vary. The following proposition shows that β and β_{ind} differ in how they average effect $\beta_{\ell n}$ (here treated as non-stochastic) across the (ℓ, n) cells. The weights corresponding to the SSIV estimand β are relatively higher for cells that represent a larger fraction of the regional economy. This follows because in the regional regression $s_{\ell n}$ determines the cell's weight in both the outcome and the shift-share instrument, while in the industry regression only the former argument applies. Heterogeneity in the $\pi_{\ell n}$ and σ_n^2, in contrast, has equivalent effects on the weighting scheme of both estimands.

Proposition A4 In the casual model (A11),

$$\hat{\beta} = \frac{\sum_{\ell,n} E_{\ell n} s_{\ell n} \pi_{\ell n} \sigma_n^2 \cdot \beta_{\ell n}}{\sum_{\ell,n} E_{\ell n} s_{\ell n} \pi_{\ell n} \sigma_n^2} + o_p(1) \tag{A12}$$

and

$$\hat{\beta}_{\text{ind}} = \frac{\sum_{\ell,n} E_{\ell n} \pi_{\ell n} \sigma_n^2 \cdot \beta_{\ell n}}{\sum_{\ell,n} E_{\ell n} \pi_{\ell n} \sigma_n^2} + o_p(1), \tag{A13}$$

PROOF See Appendix B.6.

A.4 Connection to Rotemberg Weights

In this appendix we rewrite the decomposition of the SSIV coefficient $\hat{\beta}$ from Goldsmith-Pinkham et al. (2020) that gives rise to their "Rotemberg weight" interpretation, and show that these weights measure the leverage of shocks in our equivalent shock-level IV regression. We then show that, in our framework, skewed Rotemberg weights do not measure sensitivity to misspecification (of share exogeneity) and do not pose a problem for SSIV consistency. We finally discuss the implications of high-leverage observations for SSIV inference.

Proposition 1 implies the following decomposition:

$$\hat{\beta} = \frac{\sum_n s_n g_n \bar{y}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp} = \sum_n \alpha_n \hat{\beta}_n, \tag{A14}$$

where

$$\hat{\beta}_n = \frac{\bar{y}_n^\perp}{\bar{x}_n^\perp} = \frac{\sum_\ell e_\ell s_{\ell n} y_\ell^\perp}{\sum_\ell e_\ell s_{\ell n} x_\ell^\perp} \tag{A15}$$

and

$$\alpha_n = \frac{s_n g_n \bar{x}_n^\perp}{\sum_{n'} s_{n'} g_{n'} \bar{x}_{n'}^\perp}. \tag{A16}$$

This is a shock-level version of the decomposition discussed in Goldsmith-Pinkham et al. (2020): $\hat{\beta}_n$ is the IV estimate of β that uses share $s_{\ell n}$ as the instrument, and α_n is the so-called Rotemberg weight.

To see the connection with leverage (defined, typically in the context of OLS, as the derivative of each observation's fitted value with respect to its outcome) in our equivalent IV regression, note that

$$\frac{\partial \left(\bar{x}_n^\perp \hat{\beta}\right)}{\partial \bar{y}_n^\perp} = \bar{x}_n^\perp \frac{s_n g_n}{\sum_{n'} s_{n'} g_{n'} \bar{x}_{n'}^\perp} = \alpha_n. \tag{A17}$$

In this way, α_n measures the sensitivity of $\hat{\beta}$ to $\hat{\beta}_n$.

In the preferred interpretation of Goldsmith-Pinkham et al. (2020), exposure to each shock is a valid instrument such that $\hat{\beta}_n \xrightarrow{p} \beta$ for each n. However, in our framework deviations of $\hat{\beta}_n$ from β reflect nonzero $\bar{\varepsilon}_n$ in large samples, and such share endogeneity is not ruled out; thus α_n does not have the same sensitivity-to-misspecification interpretation. Moreover, a high leverage of certain shocks ("skewed Rotemberg weights," in the language of Goldsmith-Pinkham et al. (2020)) is not a problem for consistency in our framework, provided it results from a heavy-tailed and high-variance distribution of shocks (that still satisfies our regularity conditions, such as finite shock variance), and each s_n is small as required by Assumption 2.

Nevertheless, skewed α_n may cause issues with SSIV inference, as would high leverage observations in any regression. In general, the estimated residuals $\hat{\bar{\varepsilon}}_n^\perp$ of high-leverage observations will tend to be biased toward zero, which may lead to underestimation of the residual variance and too small standard errors (e.g., Cameron and Miller 2015). This issue can be addressed, for instance, by computing confidence intervals with the null imposed, as Adão et al. (2019) recommend and as we discuss in Section 5.1. In practice our Monte-Carlo simulations in Appendix A.11 find that the coverage of conventional exposure-robust confidence intervals to be satisfactory even with Rotemberg weights as skewed as those reported in the applications of Goldsmith-Pinkham et al. (2020) analysis.

A.5 Consistency of Control Coefficients

This appendix shows how the control coefficient γ, defined in main text footnote 5, can be consistently estimated as required in Proposition 3 (Assumption B2). We discuss conditions for $\sum_\ell e_\ell w_\ell \varepsilon_\ell \xrightarrow{p} \mathbb{E}\left[\sum_\ell e_\ell w_\ell \varepsilon_\ell\right]$, where by definition $\mathbb{E}\left[\sum_\ell e_\ell w_\ell \varepsilon_\ell\right] = 0$. Consistency of the estimator $\hat{\gamma} = \gamma + \left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1} \sum_\ell e_\ell w_\ell \varepsilon_\ell$ follows, provided the elements of $\left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1}$ are stochastically bounded (i.e., $O_p(1)$). For simplicity we consider control vectors w_ℓ of fixed length.

The argument for convergence of $\sum_\ell e_\ell w_\ell \varepsilon_\ell$ depends on the source of randomness in w_ℓ and ε_ℓ. We consider two characteristic cases. In the first case, $(e_\ell, w_\ell', \varepsilon_\ell)'$ can be viewed as *iid* or clustered in a conventional way. For example, w_ℓ and ε_ℓ may contain observed and unobserved local labor supply shocks which are uncorrelated across markets, clusters of markets (e.g. states), or beyond a given distance threshold. In this case conventional laws of large numbers can be used to establish $\sum_\ell e_\ell w_\ell \varepsilon_\ell \xrightarrow{p} 0$. For instance if $(e_\ell, w_\ell', \varepsilon_\ell)'$ is *iid* then $\sum_\ell e_\ell w_\ell \varepsilon_\ell$ gives a vector of sample averages of mutually uncorrelated mean-zero random variables, which weakly converge to zero when the e_ℓ weights are asymptotically dispersed ($\mathbb{E}\left[\sum_\ell e_\ell^2\right] \to 0$) and when $\mathbb{E}\left[w_\ell^2 \varepsilon_\ell^2 \mid e\right]$ is uniformly bounded.

In the second case, either w_ℓ or ε_ℓ has a shift-share structure like z_ℓ: i.e. $w_\ell = \sum_n s_{\ell n} q_n$ for an observed q_n (in line with our Proposition 3) or $\varepsilon_\ell = \sum_n s_{\ell n} \nu_n$ for an unobserved ν_n (capturing, for example, a set of unobserved industry-level factors averaged with the employment weights $s_{\ell n}$). In this case convergence of $\sum_\ell e_\ell w_\ell \varepsilon_\ell$ can be shown to follow similarly to the convergence of the sample analog of the instrument moment condition (3). If, for instance, $\varepsilon_\ell = \sum_n s_{\ell n} \nu_n$ with $\mathbb{E}\left[\nu_n \mid s, w\right] = 0$ and $\mathrm{Cov}\left[\nu_n, \nu_m \mid s, w\right] = 0$ for $w = \{w_n\}_n$, then for each control $\sum_\ell e_\ell w_{\ell m} \varepsilon_\ell = \sum_n s_n \nu_n \bar{w}_{nm}$ weakly converges when the s_n weights are dispersed ($\mathbb{E}\left[\sum_n s_n^2\right] \to 0$) and both $\mathrm{Var}\left[\nu_n \mid s, w\right]$ and $\mathbb{E}\left[\bar{w}_{nm}^2 \mid s\right]$ are uniformly bounded. This argument can be extended to the case where either w_ℓ or ε_ℓ is formed from different exposure shares $\tilde{s}_{\ell k}$, perhaps defined over a different range of K observed q_k or unobserved ν_k, and when the q_k or ν_k are clustered or otherwise weakly mutually correlated.

More generally, the two cases can be combined to settings where $w_\ell = \sum_k \tilde{s}_{\ell k} q_k + \check{w}_\ell$ and $\varepsilon_\ell = \sum_{k'} \tilde{s}_{\ell k'} \nu_{k'} + \check{\varepsilon}_\ell$ where $(e_\ell, \check{w}_\ell', \check{\varepsilon}_\ell)'$ is *iid* or conventionally clustered and where q_k and $\nu_{k'}$ are many weakly correlated random shocks or, even more generally, allowing for multiple shift-share terms with different exposure shares.

A.6 Estimated Shocks

This appendix establishes the formal conditions for the SSIV estimator, with or without a leave-one-out correction, to be consistent when shocks g_n are noisy estimates of some latent g_n^* satisfying Assumptions 1 and 2. We also propose a heuristic measure that indicates whether the leave-one-out correction is likely to be important and compute it for the Bartik (1991) setting. Straightforward extensions to other split-sample estimators follow.

Suppose a researcher estimates shocks via a weighted average of variables $g_{\ell n}$. That is, given weights $\omega_{\ell n} \geq 0$ such that $\sum_\ell \omega_{\ell n} = 1$ for all n, she computes

$$g_n = \sum_\ell \omega_{\ell n} g_{\ell n}. \tag{A18}$$

A leave-one-out (LOO) version of the shock estimator is instead

$$g_{n,-\ell} = \frac{\sum_{\ell' \neq \ell} \omega_{\ell' n} g_{\ell' n}}{\sum_{\ell' \neq \ell} \omega_{\ell' n}}. \tag{A19}$$

We assume that each $g_{\ell n}$ is a noisy version of the same latent shock g_n^*:

$$g_{\ell n} = g_n^* + \psi_{\ell n}, \tag{A20}$$

where g_n^* satisfies Assumptions 1 and 2 and $\psi_{\ell n}$ is estimation error (in Section 4.1 we considered the special case of $\psi_{\ell n} \propto \varepsilon_\ell$). This implies a feasible shift-share instrument of $z_\ell = z_\ell^* + \psi_\ell$ and its LOO version $z_\ell^{LOO} = z_\ell^* + \psi_\ell^{LOO}$, where $z_\ell^* = \sum_n s_{\ell n} g_n^*$, $\psi_\ell = \sum_n s_{\ell n} \sum_{\ell'} \omega_{\ell' n} \psi_{\ell' n}$, and $\psi_\ell^{LOO} = \sum_n s_{\ell n} \frac{\sum_{\ell' \neq \ell} \omega_{\ell' n} \psi_{\ell' n}}{\sum_{\ell' \neq \ell} \omega_{\ell' n}}$. Consistency with these instruments, given a first stage, requires that $\sum_\ell e_\ell \varepsilon_\ell \psi_\ell \xrightarrow{p} 0$ and $\sum_\ell e_\ell \varepsilon_\ell \psi_\ell^{LOO} \xrightarrow{p} 0$ respectively.

We now present three sets of results. First, we establish a simple sufficient condition under which the LOO instrument satisfies $\sum_\ell e_\ell \varepsilon_\ell \psi_\ell^{LOO} \xrightarrow{p} 0$. We also propose stronger conditions that guarantee consistency of LOO-SSIV. Second, we explore the conditions under which the covariance between ε_ℓ and $\psi_{\ell n}$ is ignorable, i.e. asymptotically does not lead to a "mechanical" bias of the conventional non-leave-one-out estimator. We propose a heuristic measure that is large when the bias is likely to be small. Lastly, we apply these ideas to the setting of Bartik (1991) using the data from Goldsmith-Pinkham et al. (2020). In line with previous appendices, we condition on $s_{\ell n}$, $\omega_{\ell n}$, and e_ℓ and treat them as non-stochastic for notational convenience. We also assume the SSIV regressions are estimated without controls w_ℓ.

LOO Identification and Consistency The following proposition establishes three results. The first is the most important one, providing the condition for orthogonality to hold . The second strengthens this condition so that the estimator converges, which naturally requires that most shocks are estimated with sufficient amount of data. A tractable case of complete specialization is considered in last part, where there should be many more observations than shocks.

Proposition A5

1. If $\mathbb{E}\left[\varepsilon_\ell \psi_{\ell' n}\right] = 0$ for all $\ell \neq \ell'$ and n, then $\mathbb{E}\left[\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell, LOO}\right] = 0$.

2. If $\mathbb{E}\left[(\varepsilon_\ell, \psi_{\ell n}) \mid \{(\varepsilon_{\ell'}, \psi_{\ell' n'})\}_{\ell' \neq \ell, n'}\right] = 0$ for all ℓ and n, then the LOO estimator is consistent, provided it has a first stage and two regularity conditions hold: $\mathbb{E}\left[\left|\varepsilon_{\ell_1} \varepsilon_{\ell_2} \psi_{\ell_1' n_1} \psi_{\ell_2' n_2}\right|\right] \leq B$ for a constant B and all $(\ell_1, \ell_2, \ell_1', \ell_2', n_1, n_2)$ and

$$\sum_{\substack{(\ell_1, \ell_2, \ell_1', \ell_2') \in \mathcal{J}, \\ n_1, n_2}} e_{\ell_1} e_{\ell_2} s_{\ell_1 n_1} s_{\ell_2 n_2} \frac{\omega_{\ell_1' n_1}}{\sum_{\ell \neq \ell_1} \omega_{\ell n_1}} \frac{\omega_{\ell_2' n_2}}{\sum_{\ell \neq \ell_2} \omega_{\ell n_2}} \to 0, \tag{A21}$$

   with $\mathcal{J}$ denoting the set of tuples $(\ell_1, \ell_2, \ell_1', \ell_2')$ for which one of the two conditions hold: (i) $\ell_1 = \ell_2$ and $\ell_1' = \ell_2' \neq \ell_1$, (ii) $\ell_1 = \ell_2'$ and $\ell_2 = \ell_1' \neq \ell_1$.

3. Condition (A21) is satisfied if $\frac{N}{L} \to 0$ in the special case where each region is specialized in

one industry, i.e. $s_{\ell n} = \mathbf{1}[n = n(\ell)]$ for some $n(\cdot)$, there are no importance weights ($e_\ell = \frac{1}{L}$), and shocks estimated by simple LOO averaging among observations exposed to a given shock ($\omega_{\ell n} = \frac{1}{L_n}$ for $L_n = \sum_\ell \mathbf{1}[n(\ell) = n]$), assuming further that $L_n \geq 2$ for each n so that the LOO estimator is well-defined.

Proof See Appendix B.7.

The condition in the first part of Proposition A5 would be quite innocuous in random samples of ℓ – the environment in which leave-one-out adjustments are often considered (e.g. Angrist et al. (1999)) – but is strong without random sampling. It requires ε_ℓ and $\psi_{\ell' n}$ to be uncorrelated for $\ell' \neq \ell$, which may easily be violated when both ℓ and ℓ' are exposed to the same shocks—a situation in which excluding own observation is not sufficient. Moreover, since we have conditioned on the exposure shares throughout, $\mathbb{E}[\varepsilon_\ell \psi_{\ell' n}] = 0$ generally requires either ε_ℓ or $\psi_{\ell' n}$ to have a zero *conditional* mean—the share exogeneity assumption applied to either the residuals or the estimation error. At the same time, this condition does not require $\mathbb{E}[\varepsilon_\ell \psi_{\ell' n}] = 0$ for $\ell = \ell'$, which reflects the benefit of LOO: eliminating the mechanical bias from the residual directly entering shock estimates.

Heuristic for Importance of LOO Correction We now return to the non-LOO SSIV estimator. As in Proposition A5, we assume that $\mathbb{E}[\varepsilon_\ell \psi_{\ell' n}] = 0$ for $\ell' \neq \ell$ and all n, so the LOO estimator is consistent under the additional regularity conditions. We also assume, without loss of generality, that z_ℓ is mean-zero. Then the "mechanical bias" mentioned in Section 4.1 is the only potential problem: under appropriate regularity conditions (similar to those in part 2 of Proposition A5),

$$\begin{aligned}\hat{\beta} - \beta &= \frac{\mathbb{E}\left[\sum_\ell e_\ell \varepsilon_\ell \psi_\ell\right]}{\mathbb{E}\left[\sum_\ell e_\ell z_\ell x_\ell\right]} + o_p(1) \\ &= \frac{\sum_{\ell,n} e_\ell s_{\ell n} \omega_{\ell n} \mathbb{E}\left[\varepsilon_\ell \psi_{\ell n}\right]}{\mathbb{E}\left[\sum_\ell e_\ell z_\ell x_\ell\right]} + o_p(1). \end{aligned} \tag{A22}$$

With $|\mathbb{E}[\varepsilon_\ell \psi_{\ell n}]|$ bounded by some $B_1 > 0$ for all ℓ and n, the numerator of (A22) is bounded by $H_N B_1$, for an observable composite of the relevant shares $H_N = \sum_{\ell,n} e_\ell s_{\ell n} \omega_{\ell n}$. The structure of the shares also influences the strength of the first stage in the denominator. Imposing our standard model of the first stage from Section 3.1 (but specified based on the latent shock g_n^*), i.e. $x_\ell = \sum_n s_{\ell n} x_{\ell n}$ for $x_{\ell n} = \pi_{\ell n} g_n^* + \eta_{\ell n}$, $\eta_{\ell n}$ mean-zero and uncorrelated with $g_{n'}^*$ for all ℓ, n, n', $\mathrm{Var}[g_n^*] \geq \bar{\sigma}_g^2 > 0$ and $\pi_{\ell n} \geq \bar{\pi} > 0$, yields:

$$\begin{aligned}\mathbb{E}\left[\sum_\ell e_\ell z_\ell x_\ell\right] &= \sum_\ell e_\ell \mathbb{E}\left[\left(\sum_n s_{\ell n}\left(g_n^* + \psi_{\ell n}\right)\right)\left(\sum_{n'} s_{\ell n'}\left(\pi_{\ell n} g_{n'}^* + \eta_{\ell n'}\right)\right)\right] \\ &= \sum_{\ell,n} e_\ell s_{\ell n}^2 \cdot \pi_{\ell n} \mathrm{Var}\left[g_n^*\right] + \sum_\ell e_\ell \sum_{n,n'} s_{\ell n} s_{\ell n'} \mathbb{E}\left[\psi_{\ell n}\left(\pi_{\ell n} g_{n'}^* + \eta_{\ell n'}\right)\right]. \end{aligned} \tag{A23}$$

Excepting knife-edge cases where the two terms in (A23) cancel out, $\mathbb{E}\left[\sum_\ell e_\ell z_\ell^\perp x_\ell\right] \not\to 0$ provided $H_L = \sum_{\ell,n} e_\ell s_{\ell n}^2 \geq \bar{H}$ for some fixed $\bar{H} > 0$.

We thus define the following heuristic:

$$H = \frac{H_L}{H_N} = \frac{\sum_{\ell,n} e_\ell s_{\ell n}^2}{\sum_{\ell,n} e_\ell s_{\ell n}\omega_{\ell n}}. \tag{A24}$$

When H is large, we expect the non-LOO SSIV estimator to be relatively insensitive to the mechanical bias generated by the average covariance between $\psi_{\ell n}$ and ε_ℓ, and thus similar to the LOO estimator.

We note an important special case. Suppose all weights are derived from variable $E_{\ell n}$ (e.g. lagged employment level in region ℓ and industry n) as $s_{\ell n} = \frac{E_{\ell n}}{E_\ell}$, $\omega_{\ell n} = \frac{E_{\ell n}}{E_n}$, and $e_\ell = \frac{E_\ell}{E}$, for $E_\ell = \sum_n E_{\ell n}$, $E_n = \sum_\ell E_{\ell n}$, and $E = \sum_\ell E_\ell$. Then

$$H_N = \sum_{\ell,n} \frac{E_\ell}{E}\frac{E_{\ell n}}{E_\ell}\frac{E_{\ell n}}{E_n} = \sum_{\ell,n} \frac{E_n}{E}\left(\frac{E_{\ell n}}{E_n}\right)^2 = \sum_n s_n \sum_\ell \omega_{\ell n}^2, \tag{A25}$$

where $s_n = \frac{E_n}{E}$ is the weight in our equivalent shock-level regression. Therefore, H_N is the weighted average across n of n-specific Herfindahl concentration indices, while H_L is the weighted average across ℓ of ℓ-specific Herfindahl indices. With $E_{\ell n}$ denoting lagged employment, H is high (and thus we expect the LOO correction to be unnecessary) when employment is much more concentrated across industries in a typical region than it is concentrated across regions for a typical industry.

The formula simplifies further with $E_{\ell n} = \mathbf{1}[n = n(\ell)]$ for all ℓ, n, corresponding to the case of complete specialization of observations in shocks with no regression or shock estimation weights, as in part 3 of Proposition A5. In that case,

$$H = \frac{1}{\sum_\ell \frac{1}{L}\frac{1}{L_{n(\ell)}}} = \frac{1}{\frac{1}{L}\sum_n \sum_{\ell:\, n(\ell)=n} \frac{1}{L_n}} = \frac{L}{N}. \tag{A26}$$

Our heuristic is therefore large when there are many observations per estimated shock.[4]

Application to Bartik (1991) We finally apply our insights to the Bartik (1991) setting, using the Goldsmith-Pinkham et al. (2020) replication code and data. Table C6 reports the results. Column 1 shows the estimates of the inverse local labor supply elasticity using SSIV estimators with and without the LOO correction and using population weights, replicating Table 3, column 2, of Goldsmith-Pinkham et al. (2020) except with employment on the left-hand side and wages on the right-hand side.[5] Column 2 repeats the analysis without the population weights.[6] We find all estimates to range

[4]Here $1/H = N/L$ is proportional to the "bias" of the non-LOO estimator, which is similar to how the finite-sample bias of conventional 2SLS is proportional to the number of instruments over the sample size (Nagar 1959).

[5]Goldsmith-Pinkham et al. (2020) estimate the inverse labor supply elasticity. By properties of IV estimation, our coefficient is the inverse of theirs.

[6]Industry growth shocks in this column are the same as in Column 1, again estimated with employment weights.

between 1.2 and 1.3, showing that in practice for Bartik (1991) the LOO correction does not play a substantial role.

This is however especially true without weights, where the LOO and conventional SSIV estimators are 1.30 and 1.29, respectively. Our heuristic provides an explanation: H is almost 8 times bigger when computed without weights. The intuition is that large commuting zones, such as Los Angeles and New York, may constitute a substantial fraction of employment in industries of their comparative advantage. This generates a potential for the mechanical bias: labor supply shocks in those regions affect shock estimates; this bias is avoided by LOO estimators. However, the role of the largest commuting zones is only significant in weighted regressions (by employment or, as in Goldsmith-Pinkham et al. (2020), population).

A.7 Equilibrium Industry Growth in a Model of Local Labor Markets

This appendix develops a simple model of regional labor supply and demand, similar to the model in Adão et al. (2020). Our goal is to show how the national growth rate of industry employment can be viewed as a noisy version of the national industry-specific labor demand shocks, and how regional labor supply shocks (along with some other terms) generate the "estimation error."

Consider an economy that consists of a set of L regions. In each region ℓ there is a prevailing wage W_ℓ, and labor supply has constant elasticity ϕ:

$$E_\ell = M_\ell W_\ell^\phi, \tag{A27}$$

where E_ℓ is total regional employment and M_ℓ is the supply shifter that depends on the working-age population, the outside option, and other factors. Labor demand in each industry n is given by a constant-elasticity function

$$E_{\ell n} = A_n \xi_{\ell n} W_\ell^{-\sigma}, \tag{A28}$$

where $E_{\ell n}$ is employment, A_n is the national industry demand shifter, $\xi_{\ell n}$ is its idiosyncratic component, and σ is the elasticity of labor demand. The equilibrium is given by

$$\sum_n E_{\ell n} = E_\ell. \tag{A29}$$

Now consider small changes in fundamentals A_n, $\xi_{\ell n}$ and M_ℓ. We use log-linearization around the observed equilibrium and employ the Jones (1965) hat algebra notation, with $\hat{v}$ denoting the relative change in v between the equilibria. We then establish:

Proposition A6 After a set of small changes to fundamentals, the national industry employment growth is characterized by

$$g_n = \sum_\ell \omega_{\ell n} g_{\ell n}, \tag{A30}$$

for $\omega_{\ell n} = E_{\ell n} / \sum_{\ell'} E_{\ell' n}$ denoting the share of region ℓ in industry employment, and the change in region-by-industry employment $g_{\ell n}$ is characterized by

$$g_{\ell n} = g_n^* + \frac{\sigma}{\sigma + \phi} \varepsilon_\ell + \hat{\xi}_{\ell n} - \frac{\sigma}{\sigma + \phi} \sum_n s_{\ell n} \left(g_n^* + \hat{\xi}_{\ell n} \right), \tag{A31}$$

where $g_n^* = \hat{A}_n$ is the national industry labor demand shock, $\varepsilon_\ell = \hat{M}_\ell$ is the regional labor supply shock, and $s_{\ell n} = E_{\ell n} / \sum_{n'} E_{\ell n'}$.

Proof See Appendix B.8.

The first term in (A31) justifies our interpretation of the observed industry employment growth as a noisy estimate of the latent labor demand shock g_n^*. The other terms constitute the "estimation error." The first of them is proportional to the residual of the labor supply equation, ε_ℓ; we have previously established the conditions under which it may or may not confound SSIV estimation. The other terms, that we abstracted away from in Section 4.1, include the idiosyncratic demand shock $\hat{\xi}_{\ell n}$ and shift-share averages of both national and idiosyncratic demand shocks. If the model is correct, all of these are uncorrelated with ε_ℓ, thus not affecting Assumption 1.

A.8 SSIV Consistency in Short Panels

This appendix shows how alternative shock exogeneity assumptions imply the consistency of panel SSIV regressions with many fixed effect coefficients. We consider the incidental parameters problem in "short" panels, with fixed T and $L \to \infty$ and with unit fixed effects, in which case the control coefficient γ cannot be consistently estimated with the fixed effects included in w_ℓ. We show how an analog of Assumption 3 can be instead applied to a demeaned shock-level unobservable that partials out the fixed effect nuisance coefficients. A similar argument applies to period fixed effects in the fixed L and $T \to \infty$ asymptotic.

Suppose for the linear causal model $y_{\ell t} = \beta x_{\ell t} + \epsilon_{\ell t}$ and control vector $w_{\ell t}$ (which includes unit FEs), we define $\gamma = \mathbb{E}\left[\sum_\ell e_{\ell t} w_{\ell t}^\Delta w_{\ell t}^{\Delta\prime}\right]^{-1} \mathbb{E}\left[\sum_\ell e_{\ell t} w_{\ell t}^\Delta \epsilon_{\ell t}^\Delta\right]$ where $v_{\ell t}^\Delta$ is a subvector of the (weighted) unit-demeaned observation of variable $v_{\ell t}$, $v_{\ell t} - \frac{\sum_\tau e_{\ell \tau} v_{\ell \tau}}{\sum_\tau e_{\ell \tau}}$, that drops any elements that are identically zero (e.g. those corresponding to the unit FEs in $w_{\ell t}$). Note we have assumed no perfect multicollinearity in the remaining elements such that $\mathbb{E}\left[\sum_\ell e_{\ell t} w_{\ell t}^\Delta w_{\ell t}^{\Delta\prime}\right]$ is invertible. We can then write $y_{\ell t}^\Delta = \beta x_{\ell t}^\Delta + w_{\ell t}^{\Delta\prime} \gamma + \varepsilon_{\ell t}^\Delta$. Suppose also that $\sum_\ell e_{\ell t} z_{\ell t} x_{\ell t}^\perp \xrightarrow{p} \pi$ for some $\pi \neq 0$ and the analog of Assumption B2 for unit-demeaned controls holds. Then, following the proof to Proposition 3, $\hat{\beta}$ is consistent if and only if

$$\sum_{n,t} s_{nt} g_{nt} \bar{\varepsilon}_{nt}^{\Delta} \xrightarrow{p} 0, \tag{A32}$$

where $s_{nt} = \sum_\ell e_{\ell t} s_{\ell nt}$ and $\bar{\varepsilon}_{nt}^{\Delta} = \frac{\sum_\ell e_{\ell t} s_{\ell nt} \varepsilon_{\ell t}^{\Delta}}{\sum_\ell e_{\ell t} s_{\ell nt}}$. This condition is satisfied when analogs of Assumptions 1,2, and B1 hold, or under the various extensions discussed in Section 3. In particular when $w_{\ell t}$ contains t-specific FE the key assumption of quasi-experimental shock assignment is $\mathbb{E}\left[g_{nt} \mid \bar{\varepsilon}^{\Delta}, s\right] = \mu_t$, for all n and t, allowing endogenous period-specific shock means μ_t via Proposition 4. This assumption avoids the incidental parameters problem by considering shocks as-good-as-randomly assigned given the set of unobserved $\bar{\varepsilon}_{nt}^{\Delta}$, each of which is a function of the time-varying $\varepsilon_{\ell p}$ across all periods p.

An intuitive special case is when the exposure shares and importance weights are time-invariant: $s_{\ell nt} = s_{\ell n0}$ and $e_{\ell t} = e_{\ell 0}$. Then the weights in (A32) are also time-invariant, $s_{nt} = s_{n0}$, and

$$\begin{aligned} \bar{\varepsilon}_{nt}^{\Delta} &= \frac{\sum_\ell e_{\ell 0} s_{\ell n0} \varepsilon_{\ell t}^{\Delta}}{\sum_\ell e_\ell s_{\ell n0}} \\ &= \frac{\sum_\ell e_{\ell 0} s_{\ell n0} \left(\varepsilon_{\ell t} - \frac{1}{T}\sum_\tau \varepsilon_{\ell \tau}\right)}{\sum_\ell e_{\ell 0} s_{\ell n0}} \\ &= \bar{\varepsilon}_{nt} - \frac{1}{T}\sum_\tau \bar{\varepsilon}_{n\tau}, \end{aligned} \tag{A33}$$

where $\bar{\varepsilon}_{nt} = \frac{\sum_\ell e_{\ell 0} s_{\ell n0} \varepsilon_{\ell t}}{\sum_\ell e_{\ell 0} s_{\ell n0}}$ is an aggregate of period-specific unobservables $\varepsilon_{\ell t}$. It is then straightforward to extend Propositions 3 and 4 under a shock-level assumption of strong exogeneity, i.e. that $\mathbb{E}\left[g_{nt} \mid \bar{\varepsilon}, s\right] = \mu_n + \zeta_t$ for all n and t. Here endogenous n-specific shock means are permitted by the observation in Section 4.3, that share-weighted n-specific FEs at the shock level are subsumed by ℓ-specific FEs in the SSIV regression when shares and weights are time-invariant.

A.9 SSIV Relevance with Panel Data

This appendix shows that holding the exposure shares fixed in a pre-period is likely to weaken the SSIV first-stage in panel regressions. Consider a panel extension of the first stage model used in Section 3.1, where $x_{\ell t} = \sum_n s_{\ell nt} x_{\ell nt}$ with $x_{\ell nt} = \pi_{\ell nt} g_{nt} + \eta_{\ell nt}$, $\pi_{\ell n} \geq \bar{\pi}$ for some fixed $\bar{\pi} > 0$, and the g_{nt} are mutually independent and mean-zero with variance $\sigma_{nt}^2 \geq \bar{\sigma}_g^2$ for fixed $\sigma_g^2 > 0$, independently of $\{\eta_{\ell nt}\}_{\ell,n,t}$. As in other appendices, we here treat $s_{\ell nt}$, $e_{\ell t}$, and $\pi_{\ell nt}$ as non-stochastic. Then an SSIV regression with $z_{\ell t} = \sum_n s_{\ell nt}^* g_{nt}$ as an instrument, where $s_{\ell nt}^*$ is either $s_{\ell nt}$ (updated shares) or $s_{\ell n0}$ (fixed shares), yields a first-stage of

$$\mathbb{E}\left[\sum_\ell \sum_t e_{\ell t} z_{\ell t} x_{\ell t}\right] \geq \bar{\sigma}_g^2 \bar{\pi} \sum_\ell \sum_t e_{\ell t} \sum_n s_{\ell nt}^* s_{\ell tn}. \tag{A34}$$

For panel SSIV relevance we require the $e_{\ell t}$-weighted average of $\sum_n s^*_{\ell nt} s_{\ell nt}$ to not vanish asymptotically. With updated shares this is satisfied when the Herfindahl index of an average observation-period (across shocks) is non-vanishing, while in the fixed shares case the overlap of shares in periods 0 and t, $\sum_n s_{\ell n0} s_{\ell nt}$, may become weak or even vanish as $T \to \infty$, on average across observations.

A.10 SSIV with Multiple Endogenous Variables or Instruments

This appendix first generalizes our equivalence result to SSIV regressions with multiple endogenous variables and instruments, and discusses corresponding extensions of our quasi-experimental framework via the setting of Jaeger et al. (2018). We also describe how to construct the effective first-stage F-statistic of Montiel Olea and Pflueger (2013) for SSIV with one endogenous variable but multiple instruments. We then consider new shock-level IV procedures in this framework, which can be used for efficient estimation and specification testing. Finally, we illustrate these new procedures in the Autor et al. (2013) setting.

Generalized Equivalence and SSIV Consistency We consider a class of SSIV estimators of an outcome model with multiple treatment channels,

$$y_\ell = \beta' x_\ell + \gamma' w_\ell + \varepsilon_\ell, \tag{A35}$$

where $x_\ell = (x_{1\ell}, \ldots, x_{K\ell})'$ is instrumented by $z_\ell = (z_{1\ell}, \ldots, z_{J\ell})'$, for $z_{j\ell} = \sum_n s_{\ell n} g_{jn}$ and $J \geq K$, and observations are weighted by e_ℓ. Members of this class are parameterized by a (possibly stochastic) full-rank $K \times J$ matrix $\boldsymbol{c}$, which is used to combine the instruments into a vector of length J, $\boldsymbol{c} z_\ell$. For example the two-stage least squares (2SLS) estimator sets $\boldsymbol{c} = \boldsymbol{x}^{\perp\prime} \boldsymbol{e} \boldsymbol{z} (\boldsymbol{z}^{\perp\prime} \boldsymbol{e} \boldsymbol{z}^{\perp})^{-1}$, where $\boldsymbol{z}^\perp$ stacks observations of the residualized $z_\ell^{\perp\prime}$. IV estimates using a given combination are written as

$$\hat{\beta} = (\boldsymbol{c}\boldsymbol{z}'\boldsymbol{e}\boldsymbol{x}^\perp)^{-1} \boldsymbol{c}\boldsymbol{z}'\boldsymbol{e}\boldsymbol{y}^\perp, \tag{A36}$$

where $\boldsymbol{y}^\perp$ and $\boldsymbol{x}^\perp$ stack observations of the residualized $y_\ell^\perp$ and $x_\ell^{\perp\prime}$, $\boldsymbol{z}$ stacks observations of z_ℓ', and $\boldsymbol{e}$ is an $L \times L$ diagonal matrix of e_ℓ weights. In just-identified IV models (i.e. $J = K$) the two $\boldsymbol{c}$'s cancel in this expression and all IV estimators are equivalent. Note that while the shocks g_{jn} are different across the multiple instruments, we assume here that the exposure shares $s_{\ell n}$ are all the same.

As in Proposition 1, $\hat{\beta}$ can be equivalently obtained by a particular shock-level IV regression. Intuitively, when the shares are the same, $\boldsymbol{c} z_\ell$ also has a shift-share structure based on a linear combination of shocks $\boldsymbol{c} g_n$, and thus Proposition 1 extends. Formally, write $\boldsymbol{z} = \boldsymbol{s}\boldsymbol{g}$ where $\boldsymbol{s}$ is an

$L \times N$ matrix of exposure shares and $\boldsymbol{g}$ stacks observations of the shock vector g_n'; then,

$$\begin{aligned}\hat{\beta} &= (\boldsymbol{c}\boldsymbol{g}'\boldsymbol{s}'\boldsymbol{e}\boldsymbol{x}^{\perp})^{-1}\boldsymbol{c}\boldsymbol{g}'\boldsymbol{s}'\boldsymbol{e}\boldsymbol{y}^{\perp} \\ &= (\boldsymbol{c}\boldsymbol{g}'\boldsymbol{S}\bar{\boldsymbol{x}}^{\perp})^{-1}(\boldsymbol{c}\boldsymbol{g}'\boldsymbol{S}\bar{\boldsymbol{y}}^{\perp}), \end{aligned} \tag{A37}$$

where $\boldsymbol{S}$ is an $N \times N$ diagonal matrix with elements s_n, $\bar{\boldsymbol{x}}^{\perp}$ is an $N \times K$ matrix with elements $\bar{x}_{kn}^{\perp}$, and $\bar{\boldsymbol{y}}^{\perp}$ is an $N \times 1$ vector of $\bar{y}_n^{\perp}$. This is the formula for an s_n-weighted IV regression of $\bar{y}_n^{\perp}$ on $\bar{x}_{1n}^{\perp}, \ldots, \bar{x}_{Kn}^{\perp}$ with shocks as instruments, no constant, and the same $\boldsymbol{c}$ matrix. Furthermore, as in Proposition 1,

$$\iota'\boldsymbol{S}\bar{\boldsymbol{y}}^{\perp} = \sum_n s_n \bar{y}_n^{\perp} = \sum_{\ell} e_{\ell} \left(\sum_n s_{\ell n}\right) y_{\ell}^{\perp} = \sum_{\ell} e_{\ell} y_{\ell}^{\perp} = 0, \tag{A38}$$

and similarly for $\iota'\boldsymbol{S}\bar{\boldsymbol{x}}'$, where ι is a $N \times 1$ vector of ones. Therefore, the same estimate is obtained by including a constant in this IV procedure (and the same result holds including a shock-level control vector q_n provided $\sum_n s_{\ell n}$ has been included in w_{ℓ}, as in Proposition 5). The $\boldsymbol{c}$ matrix is again redundant in the just-identified case.

A natural generalization of the quasi-experimental framework of Section 3 follows. Rather than rederiving all of these results, we discuss them intuitively in the setting of Jaeger et al. (2018). Here y_{ℓ} denotes the growth rate of wages in region ℓ in a given period (residualized on Mincerian controls), $x_{1\ell}$ is the immigrant inflow rate in that period, and $x_{2\ell}$ is the previous period's immigration rate. The residual ε_{ℓ} captures changes to local productivity and other regional unobservables. Jaeger et al. (2018, Table 5) estimate this model with two "past settlement" instruments $z_{1\ell} = \sum_n s_{\ell n} g_{1n}$ and $z_{2\ell} = \sum_n s_{\ell n} g_{2n}$, where $s_{\ell n}$ is the share of immigrants from country of origin n in location ℓ at a previous reference date and $\boldsymbol{g}_n = (g_{1n}, g_{2n})'$ gives the current and previous period's national immigration rate from n. When this path of immigration shocks is as-good-as-randomly assigned with respect to the aggregated productivity shocks $\bar{\varepsilon}_n$ (satisfying a generalized Assumption 1), the $\boldsymbol{g}_n$ are uncorrelated across countries and $\mathbb{E}\left[\sum_n s_n^2\right] \to 0$ (satisfying a generalized Assumption 2), and appropriately generalized regularity conditions hold, the multiple-treatment shock orthogonality condition is satisfied: $\sum_n s_n g_{kn} \bar{\varepsilon}_n \xrightarrow{p} 0$ for each k. Then under the relevance condition from Proposition 2, again appropriately generalized, the SSIV estimates are consistent: $\hat{\beta} \xrightarrow{p} \beta$.

Effective First-Stage $\boldsymbol{F}$-statistics With one endogenous variable and multiple instruments, the Montiel Olea and Pflueger (2013) effective first-stage F-statistic provides a state-of-art heuristic for detecting a weak first-stage. Here we describe a correction to it for SSIV that generalizes the F-statistic in the single instrument case discussed in Section 5.2. The Stata command *weakssivtest*, provided with our replication archive, implements this correction.[7]

[7] Our package extends the *weakivtest* command developed by Pflueger and Wang (2015).

Consider a structural first stage with multiple instruments and one endogenous variable:

$$x_\ell = \pi' z_\ell + \rho w_\ell + \eta_\ell. \tag{A39}$$

Suppose each of the shocks satisfies Assumption 3, i.e. $\mathbb{E}\left[g_{jn} \mid \bar{\varepsilon}, q, s\right] = \mu_j' q_n$, where $\sum_n s_{\ell n} q_n$ is included in w_ℓ, and the residual shocks $g_{jn}^* = g_{jn} - \mu_j' q_n$ are independent from $\{\eta_\ell\}_\ell$. The Montiel Olea and Pflueger (2013) effective F-statistic for the 2SLS regression of y_ℓ on x_ℓ, instrumenting with $z_{1\ell}, \ldots, z_{J\ell}$, controlling for w_ℓ, and weighting by e_ℓ, is given by

$$F_{\text{eff}} = \frac{\left(\sum_\ell e_\ell x_\ell^\perp z_\ell^\perp\right)' \left(\sum_\ell e_\ell x_\ell^\perp z_\ell^\perp\right)}{\operatorname{tr}\left(\hat{V}\right)}, \tag{A40}$$

where $\hat{V}$ estimates $V = \operatorname{Var}\left[\sum_\ell e_\ell z_\ell^\perp \eta_\ell\right]$. Note that, as before, the first-stage covariance of the original SSIV regression equals that of the equivalent shock-level one from Proposition 5:

$$\sum_\ell e_\ell x_\ell^\perp z_\ell^\perp = \sum_\ell e_\ell x_\ell^\perp z_\ell = \sum_n s_n g_n \bar{x}_n^\perp = \sum_n s_n g_{n\perp} \bar{x}_n^\perp, \tag{A41}$$

where $g_{n\perp}$ is the residuals from an s_n-weighted projection of g_n on q_n, which consistently estimates g_n^*. A natural extension of Proposition 5 to many mutually-uncorrelated shocks further implies that V is well-approximated by

$$\hat{V} = \sum_n s_n^2 g_{n\perp} g_{n\perp}' \bar{\hat{\eta}}_n^2, \tag{A42}$$

where, per the discussion in Section 5.2, $\bar{\hat{\eta}}_n$ denotes the residuals from an IV regression of $\bar{x}_n^\perp$ on $\bar{z}_{1n}^\perp, \ldots, \bar{z}_{Jn}^\perp$, instrumented with $g_{1n}, \ldots, g_{Jn}$, weighted by s_n and controlling for q_n. Plugging this $\hat{V}$ into (A40) yields the corrected effective first-stage F-statistic.

Efficient Shift-Share GMM In overidentified settings ($J > K$), it is natural to consider which estimators are most efficient; for quasi-experimental SSIV, this can be answered by combining the asymptotic results of Adão et al. (2019) with the classic generalized methods of moments (GMM) theory of Hansen (1982). Here we show how standard shock-level IV procedures (such as 2SLS) may yield efficient coefficient estimates $\hat{\beta}^*$, depending on the variance structure of multiple quasi-randomly assigned shocks.

We first note that the equivalence result (A37) applies to SSIV-GMM estimators as well:

$$\begin{aligned}\hat{\beta} &= \arg\min_b \left(\boldsymbol{y}^\perp - \boldsymbol{x}^\perp b\right)' \boldsymbol{ezWz'e} \left(\boldsymbol{y}^\perp - \boldsymbol{x}^\perp b\right) \\ &= \arg\min_b \left(\bar{\boldsymbol{y}}^\perp - \bar{\boldsymbol{x}}^\perp b\right)' \boldsymbol{SgWg'S} \left(\bar{\boldsymbol{y}}^\perp - \bar{\boldsymbol{x}}^\perp b\right),\end{aligned} \tag{A43}$$

where $\boldsymbol{W}$ is an $J \times J$ moment-weighting matrix. This leads to an IV estimator with $\boldsymbol{c} = \bar{\boldsymbol{x}}^{\perp\prime} \boldsymbol{SgW}$.

For 2SLS estimation, for example, $\boldsymbol{W} = (\boldsymbol{z}^{\perp\prime}\boldsymbol{e}\boldsymbol{z}^{\perp})^{-1}$. Under appropriate regularity conditions, the efficient choice of $\boldsymbol{W}^*$ consistently estimates the inverse asymptotic variance of $\boldsymbol{z}'\boldsymbol{e}\left(\boldsymbol{y}^{\perp} - \boldsymbol{x}^{\perp}\beta\right) = \boldsymbol{g}'S\bar{\varepsilon} + o_p(1)$. Generalizations of results in Adão et al. (2019) can then be used to characterize this $\boldsymbol{W}^*$ when shocks are as-good-as-randomly assigned with respect to $\bar{\varepsilon}$. Given an estimate $\hat{\boldsymbol{W}}^*$, an efficient coefficient estimate $\hat{\beta}^*$ is given by shock-level IV regressions (A37) that set $\boldsymbol{c}^* = \bar{\boldsymbol{x}}^{\perp\prime}\boldsymbol{S}\boldsymbol{g}\hat{\boldsymbol{W}}^*$. A χ^2_{J-K} test statistic based on the minimized objective in (A43) can be used for specification testing.

As an example, suppose shocks are conditionally homoskedastic with the same variance-covariance matrix across n, $\text{Var}\left[\boldsymbol{g}_n \mid \bar{\varepsilon}, s\right] = \boldsymbol{G}$ for a constant $J \times J$ matrix $\boldsymbol{G}$. Then the optimal $\hat{\beta}^*$ is obtained by a shock-level 2SLS regression of $\bar{y}_n^{\perp}$ on all $\bar{x}_{kn}^{\perp}$ (instrumented by g_{jn} and weighted by s_n). We show this in the case of no controls (and mean-zero shocks) for notational simplicity. Then,

$$\begin{aligned}
\text{Var}\left[\boldsymbol{g}'S\left(\bar{\boldsymbol{y}}^{\perp} - \bar{\boldsymbol{x}}^{\perp}\beta\right)\right] &= \mathbb{E}\left[\bar{\varepsilon}'\boldsymbol{S}\boldsymbol{g}\boldsymbol{g}'\boldsymbol{S}\bar{\varepsilon}\right] \\
&= \text{tr}\left(\mathbb{E}\left[\bar{\varepsilon}'\boldsymbol{S}\boldsymbol{G}\boldsymbol{S}\bar{\varepsilon}\right]\right) \\
&= k\boldsymbol{G} \qquad \text{(A44)}
\end{aligned}$$

for $k = tr\left(\mathbb{E}\left[\boldsymbol{S}\bar{\varepsilon}\bar{\varepsilon}'\boldsymbol{S}\right]\right)$. The optimal weighting matrix thus should consistently estimate $\boldsymbol{G}$, which is satisfied by $\hat{\boldsymbol{G}} = \boldsymbol{g}'\boldsymbol{S}\boldsymbol{g}$. Under appropriate regularity conditions, a feasible optimal GMM estimate is thus given by

$$\begin{aligned}
\hat{\beta}^* &= (\bar{\boldsymbol{x}}^{\perp\prime}\boldsymbol{S}\boldsymbol{g}\hat{\boldsymbol{G}}^{-1}\boldsymbol{g}'\boldsymbol{S}\bar{\boldsymbol{x}}^{\perp})^{-1}(\bar{\boldsymbol{x}}^{\perp\prime}\boldsymbol{S}\boldsymbol{g}\hat{\boldsymbol{G}}^{-1}\boldsymbol{g}'\boldsymbol{S}\bar{\boldsymbol{y}}^{\perp}) \\
&= \left(\left(P_{\boldsymbol{g}}\bar{\boldsymbol{x}}^{\perp}\right)'\boldsymbol{S}\bar{\boldsymbol{x}}^{\perp}\right)^{-1}\left(P_{\boldsymbol{g}}\bar{\boldsymbol{x}}^{\perp}\right)'\boldsymbol{S}\bar{y}^{\perp}, \qquad \text{(A45)}
\end{aligned}$$

where $P_{\boldsymbol{g}} = \boldsymbol{g}(\boldsymbol{g}'\boldsymbol{S}\boldsymbol{g})^{-1}\boldsymbol{g}'\boldsymbol{S}$ is an s_n-weighted shock projection matrix. This is the formula for an s_n-weighted IV regression of $\bar{y}_n^{\perp}$ on the fitted values from projecting the $\bar{x}_{kn}^{\perp}$ on the shocks, corresponding to the 2SLS regression above. Straightforward extensions of this equivalence between optimally-weighted estimates of β and shock-level overidentified IV procedures follow in the case of heteroskedastic or clustered shocks, in which case the 2SLS estimator (A45) is replaced by the estimator of White (1982). We emphasize that these shock-level estimators are generally different than 2SLS or White (1982) estimators at the level of original observations, which are optimal under conditional homoskedasticity and independence assumptions placed on the residual ε_ℓ (assumptions which are generally violated in our quasi-experimental framework).

Many Shocks in Autor et al. (2013) Appendix Table C5 illustrates different shock-level overidentified IV estimators in the setting of Autor et al. (2013), introduced in Section 6.2.1. ADH construct their shift-share instrument based on the growth of Chinese imports in eight economies comparable to the U.S., together. We separate them to produce eight sets of industry shocks g_{jn}, $j = 1, \ldots, 8$, each reflecting the growth of Chinese imports in one of those countries. As in Section 6.2, the outcome of interest is a commuting zone's growth in total manufacturing employment with the single treatment variable measuring a commuting zone's local exposure to the growth of imports from China (see footnote 39 in the main text for precise variable definitions). The vector of controls coincides with that of column 3 of Table 4, isolating within-period variation in manufacturing shocks. Per Section 5.1, exposure-robust standard errors are obtained by controlling for period main effects in the shock-level IV procedures, and we report corrected first stage F-statistics constructed as detailed above.

Column 1 reports estimates of the ADH coefficient β using the industry-level two-stage least squares procedure (A45). At -0.238, this estimate it is very similar to the just-identified estimate in column 3 of Table 4. Column 2 shows that we also obtain a very similar coefficient of -0.247 with an industry-level limited information maximum likelihood (LIML) estimator. Finally, in column 3 we report a two-step optimal IV estimate of β using an industry-level implementation of the White (1982) estimator. Both the coefficient and standard error fall somewhat, with the latter consistent with the theoretical improvement in efficiency relative to columns 1 and 2. From this efficient estimate we obtain an omnibus overidentification test statistic of 10.92, distributed as chi-squared with seven degrees of freedom under the null of correct specification. This yields a p-value for the test of joint orthogonality of all eight ADH shocks of 0.142. Table C5 also reports the corrected effective first-stage F-statistic which measures the strength of the relationship between the endogenous variable and the eight shift-share instruments across regions. At 15.10 it is substantially lower than with one instrument in column 3 of Table 4 but still above the conventional heuristic threshold of 10.

A.11 Finite-Sample Performance of SSIV: Monte-Carlo Evidence

In this appendix we study the finite-sample performance of the SSIV estimator via Monte-Carlo simulation. We base this simulation on the data of Autor et al. (2013), as described in Section 6.2. For comparison, we also simulate more conventional shock-level IV estimators, similar to those used in Acemoglu et al. (2016), which also estimate the effects of import competition with China on U.S. employment. We begin by describing the design of these simulations and the benchmark Monte-Carlo results. We then explore how the simulation results change with various deviations from the benchmark: with different levels of industry concentration, different numbers of industries and regions, and with many shock instruments. Besides showing the general robustness of our framework, these extensions allow us to see how informative some conventional rules of thumb are on the finite-sample performance of shift-share estimators.[8]

[8] Naturally, these simulation results may be specific to the data-generating process we consider here, modeled after the "China shock" setting of Autor et al. (2013). In practice, we recommend that researchers perform similar simulations based on their data if they are concerned with the quality of asymptotic approximation—a suggestion that of course applies to conventional shock-level IV analyses as well.

Simulation design We base our benchmark data-generating process for SSIV on the specification in column 3 of Table 4. The outcome variable $y_{\ell t}$ corresponds to the change in manufacturing employment as a fraction of working-age population of region ℓ in period t, treatment $x_{\ell t}$ is a measure of regional import competition with China, and the shift-share instrument is constructed by combining the industry-level growth of China imports in eight developed economies, g_{nt}, with lagged regional employment weights of different industries $s_{\ell nt}$. We also include pre-treatment controls $w_{\ell t}$ as in column 3 of Table 4 and and estimate regressions with regional employment weights $e_{\ell t}$; see Section 6.2.1 for more detail on the Autor et al. (2013) setting.

In a first step we obtain an estimated SSIV second and first stage of

$$y_{\ell t} = \hat{\beta} x_{\ell t} + \hat{\gamma}' w_{\ell t} + \hat{\varepsilon}_{\ell t}, \tag{A46}$$

$$x_{\ell t} = \hat{\pi} z_{\ell t} + \hat{\rho}' w_{\ell t} + \hat{u}_{\ell t}. \tag{A47}$$

We then generate 10,000 simulated samples by drawing shocks g^*_{nt}, as detailed below, and constructing the simulated shift-share instrument $z^*_{\ell t} = \sum_n s_{\ell nt} g^*_{nt}$ and treatment $x^*_{\ell t} = \hat{\pi} z^*_{\ell t} + \hat{u}_{\ell t}$. Imposing a true causal effect of $\beta^* = 0$, we use the same $y^*_{\ell t} \equiv \hat{\varepsilon}_{\ell t}$ as the outcome in each simulation (note that it is immaterial whether we include $\hat{\pi}' w_{\ell t}$ and $\hat{\rho}' w_{\ell t}$, since all our specifications control for $w_{\ell t}$). By keeping $\hat{\varepsilon}_{\ell t}$ and $\hat{u}_{\ell t}$ fixed, we study the finite sample properties of the estimator that arises from the randomness of shocks, which is the basis of the inferential framework of Adão et al. (2019); we also avoid having to take a stand on the joint data generating process of $(\varepsilon_{\ell t}, u_{\ell t})$, which this inference framework does not restrict.

We estimate SSIV specifications that parallel (A46)-(A47) from the simulated data

$$y^*_{\ell t} = \beta^* x^*_{\ell t} + \gamma^{*\prime} w_{\ell t} + \varepsilon^*_{\ell t}, \tag{A48}$$

$$x^*_{\ell t} = \pi^* z^*_{\ell t} + \rho^{*\prime} w_{\ell t} + u^*_{\ell t}. \tag{A49}$$

using the original weights $e_{\ell t}$ and controls $w_{\ell t}$. We then test the (true) hypothesis $\beta^* = 0$ using either the heteroskedasticity-robust standard errors from the equivalent industry-level regression or their version with the null imposed, as in Section 5.1.[9] As in column 3 of Table 4, we control for period indicators as q_{nt} in the industry-level regression.

Our comparison estimator is a conventional industry-level IV inspired by Acemoglu et al. (2016). However, we try to keep the IV regression as similar to the SSIV as possible, thus diverging from Acemoglu et al. (2016) in some details. Specifically, the outcome y_{nt} is the industry employment growth as measured by these authors. It is defined for 392 out of the 397 industries in Autor et

[9]Note that there is no need for clustering since we generate the shocks independently across industries in all simulations. We have verified, however, that allowing for correlation in shocks within industry groups and using clustered standard errors yields similar results.

al. (2013), so we drop the remaining five industries in each period. The endogenous regressor $x_{nt} \equiv g_{nt}^{US}$ (growth of U.S. imports from China per worker) and the instrument g_{nt} (growth of China imports into eight developed economies) are those from which we built the shift-share endogenous regressor and treatment, respectively (see main text footnote 39). Construction of those variables differ from Acemoglu et al. (2016) who measure imports relative to domestic absorption rather than employment. We also follow our SSIV analysis in using period indicators as the only industry-level control variables q_{nt} and taking identical regression importance weights s_{nt}.

The Monte-Carlo strategy for the conventional shock-level IV parallels the one for SSIV; we obtain an estimated industry-level second and first stage of

$$y_{nt} = \hat{\beta}_{\text{ind}} x_{nt} + \hat{\gamma}' q_{nt} + \hat{\varepsilon}_{nt}, \tag{A50}$$

$$x_{nt} = \hat{\pi}_{\text{ind}} g_{nt} + \hat{\rho}' q_{nt} + \hat{u}_{nt}. \tag{A51}$$

using the s_{nt} importance weights. We then perform 10,000 simulations where we regenerate shocks g_{nt}^* and regress $y_{nt}^* = \hat{\varepsilon}_{nt}$ (consistent with a true causal effect of $\beta^{\text{ind}} = 0$, given that we control for q_{nt}) on $x_{nt}^* = \hat{\pi}_{\text{ind}} g_{nt}^* + \hat{u}_{nt}$, instrumenting by g_{nt}^*, controlling for q_{nt}, and weighting by s_{nt}. We test $\beta_{\text{ind}} = 0$ by using robust standard errors in this IV regression or the version with the null imposed, which corresponds to a standard Lagrange Multiplier test for this true null hypothesis.

In both simulations we report the rejection rate of nominal 5% level tests for $\beta = 0$ and $\beta_{\text{ind}} = 0$ to gauge the quality of each asymptotic approximation. We do not report the bias of the estimators because they are all approximately unbiased (more precisely, the simulated median bias is at most 1% of the estimator's standard deviation). However we return to the question of bias at the end of the section, where we extend the analysis to having many instruments with a weak first stage.

Main results Table C7 reports the rejection rates for shift-share IV (columns 1 and 2) and conventional industry-level IV (columns 3 and 4) in various simulations. Specifically, column 1 corresponds to using exposure-robust standard errors from the equivalent industry-level IV, and column 2 implements the version with the null hypothesis imposed. Columns 3 and 4 parallel columns 1 and 2 when applied to conventional IV: the former uses heteroskedasticity-robust standard errors and the latter tests $\beta_{\text{ind}} = 0$ with the null imposed, which amounts to using the Lagrange multiplier test.

The simulations in Panel A vary the data-generating process of the shocks. Following Adão et al. (2019) in row (a) we draw the shocks *iid* from a normal distribution with the variance matched to the sample variance of the shocks in the data after de-meaning by year. The rejection rate is close to the nominal rate of 5% for both SSIV and conventional IV (7.6% and 6.8%, respectively), and in both cases it becomes even closer when the null is imposed (5.2% and 5.0%).

This simulation may not approximate the data-generating process well because of heteroskedasticity: smaller industries have more volatile shocks.[10] To match unrestricted heteroskedasticity, in row (b) we use wild bootstrap, generating $g^*_{nt} = g_{nt}\nu^*_{nt}$ by multiplying the year-demeaned observed shocks g_{nt} by $\nu^*_{nt} \overset{iid}{\sim} \mathcal{N}(0,1)$ (Liu 1988). This approach also provides a better approximation for the marginal distribution of shocks than the normality assumption. Here the relative performance of SSIV is even better: the rejection rate is 8.0% vs. 14.2% for conventional IV.

We now depart from the row (b) simulation in several directions, as a case study for the sensitivity of the asymptotic approximation to different features of the SSIV setup. Specifically, we study the role of the Herfindahl concentration index across industries, the number of regions and industries, and the many weak instrument bias. We uniformly find that the performance of the SSIV estimator is similar to that of industry-level IV. Our results also suggest that the Herfindahl index is a useful statistic for measuring the effective number of industries in SSIV, and the first-stage F-statistic is informative about the weak instrument bias, as usual.

The Role of Industry Concentration Since Assumption 2 requires small concentration of industry importance weights, measured using the Herfindahl index $\sum_{n,t} s^2_{nt} / \left(\sum_{n,t} s_{nt}\right)^2$, Panel B of Table C7 studies how increasing the skewness of s_{nt} towards the bigger industries affects coverage of the tests.[11] For conventional IV this simply amounts to reweighting the regression. Specifically, for a parameter $\alpha > 1$, we use weights

$$\tilde{s}_{nt} = s^{\alpha}_{nt} \cdot \frac{\sum_{n',t'} s_{n't'}}{\sum_{n',t'} s^{\alpha}_{n't'}}.$$

We choose the unique α to match the target level of $\widetilde{HHI}$ by solving, numerically,

$$\frac{\sum_{n,t} \left(\tilde{s}_{nt}\right)^2}{\left(\sum_{n,t} \tilde{s}_{nt}\right)^2} = \widetilde{HHI}. \tag{A52}$$

Matching the Herfindahl index in SSIV is more complicated since we need to choose how exactly to amend shares $\tilde{s}_{\ell nt}$ and regional weights $\tilde{e}_{\ell t}$ that would yield $\tilde{s}_{nt}$ from (A52). We proceed as follows: we consider the lagged level of manufacturing employment by industry $E_{\ell nt} = e_{\ell t} s_{\ell nt}$ and the total regional non-manufacturing employment $E_{\ell 0t} = e_{\ell t}\left(1 - \sum_n s_{\ell nt}\right)$.[12] We then define $\tilde{E}_{\ell nt} = E_{\ell nt} \cdot \frac{\tilde{s}_{nt}}{s_{nt}}$ for manufacturing industries (and leave non-manufacturing employment unchanged, $\tilde{E}_{\ell 0t} = E_{\ell 0t}$). This increases employment in large manufacturing industries proportionately in all regions, while

[10] This is established by unreported regressions of $|g_{nt}|$ on s_{nt}, for year-demeaned g_{nt} from ADH, with or without weights. The negative relationship is significant at conventional levels.

[11] Note that in ADH $\sum_n s_{\ell nt}$ equals the lagged share of regional manufacturing employment, which is below one. We thus renormalize the shares when computing the Herfindahl.

[12] The interpretation of $E_{\ell nt}$ as the lagged level is approximate since $e_{\ell t}$ is measured at the beginning of period in ADH, while $s_{\ell nt}$ is lagged.

reducing it in smaller ones. We then recompute shares $\tilde{s}_{\ell nt}$ and weights $\tilde{e}_{\ell t}$ accordingly:

$$\tilde{e}_{\ell t} = \sum_{n=0}^{N} \sum_{t} \tilde{E}_{\ell nt},$$
$$\tilde{s}_{\ell nt} = \frac{\tilde{E}_{\ell nt}}{\tilde{e}_{\ell t}}.$$

Rows (c)–(e) of Table C7 Panel B implement this procedure for target Herfindahl levels of 1/50, 1/20, and 1/10, respectively. For comparison, the Herfindahl in the actual ADH data is 1/191.6 (Table 1, column 2). The table finds that even with the Herfindahl index of 1/20 (corresponding to the "effective" number of shocks of 20 in both periods total) the rejection rate is still around 7%, a level that may be considered satisfactory. It also shows that the rejection rate grows when the Herfindahl is even higher, at 1/10, suggesting that the Herfindahl can be used as an indicative rule of thumb. More importantly, the rejection rates are similar for SSIV and conventional industry-level IV, as before.

Varying the Number of Industries and Regions The asymptotic sequence we consider in Section 3.1 relies on both N and L growing. Here we study how the quality of the asymptotic approximation depends on these parameters.

First, to consider the case of small N, we aggregate industries in a natural way: from 397 four-digit manufacturing SIC industries into 136 three-digit ones and further into 20 two-digit ones and reconstruct the endogenous right-hand side variable and the instrument using aggregated data.[13] Rows (f) and (g) of Table C7 Panel C report simulation results based on the aggregated data. They show that rejection rates are similar to the case of detailed industries, and between SSIV and conventional IV. This does not mean that disaggregated data are not useful: the dispersion of the simulated distribution (not reported) increases with industry aggregation, reducing test power. However, standard errors correctly reflect this variability, resulting in largely unchanged test coverage rates.

Second, to study the implications of having fewer regions L, we select a random subset of them in each simulation. The results are presented in Rows (h) and (i) of Panel C for $L = 100$ and 25, compared to the original $L = 722$, respectively.[14] They show once again that rejection rates are not significantly affected (even though unreported standard errors expectedly increase).

Many Weak Instruments In this final simulation we return to the question of SSIV bias. Since our previous simulations confirm that just-identified SSIV is median-unbiased, we turn to the case of

[13] Specifically, we aggregate imports from China to the U.S. and either developed economies as well as the number of U.S. workers by manufacturing industry to construct the new g_{nt} and g_{nt}^{US}. We then aggregate the shares $s_{\ell nt}$ and $s_{\ell nt}^{\text{current}}$ to construct $x_{\ell t}$ and $z_{\ell t}$ (see main text footnote 39 for formulas). We do not change the regional outcome, controls, or importance weights. For conventional IV, we additionally reconstruct the outcome (industry employment growth) by aggregating employment levels by year in the Acemoglu et al. (2016) data and measuring growth according to their formulas.

[14] When we select regions, we always keep observations from both periods for each selected region. We keep the second- and first-stage coefficients from the full sample to focus on the noise that arises from shock randomness.

multiple instruments. We show that the problem of many weak instruments is similar between SSIV and conventional IV, and that first-stage F-statistics, when properly constructed, can serve as useful heuristics.

For clarity, we begin by describing the procedure for the conventional shock-level IV that is a small departure from Column 3 of Table C7. For a given number of instruments $J \geq 1$, in each simulation we generate g^*_{jnt}, $j = 1, \ldots, J$, independently across j using wild bootstrap (as in Table C7 Row (b)).[15] We make only the first instrument relevant by setting $x^*_{nt} = \hat{\pi}_{\text{ind}} g^*_{1nt} + \sum_{j=2}^{J} 0 \cdot g^*_{jnt} + \hat{u}_{nt}$. We then estimate the IV regression of $y^*_{nt} \equiv \hat{\varepsilon}_{nt}$ on x^*_{nt}, instrumenting with $g^*_{1nt}, \ldots, g^*_{Jnt}$, controlling for q_{nt}, and weighting by s_{nt}. We use robust standard errors and compute the effective first-stage F-statistic using the Montiel Olea and Pflueger (2013) method.

The procedure for SSIV is more complex but as usual parallels the one for the conventional shock-level IV as much as possible. Given simulated shocks g^*_{jnt}, we construct shift-share instruments $z^*_{j\ell t} = \sum_{\ell} s_{\ell nt} g^*_{jnt}$ and make only the first of them relevant, $x^*_{\ell t} = \hat{\pi} z^*_{1\ell t} + \sum_{j=2}^{J} 0 \cdot z^*_{jnt} + \hat{u}_{\ell t}$. Since the equivalence result from Section 2.3 need not hold for overidentified SSIV, we rely on the results in Appendix A.10: we estimate β^* from the industry-level regression of $\bar{y}^{*\perp}_{nt}$ (based on $y^*_{\ell t} = \hat{\varepsilon}_{\ell t}$ as before) on $\bar{x}^{*\perp}_{nt}$ by 2SLS, instrumenting by $g^*_{1nt}, \ldots, g^*_{Jnt}$, controlling for q_{nt} and weighting by s_{nt}. We compute robust standard errors from this regression to test $\beta^* = 0$. For effective first-stage F-statistics, we follow the procedure described in Appendix A.10 and implemented via our *weakssivtest* command in Stata.

Table C8 reports the result for $J = 1, 5, 10, 25$, and 50, presenting the rejection rate corresponding to the 5% nominal, the median bias as a percentage of the simulated standard deviation, and the median first-stage F-statistic. Panel A corresponds to SSIV and Panel B to the conventional shock-level IV. For higher comparability, we adjust the first-stage coefficient $\hat{\pi}_{\text{ind}}$ in the latter in order to make the F-statistics approximately match between the two panels. We find that the median bias is now non-trivial and grows with J, at the same time as the F-statistic declines. However, the level of bias is similar for the two estimators. The rejection rates tend to be higher for conventional IV than SSIV, although they converge as J grows.

[15] For computational reasons we perform only $15,000/J$ simulations when $J > 1$ (but 10,000 for $J = 1$ as before).

B Appendix Proofs

B.1 Proposition 4 and Extensions

This section proves Proposition 4 and extensions that allow for certain forms of mutual shock dependence (Assumptions 5 and 6). Proposition 3 is obtained as a special case, where $q_n = 1$. In addition to Assumptions 3 and 4 and the relevance condition of $\sum_\ell e_\ell z_\ell x_\ell^\perp \xrightarrow{p} \pi$ with $\pi \neq 0$, the proof of Proposition 4 uses two regularity conditions:

Assumption B1: $\mathbb{E}\left[\tilde{g}_n^2 \mid \bar{\varepsilon}, q, s\right]$ and $\mathbb{E}\left[\bar{\varepsilon}_n^2 \mid s\right]$ are uniformly bounded by some fixed B_g and B_ε.

Assumption B2: $\left\|\sum_\ell e_\ell w_\ell \varepsilon_\ell\right\|_1 = o_p(1)$, $\max\left|\left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1}\right| = O_p(1)$, and $\max\left|\sum_\ell e_\ell w_\ell z_\ell\right| = O_p(1)$.

The first of these is a weak condition on the second moments of shocks and shock-level unobservables which we show below permits a shock-level law of large numbers. The second condition ensures the consistency of the IV estimate of the control coefficient, $\hat{\gamma} = \left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1} \sum_\ell e_\ell w_\ell \epsilon_\ell = \gamma + \left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1} \sum_\ell e_\ell w_\ell \varepsilon_\ell$ (see footnote 5 in the main text), and stochastic boundedness of the weighted average $\sum_\ell e_\ell w_{\ell m} z_\ell$, while generally allowing the length of the control vector to increase with L. We discuss low-level conditions for the consistency of $\hat{\gamma}$ in Appendix A.5.

To prove Proposition 4, we first note that under Assumption B2,

$$\begin{aligned}
\sum_n s_n g_n \bar{\varepsilon}_n - \sum_n s_n g_n \bar{\varepsilon}_n^\perp &= \sum_\ell e_\ell z_\ell \left(\varepsilon_\ell - \varepsilon_\ell^\perp\right) \\
&= \left(\sum_\ell e_\ell z_\ell w_\ell'\right)(\hat{\gamma} - \gamma) \\
&= \left(\sum_\ell e_\ell z_\ell w_\ell'\right)\left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1} \sum_\ell e_\ell w_\ell \varepsilon_\ell \xrightarrow{p} 0, \qquad \text{(B1)}
\end{aligned}$$

so that, when the relevance condition holds,

$$\begin{aligned}
\hat{\beta} - \beta &= \frac{\sum_n s_n g_n \bar{\varepsilon}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp} \\
&= \pi^{-1} \sum_n s_n g_n \bar{\varepsilon}_n \left(1 + o_p(1)\right). \qquad \text{(B2)}
\end{aligned}$$

Furthermore, since $\sum_n s_{\ell n} = 1$, we also have under Assumption B2 that

$$\sum_n s_n q_n' \mu \bar{\varepsilon}_n = \left(\sum_\ell e_\ell \tilde{w}_\ell \varepsilon_\ell\right)' \mu \xrightarrow{p} 0. \qquad \text{(B3)}$$

Thus

$$\sum_n s_n g_n \bar{\varepsilon}_n = \sum_n s_n \tilde{g}_n \bar{\varepsilon}_n + o_p(1), \tag{B4}$$

with

$$\mathbb{E}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] = 0 \tag{B5}$$

under Assumption 3.

To prove consistency of $\hat{\beta}$, it remains to show that $\operatorname{Var}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] \to 0$. Since

$$\mathbb{E}\left[\tilde{g}_n \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right] = \operatorname{Cov}\left[\tilde{g}_n, \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right] = 0 \tag{B6}$$

under Assumptions 3 and 4,

$$\begin{aligned}
\operatorname{Var}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] &= \mathbb{E}\left[\left(\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right)^2\right] \\
&= \sum_n \sum_{n'} \mathbb{E}\left[s_n s_{n'} \tilde{g}_n \tilde{g}_{n'} \bar{\varepsilon}_n \bar{\varepsilon}_{n'}\right] \\
&= \sum_n \mathbb{E}\left[s_n^2 \mathbb{E}\left[\mathbb{E}\left[\tilde{g}_n^2 \mid \bar{\varepsilon}, q, s\right] \bar{\varepsilon}_n^2 \mid s\right]\right].
\end{aligned} \tag{B7}$$

Then, by Assumption B1 and the Cauchy-Schwartz inequality:

$$\operatorname{Var}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] \le B_g B_\varepsilon \mathbb{E}\left[\sum_n s_n^2\right] \to 0. \tag{B8}$$

Extensions Similar steps establish equation (B8) when Assumption 4 is replaced by either Assumption 5 or 6. Under Assumption 5 we have, for $N(c) = \{n \colon c(n) = c\}$,

$$\begin{aligned}
\operatorname{Var}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] &= \mathbb{E}\left[\left(\sum_c \sum_{n \in N(c)} s_n \tilde{g}_n \bar{\varepsilon}_n\right)^2\right] \\
&= \mathbb{E}\left[\sum_c s_c^2 \mathbb{E}\left[\left(\sum_{n \in N(c)} \frac{s_n}{s_c} \tilde{g}_n \bar{\varepsilon}_n\right)^2 \mid s\right]\right] \\
&= \mathbb{E}\left[\sum_c s_c^2 \sum_{n, n' \in N(c)} \frac{s_n}{s_c} \frac{s_{n'}}{s_c} \mathbb{E}\left[\tilde{g}_n \tilde{g}_{n'} \bar{\varepsilon}_n \bar{\varepsilon}_{n'} \mid s\right]\right] \\
&\le B_g B_\varepsilon \mathbb{E}\left[\sum_c s_c^2\right] \to 0.
\end{aligned} \tag{B9}$$

Here the last line used Assumption B1 and the Cauchy-Schwartz inequality twice: to establish, for $n, n' \in N(c)$,

$$\begin{aligned} \mathbb{E}\left[\tilde{g}_n \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right] &\leq \sqrt{\mathbb{E}\left[\tilde{g}_n \mid \bar{\varepsilon}, q, s\right] \mathbb{E}\left[\tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right]} \\ &\leq B_g \end{aligned} \tag{B10}$$

and

$$\begin{aligned} \mathbb{E}\left[|\bar{\varepsilon}_n| \, |\bar{\varepsilon}_{n'}| \mid s_c\right] &\leq \sqrt{\mathbb{E}\left[\bar{\varepsilon}_n^2 \mid s\right] \mathbb{E}\left[\bar{\varepsilon}_{n'}^2 \mid s\right]} \\ &\leq B_\varepsilon. \end{aligned} \tag{B11}$$

If we instead replace Assumption 4 with Assumption 6, we have

$$\begin{aligned} \mathrm{Var}\left[\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right] &= \mathbb{E}\left[\left(\sum_n s_n \tilde{g}_n \bar{\varepsilon}_n\right)^2\right] \\ &= \sum_n \sum_{n'} \mathbb{E}\left[s_n s_{n'} \mathbb{E}\left[\tilde{g}_n \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right] \bar{\varepsilon}_n \bar{\varepsilon}_{n'}\right] \\ &\leq B_L \sum_n \sum_{n'} f\left(|n' - n|\right) \mathbb{E}\left[|s_n \bar{\varepsilon}_n s_{n'} \bar{\varepsilon}_{n'}|\right] \\ &= B_L \left(\sum_n \mathbb{E}\left[(s_n \bar{\varepsilon}_n)^2\right] f(0) + 2 \sum_{r=1}^{N-1} \sum_{n=1}^{N-r} \mathbb{E}\left[|s_{n+r} \bar{\varepsilon}_{n+r}| \cdot |s_n \bar{\varepsilon}_n|\right] f(r)\right) \\ &\leq \left(B_L \sum_n \mathbb{E}\left[s_n^2 \mathbb{E}\left[\bar{\varepsilon}_n^2 \mid s\right]\right]\right) \left(f(0) + 2 \sum_{r=1}^{N-1} f(r)\right) \\ &\leq B_\varepsilon \left(f(0) + 2 \sum_{r=1}^{N-1} f(r)\right) \left(B_L \mathbb{E}\left[\sum_n s_n^2\right]\right) \to 0, \end{aligned} \tag{B12}$$

using $\mathbb{E}\left[\bar{\varepsilon}_n^2 \mid s_n\right] < B_\varepsilon$ in the last line. Here the second-to-last line follows because for any sequence of numbers $a_1, \ldots, a_N$ and any $r > 0$,

$$\begin{aligned} \sum_n a_n^2 &\geq \frac{1}{2} \left(\sum_{n=1}^{N-r} a_n^2 + \sum_{n=1}^{N-r} a_{n+r}^2\right) \\ &= \frac{1}{2} \sum_{n=1}^{N-r} \left(a_n - a_{n+r}\right)^2 + \sum_{n=1}^{N-r} a_n a_{n+r} \\ &\geq \sum_{n=1}^{N-r} a_n a_{n+r}, \end{aligned} \tag{B13}$$

and the same is true in expectation if $a_n = |s_n \bar{\varepsilon}_n|$ are random variables. We note that allowing B_L to grow in the asymptotic sequence imposes much weaker conditions on the correlation structure of shocks. For example, with shock importance weights s_n approximately equal, i.e. $\sum_n s_n^2 = O_p\left(1/N\right)$, it is enough to have $|\mathrm{Cov}\left[\tilde{g}_n, \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s\right]| \leq B_1/N^{\alpha}$ for any $\alpha > 0$: in this case one can satisfy Assumption 6 by setting $B_L = B_1 N^{1-\alpha/2}$ and $f(r) = r^{-1-\alpha/2}$.

B.2 Proposition 5 and Related Results

This section proves Proposition 5 and then establishes several additional results mentioned in Section 5.1. First, we show the heteroskedasticity-robust standard error from estimating equation (10) is numerically equivalent to the baseline IV standard error of Adão et al. (2019) when w_ℓ contains only a constant. Second, we show that when Assumption B4 on the structure of controls is relaxed, the standard errors from Proposition 5 are conservative. We also discuss the likely difference between our standard error estimates and those of Adão et al. (2019) when Assumption B4 holds. Finally, we show how the alternative null-imposed inference procedure of Adão et al. (2019) is also conveniently obtained from our equivalent shock-level regression.

We prove Proposition 5 under additional assumptions that largely follow Adão et al. (2019):

Assumption B3: The first stage satisfies $x_\ell = \sum_n s_{\ell n} \pi_{\ell n} g_n + \eta_\ell$, for all ℓ.

Assumption B4: The control vector can be partitioned as $w_\ell = [\tilde{w}_\ell', u_\ell']'$, for $\tilde{w}_\ell = \sum_n s_{\ell n} q_n$. The vector q_n captures all sources of shock confounding: $\mathbb{E}\left[g_n \mid \mathcal{I}_L\right] = q_n'\mu$, for all n and $\mathcal{I}_L = \left\{\{q_n\}_n, \{u_\ell, \epsilon_\ell, \eta_\ell, \{s_{\ell n}, \pi_{\ell n}\}_n, e_\ell\}_\ell\right\}$.

Assumption B5: The g_n are mutually independent given $\mathcal{I}_L$, $\max_n s_n \to 0$, and $\max_n \frac{s_n^2}{\sum_{n'} s_{n'}^2} \to 0$.

Assumption B6: $\mathbb{E}\left[|g_n|^{4+v} \mid \mathcal{I}_L\right]$ is uniformly bounded for some $v > 0$ and $\sum_\ell e_\ell \sum_n s_{\ell n}^2 \mathrm{Var}\left[g_n \mid \mathcal{I}_L\right] \pi_{\ell n} \neq 0$ almost surely. The support of $\pi_{\ell n}$ is bounded, the fourth moments of ϵ_ℓ, η_ℓ, u_ℓ, q_n, and $\tilde{g}_n$ exist and are uniformly bounded, $\sum_\ell e_\ell w_\ell w_\ell' \xrightarrow{p} \Omega_{ww}$ for positive definite Ω_{ww}, and $\sum s_n q_n q_n' \xrightarrow{p} \Omega_{qq}$ for positive definite Ω_{qq}. The control vector γ is consistently estimated by $\hat{\gamma} = \left(\sum_\ell e_\ell w_\ell w_\ell'\right)^{-1} \sum_\ell e_\ell w_\ell \epsilon_\ell$.

We note that Assumption B5 both strengthens our baseline Herfindahl index condition in Assumption 4 and implicitly treats the set of s_n as non-stochastic, following Assumption 2 of Adão et al. (2019). The regularity condition B6 includes the relevant conditions from Assumptions 4 and A.3 of Adão et al. (2019). These assumptions strengthen those of Proposition 4: Assumptions B3–B6 imply our Assumptions 3, 4, B1, and B2. Relative to Adão et al. (2019), we do not impose that $L > N$ or that the shares are non-collinear.

To establish the equivalence of IV coefficients in Proposition 5, note that when $\sum_n s_{\ell n} q_n$ is included in w_ℓ

$$\sum_n s_n q_n \bar{y}_n^\perp = \sum_\ell e_\ell y_\ell^\perp \left(\sum_n s_{\ell n} q_n\right) = 0 \tag{B14}$$

and similarly for $\sum_n s_n q_n \bar{x}_n^\perp$. The s_n-weighted regression of $\bar{y}_n^\perp$ and $\bar{x}_n^\perp$ on q_n thus produces a coefficient vector that is numerically zero, implying the s_n-weighted and g_n-instrumented regression of $\bar{y}_n^\perp$ on $\bar{x}_n^\perp$ is unchanged with the addition of q_n controls. Proposition 1 shows that the IV coefficient from this regression is equivalent to the SSIV estimate $\hat{\beta}$.

To establish validity of the standard errors, note that the conventional heteroskedasticity-robust standard error from for the s_n-weighted shock-level IV regression of $\bar{y}_n^\perp$ on $\bar{x}_n^\perp$ and q_n, instrumented by g_n, is given by

$$\widehat{se}_{\text{equiv}} = \frac{\sqrt{\sum_n s_n^2 \hat{\varepsilon}_n^2 \hat{g}_n^2}}{\left|\sum_n s_n \bar{x}_n^\perp g_n\right|}, \tag{B15}$$

where $\hat{\varepsilon}_n = \bar{y}_n^\perp - \hat{\beta}\bar{x}_n^\perp$ is the estimated shock-level regression residual (where we used the fact that the estimated coefficients on q_n in that regression are numerically zero) and $\hat{g}_n = g_n - \hat{\mu} q_n$, where $\hat{\mu} = \left(\sum_n s_n q_n q_n'\right)^{-1} \sum_n s_n q_n g_n$, is the residual from a projection of the instrument in equation (10) on the control vector q_n. By Proposition 1, $\hat{\varepsilon}_n$ coincides with the share-weighted aggregate of the SSIV estimated residuals $\hat{\varepsilon}_\ell = y_\ell^\perp - \hat{\beta} x_\ell^\perp$:

$$\hat{\varepsilon}_n = \frac{\sum_\ell e_\ell s_{\ell n} y_\ell^\perp}{\sum_\ell e_\ell s_{\ell n}} - \hat{\beta} \cdot \frac{\sum_\ell e_\ell s_{\ell n} x_\ell^\perp}{\sum_\ell e_\ell s_{\ell n}} = \frac{\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell}{\sum_\ell e_\ell s_{\ell n}}. \tag{B16}$$

The squared numerator of (B15) can thus be rewritten

$$\sum_n s_n^2 \hat{\varepsilon}_n^2 \hat{g}_n^2 = \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2. \tag{B17}$$

The expression in the denominator of (B15) estimates the magnitude of the shock-level first-stage covariance, which matches the e_ℓ-weighted sample covariance of x_ℓ and z_ℓ:

$$\sum_n s_n \bar{x}_n^\perp g_n = \sum_n \left(\sum_\ell e_\ell s_{\ell n} x_\ell^\perp\right) g_n = \sum_\ell e_\ell x_\ell^\perp z_\ell. \tag{B18}$$

Thus

$$\widehat{se}_{\text{equiv}} = \frac{\sqrt{\sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2}}{\left|\sum_\ell e_\ell x_\ell^\perp z_\ell\right|}. \tag{B19}$$

We now compare this expression to the standard error formula from Adão et al. (2019), incorporating the e_ℓ importance weights. Equation (39) in their paper yields

$$\widehat{se}_{\text{AKM}} = \frac{\sqrt{\sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \ddot{g}_n^2}}{\left|\sum_\ell e_\ell x_\ell^\perp z_\ell\right|}, \tag{B20}$$

where $\ddot{g}_n$ denotes the coefficients from regressing the residualized instrument $z_\ell^\perp$ on all shares $s_{\ell n}$, without a constant; note that to compute this requires $L > N$ and that the matrix of exposure shares $s_{\ell n}$ is full rank. The formulas for $\widehat{se}_{\text{equiv}}$ and $\widehat{se}_{\text{AKM}}$ therefore differ only in the construction of shock residuals, $\hat{g}_n$ versus $\ddot{g}_n$.

We establish the general asymptotic equivalence of $\widehat{se}^2_{\text{equiv}}$ and $\widehat{se}^2_{\text{AKM}}$, and thus the asymptotic validity of $\widehat{se}_{\text{equiv}}$, by showing that both capture the conditional asymptotic variance of $\hat{\beta}$ given $\mathcal{I}_L$ under Assumptions B3-B6. Both of the resulting confidence intervals are then asymptotically valid unconditionally, since if $Pr(\beta \in \widehat{CI} \mid \mathcal{I}_L) = \alpha$ then $Pr(\beta \in \widehat{CI}) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}[\beta \in \widehat{CI} \mid \mathcal{I}_L\right]\right] = \alpha$ by the law of iterated expectations. Under Assumptions B3-B6, Proposition A.1 of Adão et al. (2019) applies and shows that

$$\sqrt{r_L}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathcal{V}}{\pi^2}\right) \tag{B21}$$

for $\mathcal{V} = \text{plim}_{L\to\infty} r_L \mathcal{V}_L$, where $r_L = 1/\left(\sum_n s_n^2\right)$ and $\mathcal{V}_L = \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell\right)^2 \text{Var}\left[g_n \mid \mathcal{I}_L\right]$, provided such a limit exists. To establish the asymptotic validity of $\widehat{se}_{\text{AKM}}$, i.e. that $r_L \left(\sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \ddot{g}_n^2 - \mathcal{V}_L\right) \xrightarrow{p} 0$, Adão et al. (2019) further assume that $L \geq N$, the matrix of $s_{\ell n}$ is always full rank, and additional regularity conditions (see their Proposition 5). We establish $r_L \left(\sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2 - \mathcal{V}_L\right) \xrightarrow{p} 0$, and thus $r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2 \xrightarrow{p} \mathcal{V}$, without imposing those assumptions.

To start, we write $\tilde{g}_n = g_n - q_n' \mu$ and decompose

$$\begin{aligned} r_L \left(\sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2 - \mathcal{V}_L\right) =& r_L \left(\sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell\right)^2 \tilde{g}_n^2 - \mathcal{V}_L\right) \\ &+ r_L \sum_n \left(\left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 - \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell\right)^2\right) \tilde{g}_n^2 \\ &+ r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \left(\hat{g}_n^2 - \tilde{g}_n^2\right). \end{aligned} \tag{B22}$$

Adão et al. (2019) show that the second term of this expression is $o_p(1)$ under our assumptions, using the fact (their Lemma A.3, again generalized to include importance weights) that for a triangular array $\{A_{L1}, \dots, A_{LL}, B_{L1}, \dots, B_{LL}, C_{L1}, \dots, C_{LN_L}\}_{L=1}^{\infty}$ with $\mathbb{E}\left[A_{L\ell}^4 \mid \{\{s_{\ell' n}\}_n, e_{\ell'}\}_{\ell'}\right]$, $\mathbb{E}\left[B_{L\ell}^4 \mid \{\{s_{\ell' n}\}_n, e_{\ell'}\}_{\ell'}\right]$, and $\mathbb{E}\left[C_{Ln}^2 \mid \{\{s_{\ell n'}\}_{n'}, e_\ell\}_\ell\right]$ uniformly bounded,

$$r_L \sum_\ell \sum_{\ell'} \sum_n e_\ell e_{\ell'} s_{\ell n} s_{\ell' n} A_{L\ell} B_{L\ell'} C_{Ln} = O_p(1). \tag{B23}$$

Here with $D_\ell = (z_\ell, w_\ell')'$, $\theta = (\beta, \gamma')'$, and $\hat{\theta} = (\hat{\beta}, \hat{\gamma}')'$ we can write

$$\begin{aligned} \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 =& \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell\right)^2 + 2 \sum_\ell \sum_{\ell'} e_\ell e_{\ell'} s_{\ell n} s_{\ell' n} D_\ell' \left(\theta - \hat{\theta}\right) \varepsilon_{\ell'} \\ &+ \sum_\ell \sum_{\ell'} e_\ell e_{\ell'} D_\ell' \left(\theta - \hat{\theta}\right) D_{\ell'}' \left(\theta - \hat{\theta}\right), \end{aligned} \tag{B24}$$

and both D_ℓ and ε_ℓ have bounded fourth moments by the assumption of bounded fourth moments of

ϵ_ℓ, η_ℓ, u_ℓ, and q_n, and g_n in Assumption B6. Thus by the lemma

$$r_L \sum_n \left(\left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 - \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \right) \tilde{g}_n^2 = 2 \left(\theta - \hat{\theta}\right)' \left(r_L \sum_\ell \sum_{\ell'} \sum_n e_\ell e_{\ell'} s_{\ell n} s_{\ell' n} \tilde{g}_n^2 D_\ell \varepsilon_{\ell'} \right)$$
$$+ \left(\theta - \hat{\theta}\right)' \left(r_L \sum_\ell \sum_{\ell'} \sum_n e_\ell e_{\ell'} s_{\ell n} s_{\ell' n} \tilde{g}_n^2 D_\ell D_{\ell'}' \right) \left(\theta - \hat{\theta}\right)$$
$$= \left(\theta - \hat{\theta}\right)' O_p(1) + \left(\theta - \hat{\theta}\right)' O_p(1) \left(\theta - \hat{\theta}\right), \qquad \text{(B25)}$$

which is $o_p(1)$ by the consistency of $\hat{\theta}$ (implied by Assumptions B3-B6). Adão et al. (2019) further show the first term of equation (B22) is $o_p(1)$, without using the additional regularity conditions of their Proposition 5.

It thus remains for us to show the third term of (B22) is also $o_p(1)$. Note that

$$\hat{g}_n^2 = \left(g_n - q_n' \hat{\mu}\right)^2 = \tilde{g}_n^2 + \left(q_n' \left(\hat{\mu} - \mu\right)\right)^2 - 2 \tilde{g}_n q_n' \left(\hat{\mu} - \mu\right), \qquad \text{(B26)}$$

so that

$$r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \left(\hat{g}_n^2 - \tilde{g}_n^2\right)$$
$$= r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \left(q_n' \left(\hat{\mu} - \mu\right) - 2\tilde{g}_n\right) q_n' \left(\hat{\mu} - \mu\right)$$
$$= r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \left(q_n' \left(\hat{\mu} - \mu\right) - 2\tilde{g}_n\right) q_n' \left(\hat{\mu} - \mu\right)$$
$$+ r_L \sum_n \left(\left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 - \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \right) \left(q_n' \left(\hat{\mu} - \mu\right) - 2\tilde{g}_n\right) q_n' \left(\hat{\mu} - \mu\right). \qquad \text{(B27)}$$

Using the previous lemma, the first term of this expression is $O_p(1)\left(\hat{\mu} - \mu\right)$ since ε_ℓ, q_n, and $\tilde{g}_n$ have bounded fourth moments under Assumption B6. The second term is similarly $O_p(1)\left(\hat{\mu} - \mu\right)$ by the lemma and the decomposition used in equation (B25). Noting that $\hat{\mu} - \mu = \left(\sum_n s_n q_n q_n'\right)^{-1} \sum_n s_n q_n \tilde{g}_n \xrightarrow{p} 0$ under the assumptions completes the proof.

The Case of No Controls We show that when there are no controls besides a constant, i.e. $w_\ell = g_n = 1$, the standard errors are numerically the same. To prove this, it suffices to show that $\hat{g}_n = \ddot{g}_n$. Absent controls, $\hat{g}_n = g_n - \sum_n s_n g_n$ is the s_n-weighted demeaned shock. The $\ddot{g}_n$ are obtained as the projection coefficient of $z_\ell^\perp = z_\ell - \sum_\ell e_\ell z_\ell$ on the N shares. Note that

$$\sum_\ell e_\ell z_\ell = \sum_\ell e_\ell \sum_n s_{\ell n} g_n = \sum_n s_n g_n, \qquad \text{(B28)}$$

so that, with $\sum_n s_{\ell n} = 1$,

$$z_\ell - \sum_\ell e_\ell z_\ell = \sum_n s_{\ell n} g_n - \sum_n s_n g_n = \sum_\ell s_{\ell n} \hat{g}_n. \tag{B29}$$

This means that the projection in Adão et al. (2019) has exact fit and produces $\ddot{g}_n = \hat{g}_n$.

Relaxing Assumption B4 We now show that the standard errors from our equivalent regression in Proposition 5 are asymptotically conservative under a weaker assumption on the structure of controls than Assumption B4:

Assumption B4′: There exists a K-dimensional vector p_n, with uniformly bounded fourth moments, such that $w_\ell = \sum_n s_{\ell n} p_n + u_\ell$ for some K-dimensional vector u_ℓ and $\mathbb{E}\left[g_n \mid \mathcal{I}_L\right] = p_n' \mu$ for all n and for $\mathcal{I}_L = \left\{\{p_n\}_n, \{u_\ell, \epsilon_\ell, \eta_\ell, \{s_{\ell n}, \pi_{\ell n}\}_n, e_\ell\}_\ell\right\}$.

This assumption requires that the controls w_ℓ can be represented as noisy versions of some latent shift-share confounding variables $\sum_n s_{\ell n} p_n$. Since the variance of u_ℓ is unrestricted, this assumption relaxes not only our Assumption B4 but also the assumption of approximate shift-share controls in Adão et al. (2019).

Specifically, we show that under Assumptions B3, B4′, B5, and B6 the shock-level regression from Proposition 5 that controls for a subvector of confounders $q_n \subseteq p_n$ yields asymptotically conservative standard errors. Consider

$$\widehat{\Delta}_L = r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \hat{g}_n^2 - r_L \mathcal{V}_L \tag{B30}$$

with $\hat{g}_n$ still denoting the s_n-weighted projection of g_n on q_n (only), and $\mathcal{V}_L = \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell\right)^2 \operatorname{Var}\left[g_n \mid \mathcal{I}_L\right]$ where $\mathcal{I}_L$ is the expanded set from Assumption B4′. Write

$$\begin{aligned} \hat{g}_n &= g_n - p_n' \hat{\mu} \\ &= \tilde{g}_n + p_n' \left(\mu - \hat{\mu}\right), \end{aligned} \tag{B31}$$

where the non-zero elements of $\hat{\mu}$ correspond to the q_n subvector. We show that

$$\widehat{\Delta}_L - \Delta_L \xrightarrow{p} 0 \tag{B32}$$

for the non-negative

$$\Delta_L = r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell\right)^2 \left(p_n' \left(\mu - \hat{\mu}\right)\right)^2, \tag{B33}$$

when $\hat{\mu} = O_p(1)$.[16]

[16] We note this is a weaker condition than convergence of the incorrect shock-level projection (i.e. that $\hat{\mu} = \overline{\mu} + o_p(1)$

First note by equation (B31) that, for $\tilde{g}_n = g_n - p_n'\mu$,

$$\hat{g}_n^2 = \tilde{g}_n^2 + 2\tilde{g}_n p_n' (\mu - \hat{\mu}) + (p_n' (\mu - \hat{\mu}))^2 .$$

Thus

$$\begin{aligned}\widehat{\Delta}_L - \Delta_L =& r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \tilde{g}_n^2 - r_L \mathcal{V}_L \\ &+ 2 r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \tilde{g}_n p_n' (\mu - \hat{\mu}) . \end{aligned} \quad \text{(B34)}$$

We showed that the first term is $o_p(1)$ in the proof of Proposition 5. It remains to show the second term is also $o_p(1)$. To see this, write

$$r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \tilde{g}_n p_n = r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \tilde{g}_n p_n \quad \text{(B35)}$$

$$+ r_L \sum_n \left(\left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 - \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \right) \tilde{g}_n p_n . \quad \text{(B36)}$$

We have

$$\mathbb{E} \left[r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \tilde{g}_n p_n \right] = \mathbb{E} \left[r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \mathbb{E} \left[\tilde{g}_n \mid \mathcal{I}_L \right] p_n \right] = 0 \quad \text{(B37)}$$

since $\mathbb{E} [\tilde{g}_n \mid \mathcal{I}_L]$. Furthermore,

$$\text{Var} \left[r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^2 \tilde{g}_n p_n \right] = r_L^2 \sum_n \mathbb{E} \left[\left(\sum_\ell e_\ell s_{\ell n} \varepsilon_\ell \right)^4 \text{Var} \left[\tilde{g}_n \mid \mathcal{I}_L \right] p_n p_n' \right] \to 0, \quad \text{(B38)}$$

implying the first term of (B35) is $o_p(1)$. The second term of this expression can also be shown to be $o_p(1)$ by applying the lemma from Adão et al. (2019) and the representation used in equation (B25). Thus

$$2 r_L \sum_n \left(\sum_\ell e_\ell s_{\ell n} \hat{\varepsilon}_\ell \right)^2 \tilde{g}_n p_n' (\mu - \hat{\mu}) = o_p(1)' O_p(1), \quad \text{(B39)}$$

completing the proof.

Comparison of Standard Errors under Assumption B4 The characterization of the standard errors in equations (B19)–(B20) also offers insights into how these standard errors may differ in

for some $\overline{\mu}$).

presence of controls, when both standard error calculations are asymptotically valid. We argue that under the conditions of Proposition 5, our standard errors are likely smaller in finite samples. More precisely, we show that the homoskedastic version of (B19) is smaller than the homoskedastic version of (B20). This is suggestive of the comparison under heteroskedasticity, but is not a proof.

To see this, consider versions of the two standard error formulas obtained under shock homoskedasticity (i.e. $\text{Var}\left[g_n \mid \mathcal{I}_L\right] = \sigma_g^2$):

$$\widehat{se}_{\text{equiv}}^{\text{homo}} = \frac{\sqrt{\left(\sum_n s_n^2 \hat{\varepsilon}_n^2\right)\left(\sum_n s_n \hat{g}_n^2\right)}}{\left|\sum_n s_n \bar{x}_n^{\perp} g_n\right|} \tag{B40}$$

$$\widehat{se}_{\text{AKM}}^{\text{homo}} = \frac{\sqrt{\left(\sum_n s_n^2 \hat{\varepsilon}_n^2\right)\left(\sum_n s_n \ddot{g}_n^2\right)}}{\left|\sum_\ell e_\ell x_\ell^{\perp} z_\ell\right|}, \tag{B41}$$

which differ by a factor of $\sqrt{\sum_n s_n \hat{g}_n^2 / \sum_n s_n \ddot{g}_n^2}$.

When the SSIV controls have an exact shift-share structure, $w_\ell = \sum_n s_{\ell n} q_n$, the share projection producing $\ddot{g}_n$ has exact fit such that one can represent $\ddot{g}_n = g_n - q_n' \hat{\mu}_{AKM}$ for some $\hat{\mu}_{AKM}$. In this case the s_n-weighted sum of squares of shock residuals is lower in our equivalent regression by construction of $\hat{\mu}$: $\sum_n s_n \hat{g}_n^2 \le \sum_n s_n \ddot{g}_n^2$ (with strict inequality when $\hat{\mu}_{\text{AKM}} \neq \hat{\mu}$). Similarly, when w_ℓ instead contains controls that are included for efficiency only and are independent of the shocks, projection of z_ℓ on the shares produces a noisy estimate of $g_n - \sum_n s_n g_n$, which again has a higher weighted sum of squares.

Null-Imposed Inference Procedure Finally, our shock-level equivalence provides a convenient implementation for the alternative inference procedure that may have superior finite-sample performance. Adão et al. (2019) show how standard errors that impose a given null hypothesis $\beta = \beta_0$ in estimating the residual ε_ℓ can generate confidence intervals with better coverage in situations with few shocks (and a similar argument can be made in the case of shocks with a heavy-tailed distribution).[17] Building on Proposition 5, such confidence intervals can be constructed in the same way as in any regular shock-level IV regression. To test $\beta = \beta_0$, one regresses $\bar{y}_n^{\perp} - \beta_0 \bar{x}_n^{\perp}$ on the shocks g_n (weighting by s_n and including any relevant shock-level controls q_n) and uses a null-imposed residual variance estimate. This procedure corresponds to the standard shock-level Lagrange multiplier test for $\beta = \beta_0$ that can be implemented by standard statistical software.[18] The confidence interval for β is constructed by collecting all candidate β_0 that are not rejected.

[17] As explained by Adão et al. (2019), the problem that this "AKM0" confidence interval addresses generalizes the standard finite-sample bias of cluster-robust standard errors with few clusters (Cameron and Miller 2015). With few or heavy-tailed shocks, estimates of the residual variance will tend to be biased downwards, leading to undercoverage of confidence intervals based on standard errors that do not impose the null.

[18] For example in Stata one can use the *ivreg2* overidentification test statistic from regressing $\bar{y}_n^{\perp} - \beta_0 \bar{x}_n^{\perp}$ on q_n with no endogenous variables and with g_n specified as the instrument (again with s_n weights).

B.3 Proposition A1

We consider each expectation in equation (A2) in turn. For each n, write

$$\kappa_n(g_{-n}, \varepsilon_\ell, \eta_\ell) = \lim_{g_n \to -\infty} y(x_1([g_n; g_{-n}], \eta_{\ell 1}), \ldots, x_R([g_n; g_{-n}], \eta_{\ell R}), \varepsilon_\ell) \tag{B42}$$

such that

$$\begin{aligned} s_{\ell n} e_\ell g_n y_\ell =& s_{\ell n} e_\ell g_n \kappa_n(g_{-n}, \varepsilon_\ell, \eta_\ell) \\ &+ s_{\ell n} e_\ell g_n \int_{-\infty}^{g_n} \frac{\partial}{\partial g_n} y(x_1([\gamma; g_{-n}], \eta_{\ell 1}), \ldots x_R([\gamma; g_{-n}], \eta_{\ell R}), \varepsilon_\ell) d\gamma. \end{aligned} \tag{B43}$$

By as-good-as-random shock assignment, the expectation of the first term is

$$\mathbb{E}\left[s_{\ell n} e_\ell g_n \kappa_n(g_{-n}, \varepsilon_\ell, \eta_\ell)\right] = \mathbb{E}\left[s_{\ell n} e_\ell \mathbb{E}\left[g_n \mid s, e, g_{-n}, \varepsilon, \eta_\ell\right] \kappa_n(g_{-n}, \varepsilon_\ell, \eta_\ell)\right] = 0, \tag{B44}$$

while the expectation of the second is

$$\begin{aligned} &\mathbb{E}\left[s_{\ell n} e_\ell g_n \int_{-\infty}^{g_n} \frac{\partial}{\partial g_n} y(x_1([\gamma; g_{-n}], \eta_{\ell 1}), \ldots x_R([\gamma; g_{-n}], \eta_{\ell R}), \varepsilon_\ell) d\gamma\right] \\ &= \mathbb{E}\left[s_{\ell n} e_\ell \int_{-\infty}^{\infty} \int_{-\infty}^{g_n} g_n \frac{\partial}{\partial g_n} y(x_1([\gamma; g_{-n}], \eta_{\ell 1}), \ldots x_R([\gamma; g_{-n}], \eta_{\ell R}), \varepsilon_\ell) d\gamma dF_n(g_n \mid \mathcal{I})\right] \\ &= \mathbb{E}\left[s_{\ell n} e_\ell \int_{-\infty}^{\infty} \frac{\partial}{\partial g_n} y(x_1([\gamma; g_{-n}], \eta_{\ell 1}), \ldots x_R([\gamma; g_{-n}], \eta_{\ell R}), \varepsilon_\ell) \int_{\gamma}^{\infty} g_n dF_n(g_n \mid \mathcal{I}) d\gamma\right] \end{aligned} \tag{B45}$$

where $F_n(\cdot \mid \mathcal{I})$ denotes the conditional distribution of g_n. Thus

$$\begin{aligned} \mathbb{E}\left[s_{\ell n} e_\ell g_n y_\ell\right] &= \mathbb{E}\left[s_{\ell n} e_\ell \int_{-\infty}^{\infty} \frac{\partial}{\partial g_n} y(x_1([\gamma; g_{-n}], \eta_\ell), \ldots x_R([\gamma; g_{-n}], \eta_\ell), \varepsilon_\ell) \mu_n(\gamma \mid \mathcal{I}) d\gamma\right] \\ &= \sum_r \mathbb{E}\left[\int_{-\infty}^{\infty} s_{\ell n} e_\ell \alpha_{\ell r} \pi_{\ell r n}([\gamma; g_{-n}]) \mu_n(\gamma \mid \mathcal{I}) \tilde{\beta}_{\ell r n}(\gamma) d\gamma\right] \end{aligned} \tag{B46}$$

where

$$\begin{aligned} \mu_n(\gamma \mid \mathcal{I}) \equiv& \int_{\gamma}^{\infty} g_n dF_n(g_n \mid \mathcal{I}). \\ &= \left(\mathbb{E}\left[g_n \mid g_n \geq \gamma, \mathcal{I}\right] - \mathbb{E}\left[g_n \mid g_n < \gamma, \mathcal{I}\right]\right) Pr\left(g_n \geq \gamma \mid \mathcal{I}\right)\left(1 - Pr\left(g_n \geq \gamma \mid \mathcal{I}\right)\right) \geq 0 \ a.s. \end{aligned} \tag{B47}$$

Similarly

$$\mathbb{E}\left[s_{\ell n} e_\ell g_n x_\ell\right] = \sum_r \mathbb{E}\left[\int_{-\infty}^{\infty} s_{\ell n} e_\ell \alpha_{\ell r} \pi_{\ell r n}([\gamma; g_{-n}]) \mu_n(\gamma \mid \mathcal{I}) d\gamma\right]. \tag{B48}$$

Combining equations (B46) and (B48) completes the proof, with

$$\omega_{\ell rn}(\gamma) = s_{\ell n} e_\ell \alpha_{\ell r} \mu_n(\gamma \mid \mathcal{I}) \pi_{\ell rn}([\gamma; g_{-n}]) \geq 0 \text{ a.s.} \tag{B49}$$

B.4 Proposition A2

By definition of $\bar{\varepsilon}_n$,

$$\begin{aligned} \bar{\varepsilon}_n &= \frac{\sum_\ell e_\ell s_{\ell n} \left(\sum_{n'} s_{\ell n'} \nu_{n'} + \check{\varepsilon}_\ell\right)}{\sum_\ell e_\ell s_{\ell n}} \\ &\equiv \sum_{n'} \alpha_{nn'} \nu_{n'} + \bar{\check{\varepsilon}}_n, \end{aligned} \tag{B50}$$

for $\alpha_{nn'} = \frac{\sum_\ell e_\ell s_{\ell n} s_{\ell n'}}{\sum_\ell e_\ell s_{\ell n}}$ and $\bar{\check{\varepsilon}}_n = \frac{\sum_\ell e_\ell s_{\ell n} \check{\varepsilon}_\ell}{\sum_\ell e_\ell s_{\ell n}}$. Therefore,

$$\begin{aligned} \operatorname{Var}\left[\bar{\varepsilon}_n\right] &= \sum_{n'} \sigma_{n'}^2 \alpha_{nn'}^2 + \operatorname{Var}\left[\bar{\check{\varepsilon}}_n\right] \\ &\geq \sigma_\nu^2 \alpha_{nn}^2, \end{aligned} \tag{B51}$$

and

$$\max_n \operatorname{Var}\left[\bar{\varepsilon}_n\right] \geq \sigma_\nu^2 \max_n \alpha_{nn}^2. \tag{B52}$$

To establish a lower bound on this quantity, observe that the s_n-weighted average of α_{nn} satisfies:

$$\begin{aligned} \sum_n s_n \alpha_{nn} &= \sum_n s_n \frac{\sum_\ell e_\ell s_{\ell n}^2}{s_n} \\ &= H_L. \end{aligned} \tag{B53}$$

Since $\sum_n s_n = 1$, it follows that $\max_n \alpha_{nn} \geq H_L$ and therefore $\max_n \operatorname{Var}\left[\bar{\varepsilon}_n\right] \geq \sigma_\nu^2 H_L^2$. Since $H_L \to \bar{H} > 0$, we conclude that, for sufficiently large L, $\max_n \operatorname{Var}\left[\bar{\varepsilon}_n\right]$ is bounded from below by any positive $\delta < \sigma_\nu^2 \bar{H}^2$.

B.5 Proposition A3

To prove (A9), we aggregate (A8) across industries within a region using $E_{\ell n}$ weights:

$$y_\ell = \left(\beta_0 - \beta_1\right) x_\ell + \varepsilon_\ell, \tag{B54}$$

where $\varepsilon_\ell = \sum_n s_{\ell n}\varepsilon_{\ell n}$. The shift-share instrument z_ℓ is relevant because

$$\begin{aligned}\mathbb{E}\left[\sum_\ell e_\ell x_\ell z_\ell\right] &= \sum_\ell e_\ell \mathbb{E}\left[\sum_n s_{\ell n}\left(\bar{\pi} g_n + \eta_{\ell n}\right)\cdot \sum_{n'} s_{\ell n'} g_{n'}\right] \\ &= \sum_{\ell,n} e_\ell s_{\ell n}^2 \bar{\pi}\sigma_g^2 \\ &\geq \bar{H}_L \bar{\pi}\sigma_g^2, \end{aligned} \tag{B55}$$

while exclusion holds because

$$\begin{aligned}\mathbb{E}\left[\sum_\ell e_\ell z_\ell \varepsilon_\ell\right] &= \sum_\ell e_\ell \mathbb{E}\left[\sum_n s_{\ell n}\varepsilon_{\ell n}\cdot \sum_{n'} s_{\ell n'} g_{n'}\right] \\ &= 0. \end{aligned} \tag{B56}$$

Thus by an appropriate law of large numbers, $\hat{\beta} = \beta_0 - \beta_1 + o_p(1)$.

To study $\hat{\beta}_{\text{ind}}$, we aggregate (A8) across regions (again with $E_{\ell n}$ weights):

$$y_n = \beta_0 x_n - \beta_1 \sum_\ell \omega_{\ell n} \sum_{n'} s_{\ell n'} x_{\ell n'} + \varepsilon_n, \tag{B57}$$

for $\varepsilon_n = \sum_\ell \omega_{\ell n}\varepsilon_{\ell n}$. The resulting IV estimate yields

$$\begin{aligned}\hat{\beta}_{\text{ind}} - \beta_0 &= \frac{\sum_n s_n y_n g_n}{\sum_n s_n x_n g_n} - \beta_0 \\ &= \frac{\sum_n s_n \left(-\beta_1 \sum_\ell \omega_{\ell n} \sum_{n'} s_{\ell n'} x_{\ell n'} + \varepsilon_n\right) g_n}{\sum_n s_n x_n g_n}. \end{aligned} \tag{B58}$$

The expected denominator of $\hat{\beta}_{\text{ind}}$ is non-zero:

$$\begin{aligned}\mathbb{E}\left[\sum_n s_n x_n g_n\right] &= \sum_n s_n \mathbb{E}\left[\sum_\ell \omega_{\ell n}\left(\bar{\pi} g_n + \eta_{\ell n}\right) g_n\right] \\ &= \sum_n s_n \omega_{\ell n} \bar{\pi}\sigma^2 \\ &= \sum_n \frac{E_n}{E}\cdot \frac{E_{\ell n}}{E} \bar{\pi}\sigma^2 \\ &= \bar{\pi}\sigma^2, \end{aligned} \tag{B59}$$

while the expected numerator is

$$\begin{aligned}\mathbb{E}\left[\sum_n s_n \left(-\beta_1 \sum_\ell \omega_{\ell n} \sum_{n'} s_{\ell n'} x_{\ell n'} + \varepsilon_n\right) g_n\right] &= -\beta_1 \sum_{n,\ell} s_n \omega_{\ell n} s_{\ell n} \bar{\pi}\sigma^2 \\ &= -\beta_1 H_L \bar{\pi}\sigma^2, \end{aligned} \tag{B60}$$

where the last equality follows because

$$\begin{aligned}\sum_{n,\ell} s_n \omega_{\ell n} s_{\ell n} &= \sum_{n,\ell} \frac{E_n}{E} \frac{E_{\ell n}}{E_n} \frac{E_{\ell n}}{E_\ell} \\ &= \sum_{n,\ell} \frac{E_\ell}{E} \left(\frac{E_{\ell n}}{E_\ell} \right)^2 \\ &= \sum_{n,\ell} e_\ell s_{\ell n}^2 \\ &= H_L. \end{aligned} \tag{B61}$$

Thus by an appropriate law of large numbers,

$$\hat{\beta}_{\text{ind}} = \beta_0 - \beta_1 H_L + o_p(1). \tag{B62}$$

B.6 Proposition A4

By appropriate laws of large numbers,

$$\begin{aligned}\hat{\beta} &= \frac{\mathbb{E}\left[\sum_\ell E_\ell \left(\sum_n s_{\ell n} y_{\ell n}\right) \left(\sum_{n'} s_{\ell n'} g_{n'}\right)\right]}{\mathbb{E}\left[\sum_\ell E_\ell \left(\sum_n s_{\ell n} x_{\ell n}\right) \left(\sum_{n'} s_{\ell n'} g_{n'}\right)\right]} + o_p(1) \\ &= \frac{\sum_{\ell,n} E_\ell s_{\ell n}^2 \pi_{\ell n} \sigma_n^2 \beta_{\ell n}}{\sum_{\ell,n} E_\ell s_{\ell n}^2 \pi_{\ell n} \sigma_n^2} + o_p(1) \\ &= \frac{\sum_{\ell,n} E_{\ell n} s_{\ell n} \pi_{\ell n} \sigma_n^2 \beta_{\ell n}}{\sum_{\ell,n} E_{\ell n} s_{\ell n} \pi_{\ell n} \sigma_n^2} + o_p(1) \end{aligned} \tag{B63}$$

while

$$\begin{aligned}\hat{\beta}_{\text{ind}} &= \frac{\sum_n E_n y_n g_n}{\sum_n E_n x_n g_n} \\ &= \frac{\mathbb{E}\left[\sum_n E_n \left(\sum_\ell \omega_{\ell n} y_{\ell n}\right) g_n\right]}{\mathbb{E}\left[\sum_n E_n \left(\sum_\ell \omega_{\ell n} x_{\ell n}\right) g_n\right]} + o_p(1) \\ &= \frac{\sum_{\ell,n} E_n \omega_{\ell n} \pi_{\ell n} \sigma_n^2 \beta_{\ell n}}{\sum_{\ell,n} E_n \omega_{\ell n} \pi_{\ell n} \sigma_n^2} + o_p(1) \\ &= \frac{\sum_{\ell,n} E_{\ell n} \pi_{\ell n} \sigma_n^2 \beta_{\ell n}}{\sum_{\ell,n} E_{\ell n} \pi_{\ell n} \sigma_n^2} + o_p(1). \end{aligned} \tag{B64}$$

B.7 Proposition A5

We prove each part of this proposition in turn.

1. Expanding the moment condition yields:

$$\begin{aligned}\mathbb{E}\left[\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell,LOO}\right] &= \sum_\ell \mathbb{E}\left[e_\ell \varepsilon_\ell \sum_n s_{\ell n} \frac{\sum_{\ell'\neq\ell}\omega_{\ell' n}\psi_{\ell' n}}{\sum_{\ell'\neq\ell}\omega_{\ell' n}}\right] \\ &= \sum_\ell e_\ell \sum_n s_{\ell n} \frac{\sum_{\ell'\neq\ell}\omega_{\ell' n}\mathbb{E}\left[\varepsilon_\ell \psi_{\ell' n}\right]}{\sum_{\ell'\neq\ell}\omega_{\ell' n}} \\ &= 0. \end{aligned} \tag{B65}$$

2. The assumption of part (1) is satisfied here, so $\mathbb{E}\left[\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell,LOO}\right] = 0$. We now establish that $\mathbb{E}\left[\left(\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell,LOO}\right)^2\right] \to 0$, which implies $\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell,LOO} \xrightarrow{p} 0$ and thus consistency of the LOO SSIV estimator provided it has a first stage:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_\ell e_\ell \varepsilon_\ell \psi_{\ell,LOO}\right)^2\right] &= \sum_{\substack{\ell_1,\ell_2,n_1,n_2,\\ \ell_1'\neq\ell_1,\ell_2'\neq\ell_2}} e_{\ell_1} e_{\ell_2} s_{\ell_1 n_1} s_{\ell_2 n_2} \frac{\omega_{\ell_1' n_1}}{\sum_{\ell\neq\ell_1}\omega_{\ell n_1}} \frac{\omega_{\ell_2' n_2}}{\sum_{\ell\neq\ell_2}\omega_{\ell n_2}} \mathbb{E}\left[\varepsilon_{\ell_1}\varepsilon_{\ell_2}\psi_{\ell_1' n_1}\psi_{\ell_2' n_2}\right] \\ &\leq \sum_{\substack{(\ell_1,\ell_2,\ell_1',\ell_2')\in\mathcal{J},\\ n_1,n_2}} e_{\ell_1} e_{\ell_2} s_{\ell_1 n_1} s_{\ell_2 n_2} \frac{\omega_{\ell_1' n_1}}{\sum_{\ell\neq\ell_1}\omega_{\ell n_1}} \frac{\omega_{\ell_2' n_2}}{\sum_{\ell\neq\ell_2}\omega_{\ell n_2}} \cdot B \to 0. \end{aligned} \tag{B66}$$

Here the second line used the first regularity condition, which implies that $\mathbb{E}\left[\varepsilon_{\ell_1}\varepsilon_{\ell_2}\psi_{\ell_1' n_1}\psi_{\ell_2' n_2}\right] = 0$ whenever there is at least one index among $\{\ell_1,\ell_2,\ell_1',\ell_2'\}$ which is not equal to any of the others, i.e. for all $(\ell_1,\ell_2,\ell_1',\ell_2') \notin \mathcal{J}$.

3. We show that under the given assumptions on $s_{\ell n}$, e_ℓ, and $\omega_{\ell n}$, the expression in (A21) is bounded by $4N/L$:

$$\begin{aligned}&\sum_{\substack{(\ell_1,\ell_2,\ell_1',\ell_2')\in\mathcal{J},\\ n_1,n_2}} e_{\ell_1} e_{\ell_2} s_{\ell_1 n_1} s_{\ell_2 n_2} \frac{\omega_{\ell_1' n_1}}{\sum_{\ell\neq\ell_1}\omega_{\ell n_1}} \frac{\omega_{\ell_2' n_2}}{\sum_{\ell\neq\ell_2}\omega_{\ell n_2}} \\ &= \sum_{(\ell_1,\ell_2,\ell_1',\ell_2')\in\mathcal{J}} \frac{1}{L^2} \frac{\omega_{\ell_1' n(\ell_1)}}{\sum_{\ell\neq\ell_1}\omega_{\ell n(\ell_1)}} \frac{\omega_{\ell_2' n(\ell_2)}}{\sum_{\ell\neq\ell_2}\omega_{\ell n(\ell_2)}} \\ &= \frac{1}{L^2} \sum_{\substack{(\ell_1,\ell_2,\ell_1',\ell_2')\in\mathcal{J}\\ n(\ell_1')=n(\ell_1),\\ n(\ell_2')=n(\ell_2)}} \frac{1}{L_{n(\ell_1)}-1} \frac{1}{L_{n(\ell_2)}-1} \\ &= \frac{1}{L^2} \sum_n \frac{2L_n\left(L_n-1\right)}{\left(L_n-1\right)^2} \leq 4\frac{N}{L}. \end{aligned} \tag{B67}$$

Here the second line plugs in the expressions for $s_{\ell n}$ and e_ℓ, and the third line plugs in $\omega_{\ell n}$. The last line uses the fact that any tuple $(\ell_1,\ell_2,\ell_1',\ell_2') \in \mathcal{J}$ such that $n(\ell_1') = n(\ell_1)$ and $n(\ell_2') = n(\ell_2)$ has all four elements exposed to the same shock n. Moreover, it is easily verified that all of these tuples have a structure $(\ell_A, \ell_B, \ell_A, \ell_B)$ or $(\ell_A, \ell_B, \ell_B, \ell_A)$ for any $\ell_A \neq \ell_B$ exposed to the same shock. Therefore, there are $2L_n(L_n - 1)$ of them for each n. Finally, $\frac{L_n}{L_n - 1} \leq 2$ as $L_n \geq 2$.

B.8 Proposition A6

National industry employment satisfies $E_n = \sum_\ell E_{\ell n}$; log-linearizing this immediately implies (A30). To solve for $g_{\ell n}$, log-linearize (A27), (A28), and (A29):

$$\hat{E}_\ell = \phi \hat{W}_\ell + \varepsilon_\ell, \tag{B68}$$

$$g_{\ell n} = g_n^* + \hat{\xi}_{\ell n} - \sigma \hat{W}_\ell, \tag{B69}$$

$$\hat{E}_\ell = \sum_n s_{\ell n} g_{\ell n}. \tag{B70}$$

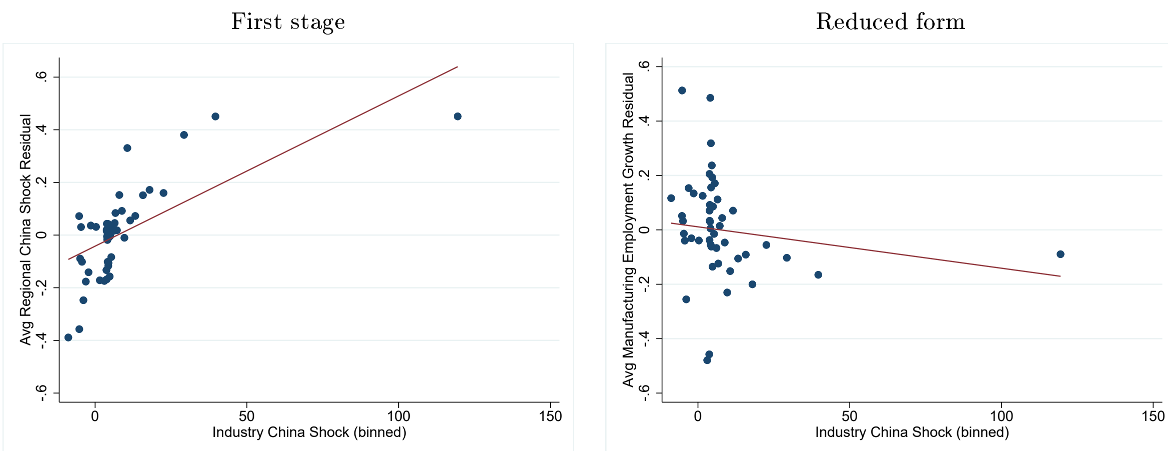
Solving this system of equations yields

$$\hat{W}_\ell = \frac{1}{\sigma + \phi} \left(\sum_n s_{\ell n} \left(g_n^* + \hat{\xi}_{\ell n} \right) - \varepsilon_\ell \right) \tag{B71}$$

and expression (A31).

C Appendix Figures and Tables

Figure C1: Industry-Level Variation in the Autor et al. (2013) Setting



Notes: This figure shows binned scatterplots of shock-level outcome and treatment residuals, $\bar{y}_{nt}^{\perp}$ and $\bar{x}_{nt}^{\perp}$, corresponding to the SSIV specification in column 3 of Table 4. The manufacturing industry shocks, g_{nt}, are residualized on period indicators (with the full-sample mean added back) and grouped into fifty weighted bins, with each bin representing around 2% of total share weight s_{nt}. Lines of best fit, indicated in red, are weighted by the same s_{nt}. The slope coefficients equal 5.71×10^{-3} and -1.52×10^{-3}, respectively, with the ratio (-0.267) equaling the SSIV coefficient in column 3 of Table 4.

Table C1: Shift-Share IV Estimates of the Effect of Chinese Imports on Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unemployment growth	0.221	0.217	0.063	-0.014	0.104	0.107	0.235
	(0.049)	(0.046)	(0.060)	(0.079)	(0.079)	(0.083)	(0.178)
NILF growth	0.553	0.534	0.098	0.149	0.142	0.117	0.187
	(0.185)	(0.183)	(0.133)	(0.083)	(0.155)	(0.161)	(0.297)
Log weekly wage growth	-0.759	-0.607	0.227	0.320	0.145	0.063	-0.211
	(0.258)	(0.226)	(0.242)	(0.209)	(0.264)	(0.260)	(0.651)
# of industry-periods	796	794	794	794	794	794	794
# of region-periods	1,444	1,444	1,444	1,444	1,444	1,444	1,444

Notes: This table extends the analysis of Table 4 to different regional outcomes in Autor et al. (2013): unemployment growth, labor force non-participation (NILF) growth, and log average weekly wage growth. The specifications are otherwise the same as in the corresponding columns of Table 4. SIC3-clustered exposure-robust standard errors are computed using equivalent industry-level IV regressions and reported in parentheses.

Table C2: Alternative Standard Errors in the Autor et al. (2013) Setting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Coefficient	-0.596	-0.489	-0.267	-0.314	-0.310	-0.290	-0.432
Table 4 SE	(0.114)	(0.100)	(0.099)	(0.107)	(0.134)	(0.129)	(0.205)
State-clustered SE	(0.099)	(0.086)	(0.086)	(0.097)	(0.104)	(0.101)	(0.193)
Adão et al. (2019) SE	(0.126)	(0.116)	(0.113)	(0.107)	(0.143)	(0.140)	(0.192)
Confidence interval with	[-1.059,	[-0.832,	[-0.568,	[-0.637,	[-0.705,	[-0.699,	[-1.207,
the null imposed	-0.396]	-0.309]	-0.028]	-0.018]	-0.002]	0.002]	0.122]

Notes: This table extends the analysis of Table 4 by reporting conventional state-clustered standard errors, the Adão et al. (2019) SIC3-clustered standard errors, and confidence intervals based on the equivalent industry-level IV regression with the null imposed, as discussed in Section 5.1. The specifications are the same as those in the corresponding columns of Table 4; for comparison we repeat the coefficient estimates and exposure-robust standard errors from that table.

Table C3: Period-Specific Effects in the Autor et al. (2013) Setting

	(1) Mfg. emp.	(2) Unemp.	(3) NILF	(4) Wages
Coefficient (1990s)	-0.491	0.329	1.209	-0.649
	(0.266)	(0.155)	(0.347)	(0.571)
Coefficient (2000s)	-0.225	0.014	-0.109	0.391
	(0.103)	(0.083)	(0.123)	(0.288)

Notes: This table reports coefficient estimates for versions of the shift-share IV specification in column 3 of Tables 4 and C1, allowing the treatment coefficient to vary by period. This specification uses two endogenous treatment variables (treatment interacted with period indicators) and two corresponding shift-share instruments. The controls are the same as in column 3 of Table 4. SIC3-clustered exposure-robust standard errors are obtained by the equivalent shock-level regressions and reported in parentheses.

Table C4: Robustness to Acemoglu et al. (2016) Controls in the Autor et al. (2013) Setting

	(1)	(2)	(3)	(4)
Coefficient	-0.200	-0.293	-0.241	-0.232
	(0.093)	(0.125)	(0.115)	(0.122)
Regional controls ($w_{\ell t}$)				
Autor et al. (2013) controls	✓	✓	✓	✓
Period-specific lagged mfg. share	✓	✓	✓	✓
Lagged 10-sector shares		✓	✓	✓
Local Acemoglu et al. (2016) controls		✓		✓
Local Acemoglu et al. (2016) pre-trends	✓		✓	✓
SSIV first stage F-stat.	118.9	53.3	65.9	56.6
# of region-periods		1,444		
# of industry-periods		794		

Notes: This table extends Table 4 by adding exposure-weighted sums of the other industry-level controls in Table 3 of Acemoglu et al. (2016). Pre-trends controls refer to the changes in industry log average wages and in the industry share of total U.S. employment over 1976–91; see the notes to Table 4 notes for details on the other controls and calculation of the SIC3-clustered exposure-robust standard errors (in parentheses) and first-stage F-statistics.

Table C5: Overidentified Shift-Share IV Estimates of the Effect of Chinese Imports on Manufacturing Employment

	(1)	(2)	(3)
Coefficient	-0.238	-0.247	-0.158
	(0.099)	(0.105)	(0.078)
Shock-level estimator	2SLS	LIML	GMM
Effective first stage F-statistic		15.10	
$\chi^2(7)$ overid. test stat. [p-value]		10.92 [0.142]	

Notes: Column 1 of this table reports an overidentified estimate of the coefficient corresponding to column 3 of Table 4, obtained from a two-stage least squares regression of shock-level average manufacturing employment growth residuals $\bar{y}_{nt}^{\perp}$ on shock-level average Chinese import competition growth residuals $\bar{x}_{nt}^{\perp}$, instrumenting by the growth of imports (per U.S. worker) in each of the eight non-U.S. countries from ADH, g_{nk} for $k = 1, \ldots, 8$, controlling for period fixed effects q_{nt}, and weighting by average industry exposure s_{nt}. Column 2 reports the corresponding limited information maximum likelihood estimate, while column 3 reports a two-step optimal generalized method of moments estimate. Standard errors, the optimal weight matrix, and the Hansen (1982) χ^2 test of overidentifying restrictions all allow for clustering of shocks at the SIC3 industry group level. The first-stage F-statistic is computed by a shift-share version of the Montiel Olea and Pflueger (2013) method described in Appendix A.10.

Table C6: Bartik (1991) Application

	(1)	(2)
Leave-one-out estimator	1.277	1.300
	(0.150)	(0.124)
Conventional estimator	1.215	1.286
	(0.139)	(0.121)
H heuristic	1.32	10.50
Population weights	✓	
# of region-periods	2,166	

Notes: Column 1 replicates column 2 of Table 3 from Goldsmith-Pinkham et al. (2020), reporting two SSIV estimators of the inverse labor supply elasticity, with and without the leave-one-out adjustment. Regions are U.S. commuting zones; periods are 1980s, 1990s, and 2000s; all specifications include controls for 1980 regional characteristics interacted with period indicators (see Goldsmith-Pinkham et al. (2020) for more details). Standard errors allow for clustering by commuting zones. Column 1 uses 1980 population weights, while column 2 repeats the same analysis without population weights. The table also reports the H heuristic for the importance of the leave-one-out adjustment proposed in Appendix A.6 (equation (A24)).

Table C7: Simulated 5% Rejection Rates for Shift-Share and Conventional Shock-Level IV

		SSIV Exposure-Robust SE		Shock-level IV Robust SE	
		Null not Imposed	Null Imposed	Null not Imposed	Null Imposed
		(1)	(2)	(3)	(4)
	Panel A: Benchmark Monte-Carlo Simulation				
(a)	Normal shocks	7.6%	5.2%	6.8%	5.0%
(b)	Wild bootstrap (benchmark)	8.0%	4.9%	14.2%	4.0%
	Panel B: Higher Industry Concentration				
(c)	$1/HHI = 50$	5.6%	4.9%	8.4%	6.1%
(d)	$1/HHI = 20$	7.3%	5.5%	7.0%	10.7%
(e)	$1/HHI = 10$	9.0%	8.2%	14.8%	23.8%
	Panel C: Smaller Numbers of Industries or Regions				
(f)	$N = 136$ (SIC3 industries)	5.4%	4.5%	7.7%	4.3%
(g)	$N = 20$ (SIC2 industries)	7.7%	3.7%	7.9%	3.2%
(h)	$L = 100$ (random regions)	9.7%	4.5%	N/A	
(i)	$L = 25$ (random regions)	10.4%	4.3%	N/A	

Notes: This table summarizes the results of the Monte-Carlo analysis described in Appendix A.11, reporting the rejection rates for a nominal 5% level test of the true null that $\beta^* = 0$. In all panels, columns 1 and 2 are simulated from the SSIV design based on Autor et al. (2013), as in column 3 of Table 4, while columns 3 and 4 are based on the conventional industry-level IV in Acemoglu et al. (2016). Column 1 uses exposure-robust standard errors from the equivalent industry-level IV and column 2 implements the version with the null hypothesis imposed. Columns 3 and 4 parallel columns 1 and 2 when applied to conventional IV. In Panel A, the simulations approximate the data-generating process using a normal distribution in row (a), with the variance matched to the sample variance of the shocks in the data after de-meaning by year, while wild bootstrap is used in row (b), following Liu (1988). Panel B documents the role of the Herfindahl concentration index across industries, varying 1/HHI from 50 to 10 in rows (c) to (e), compared with 191.6 for shift-share IV and 189.7 for conventional IV. Panel C documents the role of the number of regions and industries. We aggregate industries from 397 four-digit manufacturing SIC industries into 136 three-digit industries in row (f) and further into 20 two-digit industries in row (g). In rows (h) and (i), we select a random subset of region in each simulation. See Appendix A.11 for a complete discussion.

Table C8: First Stage F-statistics as a Rule of Thumb: Monte-Carlo Evidence

	Number of Instruments				
	1	5	10	25	50
	(1)	(2)	(3)	(4)	(5)
Panel A: SSIV					
5% rejection rate	8.0%	8.9%	11.5%	15.0%	23.0%
Median bias, % of std. dev.	0.3%	14.6%	28.3%	43.2%	72.1%
Median first-stage F	54.3	14.8	9.1	6.4	7.7
Panel B: Conventional Shock-Level IV					
5% rejection rate	13.6%	13.9%	14.9%	17.7%	22.0%
Median bias, % of std. dev.	-0.3%	10.1%	27.1%	57.0%	80.2%
Median first-stage F	59.4	19.4	13.2	10.0	11.2
Number of simulations	10,000	3,000	1,500	500	300

Notes: This table reports the results of the Monte-Carlo analysis with many weak instruments, described in Appendix A.11. Panel A is simulated from the SSIV design based on Autor et al. (2013), as in column 3 of Table 4, while Panel B is based on the conventional industry-level IV in Acemoglu et al. (2016). The five columns increase the number of shocks $J = 1, 5, 10, 25$, and 50, with only one shock relevant to treatment. The table reports the rejection rates corresponding to a nominal 5% level test of the true null that $\beta^* = 0$, the median bias of the estimator as a percentage of the simulated standard deviation, and the median first-stage F-statistic obtained via the Montiel Olea and Pflueger (2013) method (extended to shift-share IV in Panel A, following Appendix A.10). See Appendix A.11 for a complete discussion.

References

Acemoglu, Daron, David H. Autor, David Dorn, Gordon H. Hanson, and Brendard Price. 2016. "Import Competition and the Great U.S . Employment Sag of the 2000s." *Journal of Labor Economics* 34:S141–S198.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2020. "General Equilibrium Indirect Effects in Space: Theory and Measurement." *Working paper.*

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

Angrist, Joshua D., Kate Graddy, and Guido W. Imbens. 2000. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *The Review of Economic Studies* 67:499–527.

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

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

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

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *The Journal of Human Resources* 50:317–372.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110:2586–2624.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50:1029–1054.

Jaeger, David A, Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration." Working Paper.

Jones, Ronald W. 1965. "The Structure of Simple General Equilibrium Models." *Journal of Political Economy* 73:557–572.

Liu, Regina Y. 1988. "Bootstrap Procedures under Some Non-I.I.D. Models." *The Annals of Statistics* 16:1696–1708.

Montiel Olea, Jose Luis, and Carolin Pflueger. 2013. "A robust test for weak instruments." *Journal of Business and Economic Statistics* 31:358–369.

Nagar, Anirudh L. 1959. "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations." *Econometrica* 27:575–595.

Pflueger, Carolin E., and Su Wang. 2015. "A robust test for weak instruments in Stata." *Stata Journal* 15:216–225.

White, Halbert. 1982. "Instrumental Variables Regression with Independent Observations." *Econometrica* 50:483–499.