SHIFT-SHARE DESIGNS: THEORY AND INFERENCE[*]

Rodrigo Adão

Michal Kolesár[†]

Eduardo Morales

August 9, 2019

Abstract

We study inference in shift-share regression designs, such as when a regional outcome is regressed on a weighted average of sectoral shocks, using regional sector shares as weights. We conduct a placebo exercise in which we estimate the effect of a shift-share regressor constructed with randomly generated sectoral shocks on actual labor market outcomes across U.S. Commuting Zones. Tests based on commonly used standard errors with 5% nominal significance level reject the null of no effect in up to 55% of the placebo samples. We use a stylized economic model to show that this overrejection problem arises because regression residuals are correlated across regions with similar sectoral shares, independently of their geographic location. We derive novel inference methods that are valid under arbitrary cross-regional correlation in the regression residuals. We show using popular applications of shift-share designs that our methods may lead to substantially wider confidence intervals in practice.

JEL codes: C12, C21, C26, F16, F22

1 Introduction

We study how to perform inference in shift-share designs: regression specifications in which one studies the impact of a set of shocks, or "shifters", on units differentially exposed to them, with the exposure measured by a set of weights, or "shares". Specifically, shift-share regressions have the form

$$Y_i = \beta X_i + Z_i'\delta + \epsilon_i, \qquad \text{where} \quad X_i = \sum_{s=1}^{S} w_{is}\mathcal{X}_s, \quad w_{is} \geq 0 \text{ for all } s, \quad \text{and} \quad \sum_{s=1}^{S} w_{is} \leq 1. \qquad (1)$$

For example, in an investigation of the impact of sectoral demand shifters on regional employment changes, Y_i is the change in employment in region i, the shifter $\mathcal{X}_s$ is a measure of the change in demand for the good produced by sector s, and the share w_{is} may be measured as the initial share of region i's employment in sector s. Other observed characteristics of region i are captured by the vector Z_i, which includes the intercept, and ϵ_i is the regression residual. Shift-share specifications are increasingly common in many contexts (see, e.g., Bartik (1991), Blanchard and Katz (1992), Card (2001), or Autor, Dorn and Hanson (2013)). However, their formal properties are relatively understudied.

Our starting point is the observation that usual standard error formulas may substantially understate the true variability of OLS estimators of β in eq. (1). We illustrate the importance of this issue through a placebo exercise. As outcomes, we use 2000–2007 changes in employment rates and average wages for 722 Commuting Zones in the United States. We build a shift-share regressor by combining actual sectoral employment shares in 1990 with randomly drawn sector-level shifters for 396 4-digit SIC manufacturing sectors. The placebo samples thus differ exclusively in the randomly drawn sectoral shifters. For each sample, we compute the OLS estimate of β in eq. (1) and test if its true value is zero. Since the shifters are randomly generated, their true effect is indeed zero. Valid 5% significance level tests should therefore reject the null of no effect in at most 5% of the placebo samples. We find, however, that usual standard errors—clustering on state as well as heteroskedasticity-robust errors— are much smaller than the standard deviation of the OLS estimator and, as a result, lead to severe overrejection. Depending on the labor market outcome used, the rejection rate for 5% level tests can be as high as 55% for heteroskedasticity-robust standard errors and 45% for standard errors clustered on state, and it is never below 16%.

To explain the source of this overrejection problem, we introduce a stylized economic model featuring multiple regions, each of which produces output in multiple sectors. The key ingredients of our model are a sector- and region-specific labor demand and a regional labor supply. We assume that labor demand in each sector-region pair has a sector-specific elasticity with respect to wages and an intercept that aggregates several sector-specific components (e.g. sectoral productivities and demand shifters for the corresponding sectoral good). Labor supply in each region is upward-sloping and has a region-specific intercept that may aggregate group-specific labor supply shifters (e.g. push factors that raise immigration from different countries of origin). Up to a first-order approximation, the impact of sector-level shocks on labor market outcomes takes the form of a shift-share specification similar to that in eq. (1).

A key insight of our model is that the regression residual ϵ_i in eq. (1) will generally account for shift-share components that aggregate all unobserved sector-level shocks using the same shares

1

w_{is} that enter the construction of the regressor X_i, as well as shift-share components that aggregate unobserved group-specific labor supply shifters using exposures $\tilde{w}_{ig}$ of region i to group-g specific shocks. Thus, the residual may incorporate multiple shift-share terms with shares correlated with those defining the shift-share regressor X_i. Consequently, whenever two regions have similar shares, they will not only have similar exposure to the shifters $\mathcal{X}_s$, but will also tend to have similar values of the residuals ϵ_i. While traditional inference methods allow for some forms of dependence between the residuals, such as spatial dependence within a state, they do not directly address the possible dependence between residuals generated by unobserved shift-share components. This is why, in our placebo exercise, traditional inference methods underestimate the variance of the OLS estimator of β, creating the overrejection problem.

We then establish the large-sample properties of the OLS estimator of β in eq. (1) under repeated sampling of the shifters $\mathcal{X}_s$, conditioning on the realized shares w_{is}, controls Z_i, and residuals ϵ_i. This sampling approach is motivated by our economic model: we are interested in what would have happened to outcomes if the sector-level shocks $\mathcal{X}_s$ had taken different values, holding everything else constant. Our framework allows for heterogeneous effects of the shifters: one unit increase in $\mathcal{X}_s$ causes the outcome in region i to increase by $w_{is}\beta_{is}$, where β_{is} is an unknown parameter.

Our key assumption is that, conditional on the controls and the shares, the shifters are as good as randomly assigned and independent across sectors. An advantage of this assumption is that it allows us to do inference conditionally on ϵ_i; as a result, we can allow for *any* correlation structure of the regression residuals across regions.[1] In contrast, if, instead of assuming independence of the shifters across sectors, we modeled the correlation structure in the residual, as in the spatial econometrics literature (e.g. Conley, 1999) or in the interactive fixed effects literature (e.g. Bai, 2009; Gobillon and Magnac, 2016), the resulting inference would be sensitive to the validity of the modeling assumptions. We show that the regression estimand β in eq. (1) corresponds to a weighted average of the heterogeneous parameters β_{is} and derive novel confidence intervals that are valid in samples with many regions and sectors. We also derive an analogous formula when X_i is used as an instrument in an instrumental variables regression, which follows directly from the fact that the associated first-stage and reduced-form regressions take the form in eq. (1).

To gain intuition for our formula, it is useful to consider the special case in which each region is fully specialized in one sector (i.e. for every i, $w_{is} = 1$ for some sector s). In this case, our procedure is identical to using the usual clustered standard error formula, but with clusters defined as groups of regions specialized in the same sector. This is in line with the rule of thumb that one should "cluster" at the level of variation of the regressor of interest. In the general case, our standard error formula essentially forms sectoral clusters, the variance of which depends on the variance of a weighted sum of the regression residuals ϵ_i, with weights that correspond to the shares w_{is}.

We extend our baseline results in three ways. We provide versions of our standard errors that only require the shifters to be independent across "clusters" of sectors, allowing for arbitrary correlation

---

[1]This is similar to the insight in Barrios et al. (2012), who consider cross-section regressions estimated at an individual level when the variable of interest varies only across groups of individuals. They show that, as long as the regressor of interest is as good as randomly assigned and independent across the groups, standard errors clustered on groups are valid under any correlation structure of the residuals.

among sectors belonging to the same "cluster." We also show how to apply our framework to panel data settings in which we have multiple observations of each region over time. Finally, we cover applications in which the shifter is unobserved, but can be estimated using observable local shocks.

We illustrate the finite-sample properties of our novel inference procedure in the same placebo exercise that we use to show the bias of the usual standard error formulas. Our new formulas give a good approximation to the variability of the OLS estimator across the placebo samples; consequently, they yield rejection rates that are close to the nominal significance level. As predicted by the theory, our standard error formula remains accurate under alternative distributions of both the shifters and the regression residuals. When the number of sectors is small or there is a sector that is significantly larger than the rest, our method overrejects, although the overrejection is milder in comparison with the usual standard error formulas. If the shifters are not independent across sectors, we show that it is important to properly account for their correlation structure.

In the final part of the paper, we illustrate the implications of our new inference procedure for two popular applications of shift-share regressions. First, we study the effect of changes in sector-level Chinese import competition on labor market outcomes across U.S. Commuting Zones, as in Autor, Dorn and Hanson (2013). Second, we use changes in sector-level national employment to estimate the regional inverse labor supply elasticity, as in Bartik (1991).[2] Our new confidence intervals for the effects of Chinese competition on local labor markets increase by 23%–66% relative to those implied by state-clustered or heteroskedasticity-robust standard errors, although these effects remain statistically significant. In contrast, our confidence intervals for the inverse labor supply elasticity estimated using the procedure in Bartik (1991) are very similar to those constructed using standard approaches.

Shift-share designs have been applied to estimate the effect of a wide range of shocks. For example, in seminal papers, Bartik (1991) and Blanchard and Katz (1992) use shift-share designs to analyze the impact on local labor markets of shifters measured as changes in national sectoral employment. More recently, shift-share strategies have been applied to investigate the local labor market impact of various shocks, including international trade competition (Topalova, 2007, 2010; Kovak, 2013; Autor, Dorn and Hanson, 2013; Dix-Carneiro and Kovak, 2017; Pierce and Schott, 2018), credit supply (Greenstone, Mas and Nguyen, 2015), technological change (Acemoglu and Restrepo, 2019, 2018), and industry reallocation (Chodorow-Reich and Wieland, 2018). Shift-share regressors have been used as well to estimate the impact of immigration on labor markets, as in Card (2001) and many other papers following his approach; see reviews in Lewis and Peri (2015) and Dustmann, Schönberg and Stuhler (2016). Furthermore, recent papers use shift-share strategies to estimate how firms respond to changes in outsourcing costs and foreign demand (Hummels et al., 2014; Aghion et al., 2018).[3]

Our paper is related to two other papers studying the statistical properties of shift-share instrumental variables. First, Goldsmith-Pinkham, Sorkin and Swift (2018) consider using the full vector of

---

[2]Additionally, in Online Appendix F, we use changes in the stock of immigrants from various origin countries to investigate the impact of immigration on employment and wages, following Altonji and Card (1991) and Card (2001).

[3]Shift-share regressors have also been used to study the impact of sectoral shocks on political preferences (Autor et al., 2017; Che et al., 2017; Colantone and Stanig, 2018), marriage patterns (Autor, Dorn and Hanson, 2018), crime levels (Dix-Carneiro, Soares and Ulyssea, 2018), and innovation (Acemoglu and Linn, 2004; Autor et al., 2019). In addition to using shift-share designs to estimate the overall impact of a shifter of interest, other work has used them as part of a more general structural estimation approach; see Diamond (2016), Adão (2016), Galle, Rodríguez-Clare and Yi (2018), Burstein et al. (2018), Bartelme (2018). Baum-Snow and Ferreira (2015) review additional applications in the context of urban economics.

shares $(w_{i1}, \ldots, w_{iS})$ as an instrument for endogenous treatment. They conclude that this approach requires the entire vector of shares to be as good as randomly assigned conditional on the shifters. Second, Borusyak, Hull and Jaravel (2018), focusing on the use of a shift-share regressor as an instrument, show it is a valid instrument if the set of shifters is as good as randomly assigned conditional on the shares, and discuss consistency of the instrumental variables estimator in this context. We follow Borusyak, Hull and Jaravel (2018) by modeling the shifters as randomly assigned, since this approach follows naturally from our economic model. Using this assumption, we point out the potential bias of standard inference procedures when applied to shift-share designs, and provide a novel inference procedure that is valid in this context.

While our paper focuses on the statistical properties of the OLS estimator of β in eq. (1), there exists a prior literature that has focused on studying the validity of different economic interpretations that one may attach to the estimand β. For example, this prior literature has studied how this interpretation may be affected by the presence of cross-regional general equilibrium effects (Beraja, Hurst and Ospina, 2019; Adão, Arkolakis and Esposito, 2019), slow adjustment of labor market outcomes to the shifters $\mathcal{X}_s$ (Jaeger, Ruist and Stuhler, 2018), and heterogeneous effects of the shifters across sectors and regions (Monte, Redding and Rossi-Hansberg, 2018).

The rest of this paper is organized as follows. Section 2 presents a placebo exercise illustrating the properties of the usual inference procedures. Section 3 introduces a stylized economic model and maps its implications into a potential outcome framework. Section 4 establishes the asymptotic properties of the OLS estimator of β in eq. (1), as well as the properties of an instrumental variables estimator that uses a shift-share variable as an instrument. Section 5 discusses extensions of our baseline framework. Section 6 examines the performance of our novel inference procedures in a series of placebo exercises. Section 7 revisits two prior applications of shift-share designs, and Section 8 concludes. Proofs and additional results are collected in an Online Appendix.

2 Overrejection of usual standard errors: placebo evidence

In this section, we implement a placebo exercise to evaluate the finite-sample performance of the two inference methods most commonly applied in shift-share regression designs: (a) Eicker-Hubert-White—or heteroskedasticity-robust—standard errors, and (b) standard errors clustered on groups of regions geographically close to each other. In our placebo, we regress observed changes in U.S. regional labor market outcomes on a shift-share regressor that is constructed by combining actual data on initial sectoral employment shares for each region with randomly generated sector-level shocks. We describe the setup in Section 2.1 and discuss the results in Section 2.2.

2.1 Setup and Data

We generate $30,000$ placebo samples indexed by m. Each of them contains $N = 722$ regions and $S = 396$ sectors. We identify each region i with a U.S. Commuting Zone (CZ) and each sector s with a 4-digit SIC manufacturing industry.

Using the notation from eq. (1), the shares $\{w_{is}\}_{i=1,s=1}^{N,S}$, and the outcomes $\{Y_i\}_{i=1}^{N}$ are identical

in each placebo sample. The shares correspond to employment shares in 1990, and the outcomes correspond to changes in employment rates and average wages for different subsets of the population between 2000 and 2007. Our source of data on employment shares is the County Business Patterns, and our measures of changes in employment rates and average wages are based on data from the Census Integrated Public Use Micro Samples in 2000 and the American Community Survey for 2006 through 2008. Given these data sources, we construct our variables following the procedure described in the Online Appendix of Autor, Dorn and Hanson (2013).

The placebo samples differ exclusively in the shifters $\{\mathcal{X}_s^m\}_{s=1}^N$, which are drawn i.i.d. from a normal distribution with zero mean and variance equal to five in each placebo sample m. Since the shifters are independent of both the outcomes and the shares, the parameter β is zero; this is true irrespective of the dependence structure between the outcomes and the shares.

For each placebo sample m, given the observed outcome Y_i, the generated shift-share regressor X_i^m and a vector of controls Z_i including only an intercept, we compute the OLS estimate of β, the heteroskedasticity-robust standard error (which we label *Robust*), and the standard error that clusters CZs in the same state (labeled *Cluster*).

2.2 Results

Table 1 presents the median and standard deviation of the empirical distribution of the OLS estimates of β across the 30,000 placebo samples, along with the median standard error estimates, and rejection rates for 5% significance level tests of the null hypothesis $H_0: \beta = 0$. We present these statistics for several outcome variables, which are listed in the leftmost column.

Column (1) of Table 1 shows that, up to simulation error, the average of the OLS estimates is zero for all outcomes. Column (2) reports the standard deviation of the estimated coefficients. This dispersion is the target of the estimators of the standard error of the OLS estimator.[4] Columns (3) and (4) report the median standard error estimates for the *Robust* and *Cluster* procedures, respectively, and show that both standard error estimators are downward biased. On average across all outcomes, the median magnitudes of the heteroskedasticity-robust and state-clustered standard errors are, respectively, 55% and 46% lower than the standard deviation.

The downward bias in the *Robust* and *Cluster* standard errors translates into a severe overrejection of the null hypothesis $H_0: \beta = 0$. Since the true value of β equals 0 by construction, a correctly behaved test with significance level 5% should have a 5% rejection rate. Columns (5) and (6) in Table 1 show that traditional standard error estimators yield much higher rejection rates. For example, when the outcome variable is the CZ's employment rate, the rejection rate is 48.5% and 38.1% when *Robust* and *Cluster* standard errors are used, respectively. These rejection rates are very similar when the dependent variable is instead the change in the average log weekly wage.

These results are quantitatively important. To see this, consider the following thought-experiment. Suppose we were to provide the 30,000 simulated samples to 30,000 researchers without disclosing the origin of the data to them. Instead, we would tell them that the shifters correspond to changes in a

---

[4]Figure D.1 in Online Appendix D.1 reports the empirical distribution of the OLS estimates when the dependent variable is the change in each CZ's employment rate. Its distribution resembles a normal distribution centered around $\beta = 0$.

Table 1: Standard errors and rejection rate of the hypothesis $H_0\colon \beta = 0$ at 5% significance level.

	Estimate		Median std. error		Rejection rate	
	Mean (1)	Std. dev. (2)	Robust (3)	Cluster (4)	Robust (5)	Cluster (6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	0.73	0.92	48.5%	38.1%
Employed in manufacturing	−0.01	1.88	0.60	0.76	55.7%	44.8%
Employed in non-manufacturing	0.00	0.94	0.58	0.67	23.2%	17.6%
Panel B: Change in average log weekly wage						
Employed	−0.03	2.66	1.01	1.33	47.3%	34.2%
Employed in manufacturing	−0.03	2.92	1.68	2.11	26.7%	16.8%
Employed in non-manufacturing	−0.02	2.64	1.05	1.33	45.4%	33.7%

Notes: For the outcome variable indicated in the leftmost column, this table indicates the mean and standard deviation of the OLS estimates of β in eq. (1) across the placebo samples (columns (1) and (2)), the median standard error estimates (columns (3) and (4)), and the percentage of placebo samples for which we reject the null hypothesis $H_0\colon \beta = 0$ using a 5% significance level test (columns (5) and (6)). *Robust* is the Eicker-Huber-White standard error, and *Cluster* is the standard error that clusters CZs in the same state. Results are based on 30,000 placebo samples.

sectoral shock of interest—for instance, trade flows, tariffs, or national employment. If the researchers set out to test the null that the impact of this shock is zero using standard inference procedures at a 5% significance level, then over a third of them would conclude that our computer generated shocks had a statistically significant effect on the evolution of employment rates between 2000 and 2007.

The following remark summarizes the results of our placebo exercise.

Remark 1. *In shift-share regressions, traditional inference methods may suffer from a severe overrejection problem, and yield confidence intervals that are too short.*

To understand the source of this overrejection problem, note that the standard error estimators reported in Table 1 assume that the regression residuals are either independent across all regions (for *Robust*), or between geographically defined groups of regions (for *Cluster*). Given that shift-share regressors are correlated across regions with similar employment shares $\{w_{is}\}_{s=1}^S$, these methods generally lead to a downward bias in the standard error estimate whenever regions with similar employment shares $\{w_{is}\}_{s=1}^S$ also have similar regression residuals. In the next section, we show how such correlations between regression residuals may arise.

3   Stylized economic model

This section presents a stylized economic model mapping labor demand and labor supply shocks to labor market outcomes for a set of regional economies. The aim of the model is twofold. First, it illustrates the economic mechanisms behind the overrejection problem documented in Section 2.2. Second, it provides guidance on how to estimate: (i) the impact of sector-specific labor demand shifters on regional labor market outcomes; and (ii) the regional inverse labor supply elasticity. We

describe the model fundamentals in Section 3.1, discuss its main implications in Section 3.2, and map these implications to a potential outcome framework in Section 3.3.

3.1 Environment

We consider an economy with multiple sectors $s = 1, \ldots, S$ and multiple regions $i = 1, \ldots, N$. We assume that the labor demand in sector s and region i, L_{is}, is given by

$$\log L_{is} = -\sigma_s \log \omega_i + \log D_{is}, \qquad \sigma_s > 0, \tag{2}$$

where ω_i is the wage rate in region i, σ_s is the labor demand elasticity in sector s, and D_{is} is a region- and sector-specific labor demand shifter. This shifter may account for multiple sectoral components. Specifically, we decompose D_{is} into a sectoral shifter of interest χ_s, other shifters that vary by sector μ_s, and a residual region- and sector-specific shifter η_{is}:

$$\log D_{is} = \rho_s \log \chi_s + \log \mu_s + \log \eta_{is}. \tag{3}$$

We assume that the labor supply in region i is given by

$$\log L_i = \phi \log \omega_i + \log v_i, \qquad \phi > 0, \tag{4}$$

where ϕ is the labor supply elasticity, and v_i is a region-specific labor supply shifter. We allow this shifter to have a shift-share structure that yields region-specific aggregates of group-specific labor supply shocks. In particular, indexing labor groups by $g = 1, \ldots, G$, we decompose

$$\log v_i = \sum_{g=1}^{G} \tilde{w}_{ig} \log \nu_g + \log \upsilon_i, \tag{5}$$

where ν_g is a group-specific labor supply shifter, $\tilde{w}_{ig}$ measures the exposure of region i to group g labor supply shifter, and υ_i captures region-specific factors affecting labor supply. The variable ν_g captures factors that affect the supply of labor of group g in all regions in the population of interest. Workers may be classified into groups according to their education level, gender, or country of origin.

We assume that workers cannot move across regions but are freely mobile across sectors. Thus, labor markets clear if

$$L_i = \sum_{s=1}^{S} L_{is}, \qquad i = 1, \ldots, N. \tag{6}$$

3.2 Labor market equilibrium

We assume that, in each period, the model described by eqs. (2) to (6) characterizes the labor market equilibrium in every region, and that, across periods, changes in the labor market outcomes $\{\omega_i, L_i\}_{i=1}^{N}$ are due to changes in either the labor demand shifters, $\{\chi_s, \mu_s\}_{s=1}^{S}$ and $\{\eta_{is}\}_{i=1,s=1}^{N,S}$, or the labor supply shifters, $\{\nu_g\}_{g=1}^{G}$ and $\{\upsilon_i\}_{i=1}^{N}$.

We use $\hat{z} = \log(z^t/z^0)$ to denote log-changes in a variable z between a period $t = 0$ and some other period t. We assume that the realized changes between any two periods in all labor demand and supply shifters are draws from a joint distribution $F(\cdot)$:

$$\left(\{\hat{\chi}_s, \hat{\mu}_s\}_{s=1}^S, \{\hat{\eta}_{is}\}_{i=1,s=1}^{N,S}, \{\hat{v}_g\}_{g=1}^G \{\hat{v}_i\}_{i=1}^N\right) \sim F(\cdot). \tag{7}$$

Up to a first-order approximation around the initial equilibrium, eqs. (2) to (6) imply that the changes in employment and wages in region i are given by

$$\hat{L}_i = \sum_{s=1}^S l_{is}^0(\theta_{is}\hat{\chi}_s + \lambda_i\hat{\mu}_s + \lambda_i\hat{\eta}_{is}) + (1 - \lambda_i)\left(\sum_{g=1}^G \tilde{w}_{ig}\hat{v}_g + \hat{v}_i\right), \tag{8}$$

$$\hat{\omega}_i = \phi^{-1}\sum_{s=1}^S l_{is}^0(\theta_{is}\hat{\chi}_s + \lambda_i\hat{\mu}_s + \lambda_i\hat{\eta}_{is}) - \phi^{-1}\lambda_i\left(\sum_{g=1}^G \tilde{w}_{ig}\hat{v}_g + \hat{v}_i\right), \tag{9}$$

where $l_{is}^0 = L_{is}^0/L_i^0$ is the initial employment share of sector s in region i, $\lambda_i = \phi\left[\phi + \sum_{s=1}^S l_{is}^0\sigma_s\right]^{-1}$, and $\theta_{is} = \rho_s\lambda_i$.

Consider first the model's implications for the impact on regional labor market outcomes of changes in sector-specific labor demand. We focus here on the impact of the demand shocks $\{\hat{\chi}_s, \hat{\mu}_s\}_{s=1}^S$ on the change in the employment rate $\hat{L}_i$; however, given the symmetry between eqs. (8) and (9), the model's implications for the impact of these shocks on the change in the wage level $\hat{\omega}_i$ are analogous.

According to eq. (8), the change in the employment rate in region i depends on two shift-share components that aggregate the impact of the sector-specific labor demand shocks. In both components, the "share" term is the initial employment share l_{is}^0; the "shift" term corresponds in each of them to one of the two sector-specific labor demand shocks, $\hat{\chi}_s$ or $\hat{\mu}_s$. Furthermore, $\hat{L}_i$ also depends on additional shift-share terms that aggregate the impact of group-specific labor supply shocks. In this case, the "share" term is the region's exposure to each group-specific shock, $\tilde{w}_{ig}$. Conditional on a sector s and a labor group g, the shares $\{l_{is}^0\}_{i=1}^N$ and $\{\tilde{w}_{ig}\}_{i=1}^N$ may be correlated. Settings in which the outcome of interest depends on multiple shift-share terms with potentially correlated shares is central to understanding the placebo results presented in Section 2.

Another implication of eq. (8) is that, even conditional on the initial employment share l_{is}^0, the impact of sectoral labor demand shocks on regional employment may be heterogeneous across sectors and regions; e.g., the impact of $\hat{\chi}_s$ on $\hat{L}_i$ depends not only on l_{is}^0 but also on θ_{is}, which may vary across i and s. While datasets usually contain information on the initial employment shares for every sector and region $\{l_{is}^0\}_{i=1,s=1}^{N,S}$, the parameters $\{\theta_{is}\}_{i=1,s=1}^{N,S}$ are not generally known.

We summarize the discussion in the last two paragraphs in the following remark:

Remark 2. *In our model, the equilibrium equations for the change in regional labor market outcomes combines multiple shift-share terms, and the shifter effects depend on unknown parameters that may be heterogeneous.*

Online Appendices B and C show that there are multiple microfoundations consistent with the insights summarized in Remark 2. Alternative microfoundations may differ in the mapping between the labor demand and supply elasticities, σ_s and ϕ, and structural parameters, or in the interpre-

tation of the different terms entering the labor demand shifter D_{is} in eq. (3).[5] In addition, Online Appendix C.3 shows that similar insights arise in a model that allows for migration across regions. In this case, the change in regional employment depends not only on the region's own shift-share terms included in eq. (8), but also on a component, common to all regions, that combines the shift-share terms corresponding to all N regions. In this environment, $l_{is}^0 \theta_{is}$ is the partial effect of the shifter $\hat{\chi}_s$ on $\hat{L}_i$ conditional on a fixed effect that absorbs cross-regional spillovers created by migration.

Turning to the estimation of the inverse labor supply elasticity, eqs. (4) and (5) imply that

$$\hat{\omega}_i = \tilde{\phi} \hat{L}_i - \tilde{\phi} (\sum_{g=1}^{G} \tilde{w}_{ig} \hat{v}_g + \hat{v}_i) \quad \text{with} \quad \tilde{\phi} = \phi^{-1}. \tag{10}$$

It follows from eq. (8) that the change in region i's employment rate, $\hat{L}_i$, also depends on the term $\sum_{g=1}^{G} \tilde{w}_{ig} \hat{v}_g + \hat{v}_i$. Thus, the two terms on the right-hand side of eq. (10) are correlated with each other, creating an endogeneity problem. The instrumental variables solution to this problem relies on the observation that using eqs. (8) and (9), one can write the inverse labor supply elasticity as the ratio of the impact of a sector-specific labor demand shock (e.g. $\hat{\chi}_s$) on wages to that on employment:

$$\tilde{\phi} = \frac{\partial \hat{\omega}_i}{\partial \hat{\chi}_s} \bigg/ \frac{\partial \hat{L}_i}{\partial \hat{\chi}_s}.$$

In Sections 4 and 5, we use the model described here to provide an economic interpretation for the econometric assumptions we impose when discussing identification and estimation in shift-share designs. These assumptions imply restrictions on the distribution of labor supply and demand shocks $F(\cdot)$ introduced in eq. (7). In Section 7, we return to this economic model when interpreting empirical estimates of the impact of sector-specific labor demand shifters on regional labor market outcomes (Section 7.1); and the regional inverse labor supply elasticity (Section 7.2).

3.3 From economic model's equilibrium conditions to a potential outcome framework

We build on the results in Section 3.2 to propose a general framework for the estimation of the impact of shifters on outcomes measured at a different unit of observation. For concreteness, we refer to the level at which shifters vary as sectors and to the level at which the outcome varies as regions.

To make precise what we mean by "the effect of shifters on an outcome", we use the potential outcomes notation, writing $Y_i(x_1, \ldots, x_S)$ to denote the potential (counterfactual) outcome that would occur in region i if the shocks to the S sectors were exogenously set to $\{x_s\}_{s=1}^S$. Consistently with eqs. (8) and (9), we assume that the potential outcomes are linear in the shocks,

$$Y_i(x_1, \ldots, x_S) = Y_i(0) + \sum_{i=1}^{S} w_{is} x_s \beta_{is}, \qquad \text{where} \qquad w_{is} \geq 0 \text{ for all } s, \qquad \sum_{s=1}^{S} w_{is} \leq 1, \tag{11}$$

---

[5]In Online Appendix B, we derive eqs. (8) and (9) from a multisector gravity model with endogenous labor supply that follows closely that in Adão, Arkolakis and Esposito (2019). In Online Appendix C.1, we show that Remark 2 is consistent with a Jones (1971) model featuring sector-specific production inputs, as in Kovak (2013). In Online Appendix C.2, we show that it is also consistent with a Roy (1951) model featuring workers with heterogeneous preferences for employment across sectors, as in Galle, Rodríguez-Clare and Yi (2018), Lee (2018) and Burstein, Morales and Vogel (2019).

and $Y_i(0) = Y_i(0, \ldots, 0)$ denotes the potential outcome in region i when all shocks $\{x_s\}_{s=1}^{S}$ are set to zero. Thus, increasing x_s by one unit, holding the shocks to the other sectors constant, leads to an increase in region i's outcome of $w_{is}\beta_{is}$ units. This is the treatment effect of x_s on $Y_i(x_1, \ldots, x_S)$. The actual (observed) outcome is given by $Y_i = Y_i(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, which depends on the realization of the shifters, $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$.

If the shifters of interest are the sectoral labor demand shocks $\{\hat{\chi}_s\}_{s=1}^{S}$, and the outcome of interest is the employment change $\hat{L}_i$, we can map eq. (8) into eq. (11) by defining

$$Y_i = \hat{L}_i, \ w_{is} = l_{is}^0, \ x_s = \hat{\chi}_s, \ \beta_{is} = \theta_{is}, \ Y_i(0) = \lambda_i \sum_{s=1}^{S} w_{is}(\hat{\mu}_s + \hat{\eta}_{is}) + (1 - \lambda_i)(\sum_{g=1}^{G} \tilde{w}_{ig}\hat{v}_g + \hat{v}_i). \quad (12)$$

Observe that $Y_i(0)$ aggregates all shifters other than the sectoral shocks of interest $\{\hat{\chi}_s\}_{s=1}^{S}$.[6]

We are interested in the properties of the OLS estimator $\hat{\beta}$ of the coefficient on the shift-share regressor $X_i = \sum_{s=1}^{S} w_{is}\mathcal{X}_s$ in a regression of Y_i onto X_i.[7] To focus on the key conceptual issues, we abstract away from any additional covariates or controls for now, and assume that $\mathcal{X}_s$ and Y_i have been demeaned, so that we can omit the intercept in a regression of Y_i on X_i (see Section 4.2 for the case with controls). In this simplified setting, the OLS estimator of the coefficient on X_i is given by

$$\hat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2}, \quad (13)$$

and we can write the regression equation as

$$Y_i = \beta X_i + \epsilon_i, \qquad \text{where} \quad X_i = \sum_{s=1}^{S} w_{is}\mathcal{X}_s. \quad (14)$$

The definition of the estimand β in eq. (14) and the properties of the estimator $\hat{\beta}$ will depend on: (a) what is the population of interest; and (b) how we think about repeated sampling. For (a), we define the population of interest to be the observed set of N regions, as opposed to focusing on a large superpopulation of regions from which the N observed regions are drawn. Consequently, we are interested in the parameters $\{\beta_{is}\}_{i=1,s=1}^{N,S}$ and the treatment effects $\{w_{is}\beta_{is}\}_{i=1,s=1}^{N,S}$ themselves, rather than the distributions from which they are drawn, which would be the case if we were interested in a superpopulation of regions.[8] For (b), given our interest on estimating the *ceteris paribus* impact of a specific set of shocks $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, we consider repeated sampling of these shocks, while holding the shares $\{w_{is}\}_{i=1,s=1}^{N,S}$, the parameters $\{\beta_{is}\}_{i=1,s=1}^{N,S}$, and the potential outcomes $\{Y_i(0)\}_{i=1}^{N}$ fixed.

---

[6] Given the mapping in eq. (12), the expression in eq. (11) captures the first-order impact of the labor demand shocks $\{\hat{\chi}_s\}_{s=1}^{S}$ on changes in the employment rate. We focus on this first-order impact because it helps connecting our analysis to linear specifications used extensively in the shift-share literature. See Online Appendix D.5 for a discussion of the approximation error arising from the linear specification imposed in eq. (8).

[7] We assume for now that the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$ are directly observable. In Section 5.3, we consider the case in which we only observe noisy estimates of these shifters.

[8] Treating the set of observed regions as the population of interest is common in applications of the shift-share approach. For example, the abstract of Autor, Dorn and Hanson (2013) reads: "We analyze the effect of rising Chinese import competition between 1990 and 2007 *on U.S. local labor markets*". Similarly, the abstract of Dix-Carneiro and Kovak (2017) reads: "We study the evolution of trade liberalization's effects *on Brazilian local labor markets*" (emphases added).

Given these assumptions, the estimand β is defined as the population analog of eq. (13) under repeated sampling of the shocks $\mathcal{X}_s$,

$$\beta = \frac{\sum_{i=1}^{N} E[X_i Y_i \mid \mathcal{F}_0]}{\sum_{i=1}^{N} E[X_i^2 \mid \mathcal{F}_0]}, \qquad \text{with} \qquad \mathcal{F}_0 = \{Y_i(0), \beta_{is}, w_{is}\}_{i=1,s=1}^{N,S}, \tag{15}$$

and, given eqs. (11) and (14), the regression error ϵ_i is then defined as the residual

$$\epsilon_i = Y_i - X_i \beta = Y_i(0) + \sum_{i=1}^{S} w_{is} \mathcal{X}_s (\beta_{is} - \beta). \tag{16}$$

Thus, the statistical properties of the regression residual ϵ_i depend on the properties of the potential outcome $Y_i(0)$, the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$, the shares $\{w_{is}\}_{s=1}^{S}$, and the difference between the parameters $\{\beta_{is}\}_{s=1}^{S}$ and the estimand β. Importantly, as illustrated in eq. (12), the potential outcome $Y_i(0)$ will generally incorporate terms that have a shift-share structure with shares that are either identical to (e.g. the term $\sum_{s=1}^{S} w_{is} \hat{\mu}_s$) or different from but potentially correlated with (e.g. the term $\sum_{g=1}^{G} \tilde{w}_{ig} \hat{v}_g$) the shares $\{w_{is}\}_{s=1}^{S}$ that define the shift-share regressor X_i. It then follows from eq. (16) that the residuals ϵ_i and $\epsilon_{i'}$ will generally be correlated for any pair of regions i and i' with similar values of the shift-share regressor.

We summarize this discussion in the following remark.

Remark 3. *Correct inference for the coefficient on a shift-share regressor requires taking into account potential cross-regional correlation in residuals across observations with similar values of the shift-share covariate of interest. One possible source of such correlation is the presence in these residuals of shift-share components with shares identical to or correlated with those entering the covariate of interest.*

Remark 3 has important implications for estimating the sampling variability of $\hat{\beta}$. In particular, traditional inference procedures do not account for correlation in ϵ_i among regions with similar shares and, therefore, tend to underestimate the variability of $\hat{\beta}$. As we formalize in the next section, this is the main reason for the overrejection problem described in Section 2.

4 Asymptotic properties of shift-share regressions

In this section, we formulate the statistical assumptions that we impose on the data generating process (DGP), use them to derive asymptotic results, and provide an economic interpretation of these assumptions using the model introduced in Section 3. In Section 4.1, we consider the case in which there is a single shift-share regressor and no controls. We account for controls in Section 4.2. In Section 4.3, we consider using the shift-share variable as an instrument for a regional treatment variable. All proofs and technical details are collected in Online Appendix A.

We follow the notation from eq. (1) by writing sector-level variables (such as the shifter $\mathcal{X}_s$) in script font style and region-level aggregates (such as X_i) in normal style. We use standard matrix and vector notation. In particular, for a (column) L-vector A_i that varies at the regional level, A denotes the $N \times L$ matrix with the ith row given by A_i'. For an L-vector $\mathcal{A}_s$ that varies at the sectoral level, $\mathcal{A}$

denotes the $S \times L$ matrix with the sth row given by $\mathscr{A}'_s$. If $L = 1$, then A and $\mathscr{A}$ are an N-vector and an S-vector, respectively. Let W denote the $N \times S$ matrix of shares, so that its (i, s) element is given by w_{is}, and let B denote the $N \times S$ matrix with (i, s) element given by β_{is}.

4.1 Simple case without controls

We focus here on the statistical properties of the OLS estimator $\hat{\beta}$ defined in eq. (13).

Assumptions

We consider large-sample properties of $\hat{\beta}$ as the number sectors goes to infinity, $S \to \infty$. The assumptions below imply that $N \to \infty$ as $S \to \infty$. To assess how large S needs to be in order that these asymptotics provide a good approximation to the finite sample distribution of $\hat{\beta}$, we conduct a series of placebo simulations in Section 6. We describe here the main substantive assumptions, and collect technical regularity conditions in Online Appendix A.1.1. As in eq. (15), let $\mathscr{F}_0 = (Y(0), B, W)$.

Assumption 1 (Identification). (i) The observed outcome is given by $Y_i = Y_i(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, such that eq. (11) holds; (ii) The shifters are as good as randomly assigned conditional on $\mathscr{F}_0$ in the sense that, for all $s = 1, \ldots, S$,

$$E[\mathcal{X}_s \mid \mathscr{F}_0] = 0. \tag{17}$$

Assumption 1(i) requires that the potential outcomes are linear in the shifters $\{\mathcal{X}_s\}_{s=1}^S$. As discussed in Section 3.3, one can generate such linear specification from a first-order approximation of the impact of the shifters $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ on the outcome Y_i. This approximation may be subject to error. In Online Appendix A.1.1, we generalize eq. (11) to allow for a linearization error and derive restrictions on this error under which our inference procedures remain valid.

Assumption 1(ii) imposes that the sectoral shifters $\mathcal{X}$ are mean independent of the shares W, potential outcomes $Y(0)$, and parameters B; the assumption that the shifters are mean zero is a normalization to allow us to drop the intercept; we relax it in Section 4.2. This random assignment assumption is a key assumption for identifying the causal impact of a shift-share covariate; a version of this assumption has been previously proposed by Borusyak, Hull and Jaravel (2018).

If we are interested in studying the effect of labor demand shifters in the context of the model in Section 3 (i.e. $\mathcal{X}_s = \hat{\chi}_s$), Assumption 1(ii) will hold if the shifters $\{\hat{\chi}_s\}_{s=1}^S$ are mean independent of the other labor demand shifters, $\{\hat{\mu}_s\}_{s=1}^S$ and $\{\hat{\eta}_{is}\}_{i=1,s=1}^{N,S}$, and of the labor supply shifters, $\{\hat{v}_g\}_{g=1}^G$ and $\{\hat{v}_i\}_{i=1}^N$. The plausibility of this restriction depends on the specific empirical application. For example, if all N regions in the sample are regions within a small open economy, $\hat{\chi}_s$ denotes changes in international prices in sector s, and $\hat{\mu}_s$ denotes changes in the tariffs that this small open economy charges on its sector s imports; then, Assumption 1(ii) requires these changes in tariffs to be independent of the changes in tariffs in any country that is large enough for their tariff changes to affect international prices (see Online Appendix B.4 for additional details).

Assumption 2 (Consistency and Inference). (i) The shifters $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ are independent conditional on $\mathscr{F}_0$; (ii) $\max_s n_s / \sum_{t=1}^S n_t \to 0$, where $n_s = \sum_{s=1}^S w_{is}$ denotes the total share of sector s; (iii) $\max_s n_s^2 / \sum_{t=1}^S n_t^2 \to 0$.

Assumption 2(i) requires the shifters to be independent. It adapts to our setting the assumption underlying randomization-style inference in randomized controlled trials that the treatment assignment is independent across entities (see Imbens and Rubin, 2015, for a review). An independence or a weak dependence assumption of this type is generally necessary in order to do inference.[9] One could alternatively impose assumptions on the correlation structure of the regression residuals, either by imposing a particular structure on them, as in the literature on interactive fixed effects (e.g. Gobillon and Magnac, 2016), or by imposing a distance metric on the observations, as in the spatial econometrics literature (e.g. Conley, 1999). However, as the economic model in Section 3 shows, the structure of the residuals may be very complex. The residuals may include potentially correlated region-specific terms as well as several shift-share terms, which may or may not use the same shares as the covariate of interest X_i. It is thus difficult to conceptualize which exact restriction on their joint distribution one should impose.

By instead imposing restrictions on the distribution of the vector of shifters $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ conditional on $\mathcal{F}_0 = (Y(0), B, W)$, Assumption 2(i) ensures that the standard errors we derive remain valid under *any* dependence structure between the shares w_{is} across sectors and regions, and under *any* correlation structure of the potential outcomes $Y_i(0)$ or, equivalently, of the regression errors ϵ_i, across regions.[10] We thus do not have to worry about correctly specifying this correlation structure, as one would under the alternative approaches mentioned above. Our approach allows (but does not require) the residual to have a shift-share structure; it similarly allows all $\{w_{is}\}_{i=1,s=1}^{N,S}$ to be equilibrium objects responding to the same economic shocks, and thus be correlated across regions and sectors.[11] In Section 5.1, we relax Assumption 2(i) and allow for a non-zero correlation in the shifters $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ within clusters of sectors; we only require that the shifters are independent across the clusters. Additionally, in the context of the empirical application in Section 7.1, we discuss how to perform inference in a setting in which all shifters of interest are generated by a common shock that has heterogeneous effects across sectors.

In the economic model in Section 3, if $\mathcal{X}_s = \hat{\chi}_s$ and we interpret these shocks as, for example, sector-specific productivity shocks, Assumption 2(i) requires that there is no common component driving the changes in sectoral productivities. Our approach does not require the shifters $\{\mathcal{X}_s\}_{s=1}^S$ to be identically distributed; we allow, for example, the variance of the shock to differ across sectors.

Assumptions 2(ii) and 2(iii) are our main regularity conditions.[12] Assumption 2(ii) is needed for

---

[9]For example, for inference on average treatment effects, which is commonly the goal when running a regression, one typically assumes that the sample is a random sample from the population of interest and, thus, that the treatment variable is independent across the individuals in the sample.

[10]Since our inference is valid conditional on $\{\epsilon_i\}_{i=1}^N$, it accounts for any correlation structure they may have, including spatial, or, in applications with multiple periods, temporal correlations. See Section 5.2 for settings with multiple periods.

[11]This conceptualization of all the shares w_{is} as equilibrium objects that respond (at least partly) to the same set of shocks is consistent with the model in Section 3. As shown in eq. (12), each share w_{is} corresponds to the share of workers in region i employed in sector s in an initial equilibrium, l_{is}^0. Furthermore, each of these initial employment shares will be a function of the same sector-specific demand shocks and group-specific labor supply shocks; consequently l_{is}^0 will generally be correlated with $l_{i's'}^0$ even for $i \neq i'$ and $s \neq s'$.

[12]In the context of a shift-share instrumental variables regression, Goldsmith-Pinkham, Sorkin and Swift (2018) discuss similar conditions stated in terms of Rotemberg weights. This is convenient under the baseline assumption considered in Goldsmith-Pinkham, Sorkin and Swift (2018) that the vector of shares $(w_{i1}, \ldots, w_{iS})$ is exogenous, because the Rotemberg weights determine the asymptotic bias of the estimator under local failures of this exogeneity condition. Since we do not assume exogeneity of the shares, this interpretation is not available under our setup.

consistency: it requires that the size of each sector, n_s, is asymptotically negligible. This assumption is analogous to the standard consistency condition in the clustering literature that the largest cluster be asymptotically negligible. To see the connection, consider the special case with "concentrated sectors", in which each region i specializes in one sector $s(i)$; i.e. $w_{is} = 1$ if $s = s(i)$ and $w_{is} = 0$ otherwise, and n_s is thus the number of regions that specialize in sector s. In this case, $X_i = \mathcal{X}_{s(i)}$, so that, if eq. (17) holds, $\hat{\beta}$ is equivalent to an OLS estimator in a randomized controlled trial in which the treatment varies at a cluster level; here the sth cluster consists of regions that specialize in sector s. The condition $\max_s n_s / \sum_{t=1}^{S} n_t \to 0$ then reduces to the assumption that the largest cluster be asymptotically negligible. Assumption 2(iii) is needed for asymptotic normality—it ensures that the Lindeberg condition holds. It strengthens Assumption 2(ii) slightly by requiring that the contribution of each sector to the asymptotic variance is asymptotically negligible; otherwise the estimator will not generally be asymptotically normal, even if it is consistent.

In terms of the economic model introduced in Section 3, Assumptions 2(ii) and 2(iii) require that no sector dominates the rest in terms of initial employment at the national level; i.e. $\sum_{i=1}^{N} l_{is}^0$ is not too large for any sector. Section 6.1 shows that this assumption is reasonable for the U.S. if the S sectors used to construct the treatment of interest X_i correspond to the 396 4-digit manufacturing sectors (see Section 2.1). In Section 6.2, we illustrate the consequences of the failure of this assumption due to the inclusion of a large aggregate sector, the non-manufacturing sector, in X_i.

Asymptotic theory

We now establish that the OLS estimator in eq. (13) is consistent and asymptotically normal.

Proposition 1. *Suppose Assumption 1, Assumptions 2(i) and 2(ii), and Assumptions A.1(i) to A.1(iii) in Online Appendix A.1.1 hold. Then*

$$\beta = \frac{\sum_{i=1}^{N} \sum_{s=1}^{S} \pi_{is} \beta_{is}}{\sum_{i=1}^{N} \sum_{s=1}^{S} \pi_{is}}, \qquad and \qquad \hat{\beta} = \beta + o_p(1), \tag{18}$$

where $\pi_{is} = w_{is}^2 \, \text{var}(\mathcal{X}_s \mid \mathcal{F}_0)$.

This proposition gives two results. First, it shows that the estimand β in eq. (15) can be expressed as a weighted average of the region- and sector-specific parameters $\{\beta_{is}\}_{i=1,s=1}^{N,S}$, with the weight π_{is} increasing in the share w_{is} and in the conditional variance of the shifter $\text{var}(\mathcal{X}_s \mid \mathcal{F}_0)$. Second, it states that the OLS estimator $\hat{\beta}$ converges to this estimand as $S \to \infty$. The special case with concentrated sectors is again useful in interpreting Proposition 1. In this case, $\sum_{s=1}^{S} \pi_{is} \beta_{is} = \text{var}(\mathcal{X}_{s(i)} \mid \mathcal{F}_0) \beta_{is(i)}$ and, therefore, the first result in Proposition 1 reduces to the standard result from the randomized controlled trials literature with cluster-level randomization (with each "cluster" defined as all regions specialized in the same sector) that the weights are proportional to the variance of the shock.

The estimand β *does not* in general equal a weighted average of the heterogeneous treatment effects. As discussed in Section 3.3, the effect on the outcome in region i of increasing the value of the sector s shock in one unit is equal to $w_{is} \beta_{is}$; weighting this effect using a set of region- and

sector-specific weights $\{\xi_{is}\}_{i=1,s=1}^{N,S}$, yields the weighted average treatment effect

$$\tau_\xi = \frac{\sum_{i=1}^N \sum_{s=1}^S \xi_{is} w_{is} \beta_{is}}{\sum_{i=1}^N \sum_{s=1}^S \xi_{is}}.$$

Alternatively, the total effect of increasing the shifters simultaneously in every sector by one unit is $\sum_{s=1}^S w_{is}\beta_{is}$; weighting it using a set of region-specific weights $\{\zeta_i\}_{i=1}^N$ yields the weighted to-tal treatment effect $\tau_\zeta^T = \sum_{i=1}^N \zeta_i \sum_{s=1}^S w_{is}\beta_{is} / \sum_{i=1}^N \zeta_i$. If β_{is} is constant across i and s, then $\beta = \tau_\zeta^T$, provided $\sum_{s=1}^S w_{is} = 1$ in every region i; otherwise, we can consistently estimate τ_ζ^T by $\hat\beta \cdot \sum_{i=1}^N \zeta_i \sum_{s=1}^S w_{is} / \sum_{i=1}^N \zeta_i$. Similarly, if β_{is} is constant across i and s, τ_ξ is consistently estimated by $\hat\beta \cdot \sum_{i=1}^N \sum_{s=1}^S \xi_{is} w_{is} / \sum_{i=1}^N \sum_{s=1}^S \xi_{is}$. On the other hand, if β_{is} varies across regions and sectors, then it is not clear in general how to exploit knowledge of the estimand β defined in eq. (18) to learn something about τ_ξ or τ_ζ^T. A special case in which it is possible to consistently estimate τ_ξ even if β_{is} varies across i or s arises when $\mathcal{X}_s$ is homoskedastic, $\mathrm{var}(\mathcal{X}_s \mid \mathcal{F}_0) = \sigma^2$, and $\xi_{is} = w_{is}$; in this case, a consistent estimate of τ_ξ is given by $\hat\beta \sum_{i=1}^N \sum_{s=1}^S w_{is}^2 / \sum_{i=1}^N \sum_{s=1}^S w_{is}$.[13]

Proposition 2. *Suppose Assumptions 1 and 2, and Assumption A.1 in Online Appendix A.1.1 hold. Suppose also that*

$$V_N = \frac{1}{\sum_{s=1}^S n_s^2} \mathrm{var}\left(\sum_{i=1}^N X_i \epsilon_i \mid \mathcal{F}_0\right)$$

converges in probability to a non-random limit. Then

$$\frac{N}{\sqrt{\sum_{s=1}^S n_s^2}} (\hat\beta - \beta) = n\left(0, \frac{V_N}{\left(\frac{1}{N}\sum_{i=1}^N X_i^2\right)^2}\right) + o_p(1).$$

This proposition shows that $\hat\beta$ is asymptotically normal, with a rate of convergence equal to $N(\sum_{s=1}^S n_s^2)^{-1/2}$. If all sector sizes n_s are of the order N/S, the rate of convergence equals $\sqrt{S}$. How-ever, if the sizes are unequal, the rate may be slower.

According to Proposition 2, the asymptotic variance formula has the usual "sandwich" form. Since X_i is observed, to construct a consistent standard error estimate, it suffices to construct a consistent estimate of V_N, the middle part of the sandwich. To motivate our standard error formula, suppose that β_{is} is constant across i and s, $\beta_{is} = \beta$. Then it follows from eq. (17) and Assumption 2(i) that

$$V_N = \frac{\sum_{s=1}^S \mathrm{var}(\mathcal{X}_s \mid \mathcal{F}_0) R_s^2}{\sum_{s=1}^S n_s^2}, \qquad R_s = \sum_{i=1}^N w_{is}\epsilon_i. \tag{19}$$

---

[13]In general, one can consistently estimate τ_ξ or τ_ζ^T by imposing a mapping between β_{is} and structural parameters, and obtaining consistent estimates of these structural parameters. However, since this mapping will vary across models, the consistency of such estimator will not be robust to alternative modeling assumptions, even if all these assumptions predict an equilibrium relationship like that in eq. (8); e.g. see Online Appendix B and Online Appendices C.1 and C.2 for examples of this mapping in different models.

Replacing $\text{var}(\mathcal{X}_s \mid \mathcal{F}_0)$ by $\mathcal{X}_s^2$, and ϵ_i by the regression residual $\hat{\epsilon}_i = Y_i - X_i\hat{\beta}$, we obtain the estimate

$$\widehat{V}_{AKM}(\hat{\beta}) = \frac{\hat{\mathcal{V}}_{AKM}(\hat{\beta})}{\left(\sum_{i=1}^N X_i^2\right)^2}, \qquad \hat{\mathcal{V}}_{AKM}(\hat{\beta}) = \sum_{s=1}^S \mathcal{X}_s^2 \hat{R}_s^2, \qquad \hat{R}_s = \sum_{i=1}^N w_{is}\hat{\epsilon}_i. \tag{20}$$

When $\beta_{is} = \beta$, we show formally that this variance estimate leads to valid inference under regularity conditions in Section 4.2. In Online Appendix A.1.6 we show that this variance estimate remains valid under heterogeneous β_{is} under further regularity conditions.

To gain intuition for the variance estimate in eq. (20), consider the case with concentrated sectors. Then the numerator in eq. (20) becomes $\sum_{s=1}^S \mathcal{X}_s^2 \hat{R}_s^2 = \sum_{s=1}^S (\sum_{i=1}^N \mathbb{I}\{s(i) = s\} X_i \hat{\epsilon}_i)^2$, so that eq. (20) reduces to the cluster-robust variance estimate that clusters on the sector that each region is specialized. This is consistent with the rule of thumb that one should "cluster" at the level of variation of the regressor of interest. More generally, the variance estimate essentially forms sectoral clusters with variance that depends on the variance of $\hat{R}_s$, a weighted sum of the regression residuals $\{\hat{\epsilon}_i\}_{i=1}^N$, with weights that correspond to the shares $\{w_{is}\}_{i=1}^N$. An important advantage of $\widehat{V}_{AKM}(\hat{\beta})$ is that it allows for an arbitrary structure of cross-regional correlation in residuals:

Remark 4. *In the expression for $\mathcal{V}_N$ in eq. (19), the expectation is only taken over $\{\mathcal{X}_s\}_{s=1}^S$—we do not take any expectation over the shares $\{w_{is}\}_{i=1,s=1}^{N,S}$ or the residuals $\{\epsilon_i\}_{i=1}^N$. This is because our inference is conditional on the realized values of the shares and on the potential outcomes $\{Y_i(0)\}_{i=1}^N$. In terms of the regression in eq. (14), this means that we consider properties of $\hat{\beta}$ under repeated sampling of $X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$ conditional on the shares $\{w_{is}\}_{i=1,s=1}^{N,S}$ and on the residuals $\{\epsilon_i\}_{i=1}^N$ (as opposed to, say, considering properties of $\hat{\beta}$ under repeated sampling of the residuals conditional on $\{X_i\}_{i=1}^N$). As a result, our inference method allows for arbitrary dependence between the residuals $\{\epsilon_i\}_{i=1}^N$.*

To understand the source of the overrejection problem discussed in Section 2, let us compare the variance estimate $\widehat{V}_{AKM}(\beta)$ with the cluster-robust variance estimate when the residuals $\hat{\epsilon}_i$ are computed at the true β (so that $\hat{\epsilon}_i = \epsilon_i$). These variance estimates differ in the middle sandwich, with the cluster-robust estimate replacing $\hat{\mathcal{V}}_{AKM}(\beta)$ in eq. (20) with $\hat{\mathcal{V}}_{CL}(\beta) = \sum_{i=1}^N \sum_{j=1}^N \mathbb{I}\{c(i) = c(j)\} X_i X_j \epsilon_i \epsilon_j$, where $c(i)$ denotes the cluster that region i belongs to (the comparison with heteroskedasticity-robust standard errors obtains as a special case if $c(i) = i$, so that each region belongs to its own cluster). Assuming for simplicity that the conditional variance of $\mathcal{X}_s$ does not depend on $Y(0)$, it follows by simple algebra that the expectation of the difference between these terms is given by

$$E[\hat{\mathcal{V}}_{AKM}(\beta) - \hat{\mathcal{V}}_{CL}(\beta) \mid W] = \sum_{s=1}^S \text{var}(\mathcal{X}_s \mid W) \sum_{i=1}^N \sum_{j=1}^N \mathbb{I}\{c(i) \neq c(j)\} w_{is} w_{js} E[\epsilon_i \epsilon_j \mid W]. \tag{21}$$

This expression is non-negative so long as the correlation between the residuals is non-negative. The magnitude of the difference will be large if regions located in different clusters (so that $c(i) \neq c(j)$) that have similar shares (i.e. large values of $\sum_{s=1}^S w_{is} w_{js}$) also tend to have similar residuals (i.e. large values of $E[\epsilon_i \epsilon_j \mid W]$). For illustration, consider a simplified version of the model described in Section 3 in which: (a) $\sigma_s \geq 0$ for all s and $\phi \geq 0$, so that $0 \leq \lambda_i \leq 1$; (b) region-specific labor demand and supply shocks $\{\hat{\eta}_{is}\}_{s=1}^S$ and $\hat{v}_i$ are independent across regions; and (c) all labor demand and supply shocks

are independent of each other. Then, it follows from eqs. (12) and (16) that, for any $i \neq j$,

$$E[\epsilon_i \epsilon_j \mid W, \tilde{W}] = \lambda_i \lambda_j \sum_{s=1}^{S} w_{is} w_{js} E[\hat{\mu}_s^2 \mid W, \tilde{W}] + (1 - \lambda_i)(1 - \lambda_j) \sum_{g=1}^{G} \tilde{w}_{jg} \tilde{w}_{ig} E[\hat{v}_g^2 \mid W, \tilde{W}] \geq 0, \qquad (22)$$

which by the law of iterated expectations implies that $E[\hat{V}_{AKM}(\beta) - \hat{V}_{CL}(\beta) \mid W] \geq 0$. This expression illustrates that regions with similar shares will tend to have similar residuals in two cases. First, if the variance of the unobserved shifter $\hat{\mu}_s$ is large, so that $E[\hat{\mu}_s^2 \mid W, \tilde{W}]$ is large. In other words, standard inference methods lead to overrejection if the residual contains important shift-share terms that affect the outcome of interest through the same shares $\{w_{is}\}_{s=1}^{S}$ as those defining the covariate of interest X_i. Second, if the variance of the unobserved shifter $\hat{v}_g$ is large, so that $E[\hat{v}_g^2 \mid W, \tilde{W}]$ is large, and the shares $\tilde{w}_{ig}$ through which these shifters affect the outcome variable have a correlation structure that is similar to that of w_{is} (so that $\sum_{g=1}^{G} \tilde{w}_{ig} \tilde{w}_{jg}$ is large whenever $\sum_{s=1}^{S} w_{is} w_{js}$ is large). Thus, standard inference methods may overreject even when the unobserved shifters contained in the residual vary along a different dimension than the shift-share covariate of interest.

4.2 General case with controls

We now study the properties of the OLS estimator $\hat{\beta}$ of the coefficient on X_i in a regression of Y_i onto X_i and a K-vector of controls Z_i. To this end, let Z denote the $N \times K$ matrix with i-th row given by $Z_i' = (Z_{i1}, \ldots, Z_{iK})$, and let $\ddot{X} = X - Z(Z'Z)^{-1}Z'X$ denote an N-vector with i-th element equal to the regressor X_i with the controls Z_i partialled out (i.e. the residual from regressing X_i onto Z_i). Then, by the Frisch–Waugh–Lovell theorem, $\hat{\beta}$ can be written as

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_i}{\sum_{i=1}^{N} \ddot{X}_i^2} = \frac{\ddot{X}'Y}{\ddot{X}'\ddot{X}}. \qquad (23)$$

The controls may play two roles. First, they may be included to increase the precision of $\hat{\beta}$. Second, and more importantly, they may be included because one may worry that the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$ are correlated with the potential outcomes $\{Y_i(0)\}_{i=1}^{N}$, violating Assumption 1(ii). To formalize how Z_i, a regional variable, may be a control variable for the shifters, which vary at a sectoral level, we project Z_i onto the sectoral space using the same shares as those defining the shift-share regressor X_i,

$$Z_i = \sum_{s=1}^{S} w_{is} \mathcal{Z}_s + U_i. \qquad (24)$$

We think of $\{\mathcal{Z}_s\}_{s=1}^{S}$ as latent sector-level shocks that may have an independent effect on the outcome Y and may also be correlated with the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$, with U_i, the residual in this projection, mean-independent of the shifters. If the kth control Z_{ik} is included for precision, then the sector-level shocks $\{\mathcal{Z}_{sk}\}_{s=1}^{S}$ and, thus, Z_{ik}, are uncorrelated with X_i. If Z_{ik} is included because one worries that otherwise X_i may not be as good as randomly assigned, we interpret Z_{ik} as a proxy for the confounding sector-level shocks $\{\mathcal{Z}_{sk}\}_{s=1}^{S}$, and think of U_{ik} as a measurement error in this proxy.

To make this concrete, consider the model in Section 3, with the equivalences in eq. (12). Then

we may include $Z_{ik} = \sum_{s=1}^{S} l_{is}^{0} \hat{\mu}_s$ as a control. Here the measurement error in eq. (24) is zero, and $\mathcal{Z}_{sk} = \hat{\mu}_s$. If the shifters $\{\hat{\chi}_s\}_{s=1}^{S}$ are correlated with the demand shocks $\{\hat{\mu}_s\}_{s=1}^{S}$, then not including this control will generate omitted variable bias. Alternatively, we may include $Z_{ik} = \sum_{s=1}^{S} w_{is} \hat{\eta}_{is}$ as a control. Here $\mathcal{Z}_{sk} = 0$, and $U_{ik} = Z_{ik}$ is a regional aggregation of idiosyncratic region- and sector-specific labor-demand shocks that are independent of $\mathcal{X}_s$. In this case, if the shifters $\{\hat{\chi}_s\}_{s=1}^{S}$ are independent of the demand shocks $\{\eta_{is}\}_{i=1,s=1}^{N,S}$, then including the control will help increase the precision of $\hat{\beta}$, but it is not necessary for consistency.

Assumptions

For clarity of exposition, we focus here on the main substantive assumptions and relegate technical regularity conditions to Online Appendix A.1.1. Let $\mathcal{F}_0 = (Y(0), W, B, \mathcal{Z}, U)$; without controls, this set of variables reduces to $(Y(0), B, W)$, as in Section 4.1. Here, $\mathcal{Z}$ denotes the $S \times K$ matrix with sth row given by $\mathcal{Z}_s'$, and U denotes the $N \times K$ matrix with i-th element given by U_i'.

We maintain Assumption 2 with $\mathcal{F}_0 = (Y(0), W, B, \mathcal{Z}, U)$. The inclusion of controls allows us to weaken Assumption 1 and instead impose the following identification assumption:

Assumption 3 (Identification with controls). (i) The observed outcome satisfies $Y_i = Y_i(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, such that eq. (11) holds, and the controls Z_i satisfy eq. (24); (ii) The shifters are as good as randomly assigned in the sense that, for every s,

$$E[\mathcal{X}_s \mid \mathcal{F}_0] = E[\mathcal{X}_s \mid \mathcal{Z}_s], \tag{25}$$

and the right-hand side is linear in $\mathcal{Z}_s$,

$$E[\mathcal{X}_s \mid \mathcal{Z}_s] = \mathcal{Z}_s' \gamma; \tag{26}$$

(iii) For elements k such that $\gamma_k \neq 0$, $N^{-1} \sum_{i=1}^{N} E[U_{ik}^2] \to 0$; (iv) For elements k such that $\gamma_k \neq 0$, $(\sum_{s=1}^{S} n_s^2)^{-1/2} \sum_{i=1}^{N} E[U_{ik}^2] \to 0$.

Assumption 3(ii) weakens Assumption 1(ii) by only requiring the shifters to be as good as randomly assigned conditional on $\mathcal{Z}$, in the sense that eq. (25) holds. To interpret this restriction, consider a projection of the regional potential outcomes onto the sectoral space. For simplicity, consider the case with constant effects, $\beta_{is} = \beta$ for all i and s, and project $Y_i(0)$ onto the shares $(w_{i1}, \ldots, w_{iS})$, so that we may write $Y_i(0) = \sum_{s=1}^{S} w_{is} \mathcal{Y}_s(0) + \kappa_i$. Then, eq. (25) holds if (i) $\mathcal{Y}_s(0)$ is spanned by the vector of controls $\mathcal{Z}_s$; and (ii) $\{\mathcal{X}_s\}_{s=1}^{S}$ is mean-independent of the projection residuals $\{\kappa_i\}_{i=1}^{N}$.

As an example, consider again the model in Section 3, with the outcomes Y_i generated by eq. (12). Then eq. (25) holds, for example, if we set $\mathcal{Z}_s = \mathcal{Y}_s(0) = \hat{\mu}_s$ and if, conditional on the sector-specific labor demand shocks $\{\hat{\mu}_s\}_{s=1}^{S}$, the shifters of interest $\{\hat{\chi}_s\}_{s=1}^{S}$ are mean independent of the sector- and region-specific labor demand shocks $\{\hat{\eta}_{is}\}_{i=1,s=1}^{N,S}$ and of the labor supply shocks $\{\hat{v}_g\}_{g=1}^{G}$ and $\{\hat{v}_i\}_{i=1}^{N}$. Suppose, for instance, the shocks of interest $\{\hat{\chi}_s\}_{s=1}^{S}$ are changes in tariffs (e.g. Kovak, 2013) and that other potential labor demand shocks are those induced by automation and robots (e.g Acemoglu and Restrepo, 2019). Splitting the impact of automation into nationwide sector-specific effects, as captured

18

by $\{\hat{\mu}_s\}_{s=1}^S$, and sector- and region-specific deviations from the nationwide effects, as captured by $\{\hat{\eta}_{is}\}_{i=1,s=1}^{N,S}$, eq. (25) allows the political entity responsible for setting the tariffs to do so influenced by the nationwide sector-specific effects of automation, but not by any region-specific deviation from those national effects. In contrast, Assumption 1(ii) would require that the tariffs are also independent of the nationwide effects of automation.

Under eq. (25), one generally needs to include the controls non-parametrically; by imposing eq. (26), we ensure that it suffices to include the controls as additional covariates in a linear regression. If the shifters $\mathcal{X}_s$ are not mean zero (in the sense that the regression intercept on the right-hand side of eq. (26) is non-zero), eq. (26) requires that we include a constant $\mathcal{Z}_{sk} = 1$ as one of the controls. If the shares sum to one, $\sum_{s=1}^S w_{is} = 1$, this amounts to including an intercept $Z_{ik} = 1$ as a control in the regression. Importantly, if the shares do not sum to one, this amounts to including $\sum_{s=1}^S w_{is}$ as a control (see Borusyak, Hull and Jaravel, 2018, for a more extensive discussion of this point). For instance, if the shares w_{is} correspond to labor shares in different manufacturing sectors, one needs to include the size of the manufacturing sector $\sum_{s=1}^S w_{is}$ in each region as a control.

Given Assumption 3(ii), if we observed $\{\mathcal{Z}_s\}_{s=1}^S$ directly, we could include the vector $Z_i^* = \sum_{s=1}^S w_{is}\mathcal{Z}_s$ directly as control. However, the definition of each regional control Z_i in eq. (24) allows for Z_i^* to be observed with measurement error U_i. If $\gamma_k = 0$, such as when Z_{ik} is included for precision, then this measurement error in Z_{ik}^* does not matter; if $\gamma_k \neq 0$, this measurement error will in general induce a bias in $\hat{\beta}$. This is analogous to the classic linear regression result that measurement error in a control variable generally leads to a bias in the estimate of the coefficient on the variable of interest. Assumption 3(iii) ensures that any such bias disappears in large samples by imposing that the variance of the measurement error for controls that matter (i.e. those with $\gamma_k \neq 0$) converges to zero as $S \to \infty$. This ensures consistency of $\hat{\beta}$. For asymptotic normality, we need to strengthen this condition in Assumption 3(iv) by requiring that the variance of the measurement error converges to zero sufficiently fast. Assumption 3(iv) holds, for instance, if $U_i = S^{-1}\sum_{s=1}^S \psi_{is}$, where ψ_{is} is an idiosyncratic measurement error that is independent across s. In intuitive terms, this condition guarantees that Z_i is a sufficiently good proxy for the confounding latent shocks $\{\mathcal{Z}_s\}_{s=1}^S$.

Asymptotic theory

The following result generalizes Proposition 1:

Proposition 3. *Suppose Assumptions 2(i) and 2(ii) and Assumptions A.1(i) to A.1(iii) in Online Appendix A.1.1 hold with $\mathcal{F}_0 = (\mathcal{Z}, U, Y(0), B, W)$. Suppose also that Assumptions 3(i) to 3(iii) and Assumptions A.2(i) and A.2(ii) in Online Appendix A.1.1 hold. Then*

$$\beta = \frac{\sum_{i=1}^N \sum_{s=1}^S \pi_{is}\beta_{is}}{\sum_{i=1}^N \sum_{s=1}^S \pi_{is}}, \qquad and \qquad \hat{\beta} = \beta + o_p(1), \tag{27}$$

where $\pi_{is} = w_{is}^2 \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)$.

The only difference in the characterization of the probability limit relative to Proposition 1 is that the weights π_{is} now reflect the variance of $\mathcal{X}_s$ that also conditions on the controls.

To state the asymptotic normality result, define $\delta = E[Z'Z]^{-1}E[Z'(Y - X\beta)]$, so that we can define the regression residual in eq. (1) as $\epsilon_i = Y_i - X_i\beta - Z_i'\delta$.

Proposition 4. *Suppose Assumptions 2 and 3 and Assumptions A.1 and A.2 in Online Appendix A.1.1 hold with $\mathcal{F}_0 = (\mathcal{Z}, U, Y(0), B, W)$. Suppose, in addition, that*

$$\mathcal{V}_N = \frac{1}{\sum_{s=1}^S n_s^2} \operatorname{var}\left(\sum_{i=1}^N (X_i - Z_i'\gamma)\epsilon_i \mid \mathcal{F}_0 \right)$$

converges in probability to a non-random limit. Then

$$\frac{N}{\sqrt{\sum_{s=1}^S n_s^2}}(\hat{\beta} - \beta) = n\left(0, \frac{\mathcal{V}_N}{\left(\frac{1}{N}\sum_{i=1}^N \ddot{X}_i^2 \right)^2} \right) + o_p(1).$$

Relative to Proposition 2, the main difference is that X_i in the definition of $\mathcal{V}_N$ is replaced by $X_i - Z_i'\gamma$, and that X_i is replaced by $\ddot{X}_i$ in the outer part of the "sandwich." To motivate our standard error formula, suppose that $\beta_{is} = \beta$ for all i and s. Under $\beta_{is} = \beta$, it follows from eq. (25) and Assumption 2(i) that

$$\mathcal{V}_N = \frac{\sum_{s=1}^S \operatorname{var}(\tilde{\mathcal{X}}_s \mid \mathcal{F}_0)R_s^2}{\sum_{s=1}^S n_s^2}, \qquad R_s = \sum_{i=1}^N w_{is}\epsilon_i, \qquad \tilde{\mathcal{X}}_s = \mathcal{X}_s - \mathcal{Z}_s'\gamma.$$

A plug-in estimate of R_s can be constructed by replacing ϵ_i with the estimated regression residuals $\hat{\epsilon}_i = Y_i - X_i\hat{\beta} - Z_i\hat{\delta}$, where $\hat{\delta} = (Z'Z)^{-1}Z'(Y - X\hat{\beta})$ is an OLS estimate of δ. We can estimate the variance $\operatorname{var}(\tilde{\mathcal{X}}_s \mid \mathcal{F}_0)$ by $\widehat{\mathcal{X}}^2$, where

$$\widehat{\mathcal{X}} = (W'W)^{-1}W'\ddot{X} \tag{28}$$

projects the estimate $\ddot{X}$ of $X - Z'\gamma$ onto the sectoral space by regressing it onto the shares W. To carry out the regression in eq. (28), W must be full rank; this requires that there are more regions than sectors, $N \geq S$. These steps lead to the standard error estimate

$$\widehat{se}(\hat{\beta}) = \frac{\sqrt{\sum_{s=1}^S \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{\sum_{i=1}^N \ddot{X}_i^2}, \qquad \hat{R}_s = \sum_{i=1}^N w_{is}\hat{\epsilon}_i. \tag{29}$$

The next remark summarizes the steps needed for the construction of the standard error $\widehat{se}(\hat{\beta})$:

Remark 5. *To construct the standard error estimate in eq. (29):*

1. *Obtain the estimates $\hat{\beta}$ and $\hat{\delta}$ by regressing Y_i onto $X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$ and the controls Z_i. The estimate $\hat{\epsilon}_i$ corresponds to the estimated regression residuals.*

2. *Construct $\ddot{X}_i$, the residuals from regressing X_i onto Z_i. Compute $\widehat{\mathcal{X}}_s$, the regression coefficients from regressing $\ddot{X}$ onto W.*

3. *Plug the estimates $\hat{\epsilon}_i$, $\ddot{X}_i$, and $\widehat{\mathcal{X}}_s$ into the standard error formula in eq. (29).*

To gain intuition for the procedure in Remark 5, it is useful to consider again the case with concentrated sectors. Suppose that $U_i = 0$ for all i, so that the regression of Y_i onto X_i and Z_i is identical to the regression of Y_i onto $\mathcal{X}_{s(i)}$ and $\mathcal{Z}_{s(i)}$. Then the standard error formula in eq. (29) reduces to the usual cluster-robust standard error, with clustering on $s(i)$.

The cluster-robust standard error is generally biased due to estimation noise in estimating ϵ_i, which can lead to undercoverage, especially in cases with few clusters (see Cameron and Miller, 2014 for a survey). Since the standard error in eq. (29) can be viewed as generalizing the cluster-robust formula, similar concerns arise in our setting. We thus consider a modification $\widehat{se}_{\beta_0}(\hat{\beta})$ of $\widehat{se}(\hat{\beta})$ that imposes the null hypothesis when estimating the regression residuals to reduce the estimation noise in estimating ϵ_i.[14] To calculate the standard error $\widehat{se}_{\beta_0}(\hat{\beta})$ for testing the hypothesis $H_0 \colon \beta = \beta_0$ against a two-sided alternative at significance level α, one replaces $\hat{\epsilon}_i$ with $\hat{\epsilon}_{\beta_0,i}$, the residual from regressing $Y_i - X_i\beta_0$ onto Z_i ($\hat{\epsilon}_{\beta_0,i}$ is an estimate of the residuals with the null imposed). The null is rejected if the absolute value of the t-statistic $(\hat{\beta} - \beta_0)/\widehat{se}_{\beta_0}(\hat{\beta})$ exceeds $z_{1-\alpha/2}$, the $1 - \alpha/2$ quantile of a standard normal distribution (1.96 for $\alpha = 0.05$). To construct a confidence interval (CI) with coverage $1 - \alpha$, one collects all hypotheses β_0 that are not rejected. The endpoints of this CI are a solution to a quadratic equation, and are thus available in closed form—one does not have to numerically search for all the hypotheses that are not rejected. The next remark summarizes this procedure.

Remark 6 (Confidence interval with null imposed). *To test the hypothesis $H_0 \colon \beta = \beta_0$ with significance level α or, equivalently, to check whether β_0 lies in the confidence interval with confidence level $1 - \alpha$:*

1. *Obtain the estimate $\hat{\beta}$ by regressing Y_i onto $X_i = \sum_{s=1}^{S} w_{is}\mathcal{X}_s$ and the controls Z_i. Obtain the restricted regression residuals $\hat{\epsilon}_{\beta_0,i}$ as the residuals from regressing $Y_i - X_i\beta_0$ onto Z_i.*

2. *Construct $\ddot{X}_i$, the residuals from regressing X_i onto Z_i. Compute $\widehat{\mathcal{X}}_s$, the regression coefficients from regressing $\ddot{X}$ onto W (this step is identical to step 2 in Remark 5).*

3. *Compute the standard error as*

$$\widehat{se}_{\beta_0}(\hat{\beta}) = \frac{\sqrt{\sum_{s=1}^{S} \widehat{\mathcal{X}}_s^2 \hat{R}_{\beta_0,s}^2}}{\sum_{i=1}^{N} \ddot{X}_i^2}, \qquad \hat{R}_{\beta_0,s} = \sum_{i=1}^{N} w_{is}\hat{\epsilon}_{\beta_0,i}. \tag{30}$$

4. *Reject the null if $|(\hat{\beta} - \beta_0)/\widehat{se}_{\beta_0}(\hat{\beta})| > z_{1-\alpha/2}$. A confidence set with coverage $1 - \alpha$ is given by all nulls that are not rejected, $CI_{1-\alpha} = \{\beta_0 \colon |(\hat{\beta} - \beta_0)/\widehat{se}_{\beta_0}(\hat{\beta})| < z_{1-\alpha/2}\}$. This set is an interval with endpoints given by*

$$\hat{\beta} - A \pm \sqrt{A^2 + \frac{\widehat{se}(\hat{\beta})^2}{Q/(\ddot{X}'\ddot{X})^2}}, \qquad A = \frac{\sum_{s=1}^{S} \widehat{\mathcal{X}}_s^2 \hat{R}_s \sum_{i=1}^{N} w_{is}\ddot{X}_i}{Q}, \tag{31}$$

where $Q = (\ddot{X}'\ddot{X})^2/z_{1-\alpha/2}^2 - \sum_{s=1}^{S} \widehat{\mathcal{X}}_s^2 (\sum_i w_{is}\ddot{X}_i)^2$ and $\widehat{se}(\hat{\beta})$ and $\hat{R}_s$ are given in eq. (29).

---

[14] Alternatively, one could construct a bias-corrected variance estimate; see, for example, Bell and McCaffrey (2002) for an example of this approach in the context of cluster-robust inference.

Proposition 5. *Suppose that the assumptions of Proposition 4 hold, and that $\beta_{is} = \beta$. Suppose also that $N \geq S$, W is full rank, and that either $\max_s \sum_{i=1}^N |((W'W)^{-1}W')_{si}|$ is bounded and $\max_i E[(U_i'\gamma)^4 \mid W] \to 0$, or else that $U_i = 0$ for $i = 1, \ldots, N$. Define $\widehat{\mathcal{X}}$ as in eq. (28), and let $\hat{R}_s = \sum_{i=1}^N w_{is}\tilde{\epsilon}_i$, where $\tilde{\epsilon}_i = Y_i - X_i\tilde{\beta} - Z_i'\tilde{\delta}$, and $\tilde{\beta}$ and $\tilde{\delta}$ are consistent estimators of δ and β. Then*

$$\frac{\sum_{s=1}^S \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}{\sum_{s=1}^S n_s^2} = \mathcal{V}_N + o_p(1). \tag{32}$$

Since in both $\hat{\epsilon}_i$ and $\hat{\epsilon}_{\beta_0,i}$ are consistent estimates of the residuals, this proposition shows that the procedures in Remarks 5 and 6 both yield asymptotically valid confidence intervals. The additional assumptions of Proposition 5 ensure that the estimation error in $\widehat{\mathcal{X}}_s$ that arises from having to back out the sector-level shocks $\mathcal{Z}_s$ from the controls Z_i is not too large. If the sectors are concentrated, then $((W'W)^{-1}W')_{si} = \mathbb{I}\{s(i) = s\}/n_s$, so that $\max_s \sum_{i=1}^N |((W'W)^{-1}W')_{si}| = 1$, and the assumption always holds. We show in Online Appendix A.1.6 that the procedures in Remarks 5 and 6 continue to yield valid inference if β_{is} is heterogeneous across regions and sectors, as long as further regularity conditions hold.

Although both standard errors $\widehat{se}_{\beta_0}(\hat{\beta})$ and $\widehat{se}(\hat{\beta})$ are consistent (and one could further show that the resulting confidence intervals are asymptotically equivalent), they will in general differ in finite samples. In particular, it can be seen from eq. (31) that the confidence interval with the null imposed is not symmetric around $\hat{\beta}$, but its center is shifted by A.[15] As we show in Section 6, this recentering tends to improve the finite-sample coverage properties of the confidence interval. On the other hand, the confidence interval described in Remark 6 tends to be longer on average than that in Remark 5.

4.3 Instrumental variables regression

We now turn to the problem of estimating the effect of a regional treatment variable Y_{2i} on a regional outcome Y_{1i} using the shift-share variable $X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$ as an instrumental variable (IV). To set up the problem precisely, we again use the potential outcome framework. In particular, we assume that

$$Y_{1i}(y_2) = Y_{1i}(0) + y_2\alpha, \tag{33}$$

where α, our parameter of interest, measures the causal effect of Y_{2i} onto Y_{1i}. We assume for simplicity that this causal effect is linear and constant across regions.[16] In analogy with eq. (11), we denote the region-i treatment level that would occur if the region received shocks $(x_1, \ldots, x_S)$ as

$$Y_{2i}(x_1, \ldots, x_S) = Y_{2i}(0) + \sum_{s=1}^S w_{is}x_s\beta_{is}. \tag{34}$$

---

[15]This is analogous to the differences in likelihood models between confidence intervals based on the Lagrange multiplier test (which imposes the null and is not symmetric around the maximum likelihood estimate) and the Wald test (which does not impose the null and yields the usual confidence interval).

[16]If we weaken the assumption of constant treatment effects and instead assume $Y_{1i}(y_2) = Y_{1i}(0) + y_2\alpha_i$, then it follows by a mild extension of the results in Online Appendix A.2 that our methods would deliver inference on the estimand $\sum_{i=1}^N \pi_i\alpha_i / \sum_{i=1}^N \pi_i$, with $\pi_i = \sum_{s=1}^S w_{is}^2 \text{var}(\mathcal{X}_s \mid \mathcal{F}_0)\beta_{is}$, where $\mathcal{F}_0 = (\mathcal{Z}, U, Y_1(0), Y_2(0), B, \alpha, W)$, and β_{is} is defined in eq. (34).

The observed outcome and treatment variables are given by $Y_{1i} = Y_{1i}(Y_{2i})$ and $Y_{2i} = Y_{2i}(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, respectively.

The framework in eqs. (33) and (34) maps directly to the problem of estimating the regional inverse labor supply elasticity. In particular, in the context of the model in Section 3, eqs. (8) and (10) map directly into eqs. (33) and (34) if we define

$$Y_{1i} = \hat{\omega}_i, \ Y_{2i} = \hat{L}_i, \ \alpha = \tilde{\phi}, \ Y_{1i}(0) = -\tilde{\phi}(\sum_{g=1}^{G} \tilde{w}_{ig} \hat{v}_g + \hat{v}_i), \ w_{is} = l_{is}^0, \ \mathcal{X}_s = \hat{\chi}_s, \ \beta_{is} = \theta_{is}, \qquad (35)$$

and $Y_{2i}(0)$ is given by the expression for $Y_i(0)$ in eq. (12).[17] As this mapping illustrates, the potential outcome $Y_{1i}(0)$ will generally have a shift-share structure, with the shifters being group-specific labor supply shocks (e.g. growth in the number of workers by education group). Consequently, the regression residual in the structural equation will generally have a shift-share structure. Similarly, as eq. (12) illustrates, the potential outcome $Y_{2i}(0)$ will also generally include several shift-share components, with the shifters being either sector-specific labor demand shocks or the same group-specific labor supply shocks appearing in $Y_{1i}(0)$. Thus, the regression residual in the first-stage regression of Y_{2i} onto X_i will also generally have a shift-share structure.

Our estimate of α is given by an IV regression of Y_{1i} onto Y_{2i} and a K-vector of controls Z_i, with X_i used as an instrument for Y_{2i}. This IV estimate can be written as

$$\hat{\alpha} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_{1i}}{\sum_{i=1}^{N} \ddot{X}_i Y_{2i}}, \qquad (36)$$

where, as in Section 4.2, $\ddot{X}_i$ denotes the residual from regressing X_i onto Z_i.

Assumptions

Assumption 4 is a generalization of Assumption 3. Let $\mathcal{F}_0 = (\mathcal{Z}, U, Y_1(0), Y_2(0), B, W)$.

Assumption 4 (IV Identification). (i) The observed outcome and treatment variables satisfy $Y_{1i} = Y_{1i}(Y_{2i})$ and $Y_{2i} = Y_{2i}(\mathcal{X}_1, \ldots, \mathcal{X}_s)$ such that eqs. (33) and (34) hold, and the controls Z_i satisfy eq. (24); (ii) The shifters are exogenous in the sense that, for every s,

$$E[\mathcal{X}_s \mid \mathcal{F}_0] = E[\mathcal{X}_s \mid \mathcal{Z}_s], \qquad (37)$$

and the right-hand side satisfies eq. (26); (iii) Assumptions 3(iii) and 3(iv) hold; (iv) $\sum_{i=1}^{N} \sum_{s=1}^{S} w_{is}^2 \cdot \text{var}(\mathcal{X}_s \mid \mathcal{F}_0) \beta_{is} \neq 0$.

Assumption 4(ii) adapts the standard instrument exogeneity condition (see, e.g., Condition 1 in Imbens and Angrist, 1994) to our setting. Our approach follows Borusyak, Hull and Jaravel (2018), who impose a similar identification condition. To illustrate the restrictions that Assumption 4(ii) may impose, consider again the problem of estimating the inverse labor supply elasticity within the

---

[17]In some applications of shift-share IVs, the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$ are unobserved and have to be estimated. We assume here that $\mathcal{X}_s$ is directly measurable for every sector s, and study the case with estimated shifters in Section 5.3.

context of the model in Section 3, with the mapping between this model and the potential outcomes in eqs. (33) and (34) given in eqs. (12) and (35). If the controls $\{\mathcal{Z}_s\}_{s=1}^S$ correspond to the shocks $\{\hat{\mu}_s\}_{s=1}^S$, then eq. (37) requires that, conditional on $\{\hat{\mu}_s\}_{s=1}^S$, the labor demand shocks $\{\hat{\chi}_s\}_{s=1}^S$ used to construct our IV are mean-independent of the idiosyncratic labor demand shocks $\{\hat{\eta}_{is}\}_{i=1,s=1}^{N,S}$ and of the labor supply shifters $\{\hat{\nu}_i\}_{i=1}^N$ and $\{\hat{\nu}_g\}_{g=1}^G$.[18] For example, if $\{\hat{\chi}_s\}_{s=1}^S$ are sectoral productivity shocks, then these productivity shocks need to be independent of shocks to individuals' willingness to work in different groups and regions. Assumption 4(iv) requires that the coefficient on the instrument in the first-stage equation, which can be written as $\beta = \sum_{i=1}^N \sum_{s=1}^S w_{is}^2 \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)\beta_{is} / \sum_{i=1}^N \sum_{s=1}^S w_{is}^2 \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)$, is non-zero—this is the standard IV relevance assumption. For consistency and inference, in an analogy to the OLS case, we assume that Assumption 2 holds with $\mathcal{F}_0 = (\mathcal{Z}, U, Y_1(0), Y_2(0), B, W)$.

In a recent paper, Goldsmith-Pinkham, Sorkin and Swift (2018) explore a different approach to identification and inference on the treatment effect α. Focusing here for simplicity on the case without controls, in place of Assumption 4(ii), they assume that the shares $\{w_{is}\}_{s=1}^S$ are as good as randomly assigned conditional on the shifters $\{\mathcal{X}_s\}_{s=1}^S$; so that they are mean-independent of the potential outcomes $Y_1(0)$ and $Y_2(0)$ conditional on $\mathcal{X}$. As Goldsmith-Pinkham, Sorkin and Swift (2018) show, under this alternative assumption, one can replace the shift-share instrument $X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$ by the full vector of shares $(w_{i1}, \ldots, w_{iS})$ in the first-stage equation. For estimation and inference, this alternative approach requires that, conditionally on the shifters, either the shares $(w_{i1}, \ldots, w_{iS})$ or else the structural residuals be independent across regions or clusters of regions.

For estimating the inverse labor supply elasticity in the context of the model in Section 3, eq. (35) illustrates that this alternative identification assumption requires that, conditional on $\{\hat{\chi}_s\}_{s=1}^S$, the region-specific employment shares in the initial equilibrium $\{l_{is}^0\}_{s=1}^S$ are mean-independent of both the region-specific exposure shares $\{\tilde{w}_{ig}\}_{g=1}^G$, and the region-specific labor supply shock ν_i. This assumption is violated if regions more exposed to labor demand shocks in a sector s (e.g. to changes in tariffs in the food sector) are also more exposed to labor supply shocks affecting workers of a group g (e.g. currency crisis in Mexico affecting the number of Mexican migrants; see Monras, 2018).[19]

In terms of inference, since the structural residuals will not be independent across regions unless they contain no shift-share component (which, according to the economic model in Section 3, is unlikely), the approach in Goldsmith-Pinkham, Sorkin and Swift (2018) generally requires that the shares are independent across (clusters of) regions. This assumption is, from the perspective of the model in Section 3, conceptually very different from assuming independence of the shifters $\mathcal{X}_s$ across

---

[18]If, instead of eq. (34), we defined the first stage as simply the projection of Y_{2i} onto the shift-share instrument, we could further relax this condition and only require $\{\hat{\chi}_s\}_{s=1}^S$ to be mean-independent of the labor supply shifters. An advantage of the current setup is that it allows us to derive primitive conditions for the consistency of the estimates of the first-stage regression and, thus, of the IV estimator.

[19]To allow for a shift-share component in the structural residual, Goldsmith-Pinkham, Sorkin and Swift (2018) view the shares $(w_{i1}, \ldots, w_{iS})$ as "invalid" instruments, since, in this case, $E[\epsilon_i w_{is} \mid \mathcal{X}] \neq 0$, where ϵ_i denotes the structural error. Goldsmith-Pinkham, Sorkin and Swift (2018) show that if these shares are used to construct a single shift-share instrument X_i, the bias in the IV estimator coming from the correlation between any w_{is} and the structural residual averages out under certain conditions as $S \to \infty$, as in the many invalid instrument setting studied in Kolesár et al. (2015). Under the current setup, in contrast, eq. (37) implies that X_i is a valid instrument for any fixed S. Leveraging exogeneity of $\mathcal{X}_s$ is a key difference between our approach and that in Kolesár et al. (2015) and Goldsmith-Pinkham, Sorkin and Swift (2018). It allows us to do inference without imposing a particular correlation structure on the residuals ϵ_i, and it allows us to achieve identification without requiring $S \to \infty$; the latter is only needed for consistency and inference.

sectors. Since the shifters $\mathcal{X}_s = \hat{\chi}_s$ are exogenous, the latter only involves assumptions on model fundamentals by restricting the distribution in eq. (7). In contrast, each share $w_{is} = l^0_{is}$ corresponds to the employment allocation across sectors in a region i in an initial equilibrium, so that the former involves imposing restrictions on an endogenous outcome of the model. Furthermore, since all the shares $\{w_{is}\}^{N,S}_{i=1,s=1}$ depend on the same set of sector-specific labor demand shifters $\{(\chi_s, \mu_s)\}^S_{s=1}$, they will generally be correlated across regions.[20]

Which identification and inference approach is more attractive depends on the context of each particular empirical application. While the economic model in Section 3 motivates the approach we pursue here, this does not mean that our approach is generally more attractive. In other empirical applications (e.g. when the shares are exogenous variables from the perspective of an economic framework), the approach of Goldsmith-Pinkham, Sorkin and Swift (2018) may be more appropriate.

Asymptotic theory

It follows by adapting the arguments in the proof of Proposition 4 that, if Assumption 4 holds, and Assumption 2 holds with $\mathcal{F}_0 = (\mathcal{Z}, U, Y_1(0), Y_2(0), B, W)$, then, under mild technical regularity conditions (see Online Appendix A.2 for details and proof),

$$\frac{N}{\sqrt{\sum^S_{s=1} n^2_s}} (\hat{\alpha} - \alpha) = n \left(0, \frac{\mathcal{V}_N}{(\frac{1}{N} \sum^N_{i=1} \ddot{X}_i Y_{2i})^2} \right) + o_p(1), \; \mathcal{V}_N = \frac{\sum^S_{s=1} \text{var}(\tilde{\mathcal{X}}_s \mid \mathcal{F}_0) R^2_s}{\sum^S_{s=1} n^2_s}, \; R_s = \sum^N_{i=1} w_{is} \epsilon_i,$$

(38)

where $\epsilon_i = Y_{1i} - Y_{2i}\alpha - Z'_i \delta$ is the residual in the structural equation, with $\delta = E[Z'Z]^{-1} E[Z'(Y_1 - Y_2\alpha)]$. This suggests the standard error estimate

$$\hat{se}(\hat{\alpha}) = \frac{\sqrt{\sum^S_{s=1} \hat{\mathcal{X}}^2_s \hat{R}^2_s}}{|\sum^N_{i=1} \ddot{X}_i Y_{2i}|} = \frac{\sqrt{\sum^S_{s=1} \hat{\mathcal{X}}^2_s \hat{R}^2_s}}{\sum^N_{i=1} \ddot{X}^2_i |\hat{\beta}|}, \qquad \hat{R}_s = \sum^N_{i=1} w_{is} \hat{\epsilon}_i,$$

(39)

where $\hat{\mathcal{X}}_s$ is constructed as in Remark 5, $\hat{\epsilon} = Y_1 - Y_2\hat{\alpha} - Z'(Z'Z)^{-1} Z'(Y_1 - Y_2\hat{\alpha})$ is the estimated residual of the structural equation, and $\hat{\beta} = \sum^N_{i=1} \ddot{X}_i Y_{2i} / \sum^N_{i=1} \ddot{X}^2_i$ is the first-stage coefficient.

The difference between the IV standard error formula in eq. (39) and the OLS version in eq. (29) is analogous to the difference between IV standard errors and OLS heteroskedasticity-robust standard errors for the corresponding reduced-form specification: the residual $\hat{\epsilon}_i$ corresponds to the residual in the structural equation, and the denominator is scaled by the first-stage coefficient. To obtain the IV analog of the standard error estimator under the null $H_0: \alpha = \alpha_0$, we use the formula in eq. (39) except that, instead of $\hat{\epsilon}_i$, we use the structural residual computed under the null, $\hat{\epsilon}_{\alpha_0} = (I - Z'(Z'Z)^{-1}Z')(Y_1 - Y_2\alpha_0)$. The resulting confidence interval is a generalization of the Anderson and Rubin (1949) confidence interval (which assumes that the structural errors are independent). For this reason, this confidence interval will remain valid even if the shift-share instrument is weak.

[20]For instance, if $\sigma_s = \sigma$ for all s, then $l^0_{is} = D^0_{is} / (\sum^S_{t=1} D^0_{it})$, where D^0_{is} is the labor demand shifter of sector s in region i in the initial equilibrium. According to eq. (3), for any s, all shifters $\{D^0_{is}\}^N_{i=1}$ depend on the same sector-level demand shocks, $\{(\chi_s, \mu_s)\}^S_{s=1}$ and, thus, the labor shares l^0_{is} will generally be correlated across all regions for any given sector.

5 Extensions

We now discuss three extensions to the basic setup. In Section 5.1, we relax the assumption that the shifters $\{\mathcal{X}_s\}_{s=1}^S$ are independent, allowing them to be correlated within clusters of sectors. Section 5.2 generalizes our results to settings in which we have multiple observations for each region. Section 5.3 considers the case in which the shifters are not directly observed, and have to be estimated.

5.1 Clusters of sectors

Suppose that the sectors can be grouped into larger units, which we refer to as "clusters", with $c(s) \in \{1, \ldots, C\}$ denoting the cluster that sector s belongs to; e.g., if each s corresponds to a four-digit industry code, $c(s)$ may correspond to a three-digit code. With this structure, we replace Assumption 2(i) with the weaker assumption that, conditional on $\mathcal{F}_0$, the shocks $\mathcal{X}_s$ and $\mathcal{X}_k$ are independent if $c(s) \neq c(k)$, and we replace Assumption 2(iii) with the assumption that, as $C \to \infty$, the largest cluster makes an asymptotically negligible contribution to the asymptotic variance; i.e. $\max_c \tilde{n}_c^2 / \sum_{d=1}^C \tilde{n}_d^2 \to 0$, where $\tilde{n}_c = \sum_{s=1}^S \mathbb{I}\{c(s) = c\} n_s$ is the total share of cluster c.

Under this setup, by generalizing the arguments in Section 4.2, one can show that, as $C \to \infty$,

$$\frac{N}{\sqrt{\sum_{c=1}^C \tilde{n}_c^2}} (\hat{\beta} - \beta) = n \left(0, \frac{\mathcal{V}_N}{\left(\frac{1}{N} \sum_{i=1}^N \ddot{X}_i^2 \right)^2} \right) + o_p(1),$$

and, assuming that $\beta_{is} = \beta$ for every region and sector, the term $\mathcal{V}_N$ is now given by

$$\mathcal{V}_N = \frac{\sum_{c=1}^C \sum_{s=1,t=1}^{S,S} \mathbb{I}\{c(s) = c(t) = c\} E[\tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t \mid W, \mathcal{Z}] R_s R_t}{\sum_{c=1}^C \tilde{n}_c^2}, \qquad R_s = \sum_{i=1}^N w_{is} \epsilon_i, \quad \tilde{\mathcal{X}}_s = \mathcal{X}_s - \mathcal{Z}_s' \gamma.$$

As a result, we replace the standard error estimate in eq. (29) with a version that clusters $\widehat{\mathcal{X}}_s \hat{R}_s$,

$$\hat{se}(\hat{\beta}) = \frac{\sqrt{\sum_{c=1}^C \sum_{s,t} \mathbb{I}\{c(s) = c(t) = c\} \widehat{\mathcal{X}}_s \hat{R}_s \widehat{\mathcal{X}}_t \hat{R}_t}}{\sum_{i=1}^N \ddot{X}_i^2}, \qquad \hat{R}_s = \sum_{i=1}^N w_{is} \hat{\epsilon}_i, \tag{40}$$

where $\widehat{\mathcal{X}}_s$ is defined as in Remark 5. Confidence intervals with the null imposed can be constructed as in Remark 6, replacing $\hat{\epsilon}_i$ with $\hat{\epsilon}_{\beta_0, i}$ in eq. (40). In the IV setting considered in Section 4.3, the standard error for $\hat{\alpha}$ is analogous to that in eq. (40), except that $\hat{\epsilon}_i$ denotes the residual in the structural equation, and we divide the expression by the absolute value of the first-stage coefficient, $\sum_{i=1}^N \ddot{X}_i Y_{2i} / \sum_{i=1}^N \ddot{X}_i^2$.

5.2 Panel data

Consider a setting with $j = 1, \ldots, J$ regions, $k = 1, \ldots, K$ sectors, and $t = 1, \ldots, T$ periods. For each period t, we have data on shifters $\{\mathcal{X}_{kt}\}_{k=1}^K$, outcomes $\{Y_{jt}\}_{j=1}^J$, and shares $\{w_{jkt}\}_{j=1,k=1}^{J,K}$. This setup maps into the potential outcome framework in eq. (11) if we identify a "sector" with a sector-period pair $s = (k, t)$, and a "region" with a region-period pair $i = (j, t)$, so that we can index outcomes and

26

shifters as $Y_i = Y_{jt}$ and $\mathcal{X}_s = \mathcal{X}_{kt}$, with the shares given by

$$
w_{is} = \begin{cases} w_{jkt} & \text{if } i = (j,t) \text{ and } s = (k,t), \\ 0 & \text{if } i = (j,t), \, s = (k,t'), \text{ and } t \neq t'. \end{cases} \tag{41}
$$

If the shifters $\mathcal{X}_{kt}$ are independent across time and sectors, Propositions 3 and 4 immediately give the large-sample distribution of the OLS estimator. In general, however, it will be important to allow the shifters $\mathcal{X}_{kt}$ to be correlated across time within each sector k. In this case, one can use the clustered standard error derived in Section 5.1 by grouping observations over time for each sector k into a common cluster, so that $c(k,t) = c(k',t')$ if $k = k'$. We can then apply the formula in eq. (40) to allow for any arbitrary time-series correlation in the sector-level shocks $\mathcal{X}_{kt}$ for any given sector k. Regardless of whether the sector-period pairs (k,t) are clustered, as discussed in Remark 4, our standard error formulas allow for arbitrary dependence patterns in the regression residuals—in particular, they account for potential serial dependence in the regression residuals.

If the shift-share regressor is used as an IV in a regression of an outcome Y_{1jt} onto a treatment Y_{2jt}, the mapping to eqs. (33) and (34) is analogous, and one can use an IV version of the formula in eq. (40) for inference.

5.3 IV with estimated shifters

We now consider a setting in which the sectoral shifters $\{\mathcal{X}_s\}_{s=1}^S$ that define the shift-share IV studied in Section 4.3 are not directly observed. We follow the setup in Section 4.3 but assume that, instead of observing $\mathcal{X}_s$ directly, we only observe a noisy measure of it,

$$
X_{is} = \mathcal{X}_s + \psi_{is} \tag{42}
$$

for each sector-region pair. We consider IV regressions that use two different estimates of $X_i = \sum_{s=1}^S w_{is}\mathcal{X}_s$. First, an estimate that replaces $\mathcal{X}_s$ with an estimate $\hat{\mathcal{X}}_s = \sum_{i=1}^N \breve{w}_{is}X_{is}/\breve{n}_s$, where $\breve{n}_s = \sum_{i=1}^N \breve{w}_{is}$ and the weights $\breve{w}_{is}$ are not necessarily related to w_{is}. The resulting estimate of X_i is

$$
\hat{X}_i = \sum_{s=1}^S w_{is}\hat{\mathcal{X}}_s = \sum_{s=1}^S w_{is}\frac{1}{\breve{n}_s}\sum_{j=1}^N \breve{w}_{js}X_{js}, \tag{43}
$$

and it yields the IV estimate $\tilde{\alpha} = \ddot{\hat{X}}'Y_1/\ddot{\hat{X}}'Y_2$, where $\ddot{\hat{X}} = \hat{X} - Z(Z'Z)^{-1}Z'\hat{X}$ is the residual from regressing $\hat{X}_i$ onto Z_i. Second, we consider the leave-one-out estimator

$$
\hat{X}_{i,-} = \sum_{s=1}^S w_{is}\hat{\mathcal{X}}_{s,-i} = \sum_{s=1}^S w_{is}\frac{1}{\breve{n}_{s,-i}}\sum_{j=1}^N \mathbb{I}\{j \neq i\}\breve{w}_{js}X_{js}, \qquad \breve{n}_{s,-i} = \sum_{j=1}^N \mathbb{I}\{j \neq i\}\breve{w}_{js}, \tag{44}
$$

where $\hat{\mathcal{X}}_{s,-i} = \sum_{j=1}^N \mathbb{I}\{j \neq i\}\breve{w}_{js}X_{js}/\breve{n}_{s,-i}$ is an estimate of $\mathcal{X}_s$ that excludes region i. A version of this estimator has been used in Autor and Duggan (2003). This leave-one-out estimator of the shift-share instrument X_i yields the IV estimate $\hat{\alpha}_- = \ddot{\hat{X}}'_-Y_1/\ddot{\hat{X}}'_-Y_2$, where $\ddot{\hat{X}}_- = \hat{X}_- - Z'(Z'Z)^{-1}Z'\hat{X}_-$.

While we assume that $\mathcal{X}_s$ satisfies the exogeneity restriction in Assumption 4(ii) for every s, we

allow the measurement errors $\psi_i = (\psi_{i1}, \ldots, \psi_{iS})'$ to be potentially correlated with the potential outcomes $Y_{1i}(0)$ and $Y_{2i}(0)$ in the same region i. We assume, however, that ψ_i is independent of the errors ψ_j and of the potential outcomes $Y_{1j}(0)$ and $Y_{2j}(0)$ for any region $j \neq i$ (see Online Appendix A.2 for a formal statement). In Online Appendix E.2.3, we use the model in Section 3 to discuss these assumptions in the context of estimating the inverse labor supply elasticity.[21]

The potential correlation between ψ_i and the potential outcomes in region i implies that the estimation error in $\hat{X}_i$, which is a function on ψ_i, may be correlated with the residual in the structural equation. Thus, including the ith observation in the construction of $\hat{X}_i$ induces an own-observation bias in the IV estimator $\tilde{\alpha}$ of α. See Goldsmith-Pinkham, Sorkin and Swift (2018) and Borusyak, Hull and Jaravel (2018) for a discussion. This bias is analogous to the bias of the two-stage least squares estimator in settings with many instruments (e.g. Bekker, 1994; Angrist, Imbens and Krueger, 1999), such as when one uses group indicators as instruments.[22] We show in Online Appendix A.2 that the magnitude of the bias is of the order $\frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \frac{w_{is} \breve{w}_{is}}{\breve{n}_s} \leq S/N$, so that consistency of $\tilde{\alpha}$ generally requires the number of sectors to grow more slowly than the number of regions. Furthermore, to ensure that the asymptotic bias in $\tilde{\alpha}$ does not induce undercoverage of the resulting confidence intervals, one generally requires $S^{3/2}/N \to 0$.

The estimator $\hat{\alpha}_-$, which can be thought of as a shift-share analog of the jackknife IV estimator studied in Angrist, Imbens and Krueger (1999), remains consistent, as shown in Borusyak, Hull and Jaravel (2018) and in Online Appendix A.2. We also show in this appendix that, under regularity conditions, its asymptotic distribution is given by

$$\frac{N}{\sqrt{\sum_{s=1}^{S} n_s^2}} (\hat{\alpha}_- - \alpha) = n \left(0, \frac{\mathcal{V}_N + \mathcal{W}_N}{\left(\frac{1}{N} \sum_{i=1}^{N} \ddot{X}_i Y_{2i} \right)^2} \right) + o_p(1), \tag{45}$$

with $\mathcal{V}_N$ defined as in eq. (38), and

$$\mathcal{W}_N = \frac{1}{\sum_{s=1}^{S} n_s^2} \left(\sum_{j=1}^{N} \left(\sum_{i=1}^{N} \mathcal{S}_{ij} \right)^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{S}_{ij} \mathcal{S}_{ji} \right), \qquad \mathcal{S}_{ij} = \sum_{s=1}^{S} \mathbb{I}\{i \neq j\} \frac{w_{is} \breve{w}_{js} \psi_{js} \epsilon_i}{\breve{n}_{s,-i}}.$$

The term $\mathcal{W}_N$ accounts for the additional uncertainty stemming from the fact that the shift-share IV is estimated. It is analogous to the many-instrument term in the jackknife IV estimator under many instrument asymptotics (see Chao et al., 2012). Using simulations, we show in Online Appendix E.2.4 several designs in which, while correcting for the own-observation bias by using $\hat{\alpha}_-$ instead of $\tilde{\alpha}$ is quantitatively important, accounting for the additional variance term $\mathcal{W}_N$ is less important.

---

[21]Specifically, we show in Online Appendix E.2.3 that, if X_{is} corresponds to employment growth rates, then ψ_i will generally not be independent of $(\psi_j, Y_{1j}(0), Y_{2j}(0))$ in others regions $j \neq i$, unless one makes restrictive assumptions about the demand elasticities σ_s, such as $\sigma_s = 0$. We also construct alternative shift-share IVs that satisfy this independence assumption under weaker restrictions on σ_s, but require adjusting the shifter used in estimation.

[22]See, e.g., Maestas, Mullen and Strand (2013); Dobbie and Song (2015); Aizer and Doyle (2015), or Silver (2016).

6 Performance of new methods: placebo evidence

In Section 6.1, we revisit the placebo exercise in Section 2 to examine the finite-sample properties of the inference procedures described in Remarks 5 and 6. In Section 6.2, we show that our baseline placebo results are robust to several changes in the placebo design.

6.1 Baseline specification

We first consider the performance of the standard error estimator in eq. (29) (which we label *AKM*), and the standard error and confidence interval in eqs. (30) and (31) (with label *AKM0*) in the baseline placebo design described in Section 2.[23,24]

For the *AKM* and *AKM0* inference procedures, Table 2 presents median standard error estimates and rejection rates for 5% significance level tests of the null hypothesis $H_0 \colon \beta = 0$. In the case of *AKM0*, since the standard error depends on the null being tested, the table reports the median "effective standard error", defined as the length of the 95% confidence interval divided by 2×1.96.

The results in Table 2 show that the inference procedures introduced in Section 4 perform well. The median *AKM* standard error is slightly lower than the standard deviation of $\hat{\beta}$, by about 5% on average across all outcomes. The median *AKM0* effective standard error is slightly larger than the standard deviation of $\hat{\beta}$, by about 11% on average. The implied rejection rates are close to the 5% nominal rate: the *AKM* procedure has rejection rates between 7.5% and 9.1% and the *AKM0* rejection rates are always between 4.3% and 4.5%. As discussed in Section 4.2, the *AKM* and *AKM0* confidence intervals are asymptotically equivalent. The differences in rejection rates between the *AKM* and *AKM0* inference procedures are thus due to differences in finite-sample performance. As noted in other contexts (see, e.g., Lazarus et al., 2018), imposing the null can lead to improved finite-sample size control. The better size control of the *AKM0* procedure is consistent with these results.

6.2 Alternative placebo specifications

In Section 4, we show theoretically that the *AKM* and *AKM0* inference procedures are valid in large samples only if: (a) the number of sectors goes to infinity; (b) all sectors are asymptotically "small"; (c) the sectoral shocks are independent across sectors. Given these conditions, these inference procedures remain valid under (d) any distribution of the sectoral shifters; and (e) arbitrary correlation structure of the regression residuals. In this section, we evaluate the sensitivity of these inference procedures to requirements (a) to (c) above, and illustrate points (d) and (e) by documenting the robustness of these procedures to alternative distributions of the shifters and the residuals. In all cases, we also report *Robust* and *Cluster* standard errors estimates and rejection rates. We focus on the change in the share of working-age population employed as the outcome variable of interest.

We first evaluate how the performance of different inference procedures depends on the number of sectors. Panel A of Table 3 shows that the overrejection problem affecting standard inference

---

[23]We fix the matrix Z to be a column of ones when implementing the formulas in eqs. (29) and (31).

[24]In Online Appendix D.8, we explore the sensitivity of our results to using counties (instead of CZs) as the regional unit of analysis, and occupations (instead of sectors) as the unit at which the shifter is defined.

Table 2: Median standard errors and rejection rates for $H_0 \colon \beta = 0$ at 5% significance level.

	Estimate		Median eff. s.e.		Rejection rate	
	Mean	Std. dev	AKM	AKM0	AKM	AKM0
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	1.90	2.21	7.8%	4.5%
Employed in manufacturing	−0.01	1.88	1.77	2.06	8.0%	4.3%
Employed in non-manufacturing	0.00	0.94	0.89	1.04	8.2%	4.5%
Panel B: Change in average log weekly wage						
Employed	−0.03	2.66	2.57	2.99	7.5%	4.3%
Employed in manufacturing	−0.03	2.92	2.74	3.18	9.1%	4.5%
Employed in non-manufacturing	−0.02	2.64	2.55	2.96	7.8%	4.5%

Notes: For the outcome variable indicated in the leftmost column, this table indicates the mean and standard deviation of the OLS estimates of β in eq. (1) across the placebo samples (columns (1) and (2)), the median effective standard error estimates (columns (3) and (4)), and the percentage of placebo samples for which we reject the null hypothesis $H_0 \colon \beta = 0$ using a 5% significance level test (columns (5) and (6)). *AKM* is the standard error in Remark 5; and *AKM0* is the confidence interval in Remark 6. The median effective standard error is equal to the median length of the corresponding 95% confidence interval divided by 2×1.96. Results are based on 30,000 placebo samples.

procedures worsens when the number of sectors decreases: the rejection rates of 5% significance level tests based on *Robust* and *Cluster* standard errors reach 70.6% and 56.1%, respectively, when we construct the shift-share covariate using 20 2-digit SIC sectors (instead of the 396 4-digit SIC sectors we use in the baseline placebo). In line with the findings of the literature on clustered standard errors with few clusters, the rejection rates of hypothesis tests that rely on *AKM* standard errors also increase to 12%, but rejection rates for hypothesis tests that apply the *AKM0* inference procedure remain very close to the nominal 5% significance level.

Panels B to D of Table 3 examine the robustness of the results in Tables 1 and 2 to alternative distributions of the shifters. In Panel B, as in our baseline placebo exercise, the shifters are drawn i.i.d. from a normal distribution, but we change the variance to both a lower ($\sigma^2 = 0.5$) and a higher value ($\sigma^2 = 10$) than in the baseline ($\sigma^2 = 5$). In Panel C, we draw the shifters from a log-normal distribution re-centered to have mean zero and scaled to have the same variance as in the baseline. Panel D investigates the robustness of our results to heteroskedasticity in the sector-level shocks. We set variance of the shock in each sector s, to $\sigma_s^2 = 5 + \lambda(n_s - S/N)$. Thus, the cross-sectional average of the variance of the sector-level shocks is the same as in the baseline (which corresponds to setting $\lambda = 0$), but this variance now varies across sectors. Comparison of the results in Panels B to D of Table 3 to those in Tables 1 and 2 suggests that our baseline results are not sensitive to specific details of the distribution of sector-level shifters. This is consistent with the claim (d) above.

Panels E and F of Table 3 explore the robustness of our baseline results to different patterns of correlation in the regression residuals. In the baseline placebo, since $\beta = 0$, the regression residuals inherit the correlation patterns in the outcome variable. Here, we modify these patterns by adding a random shock η_i^m in each placebo sample m to the outcome Y_i. Panel E explores the impact of increasing the correlation between the regression residuals of CZs that belong to the same state.

Table 3: Alternative number of sectors, shifter distributions and residuals' correlation patterns

	Estimate		Median eff. s.e.				Rejection rate			
	Mean (1)	Std. dev (2)	Robust (3)	Cluster (4)	AKM (5)	AKM0 (6)	Robust (7)	Cluster (8)	AKM (9)	AKM0 (10)
Panel A: Sensitivity to the number of sectors										
2-digit ($S = 20$)	-0.01	3.19	0.65	0.96	2.84	6.06	70.6%	56.1%	12.0%	5.8%
3-digit ($S = 136$)	0.00	2.25	0.73	0.94	2.18	2.72	54.2%	42.5%	7.5%	4.5%
Panel B: Sensitivity to the variance of the shifters										
$\sigma^2 = 0.5$	-0.04	6.33	2.33	2.91	6.04	7.02	48.5%	38.0%	7.9%	4.5%
$\sigma^2 = 10$	0.00	1.41	0.52	0.65	1.35	1.57	48.1%	37.8%	7.5%	4.5%
Panel C: Log-normal shifters										
$\sigma^2 = 5$	0.27	2.26	0.86	1.05	2.17	3.7	44.6%	35.3%	7.7%	5.2%
Panel D: Heteroskedastic shifters										
$\lambda = 3$	-0.01	1.63	0.55	0.72	1.51	2.14	52.1%	40.1%	8.7%	4.0%
$\lambda = 7$	0.01	1.38	0.44	0.58	1.23	2.01	53.7%	41.1%	9.5%	4.2%
Panel E: Simulated state-level shocks in regression residual										
	0.00	2.11	0.86	1.11	1.99	2.32	42.8%	30.4%	7.9%	4.6%
Panel F: Simulated 'large' sector shifter in regression residual										
	-0.01	2.01	0.74	0.92	1.90	2.21	48.4%	37.8%	7.9%	4.6%
Panel G: Including a 'large' sector in shift-share regressor										
	-0.02	4.25	0.59	0.76	1.18	1.34	92.0%	89.6%	77.2%	76.3%

Notes: All estimates in this table use the change in the share of the working-age population employed in each CZ as the outcome variable Y_i in eq. (1). This table indicates the mean and standard deviation of the OLS estimates of β in eq. (1) across the placebo samples (columns (1) and (2)), the median effective standard error estimates (columns (3) to (6)), and the percentage of placebo samples for which we reject the null hypothesis $H_0: \beta = 0$ using a 5% significance level test (columns (7) to (10)). *Robust* is the Eicker-Huber-White standard error; *Cluster* is the standard error that clusters CZs in the same state; *AKM* is the standard error in Remark 5; *AKM0* is the confidence interval in Remark 6. For each inference procedure, the median effective standard error is equal to the median length of the corresponding 95% confidence interval divided by 2×1.96. Results are based on 30,000 placebo samples. This table presents results for placebo simulations that depart from the baseline; the results should thus be compared to those in Tables 1 and 2. In Panel A, we reduce the number of sectors relative to the baseline. In Panel B, we change the variance of the distribution from which all shifters are drawn. In Panel C, we assume that the distribution from which all shifters are drawn is log-normal (re-centered at zero) with variance equal to five. In Panel D, we allow the variance of the shock in each sector to be heteroskedastic, $\sigma_s^2 = 5 + \lambda(n_s - S/N)$. In Panel E, we simulate state-level shocks and include them in our regression residual. In Panels F and G, we simulate a shifter for the non-manufacturing sector and include it in our regression residual and in our shift-share regressor, respectively.

Specifically, we generate a random variable $\tilde{\eta}_k^m$ for each state k and simulation m such that $\tilde{\eta}_k^m \sim \mathcal{N}(0, 6)$. We then set $\eta_i^m = \tilde{\eta}_{k(i)}^m$ where $k(i)$ is the state of CZ i. Since we have now increased the relative importance of the correlation pattern accounted for by *Cluster* standard errors, the resulting overrejection decreases from 38.3% to 30.4%. In line with claim (e) above, the rejection rates of the *AKM* and *AKM0* inference procedures are not affected. In Panel F, we evaluate the robustness of our results to adding a shock to the non-manufacturing sector that is included in the regression residual. Specifically, in each simulation m, we set $\eta_i^m = (1 - \sum_{s=1}^S w_{is})\hat{\eta}_S^m$ with $\hat{\eta}_S^m \sim \mathcal{N}(0, 5)$, where $\sum_{s=1}^S w_{is}$ is the 1990 aggregate employment share of the 396 4-digit SIC manufacturing sectors included in the definition of the shift-share regressor of interest. The results in Panel F of Table 3 show that adding

this component to the regression residual does not affect the rejection rates.

Lastly, Panel G in Table 3 explores the consequences of adding the non-manufacturing sector to the shift-share regressor. In Panel F, the shock to the non-manufacturing sector is part of the regression residual; in Panel G, we use this shock, in combination with the shocks to all manufacturing sectors, to construct the shift-share regressor. Across CZs, the average initial employment share in the non-manufacturing sector is 77.5%; i.e. $N^{-1}\sum_{i=1}^{N}(1 - \sum_{s=1}^{S}w_{is}) = 77.5\%$. Including such a large sector in the shift-share regressor violates Assumptions 2(ii) and 2(iii). As a result, the *AKM* and *AKM0* inference procedures overreject severely; standard inference procedures fare even worse, with rejection rates reaching up to 92%. The results in Panels F and G suggest that, provided that the shifters are independent across sectors, it is better to exclude large sectors from the shift-share regressor of interest, and thus let the shocks associated with them enter the regression residual. One should, however, bear in mind that, if β_{is} in eq. (11) varies across sectors, excluding large sectors from the shift-share regressor will change the estimand β (see Proposition 3).

In the placebo simulations described in Tables 1 to 3, we have drawn the shifters independently from a mean-zero distribution. In Table 4, we allow for non-zero correlation in the shifters within "clusters" of sectors.[25] Specifically, we report results from placebo exercises in which the shifters are drawn from the joint distribution $(\mathcal{X}_1^m, \ldots, \mathcal{X}_S^m) \sim \mathcal{N}(0, \Sigma)$, where Σ is an $S \times S$ covariance matrix with elements $\Sigma_{sk} = (1 - \rho)\sigma\,\mathbb{I}\{s = k\} + \rho\sigma\,\mathbb{I}\{c(s) = c(k)\}$ and $c(s)$ indicates the "cluster" that industry s belongs to. In panels A, B, and C, these clusters correspond to the 3-, 2-, and 1-digit SIC sector that the 4-digit SIC sector s belongs to, respectively.

Panel A of Table 4 shows that introducing correlation within 3-digit SIC sectors has a moderate effect on the rejection rates of both the traditional methods and versions of the *AKM* and *AKM0* methods that assume that the sectoral shocks are independent. Rejection rates close to 5% are obtained with versions of the *AKM* and *AKM0* inference procedures that cluster the shifters at a 2-digit SIC level (see Section 5.1). As shown in Panel B, the overrejection problem affecting both traditional inference procedures and versions of the *AKM* and *AKM0* procedures that assume independence of shifters is more severe when the shifters are correlated at the 2-digit level. However, the last two columns show that, in this case, the versions of *AKM* and *AKM0* that cluster the sectoral shocks at the 2-digit level achieve rejection rates close to the nominal level. Finally, Panel C shows that the overrejection problem is much more severe in the presence of high correlation in shifters within the two 1-digit aggregate sectors, and this problem is not solved by clustering at the 2-digit level.

The last panel in Table 4 illustrates the inferential problems that arise in empirical applications of shift-share designs when all shifters are correlated with each other. Such correlations also arise, for example, when all shifters are generated (at least in part) by a common shock with potentially heterogeneous effects across sectors.[26] As simulations presented in Table E.2 in Online Appendix E.1.1 illustrate, if there is a common component affecting all shifters, it is important to first estimate this common component and to control for it in the shift-share regression of interest. Otherwise, hy-

---

[25]In Online Appendix D.2, we study the impact of drawing the shifters from a distribution with non-zero mean. We show that, in line with the discussion in Section 4.2, it is important to control for the region-specific sum of shares $\sum_{s=1}^{S}w_{is}$.

[26]There is an extensive empirical literature documenting the importance of common factors driving changes in sector-specific variables such as sectoral industrial production, employment and value added (see, e.g., Altonji and Ham, 1990; Shea, 2002; Foerster, Sarte and Watson, 2011).

Table 4: Correlation in sectoral shocks

	Estimate		Median eff. s.e.								Rejection rate				
					Independent		2-digit SIC				Independent		2-digit SIC	
	Mean	Std. dev	Robust	Cluster	AKM	AKM0	AKM	AKM0	Robust	Cluster	AKM	AKM0	AKM	AKM0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Panel A: Simulated shifters with correlation within 3-digit SIC sectors														
$\rho = 0.00$	−0.01	2.00	0.74	0.92	1.91	2.22	1.86	2.64	48.6%	37.8%	7.9%	4.6%	8.9%	4.7%
$\rho = 0.50$	0.02	2.14	0.77	1.07	2.03	2.16	2.24	2.87	49.2%	32.3%	6.4%	7.1%	4.4%	4.7%
$\rho = 1.00$	0.01	2.27	0.76	1.08	1.99	2.10	2.38	3.14	52.2%	35.0%	8.6%	9.7%	4.7%	4.8%
Panel B: Simulated shifters with correlation within 2-digit SIC sectors														
$\rho = 0.00$	−0.01	1.99	0.73	0.92	1.90	2.22	1.86	2.65	48.2%	37.7%	7.6%	4.5%	8.8%	4.5%
$\rho = 0.50$	−0.01	2.73	0.73	1.13	1.82	1.89	2.92	4.07	62.3%	43.2%	20.5%	23.5%	5.8%	5.0%
$\rho = 1.00$	0.01	3.20	0.69	1.18	1.67	1.63	3.38	6.16	68.4%	48.1%	31.2%	35.7%	6.3%	5.2%
Panel C: Simulated shifters with correlation within 1-digit SIC sectors														
$\rho = 0.00$	0.01	1.98	0.73	0.92	1.90	2.21	1.85	2.65	48.5%	37.7%	7.3%	4.4%	8.5%	4.4%
$\rho = 0.50$	0.02	4.95	0.74	1.41	1.88	1.59	2.42	2.36	84.2%	72.3%	58.3%	64.7%	42.7%	53.2%
$\rho = 1.00$	0.42	59.63	0.71	1.74	1.86	1.02	2.94	1.67	90.2%	78.3%	75.2%	82.7%	52.1%	65.6%

Notes: All estimates in this table use the change in the share of the working-age population employed in each CZ as the outcome variable Y_i in eq. (1). This table indicates the mean and standard deviation of the OLS estimates of β in eq. (1) across the placebo samples (columns (1) and (2)), the median effective standard error estimates (columns (3) to (8)), and the percentage of placebo samples for which we reject the null hypothesis $H_0: \beta = 0$ using a 5% significance level test (columns (9) to (14)). *Robust* is the Eicker-Huber-White standard error; *Cluster* is the standard error that clusters CZs in the same state; in columns (5) and (11), *AKM* is the standard error in Remark 5; in columns (7), and (13), *AKM* is the standard error in eq. (40) for 2-digit SIC sector clusters; in columns (6) and (12), *AKM0* is the confidence interval in Remark 6; in columns (8) and (14), *AKM0* adjusts the confidence interval in Remark 6 as indicated in Section 5.1. For each inference procedure, the median effective standard error is equal to the median length of the corresponding 95% confidence interval divided by 2×1.96. Results are based on 30,000 placebo samples.

pothesis tests based on standard inference procedures as well as on the *AKM* and *AKM0* inference procedures may suffer from an overrejection problem.

We summarize the conclusions from Tables 3 and 4 in the following remark.

Remark 7. *In shift-share regressions, overrejection of the usual inference procedures is more severe when there is a small number sectors. In this case, the methods we provide attenuate the overrejection problem, but may still overreject when the number of sectors is very small. Our methods perform well under different distributions of shifters and regression residuals, but they lead to an overrejection problem when the shift-share covariate aggregates over a large sector. Finally, when the shifters are not independent across sectors, it is important to properly account for their correlation structure.*

In Online Appendices D.3 to D.7 we present results from additional placebo simulations in which we investigate the consequences of: (a) the violation of the assumption that the shifters of interest are as good as randomly assigned; (b) the presence of serial correlation in both the shifters of interest and the regression residuals, in panel data settings; (c) the true potential outcome function being nonlinear, implying that the linearly additive potential outcome framework in eq. (11) is misspecified; (d) the presence in the regression residuals of shift-share components with shares correlated in different degrees with those entering the shift-share covariate of interest; and, (e) the presence of treatment heterogeneity across regions and sectors.

7 Empirical applications

We now apply the *AKM* and *AKM0* inference procedures to two empirical applications. First, the effect of Chinese competition on U.S. local labor markets, as in Autor, Dorn and Hanson (2013). Second, the estimation of the local inverse elasticity of labor supply, as in Bartik (1991). Additionally, in Online Appendix F, we apply the *AKM* and *AKM0* inference procedures to the study of the impact of immigration on labor market outcomes of U.S. natives.

7.1 Effect of Chinese exports on U.S. labor market outcomes

Autor, Dorn and Hanson (2013, henceforth ADH), explore the impact of exports from China on labor market outcomes across U.S. CZs. Specifically, ADH present IV estimates for a specification that fits within the panel data setting described in Section 5.2, with each region $j = 1, \ldots, 722$ denoting a CZ, each sector $k = 1, \ldots, 396$ denoting a 4-digit SIC industry, and each period $t = 1, 2$ denoting either 1990–2000 changes or 2000–2007 changes. As in Section 5.2, we index here the intersection of a region j and a period t by i, and the intersection of a sector k and a period t by s. In ADH, the outcome Y_{1i} is a ten-year equivalent change in a labor-market outcome, the endogenous treatment is $Y_{2i} = \sum_{s=1}^{S} \bar{w}_{is} \mathcal{X}_s^{US}$, where $\mathcal{X}_s^{US}$ is the change in U.S. imports from China normalized by the start-of-period total U.S. employment in the sector, and $\bar{w}_{is}$ is the start-of-period employment share of a sector in a CZ. ADH use the shift-share IV $X_i = \sum_{s=1}^{S} w_{is} \mathcal{X}_s$, where $\mathcal{X}_s$ denotes imports from China by high-income countries other than the U.S. normalized by a ten-year-lag of the start-of-period total U.S. employment in the sector, and w_{is} is the ten-year-lag of the employment share $\bar{w}_{is}$. To measure

these variables, we use the data sources described in Section 2.1. In all regression specifications, we include a vector of controls Z_i corresponding to the largest set of controls used in ADH.[27]

Table 5 reports 95% CIs computed using different methodologies for the specifications in Tables 5 to 7 in ADH. Panels A, B, and C present the IV, reduced-form and first-stage estimates, respectively. Following Autor et al. (2014), the *AKM* and *AKM0* CIs cluster the shifters $\{\mathcal{X}_s\}_{s=1}^S$ by 3-digit SIC industry; thus, the *AKM* and *AKM0* CIs we report are robust to serial correlation in the shifters as well as to cross-sectoral correlation in the shifters within 3-digit SIC industries. Tables E.4 and E.5 in Online Appendix E.1 report *AKM* and *AKM0* CIs for alternative definitions of clusters.

In Online Appendix E.1, we present placebo simulations that depart from our baseline placebo design in ways that explore specific features of the empirical setting studied in this section. In Table E.1, we draw the shifters from the empirical distribution of shifters used to construct the ADH IV (instead of drawing them from a normal distribution); the resulting rejection rates are very similar to those in the baseline simulation. In Table E.2, we draw shifters that have a common component with factor structure; since the resulting correlation structure cannot be captured by clustering, we show that it is important in this case to include an estimate of the common factor component as an additional control.[28]

In Table 5, state-clustered CIs are very similar to the heteroskedasticity-robust ones. In contrast, our proposed CIs are wider than those implied by state-clustered standard errors. For the IV estimates reported in Panel A, the average increase across all outcomes in the length of the 95% CI is 24% with the *AKM* procedure and 65% with the *AKM0* procedure. When the outcome is the change in the manufacturing employment rate, the length of the 95% CI increases by 26% with the *AKM* procedure and by 65% with the *AKM0* procedure. In light of the lack of impact of state-clustering on the 95% CI, the wider intervals implied by our inference procedures indicate that cross-region residual correlation is driven by similarity in sectoral compositions rather than by geographic proximity.

Panel B of Table 5 reports CIs for the reduced-form specification. In this case, the increase in the CI length is slightly larger than for the IV estimates: across outcomes, it increases on average by 54% for *AKM* and 130% for *AKM0*. The smaller relative increase in the CI length for the IV estimate relative to its increase for the reduced-form estimate is a consequence of the fact that all inference procedures yield similar CIs for the first-stage estimate, as reported in Panel C.

As discussed in Section 6, the differences between *AKM* (or *AKM0*) CIs and state-clustered CIs are related to the importance of shift-share components in the regression residual. The results in Panel C suggest that, once we account for changes in sectoral imports from China to other high-income countries, there is not much sectoral variation left in the first-stage regression residual; i.e., there are no other sectoral variables that are important to explain changes in sectoral imports from China to the

---

[27]See column (6) of Table 3 in ADH. The vector Z_i aims to control for labor supply shocks and labor demand shocks other than the changes in imports from China, and it includes the start-of-period percentage of employment in manufacturing. The discussion in Section 4.2 implies that one should instead control for the ten-year-lagged of the start-of-period employment share in manufacturing, to match the shares that enter the definition of the shift-share IV. However, to facilitate the comparison with the original results in ADH, we use their vector of controls. As shown in Borusyak, Hull and Jaravel (2018), controlling for the ten-year-lagged manufacturing employment shares does not substantively affect the estimates.

[28]For placebo simulation evidence under our baseline assumption that the shifters are independent across 3-digit clusters, using data for outcomes Y_{1i} and shares w_{is} identical to that used in this section, see Online Appendix D.4.

Table 5: Effect of Chinese exports on U.S. commuting zones—Autor, Dorn and Hanson (2013)

	Change in the employment share			Change in avg. log weekly wage		
	All	Manuf.	Non-Manuf.	All	Manuf.	Non-Manuf.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 2SLS Regression						
$\hat{\beta}$	-0.77	-0.60	-0.18	-0.76	0.15	-0.76
Robust	$[-1.10, -0.45]$	$[-0.78, -0.41]$	$[-0.47, 0.12]$	$[-1.23, -0.29]$	$[-0.81, 1.11]$	$[-1.27, -0.25]$
Cluster	$[-1.12, -0.42]$	$[-0.79, -0.40]$	$[-0.45, 0.10]$	$[-1.26, -0.26]$	$[-0.81, 1.11]$	$[-1.28, -0.24]$
AKM	$[-1.25, -0.30]$	$[-0.84, -0.35]$	$[-0.54, 0.18]$	$[-1.37, -0.15]$	$[-0.81, 1.11]$	$[-1.42, -0.10]$
AKM0	$[-1.69, -0.39]$	$[-1.01, -0.36]$	$[-0.84, 0.14]$	$[-1.77, -0.17]$	$[-1.49, 1.05]$	$[-1.97, -0.19]$
Panel B: OLS Reduced-Form Regression						
$\hat{\beta}$	-0.49	-0.38	-0.11	-0.48	0.10	-0.48
Robust	$[-0.71, -0.27]$	$[-0.48, -0.28]$	$[-0.31, 0.08]$	$[-0.80, -0.16]$	$[-0.50, 0.69]$	$[-0.83, -0.13]$
Cluster	$[-0.64, -0.34]$	$[-0.45, -0.30]$	$[-0.27, 0.05]$	$[-0.78, -0.18]$	$[-0.51, 0.70]$	$[-0.81, -0.15]$
AKM	$[-0.81, -0.17]$	$[-0.52, -0.23]$	$[-0.35, 0.12]$	$[-0.88, -0.07]$	$[-0.50, 0.69]$	$[-0.93, -0.03]$
AKM0	$[-1.24, -0.24]$	$[-0.67, -0.25]$	$[-0.64, 0.08]$	$[-1.27, -0.10]$	$[-1.16, 0.61]$	$[-1.47, -0.11]$
Panel C: 2SLS First-Stage						
$\hat{\beta}$		0.63				
Robust		$[0.46, 0.80]$				
Cluster		$[0.45, 0.81]$				
AKM		$[0.53, 0.73]$				
AKM0		$[0.54, 0.84]$				

Notes: $N = 1,444$ (722 CZs $\times$ 2 time periods). Observations are weighted by the start of period CZ share of national population. All regressions include the full vector of baseline controls in ADH; i.e. those in column (6) of Table 3 in Autor, Dorn and Hanson (2013). 95% confidence intervals are reported in square brackets. *Robust* is the Eicker-Huber-White standard error; *Cluster* is the standard error that clusters of CZs in the same state; *AKM* is the standard error in eq. (40) with 3-digit SIC clusters; *AKM0* is the confidence interval with 3-digit SIC clusters described in the last sentence of Section 5.1.

U.S.[29] To investigate this claim, Table E.3 in Online Appendix E.1 reports the rejection rates implied by a placebo exercise designed to match the first-stage specification reported in Panel C of Table 5. The placebo results show that, while traditional methods still suffer from severe overrejection when no controls are included, the overrejection is attenuated once we include as controls the shift-share IV and the control vector Z_i we use in Table 5, indicating that these variables soak up much of the cross-CZ correlation in the treatment variable used in ADH.

Overall, Table 5 shows that, despite the wider confidence intervals obtained with our procedures, the qualitative conclusions in ADH remain valid at usual significance levels. However, the increased width of the 95% CI shows that the uncertainty regarding the magnitude of the impact of Chinese import exposure on U.S. labor markets is greater than that implied by usual inference procedures. In particular, the *AKM0* CI is much wider than that based on state-clustered standard errors; furthermore due to its asymmetry around the point estimate, using the *AKM0* CI, we cannot rule out impacts of

---

[29]This is analogous to what we would observe in a regression in which the regressor of interest varies at the state level, and we control for all state-specific covariates affecting the outcome variable: state-clustered standard errors would be similar to heteroskedasticity-robust standard errors, since there is little within-state correlation left in the residuals.

the China shock that are two to three times larger than the point estimates of these effects.[30]

7.2 Estimation of inverse labor supply elasticity

In our second application, we estimate the inverse labor supply elasticity. Specifically, using the notation of Section 3, we estimate the parameter $\tilde{\phi}$ in the equation

$$\hat{\omega}_i = \tilde{\phi}\hat{L}_i + \delta Z_i + \epsilon_i, \qquad \tilde{\phi} = \phi^{-1}, \tag{46}$$

where $\hat{L}_i$ denotes the log change in the employment rate in CZ i, $\hat{\omega}_i$ denotes the log change in wages, Z_i is a vector of controls, and ϵ_i is a regression residual. We use the same sample, data sources, and vector of controls Z_i as in Section 7.1.[31]

The model in Section 3 has implications for the properties of different strategies for estimating the inverse labor supply elasticity $\tilde{\phi}$. By eq. (10), the residual ϵ_i in eq. (46) accounts for changes in labor supply shocks, $\sum_{g=1}^{G} \tilde{w}_{ig}\hat{v}_g + \hat{v}_i$, not controlled for by the vector Z_i. Second, it follows from eq. (8) that, up to a first-order approximation around an initial equilibrium, changes in regional employment rates, $\hat{L}_i$, can be written as a function of both shift-share aggregators of sectoral labor demand shocks and the same labor supply shocks potentially entering ϵ_i in eq. (10), $\sum_{g=1}^{G} \tilde{w}_{ig}\hat{v}_g + \hat{v}_i$. Thus, $\hat{L}_i$ and ϵ_i will generally be correlated and the OLS estimator of $\tilde{\phi}$ in eq. (46) will be biased. However, as discussed in Section 4.3, the model in Section 3 also implies that we can instrument for $\hat{L}_i$ using shift-share aggregators of sectoral labor demand shocks that are independent of the unobserved labor supply shocks (see Online Appendix E.2 for more details).

In this section, we use three different shift-share IVs to estimate $\tilde{\phi}$ in eq. (46). For each of them, Table 6 presents the reduced-form, first-stage and 2SLS estimates. First, in Panel A, we use the instrumental variable in Bartik (1991); i.e. $\hat{X}_i = \sum_{i=1}^{N} w_{is}\hat{L}_s$, where $\hat{L}_s$ denotes the nation-wide employment growth in sector s. Second, in Panel B, we use the leave-one-out version of this instrument; i.e. $\hat{X}_i = \sum_{i=1}^{N} w_{is}\hat{L}_{s,-i}$, where $\hat{L}_{s,-i}$ denotes the employment growth in sector s over all CZs excluding CZ i.[32] Third, in Panel C, we use the IV used in Autor, Dorn and Hanson (2013), which we denote as ADH IV and describe in detail in Section 7.1.[33] As in Section 7.1, we report versions of the *AKM* and *AKM0* CIs with shifters clustered at the 3-digit SIC industry for all periods.

Column (3) of Table 6 shows that the estimates of the inverse labor supply elasticity are similar no matter which IV we use: 0.80 when using the original Bartik IV, 0.82 when using the leave-one-out

---

[30]It follows from Remark 6 (see the expression for the quantity A) that the asymmetry in the *AKM0* CI comes from the correlation between the regression residuals $\hat{R}_s$ and the shifters cubed. In large samples, this correlation is zero and the *AKM* and *AM0* CIs are asymptotically equivalent. The differences between both CIs in Table 5 thus reflect differences in their finite-sample properties. This notwithstanding, the placebo exercise presented in Online Appendix D.4 shows that both inference procedures yield close to correct rejection rates in a sample analogous to that used in ADH.

[31]Table E.7 in Online Appendix E.2 investigates the robustness of our results to alternative sets of controls.

[32]The leave-one-out version of the instrument in Bartik (1991) was originally proposed by Autor and Duggan (2003). In Online Appendix E.2.3, we clarify the assumptions under which the model in Section 3 is consistent with the validity of the leave-one-version of the Bartik IV. Online Appendix E.2.4 presents placebo exercises attesting that the *AKM* and *AKM0* CIs reported in this section have appropriate coverage in the context of this empirical application.

[33]The effect of these IVs on the changes in the employment rate may be heterogeneous across regions and sectors (see eq. (8)). This does not affect the validity of our inference procedures since, as discussed in Section 4.3, we allow for heterogeneous effects in the first-stage regression.

Table 6: Estimation of inverse labor supply elasticity

Dependent variable:	First-Stage $\hat{L}_i$ (1)	Reduced-Form $\hat{\omega}_i$ (2)	2SLS $\hat{\omega}_i$ (3)
Panel A: Bartik IV—Not leave-one-out estimator			
$\hat{\beta}$	0.90	0.73	0.80
Robust	$[0.70, 1.10]$	$[0.54, 0.91]$	$[0.64, 0.97]$
Cluster	$[0.64, 1.16]$	$[0.47, 0.98]$	$[0.60, 1.01]$
AKM	$[0.65, 1.16]$	$[0.49, 0.96]$	$[0.62, 0.98]$
AKM0	$[0.61, 1.17]$	$[0.44, 0.96]$	$[0.59, 1.02]$
Panel B: Bartik IV—Leave-one-out estimator			
$\hat{\beta}$	0.87	0.71	0.82
Robust	$[0.68, 1.06]$	$[0.53, 0.89]$	$[0.65, 0.98]$
Cluster	$[0.62, 1.12]$	$[0.46, 0.96]$	$[0.60, 1.03]$
AKM (*leave-one-out*)	$[0.59, 1.15]$	$[0.47, 0.94]$	$[0.61, 1.02]$
AKM0 (*leave-one-out*)	$[0.53, 1.15]$	$[0.42, 0.94]$	$[0.59, 1.09]$
Panel C: ADH IV			
$\hat{\beta}$	-0.72	-0.48	0.67
Robust	$[-1.04, -0.39]$	$[-0.80, -0.16]$	$[0.36, 0.98]$
Cluster	$[-0.93, -0.50]$	$[-0.78, -0.18]$	$[0.35, 0.99]$
AKM	$[-1.19, -0.24]$	$[-0.88, -0.07]$	$[0.27, 1.07]$
AKM0	$[-1.83, -0.35]$	$[-1.27, -0.10]$	$[0.18, 1.14]$

Notes: $N = 1,444$ (722 CZs × 2 time periods). The variable $\hat{L}_i$ denotes the log-change in the employment rate in CZ i. The variable $\hat{\omega}_i$ denotes the log change in mean weekly earnings. Observations are weighted by the start of period CZ share of national population. All regressions include the full vector of baseline controls in ADH; i.e. those in column (6) of Table 3 in Autor, Dorn and Hanson (2013). 95% confidence intervals in square brackets. *Robust* is the Eicker-Huber-White standard error; *Cluster* is the standard error that clusters of CZs in the same state; *AKM* is the standard error in eq. (40) with 3-digit SIC clusters; *AKM0* is the confidence interval with 3-digit SIC clusters described in the last sentence of Section 5.1; *AKM (leave-one-out)* is the standard error in Section 5.3 with 3-digit SIC clusters; *AKM0 (leave-one-out)* is the confidence interval with 3-digit SIC clusters described in Section 5.3.

version of this estimator, and 0.67 when using the ADH IV.[34] In both Panel A and Panel B, the *AKM* and *AKM0* CIs are very similar to the state-clustered CI. In Panel C, the *AKM* and *AKM0* CIs are only moderately wider than those obtained with state-clustered standard errors.

Columns (1) and (2) of Table 6 show the first-stage and reduced-form estimates, respectively. In Panel A and Panel B, the *AKM0* CIs are similar to the state-clustered CIs; in contrast, in Panel C, the first-stage and reduced-form *AKM0* CIs more twice as wide, and more than three times as wide as the state-clustered CI, respectively. Thus, the first-stage and reduced-form *AKM* and *AKM0* CIs differ more from the state-clustered CI when the ADH IV is used than when the Bartik IV is used. A possible explanation for this finding is that the shift-share component of the first-stage and reduced-form regression residuals is much smaller in the latter than in the former case. The Bartik IV absorbs

---

[34]One explanation for the similarity between the leave-one-out and the original Bartik IV is that, as discussed in Section 5.3, the bias of the original Bartik IV is of the order $\frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \frac{w_{is} \hat{w}_{is}}{\hat{n}_s}$. This quantity equals 0.004 in this application, indicating that the own-observation bias is likely to be small.

the bulk of the shift-share covariates that affect the change in the employment rate and wages across CZs. In contrast, the ADH IV is just one of the possibly various shift-share terms affecting the change in the outcome and endogenous treatment of interest. With the remaining shift-share entering the regression residual, it becomes quantitatively important to use our inference procedures to obtain CIs with the right coverage.

8   Concluding remarks

This paper studies inference in shift-share designs. We show that standard economic models predict that changes in regional outcomes depend on observed and unobserved sector-level shocks through several shift-share terms. Our model thus implies that the residual in shift-share regressions is likely to be correlated across regions with similar sectoral composition, independently of their geographic location, due to the presence of unobserved shift-share terms. Such correlations are not accounted for by inference procedures typically used in shift-share regressions, such as when standard errors are clustered on geographic units. To illustrate the importance of this shortcoming, we conduct a placebo exercise in which we study the effect of randomly generated sector-level shocks on actual changes in labor market outcomes across CZs in the United States. We find that traditional inference procedures severely overreject the null hypothesis of no effect. We derive two novel inference procedures that yield correct rejection rates.

It has become standard practice to report cluster-robust standard errors in regression analysis whenever the variable of interest varies at a more aggregate level than the unit of observation. This practice guards against potential correlation in the residuals that arises whenever these residuals contain unobserved shocks that also vary at the same level as the variable of interest. In the same way, we recommend that researchers report confidence intervals in shift-share designs that allow for a shift-share structure in the residuals, such as one of the two confidence intervals that we propose.

University of Chicago Booth School of Business

Princeton University

Princeton University

References

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

⎯ **and Pascual Restrepo**, "Demographics and Automation," 2018. Unpublished manuscript, MIT.

⎯ **and** ⎯ , "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2019, *forthcoming*.

Adão, Rodrigo, "Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil," 2016. Unpublished manuscript, University of Chicago.

_ , **Costas Arkolakis, and Federico Esposito**, "Spatial linkages, Global Shocks, and Local Labor Markets: Theory and Evidence," 2019. Unpublished manuscript, University of Chicago.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc J. Melitz, "The Impact of Exports on Innovation: Theory and Evidence," 2018. Unpublished manuscript, Harvard University.

Aizer, Anna and Jr Doyle Joseph J., "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *The Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

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

Altonji, Joseph G. and John C. Ham, "Variation in Employment Growth in Canada: The Role of External, National, Regional, and Industrial Factors," *Journal of Labor Economics*, 1990, *8* (1), 198–236.

Anderson, Theodore W. and Herman Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 1949, *20* (1), 46–63.

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

Autor, D. H. and M. G. Duggan, "The Rise in the Disability Rolls and the Decline in Unemployment," *The Quarterly Journal of Economics*, 2003, *118* (1), 157–206.

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

_ , _ , **and** _ , "When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men," *American Economic Review: Insights*, 2018, *forthcoming.*

_ , _ , _ , **and Jae Song**, "Trade adjustment: Worker-level evidence," *The Quarterly Journal of Economics*, 2014, *129* (4), 1799–1860.

_ , _ , _ , **and Kaveh Majlesi**, "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," 2017. Unpublished manuscript, University of Zurich.

_ , _ , _ , **Gary Pisano, and Pian Shu**, "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents," *American Economic Review: Insights*, 2019, *forthcoming.*

Bai, Jushan, "Panel Data Model with Interactive Fixed Effects," *Econometrica*, 2009, *77* (4), 1229–1279.

Barrios, Thomas, Rebecca Diamond, Guido W. Imbens, and Michal Kolesár, "Clustering, Spatial Correlation, and Randomization Inference," *Journal of the American Statistical Association*, 2012, *107* (498), 578–591.

Bartelme, Dominick, "Trade Costs and Economic Geography: Evidence from the U.S.," 2018. Unpublished manuscript, University of Michigan.

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

Baum-Snow, Nathaniel and Fernando Ferreira, "Causal inference in urban and regional economics," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 3–68.

Bekker, Paul A., "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 1994, *62* (3), 657–681.

Bell, Robert M. and Daniel F. McCaffrey, "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples," *Survey Methodology*, 2002, *28* (2), 169–181.

Beraja, Martin, Erik Hurst, and Juan Ospina, "The Aggregate Implications of Regional Business Cycles," Technical Report 21956, National Bureau of Economic Research February 2019.

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental Shift-share Research Designs," September 2018. arXiv:1806.01221 [econ.EM].

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel, "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 348–400.

_ , **Gordon Hanson, Lin Tian, and Jonathan Vogel**, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the U.S.," 2018. Unpublished manuscript, University of California Los Angeles.

Cameron, Colin A. and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2014, *50* (2), 317–372.

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Chao, John C., Norman R. Swanson, Jerry A. Hausman, Whitney K. Newey, and Tiemen Woutersen, "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments," *Econometric Theory*, 2012, *12* (1), 42–86.

Che, Yi, Yi Lu, Justin R. Pierce, Peter K. Schott, and Tao Zhigang, "Did Trade Liberalization with China Influence U.S. Elections?," 2017. Unpublished manuscript, Yale University.

Chodorow-Reich, Gabriel and Johannes Wieland, "Secular Labor Reallocation and Business Cycles," *Journal of Political Economy*, 2018, *forthcoming.*

Colantone, Italo and Piero Stanig, "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe," *American Journal of Political Science*, October 2018, *62* (4), 936–953.

Conley, Timothy G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, September 1999, *92* (1), 1–45.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dix-Carneiro, Rafael and Brian K. Kovak, "Trade Liberalization and Regional Dynamics," *American Economic Review*, October 2017, *107* (10), 2908–46.

_ , Rodrigo Soares, and Gabriel Ulyssea, "Economic Shocks and Crime: Evidence From The Brazilian Trade Liberalization," *American Economic Journal: Applied Economics*, 2018, *10* (4), 158–195.

Dobbie, Will and Jae Song, "Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection," *American Economic Review*, 2015, *105* (3), 1272–1311.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler, "The Impact of Immigration: Why Do Studies Reach Such Different Results?," *Journal of Economic Perspectives*, 2016, *30* (4), 31–56.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson, "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 2011, *119* (1), 1–38.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi, "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," October 2018. Unpublished manuscript, University of California Berkeley.

Gobillon, Laurent and Thierry Magnac, "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, July 2016, *98* (3), 535–551.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," Technical Report 24408, National Bureau of Economic Analysis, 2018.

Greenstone, Michael, Alex Mas, and Hoai-Luu Nguyen, "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times," *American Economic Journal: Economic Policy*, 2015, *forthcoming.*

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang, "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 2014, *104* (6), 1597–1629.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: an Introduction*, New York, NY: Cambridge University Press, 2015.

Imbens, Guido Wilhelmus and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Jaeger, David A, Joakim Ruist, and Jan Stuhler, "Shift-share instruments and the impact of immigration," Technical Report 24285, National Bureau of Economic Research February 2018.

Jones, Ronald W., "A Three-Factor Model in Theory, Trade and History," in Jagdish Bhagwati, Ronald Jones, Robert Mundell, and Jaroslav Vanek, eds., *Trade, Balance of Payments and Growth*, Amsterdam: North-Holland, 1971.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens, "Identification and Inference With Many Invalid Instruments," *Journal of Business & Economic Statistics*, October 2015, *33* (4), 474–484.

Kovak, Brian K, "Regional effects of trade reform: What is the correct measure of liberalization?," *American Economic Review*, 2013, *103* (5), 1960–76.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson, "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, October 2018, *36* (4), 541–559.

Lee, Eunhee, "Trade, Inequality, and the Endogenous Sorting ofHeterogeneous Workers," 2018. Unpublished manuscript, University of Maryland.

Lewis, Ethan and Giovanni Peri, "Immigration and the Economy of Cities and Regions," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 625–685.

Maestas, Nicole, Kathleen J Mullen, and Alexander Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, August 2013, *103* (5), 1797–1829.

Monras, Joan, "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," December 2018. Unpublished manuscript, CEMFI.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Pierce, Justin R. and Peter K. Schott, "Trade Liberalization and Mortality: Evidence from US Counties," *American Economic Review: Insights*, 2018, *forthcoming.*

Roy, Andrew D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, *3* (2), 135–146.

Shea, John, "Complementarities and Comovements," *Journal of Money, Credit and Banking*, 2002, *34* (2), 412–433.

Silver, David, "Haste or Waste? Peer Pressure and the Distribution of Marginal Returns to Health Care," 2016. Working Paper, Princeton University.

Topalova, Petia, "Trade Liberalization, Poverty and Inequality: Evidence from Indian Districts," in Ann Harrison, ed., *Globalization and Poverty*, University of Chicago Press, 2007, pp. 291–336.

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