Shift-Share Designs: Theory and Inference*

Rodrigo Adão† Michal Kolesár‡ Eduardo Morales§

June 17, 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.

*We thank Kirill Borusyak, Peter Egger, Gordon Hanson, Bo Honoré, seminar participants at Carleton University, CEMFI, Chicago Fed, EESP-FGV, Fed Board, GTDW, Johns Hopkins University, Princeton University, PUC-Rio, Stanford University, UC Irvine, Unil, University of Michigan, University of Nottingham, University of Virginia, University of Warwick, and Yale University, and conference participants at the CEPR-CURE conference, CEPR-ERWIT conference, EIIT-FREIT, Globalization & Inequality BFI conference, NBER-ITI winter meeting, and Princeton-IES Workshop, for very useful comments and suggestions. We thank Juan Manuel Castro Vincenzi for excellent research assistance. We thank David Autor, David Dorn and Gordon Hanson for sharing their code and data. All errors are our own.

†University of Chicago Booth School of Business. Email: rodrigo.adao@chicagobooth.edu

‡Princeton University. Email: mkolesar@princeton.edu

§Princeton University. Email: ecmorale@princeton.edu

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

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

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. Although shift-share specifications are increasingly common in many contexts, and they have been used in numerous influential studies, including Bartik (1991), Blanchard and Katz (1992), Card (2001) and Autor, Dorn and Hanson (2013), 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 unclustered errors—are much smaller than the true 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% if heteroskedasticity-robust standard errors are used 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 stylized model are a sector-region 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 depends on a region-specific intercept, which may aggregate group-specific labor supply shifters (e.g. push factors that raise immigration from different countries of origin).

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 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 the presence of shift-share components in the residuals. 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 sector-level shocks $\mathcal{X}_s$, conditioning on the realized shares w_{is}, controls Z_i, and residuals ϵ_i. This sampling approach is directly motivated by our economic model: we are interested in what would have happened to outcomes if the sector-level shock of interest had taken different values, holding everything else constant. Our framework allows for heterogeneous effects of the shifters: a 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 a large number of 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 observe multiple observations of each region over time. Finally, we cover applications in which the shifter is unobserved, but can be estimated using 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 deliver estimates that are close to the true standard deviation 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 of interest 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 relative to the nominal significance level, although it still attenuates the overrejection problem in comparison with the usual standard error formulas. If the shifters are not independent across sectors, we show that it is important to allow for clustering of the shifters at the appropriate level.

In the final part of the paper, we illustrate the implications of our new inference procedure for three popular applications of shift-share regressions. First, we study of 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). Lastly, we use changes in the stock of immigrants from various origin countries to investigate the impact of immigration on employment and wages across occupations, education groups, and Commuting Zones in the United States, as in the literature following Altonji and Card (1991) and Card (2001).

Our proposed confidence intervals for the estimated 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 almost identical to those constructed using standard approaches. In our last application, how much our method matters for the confidence interval of the estimated effect of immigration shocks on natives' local labor market outcomes depends on the unit of analysis at which the outcome is measured. We find that confidence intervals do not change significantly when outcomes are measured at the intersection of local labor markets and disaggregated occupations, but they increase when measuring outcomes by regions and education groups, or by regions and aggregate occupations.

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 strategies 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 consequences of various observable 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, 2017, 2018), and industry reallocation (Chodorow-Reich and Wieland, 2018). Shift-share regressors have been extensively 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 of this literature in Lewis and Peri (2015) and Dustmann, Schönberg and Stuhler (2016). Furthermore, recent papers have explored versions of shift-share strategies to estimate the effect on firms of shocks to outsourcing costs and foreign demand (Hummels et al., 2014; Aghion et al., 2018).[2]

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 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, as this approach follows naturally from our economic model.[3] 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 in 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 et al., 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 several prior applications of shift-share designs, and Section 8 concludes. Appendix A contains proofs for all propositions in Section 4. Additional results are collected in Online Appendices B, C, D, E and F.

2 Overrejection of usual standard errors: placebo evidence

In this section, we implement a placebo exercise to evaluate the finite-sample performance of two inference methods most commonly applied in shift-share regression designs: (a) Eicker-Hubert-White—

[2]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.

[3]In contrast, the approach in Goldsmith-Pinkham, Sorkin and Swift (2018) is less appealing in the context our model, in which the shares w_{is} are endogenous equilibrium objects.

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).[4]

The placebo samples differ exclusively in the vector of shifters $(\mathcal{X}_1^m, \dots, \mathcal{X}_S^m)$, which is 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 as *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 length of the different standard error estimates, and rejection rates for 5% significance level tests of the null hypothesis $H_0\colon \beta = 0$. We present these statistics for several outcome variables Y_i, which are listed in the leftmost column.

Column (1) of Table 1 shows that, up to simulation error, the average of the estimated coefficients is indeed 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.[5] 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

[4]We are very grateful to the authors for sharing their code and datasets with us.

[5]Figure E.1 in Online Appendix E.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 median 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.

all outcomes, the median magnitudes of the heteroskedasticity-robust and state-clustered standard errors are, respectively, 55% and 46% lower than the true standard deviation.

The downward bias in the *Robust* and *Cluster* standard errors translates into a severe overrejection of the null hypothesis $H_0\colon \beta = 0$. 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 sectoral shock of interest—for instance, trade flows, tariffs, national employment or the number of foreign workers employed in an industry. If the researchers set out to evaluate the impact of these shocks 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 for two applications: (i) the estimation of the impact of sector-specific labor demand shifters or labor group-specific labor supply shifters on regional labor market outcomes; and (ii) the estimation of the regional inverse labor supply elasticity. We describe the model fundamentals in Section 3.1, discuss its main implications for these applications 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. The shifter D_{is} 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}:

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

Any of the terms on the right-hand side of this equation may equal zero and, thus, this decomposition is without loss of generality.

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 such that regional markets are differentially affected by group-specific labor supply shocks. In particular, indexing labor groups by $g = 1, \ldots, G$, we decompose

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

where ν_g is a group-specific labor supply shifter, $\tilde{w}_{ig}$ measures the exposure of region i to group g-specific labor supply shifter, and ν_i captures region-specific factors affecting labor supply. The variable ν_g thus captures factors that affect the supply of labor of group g across all regions in the population of interest. Workers may be classified into groups according to their education level, gender, age, 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 $i = 1, \ldots, N$ 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 $\{\nu_i\}_{i=1}^{N}$.

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

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

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

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

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

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

Consider first the model's implications for the impact of changes in sector-specific labor demand and group-specific labor supply shifters on regional labor market outcomes. We focus here on the impact of the demand shocks $\{\hat{\chi}_s, \hat{\mu}_s\}_{s=1}^{S}$ and the supply shocks $\{\hat{\nu}_g\}_{g=1}^{G}$ 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 a 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 always the initial employment share l_{is}^{0}, and the "shift" term corresponds 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 g-specific shock, $\tilde{\omega}_{ig}$. Conditional on any sector s and any labor group g, the shares $\{l^0_{is}\}_{i=1}^N$ and $\{\tilde{\omega}_{ig}\}_{i=1}^N$ 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^0_{is}, the impact of a sectoral labor demand shocks on a region's 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^0_{is} but also on θ_{is}, which may vary across i and s. Similarly, even conditional on the exposure measure $\tilde{\omega}_{ig}$, the impact of group-specific labor supply shocks may be heterogeneous across regions; i.e. the impact of $\hat{v}_g$ on $\hat{L}_i$ depends not only on $\tilde{\omega}_{ig}$ but also on λ_i. While datasets usually contain information on the initial employment shares for every sector and region $\{l^0_{is}\}_{i=1,s=1}^{N,S}$, and on the exposure measures $\{\tilde{\omega}_{ig}\}_{i=1,g=1}^{N,G}$ for every group and region, the parameters $\{\theta_{is}\}_{i=1,s=1}^{N,S}$ and $\{\lambda_i\}_{i=1}^N$ are not generally known.

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

Remark 2. *In our model, the equilibrium equation 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, $\{\sigma_s\}_{s=1}^S$ and ϕ, and structural parameters, or in the interpretation of the different terms entering the labor demand shifter D_{is} or the labor supply shifter v_i.[6] In addition, Online Appendix C.3 shows that similar insights arise in a model featuring migration across regions. In this case, the change in regional employment $\hat{L}_i$ 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 regions $i = 1, \dots, N$. In this environment, for example, $l^0_{is}\theta_{is}$ 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 second to the estimation of the inverse labor supply elasticity, eqs. (4) and (5) imply that

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

It follows from eq. (8) that the change in region i's employment rate, $\hat{L}_i$, also depends on the term $\sum_{g=1}^{G} \tilde{\omega}_{ig}\hat{v}_g + \hat{v}_i$. 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, according to eqs. (8) and (9), one can write the inverse labor supply elasticity as the ratio of the impact of any given sector-specific demand shock (e.g. $\hat{\chi}_s$) on wages to that on

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

employment:

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

In Sections 4 and 5, we use the model described here to provide an economic interpretation for the econometric assumptions that we impose when discussing the identification and estimation in general shift-share designs. These assumptions imply restrictions on the joint distribution of labor supply and demand shocks $F(\cdot)$ introduced in eq. (7). In Section 7, we return again to this economic model when interpreting empirical estimates of (a) the impact of sector-specific labor demand shifters (Section 7.1) and group-specific labor supply shifters (Section 7.3) on regional labor market outcomes; and, (b) 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,

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

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. According to eq. (11), 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 shifter of interest is the sectoral labor demand shock $\hat{\chi}_s$, and the outcome of interest is the employment change $\hat{L}_i$, we can map eq. (8) into eq. (11) by defining

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

Observe that $Y_i(0)$ aggregates all shifters other than the sectoral shifter of interest $\hat{\chi}_s$.[7]

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 using data on the N regions.[8] To help us focus

[7] 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 as it helps connect our analysis to the linear econometric specification used extensively in the shift-share literature. See Section 6.3.3 and Online Appendix E.4 for a discussion of the approximation error arising from the linear specification imposed in eq. (8).

[8] 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

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). The OLS estimator of the coefficient on X_i in this simplified setting is given by

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

and we can write the regression equation as

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

The definition of the estimand β in eq. (14) and the properties of the estimator $\hat{\beta}$ will depend on: (a) what the population of interest is; 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.[9] 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.

Given our assumptions on the population of interest and on the type of repeated sampling, 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 $\sum_{g=1}^{G} \tilde{w}_{ig}\hat{\nu}_g$) 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.

only observe noisy estimates of the shifters.

[9] 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).

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 Appendix A.

When presenting our assumptions and results, 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 $\mathcal{A}_s'$. If $L = 1$, then A and $\mathcal{A}$ 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).

4.1.1 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 $S \to \infty$.[10] 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 Appendix A.1. As in eq. (15), let $\mathcal{F}_0 = (Y(0), B, W)$.

[10] We do not restrict the relative rates of growth of N and S, so that the number of regions may grow at a faster or a slower rate than the number of sectors.

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

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

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 $(\mathcal{X}_1, \dots, \mathcal{X}_S)$ on the outcome Y_i. This approximation may be subject to error. In Appendix A.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 $\{\mathcal{X}_s\}_{s=1}^S$ 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 economic model in Section 3 (so that $\mathcal{X}_s = \hat{\chi}_s$), 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 the labor supply shifters, $\{\nu_g\}_{g=1}^G$ and $\{\nu_i\}_{i=1}^N$. The plausibility of this restriction depends on the particular details of a specific empirical application. For example, if all N regions in the sample are regions within a small open economy, one may think of $\hat{\chi}_s$ as capturing changes in international prices in sector s, and $\hat{\mu}_s$ as capturing either changes in the tariffs that this small open economy charges on its sector s imports, or sector s-specific productivity changes common to all regions in this small open economy; in this case, Assumption 1(ii) requires changes in tariffs and changes in productivity in the regions forming the population of interest to be independent of the changes in international prices (see Online Appendix B.4 for additional details).

Assumption 2 (Consistency and Inference)**.** (i) The shifters $(\mathcal{X}_1, \dots, \mathcal{X}_S)$ are independent conditional on $\mathcal{F}_0$; (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.[11] 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 residual may be very complex—see in particular eqs. (12), (15) and (16). The residual

[11]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.

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 the joint distribution of these regression residuals 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. We do not have to worry about correctly specifying this correlation structure, as one would under the alternative approach above. Our approach allows (but does not require) the residual to have a shift-share structure;[12] 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.[13] 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 these shifters are independent across the clusters. We evaluate the quantitative importance of several violations of this assumption through placebo simulations in Section 6.2. Additionally, in the context of the empirical application discussed 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 all 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. Assumption 2(ii) is needed for 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 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.[14]

[12] Since our inference is valid conditionally on the residuals $\{\epsilon_i\}_{i=1}^{N}$, it automatically accounts for whatever correlation structure they may have, including spatial, or, as in applications with multiple periods, temporal correlations. We study the case with multiple periods in Section 5.2.

[13] This conceptualization of the shares as equilibrium objects that respond (at least partly) to the same set of economic shocks is consistent with the economic 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 $s \neq s'$.

[14] Since $\max_s n_s^2 / \sum_{t=1}^{S} n_t^2 \geq \max_s n_s^2 / \max_{s'} n_{s'} \sum_{t=1}^{N} n_t = \max_s n_s / \sum_{t=1}^{N} n_t$, Assumption 2(iii) implies Assumption 2(ii).

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.

4.1.2 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 5(i) to 5(iii) in Appendix A.1 hold. Then*

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

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

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 $\operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)$. 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, $\sum_{s=1}^{S}\pi_{is}\beta_{is} = \operatorname{var}(\mathcal{X}_{s(i)} \mid \mathcal{F}_0)\beta_{is(i)}$ 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 total 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, $\operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0) = \sigma^2$, and $\xi_{is} = w_{is}$; in this case, a consistent estimate of τ_ξ is given by $\hat{\beta}\sum_{i=1}^{N}\sum_{s=1}^{S} w_{is}^2 / \sum_{i=1}^{N}\sum_{s=1}^{S} w_{is}$.[15]

Proposition 2. *Suppose Assumptions 1 and 2, and Assumption 5 in Appendix A.1 hold. Suppose also that*

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

converges in probability to a non-random limit. Then

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

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 the sector sizes n_s are all of the order N/S, the rate of convergence equals $\sqrt{S}$. However, 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 $\mathcal{V}_N$, 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

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

Replacing $\operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)$ 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

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

We show formally that this variance estimate leads to valid inference under regularity conditions in Section 4.2; and in Appendix A.6 for the case in which β_{is} is heterogeneous.

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 specializes in, in line 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:

[15] In general, one can consistently estimate τ_ξ or τ_ξ^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.

Remark 4. *In the expression for $\mathcal{V}_N$ in eq. (19), the expectation is only taken over $\mathcal{X}_s$—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 in the placebo exercise 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 only differ in the middle sandwich, with the cluster-robust variance 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

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

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 $i \neq j$,

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

which by the law of iterated expectations implies that $E[\hat{\mathcal{V}}_{AKM}(\beta) - \hat{\mathcal{V}}_{CL}(\beta) \mid W] \geq 0$. It can be seen from this expression 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 group-level 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 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 $\sum_{s=1}^{S} w_{js} w_{js}$ is large). In particular, 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 $Z_i' = (Z_{i1}, \dots, Z_{iK})$, 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

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

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,

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

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.

4.2.1 Assumptions

For clarity of exposition, we focus here on the main substantive assumptions and relegate technical regularity conditions to Appendix A.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 facilitate the interpretation of this restriction, it is useful to 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, 2017). Splitting the impact of automation into nationwide sector-specific effects, as captured 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 $\{\mathcal{Z}_s\}_{s=1}^{S}$.

4.2.2 Asymptotic theory

We now study the asymptotic properties of $\hat{\beta}$ in the setting described in Section 4.2.1. The following result generalizes Proposition 1:

Proposition 3. *Suppose Assumptions 2(i) and 2(ii) and Assumptions 5(i) to 5(iii) in Appendix A.1 hold with $\mathcal{F}_0 = (\mathcal{Z}, U, Y(0), B, W)$. Suppose also that Assumptions 3(i) to 3(iii) and Assumptions 6(i) and 6(ii) in Appendix A.1 hold. Then*

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

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 5 and 6 in Appendix A.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

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

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

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

We show in Appendix A.6 that this standard error estimate remains valid if β_{is} is heterogeneous across regions and sectors, as long as some mild regularity conditions hold.

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 therefore also 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.[16] In particular, 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

[16]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.

$\hat{\epsilon}_{\beta_0,i}$, the residual from regressing $Y_i - X_i\beta_0$ onto Z_i (that is, $\hat{\epsilon}_{\beta_0,i}$ 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 were not rejected. It follows from simple algebra that the endpoints of this CI are a solution to a quadratic equation, so that they are available in closed form—one does not have to numerically search for all the hypotheses that were 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).

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, the 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.

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.[17] 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. In analogy with eq. (11), we denote the region-i treatment level that would occur if the region received shocks $(x_1, \dots, x_S)$ as

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

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

The framework in eqs. (33) and (34) maps directly to the problem of the 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

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

and $Y_{2i}(0)$ is given by the expression for $Y_i(0)$ in eq. (12).[18] 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.

[17] 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).

[18] 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.

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

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

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

4.3.1 Assumptions

The identification restriction that we impose here 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 $Y_{2i} = Y_{2i}(\mathcal{X}_1, \dots, \mathcal{X}_s)$ 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,

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

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

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 context of the model in Section 3, with the mapping between the economic 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 instrumental variable 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 $\{\hat{\nu}_g\}_{g=1}^{G}$.[19] For example, if $\{\hat{\chi}_s\}_{s=1}^{S}$ are observed 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 $(w_{i1}, \dots, w_{iS})$ are as good as

[19] 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, and that it more readily generalizes to the case where α is heterogeneous.

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 $(w_{i1}, \dots, w_{iS})$ in the first-stage equation. For estimation and inference, this alternative approach requires that, conditionally on the shifters, either the shares $(w_{i1}, \dots, w_{iS})$ 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, it can be seen from eq. (35) 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 both of the region-specific exposure shares to each labor group $\{\tilde{w}_{ig}\}_{g=1}^G$ and of the region-specific labor supply shock ν_i. This assumption will be violated if regions that are 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 immigrants coming from a country of origin g (e.g. to changes in the number of migrants from Mexico, as caused, for example, by a currency crisis in Mexico; see Monras, 2018).[20]

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 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 $w_{is} = l_{is}^0$ 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 $\{w_{is}\}_{i=1,s=1}^{N,S}$ depend on the same set of sector-specific labor demand shifters $\{(\chi_s, \mu_s)\}_{s=1}^S$, they will generally be correlated across regions.[21]

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.

[20] To allow for a shift-share component in the structural residual, Goldsmith-Pinkham, Sorkin and Swift (2018) view the shares $(w_{i1}, \dots, w_{iS})$ 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 $\{w_{is}\}_{s=1}^S$ 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. Thus, in contrast to Kolesár et al. (2015), the condition $S \to \infty$ is not needed for identification; it is only needed for consistency and inference. Consequently, the large-sample results we obtain below are quite different from those in Kolesár et al. (2015).

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

4.3.2 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 D.2 for details and proof),

$$\frac{N}{\sqrt{\sum_{s=1}^{S} n_s^2}}(\hat{\alpha} - \alpha) = \mathcal{N}\left(0, \frac{\mathcal{V}_N}{(\frac{1}{N}\sum_{i=1}^{N}\ddot{X}_i Y_{2i})^2}\right) + o_p(1),\ \mathcal{V}_N = \frac{\sum_{s=1}^{S}\operatorname{var}(\tilde{\mathcal{X}}_s \mid \mathcal{F}_0) R_s^2}{\sum_{s=1}^{S} n_s^2},\ R_s = \sum_{i=1}^{N} w_{is}\epsilon_i, \tag{38}$$

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

$$\widehat{se}(\hat{\alpha}) = \frac{\sqrt{\sum_{s=1}^{S}\widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{|\sum_{i=1}^{N}\ddot{X}_i Y_{2i}|} = \frac{\sqrt{\sum_{s=1}^{S}\widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{\sum_{i=1}^{N}\ddot{X}_i^2 |\hat{\beta}|}, \qquad \hat{R}_s = \sum_{i=1}^{N} w_{is}\hat{\epsilon}_i, \tag{39}$$

where $\widehat{\mathcal{X}}_s$ 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 $\hat{\beta} = \sum_{i=1}^{N}\ddot{X}_i Y_{2i} / \sum_{i=1}^{N}\ddot{X}_i^2$ 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 $H_0\colon \alpha = \alpha_0$, 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.

5 Extensions

We now discuss three extensions to the basic setup. In Section 5.1, we relax the assumption that the shifters $(\mathcal{X}_1, \dots, \mathcal{X}_S)$ are independent, and allow them to be correlated within clusters of sectors. Section 5.2 generalizes our results to panel data settings in which we observe multiple observations for each region i. Finally, 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 $c(s) \in \{1, \dots, C\}$ 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 $C \to \infty$,

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

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 $\widehat{\mathcal{X}}_s \hat{R}_s$,

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

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 the formula 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 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 shares $\mathcal{X}_{kt}$ 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 shares $\mathcal{X}_{kt}$ 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 $\{\mathcal{X}_{kt}\}_{t=1}^{T}$ 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 may be unobserved. We follow the setup in Section 4.3 but we assume that, instead of observing $\mathcal{X}_s$ directly, we only observe a noisy measurement

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

for each sector s and region i. 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 the weighted estimate $\hat{\mathcal{X}}_s = \sum_{i=1}^{N} \check{w}_{is} X_{is}/\check{n}_s$, where $\check{n}_s = \sum_{i=1}^{N} \check{w}_{is}$ and the weights $\check{w}_{is}$ are not necessarily related to w_{is}. The resulting estimate of X_i is given by

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

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. We also consider the leave-one-out estimator

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

where $\hat{\mathcal{X}}_{s,-i} = \sum_{j=1}^{N} \mathbb{I}\{j \neq i\}\check{w}_{js}X_{js}/\check{n}_{s,-i}$ 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 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 $\psi_i = (\psi_{i1}, \dots, \psi_{iS})'$ 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 D.2 for a formal statement). In Online Appendix F.2, we use the model in Section 3 to show how this setup covers various approaches to the estimation of the inverse labor supply elasticity.

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. Therefore, including the ith observation in the construction of the estimate $\hat{X}_i$ 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 et al., 1999), such as when one uses group indicators as instruments.[22] We show in Online Appendix D.2 that this bias is of the order $\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{S}\frac{w_{is}\check{w}_{is}}{\check{n}_s}\leq S/N$, so that consistency of $\tilde{\alpha}$ generally requires the number of regions to grow more slowly than the number of sectors. Furthermore, unless $S^{3/2}/N\to 0$, the resulting asymptotic bias in $\tilde{\alpha}$ will induce undercoverage in confidence intervals based on $\tilde{\alpha}$.

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 Online Appendix D.2. We also show in this appendix that, under regularity conditions, its asymptotic distribution is given by

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

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

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

The term $\mathcal{W}_N$ accounts for the additional uncertainty stemming from the fact that the shift-share regressor is estimated. It is analogous to the many-instrument term in the jackknife IV estimator under many instrument asymptotics (see Chao et al., 2012). We give simulation evidence in Online Appendix F.2.4 showing that, while correcting for the own-observation bias by using the leave-one-out estimator $\hat{\alpha}_{-}$ instead of $\tilde{\alpha}$ is quantitatively important, accounting for the additional variance term $\mathcal{W}_N$ is less important, at least for the designs that we consider.

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 changes in the number of sectors and to the choice of distribution of both the shifters of interest and the regression residuals. We also illustrate the importance of controlling for the region-specific sum of shares, $\sum_{s=1}^{S}w_{is}$. In Section 6.3, we consider additional placebo exercises that examine the sensitivity of our standard errors to assumptions underlying their validity, and show that the overrejection problem of commonly used inference procedures is pervasive.

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]

[22] For empirical examples of this setting, see, for example, Maestas, Mullen and Strand (2013); Dobbie and Song (2015); Aizer and Doyle (2015), or Silver (2016).

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

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 (1)	Std. dev (2)	AKM (3)	AKM0 (4)	AKM (5)	AKM0 (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 median 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. 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.

For the *AKM* and *AKM0* inference procedures, Table 2 presents the median length of the standard errors 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. For *AKM*, the "effective standard error" is the actual standard error.

The results in Table 2 show that the inference procedures introduced in Section 4 perform well. The median standard error based on *AKM* is slightly lower than the true standard deviation of the estimator $\hat{\beta}$, by about 5% on average across all outcomes. The median effective standard error based on *AKM0* is slightly larger than the standard deviation of $\hat{\beta}$, by about 11% on average.[24] The implied rejection rates are close to the 5% nominal rate: the *AKM* procedure has a rejection rate that is between 7.5% and 9.1% and the *AKM0* procedure has a rejection rate that is 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 in Table 2 are thus due to differences in finite-sample performance. It has been noted in other contexts (see, e.g., Lazarus et al., 2018) that imposing the null can lead to improved finite-sample size control. The better size control of the *AKM0* procedure is consistent with these results. Intuitively, imposing the null reduces the estimation noise in the estimated regression residuals, which helps reduce the finite-sample bias that arises in estimating the asymptotic variance of $\hat{\beta}$.

[24]For the placebo exercise that uses the change in the employment rate as outcome variable, Figure E.2 in Online Appendix E.1 presents histograms representing the empirical distribution of the effective standard errors.

6.2 Alternative placebo specifications

As discussed in Section 4, the inference procedures described in Remarks 5 and 6 are valid in large samples if: (a) the number of sectors goes to infinity; (b) all sectors are asymptotically "small"; (c) the sectoral shocks are independent across sectors; and (d) either the sector-level shocks are mean zero or, alternatively, we control for region-specific sum of shares $\sum_{s=1}^{S} w_{is}$ (see Borusyak, Hull and Jaravel, 2018, for additional discussion of this last point). Given these conditions, the *AKM* and *AKM0* inference procedures remain valid under (e) any distribution of the sectoral shifters; and (f) arbitrary correlation structure of the regression residuals. In this section, we evaluate the sensitivity of these inference procedures to assumptions (a) to (d) above, and illustrate points (e) and (f) by documenting the robustness of these procedures to alternative distributions of the sectoral shifters and of the residuals. In all cases, we also report *Robust* and *Cluster* standard error estimates and rejection rates. To simplify the exposition of the results, 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 is affected by changes in the number of sectors. Panel A of Table 3 shows that the overrejection problem affecting standard inference 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 (instead of a 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 draw the variance of the shock in each sector s, σ_s^2, i.i.d. from a uniform distribution with support $[2, 7]$. Then, in each sample m, we draw $\mathcal{X}_s^m$ i.i.d from the distribution $\mathcal{N}(0, \sigma_s^2)$. Thus, the cross-sectoral average of the variance of the sector-level shocks is the same as in the baseline, but this variance now varies across sectors. Comparing the results in Panels B to D of Table 3 to those in Tables 1 and 2, we observe that our baseline results are not sensitive to specific details of the distribution of sector-level shifters. This is consistent with the claim (e) above.

Panels E and F of Table 3 explore the robustness of our baseline results to different patterns of cross-regional correlation in the regression residuals. In the baseline placebo, since $\beta = 0$, the regression residuals inherit the cross-regional 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 variable Y_i. Panel E explores the impact of increasing the correlation across the regression residuals of CZs that belong to the same state. Specifically, for each simulation m, we generate a random variable

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										
$\sigma_s^2 \sim U[2,7]$	0.02	2.01	0.74	0.92	1.92	2.25	48.5%	38.6%	7.9%	4.6%
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 median 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 $H_0 \colon \beta = 0$ 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. Panels A to G present 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 different shifters to be drawn from distributions with different variance. In Panel E, we simulate state-level shocks and include them in our regression residual. In Panel F, we simulate a shifter for the non-manufacturing sector and include it in our regression residual. In Panel G, we simulate a shifter for the non-manufacturing sector and include it in our shift-share regressor.

$\tilde{\eta}_k^m$ for each state k 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 the *Cluster* standard errors, the overrejection problem affecting these standard errors is less severe: it goes down from 38.3% to 30.4%. Importantly, in line with claim (f) 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 manufacturing employment share in region i in 1990 (i.e. 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 this additional component of the regression

Table 4: Controlling for the size of the residual sector in each CZ

	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: Shifters with mean equal to zero										
No controls	0.01	1.99	0.74	0.92	1.91	2.23	48.0%	37.7%	7.6%	4.5%
Control: $1-\sum_{s=1}^{S} w_{is}$	-0.02	1.43	0.74	0.84	1.31	1.52	33.6%	28.4%	11.2%	4.7%
Panel B: Shifters with mean different from zero										
No controls	-4.67	1.28	0.71	0.94	1.48	1.66	99.1%	97.8%	85.4%	87.6%
Control: $1-\sum_{s=1}^{S} w_{is}$	0.00	1.43	0.74	0.84	1.31	1.52	33.3%	27.8%	11.1%	4.6%

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 median 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 $H_0\colon \beta = 0$ 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. In Panel A, $(\mathcal{X}_1^m, \ldots, \mathcal{X}_S^m)$ is drawn i.i.d. from a normal distribution with zero mean and variance equal to five in each placebo sample. In Panel B, $(\mathcal{X}_1^m, \ldots, \mathcal{X}_S^m)$ is drawn i.i.d. from a normal distribution with mean equal to one and variance equal to five in each placebo sample. For each of the two panels, the first row presents results in which no control is accounted for in the estimating equation; the second row presents results in which we control for the size of the residual sector.

residual does not alter the rejection rates documented in our baseline results.

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%.

Panels F and G in Table 3 show that, provided that the shifters are independent across sectors, it is better to exclude large sectors from the shift-share regressor of interest and 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 depart from the mean-zero assumption; in Table 5, we allow for non-zero correlation across in the shifters across "clusters" of sectors.

As discussed in Section 4.2, controlling for the region-specific sum of shares, $\sum_{s=1}^{S} w_{is}$, is important if the shifters have non-zero mean. In our placebo setting, this is equivalent to controlling for the CZ-specific share of employment in the non-manufacturing sector in 1990, $1-\sum_{s=1}^{S} w_{is}$; we refer to this control here as the "residual sector control". Panel A in Table 4 shows that, when the shifters are mean zero, the mean of $\hat{\beta}$ is not affected by whether we include the residual sector control.[25] Panel B in Table 4 shows that, if the shifter mean is non-zero, the OLS estimate of β in eq. (1) suffers from substantial bias when the residual sector control is not included in the regression; this bias disappears once it is included.[26]

In Table 5, we report results from placebo exercises in which the shifters are drawn from the joint distribution

$$(\mathcal{X}_1^m, \ldots, \mathcal{X}_S^m) \sim \mathcal{N}\left(0, \Sigma\right), \tag{46}$$

where Σ is an $S \times S$ covariance matrix with elements $\Sigma_{sk} = (1-\rho)\sigma\,\mathbb{I}\{s = k\} + \rho\sigma\,\mathbb{I}\{c(s) = c(k)\}$. Here $c(s)$ indicates the "cluster" that industry s belongs to (recall that each industry s corresponds to a 4-digit SIC sector). In panels A, B, and C, these clusters correspond to the 3-, 2-, and 1-digit SIC sector that industry s belongs to, respectively.

Panel A of Table 5 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 (see Remarks 5 and 6). 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 these shifters are correlated at the 2-digit level. However, the last two columns of Table 5 show that, in this case, the versions of *AKM* and *AKM0* 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 sector-level shocks within the two 1-digit aggregate sectors, and this overrejection problem is not solved by clustering at the 2-digit level.

We summarize the conclusions from Tables 3 to 5 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. When the shifters are not independent across sectors, it is important to cluster the shifters at the appropriate level. Finally, if the shares do not sum to one in each region, it is important to control for the residual sector size to allow the shifters to have a non-zero mean.*

[25] However, including the residual sector control attenuates the overrejection problem of traditional inference methods. Intuitively, this control soaks part of the correlation in residuals that traditional inference methods do not take into account.

[26] In Panel B, $\mathcal{X}_s^m \sim \mathcal{N}(1,5)$, and the estimator in the first row of this panel suffers from negative bias because the positive mean of the shifters creates a positive correlation between the shift-share regressor of interest and the control $\sum_{s=1}^{S} w_{is}$, which captures the larger secular decline in the employment rate in regions initially specialized in manufacturing production.

Table 5: 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 median 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.

6.3 Additional Placebo Exercises

6.3.1 Confounding sector-level shocks: omitted variable bias and solutions

In Online Appendix E.2, we investigate the consequences of violations of the assumption that the shifters of interest are mean independent of other shocks affecting the outcome variable (Assumption 1(ii)). Our simulations show that confounding sector-level shocks induce an omitted variable bias in the OLS estimator. In the same appendix, we also illustrate the performance of two solutions to this omitted variable bias problem: (i) the inclusion of regional controls as proxies for latent sector-level shocks, and (ii) the use of a shift-share instrumental variable. These two solutions are discussed theoretically in Sections 4.2 and 4.3, respectively. Our results show that, unless the region-level controls are good proxies for the latent sector-level confounders, the bias persists and an IV is needed. Our results also show that, in this IV setting, while traditional inference methods suffer from an overrejection problem, the *AKM* and *AKM0* inference procedures yield correct rejection rates.

6.3.2 Panel data: serial correlation in residuals and shifters

In Online Appendix E.3, we extend our baseline placebo setting to two periods (1990–2000 and 2000–2007) to illustrate the consequences of serial correlation in both the shifters of interest and the regression residuals. Consistently with the discussion in Section 5.2, serial correlation in residuals does not affect the properties of the *AKM* and *AKM0* inference procedures, but serially correlated shifters can generate moderate overrejection. For good size control, one must compute the standard error formula in eq. (40) with clusters containing sector-period pairs associated with the same underlying sector.

6.3.3 Misspecification of linearly additive potential outcome framework

The potential outcome framework in eq. (11) is linear and additive in the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$. As shown in eqs. (8) and (9), according to the model in Section 3, this functional form is consistent with a first-order approximation around the initial equilibrium to the exact expressions determining the impact of labor demand and supply shocks on regional labor market outcomes. In Online Appendix E.4, we study the consequences of the presence of an approximation error on the right-hand side of eq. (9) (the conclusions would be analogous if we had focused on eq. (8)).

The extent to which the impact of sector-specific labor demand shocks on local wages predicted by the first-order approximation in eq. (9) is close to the impact predicted by the exact (non-linear) model-based expression depends on: (a) the values of the labor demand shifter D_{is} in eq. (2) in the initial equilibrium; (b) the size of the shocks. To calibrate the values of $\{D_{is}^{0}\}_{i=1,s=1}^{N,S}$, we use eq. (2) and data on $\{L_{is}^{0}\}_{i=1,s=1}^{N,S}$ and $\{\omega_i^{0}\}_{i=1}^{N}$. Given these calibrated values of the model fundamentals, we present results for various distributions of the shocks.

Our conclusions are twofold. First, when the variance of the shifters of interest is low, the treatment effects predicted by the estimated linear specification are a good approximation to the true impacts predicted by the non-linear solution to the model. However, as the variance of the shifters grows, the quality of the linear approximation deteriorates. Second, irrespective of the economic interpretation that one may be able to attach to the estimand β in eq. (1) when the true potential outcome structure is non-linear, the *AKM* and *AKM0* inference procedures adequately capture the sampling uncertainty in the OLS estimator of β.

6.3.4 Unobserved shift-share components with different shares

Through eqs. (21) and (22), we show in Section 4.1.2 that the overrejection problem affecting traditional inference procedures is more severe when regions with similar shares (regions i and j with high values of $\sum_{s=1}^{S} \mathcal{X}_s^2 w_{is} w_{js}$) also tend to have correlated regression residuals (high values of $E[\epsilon_i \epsilon_j \mid W]$). This type of correlation arises if the residuals contain shift-share terms with shares that are either identical to or correlated with the shares entering the shift-share covariate of interest. In Online Appendix E.5, we present an extension of our baseline placebo setting to illustrate this point. Instead of using observed changes in regional labor market outcomes as the variable Y_i, we simulate it in each draw m as $Y_i^m = \sum_{s=1}^{S} \tilde{w}_{is} \mathcal{A}_s^m$, where every $\tilde{w}_{is}$ is constructed by adding noise to the corresponding share w_{is} entering the shift-share term of interest, and each $\mathcal{A}_s^m$ is drawn i.i.d. from a normal distribution with mean zero and variance equal to five. Our results indicate that the overrejection problem affecting robust and state-clustered standard errors becomes less severe as we reduce the correlation between w_{is} and $\tilde{w}_{is}$. However, the rejection rates are still above 10% even when the correlation between w_{is} and $\tilde{w}_{is}$ is as low as 0.17. For all cases, the rejection rates of the *AKM* and *AKM0* inference procedures remain stable and close to the 5% nominal rate.

6.3.5 Other extensions

Online Appendix E.6 explores the sensitivity of our results to alternative definitions of the units at which the outcome variable and the shifters are measured. When using counties (instead of CZs) as the regional unit of analysis, Table E.7 shows that rejection rates are very similar to those in Tables 1 and 2. When using 331 occupations (instead of 396 sectors) to construct the shift-share term of interest, Table E.8 shows that: (a) the overrejection problem affecting traditional inference methods is even more severe than in the baseline placebo simulations reported in Section 2; (b) the *AKM* inference procedure attenuates the problem, but still yields rejections rates higher than the nominal significance level; and (c) the *AKM0* inference procedure yields tests with the correct rejection rate.

7 Empirical applications

We now apply the *AKM* and *AKM0* inference procedures to three 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). Finally, the impact of immigration on U.S. natives' labor market outcomes, as in the literature reviewed by Lewis and Peri (2015) and Dustmann, Schönberg and Stuhler (2016).

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 presents 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 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 a 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 total U.S. employment in the sector, and w_{is} is a ten-year-lag of the employment share $\bar{w}_{is}$. To obtain information on 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 6 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, so that 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 F.4 and F.5 in Online Appendix F.1 report *AKM* and *AKM0* CIs for alternative definitions of clusters. The appendix also considers a placebo exercise in which the shifters 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 (see Online Appendix F.1.1 for details).[28]

In all panels of Table 6, state-clustered CIs are very similar to the heteroskedasticity-robust ones. In contrast, our proposed confidence intervals 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 23% with the *AKM* procedure and 66% with the *AKM0* procedure. When the outcome variable is the change in the manufacturing employment rate, the length of the 95% CI increases by 37% with the *AKM* procedure and by 90% 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

[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 exports from China, and it includes the start-of-period percentage of employment in manufacturing. The discussion in Section 4.2.1 implies that one should instead control for the ten-year-lagged start-of-period employment share in manufacturing, to match the shares that enter the definition of the shift-share instrument. 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] The placebo exercise in Section 6.3.2 uses data for outcomes Y_{1i} and shares w_{is} identical to that used in this section; those placebo results are thus informative about the finite-sample properties of the different inference procedures under our baseline assumption that the shifters are independent across 3-digit clusters. In Table F.1 in Online Appendix F.1.1, we consider an alternative placebo exercise, in which we draw the shifters from the empirical distribution of the shifters used to construct the ADH IV (i.e., appropriately normalized imports from China by high-income countries other than the U.S.).

Table 6: 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 (1)	Manuf. (2)	Non-Manuf. (3)	All (4)	Manuf. (5)	Non-Manuf. (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, −0.29]	[−0.85, −0.35]	[−0.54, 0.18]	[−1.36, −0.16]	[−0.80, 1.10]	[−1.41, −0.12]
AKM0	[−1.72, −0.39]	[−1.02, −0.36]	[−0.85, 0.13]	[−1.76, −0.19]	[−1.49, 1.03]	[−1.97, −0.21]
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, −0.16]	[−0.52, −0.23]	[−0.35, 0.13]	[−0.87, −0.09]	[−0.50, 0.69]	[−0.92, −0.04]
AKM0	[−1.25, −0.24]	[−0.68, −0.25]	[−0.64, 0.08]	[−1.26, −0.12]	[−1.15, 0.60]	[−1.46, −0.13]
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. 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.

indicate that cross-region residual correlation is driven by similarity in sectoral compositions rather than by geographic proximity.

Panel B of Table 6 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 131% 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 very similar CIs for the first-stage estimate, as reported in Panel C.

As discussed in Section 6, the differences between the *AKM* (or the *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 the changes in sectoral imports from China to the U.S.[29] To investigate this claim, Table F.3 in Online Appendix F.1 reports the

[29]Intuitively, this is similar 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.

rejection rates implied by a placebo exercise designed to match the first-stage specification reported in Panel C of Table 6. 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 6, indicating that these variables soak up much of the cross-CZ correlation in the treatment variable used in ADH.

Overall, Table 6 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 the 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 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. Using the notation of Section 3, we are interested in the inverse labor supply elasticity $\tilde{\phi}$ in the equation

$$\hat{\omega}_i = \tilde{\phi}\hat{L}_i + \delta Z_i + \epsilon_i, \qquad \tilde{\phi} = \phi^{-1}, \tag{47}$$

where $\hat{L}_i$ denotes log changes in the employment rate in CZ i, $\hat{\omega}_i$ denotes log changes in wages, Z_i is a vector of controls, and ϵ_i is a regression residual. We rely here on the same sample and the same vector of controls Z_i as in Section 7.1 (i.e. the vector of controls listed in column (6) of Table 3 of ADH), and use the data sources described in Section 2.1 to measure every variable in eq. (47).[31]

We use the model in Section 3 to evaluate different strategies to estimate the parameter $\tilde{\phi}$ in eq. (47). Using eq. (10), we conclude that the residual ϵ_i in eq. (47) accounts for changes in local labor supply shocks, $\sum_{g=1}^{G} \tilde{w}_{ig}\hat{v}_g + \hat{v}_i$, not controlled for by the vector Z_i. At the same time, as discussed in Section 3.2, one can write, to a first-order approximation around an initial equilibrium, changes in regional employment rates, $\hat{L}_i$, as a function of both several shift-share aggregators of sectoral labor demand shocks and the same local labor supply shocks entering 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. (47) will be biased. However, as discussed in Section 4.3, the model in Section 3 also implies that we can instrument for changes in employment rates using shift-share aggregators of sectoral labor demand shocks that are independent of the unobserved regional labor supply shocks. We show this formally in Online Appendix F.2.

In this section, we use three different shift-share instrumental variables to estimate $\tilde{\phi}$ in eq. (47). For each of them, Table 7 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-

[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 6 thus reflect differences in their finite-sample properties. This notwithstanding, the placebo exercise presented in Section 6.3.2 shows that both inference procedures yield close to correct rejection rates in a sample analogous to that used in ADH.

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

Table 7: 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	$[0.68, 1.13]$	$[0.51, 0.95]$	$[0.60, 1.00]$
AKM0	$[0.66, 1.14]$	$[0.48, 0.95]$	$[0.59, 1.03]$
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*)	$[0.62, 1.11]$	$[0.49, 0.93]$	$[0.60, 1.03]$
AKM0 (*leave-one-out*)	$[0.60, 1.12]$	$[0.47, 0.93]$	$[0.60, 1.08]$
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	$[-1.16, -0.27]$	$[-0.87, -0.09]$	$[0.27, 1.07]$
AKM0	$[-1.62, -0.38]$	$[-1.16, -0.13]$	$[0.22, 1.14]$

Notes: $N = 1,444$ (722 CZs $\times$ 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 1980 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.

wide employment growth in the 4-digit SIC manufacturing 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 instrumental variable 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.

The results in column (3) of Table 7 show that the estimates of the inverse labor supply elasticity $\tilde{\phi}$

[32] The leave-one-out version of the instrument in Bartik (1991) was originally proposed by Autor and Duggan (2003) to address endogeneity concerns and possible biases in the 2SLS estimation. In Online Appendix F.2.3, we use the model in Section 3 to link the use of estimated leave-one-out shifters and the assumptions introduced in Section 5.3. Online Appendix F.2.4 presents placebo exercises attesting that the inference procedure proposed in Section 5.3 yields appropriate coverage in the context of this empirical application.

[33] As exemplified in eq. (8), the effect of either of these three instrumental variables on the changes in the employment rate may be heterogeneous across regions and sectors. This does not affect the properties of our inference procedures since the versions of *AKM* and *AKM0* CIs presented in Section 4.3 allow for heterogeneous effects in the first-stage regression.

are similar no matter which instrumental variable we use: 0.80 when using the original Bartik IV, 0.82 when using the leave-one-out version of this estimator, and 0.67 when using the ADH IV. In Panel A, the *AKM* and *AKM0* CIs are very similar to the state-clustered and heteroskedasticity-robust CIs. In Panel B, the *AKM* and *AKM0* CIs adjusted to account for the estimated shifters are also very similar to the state-clustered standard errors. In Panel C, the *AKM* and *AKM0* CIs are moderately wider than those obtained with state-clustered standard errors.

The length of *AKM* and *AKM0* CIs relative to the usual CIs is more sensitive to the instrumental variable used when we focus on the first-stage and reduced-form specifications (see columns (1) and (2) of Table 7). In both Panel A and Panel B, the *AKM0* CIs are nearly identical to the state-clustered CIs but, in Panel C, the first-stage and reduced-form *AKM0* CIs are 188% and 72% larger than the state-clustered CIs, respectively. Thus, the first-stage and reduced-form *AKM* and *AKM0* CIs differ more from the CIs based on standard inference procedures 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 the bulk of the shift-share covariates that affect the change in the employment rate and wages across CZs. In contrast, the ADH shift-share IV is just one of the possibly various shift-share terms affecting the change in outcome and endogenous treatment of interest, implying that our inference procedure has a larger impact on the length of the CIs.

7.3 Effect of immigration on U.S. local labor markets

As a third application, we estimate the impact of immigration on labor market outcomes in the United States. To this end, we estimate the linear model

$$Y_{it} = \beta \Delta ImmShare_{it} + Z_{it}'\delta + \epsilon_{it}, \tag{48}$$

where, for observation or cell i, Y_{it} is the change in a labor market outcome for native workers between years t and $t-10$, $\Delta ImmShare_{it}$ is the change in the share of immigrants in total employment between years t and $t-10$, and Z_{it} is a control vector that includes fixed effects. Following Dustmann, Schönberg and Stuhler (2016), one may classify different approaches to estimation of β in eq. (48) on the basis of the definition of the cell i: in the *skill-cell approach*, i corresponds to an education-experience cell defined at the national level (e.g. Borjas, 2003); in the *spatial approach*, i corresponds to a region (e.g. Altonji and Card, 1991); in the *mixed approach*, i corresponds to the intersection of a region and an occupation, or a region and an education group (e.g. Card, 2001). In the *spatial* and *mixed* approaches, since Altonji and Card (1991) and Card (2001), it has become common to instrument for the change in the immigrant share $\Delta ImmShare_{it}$ using a shift-share IV:

$$X_{it} = \sum_{g=1}^{G} ImmShare_{igt_0} \frac{\Delta Imm_{gt}}{Imm_{gt_0}}, \tag{49}$$

where g indexes countries (or groups of countries) of origin of immigrants, and t_0 is some pre-sample or beginning-of-the-sample time period. The variable $ImmShare_{igt_0}$ plays the role of the share w_{is} in eq. (1) and denotes the share of immigrants from origin g in total immigrant employment in cell i in year t_0; the ratio $\Delta Imm_{gt}/Imm_{gt_0}$ plays the role of the shifter $\mathcal{X}_s$ in eq. (1), with ΔImm_{gt} denoting the change in the total number of immigrants coming from origin g between years t and $t-10$, and Imm_{gt_0} denoting the total number of immigrants from region g at the national level in year t_0.

Table 8 presents results for three different implementations of the *mixed approach*.[34] In all three cases, the data comes from a three-period panel with $t = 1990, 2000, 2010$, $t_0 = 1980$, and we use information on $G = 51$ countries of origin of immigrants.[35] The implementations differ in the definition of a cell i. In columns (1) to (4) of Table 8, a cell corresponds to the intersection of a CZ and one of the 50 occupations defined in Appendix F of Burstein et al. (2018). In columns (5) and (6), we define a cell as the intersection of a CZ and one of two education groups: high school-equivalent or college-equivalent educated workers (see Card, 2009). In columns (7) to (10), a cell corresponds to the intersection of a CZ and one of seven aggregate occupations (see Card, 2001).[36]

The magnitude and statistical significance of the results is consistent across specifications. A one percentage point increase in the share of immigrants in total employment reduces the number of native workers employed by 1.19–1.49%, with all three estimates being statistically different from zero at the 5% level for all four inference procedures that we consider. For all three cell definitions and all wage outcomes (change in average weekly wages for all, only high-skill, or only low-skill native workers), the estimated impact of an increase in the immigrant share is not statistically different from zero at the 5% significance level according to the *AKM* and *AKM0* CIs. These results are consistent with those based on *Robust* and *Cluster* standard errors when observations correspond to an intersection of a CZ and an education group, but these standard inference procedures sometimes predict that immigration has a positive effect on the wages of high-skill workers when occupations are used to define the unit of analysis (see columns (3) and (9) in Table 8).[37]

There is considerable heterogeneity across the specifications in the length of the *AKM* and *AKM0* confidence intervals relative to those based on *Robust* and *Cluster* standard errors. In columns (1) to (4), which use detailed occupations to define cells, *AKM* and *AKM*0 confidence intervals tend to be very similar (in some cases, even slightly smaller) to those based on state-clustered standard errors, although they are generally much larger than those based on robust standard errors. In contrast, for

[34] Results for different versions of the *spatial approach* are presented in Online Appendix F.3.

[35] Information on all variables comes from the Census Integrated Public Use Micro Samples for 1980–2000 and the American Community Survey for 2008–2012. Although information is available for a larger set of countries or areas of origin of immigrants, to ensure Assumptions 2(ii) and 2(iii) hold, we restrict the sample to 51 areas with total immigrant share in $t_0 = 1980$ below 3%; i.e. $\sum_i ImmShare_{igt_0}/\sum_i \sum_{g'} ImmShare_{ig't_0} \leq 0.03$ (see Table F.8 in Online Appendix F.3 for a list of these areas of origin). A comparison of the results of the placebo simulations reported in Tables F.10 and F.11 in Online Appendix F.3 shows that the behavior of the different inference procedures improves when we restrict the sample to these 51 areas of origin. For completeness, Table F.17 in Online Appendix F.3 shows estimates analogous to those in Table 8 when a larger set of areas of origin of immigrants is used in the analysis.

[36] We group the 50 disaggregated occupations used in Burstein et al. (2018) into seven aggregate occupations: laborers, farm workers and low-skilled service workers; operatives and craft workers; clerical workers; sales workers; managers; professional and technical workers; and others. Although Table 8 adopts occupational definitions that build on those in Burstein et al. (2018) and Card (2001), our specifications do not exactly match their definition of shares and shifters. Thus, our estimates should not be viewed as a test of the robustness of the results presented in these studies. Also, when interpreting our estimates, one should bear in mind that, as discussed in Jaeger, Ruist and Stuhler (2018), these may conflate the short- and the long-run responses to immigration shocks.

[37] See Table F.14 in Online Appendix F.3 for p-values for the null of no effect for the specifications and inference procedures reported in Table 8.

Table 8: Effect of immigration: analysis by CZ-Occupations and CZ-Education groups (excluding large origin countries)

Outcome:	$\Delta \log E_i$	$\Delta \log w_i$			$\Delta \log E_i$	$\Delta \log w_i$	$\Delta \log E_i$	$\Delta \log w_i$		
Workers:	All	All	High-Skill	Low-Skill	All	All	All	All	High-Skill	Low-Skill
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	CZ-50 Occ. (1980 weights)				*CZ-Educ. (1980 weights)*		*CZ-7 Occ. (1980 weights)*			
Panel A: 2SLS										
$\hat{\beta}$	−1.19	0.05	0.26	−0.14	−1.49	0.18	−1.39	0.08	0.24	−0.14
Robust	[−1.55, −0.83]	[−0.09, 0.20]	[0.11, 0.41]	[−0.32, 0.03]	[−2.09, −0.90]	[−0.20, 0.56]	[−1.89, −0.90]	[−0.17, 0.33]	[0.00, 0.47]	[−0.43, 0.15]
Cluster	[−1.89, −0.49]	[−0.35, 0.46]	[−0.14, 0.67]	[−0.69, 0.40]	[−2.14, −0.85]	[−0.04, 0.39]	[−1.87, −0.92]	[−0.11, 0.27]	[0.10, 0.38]	[−0.27, −0.01]
AKM	[−1.55, −0.83]	[−0.36, 0.47]	[−0.17, 0.69]	[−0.64, 0.35]	[−2.31, −0.68]	[−0.53, 0.88]	[−2.02, −0.76]	[−0.51, 0.67]	[−0.31, 0.79]	[−0.81, 0.53]
AKM0	[−1.66, −0.72]	[−0.53, 0.54]	[−0.32, 0.81]	[−0.92, 0.39]	[−3.00, −0.14]	[−0.90, 1.60]	[−2.35, −0.54]	[−0.75, 0.94]	[−0.51, 1.07]	[−1.12, 0.80]
Panel B: Reduced-Form										
$\hat{\beta}$	−0.89	0.04	0.2	−0.11	−1.29	0.15	−1.05	0.06	0.18	−0.11
Robust	[−1.17, −0.61]	[−0.07, 0.15]	[0.06, 0.33]	[−0.23, 0.02]	[−1.86, −0.73]	[−0.19, 0.50]	[−1.38, −0.73]	[−0.13, 0.25]	[−0.01, 0.37]	[−0.32, 0.11]
Cluster	[−1.37, −0.41]	[−0.27, 0.35]	[−0.16, 0.55]	[−0.47, 0.25]	[−1.69, −0.90]	[−0.02, 0.33]	[−1.29, −0.82]	[−0.08, 0.21]	[0.07, 0.29]	[−0.20, −0.01]
AKM	[−1.35, −0.43]	[−0.28, 0.36]	[−0.18, 0.57]	[−0.44, 0.23]	[−1.99, −0.60]	[−0.48, 0.79]	[−1.54, −0.57]	[−0.39, 0.52]	[−0.26, 0.62]	[−0.59, 0.38]
AKM0	[−1.55, −0.39]	[−0.27, 0.55]	[−0.16, 0.79]	[−0.44, 0.42]	[−2.25, −0.11]	[−0.53, 1.56]	[−1.62, −0.36]	[−0.42, 0.78]	[−0.29, 0.86]	[−0.62, 0.67]
Panel C: First-Stage										
$\hat{\beta}$	0.75				0.87		0.76			
Robust	[0.56, 0.93]				[0.55, 1.18]		[0.56, 0.95]			
Cluster	[0.62, 0.88]				[0.66, 1.08]		[0.60, 0.91]			
AKM	[0.36, 1.13]				[0.44, 1.30]		[0.42, 1.09]			
AKM0	[0.38, 1.37]				[0.34, 1.70]		[0.38, 1.24]			

Notes: $\Delta \log E_i$ denotes log change in native employment; $\Delta \log w_i$ denotes log change in average weekly wages of native workers. The specifications *CZ-50 Occupations*, *CZ-2 Education Groups*, and *CZ-7 Occupations* differ in the definition of the unit of observation. In all three specifications, we use 1980 weights and three time periods, 1980–1990, 1990–2000, and 2000–2010. Thus, $N = 108,300$ (722 CZs $\times$ 50 occupations $\times$ 3 time periods) for the *CZ-50 Occupations* specification; $N = 4,332$ (722 CZs $\times$ 2 education groups $\times$ 3 time periods) for the *CZ-2 Education Groups* specification; and $N = 15,162$ (722 CZs $\times$ 7 occupations $\times$ 3 time periods) for the *CZ-7 Occupations* specification. Models are weighted by start-of-period occupation-region (or education group-region) share of national population. All regressions include occupation (or education group) and period dummies. 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 Remark 5; and *AKM0* is the confidence interval in Remark 6. We exclude from the analysis those countries of origin whose immigrant share in 1980 is larger than 3%; i.e. $\sum_i ImmShare_{igt_0} / \sum_i \sum_{g'} ImmShare_{ig't_0} > 0.03$. See Table F.8 for a list of the origin countries included in the analysis.

the other two cell definitions, the IV *AKM* and *AKM*0 CIs are on average, 200% and 356% wider than those based on state-clustered standard errors, and the reduced-form *AKM* and *AKM*0 CIs are on average 228% and 358% wider than those based on state-clustered standard errors. Similarly, the CIs for the first-stage coefficient, reported in Panel C, are more than twice as wide for *AKM* and *AKM*0 than for *Robust* and *Cluster*. These results are consistent with the placebo simulation results shown in Table F.10 in Online Appendix F.3, which show that state-clustered standard errors lead to rejection rates that are very close to the nominal level when a cell is defined as the intersection of CZs and 50 occupations, but lead to overrejection for the other two cell definitions.

To understand why this overrejection occurs, recall from the discussion in Section 4.1.2 that robust and state-clustered standard errors may be biased downward even if there is no shock in the structural residual that varies exactly at the same country-of-origin level as the shifters of interest; a downward bias will arise so long as there is a shift-share component in the residual with shares that have a correlation structure similar to that of the shares used to construct the shift-share instrument. We present simulations that illustrate this point in Section 6.3.4 and Online Appendix E.5.

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 implement 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.

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**, "Robots and Jobs: Evidence from US Labor Markets," *MIT mimeo*, 2017.

__ **and** __ , "Demographics and Automation," *MIT mimeo*, 2018.

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

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

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc J. Melitz, "The Impact of Exports on Innovation: Theory and Evidence," Technical Report 2018.

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 "Immigration, Trade, and the Labor Market," University of Chicago Press, 1991, pp. 201–234.

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," *University of Zurich mimeo*, 2017.

_ , _ , _ , 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).

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.," *University of Michigan mimeo*, 2018.

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 "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, National Bureau of Economic Research 2019.

Blanchard, Olivier and Lawrence F Katz, "Regional Evolutions, Brooking Papers on Economic Activity," *Economic Studies Program, The Brookings Institution*, 1992, *23* (1), 76.

Borjas, George J., "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 2003, *118* (4).

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental Shift-share Research Designs," 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.," *University of California Los Angeles mimeo*, 2018.

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.

—, "Immigration and Inequality," *American Economic Review: Papers & Proceedings*, 2009, *99* (2), 1–21.

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?," *Yale University mimeo*, 2017.

Chodorow-Reich, Gabriel and Johannes Wieland, "Secular Labor Reallocation and Business Cycles," *Harvard University mimeo*, 2018.

Colantone, Italo and Piero Stanig, "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe," *American Journal of Political Science*, 2018, *forthcoming.*

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.

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

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, National Bureau of Economic Research 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*, 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," *Harvard University mimeo*, 2018.

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

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. 5A, UK: North Holland, 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," *mimeo*, 2018.

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.

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

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.

Appendices

A Proofs and additional details for Section 4

Since Propositions 1 and 2 are special cases or Propositions 3 and 4, we only prove Propositions 3, 4 and 5. We give the proofs under a slightly more general setup that allows for a linearization error in the potential outcome equation. We introduce this more general setup in Appendix A.1, where we also collect the assumptions that we impose on the DGP. We collect some auxiliary Lemmata used in the proofs in Appendix A.2, and we prove these results in Appendices A.3, A.3 and A.5. Finally, Appendix A.6 discusses inference when the effects β_{is} are heterogeneous.

Throughout the appendix, we assume that $\sum_{s=1}^{S} w_{is} \leq 1$ for all i. Thus, $\sum_{s=1}^{S} n_s \leq N$, where $n_s = \sum_{i=1}^{N} w_{is}$ denotes the size of sector s. We use the notation $A_S \preceq B_S$ to denote $A_S = O(B_S)$, i.e. there exists a constant C independent of S such that $A_S \leq CB_S$. Let $\mathcal{F}_0$ denote the σ-field generated by $(\mathcal{Z}, U, Y(0), B, W)$ (for the case with no covariates, $\mathcal{F}_0$ denotes the σ-field generated by $(Y(0), B, W)$). Define $\bar{w}_{st} = \sum_{i=1}^{N} w_{is} w_{it}$, $\tilde{\mathcal{X}}_s = \mathcal{X}_s - \mathcal{Z}_s' \gamma$, and $\sigma_s^2 = \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0)$. Finally, let $r_N = 1/\sum_s n_s^2$, and let E_W denote expectation conditional on W.

A.1 General setup and assumptions

We first list and discuss the regularity conditions needed for the results in Section 4.1. We then generalize the setup from Section 4.2 by allowing for a linearization error in the potential outcome equation (11). Unless stated otherwise, all limits are taken as $S \to \infty$. We leave the dependence of the number of regions $N = N_S$ on S implicit.

For the results in Section 4.1, we assume that the observed data (Y, X, W) is generated by the variables $(Y(0), B, W, \mathcal{X})$, which we model as a triangular array, so that the distribution of the data may change with the sample size.[38] The additional regularity conditions we impose on these variables, in addition to Assumptions 1 and 2 as follows:

Assumption 5. (i) The support of β_{is} is bounded; (ii) $\frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0) w_{is}^2$ converges in probability to a strictly positive non-random limit; (iii) For some $\nu > 0$, $E[|\mathcal{X}_s|^{2+\nu} \mid \mathcal{F}_0]$ exists and is uniformly bounded, and conditional on W, the second moments of $Y_i(0)$ exist, and are bounded uniformly over i; (iv) For some $\nu > 0$, $E[|\mathcal{X}_s|^{4+\nu} \mid \mathcal{F}_0]$ is uniformly bounded, and conditional on W, the fourth moments of $Y_i(0)$ exist, and are bounded uniformly over i.

The bounded support condition on β_{is} in Assumption 5(i) is made to keep the proofs simple and can be relaxed. Assumption 5(ii) is a standard regularity condition ensuring that the shocks $\mathcal{X}$ have sufficient variation so that the denominator of $\hat{\beta}$, scaled by N, does not converge to zero. This requires that there is at least one "non-negligible" sector in most regions in the sense that its share w_{is} is bounded away from zero. This implies that $\sum_{s=1}^{S} n_s/N$ is also bounded away from zero.

[38]In other words, to allow the distribution of the data to change with the sample size S, we implicitly index the data by S. Making this index explicit, for each S, the data is thus given by the array $\{(Y_{iS}(0), \beta_{isS}, w_{isS}, \mathcal{X}_{sS}) \colon i = 1, \ldots, N_S, s = 1, \ldots, S\}$.

Assumption 5(iii) imposes some mild assumptions on the existence of moments of $\mathcal{X}$ and $Y_i(0)$. Assumption 5(iv), which is only needed for asymptotic normality, strengthens this condition.

For the results in Section 4.2, we generalize the setup in the main text by allowing for a linearization error in the expression for potential outcomes,

$$Y_i(x_1, \dots, x_S) = Y_i(0) + \sum_{s=1}^{S} w_{is} x_s \beta_{is} + L_i(x_1, \dots, x_S), \qquad \sum_{s=1}^{S} w_{is} \leq 1, \tag{A.1}$$

and we weaken Assumption 3(i) by replacing it with the assumption that the observed outcome is given by $Y_i = Y_i(\mathcal{X}_1, \dots, \mathcal{X}_S)$, such that eq. (A.1) holds with $L_i(\mathcal{X}_1, \dots, \mathcal{X}_S) = L_i$.

We assume that the observed data $(Y, \mathcal{X}, Z, W)$ is generated by the triangular array of variables $(Y(0), B, W, U, \mathcal{X}, \mathcal{Z}, L)$. Let $\check{\delta} = (Z'Z)^{-1}Z'(Y - X\beta)$ denote the regression coefficient in a regression of $Y - X\beta$ on Z, that is, the regression coefficient on Z_i in a regression in which $\hat{\beta}$ is restricted to equal to the true value β.

Assumption 6. (i) $N^{-1}\sum_{i=1}^{N} E[L_i^2]^{1/2} \to 0$, and conditional on W, the second moments of U_i and $\mathcal{Z}_s$ exist and are bounded uniformly over i and s; (ii) $Z'Z/N$ converges in probability to a positive definite non-random limit; (iii) $(\sum_s n_s^2)^{-1/2}\sum_{i=1}^{N} E[L_i^2]^{1/2} \to 0$, $\max_i E[L_i^4 \mid W] \to 0$, and conditional on W, the fourth moments of $\mathcal{Z}_s$, and U_i exist and are bounded uniformly over s and i; (iv) $\check{\delta} - \delta = O_p(q_s)$ for some sequence $q_S \to 0$; (v) $q_S^2 N / \sum_s n_s^2 \cdot \sum_i E[(U_i'\gamma)^2] \to 0$ and $\gamma' U' \epsilon = o_p((\sum_s n_s^2)^{1/2})$.

Assumption 6(i) imposes some mild moment restrictions on the controls Z_i. It also requires that on average, the variance of the linearization error L_i vanishes with sample size. This ensures that the linearization error does not impact the consistency of $\hat{\beta}$. Assumption 6(ii) ensures that the controls are not collinear.

Assumptions 6(iii) to 6(v) are only needed for asymptotic normality. Assumption 6(iii) strengthens the moment conditions in Assumption 6(i). It also imposes a stricter condition on the linearization error: it requires that, on average over N, the standard deviation of L_i is of smaller order than $(\sum_s n_s^2)^{1/2}/N$, the rate of convergence of $\hat{\beta}$. A sufficient condition is that $L_i = o_p(S^{-1/2})$. This ensures that the linearization error is of smaller order than the variance of the estimator, so that the distribution of $\hat{\beta}$ does not suffer from asymptotic bias. This formalizes the assumption that the linearization error is "small". The condition that $\max_i E[L_i^4 \mid W] \to 0$ is only needed for showing consistency of the standard error estimator; it is not needed for asymptotic normality. Assumption 6(iv) requires that $\check{\delta}$ is consistent, which ensures that the error in estimation of δ does not affect the asymptotic distribution of $\hat{\beta}$. Finally, Assumption 6(v) imposes conditions on $U_i'\gamma$, the measurement error for controls that matter, which ensure that measurement error in the controls that matter does not impact the asymptotic distribution of $\hat{\beta}$. They are stated as high-level conditions to cover a range of different cases, and depend on the rate of convergence q_S of $\check{\delta}$. In typical cases, the rate will be $q_S = (\sum_s n_s^2)^{1/2}/N$, the same as that of $\hat{\beta}$, and the condition $q_S^2 N / \sum_s n_s^2 \cdot \sum_i E[(U_i'\gamma)^2] \to 0$ is implied by Assumption 3(iii). Let U_{1i} denote the subset of elements of U_i for which $\gamma_k \neq 0$, and let U_{2i} denote the remaining elements. If U_{i1} is mean zero and independent across i conditional on the remaining variables $((Y(0), W, B, \mathcal{Z}, \mathcal{X}, U_2))$, so that these elements are pure measurement error, then the second condition is implied by Assumption 3(iv).

A.2 Auxiliary results

Lemma 1. *$\{\mathcal{A}_{S1},\dots,\mathcal{A}_{SS}\}_{S=1}^{\infty}$ be a triangular array of random variables. Fix $\eta \geq 1$, and let $A_{Si} = \sum_{s=1}^{S} w_{is}\mathcal{A}_{Ss}$, $i = 1\dots,N_S$. Suppose $E[|\mathcal{A}_{Ss}|^{\eta} \mid W]$ exists and is uniformly bounded. Then $E[|A_{Si}|^{\eta} \mid W]$ exists and is bounded uniformly over S and i.*

Proof. The result follows by triangle inequality for $\eta = 1$. Suppose therefore that $\eta > 1$. By Hölder's inequality,

$$E[|A_{Si}|^{\eta} \mid W] = E\left[\left|\sum_{s=1}^{S} w_{is}^{\frac{\eta-1}{\eta}} w_{is}^{\frac{1}{\eta}}\mathcal{A}_{Ss}\right|^{\eta} \mid W\right] \leq \left(\sum_{s=1}^{S} w_{is}\right)^{\eta-1}\sum_{s=1}^{S} w_{is}E[|\mathcal{A}_{Ss}|^{\eta} \mid W]$$
$$\leq \max_{s} E[|\mathcal{A}_{Ss}|^{\eta} \mid W]\cdot(\textstyle\sum_{s=1}^{S} w_{is})^{\eta} \leq \max_s E[|\mathcal{A}_{Ss}|^{\eta} \mid W],$$

which yields the result. □

Lemma 2. *$\{A_{S1},\dots,A_{SN_S}\}_{S=1}^{\infty}$ be a triangular array of random variables. Suppose $E[A_{Si}^2 \mid W]$ exists and is uniformly bounded. Then $\sum_{s=1}^{S} E\left[(\sum_{i=1}^{N} w_{is}A_{Si})^2 \mid W\right] \preceq \sum_s n_s^2$.*

Proof. By Cauchy-Schwarz inequality,

$$\sum_{s=1}^{S} E\left[\left(\sum_{i=1}^{N} w_{is}A_{Si}\right)^2 \;\middle|\; W\right] \leq \sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{is}w_{js}E[A_{Si}^2 \mid W]^{1/2}E[A_{Sj}^2 \mid W]^{1/2}$$
$$\preceq \sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{is}w_{js} = \sum_{s=1}^{S} n_s^2.$$

□

Lemma 3. *Let $\{A_{S1},\dots,A_{SN_S},B_{S1},\dots,B_{SN_S},\mathcal{A}_{S1},\dots,\mathcal{A}_{SS}\}_{S=1}^{\infty}$ be a triangular array of random variables. Suppose $E[A_{Si}^4 \mid W]$, $E[B_{Si}^4 \mid W]$, and $E[\mathcal{A}_{Ss}^2 \mid W]$ exist and are uniformly bounded. Then $(\sum_s n_s^2)^{-1}\cdot \sum_{i,j,s} w_{is}w_{js}A_{Si}B_{Sj}\mathcal{A}_{Ss} = O_p(1)$.*

Proof. Let $R_S = (\sum_s n_s^2)^{-1}\sum_{i,j,s} w_{is}w_{js}A_{Si}B_{Sj}\mathcal{A}_{Ss}$. By the triangle and Cauchy-Schwarz inequalities,

$$E[|R_S| \mid W] \leq \frac{1}{\sum_s n_s^2}\sum_{i,j,s} w_{is}w_{js}E[|A_{Si}B_{Sj}\mathcal{A}_{Ss}| \mid W]$$
$$\leq \frac{1}{\sum_s n_s^2}\sum_{i,j,s} w_{is}w_{js}E[|B_{Sj}|^4 \mid W]^{1/4}E[|A_{Si}|^4 \mid W]^{1/4}E[\mathcal{A}_{Ss}^2 \mid W]^{1/2} \preceq \frac{1}{\sum_s n_s^2}\sum_{i,j,s} w_{is}w_{js} = 1.$$

The result then follows by Markov inequality. □

A.3 Proof of Proposition 3

First we show that

$$Z'W\tilde{\mathcal{X}} = O_p(1/\sqrt{r_N}). \tag{A.2}$$

Conditional on W, the left-hand side has mean zero by Assumption 3(ii), and by Assumption 2(i), the variance of the kth row given by

$$\operatorname{var}\left(\sum_{i,s} w_{is}\tilde{\mathcal{X}}_s Z_{ik} \mid W\right) = \sum_s E_W \sigma_s^2 \left(\sum_i w_{is} Z_{ik}\right)^2 \preceq \sum_s E_W \left(\sum_i w_{is} Z_{ik}\right)^2.$$

By Lemma 1, Assumption 6(i), and the C_r-inequality, $E_W[Z_{ik}^2] = E_W[(\sum_s w_{is}\mathcal{Z}_{sk} + U_{ik})^2]$ is uniformly bounded. Therefore, by Lemma 2, the right-hand side is bounded by $\sum_s n_s^2$, so the result follows by Markov inequality and dominated convergence theorem.

Since $X = W\tilde{\mathcal{X}} + Z\gamma - U\gamma$, it follows from eq. (A.2) and Assumption 6(ii) that

$$\hat{\gamma} - \gamma = (Z'Z/N)^{-1} Z'W\tilde{\mathcal{X}}/N - (Z'Z/N)^{-1} Z'U\gamma/N = o_p(1), \tag{A.3}$$

where $\hat{\gamma} = (Z'Z)^{-1}Z'X$, and the last equality follows since $\sum_s n_s^2/N^2 \leq \max_s n_s/N \to 0$ by Assumption 2(ii), and since $Z'U\gamma/N = o_p(1)$ by the Cauchy-Schwarz inequality and Assumption 3(iii).

Next, we will show that

$$\ddot{X}'\ddot{X}/N = \frac{1}{N}\sum_{i,s} w_{is}^2 \sigma_s^2 + o_p(1). \tag{A.4}$$

To this end, we have

$$\begin{aligned}
\ddot{X}'\ddot{X}/N &= (W\tilde{\mathcal{X}} - U\gamma - Z(\hat{\gamma}-\gamma))'(W\tilde{\mathcal{X}} - U\gamma - Z(\hat{\gamma}-\gamma))/N \\
&= (W\tilde{\mathcal{X}})'(W\tilde{\mathcal{X}})/N + o_p(1) \\
&= \frac{1}{N}\sum_s \overline{w}_{ss}\sigma_s^2 + \frac{2}{N}\sum_{s<t} \overline{w}_{st}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t + \frac{1}{N}\sum_s \overline{w}_{ss}(\tilde{\mathcal{X}}_s^2 - \sigma_s^2) + o_p(1).
\end{aligned}$$

where the first line follows from the decomposition

$$\ddot{X} = X - Z(Z'Z)^{-1}Z'X = X - Z\hat{\gamma} = W\tilde{\mathcal{X}} - U\gamma - Z(\hat{\gamma}-\gamma), \tag{A.5}$$

the second line follows by the Cauchy-Schwarz inequality, Assumption 3(iii), and eq. (A.3), and the third line follows by expanding $(W\tilde{\mathcal{X}})'(W\tilde{\mathcal{X}})/N$. Therefore, to show eq. (A.4), it suffices to show that the second and third term in the above expression are $o_p(1)$. Since the second term has mean zero conditional on W, it suffices to show that its variance converges to zero. To that end,

$$\begin{aligned}
\operatorname{var}\left(\frac{2}{N}\sum_{s<t}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t\overline{w}_{st} \mid W\right) &= \frac{4}{N^2}\sum_{s<t} E_W[\sigma_s^2\sigma_t^2]\overline{w}_{st}^2 \preceq \frac{1}{N^2}\sum_{s,t}\overline{w}_{st}^2 = \frac{1}{N^2}\sum_{i,j,s,t} w_{is}w_{it}w_{js}w_{jt} \\
&\leq \frac{1}{N^2}\sum_{i,j,s,t} w_{is}w_{it}w_{js} \leq \frac{1}{N^2}\sum_{i,j,s} w_{is}w_{js} = \frac{1}{N^2}\sum_s n_s^2 \leq \frac{\max_t n_t \sum_s n_s}{N^2} \to 0.
\end{aligned}$$

where the convergence to 0 follows by Assumption 2(ii). By the inequality of von Bahr and Esseen, Assumption 5(iii), and the inequality $\overline{w}_{ss} \leq n_s$,

$$E[N^{-1}\left|\sum_s(\tilde{\mathcal{X}}_s^2-\sigma_s^2)\overline{w}_{ss}\right|^{1+\nu/2}\mid\mathcal{F}_0]\leq\frac{2}{N^{1+\nu/2}}\sum_s\overline{w}_{ss}^{1+\nu/2}E[\left|\tilde{\mathcal{X}}_s^2-\sigma_s^2\right|^{1+\nu/2}\mid\mathcal{F}_0]$$
$$\preceq\frac{1}{N^{1+\nu/2}}\sum_s\overline{w}_{ss}^{1+\nu/2}\leq(\max_s n_s/N)^{\nu/2},\quad\text{(A.6)}$$

which converges to zero by Assumption 2(ii). Equation (A.4) then follows by Markov inequality.

Next, we show that

$$\ddot{X}'Y/N=\frac{1}{N}\sum_{i,s}\sigma_s^2w_{is}^2\beta_{is}+o_P(1)\tag{A.7}$$

Using eq. (A.5), we can write the left-hand side as

$$\begin{aligned}\ddot{X}'Y/N&=\tilde{\mathcal{X}}'W'Y/N-\gamma'U'Y/N-Y'Z/N\cdot(\hat{\gamma}-\gamma)\\&=\tilde{\mathcal{X}}'W'Y/N+o_p(1)\\&=\frac{1}{N}\sum_{s,i}w_{is}\tilde{\mathcal{X}}_sL_i+\frac{1}{N}\sum_{s,i}w_{is}^2(\tilde{\mathcal{X}}_s\mathcal{X}_s-\sigma_s^2)\beta_{is}+\frac{1}{N}\sum_{s,i}w_{is}\tilde{\mathcal{X}}_sY_i(0)\\&\quad+\frac{1}{N}\sum_{s<t}\sum_i w_{is}w_{it}\tilde{\mathcal{X}}_s\mathcal{X}_t\beta_{it}+\frac{1}{N}\sum_{s<t}\sum_i w_{is}w_{it}\tilde{\mathcal{X}}_t\mathcal{X}_s\beta_{is}+\frac{1}{N}\sum_{s,i}w_{is}^2\sigma_s^2\beta_{is}+o_p(1)\end{aligned}$$

where the second line follows since by the C_r-inequality, Lemma 1, Assumptions 5(i), 6(i) and 5(iii), $N^{-1}\sum_i E[Y_i^2]$ is bounded, so that $Y'Z/N=O_p(1)$ and $\gamma'U'Y/N=o_p(1)$ by Cauchy-Schwarz inequality and Assumption 3(iii), and the third line follows by expanding $\tilde{\mathcal{X}}'W'Y$. We therefore need to show that the first five terms in the expression above are $o_p(1)$. By the Cauchy-Schwarz inequality, the expectation of the absolute value of the first term is bounded by

$$N^{-1}\sum_i E[L_i^2]^{1/2}(E\sum_s w_{is}^2\sigma_s^2)^{1/2}\preceq N^{-1}\sum_i E[L_i^2]^{1/2},$$

which converges to zero by Assumption 6(i). Thus, the first term is $o_p(1)$ by Markov inequality and the dominated convergence theorem. The second term is $o_p(1)$ by an argument analogous to eq. (A.6). The third to fifth terms are mean zero conditional on $\mathcal{F}_0$, so it suffices to show that their variances conditional on W converge to zero. The variance of the third summand is bounded by

$$\operatorname{var}\left(\frac{1}{N}\sum_s\tilde{\mathcal{X}}_s\sum_i w_{is}Y_i(0)\mid W\right)=\frac{1}{N^2}\sum_s E_W\sigma_s^2\left(\sum_i w_{is}Y_i(0)\right)^2\preceq\frac{1}{N^2}\sum_s E_W\left(\sum_i w_{is}Y_i(0)\right)^2,$$

which converges to zero by Lemma 2. The variance of the fourth term is bounded by

$$\begin{aligned}\operatorname{var}\left(\frac{1}{N}\sum_{s<t}\sum_i w_{is}w_{it}\tilde{\mathcal{X}}_s\mathcal{X}_t\beta_{it}\mid W\right)&=\frac{1}{N^2}\sum_{s<t,t'}\sum_{i,i'}w_{is}w_{it}\sigma_s^2E_W[\mathcal{X}_t\mathcal{X}_{t'}]\beta_{it}w_{i's}w_{i't'}\beta_{i't'}\\&\preceq\frac{1}{N^2}\sum_{s,t,t',i,i'}w_{is}w_{it}w_{i's}w_{i't'}\leq\frac{1}{N^2}\sum_s n_s^2\leq\max_s n_s/N\to 0.\end{aligned}$$

Variance of the fifth term converges to zero by analogous arguments.

Combining eq. (A.4) with eq. (A.7) and Assumption 5(ii) then yields the result.

A.4 Proof of Proposition 4

Using eq. (A.5), we have

$$
\begin{aligned}
r_N^{1/2}(\ddot{X}'\ddot{X})(\hat{\beta}-\beta) &= r_N^{1/2}X'(I-Z(Z'Z)^{-1}Z')(Z\delta+\epsilon) = r_N^{1/2}X'(I-Z(Z'Z)^{-1}Z')\epsilon \\
&= r_N^{1/2}\tilde{\mathcal{X}}'W'\epsilon - r_N^{1/2}\gamma'U'\epsilon - r_N^{1/2}(\hat{\gamma}-\gamma)'Z'\epsilon.
\end{aligned}
$$

The third term can be written as

$$
\begin{aligned}
r_N^{1/2}(\hat{\gamma}-\gamma)'Z'\epsilon &= r_N^{1/2}\epsilon'Z(Z'Z)^{-1}(Z'W\tilde{\mathcal{X}}-Z'U\gamma) = r_N^{1/2}(\check{\delta}-\delta)'(Z'W\tilde{\mathcal{X}}-Z'U\gamma) \\
&= (\check{\delta}-\delta)'(O_p(1) - r_N^{1/2}Z'U\gamma) \\
&= o_p(1) - O_p(1)\cdot q_S r_N^{1/2}Z'U\gamma = o_p(1),
\end{aligned}
$$

where the first line follows from the decomposition in eq. (A.3), the second line follows from eq. (A.2), the third line follows by Assumption 6(iv), and the last equality follows since by Cauchy-Schwarz inequality and Assumption 6(v), $q_S r_N^{1/2}E[|Z_k'U\gamma|] \preceq \sqrt{q_S^2 r_N N\sum_i E(U_i'\gamma)^2} \to 0$. Since $r_N^{1/2}\gamma'U'\epsilon = o_p(1)$ by Assumption 6(v), and since by eq. (A.4) and Assumption 5(ii), $(\ddot{X}'\ddot{X}/N)^{-1} = (1+o_p(1))\cdot(N^{-1}\sum_{i,s}\pi_{is})^{-1}$, it follows that

$$
\frac{N}{(\sum_s n_s^2)^{1/2}}(\hat{\beta}-\beta) = (1+o_p(1))\frac{1}{N^{-1}\sum_{i,s}\pi_{is}}r_N^{1/2}\sum_{s,i}\tilde{\mathcal{X}}_s w_{is}\epsilon_i + o_p(1).
$$

Therefore, it suffices to show

$$
r_N^{1/2}\sum_{s,i}\tilde{\mathcal{X}}_s w_{is}\epsilon_i = \mathcal{N}(0, \operatorname{plim}\mathcal{V}_N) + o_p(1). \tag{A.8}
$$

Define $V_i = Y_i(0) - Z_i'\delta + \sum_t w_{it}\mathcal{Z}_t'\gamma(\beta_{it}-\beta)$, and

$$
a_s = \sum_i w_{is}V_i, \qquad\qquad b_{st} = \sum_i w_{is}w_{it}(\beta_{it}-\beta). \tag{A.9}
$$

Then we can write $\epsilon_i = V_i + \sum_t w_{it}\tilde{\mathcal{X}}_t(\beta_{it}-\beta) + L_i$. Since

$$
E|r_N^{1/2}\sum_{i,s}\tilde{\mathcal{X}}_s w_{is}L_i| \leq r_N^{1/2}\sum_i(\sum_s E w_{is}^2\sigma_s^2)^{1/2}E[L_i^2]^{1/2} \preceq r_N^{1/2}\sum_i E[L_i^2]^{1/2} \to 0
$$

by Assumption 6(iii), and since $0 = \sum_{i,s}\pi_{is}(\beta_{is}-\beta) = \sum_s \sigma_s^2 b_{ss}$, we can decompose

$$
r_N^{1/2}\sum_{s,i}\tilde{\mathcal{X}}_s w_{is}\epsilon_i = r_N^{1/2}\sum_s\tilde{\mathcal{X}}_s\sum_i w_{is}\left(V_i + \sum_t w_{it}\tilde{\mathcal{X}}_t(\beta_{it}-\beta) + L_i\right) = r_N^{1/2}\sum_s\mathcal{Y}_s + o_P(1),
$$

where

$$
\mathcal{Y}_s = \tilde{\mathcal{X}}_s a_s + (\tilde{\mathcal{X}}_s^2 - \sigma_s^2)b_{ss} + \sum_{t=1}^{s-1}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t(b_{st}+b_{ts}).
$$

Observe that $\mathcal{Y}_s$ is a martingale difference array with respect to the filtration $\mathcal{F}_s = \sigma(\mathcal{X}_1,\dots,\mathcal{X}_s,\mathcal{F}_0)$.

By the dominated convergence theorem and the martingale central limit theorem, it suffices to show that $r_N^{1+\nu/4}\sum_{s=1}^S E_W[\mathcal{Y}_s^{2+\nu/2}] \to 0$ for some $\nu > 0$ so that the Lindeberg condition holds, and that the conditional variance converges,

$$r_N \sum_{s=1}^{S} E[\mathcal{Y}_s^2 \mid \mathcal{F}_{s-1}] - \mathcal{V}_N = o_p(1).$$

To verify the Lindeberg condition, by the C_r-inequality, it suffices to show that

$$r_N^2 \sum_s E_W[\tilde{\mathcal{X}}_s^4 a_s^4] \to 0, \qquad r_N^{1+\nu/4} \sum_s E_W[(\tilde{\mathcal{X}}_s^2 - \sigma_s^2)^{2+\nu/2} b_{ss}^{2+\nu/2}] \to 0,$$

$$r_N^2 \sum_s E_W \left(\sum_{t=1}^{s-1} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t b_{st} \right)^4 \to 0, \qquad r_N^2 \sum_s E_W \left(\sum_{t=1}^{s-1} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t b_{ts} \right)^4 \to 0.$$

Note that since $E(\sum_t w_{it}\mathcal{Z}_t'\gamma(\beta_{it}-\beta))^4 \preceq (\sum_t w_{it})^4 \preceq 1$, it follows from Assumptions 6(iii) and 5(iv), and the C_r inequality that the fourth moment of V_i exists and is bounded. Therefore, by arguments as in the proof of Lemma 2, $\sum_s E_W[a_s^4] \preceq \sum_s n_s^4$, so that

$$r_N^2 \sum_s E_W[\tilde{\mathcal{X}}_s^4 a_s^4] = r_N^2 \sum_s E_W[E[\tilde{\mathcal{X}}_s^4 \mid \mathcal{F}_0] a_s^4] \preceq r_N^2 \sum_s E_W[a_s^4] \preceq r_N^2 \sum_s n_s^4 \leq \max_s n_s^2 r_N \to 0 \tag{A.10}$$

by Assumption 2(iii). Second, since β_{is} is bounded by Assumption 5(i), we have $b_{ss} \preceq \sum_i w_{is}^2 \leq n_s$, so that

$$r_N^{1+\nu/4} \sum_s E_W[(\tilde{\mathcal{X}}_s^2 - \sigma_s^2)^{2+\nu/2} b_{ss}^{2+\nu/2}] \preceq r_N^{1+\nu/4} \sum_s n_s^{2+\nu/2} \leq (r_N \max_s n_s^2)^{\nu/4} \to 0.$$

Third, by similar arguments

$$\begin{aligned} r_N^2 \sum_s E_W \left(\sum_{t=1}^{s-1} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t b_{st} \right)^4 &= r_N^2 \sum_s E_W E[\tilde{\mathcal{X}}_s^4 \mid \mathcal{F}_0] E\left[\left(\sum_{t=1}^{s-1} \tilde{\mathcal{X}}_t b_{st} \right)^4 \mid \mathcal{F}_0 \right] \\ &\preceq r_N^2 \sum_s \left(\sum_{t=1}^{s-1} \sum_i w_{is} w_{it} \right)^4 \leq r_N^2 \sum_s n_s^4 \to 0. \end{aligned}$$

The claim that $r_N^2 \sum_s E_W \left(\sum_{t=1}^{s-1} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t b_{ts} \right)^4 \to 0$ follows by similar arguments.

It remains to verify that the conditional variance converges. Since $\mathcal{V}_N$ can be written as

$$\begin{aligned} \mathcal{V}_N = \frac{1}{\sum_{s=1}^S n_s^2} \operatorname{var} \left(\sum_i (X_i - Z_i'\gamma)\epsilon_i \mid \mathcal{F}_0 \right) &= r_N \sum_s E[\mathcal{Y}_s^2 \mid \mathcal{F}_0] + o_P(1) \\ &= r_N \sum_s \left[E\left[(\tilde{\mathcal{X}}_s a_s + (\tilde{\mathcal{X}}_s^2 - \sigma_s^2) b_{ss})^2 \mid \mathcal{F}_0 \right] + \sum_{t=1}^{s-1} \sigma_s^2 \sigma_t^2 (b_{st} + b_{ts})^2 \right] + o_p(1), \end{aligned}$$

we can decompose

$$r_N \sum_s E[\mathcal{Y}_s^2 \mid \mathcal{F}_{s-1}] - \mathcal{V}_N = 2D_1 + D_2 + 2D_3 + o_p(1),$$

where

$$D_1 = r_N \sum_s (\sigma_s^2 a_s + E[\tilde{\mathcal{X}}_s^3 \mid \mathcal{F}_0] b_{ss}) \sum_{t=1}^{s-1} \tilde{\mathcal{X}}_t (b_{st} + b_{ts}),$$
$$D_2 = r_N \sum_s \sigma_s^2 \sum_{t=1}^{s-1} (\tilde{\mathcal{X}}_t^2 - \sigma_t^2)(b_{st} + b_{ts})^2,$$
$$D_3 = r_N \sum_s \sigma_s^2 \sum_{t=1}^{s-1} \sum_{u=1}^{t-1} \tilde{\mathcal{X}}_t \tilde{\mathcal{X}}_u (b_{st} + b_{ts})(b_{su} + b_{us}).$$

It therefore suffices to show that $D_j = o_p(1)$ for $j = 1, 2, 3$. Since $E[D_j \mid \mathcal{F}_0] = 0$, it suffices to show that $\operatorname{var}(D_j \mid W) = E_W[\operatorname{var}(D_j \mid \mathcal{F}_0)]$ converges to zero. Since $b_{st} + b_{ts} \preceq \overline{w}_{st}$, and since $E_W[|a_s a_t|] \preceq n_s n_t$, and $|b_{ss}| \preceq \overline{w}_{ss} \leq n_s$, it follows that

$$\begin{aligned}\operatorname{var}(D_1 \mid W) &= r_N^2 \sum_t E_W \left[\sigma_t^2 \left(\sum_{s=t+1}^{S} (b_{st} + b_{ts})(\sigma_s^2 a_s + E[\tilde{\mathcal{X}}_s^3 \mid \mathcal{F}_0] b_{ss}) \right)^2 \right] \\ &\preceq r_N^2 \sum_t \left(\sum_{s=t+1}^{S} \overline{w}_{st} n_s \right)^2 \leq r_N^2 \max_s n_s^2 \sum_t \left(\sum_s \overline{w}_{st} \right)^2 = r_N \max_s n_s^2 \to 0,\end{aligned}$$

where the convergence to zero follows by Assumption 2(iii). By similar arguments, since $\overline{w}_{st} \leq n_s$

$$\begin{aligned}\operatorname{var}(D_2 \mid W) &= r_N^2 \sum_t E_W (\tilde{\mathcal{X}}_t^2 - \sigma_t^2)^2 \left(\sum_{s=t+1}^{S} \sigma_s^2 (b_{st} + b_{ts})^2 \right)^2 \preceq r_N^2 \sum_t \left(\sum_{s=t+1}^{S} \overline{w}_{st}^2 \right)^2 \\ &\leq r_N^2 \sum_t \left(\sum_{s=1}^{S} n_s \overline{w}_{st} \right)^2 \leq r_N \max_s n_s^2 \to 0.\end{aligned}$$

Finally,

$$\begin{aligned}\operatorname{var}(D_3 \mid W) &= r_N^2 \sum_t \sum_{u=t+1}^{S} E_W \sigma_t^2 \sigma_u^2 \left(\sum_{s=u+1}^{S} \sigma_s^2 (b_{st} + b_{ts})(b_{su} + b_{us}) \right)^2 \\ &\preceq r_N^2 \sum_t \sum_{u=t+1}^{S} \left(\sum_{s=u+1}^{S} \overline{w}_{st} \overline{w}_{su} \right)^2 \leq r_N^2 \sum_{s,t,u,v} \overline{w}_{st} \overline{w}_{su} \overline{w}_{vt} \overline{w}_{vu} \leq r_N \max_s n_s^2 \to 0,\end{aligned}$$

where the last line follows from the fact that since $\sum_s \overline{w}_{st} = n_t$ and $\overline{w}_{st} \leq n_s$,

$$\begin{aligned}\sum_{s,t,u,v} \overline{w}_{st} \overline{w}_{su} \overline{w}_{vt} \overline{w}_{vu} &\leq \max_s n_s \sum_{s,t,u,v} \overline{w}_{su} \overline{w}_{vt} \overline{w}_{vu} = \max_s n_s \sum_{u,v} n_u n_v \overline{w}_{vu} \\ &\leq \max_s n_s^2 \sum_{u,v} n_v \overline{w}_{vu} = \max_s n_s^2 / r_N. \quad \text{(A.11)}\end{aligned}$$

Consequently, $D_j = o_p(1)$ for $j = 1, 2, 3$, the conditional variance converges, and the theorem follows.

A.5 Proof of Proposition 5

We'll prove a more general result that doesn't assume constant treatment effects. In particular, we will show that under the conditions of the proposition when the condition $\beta_{is} = \beta$ is dropped, the variance estimator $\hat{\mathcal{V}}_N = r_N \sum_s \widehat{\mathcal{X}}_s \hat{R}_s^2$, where $r_N = 1/\sum_{s=1}^S n_s^2$ satisfies

$$\hat{\mathcal{V}}_N = r_N \sum_{s=1}^{S} E[\tilde{\mathcal{X}}_s^2 R_s^2 \mid \mathcal{F}_0] + o_p(1), \tag{A.12}$$

where, using the definitions of a_s and b_{st} in eq. (A.9),

$$R_s = \sum_{i=1}^{N} w_{is}\epsilon_i = a_s + \sum_{i=1}^{N} w_{is}L_i + \sum_{t=1}^{S} \tilde{\mathcal{X}}_t b_{st}.$$

Since under constant treatment effects, $\mathcal{V}_N = r_N \sum_{s=1}^S E[\tilde{\mathcal{X}}_s^2 R_s^2 \mid \mathcal{F}_0]$, the assertion of the proposition follows from eq. (A.12).

Throughout the proof, we write $E_{\mathcal{F}_0}[\cdot]$ and $E_W[\cdot]$ to denote expectations conditional on $\mathcal{F}_0$, and W, respectively. Let $\tilde{\theta} = (\tilde{\beta}, \tilde{\delta}')'$, $\theta = (\beta, \delta)$, $M_i = (X_i, Z_i')'$. We can decompose the variance estimator as

$$\hat{\mathcal{V}}_N = r_N \sum_s (\widehat{\mathcal{X}}_s^2 - \tilde{\mathcal{X}}_s^2)\hat{R}_s^2 + r_N \sum_s \tilde{\mathcal{X}}_s^2(\hat{R}_s^2 - R_s^2) + r_N \sum_s (\tilde{\mathcal{X}}_s^2 R_s^2 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 R_s^2]) + r_N \sum_s E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 R_s^2]. \tag{A.13}$$

We need to show that the first three terms are $o_p(1)$. Since $\tilde{\epsilon}_i = \epsilon_i + M_i'(\theta - \tilde{\theta})$, with $\epsilon_i = V_i + L_i + \sum_t w_{it}\tilde{\mathcal{X}}_t(\beta_{it} - \beta)$, we can decompose

$$\hat{R}_s^2 = \sum_{i,j} w_{is}w_{js}\tilde{\epsilon}_i\tilde{\epsilon}_j = R_s^2 + 2\sum_{i,j} w_{js}w_{is}M_i'(\theta - \tilde{\theta})\epsilon_j + \sum_{i,j} w_{is}w_{js}M_i'(\theta - \tilde{\theta})M_j'(\theta - \tilde{\theta}). \tag{A.14}$$

Therefore, the second term in eq. (A.13) satisfies

$$\begin{aligned} r_N \sum_s \tilde{\mathcal{X}}_s^2(\hat{R}_s^2 - R_s^2) &= 2(\theta - \tilde{\theta})' \left[r_N \sum_{s,i,j} w_{is}w_{js}\tilde{\mathcal{X}}_s^2 M_i\epsilon_j \right] + (\theta - \tilde{\theta})' \left[r_N \sum_{s,i,j} \tilde{\mathcal{X}}_s^2 w_{is}w_{js}M_iM_j' \right] (\theta - \tilde{\theta}) \\ &= (\theta - \tilde{\theta})'O_p(1) + (\theta - \tilde{\theta})'O_p(1)(\theta - \tilde{\theta}) = o_p(1), \end{aligned}$$

where the second line follows by applying Lemma 3 to the terms in square brackets. Next, the third term in (A.13) can be decomposed as

$$\begin{aligned} &r_N \sum_s (\tilde{\mathcal{X}}_s^2 R_s^2 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 R_s^2]) = \\ &\quad + r_N \sum_s b_{ss}^2(\tilde{\mathcal{X}}_s^4 - E_{\mathcal{F}_0}[\mathcal{X}_s^4]) + r_N \sum_{s<t}(b_{st}^2 + b_{ts}^2)(\tilde{\mathcal{X}}_s^2\tilde{\mathcal{X}}_t^2 - \sigma_s^2\sigma_t^2) + 2r_N \sum_s \sum_{t<u} b_{st}b_{su}\tilde{\mathcal{X}}_s^2\tilde{\mathcal{X}}_t\tilde{\mathcal{X}}_u \\ &\quad + r_N \sum_s (\tilde{\mathcal{X}}_s^2 - \sigma_s^2)a_s^2 + r_N \sum_{i,j,s} w_{js}w_{is}(\tilde{\mathcal{X}}_s^2 L_iL_j - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 L_iL_j]) + 2r_N \sum_{i,s} w_{is}a_s(\tilde{\mathcal{X}}_s^2 L_i - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 L_i]) \\ &\quad + 2r_N \sum_{s<t} a_sb_{st}\tilde{\mathcal{X}}_s^2\tilde{\mathcal{X}}_t + 2r_N \sum_{s<t} a_tb_{ts}\tilde{\mathcal{X}}_t^2\tilde{\mathcal{X}}_s + 2r_N \sum_s a_sb_{ss}(\tilde{\mathcal{X}}_s^3 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^3]) \end{aligned}$$

$$+ r_N \sum_{i,s,t} w_{is} b_{st} (\tilde{\mathcal{X}}_s^2 \tilde{\mathcal{X}}_t L_i - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_s^2 \tilde{\mathcal{X}}_t L_i]). \quad \text{(A.15)}$$

We will show that all terms are of the order $o_p(1)$. By the inequality of von Bahr and Esseen, since b_{ss} is bounded by a constant times $\overline{w}_{ss} \leq n_s$,

$$E_{\mathcal{F}_0} |r_N \sum_s b_{ss}^2 (\tilde{\mathcal{X}}_s^4 - E_{\mathcal{F}_0}[\mathcal{X}_s^4])|^{1+\nu/4} \preceq r_N^{1+\nu/4} \sum_s n_s^{2+\nu/2} E_{\mathcal{F}_0} |(\tilde{\mathcal{X}}_s^4 - E_{\mathcal{F}_0}[\mathcal{X}_s^4])|^{1+\nu/4} \leq (\max_s n_s^2 r_N)^{\nu/4} \to 0$$

by Assumption 2(iii), so that the first term is $o_p(1)$. The second term can be written as

$$r_N \sum_{s<t} (b_{st}^2 + b_{ts}^2)(\tilde{\mathcal{X}}_s^2 - \sigma_s^2)(\tilde{\mathcal{X}}_t^2 - \sigma_t^2) + r_N \sum_{s \neq t} (b_{st}^2 + b_{ts}^2)(\tilde{\mathcal{X}}_s^2 - \sigma_s^2)\sigma_t^2$$

The conditional variance of both summands is bounded by a constant times $r_N^2 \sum_s (\sum_t \overline{w}_{st}^2)^2 \leq r_N^2 \cdot \sum_s n_s^4 \to 0$, so that the second term is also $o_p(1)$. The third term admits the decomposition

$$2r_N \sum_s \sum_{t<u} b_{st} b_{su} \tilde{\mathcal{X}}_s^2 \tilde{\mathcal{X}}_t \tilde{\mathcal{X}}_u = 2r_N \sum_{s,t} \sum_{s \notin \{t,u\}} b_{st} b_{su} \tilde{\mathcal{X}}_s^2 \tilde{\mathcal{X}}_t \tilde{\mathcal{X}}_u + 2r_N \sum_{t \neq u} b_{tt} b_{tu} E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_t^3] \tilde{\mathcal{X}}_u$$
$$2r_N \sum_{u<t} b_{tt} b_{tu} (\tilde{\mathcal{X}}_t^3 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_t^3]) \tilde{\mathcal{X}}_u + 2r_N \sum_{t<u} b_{tt} b_{tu} (\tilde{\mathcal{X}}_t^3 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_t^3]) \tilde{\mathcal{X}}_u.$$

The conditional variance of the first summand is bounded by a constant times $r_N^2 \sum_{t,u,s,v} \overline{w}_{st} \overline{w}_{su} \overline{w}_{vt} \overline{w}_{vu}$, which converges to zero by the inequality in eq. (A.11). The conditional variance of the second summand is bounded by a constant times $r_N^2 \sum_{s,t,u} \overline{w}_{tt} \overline{w}_{tu} \overline{w}_{ss} \overline{w}_{su} \leq r_N^2 \max_s n_s^2 \sum_s n_s^2 \to 0$. Since $(\tilde{\mathcal{X}}_t^3 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_t^3]) \sum_{u=1}^{t-1} b_{tt} b_{tu} \tilde{\mathcal{X}}_u$ and $\tilde{\mathcal{X}}_u \sum_{t=1}^{u-1} b_{tt} b_{tu} (\tilde{\mathcal{X}}_t^3 - E_{\mathcal{F}_0}[\tilde{\mathcal{X}}_t^3])$ are both martingale differences, by the inequality of von Bahr and Esseen, the 4/3-th absolute moment of the last two terms is bounded by a constant times $r_N^{4/3} \sum_{s,t} \overline{w}_{tt}^{4/3} \overline{w}_{ts}^{4/3} \leq (\max_s n_s^2 r_N)^{1/3} r_N \sum_t n_t^2 \to 0$. Thus, all summands in the above display are of the order $o_p(1)$, and the third term in eq. (A.15) is therefore also $o_p(1)$. The fourth term is $o_p(1)$ by arguments in eq. (A.10). By the triangle and Cauchy-Schwarz inequalities, the conditional expectation of the absolute value of the fifth term is bounded by

$$2r_N \sum_{i,j,s} w_{js} w_{is} E_W[\tilde{\mathcal{X}}_s^4]^{1/2} E_W[L_i^4]^{1/4} E_W[L_j^4]^{1/4} \preceq \max_i E_W[L_j^4]^{1/2} \to 0.$$

Similarly, conditional expectation of the absolute value of the sixth term is bounded by

$$4r_N \sum_{i,j,s} w_{is} w_{js} E_W[V_j^4]^{1/4} E[\tilde{\mathcal{X}}_s^4]^{1/2} E_W[L_i^4]^{1/4} \preceq \max_i E_W[L_j^4]^{1/4} \to 0.$$

Thus, by the Markov inequality, the fifth and sixth terms are both of the order $o_p(1)$. The conditional variance of the seventh and eighth terms is bounded by a constant times $r_N^2 \sum_{s,t,u} n_s n_u \overline{w}_{st} \overline{w}_{ut} \leq r_N \max_s n_s^2 \to 0$, so that they are both $o_p(1)$ by Markov inequality. By the inequality of von Bahr and Esseen, the 4/3-th absolute moment of the last ninth term is bounded by a constant times $r_N^{4/3} \sum_s E_W[|a_s|^{4/3}] n_s^{4/3} \preceq (\max_s n_s^2 r_N)^{1/3} \to 0$, since by Jensen's inequality, $E|a_s|^{4/3} \leq (E a_s^2)^{2/3}$, which is bounded by a constant times $n_s^{4/3}$. Finally, the expectation of the absolute value of the last term in eq. (A.15) is bounded by a constant times

$$r_N \sum_{i,s,t} w_{is}\overline{w}_{st} E_W[\tilde{\mathcal{X}}_s^4]^{1/2} E_W[\tilde{\mathcal{X}}_t^4]^{1/4} E_W[L_i^4]^{1/4} \preceq \max_i E_W[L_i^4]^{1/4} \to 0.$$

It remains to show that the first term in eq. (A.13) is $o_p(1)$. It follows from eqs. (24) and (A.5) that

$$\widehat{\mathcal{X}} = (W'W)^{-1}W'\ddot{X} = \tilde{\mathcal{X}} - (W'W)^{-1}W'U(\hat{\gamma}-\gamma) - \mathcal{Z}(\hat{\gamma}-\gamma) - (W'W)^{-1}W'U\gamma,$$

where $\hat{\gamma} = (Z'Z)^{-1}Z'X$. Let $\mathcal{U} = (W'W)^{-1}W'U$, and denote the sth row by $\mathcal{U}_s'$. Since $\mathcal{U}_{sk}^4 = (\sum_i((W'W)^{-1}W')_{si}U_{ik})^4$, it follows by the Cauchy-Schwarz inequality that

$$E[\mathcal{U}_{sk}^4 \mid W] \leq \max_s E[(\sum_i((W'W)^{-1}W')_{si}U_{ik})^4 \mid W] \preceq \max_s(\sum_i|((W'W)^{-1}W')_{si}|)^4,$$

which is bounded assumption of the proposition. Therefore, the fourth moments of $\mathcal{U}_s$ are bounded uniformly over s. Observe also that $E_W[\epsilon_i^4]$ is bounded uniformly over s by assumptions of the proposition. Therefore, by applying Lemma 3 after using the expansion in eq. (A.14), we get

$$\begin{aligned}
r_N \sum_s (\widehat{\mathcal{X}}_s^2 - \tilde{\mathcal{X}}_s^2)\hat{R}_s^2 &= r_N \sum_s \hat{R}_s^2(\mathcal{U}_s'\gamma)^2 - 2r_N \sum_s \hat{R}_s^2 \tilde{\mathcal{X}}_s \mathcal{U}_s'\gamma \\
&\quad + r_N \sum_s \hat{R}_s^2 \left[2\mathcal{U}_s'\gamma - 2\tilde{\mathcal{X}}_s + (\mathcal{Z}_s + \mathcal{U}_s)'(\hat{\gamma}-\gamma)\right](\mathcal{Z}_s + \mathcal{U}_s)'(\hat{\gamma}-\gamma) \\
&= r_N \sum_s R_s^2(\mathcal{U}_s'\gamma)^2 - 2r_N \sum_s R_s^2 \tilde{\mathcal{X}}_s \mathcal{U}_s'\gamma + O_p(1)(\hat{\gamma}-\gamma) + o_p(1).
\end{aligned}$$

By Cauchy-Schwarz inequality,

$$r_N \sum_s E_W|R_s^2(\mathcal{U}_s'\gamma)^2| \leq r_N \sum_s (E_W[R_s^4])^{1/2}(E_W(\mathcal{U}_s'\gamma)^4)^{1/2} \preceq \max_s(E_W(\mathcal{U}_s'\gamma)^4)^{1/2} r_N \sum_s n_s^2 \to 0,$$

since $\max_s E_W[(\mathcal{U}_s'\gamma)^4] \preceq \max_i E_W(U_i'\gamma)^4 \max_s(\sum_i|((W'W)^{-1}W')_{si}|)^4$, which converges to zero by assumption of the proposition. By similar arguments, $2r_N \sum_s E_W|R_s^2\tilde{\mathcal{X}}_s\mathcal{U}_s'\gamma| \to 0$ also, so that

$$r_N \sum_s (\widehat{\mathcal{X}}_s^2 - \tilde{\mathcal{X}}_s^2)\hat{R}_s^2 = o_p(1) + O_p(1)(\hat{\gamma}-\gamma) = o_p(1),$$

where the second equality follows from eq. (A.3).

A.6 Inference under heterogeneous effects

For valid (but perhaps conservative) inference under heterogeneous effects, we need to ensure that when $\beta_{is} \neq \beta$, eq. (32) holds with inequality, that is,

$$\frac{\sum_{s=1}^S \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}{\sum_{s=1}^S n_s^2} \geq \mathcal{V}_N + o_p(1). \tag{A.16}$$

To discuss conditions under which this is the case, suppose, for simplicity, that $L_i = 0$ so that eq. (11) holds, and $R_s = \sum_s w_{is}\epsilon_i$, where $\epsilon_i = Y_i(0) - Z_i'\delta + \sum_s \mathcal{X}_s w_{is}(\beta_{is} - \beta)$ is the regression residual. Then the "middle sandwich" in the asymptotic variance sandwich formula, $\mathcal{V}_N$, as defined in Proposition 4, can be decomposed into three terms:

$$\begin{aligned}\mathcal{V}_N &= \frac{\operatorname{var}\left(\sum_s \tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0\right)}{\sum_{s=1}^S n_s^2} = \frac{\sum_s E[\tilde{\mathcal{X}}_s^2 R_s^2 \mid \mathcal{F}_0]}{\sum_{s=1}^S n_s^2} - \frac{\sum_s E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0]^2}{\sum_{s=1}^S n_s^2} + \frac{\sum_{s\neq t} \operatorname{cov}(\tilde{\mathcal{X}}_s R_s, \tilde{\mathcal{X}}_t R_t \mid \mathcal{F}_0)}{\sum_{s=1}^S n_s^2} \\ &= \frac{\sum_s E[\tilde{\mathcal{X}}_s^2 R_s^2 \mid \mathcal{F}_0]}{\sum_{s=1}^S n_s^2} - \frac{\sum_s \left(\sum_i \sigma_s^2 w_{is}^2 (\beta_{is} - \beta)\right)^2}{\sum_{s=1}^S n_s^2} + \frac{\sum_{s\neq t} \sigma_s^2 \sigma_t^2 \sum_{i,j} w_{is} w_{it} (\beta_{it} - \beta) w_{jt} w_{js} (\beta_{js} - \beta)}{\sum_{s=1}^S n_s^2}. \quad \text{(A.17)}\end{aligned}$$

As shown in the proof of Proposition 5 (see eq. (A.12)), the standard error estimator consistently estimates the first term in this decomposition. Under homogeneous effects, the second and third term are both equal to zero, and it follows that the standard error estimator is consistent. To ensure valid inference under heterogeneous effects, one needs to ensure that the second and third term converge in probability to zero, or else to a weakly negative limit. This is the case under several sufficient conditions, and we give two such conditions below.

The second term reflects the variability of the treatment effect and it is always negative. It therefore makes the variance estimate that we propose conservative if the third term equals zero. An analogous term, also reflecting the variability of the treatment effect, is present in randomized, and cluster-randomized trials, which is why the robust and cluster-robust standard error estimators yield conservative inference in these settings (see, for example Imbens and Rubin, 2015, Chapter 6). The third term reflects correlation between the treatment effects. It arises due to aggregating the sectoral shocks $\mathcal{X}_s$ to a regional level to form the shifter X_i, and it has no analog in cluster-randomized trials. Indeed, in the example with "concentrated sectors", which is analogous to cluster-randomized trials if there are no covariates, the term equals zero, since in that case $w_{is}w_{it} = 0$ for $s \neq t$. Our standard errors are thus valid, although conservative, in this case.

More generally, a sufficient condition for validity of our standard error estimator under treatment effect heterogeneity is that $T_N = \sum_{s\neq t}(\sum_i w_{is}w_{it})^2 / \sum_s n_s^2 \to 0$. Since the third term in eq. (A.17) is of the order T_N, the term is asymptotically negligible in this case. The condition $T_N \to 0$ requires that the shares are sufficiently concentrated so that not too many regions "specialize" in more than one sector (in the sense that the sectoral share w_{is} is bounded away from zero as $S \to \infty$ for more than one sector). In particular, let s_i denote the largest sector in region i. We show in Online Appendix D.1 that when $\max_{i,s\neq s_i} w_{is} \to 0$ (i.e. in each region, the share of the second-largest sector goes to zero as $S \to \infty$), then $T_N \to 0$. For illustration, in the empirical application in Section 7.1, $T_N = 0.0014$.

We also show in Online Appendix D.1 that the third term is asymptotically negligible if conditional variance of the shifters $\mathcal{X}_s$, $\sigma_s^2 = E[(\mathcal{X}_s - \mathcal{Z}_s'\gamma)^2 \mid \mathcal{F}_0]$ and the weighted treatment effects $\sigma_s^2\beta_{is}$ are mean-independent of the shares W, provided some additional mild regularity conditions are satisfied. Importantly, this condition still allows the treatment effects to depend on the controls Z, or other aspects of the model, such as $Y_i(0)$.