Shift-Share Designs: Theory and Inference*

Rodrigo Adão† Michal Kolesár‡ Eduardo Morales§

August 6, 2018

Abstract

We study inference in shift-share regression designs, such as when a regional outcome is regressed on a weighted average of observed 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 that our methods yield substantially wider confidence intervals in popular applications of shift-share regression designs.

*We thank Peter Egger, Gordon Hanson, Bo Honoré, and seminar participants at Carleton University, Yale University, the Globalization & Inequality BFI conference, IDB, GTDW, Unil, EESP-FGV, PUC-Rio, and the Princeton-IES conference for very useful comments. 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: radao@uchicago.edu

‡Princeton University. Email: mkolesar@princeton.edu

§Princeton University. Email: ecmorale@princeton.edu

1 Introduction

We study 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, and whose differential exposure depends on a set of weights, or "shares". Specifically, shift-share regressions have the form

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

For example, in an investigation of the impact of sectoral demand shifters on regional employment changes, Y_i corresponds to 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.[1]

Shift-share specifications can be very appealing in many contexts: they are simple to apply and have the potential to both circumvent complicated endogeneity issues and provide estimates of treatment effects that are robust to different microfoundations. As a result, such specifications have been applied in numerous influential studies, including Bartik (1991), Card (2001) and Autor, Dorn and Hanson (2013). At the same time, two types of concerns have been raised: first, the designs may not be appropriate in the presence of cross-regional general equilibrium effects, and second, the estimand's policy relevance is unclear when the effects of the shifters $\mathcal{X}_s$ are heterogeneous across sectors and regions. In this paper, we put these concerns aside and focus on a different question: how do we perform inference in shift-share regressions?

We find that usual standard error formulas may substantially understate the true variability of OLS estimators of β in eq. (1). To illustrate the empirical importance of this problem, we conduct 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. We construct in this way many placebo samples that 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% level significance 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 heteroscedasticity-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 as the Y_i variable in eq. (1), the rejection rate for 5% level tests can be as high as 55% if heteroscedasticity-robust standard errors are used and 45% for standard errors clustered on state, and it is never below 17%. In other words, suppose that 100 researchers received data on our randomly generated shocks, but were told instead that these are actual sectoral shocks of interest, such as

[1] For simplicity of exposition, we refer to the unit of observation at which the outcome variable is measured as a region, and the unit of observation at which the shifter is measured as a sector. However, our results apply to any regression admitting the representation in eq. (1).

changes in trade flows, tariffs, or immigrant employment. Ideally, at most 5 of them would report statistically significant, false-positive results. However, if these researchers were to use standard inference procedures, up to 55 of them would find a statistically significant effect of the randomly generated shocks on labor market outcomes across U.S. Commuting Zones. The overrejection is even more severe when 2- and 3-digit SIC codes are used to define the sectors, so that the total number of sectors is smaller.

To explain the source of this overrejection problem, we introduce a stylized economic model. Our model features 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 aggregate labor supply. We assume that labor demand in each sector-region pair has an elasticity with respect to wages that is sector-specific and an intercept that, crucially, aggregates several sector-specific components (e.g. sectoral productivities and demand shifters for the corresponding sectoral good). Aggregate labor-supply in each region is upward-sloping and depends on a region-specific intercept.[2]

We use a potential outcome framework to represent the impact of a particular sector-specific labor demand shock on changes in regional employment predicted by the model. Letting $Y_i(x_1, \dots, x_S)$ denote changes in aggregate employment in region i if the shock of interest is exogenously set to $(x_1, \dots, x_S)$, our model implies that

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

where $Y_i(0) = Y_i(0, \dots, 0)$ is region i's employment change if the shock of interest equals zero for all sectors, and $Y_i = Y_i(\mathcal{X}_1, \dots, \mathcal{X}_S)$ is the employment change for the realized shocks $(\mathcal{X}_1, \dots, \mathcal{X}_S)$.

A key insight of our model is that the potential outcome $Y_i(0)$ includes a shift-share component that, using the same shares w_{is}, measures the impact on region i of all sector-level shocks other than the shock of interest $\mathcal{X}_s$. The regression residual ϵ_i in eq. (1) will generally inherit the structure of the potential outcome $Y_i(0)$, and will thus 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. 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 similarity in the shares. This is why, in our placebo exercise, traditional inference methods underestimate the variance of the OLS estimator of β, creating the overrejection problem.

Motivated by the findings of our placebo exercise, we study the 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

[2] In Appendix A, we show that a special case of the model in Adão, Arkolakis and Esposito (2018) microfounds the labor supply and labor demand functions that we assume. In this microfoundation, every region produces a differentiated variety of each sectoral good, varieties are freely traded across regions, labor is the only factor of production, and workers are both immobile across regions and equally productive in all sectors within a region. In Online Appendix C, we provide alternative microfoundations that feature (a) sector-specific capital, as in Jones (1971) and Kovak (2013), and (b) workers with idiosyncratic sectoral productivities, as in Galle, Rodríguez-Clare and Yi (2017), Lee (2017) and Burstein, Morales and Vogel (2018a). We also discuss in this Online Appendix the implications of allowing for labor mobility across regions.

shares w_{is}, controls Z_i, and residuals ϵ_i. This sampling approach is natural given our interest in the causal effect of the shifters $\mathcal{X}_s$: we are interested in what would have happened if the sector-level shock of interest had taken different values, holding everything else constant. The key assumption we impose is that, conditional on the controls Z_i and the shares w_{is}, the shifters $\mathcal{X}_s$ are as good as randomly assigned and independent across sectors. Given this assumption, we show that the regression estimand β in eq. (1) corresponds to a weighted average of the heterogeneous parameters β_{is} in eq. (2), and derive novel confidence intervals that are valid in samples with a large number of regions and sectors under *any* correlation structure of the regression residuals across regions.[3] Our standard error formula essentially forms sectoral clusters whose variance depends on the variance of a weighted sum of the regression residuals ϵ_i, with weights that correspond to the shares w_{is}. To gain intuition on this 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.[4]

We illustrate the finite-sample properties of our novel inference procedure by implementing it on the same placebo samples that we use to illustrate the bias of 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, when applied to perform significance tests, they yield rejection rates that are close to the nominal significance level. As predicted by the theory, our standard error formula remains accurate in the presence of a state-level term in the regression residuals, and no matter whether the shifters $\mathcal{X}_s$ are homoskedastic or heteroskedastic. When the number of sectors is small or a sector is significantly larger that the other ones, our method overrejects relative to the nominal significance level, but it still attenuates the overrejection problem in comparison to usual standard error formulas.

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 the 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 elasticity of regional employment to regional average wages, 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 and Commuting Zones in the United States, as in the literature pioneered by Altonji and Card (1991) and Card (2001).

In these applications, our proposed confidence intervals are substantially wider than those implied by state-clustered or heteroscedasticity-robust standard errors. In particular, the 95% confidence

[3] This result is similar to that 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 shifter of interest is as good as randomly assigned and independent across these individuals' groups, standard errors clustered on groups are valid under any correlation structure of the residuals.

[4] In an extension, we also provide confidence intervals that are valid when the shifters $\mathcal{X}_s$ are independent only across "clusters" of sectors, allowing thus for any correlation of these shifters across sectors belonging to the same "cluster". We also extend our methodology to settings in which the shift-share regressor is not the treatment of interest but an instrument in an instrumental variables estimator.

intervals for the estimated effects of Chinese competition on local labor markets increase by 20%–70%, although these effects remain statistically significant. We obtain similar increases in the length of the 95% confidence interval for the estimated impact of immigration shocks, which are 20%–120% wider than those implied by traditional methods. In contrast, our confidence intervals for the labor supply elasticity estimated using the procedure in Bartik (1991) are almost identical to those constructed using standard approaches; intuitively, the sectoral shifter used in this application—the change in national employment by sector—soaks up most sectoral shocks affecting the outcome variable and, consequently, no shift-share structure is left in the regression residuals.[5]

Shift-share designs have been applied to estimate the effect of a variety of shocks. 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, 2017), 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 also been used to study the impact of these same shocks on alternative outcomes, such as political preferences (Autor et al., 2017a; Che et al., 2017; Colantone and Stanig, 2018), marriage patterns (Autor, Dorn and Hanson, 2018), crime levels (Dix-Carneiro, Soares and Ulyssea, 2017), and innovation (Acemoglu and Linn, 2004; Autor et al., 2017b). 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). In addition to this work using shift-share designs to estimate the overall impact of a shifter of interest, other work has used these designs as part of a more general structural estimation approach; see Diamond (2016), Adão (2016), Galle, Rodríguez-Clare and Yi (2017), Burstein et al. (2018b) and Bartelme (2018).[6] Independently of the aim of the researcher when estimating a shift-share regression, and independently of the interpretation that the researcher assigns to the estimand β in eq. (1), usual standard errors formulas will generally be biased and, as long as the restrictions we impose on the data generating process hold, our novel inference procedures will be valid.

Our paper is related to three other papers studying the statistical properties of shift-share specifications. First, Goldsmith-Pinkham, Sorkin and Swift (2018) focus on the case in which the shift-share regressor is used as an instrumental variable. Within this setting, these authors study the usage of the full vector of shares $(w_{i1}, \ldots, w_{iS})$ as an instrument for the endogenous treatment, and they conclude

[5] To illustrate this point, we estimate the same inverse labor supply elasticity using instead the shift-share instrument in Autor, Dorn and Hanson (2013). The sector shifter in this case—changes in trade flows from China to developed countries other than the U.S.—leaves in the regression residual other sectoral shocks affecting U.S. labor markets; consequently, our confidence intervals are in this case 20%–250% wider than those implied by traditional inference procedures.

[6] Several papers use a shift-share approach that treats the shifters as unobserved, and for this reason uses the shares directly as regressors. This approach has been applied to investigate the impact of technological shifters (Autor and Dorn, 2013), credit supply shifters (Huber, 2018), and immigration shifters (Card and Dinardo, 2000; Monras, 2015). We treat the sectoral shares $\mathcal{X}_s$ as observed and leave the extension to the unobserved case to future work.

that this approach requires that this vector of shares be as good as randomly assigned conditional on the shifters, and independent across regions or clusters of regions. Given our interest in exploring the impact of a specific set of shifters, rather than the impact of a set of shares, this approach is not attractive in our setting. That said, there may be other settings in which this approach is more appealing. Second, Borusyak, Hull and Jaravel (2018), also focusing on the use of a shift-share regressor as an instrumental variable, show that 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. Our approach to inference follows their identification insight; this way of thinking about the shift-share design is also natural given our economic model. Third, Jaeger, Ruist and Stuhler (2018) study complications with the shift-share instrument when it is correlated over time and there is a sluggish adjustment of the outcome variable to changes in it.

The rest of this paper is organized as follows. Section 2 presents the results of a placebo exercise illustrating the properties of inference procedures previously used in the literature on shift-share designs. Section 3 introduces our 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), and provides a consistent estimator of its standard error. Section 5 presents the results of a placebo exercise in which we illustrate the performance of our novel inference procedures. Section 6 revisits the conclusions from several prior applications of shift-share regression analysis, and Section 7 concludes. Appendix A includes a microfoundation for the stylized economic model introduced in Section 3, and Appendix B contains proofs for all propositions in Section 4. Additional results are collected in Online Appendices C, D and E.

2 Overrejection of usual standard errors: placebo evidence

In this section, we implement a placebo exercise to evaluate the finite-sample performance of the two inference methods most commonly applied in shift-share regression designs: a) Eicker-Hubert-White—or heteroskedasticity-robust—standard errors, and (b) standard errors clustered on groups of regions geographically close to each other. In our placebo, we regress observed changes in U.S. regional labor market outcomes on a shift-share regressor that is constructed by combining actual data on initial sectoral employment shares for each region with randomly generated sector-level shocks. We describe the setup in Section 2.1 and summarize 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 $S = 397$ sectors. We identify each region i with a U.S. Commuting Zone (CZ), and each sector s with either a 4-digit SIC manufacturing industry or an aggregated non-manufacturing sector. We index manufacturing industries by $s = 1, \ldots, S-1$ and the non-manufacturing sector by $s = S$.

Using the notation introduced in eq. (1), each placebo sample m has identical values of the shares $\{w_{is}\}_{i=1,s=1}^{N,S}$, the outcomes $\{Y_i\}_{i=1}^{N}$, and the non-manufacturing shifter $\mathcal{X}_S$; the placebo samples differ exclusively in the vector of shifters for the manufacturing sectors $(\mathcal{X}_1^m, \ldots, \mathcal{X}_{S-1}^m)$. Specifically, the shares correspond to employment shares in 1990, the outcomes correspond to changes in employment rates and average wages for different subsets of the population between 2000 and 2007, and the shifter for the non-manufacturing sector is always set to zero, $\mathcal{X}_S = 0$. The vector of shifters for the manufacturing sectors $(\mathcal{X}_1^m, \ldots, \mathcal{X}_{S-1}^m)$ is drawn i.i.d. from a normal distribution with zero mean and variance $\text{var}(\mathcal{X}_s^m) = 5$ in each placebo sample m. Because the shifters are independent of both the outcomes and the shares, the parameter β is zero in every placebo sample m.

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 (with label *St-cluster*).

Our main source of data on employment shares is the County Business Patterns (CBP), 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).[7]

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 $H_0\colon \beta = 0$.

The shifters have no effect on the outcomes and 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.[8] Columns (3) and (4) report the median standard error for *Robust* and *St-cluster* procedures, respectively, and show that both standard error estimators are downward biased relative to the standard deviation of the OLS estimator. On average across all outcomes, the median magnitude of the heteroskedasticity-robust and state-clustered standard errors are, respectively, 41% and 30% lower than the true standard deviation.

The downward bias in the *Robust* and *St-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 statistic should generate a rejection rate of 5%. 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 for a 5% significance level for the null hypothesis $H_0\colon \beta = 0$ is 49.1% and 38.3% when *Robust* and *St-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 our $30,000$ simulated samples to $30,000$ researchers without disclosing

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

[8] Figure D.1 in Online Appendix D.2 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)	St-cluster (4)	Robust (5)	St-cluster (6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	0.74	0.92	49.1%	38.3%
Employed in manufacturing	−0.01	1.88	0.60	0.77	55.6%	44.4%
Employed in non-manufacturing	0.01	0.94	0.58	0.67	23.0%	17.4%
Panel B: Change in average log weekly wage						
Employed	−0.02	2.68	1.02	1.34	47.2%	34.1%
Employed in manufacturing	−0.03	2.93	1.69	2.11	26.4%	16.8%
Employed in non-manufacturing	−0.02	2.65	1.05	1.33	45.4%	33.5%

Notes: For the outcome variable indicated in the first column, this table indicates the median and standard deviation of the OLS estimates across the placebo samples (columns (1) and (2)), the median standard error estimates (columns (3) and (4)), and the percentage of datasets 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 *St-cluster* is the standard error that clusters CZs in the same state. Results are based on 30,000 simulation draws.

to them the origin of the data. 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 these researchers set out to evaluate the impact of these shocks on U.S. CZs using standard inference procedures with a 5% significance level test, 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.[9]

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

To develop some intuition on the source of this overrejection problem, note that the standard error estimators commonly applied in shift-share regression designs assume that the regression residuals are either independent across all regions (for *Robust*), or between geographically defined regional groups (for *St-cluster*). Given that shift-share regressors are correlated across regions with similar sectoral employment shares $\{w_{is}\}_{s=1}^{S}$, these methods generally lead to a downward bias in the standard error estimate whenever regions with similar sectoral employment shares $\{w_{is}\}_{s=1}^{S}$ also tend to have similar regression residuals. In the next section, we consider the implications of a stylized economic model, and show that such correlations between the regression residuals are likely to arise because regions are generally exposed to unobserved sector-level shocks, in addition to the observed shocks $\mathcal{X}_s$. Consequently, whenever a researcher is running a shift-share regression, both heteroskedasticity-robust and state-clustered standard errors will generally be biased downwards.

[9] In Section 5, we extend our analysis to a number of modifications of this baseline setup, including alternative definitions of sectors and regions, allowing for a non-zero shock to the non-manufacturing sector, and allowing for correlations between the shocks to different sectors. The overrejection problem is always at least as severe as in this baseline setup.

3 Stylized economic model

This section presents a stylized economic model mapping sector-level shocks to labor market outcomes for a set of regional economies. The aim of the model is twofold. First, we show that the impact of sectoral shifters on regional labor market outcomes have a shift-share structure, with heterogeneous effects across regions and sectors. Second, we show that unobserved sectoral shifters introduce correlation in the regression residuals across regions with similar observed shares. We describe the model fundamentals in Section 3.1, discuss its main implications for the impact of sectoral shocks 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 $i = 1, \ldots, J$. We assume that the labor demand in sector s and region i, L_{is}, is given by

$$\log L_{is} = -\sigma_s \log \omega_i + \log D_{is}, \qquad \sigma_s > 0, \tag{3}$$

where ω_i is the wage rate in region i, σ_s is the sector-specific labor demand elasticity, and D_{is} are region- and sector-specific labor demand shifters. The shifter D_{is} may account for multiple sectoral components. Since our analysis focuses on the impact of one particular sectoral component, we decompose D_{is} into an observed shifter of interest, χ_s, and other (potentially unobserved) sectoral components, which are grouped into a residual shifter μ_s. That is, without loss of generality, we write

$$\log D_{is} = \rho_s \log \chi_s + \log \mu_s + \log \eta_{is}, \tag{4}$$

where η_{is} is a region- and sector-specific shifter that is mean zero across regions for each sector s.

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

$$\log L_i = \phi \log \omega_i + \log v_i, \qquad \phi > 0, \tag{5}$$

where ϕ is the labor supply elasticity, and v_i is a region-specific labor supply shifter.

Workers are assumed to be immobile across regions, but freely mobile across sectors. Thus, we define the equilibrium as the wages $\{\omega_i\}_{i=1}^{J}$ that satisfy the following market clearing condition:

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

There are multiple microfoundations that are consistent with the labor demand in eq. (3) and the labor supply in eq. (5). For our purposes, the different labor demand microfoundations are important only to the extent that they affect the interpretation of the sector- and region-specific labor demand shifter D_{is}. For example, one could assume that labor is the only factor of production and that every region i is a closed economy and, in this case, D_{is} may account both for demand shifters for sector-specific goods and for sector-specific productivity shifters. Similarly, as we show in Appendix A, we may also allow goods to be freely traded across regions and assume that a subset of the J regions are small open economies; in this case, the shifter D_{is} for these small open economies will account for the world price of sector s, which will itself capture the impact of foreign demand and productivity shocks. We also show in Appendix A that the labor supply in eq. (5) may be derived as the outcome of the utility maximization problem of individuals who, conditional on being employed, are indifferent about the sector of employment, but have heterogeneous disutilities of being employed at all.

3.2 Labor market impact of sectoral shocks

We assume that, in any period, our model characterizes the labor market equilibrium in every region $i = 1, \ldots, J$ and that, across periods, changes in the labor market outcomes $\{\omega_i, L_i\}_{i=1}^{J}$ are due to changes in the sectoral shifter of interest, $\{\chi_s\}_{s=1}^{S}$, other potential sectoral shifters $\{\mu_s\}_{s=1}^{S}$, sector- and region-specific shifters $\{\eta_{is}\}_{i=1,s=1}^{J,S}$, and labor supply shifters, $\{v_i\}_{i=1}^{J}$. Specifically, in every period, the values of these shifters correspond to draws from an unknown joint distribution $F(\cdot)$:

$$(\{\chi_s, \mu_s\}_{s=1}^{S}, \{\eta_{is}\}_{i=1,s=1}^{J,S}, \{v_i\}_{i=1}^{J}) \sim F(\cdot). \tag{7}$$

We use $\hat{z} = \log(z^t/z^0)$ to denote log-changes in a variable z between some initial period $t = 0$ and any other period t. Up to a first-order approximation around the initial equilibrium, eqs. (3) to (6) imply that the change in employment in region i is

$$\hat{L}_i = \sum_{s=1}^{S} l_{is}^0 \left[\beta_{is}\hat{\chi}_s + \lambda_i\hat{\mu}_s + \lambda_i\hat{\eta}_{is}\right] + (1 - \lambda_i)\,\hat{v}_i, \tag{8}$$

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

According to eq. (8), the impact of sectoral shifters on equilibrium employment in region i depends both on the initial sectoral employment shares $\{l_{is}^0\}_{s=1}^{S}$, and the region- and shifter-specific elasticities $\{\beta_{is}, \lambda_i\}_{s=1}^{S}$. Consequently, the employment change in eq. (8) includes several components with a shift-share structure: the "share" term is always the initial employment share in a sector l_{is}^0, and the "shift" term is either the sectoral shock of interest, $\hat{\chi}_s$, or alternative labor demand shocks, $\hat{\mu}_s$. This structure with multiple shift-share terms, some of them observed and others potentially unobserved, is central to understanding the results presented in Section 2.

Notice also that, even conditional on the initial employment share l_{is}^0, the impact of a sector s shifter on region-i's employment may be heterogeneous across sectors and regions: β_{is} may vary across i and s.[10] While standard datasets will usually contain information on the initial employment shares for every sector and region $\{l_{is}^0\}_{i-1,s=1}^{J,S}$, each parameter β_{is} is not generally known or directly observed, and thus, the impact of the sectoral shifters need to be estimated.

We summarize this discussion in the following remark:

Remark 2. *The change in regional employment will generally combine multiple shift-share terms, and the shifter effects depend on parameters that are heterogeneous across sectors and regions.*

[10] In our model, β_{is} does not vary across regions or sectors if and only if all sectors have the same labor demand elasticity, $\sigma_s = \sigma$, and shock pass-through, $\rho_s = \rho$.

The property that the impact of a shifter in sector s on employment in region i may be written as $l^0_{is}\beta_{is}$ that underlies Remark 2 does not depend on the particular microfoundation of the labor demand and labor supply expressions in eqs. (3) and (5). The only difference across these microfoundations is how β_{is} depends on the structural parameters of each microfounded model.

Besides the illustrative example of a possible microfoundation described in Appendix A, we provide alternative microfoundations in Online Appendices C.2 and C.3. Specifically, we show in Online Appendix C.2 that eq. (8) is consistent with a Jones (1971) model featuring sector-specific inputs of production. In Online Appendix C.3, we show that eq. (8) also arises in a Roy (1951) model in which workers have heterogeneous preferences for being employed in the different sectors.

We also extend our model in Online Appendix C.4 to allow for migration across regions. In this case, the change in regional employment $\hat{L}_i$ in any given region $i = 1, \ldots, J$ depends not only on the region's own shift-share terms included in eq. (8), but also on an endogenous component, common to all regions, that combines the shift-share terms corresponding to all regions $i = 1, \ldots, J$. Thus, in the presence of migration, $l^0_{is}\beta_{is}$ is the partial effect of the shifter $\hat{\chi}_s$ on local employment that ignores cross-regional spillovers; consequently, it will only capture the differential effect of the sector-specific shock $\hat{\chi}_s$ on region i relative to all other regions. However, once we condition on fixed effects that absorb these cross-regional spillovers, Remark 2 remains valid for the model with migration.

3.3 From theory to inference

We build on the insights of Section 3.2 to propose a general framework to estimate the effect of shifters on an outcome of interest that varies at a different level than these shifters. For concreteness, we refer to the level at which the shifters vary as sectors, and the level at which the outcome varies as regions, but our results do not depend on these particular labels.

To make precise what we mean by "the effect of shifters on an outcome", we use the potential outcome 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 eq. (8), we assume that the potential outcomes are linear in the shocks,

$$Y_i(x_1, \ldots, x_S) = Y_i(0) + \sum_{i=1}^{S} w_{is} x_s \beta_{is}, \qquad \text{where} \qquad \sum_{s=1}^{S} w_{is} = 1, \tag{9}$$

and $Y_i(0) \equiv Y_i(0, \ldots, 0)$ denotes the potential outcome in region i when all shocks $\{x_s\}_{s=1}^S$ are set to zero. According to eq. (9), increasing x_s by one unit while 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 $\mathcal{X}_1, \ldots, \mathcal{X}_S$. To map eq. (8) into eq. (9), define

$$Y_i = \hat{L}_i, \;\; w_{is} = l^0_{is}, \;\; x_s = \hat{\chi}_s, \quad Y_i(0) = \sum_{s=1}^{S} l^0_{is} \lambda_i (\hat{\mu}_s + \hat{\eta}_{is}) + (1 - \lambda_i)\, \hat{v}_i. \tag{10}$$

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

In the rest of the paper, we assume that we observe data for N regions and S sectors on the sectoral shifters $\mathcal{X}_s$, the regional outcomes Y_i, and the region-sector shares w_{is}.[11,12] 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. To help us focus on the key conceptual issues, we abstract away from any additional covariates or controls for now, and assume that $\mathcal{X}_s$ and Y_i have been demeaned, so that we can omit the intercept in a regression of Y_i on X_i (see Section 4.2 for the case with controls). The OLS estimator of the coefficient on X_i in this simplified setting is given by

$$\hat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2}, \tag{11}$$

and we can write the regression equation as

$$Y_i = \beta X_i + \epsilon_i, \qquad \text{where} \quad X_i \equiv \sum_{s=1}^{S} w_{is}\mathcal{X}_s, \qquad \sum_{s=1}^{S} w_{is} = 1, \tag{12}$$

where β denotes the population analog of $\hat{\beta}$.

The definition of the estimand β 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.[13] 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 fixed the shares w_{is}, the parameters β_{is}, and the potential outcomes $Y_i(0)$.

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. (11) under repeated sampling of the shocks $\mathcal{X}_s$:

$$\beta = \frac{\sum_{i=1}^{N} E[X_i Y_i \mid \mathcal{F}_0]}{\sum_{i=1}^{N} E[X_i^2 \mid \mathcal{F}_0]}, \qquad \text{with} \qquad \mathcal{F}_0 = \{Y_i(0), \beta_{is}, w_{is}\}_{i=1,s=1}^{N,S}, \tag{13}$$

and, given eqs. (9) and (12), the regression error ϵ_i is then defined as the residual

$$\epsilon_i = Y_i - X_i\beta = Y_i(0) + \sum_{i=1}^{S} w_{is}\mathcal{X}_s(\beta_{is} - \beta), \tag{14}$$

where β is defined as in eq. (13).

[11] We can think of the N observed regions as a subset of the J regions existing worldwide and whose labor market equilibrium is described in Sections 3.1 and 3.2.

[12] For simplicity, we assume that we have data on the shifters $\mathcal{X}_s$ directly, rather than possibly noisy estimates of them.

[13] This definition of 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).

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 $\{w_{is}\}_{i=1,s=1}^{N,S}$, and the difference between the parameters $\{\beta_{is}\}_{i=1,s=1}^{N,S}$ and the estimand β. Importantly, the potential outcome $Y_i(0)$ will generally incorporate terms that have a shift-share structure analogous to that of the regressor of interest, X_i. Specifically, as illustrated in eq. (10), the model introduced in Section 3.1 implies that $Y_i(0)$ includes a weighted average of unobserved sectoral labor-demand shocks, $\sum_{s=1}^{S} l_{is}^{0}\lambda_i\hat{\mu}_s$. Hence, if two regions i and i' have similar shares $\{l_{is}^{0}\}_{s=1}^{S}$ and $\{l_{i's}^{0}\}_{s=1}^{S}$, they will tend to have similar regressors X_i and $X_{i'}$ and similar potential outcomes $Y_i(0)$ and $Y_{i'}(0)$. It then follows from eq. (14) that the residuals ϵ_i and $\epsilon_{i'}$ will be correlated.[14]

We summarize this discussion in the following remark.

Remark 3. *Correctly performing inference for the OLS estimator $\hat{\beta}$ of the coefficient on a shift-share regressor requires taking into account that the regression residuals will generally inherit the same shift-share structure.*

Remark 3 has important implications for estimating the variability of $\hat{\beta}$ across samples. 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 discuss in all remaining sections of the paper, 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), present asymptotic results that we derive using these assumptions, and use the model introduced in Section 3.1 to provide an economic interpretation for these assumptions. We first consider in Section 4.1 the simple case in which there is a single regressor with a shift-share structure and no controls, as in Section 3.3. We introduce controls in Section 4.2. Section 4.3 considers further extensions. All proofs and technical details are in Appendix B.

Following the notation introduced in Section 3.3, we write sector-level variables (such as the shocks $\mathcal{X}_s$) in script font style and region-level aggregates (such as X_i) in normal style. To compactly state our assumptions and results, 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 No controls

We study here the statistical properties of the OLS estimator in eq. (11). We assume that, conditionally on the matrix of shares W, the shocks are as good as randomly assigned in that they are independent

[14] As we discuss in Section 4.2, when controls are included, this conclusion will still hold unless the controls account for *all* sectoral shocks other than $\{\mathcal{X}_s\}_{s=1}^{S}$ that affect the outcome.

of the potential outcomes $Y_i(x_1, \ldots, x_S)$. Formally, given the definition of the potential outcomes in eq. (9), we assume

$$(Y(0), B) \perp\!\!\!\perp \mathcal{X} \mid W. \tag{15}$$

In the next subsection, we weaken this assumption by assuming that the shocks are as good as randomly assigned conditionally on some controls.

As discussed in Section 3.3, we consider the statistical properties of $\hat{\beta}$ under repeated sampling of the shocks $\mathcal{X}$, and condition on the realized values of the shares and on the potential outcomes. This approach is analogous to the randomization-style inference in the literature on inference in randomized controlled trials (see Imbens and Rubin, 2015, for a review); it leverages the random assignment assumption in eq. (15), and ensures that the standard errors that we derive will 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. In particular, this approach allows (but does not require) the residual to have a shift-share structure.

We consider asymptotics with the number of sectors going to infinity, $S \to \infty$, and assume that $N \to \infty$ as $S \to \infty$. Formally, the number of regions N thus depends on S, but we keep this conditioning implicit. We do not restrict the ratio N/S, so that the number of regions may grow at a faster rate than the number of sectors. The assumptions needed for the propositions below are collected in Appendix B.1. The key assumption underlying our approach to inference is that the shocks $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ are independent across s conditional on the shares W (see Assumption 1(ii) in Appendix B.1). In contrast, $Y_i(0)$ and the shares w_{is} can be correlated in an arbitrary manner across i. We also do not require $\mathcal{X}$, or any other variables, to be identically distributed—the sectors and regions may be heterogeneous.

The main regularity condition that we need is that each sector is asymptotically negligible in the sense that $\max_s n_s/N \to 0$, where $n_s = \sum_{i=1}^{N} w_{is}$ is the aggregate "size" of sector s in the population of interest (see Assumption 2(ii) in Appendix B.1). It generalizes 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 $s(i)$. Then $w_{is} = 1$ 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. (15) 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 $\max_s n_s/N \to 0$ then reduces to the assumption that the largest cluster be asymptotically negligible.

Proposition 1. *Suppose Assumptions 1 and 2 in Appendix B.1 hold. Then*

$$\beta = \frac{\sum_{i=1}^{N} \sum_{s=1}^{S} \pi_{is} \beta_{is}}{\sum_{i=1}^{N} \sum_{s=1}^{S} \pi_{is}}, \qquad \text{and} \qquad \hat{\beta} = \beta + o_p(1), \tag{16}$$

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

This proposition gives two results. First, it shows that the estimand β in eq. (13) can be expressed as a weighted average of the region- and sector-specific parameters $\{\beta_{is}\}_{i=1,s=1}^{N,S}$, with weights that are increasing in the shares and variance of the shock. Second, it shows that the OLS estimator $\hat{\beta}$ converges to this estimand as $S \to \infty$. The special case with concentrated sectors is again useful to understand Proposition 1. Fully concentrated sectors imply that $\sum_{s=1}^{S} \pi_{is}\beta_{is} = \operatorname{var}(\mathcal{X}_{s(i)} \mid W)\beta_{is(i)}$ 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 earlier, the effect on the outcome variable of increasing the value of the sector s shock in one unit is equal to $w_{is}\beta_{is}$; consequently, for a set of region- and sector-specific weights $\{\xi_{is}\}_{i=1,s=1}^{N,S}$, the corresponding weighted average treatment effect is

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

If β_{is} is constant across i and s, $\beta_{is} = \beta$, then this weighted average can be consistently estimated as $\hat{\tau}_{\xi} = \hat{\beta} \sum_{i=1}^{N}\sum_{s=1}^{S} \xi_{is} w_{is} / \sum_{i=1}^{N}\sum_{s=1}^{S} \xi_{is}$. Furthermore, in this case, the common parameter β has the interpretation that it measures the total effect of increasing the shifters simultaneously in every sector by one unit. Conversely, when 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. (16) to learn something about τ_{ξ}. A special case in which this is possible arises when $\mathcal{X}_s$ is homoscedastic and $\xi_{is} = w_{is}$; in this case, $\hat{\tau}_{\xi} = \hat{\beta} \sum_{i=1}^{N}\sum_{s=1}^{S} w_{is}^2 \hat{\sigma}^2 / \sum_{i=1}^{N}\sum_{s=1}^{S} w_{is}\hat{\sigma}^2$, where $\hat{\sigma}^2$ is a consistent estimate of $\operatorname{var}(\mathcal{X}_s \mid W)$.[15]

Under slight strengthening of the regularity conditions (see Assumption 3 in Appendix B.1), we obtain the following distributional result:

Proposition 2. *Suppose Assumptions 1, 2 and 3 hold, and suppose that*

$$\mathcal{V}_N = \frac{1}{\sum_{s=1}^{S} n_s^2} \operatorname{var}\left(\sum_{i=1}^{N} X_i \epsilon_i \mid Y(0), B, W\right)$$

converges in probability to a non-random limit, where $n_s = \sum_{i=1}^{N} w_{is}$. 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 equal to N/S, the rate of convergence is equal to $\sqrt{S}$. However, if the sizes are unequal, the rate may be slower.

[15] In general, one could consistently estimate τ_{ξ} 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 modeling assumptions predict an equilibrium relationship like that in eq. (8); e.g. see expressions for β_{is} in Appendix A and in Online Appendices C.2 and C.3.

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. Suppose that β_{is} is common across regions and sectors, $\beta_{is} = \beta$, then it follows from eq. (15) and the assumption that $(\mathcal{X}_1, \dots, \mathcal{X}_s)$ are independent across s that[16]

$$\mathcal{V}_N = \frac{\sum_{s=1}^{S} \operatorname{var}(\mathcal{X}_s \mid W) R_s^2}{\sum_{s=1}^{S} n_s^2}, \qquad R_s = \sum_{i=1}^{N} w_{is}\epsilon_i. \tag{17}$$

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

$$\widehat{se}(\hat{\beta}) = \frac{\sqrt{\sum_{s=1}^{S} \mathcal{X}_s^2 \hat{R}_s^2}}{\sum_{i=1}^{N} X_i^2}, \qquad \hat{R}_s = \sum_{i=1}^{N} w_{is}\hat{\epsilon}_i. \tag{18}$$

To gain intuition for the expression in eq. (18), consider the case with concentrated sectors such that the formula becomes $\sum_{s=1}^{S} \mathcal{X}_s^2 \hat{R}_s^2 = \sum_{s=1}^{S} (\sum_{i=1}^{N} \mathbb{I}\{s(i) = s\} X_i \hat{\epsilon}_i)^2$. In this special case, the standard error formula in eq. (18) reduces to the usual cluster-robust standard error, allowing for arbitrary correlation across regions specialized in the same sector.

When regions are not fully specialized in a sector, the standard error in eq. (18) accounts for the fact that regions with similar sectoral composition will generally have similar errors; only in the special case in which the regression error $\epsilon_i = Y_i(0)$ has no sectoral component (so there are no unobserved sector-level shocks), it will be the case that $\operatorname{cov}(X_i\epsilon_i, X_j\epsilon_j) = 0$ for $i \neq j$. In contrast, the usual heteroscedasticity-robust standard error fails to account for this correlation. Standard errors clustered by groups of regions defined by their geographical proximity will also generally fail to account for this correlation. In fact, they will only capture it if and only if all regions are fully specialized in a single sector and the sector of specialization is the same for regions belonging to the same geographically defined cluster.

Remark 4. *In the expression for $\mathcal{V}_N$ in eq. (17), the only expectation is taken over $\mathcal{X}_s$—we do not take any expectation over the shares w_{is} or the residuals ϵ_i. This is because our inference is conditional on the realized values of the shares and on the potential outcomes. In terms of the regression in eq. (12), this means that we consider properties of $\hat{\beta}$ under repeated sampling of $X_i = \sum_s w_{is}\mathcal{X}_s$ conditional on the shares w_{is} and on the residuals ϵ_i (as opposed to, say, considering properties of $\hat{\beta}$ under repeated sampling of the residuals ϵ_i conditional on X_i). As a result, our standard errors allow for arbitrary dependence between the residuals ϵ_i.*

4.1.1 Discussion of assumptions

In general, in order to identify a relationship as causal, one needs a random assignment assumption. In order to do inference and apply a central limit theorem, one needs an independence-type assumption.[17] In our case, the key identifying assumption is that the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$ are as good as

[16]The standard error formula that we provide remains valid if β_{is} is heterogeneous across regions and sectors, as long as some mild restrictions on the form of heterogeneity apply; see Appendix B.6 for a discussion.

[17]For example, for inference on average treatment effects, which is commonly the goal when running a regression, one assumes that the treatment is as good as randomly assigned conditional on controls, and typically also that the data on

randomly assigned conditional on the shares $\{w_{is}\}_{s=1,i=1}^{S,N}$ (see eq. (15)). This identification assumption has been previously suggested by Borusyak, Hull and Jaravel (2018). For inference, we also require that the shocks are independent across sectors. As illustrated through the economic models described in Appendix A and Online Appendix C, these assumptions generally imply restrictions on the stochastic process of economic fundamentals. How strong these restrictions are will depend on the specific context. For example, in a world in which all N regions of interest are closed economies, the only sectoral shocks are either productivity or preference shocks, and the shifters of interest are the former, these assumptions require that, conditional on the shares, the productivity shocks are: (a) independent of preference shocks; and (b) independent across sectors. In Section 4.2, we illustrate how to relax assumption (a) by incorporating controls into the regression specification and, in Section 4.3.2, we show how to relax it by using instrumental variables. Additionally, we show in Section 4.3.1 how to relax assumption (b) by allowing for a non-zero correlation in the sectoral shocks of interest within clusters of sectors.

Goldsmith-Pinkham, Sorkin and Swift (2018) investigate a different approach to identification based on the assumption that the shares $(w_{i1}, \ldots, w_{iS})$ are as good as randomly assigned conditional on the shifters $\mathcal{X}_s$. For inference, this approach requires that the shares $(w_{i1}, \ldots, w_{iS})$ be independent across regions or clusters of regions. However, as illustrated through the stylized economic model presented in Section 3, these shares are generally equilibrium objects and, consequently, they are unlikely to be as good as randomly assigned. For instance, in the case of the environment described in Section 3.1, under the assumption that $\sigma_s = \sigma$ for all sectors, it holds that $l_{is}^0 = D_{is}^0/(\sum_{k=1}^S D_{ik}^0)$, where D_{is}^0 is the labor demand shifter of sector s in region i in the initial equilibrium. However, as shown in eqs. (4), (10) and (14), the regression residual ϵ_i accounts for changes in certain variables that also affect the demand shifter D_{is}^0 and, consequently, l_{is}^0 will generally be correlated with ϵ_i unless changes in those variables are independent of their past initial levels.[18] Furthermore, as the demand shifters D_{is}^0 are likely to depend on terms that vary by sector (see eq. (4)), the labor shares l_{is}^0 will generally be correlated across all regions $i = 1, \ldots, N$ for a given sector s, complicating the task of deriving valid inference procedures in this setting.

The results in Propositions 1 and 2 also require the assumption that $\max_s n_s/N \to 0$. In terms of the economic model introduced in Section 3, this assumption imposes that no one sector dominates the others in terms of initial employment at the national level; i.e. $\sum_{i=1}^N l_{is}^0$ is not too large for any one sector. As we illustrate in Section 5.2, this condition is satisfied for the U.S. when only manufacturing sectors are taken into account; it would not hold if the non-manufacturing sector is included as one of the S sectors incorporated into the analysis (unless the distribution of $\mathcal{X}_s$ for the non-manufacturing sector is degenerate at zero).[19]

Finally, Propositions 1 and 2 also require the number of sectors and the number of regions to go to

individuals is i.i.d., which implies that the treatment is independent across individuals conditional on the controls.

[18] Importantly, the correlation between the shares $\{w_{is}\}_{s=1}^S$ and the regression residuals ϵ_i does not affect the consistency of the OLS estimator of β if the shifters $\mathcal{X}_s$ are as good as randomly assigned conditional on the shares w_{is}.

[19] When analyzing the impact of international trade on regional labor market outcomes, it is standard to either set the shock of the non-manufacturing sector to zero (Topalova, 2007, 2010; Autor, Dorn and Hanson, 2013; Hakobyan and McLaren, 2016) or to remove the non-manufacturing sector from the analysis and rescale the shares of all manufacturing sectors so that they add up to one (Kovak, 2013). Either of these approaches will satisfy the restriction that $\max_s n_s/N \to 0$.

infinity. Shift-share designs are however sometimes used in settings in which the number of regions or the number of sectors is small. Through placebo exercises, we illustrate in Section 5 the finite-sample properties of the standard error estimator introduced in eq. (18): our estimates are very close to the true standard deviation of the estimator $\hat{\beta}$ for sample sizes employed in typical applications.

4.2 General case with controls

In many applications of shift-share regression designs, a K-vector of regional controls Z_i is included in the regression specification. We now study the properties of the OLS estimator of the coefficient on X_i in a regression of Y_i onto X_i and Z_i. To this end, let Z denote the $N \times K$ matrix with i-th row given by Z_i', and let $\ddot{X} = X - Z(Z'Z)^{-1}Z'X$ denote an N-vector whose i-th element is equal to the regressor X_i with the controls Z_i partialled out (i.e. the i-th residual from regressing X onto Z). Then, by the Frisch–Waugh–Lovell theorem, $\hat{\beta}$ is equivalent to

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_i}{\sum_{i=1}^{N} \ddot{X}_i^2} = \frac{\ddot{X}'Y}{\ddot{X}'\ddot{X}}, \tag{19}$$

and the OLS estimator of the coefficient on Z_i is equivalent to

$$\hat{\delta} = (Z'Z)^{-1}Z'(Y - X\hat{\beta}).$$

The controls Z may play two roles. First, controls may be included to increase the precision of $\hat{\beta}$. Second, and more importantly, they may be included to proxy for latent sector-level shocks $\{\mathcal{Z}_s\}_{s=1}^{S}$ that have an independent effect on the outcome Y and are correlated with the shifters $\{\mathcal{X}_s\}_{s=1}^{S}$. In the presence of such shocks, the shifters are only as good as randomly assigned conditional on them, and it is necessary to control for them in order to prevent omitted variable bias.

To account for the two possible roles that controls may play, we assume that the controls Z_i admit the decomposition

$$Z_i = \sum_{s=1}^{S} w_{is}\mathcal{Z}_s + U_i. \tag{20}$$

If the kth component Z_{ik} of Z_i is included for precision, then $\mathcal{Z}_{sk} = 0$ for all $s = 1, \ldots, S$, and Z_{ik} is included because $Y_i(0)$ and U_{ik} are correlated. This is the case, for instance, if $Y_i(0)$ and U_{ik} contain regional shocks that are independent of the sectoral shifters of interest $\mathcal{X}$. If, on the other hand, Z_{ik} is included to proxy for a latent shock $\mathcal{Z}_s$, then U_{ik} represents the measurement error in Z when controlling for $\mathcal{Z}$ and Z_{ik} is a perfect only if $U_{ik} = 0$.

With this setup, we replace eq. (15) with the assumption that

$$(U, Y(0), B) \perp\!\!\!\perp \mathcal{X} \mid \mathcal{Z}, W, \tag{21}$$

where $\mathcal{Z}$ denotes the $S \times K$ matrix with sth row given by $\mathcal{Z}_s'$, and U denotes the N-vector with i-th element given by U_i.

To facilitate the interpretation of the condition in eq. (21), it is useful to consider a projection of the regional potential outcomes onto the sectoral space. For simplicity, consider the case with constant effects, $\beta_{is} = \beta$, and suppose $U_i = 0$. Project $Y_i(0)$ onto the sector-level controls $\mathcal{Z}_s$, so that we can write $Y_i(0) = \sum_{s=1}^{S} w_{is}\mathcal{Z}_s'\kappa + \eta_i$. Then eq. (21) holds if the residuals η_i in this projection are independent of $\mathcal{X}$—if there are any other unobserved sector-level shocks that affect the outcomes (and are therefore in η_i), these must be unrelated to $(\mathcal{X}_s, \mathcal{Z}_s)$.

To ensure that it suffices to include the controls in the regression linearly (instead of having to control for them non-parametrically), we additionally assume that the expectation of $\mathcal{X}_s$ conditional on $\mathcal{Z}_s$ is linear in $\mathcal{Z}_s$,

$$E[\mathcal{X}_s \mid \mathcal{Z}, W] = \mathcal{Z}_s'\gamma, \tag{22}$$

where γ is a K-vector that equals 0 if and only if the scalar $\mathcal{X}_s$ is mean independent of the K-vector $\mathcal{Z}_s$. We then obtain the following generalization of Proposition 1:

Proposition 3. *Suppose Assumptions 2 and 4 in Appendix B.1 hold, and that $U_i'\gamma = 0$ for $i = 1, \dots, N$. Then,*

$$\beta = \frac{\sum_{i=1}^{N}\sum_{s=1}^{S}\pi_{is}\beta_{is}}{\sum_{i=1}^{N}\sum_{s=1}^{S}\pi_{is}}, \qquad \textit{and} \qquad \hat{\beta} = \beta + o_p(1), \tag{23}$$

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

The only difference with to Proposition 1 is that the weights π_{is} now reflect the variance of $\mathcal{X}_s$ conditional on $\mathcal{Z}$ and W, rather than just conditional on W. An additional assumption is the requirement that $U_i'\gamma = 0$ for all i. Effectively, this requires that, for each control k, either $U_{ik} = 0$ for all i, so that Z_{ik} is a perfect proxy for the sector-level variables $\mathcal{Z}_{1k}, \dots, \mathcal{Z}_{Sk}$, or else $\gamma_k = 0$, so that $\mathcal{Z}_{sk}$ is unrelated to $\mathcal{X}_s$—the proxy need not be perfect in this case, since it is not necessary to control for $\mathcal{Z}_{sk}$ in the first place (including Z_{ik} in the regression only affects the precision, but not the consistency, of $\hat{\beta}$). If $U_i'\gamma \neq 0$, then there will be omitted variable bias due to inadequately controlling for the confounders $\mathcal{Z}$. This is analogous to the classic linear regression result that measurement error in a control variable leads to a bias in the estimate of the coefficient on the variable of interest.

To state the asymptotic normality result, we need to define the residual ϵ_i in the regression equation $Y_i = X_i\beta + Z_i'\delta + \epsilon_i$. To this end, let

$$\delta = E[Z'Z]^{-1}E[Z'(Y - X\beta)]$$

denote the population regression coefficient on Z_i. We then define the regression residual as $\epsilon_i = Y_i - X_i\beta - Z_i'\delta$ and obtain the following generalization of Proposition 2:

Proposition 4. *Suppose Assumptions 2, 3, 4 and 5 in Appendix B.1 hold, and that $U_i'\gamma = 0$ for $i = 1, \dots, N$. Suppose also that*

$$\mathcal{V}_N = \frac{1}{\sum_{s=1}^{S} n_s^2} \operatorname{var}\left(\sum_i (X_i - Z_i'\gamma)\epsilon_i \mid Y(0), B, U, \mathcal{Z}, W\right)$$

converges in probability to a non-random limit, and let $n_s = \sum_{i=1}^N w_{is}$. Then

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

Relative to Proposition 2, the only 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 construct a consistent standard error estimate, similarly to the case without controls, it suffices to construct a consistent estimate of $\mathcal{V}_N$, the middle part of the sandwich. We derive the standard error formula under the assumption that $\beta_{is} = \beta$ for all i, s.[20] Under this assumption, it follows from eq. (21) and the assumption that $(\mathcal{X}_1, \dots, \mathcal{X}_s)$ are independent across s that

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

A plug-in estimate of R_s can be constructed by replacing ϵ_i with the estimated regression residuals $\hat{\epsilon}_i = Y_i - X_i\hat{\beta} - Z_i\hat{\delta}$. To construct an estimate of the variance $\operatorname{var}(\tilde{\mathcal{X}}_s \mid W, \mathcal{Z})$, we first project the consistent estimate $\ddot{X}_i$ of $X_i - Z_i'\gamma$ onto the sectoral space by regressing it onto the shares W_i,

$$\widehat{\mathcal{X}} = (W'W)^{-1}W'\ddot{X}, \tag{24}$$

and we then estimate the variance $\operatorname{var}(\tilde{\mathcal{X}}_s \mid W, \mathcal{Z})$ by $\widehat{\mathcal{X}}^2$. This leads to the standard error estimate

$$\widehat{se}(\hat{\beta}) = \frac{\sqrt{\sum_{s=1}^S \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{\sum_{i=1}^N \ddot{X}_i^2}, \qquad \hat{R}_s = \sum_{i=1}^N w_{is}\hat{\epsilon}_i. \tag{25}$$

The next remark summarizes these steps:

Remark 5. *To construct the standard error estimate in eq. (25):*

1. *Obtain the estimates $\hat{\beta}$ and $\hat{\delta}$ by regressing Y_i onto $X_i = \sum_s w_{is}\mathcal{X}_s$ 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. This requires the share matrix W to be full rank, which itself requires $N > S$.*

Plug the estimates $\hat{\epsilon}_i$, $\ddot{X}_i$, and $\widehat{\mathcal{X}}_s$ into the standard error formula in eq. (25).

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. (25) reduces to the usual cluster-robust standard error, with clustering on the sectors $s(i)$.

It has been shown that 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

[20] We discuss in Appendix B.6 the restrictions under which our standard error formula remains valid when the effects are heterogeneous.

and Miller, 2014 for a survey). Since the standard error in eq. (25) 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. 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 $\hat{\epsilon}_{\beta_0,i}$, 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 $X_i = \sum_s w_{is}\mathcal{X}_s$ 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).*

Compute the standard error as

$$\widehat{se}_{\beta_0}(\hat{\beta}) = \frac{\sqrt{\sum_{s=1}^{S}\widehat{\mathcal{X}}_s^2 \hat{R}_{\beta_0,s}^2}}{\sum_{i=1}^{N}\ddot{X}_i^2}, \qquad \hat{R}_{\beta_0,s} = \sum_{i=1}^{N} w_{is}\hat{\epsilon}_{\beta_0,i}. \tag{26}$$

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

$$\hat{\beta} - A \pm \sqrt{A^2 + \frac{\widehat{se}(\hat{\beta})^2}{Q/(\ddot{X}'\ddot{X})^2}}, \qquad A = \frac{\sum_{s=1}^{S}\widehat{\mathcal{X}}_s^2 \hat{R}_s \sum_i w_{is}\ddot{X}_i}{Q}, \tag{27}$$

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. (25).

Since in both $\hat{\epsilon}_i$ and $\hat{\epsilon}_{\beta_0,i}$ are consistent estimates of the residuals, both $\widehat{se}_{\beta_0}(\hat{\beta})$ and $\widehat{se}(\hat{\beta})$ are consistent estimates of the standard error and, consequently, yield tests and confidence intervals that are asymptotically valid. The next proposition formalizes this result.

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 $\max_s \sum_i |((W'W)^{-1}W')_{si}|$ is bounded, or else that $U_i = 0$ for $i = 1, \dots, N$. Define $\widehat{\mathcal{X}}$ as in eq. (24), 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*

$$\frac{\sum_{s=1}^{S}\widehat{\mathcal{X}}_s^2 \hat{R}_s^2}{\sum_{s=1}^{S} n_s^2} = \mathcal{V}_N + o_p(1). \tag{28}$$

The additional assumption of Proposition 5 is that either $\max_s \sum_i |((W'W)^{-1}W')_{si}|$ is bounded or, else, $U_i = 0$ for all i. This assumption ensures 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 $\max_s \sum_i |((W'W)^{-1}W')_{si}| = 1$, 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 the formula in Remark 6 that the confidence interval with the null imposed is not symmetric around $\hat{\beta}$, but its center is shifted by A.[21] As we show in Section 5, this recentering tends to improve the finite-sample coverage properties of the confidence interval. On the other hand, the confidence interval tends to be longer on average than that in Remark 5.

4.2.1 Discussion of assumptions

The role that controls play in our framework is twofold. First, the k-th element of the vector Z_i may proxy for the impact on region i of an unobserved sectoral shock $(\mathcal{Z}_{1k}, \dots, \mathcal{Z}_{Sk})$. In the context of the model in Section 3, regional labor market outcomes are not only affected by the sectoral shifters of interest $(\chi_1, \dots, \chi_S)$, but also by other sectoral shocks accounted for by the composites $(\mu_1, \dots, \mu_S)$, as illustrated in eq. (4). When the regression of Y_i on X_i does not include a vector of controls Z_i, consistent estimation of β requires assuming that the vector of sectoral shocks of interest $(\chi_1, \dots, \chi_S)$ is independent of all other sectoral shocks $(\mu_1, \dots, \mu_S)$. On the other hand, if we control for the impact of the sectoral shocks $(\mu_1, \dots, \mu_S)$ on regional labor market outcomes through a regional control $Z_i = \sum_{s=1}^{S} w_{is}\mu_s$, then the OLS estimator $\hat{\beta}$ will be consistent even if $(\mu_1, \dots, \mu_S)$ are not independent of the sectoral shocks of interest $(\chi_1, \dots, \chi_S)$.

Second, each element of the vector Z_i may proxy for regional shocks that, although independent of the sectoral shocks of interest $\mathcal{X}_s$, have an effect on the outcome variable and, thus, enter the regression residual ϵ_i in eq. (12). Controlling for these shocks is not necessary for the consistency of $\hat{\beta}$, but including them increases its precision. An example of such a shock in the context of the model in Section 3 is the region-specific labor supply shifter v_i, as long as $\{v_i\}_{i=1}^{N}$ are independent of the shocks of interest $\{\mathcal{X}_s\}_{s=1}^{S}$. If this independence condition does not hold, then it is important to control for these labor supply shocks in order to ensure consistency of $\hat{\beta}$.

Even if all other sectoral shocks are independent of the shifters of interest, including controls that proxy for them in the regression will reduce the correlation between residuals of regions with similar shares, and it may therefore attenuate the overrejection problem of traditional inference methods documented in Section 2. In the limit, if the controls soak up all sectoral shocks, so that the residuals ϵ_i are independent across i, the usual heteroscedasticity-robust confidence intervals will give correct coverage, and our confidence intervals will be asymptotically equivalent to them. However, since our inference methods are valid whether or not there is shift-share structure in the residuals, we recommend that researchers always use them, in line with the practice of always clustering the standard

[21] 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).

errors if the regressor of interest only varies at a group level.

4.3 Extensions

We now discuss two extensions of the basic setup: first, we weaken the assumption that $(\mathcal{X}_1, \ldots, \mathcal{X}_s)$ are independent across s. Second, we consider using the shift-share regressor X_i as an instrument.

4.3.1 Clusters of sectors

Suppose that the sectors can be grouped into larger units, which we refer to as "clusters", with $c(s) \in \{1, \ldots, C\}$ denoting the cluster that sector s belongs to; e.g., s may be a four-digit industry code, while $c(s)$ is a three-digit code. With this structure, we replace here the assumption that the shocks $\mathcal{X}_s$ are independent across sectors (Assumption 1(ii) in Appendix B.1 for the case without controls, and Assumption 4(ii) for the general case) with the assumption that, conditional on $\mathcal{Z}$ and W, the shocks $\mathcal{X}_s$ and $\mathcal{X}_k$ are independent if $c(s) \neq c(k)$ (for the case without controls, we just take $\mathcal{Z}$ to be a vector of ones). Also, we replace the assumption that the largest sector makes an asymptotically negligible contribution to the asymptotic variance (Assumption 2(ii) in Appendix B) 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 $C \to \infty$,

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

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

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

In other words, instead of treating $\tilde{\mathcal{X}}_s R_s$ as independent across s, the asymptotic variance formula clusters them. As a result, we replace the standard error estimate in eq. (25) with

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

with $\widehat{\mathcal{X}}_s$ 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. (29), and using this formula instead of that in eq. (26).

4.3.2 Instrumental variables regression

Consider the problem of estimating the effect of a regional treatment variable Y_{2i} on an outcome variable Y_{1i} using a shift-share regressor $X_i = \sum_s w_{is} \mathcal{X}_s$ as an instrument. We maintain the assumption that there is a K-vector of latent sectoral controls $\mathcal{Z}_s$ such that the regression specification includes a vector regional controls Z_i that have the structure in eq. (20) and such that eq. (22) holds.

We assume that the effect of Y_{2i} onto Y_{1i} is linear and constant across regions, so that the potential outcome when Y_{2i} is exogenously set to y_2 is given by

$$Y_{i1}(y_2) = Y_{i1}(0) + y_2\alpha,$$

where α is the treatment effect of Y_{2i} on Y_{1i} for every region i. The observed outcome is thus $Y_{1i} = Y_{1i}(Y_{2i})$. In analogy with eq. (9), we denote the region-i treatment level that would occur if the region received shocks $(x_1, \dots, x_S)$ as

$$Y_{2i}(x_1, \dots, x_S) = Y_{2i}(0) + \sum_{i=1}^{S} w_{is} x_s \beta_{FS}. \tag{30}$$

The observed treatment level is $Y_{2i} = Y_{2i}(\mathcal{X}_1, \dots, \mathcal{X}_S)$. For simplicity, we assume that β_{FS} does not vary across sectors or regions. Finally, we assume that, conditional on $\mathcal{Z}$, the shocks $\mathcal{X}$ are as good as randomly assigned and satisfy the exclusion restriction, so that the following restriction holds:

$$(U, Y_1(0), Y_2(0)) \perp\!\!\!\perp \mathcal{X} \mid \mathcal{Z}, W. \tag{31}$$

This restriction allows $Y_{1i}(0)$ and $Y_{2i}(0)$ to be correlated; thus, the observed treatment level Y_{2i} may be correlated with the potential outcomes (i.e. endogenous), even conditional on the controls Z_i.

Both the reduced-form regression of Y_{i1} onto X_i and Z_i and the first-stage regression of Y_{2i} onto X_i and Z_i fit into the setup of Section 4.2. Thus, by generalizing the arguments in Section 4.2, we can derive the joint asymptotic distribution of the reduced-form and first-stage coefficients on X_i:

$$\hat{\beta}_{RF} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_{1i}}{\sum_{i=1}^{N} \ddot{X}_i^2} \qquad \text{and} \qquad \hat{\beta}_{FS} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_{2i}}{\sum_{i=1}^{N} \ddot{X}_i^2}.$$

Since the IV estimate of α is given by

$$\hat{\alpha} = \frac{\sum_{i=1}^{N} \ddot{X}_i Y_{1i}}{\sum_{i=1}^{N} \ddot{X}_i Y_{2i}} = \frac{\hat{\beta}_{RF}}{\hat{\beta}_{FS}}, \tag{32}$$

we can obtain the asymptotic distribution of $\hat{\alpha}$ by the delta method. To state the result, we define the reduced-form and first-stage regression errors, $\epsilon_{1i} = Y_{1i} - Z_i'\delta_{RF} - X_i\beta_{RF}$ and $\epsilon_{2i} = Y_{2i} - Z_i'\delta_{FS} - X_i\beta_{FS}$, where $\delta_{RF} = E[Z'Z]^{-1}E[Z'(Y_1 - X\beta_{RF})]$ and $\delta_{FS} = E[Z'Z]^{-1}E[Z'(Y_2 - X\beta_{FS})]$, and it thus follows from eq. (31) that the population reduced-form coefficient on X_i is given by $\beta_{RF} = \beta_{FS}\alpha$. Then, as long as $\beta_{FS} \neq 0$, so that the shift-share instrument is relevant, it holds

$$\frac{N}{\sqrt{\sum_{s=1}^{S} n_s^2}}(\hat{\alpha} - \alpha) = \mathcal{N}\left(0, \frac{\frac{1}{\sum_{s=1}^{S} n_s^2}\sum_{s=1}^{S} \operatorname{var}(\tilde{\mathcal{X}}_s \mid \mathcal{Z}, W) R_s^2}{(\frac{1}{N}\sum_i \ddot{X}_i^2)^2 \beta_{FS}^2}\right) + o_p(1), \qquad R_s = \sum_{i=1}^{N} w_{is}(\epsilon_{1i} - \epsilon_{2i}\alpha).$$

This suggests the standard error estimate

$$\widehat{se}(\hat{\alpha}) = \frac{\sqrt{\sum_{s=1}^{S} \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{|\hat{\beta}_{FS}| \sum_{i=1}^{N} \ddot{X}_i^2} = \frac{\sqrt{\sum_{s=1}^{S} \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}}{|\sum_{i=1}^{N} \ddot{X}_i Y_{2i}|}, \qquad \hat{R}_s = \sum_{i=1}^{N} w_{is} \hat{\epsilon}_{\Delta,i}, \tag{33}$$

with $\widehat{\mathcal{X}}_s$ constructed as in Remark 5, and $\hat{\epsilon}_\Delta = Y_1 - Y_2\hat{\alpha} - Z'(Z'Z)^{-1}Z'(Y_1 - Y_2\hat{\alpha})$ corresponds to the estimate of the residual in the structural equation, $\epsilon_{1i} - \epsilon_{2i}\alpha$.[22] The difference between the IV standard error formula in eq. (33) and the OLS version in eq. (25) is analogous to the difference between IV and OLS heteroscedasticity-robust standard errors: $\hat{\epsilon}_i$ is replaced in the numerator by the estimate of the structural residual $\hat{\epsilon}_{\Delta,i}$, and the denominator is scaled by the first-stage coefficient. The IV analog of the standard error estimate with the null $H_0\colon \alpha = \alpha_0$ imposed estimates the residual as $(I - Z'(Z'Z)^{-1}Z')(Y_1 - Y_2\alpha_0)$, and the resulting confidence interval is a generalization of the Anderson and Rubin (1949) confidence interval (which assumes that the structural errors are independent). As a result, this confidence interval will remain valid even if the shift-share instrument is weak.

Faced with the problem of estimating the treatment effect α in a setting in which the instrument has a shift-share structure, our approach to identification follows Borusyak, Hull and Jaravel (2018), who impose an assumption analogous to that in eq. (31), and also discuss the extension to a setting in which β_{FS} is allowed to vary across sectors and regions and α is allowed to vary across regions. In contrast, Goldsmith-Pinkham, Sorkin and Swift (2018) consider replacing the shift-share instrument X_i with the full vector of shares $(w_{i1}, \dots, w_{iS})$. Importantly, there are settings in which X_i satisfies the exclusion restriction but the full vector $(w_{i1}, \dots, w_{iS})$ does not, and is thus not a valid instrument.[23] Intuitively, this is the case when the residual in the structural equation ϵ_Δ has a shift-share structure. Our independence restriction in eq. (31) allows for this possibility and, consequently, we adopt the approach that has been standard since Bartik (1991) and use the shift-share X_i as an instrument.

5 Performance of new methods: placebo evidence

In Section 5.1, we revisit the placebo exercise in Section 2 to illustrate the finite-sample properties of the standard error estimators introduced in Section 4. In Sections 5.2 to 5.4, we consider several extensions to illustrate the sensitivity of our standard errors to assumptions underlying their validity and to show that the overrejection problem affecting commonly used inference procedures is persistent.

5.1 Baseline specification

We first consider the performance of the standard error estimator in eq. (25) (which we label *AKM*), and the standard error and confidence interval in eqs. (26) and (27) (with label *AKM0*), in the baseline placebo samples described in Section 2. As these samples include no controls, we fix the matrix Z to

[22] Since the IV regression uses a single constructed instrument, $\hat{\epsilon}_\Delta$ is numerically equivalent to $\hat{\epsilon}_{1i} - \hat{\epsilon}_{2i}\hat{\alpha}$, where $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ are the reduced-form and first-stage residuals.

[23] For an online discussion by Tim Bartik on this point, see `https://blogs.worldbank.org/impactevaluations/comment/5042#comment-5042`. See also Borusyak, Hull and Jaravel (2018) for a discussion of different identification assumptions in this setting.

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

	Estimate		Median eff. s.e.		Rejection rate	
	Mean	Std. dev	AKM	AKM0	AKM	AKM0
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	1.90	2.21	7.9%	4.8%
Employed in manufacturing	−0.01	1.88	1.77	2.05	8.3%	4.7%
Employed in non-manufacturing	0.01	0.94	0.89	1.03	8.5%	4.8%
Panel B: Change in average log weekly wage						
Employed	−0.02	2.68	2.57	2.99	8.3%	5.1%
Employed in manufacturing	−0.03	2.93	2.75	3.18	9.0%	4.6%
Employed in non-manufacturing	−0.02	2.65	2.55	2.96	8.3%	5.1%

Notes: For the outcome variable indicated in the first column, this table indicates the median effective standard error (Median eff. s.e.) across the simulated datasets, and the percentage of datasets for which we reject the null hypothesis $H_0\colon \beta = 0$ using a 5% significance level test. *AKM* is the standard error in Remark 5; and *AKM0* is the confidence interval in Remark 6. 30,000 simulation draws.

be a column of ones when implementing the formulas in eqs. (25) and (27).

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 across the placebo samples of the "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 table shows that our proposed methods 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 of *AKM0* is slightly larger than the standard deviation of $\hat{\beta}$, by about 10% on average.[24] The implied rejection rates are close to the 5% nominal rate: the *AKM* procedure has a rejection rate that is between 7.9% and 9% and the *AKM0* procedure has a rejection rate that is always between 4.6% and 5.1%.

As discussed in Section 4.2, *AKM* and *AKM0* are asymptotically equivalent. The differences in rejection rates between *AKM* and *AKM0* in Table 2 are thus due to differences in finite-sample performance. It has been noted in other contexts (see Lazarus et al., 2018) that confidence intervals that impose the null can lead to improved finite-sample size control relative to the usual confidence intervals that do not do so. 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 and, consequently, helps reduce the finite-sample bias that arises in estimating the asymptotic variance.

5.2 Alternative number of sectors and correlated sectoral shocks

As discussed in Section 4, the inference procedures described in Remarks 5 and 6 generate tests and confidence intervals that are valid in large samples if: (a) the number of sectors goes to infinity; (b)

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

Table 3: Rejection rate of $H_0\colon \beta = 0$ at 5% significance level: sensitivity to departures from baseline specification

	Robust	St-cluster	AKM	AKM0
Panel A: Different sectoral aggregation				
2-digit ($S = 20$)	70.8%	57.0%	13.1%	5.4%
3-digit ($S = 136$)	54.6%	42.7%	7.8%	4.8%
4-digit ($S = 398$)	48.6%	37.7%	7.8%	4.5%
Panel B: Simulated non-manufacturing shocks				
	92.0%	89.5%	77.5%	76.7%
Panel C: Heteroskedastic sector-level shocks				
	48.4%	37.4%	7.4%	4.9%
Panel D: Simulated state-level shocks				
	42.6%	30.1%	7.6%	4.6%

Notes: All estimates in this table use the total employment share in each CZ as the outcome variable Y_i. The first row indicates the inference procedure employed to compute the share of the 30,000 simulated datasets for which we reject the null hypothesis $H_0\colon \beta = 0$ using a 5% significance level. *Robust* is the Eicker-Huber-White standard error; *St-cluster* is the standard error that clusters CZs in the same state; *AKM* is the standard error in Remark 5; *AKM0* is the test in Remark 6.

all sectors are asymptotically "small"; and (c) the sectoral shocks are independent across sectors. On the other hand, the inference procedures remain valid under (d) heteroscedastic sectoral shocks, and (e) arbitrary correlation structure of the residuals ϵ_i across regions. We test how sensitive different inference procedures are to assumptions (a) and (b) in panels A and B, respectively, of Table 3. Panels C and D demonstrate the robustness of *AKM* and *AKM0* to (d) and (e). We investigate robustness to violations of assumption (c) in Table 4. In all these tables, we focus on the change in the share of working-age population employed as the outcome variable of interest.

In Panel A of Table 3, we change the baseline placebo specification described in Section 2 by changing the definition of sector. The results show that the overrejection problem affecting standard inference procedures is worse when the number of sectors decreases: the rejection rates for *Robust* and *St-cluster* standard errors reach 70.8% and 57%, respectively, when the 396 4-digit SIC sectors are substituted by the 20 2-digit SIC sectors in the analysis. In line with the findings in the literature on clustering with a few clusters, the rejection rates for *AKM* also increase to 13.1%, but those for *AKM0* remain very close to the nominal 5% significance level.

In Panel B, we modify the baseline placebo setup in that we set the variance of the shock assigned to the non-manufacturing sector to a positive number; specifically, we set its variance to 5, the same as for the remaining sectoral shocks. All methods perform poorly in this case. The reason is that, across the CZs in our analysis, the non-manufacturing sector accounts on average for 77.5% of employment, with a minimum employment share of 38%. This demonstrates that it is important in practice for all sectors included in the analysis to be small.

Panel C investigates the robustness of our results to heteroscedasticity in the sector-level shocks.

Table 4: Rejection rate of $H_0 \colon \beta = 0$ at 5% significance level: correlation in sectoral shocks

Sector Cluster:			Independent		3-digit SIC		2-digit SIC	
	Robust	St-cluster	AKM	AKM0	AKM	AKM0	AKM	AKM0
Panel A: Simulated shocks with correlation within 3-digit SIC sectors								
$\rho = 0.25$	47.2%	30.4%	5.6%	6.0%	4.5%	4.8%	4.8%	5.0%
$\rho = 0.50$	49.2%	32.4%	6.8%	7.7%	4.4%	4.8%	4.6%	.05%
$\rho = 0.75$	51.0%	33.8%	8.1%	9.0%	4.5%	4.9%	4.7%	5.2%
$\rho = 1.00$	52.6%	35.6%	9.5%	10.6%	4.5%	4.9%	4.9%	5.2%
Panel B: Simulated shocks with correlation within 2-digit SIC sectors								
$\rho = 0.25$	56.6%	38.8%	14.1%	16.0%	12.3%	14.0%	5.3%	5.3%
$\rho = 0.50$	62.0%	43.0%	22.2%	25.0%	17.9%	20.6%	5.8%	5.3%
$\rho = 0.75$	66.4%	46.9%	28.8%	32.5%	22.2%	25.8%	6.2%	5.4%
$\rho = 1.00$	68.6%	48.0%	33.3%	37.0%	24.9%	28.9%	6.6%	5.5%
Panel C: Simulated shocks with correlation within 1-digit SIC sectors								
$\rho = 0.25$	78.5%	65.5%	47.9%	52.0%	46.0%	51.0%	35.4%	43.3%
$\rho = 0.50$	84.5%	73.0%	63.7%	67.3%	60.8%	65.6%	45.0%	54.6%
$\rho = 0.75$	87.8%	76.6%	73.2%	75.9%	69.8%	73.9%	51.3%	61.1%
$\rho = 1.00$	90.0%	78.4%	81.0%	83.3%	76.8%	80.6%	55.5%	66.1%

Notes: All estimates in this table use the total employment share in each CZ as the outcome variable Y_i. The second row indicates the inference procedure employed to compute the share of the 30,000 simulated datasets for which we reject the null hypothesis $H_0 \colon \beta = 0$ using a 5% significance level. The first row indicates the clustering of sector-level shocks in *AKM* and *AKM0*. *Robust* is the Eicker-Huber-White standard error; *St-cluster* is the standard error that clusters CZs in the same state; *AKM* is the standard error in Remark 5 for independent shocks and in eq. (29) for clustering; *AKM0* is the test in Remark 6 for independent shocks and in the last sentence of Section 4.3.1 for clustering.

Specifically, we set the variances $\sigma_s^2 = \text{var}(\mathcal{X}_s \mid W)$ by taking draws from a uniform distribution with support $[2, 7]$. Thus, the mean variance of the sector-level shocks is equal to 5, as in the baseline placebo, but it now varies across sectors. We then simulate repeated samples of sector-level shocks where, in each simulation draw m, we take independent draws X_s^m from a normal distribution with zero mean and the randomly drawn variance σ_s^2, which is held constant across all placebo samples. Comparing the results for *Robust* and *St-cluster* to those in the first row of Table 1, and those for *AKM* and *AKM0* to those in the first row of Table 2, we observe that allowing for heteroscedasticity in the sector-level shocks does not affect our main results.

Finally, in panel D, we explore the impact of having the residuals of CZs that belong to the same state to be correlated for reasons other than their employment sectoral composition. Specifically, for each of the 30,000 placebo samples we build, we generate a random variable η_k for each state k, imposing the distribution $\eta_k \sim \mathcal{N}(0, 6)$. We then modify the outcomes Y_i by adding the random draw $\eta_{k(i)}$ to the actual change in employment rate, where $k(i)$ denotes the state that CZ i belongs to. Since *St-cluster* now captures part of the correlation structure in the residuals, the overrejection problem is less severe: it goes down from 38.3% to 30.1%. Importantly, as the validity of the inference procedures *AKM* and *AKM0* does not rely on any assumptions about the correlation structure of the residuals, their rejection rates are not affected.

Table 4 reports the results of a placebo exercise in which we account for shifters that are correlated across sectors. Specifically, instead of assuming that the sectoral shocks $\mathcal{X}_s^m$ are independent across s, we draw the random vector $(\mathcal{X}_1^m, \ldots, \mathcal{X}_{S-1}^m)$ from the distribution

$$(\mathcal{X}_1^m, \ldots, \mathcal{X}_{S-1}^m) \sim \mathcal{N}(0, \Sigma), \tag{34}$$

where Σ is a $(S-1) \times (S-1)$ covariance matrix with $\Sigma_{sk} = (1-\rho)\sigma\,\mathbb{I}\{s = k\} + \rho\sigma\,\mathbb{I}\{c(s) = c(k)\}$ and, for every s, $c(s)$ indicates the "cluster" that industry s belongs to. In panels A, B, and C, these clusters correspond to the 3-, 2-, and 1-digit SIC sector that industry s belongs to, respectively. As in the baseline, we set the shock for the non-manufacturing sector to zero, $\mathcal{X}_S = 0$.

Panel A of Table 4 shows that introducing correlation within 3-digit SIC sectors has a moderate effect on the rejection rates of both the traditional methods and the version of the *AKM* and *AKM0* methods that assume that the sectoral shocks are independent (described in Remarks 5 and 6). For all values of the correlation within 3-digit sectors, rejection rates close to 5% are obtained with the versions of *AKM* and *AKM0* that allow for arbitrary shock correlation within 2- and 3-digit aggregate sectors (described in Section 4.3.1). As shown in Panel B, the overrejection problem affecting both traditional inference procedures and the version of our novel procedures that assumes independence of sectoral shocks is more severe when these sectoral shocks are correlated at the 2-digit level. However, the last two columns of Table 4 show that, in this case, the versions of *AKM* and *AKM0* that allow for correlation in sectoral shocks at the 2-digit level perform very well, achieving a rejection rate close to 5%. 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 level aggregate sectors.

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

Remark 7. *In shift-share regressions, overrejection of typical standard error formulas is more severe when there is a small number of large sectors. In this case, the methods we provide significantly attenuate the overrejection problem, but may still overreject relative to the nominal significance level when the number of sectors is very small. Our methods perform well when the residuals have a state-level component and when the shocks are heteroscedastic. When the shifters are not independent across sectors, it is important to allow for clustering of the shifters at the appropriate level.*

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

In Online Appendix D.1, we investigate the consequences of violations of the assumption that the shifters $(\mathcal{X}_1, \ldots, \mathcal{X}_S)$ are independent of other sectoral shocks affecting the outcome variable of interest. We also consider the properties of two solutions to this problem: (i) the inclusion of regional controls as a proxy for sector-level unobserved shocks (discussed theoretically in Section 4.2), and (ii) the use of a shift-share instrumental variable constructed as a weighted average of exogenous sector-level shocks (discussed theoretically in Section 4.3.2).

Our simulations illustrate that confounding sector-level shocks introduce bias in the OLS estimator of the coefficient on the shift-share regressor of interest. In such cases, as discussed in Section 4.2, region-level controls eliminate the bias only if they are a perfect proxy for the sector-level confounding shock. Otherwise, an IV approach is needed. Our results also illustrate that, even when the estimator

$\hat{\beta}$ is consistent, traditional inference methods suffer from a severe overrejection problem and yield confidence intervals that are too small; in contrast, the inference procedures we propose yield the correct test size and confidence intervals with the right coverage.

5.4 Other extensions

Online Appendix D.2 explores how sensitive our results are 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 D.3 shows that rejection rates are very similar to those in Tables 1 and 2. Table D.4 reports the results of a placebo exercise based on a shift-share covariate with shifters that vary at the occupation level, using occupation employment shares in 1990 and randomly drawn shifters for 331 occupations. In this case, the overrejection problem of traditional methods is even more severe. AKM attenuates the problem, but still yields rejections rates higher than the nominal significance level. In contrast, AKM0 yields the correct test size.

6 Empirical applications

We now apply our inference procedures to three empirical applications. First, the study of the effect of Chinese competition on local labor market in Autor, Dorn and Hanson (2013). Second, the estimation of the elasticity of labor supply in Bartik (1991). Finally, the estimation of the impact of immigration on employment and wages across occupations and regional markets in the United States, as in the literature reviewed by Lewis and Peri (2015) and Dustmann, Schönberg and Stuhler (2016).

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

Autor, Dorn and Hanson (2013), henceforth referred to as ADH, explore the impact of exports from China on labor market outcomes across U.S. Commuting Zones. Specifically, they present IV estimates based on eq. (32), where Y_{i1} is the ten-year equivalent change in a labor-market outcome in CZ i in either 1990–2000 or 2000–2007, w_{is} is the CZ i employment share in the 4-digit SIC sector s in the initial year of the corresponding period (either 1990 or 2000), Y_{i2} is a weighted average of the change in sectoral U.S. imports from China normalized by U.S. total employment in the corresponding sector, and X_i is analogous to Y_{i2} with the only difference that, instead of using U.S. imports from China as shifters, it uses imports from China by other high-income countries. We use the data sources described in Section 2.1 and we include in all regression specifications the largest set of controls Z_i included in ADH; see, e.g., column (6) of Table 3 in ADH.

Table 5 reports 95% CIs computed using different methodologies for the specifications in Tables 5 to 7 in ADH. Panels A, B, and C present the IV, reduced-form and first-stage estimates, respectively. These correspond to $\hat{\alpha}$, $\hat{\beta}_{RF}$ and $\hat{\beta}_{FS}$, respectively, in the notation introduced in Section 4.3.2.

In all panels, state-clustered CIs are very similar to the heteroskedasticity-robust ones. This suggests that there is not much correlation in residuals within states. 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

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

	Change in the employment share			Change in avg. log weekly wage		
	All	Manuf.	Non-Manuf.	All	Manuf.	Non-Manuf.
Panel A: 2SLS Regression						
$\hat{\beta}$	-0.77	-0.60	-0.18	-0.76	0.15	-0.76
Robust	$[-1.10, -0.45]$	$[-0.78, -0.41]$	$[-0.47, 0.12]$	$[-1.23, -0.29]$	$[-0.81, 1.11]$	$[-1.27, -0.25]$
St-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	$[-0.71, -0.27]$	$[-0.48, -0.28]$	$[-0.31, 0.08]$	$[-0.80, -0.16]$	$[-0.50, 0.69]$	$[-0.83, -0.13]$
St-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			$[0.46, 0.80]$			
St-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; *St-cluster* is the standard error that clusters of CZs in the same state; *AKM* is the standard error in eq. (29) with 3-digit SIC clusters; *AKM0* is the confidence interval with 3-digit SIC clusters described in the last sentence of Section 4.3.1.

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 27% with the *AKM* procedure and by 68% with the *AKM0* procedure.[25] In light of the lack of impact of state-clustering on the 95% CI, the wider intervals implied by our inference procedures indicate that cross-region residual correlation is driven by similarity in sectoral compositions rather than by geographic proximity.

Panel B of Table 5 reports CIs for the reduced-form specification. In this case, the increase in the CI length is slightly larger: across outcomes, it increases on average by 53% for *AKM* and 130% for *AKM0*. The smaller relative increase in the CI length for the IV estimate $\hat{\alpha}$ relative to its increase for the reduced-form estimate $\hat{\beta}_{RF}$ is a consequence of the fact that all inference procedures yield very similar CIs for the first-stage estimate $\hat{\beta}_{FS}$, as reported in Panel C.

As discussed in Section 5, 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-

[25]The *AKM* and *AKM0* estimates reported in Table 5 account for correlation in the shifters across periods and across 4-digit SIC sectors included in the same 3-digit SIC sector. Table E.2 in Online Appendix E.1 shows that similar increases in the length of the 95% CIs are implied by *AKM* and *AKM0* when we assume that sectoral shifters are: (a) independent across 4-digit SIC sectors and periods; (b) independent across 4-digit SIC sectors but possibly correlated across periods.

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.[26] To investigate this claim, Table E.3 in Online Appendix E.1 reports the rejection rates implied by a placebo exercise analogous to that described Section 5 when the outcome variable in the placebo exercise is the same as that in the first-stage specification reported in Panel C of Table 5. Panel A in Table E.3 shows that, when no controls are included, traditional methods still suffer from severe overrejection problems and our methods yield the correct test size. However, as shown in Panels B and C in Table E.3, the problem is greatly attenuated when controlling for the instrumental variable and other controls used in ADH. This indicates that the instrumental variable and additional controls included in ADH soak most of the cross-CZ correlation in the ADH treatment variable.

Overall, Table 5 shows that, despite the wider confidence intervals obtained with our procedures, the qualitative conclusions in ADH with respect to the effect of U.S. imports from China on CZs labor market outcomes remain valid at usual significance levels. However, the increase in the length of the 95% confidence interval indicates that there is more uncertainty regarding the magnitude of the impact of Chinese import exposure on U.S. labor markets. In particular, the *AKM0* confidence interval is much wider than that based on state-clustered standard errors; furthermore, it is asymmetric around the point estimate, indicating that the negative impact of the China shock could have been two to three times larger than the effect implied by the point estimates.[27]

6.2 Estimation of labor supply elasticity

In our second application, we estimate the labor supply elasticity ϕ using the following estimating equation:

$$\Delta \log E_i = \phi \Delta \log \omega_i + Z_i \delta + \epsilon_i, \tag{35}$$

where $\Delta \log E_i$ denotes changes in the employment rate in CZ i, $\Delta \log \omega_i$ denotes changes in nominal wages in i, Z_i is the same vector of controls used in Section 6.1 (i.e. the vector of controls listed in column (6) of Table 3 of ADH), and ϵ_i is the regression residual. We use the data sources described in Section 2.1 to measure the outcome, covariate and controls in eq. (35).

As illustrated through the model in Section 3, the residual ϵ_i accounts for changes in local supply shocks, $\Delta \log \nu_i$, not controlled for by the vector Z_i. These unobserved supply shocks will impact changes in both local average wages and employment rates; thus, $\Delta \log \omega_i$ and ϵ_i will be correlated and the OLS estimator of ϕ in eq. (35) will generally be biased. To circumvent this problem, a popular approach is the use of shift-share instrumental variables. In this section, we implement this strategy with two different sector-level shifters: (i) the national employment growth, as in Bartik (1991); and (ii) the increase in imports from China by a set of high-income countries that does not include the

[26] 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 no within-state correlation left in the residuals.

[27] It follows from Remark 6 (see the expression for the quantity A) that the asymmetry 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 CIs in Table 5 thus reflect differences in their finite-sample properties. This notwithstanding, the placebo exercise presented in Section 5 shows that both inference procedures yield close to correct rejection rates in a sample with the same number of regions and sectors as used in ADH.

Table 6: Estimation of labor supply elasticity

	First-Stage $\Delta \log \omega_i$	Reduced-Form $\Delta \log E_i$	2SLS $\Delta \log E_i$
Panel A: Bartik IV			
	29.66	33.85	1.14
Robust	$[20.31, 39.01]$	$[24.00, 43.70]$	$[0.90, 1.38]$
St-cluster	$[17.62, 41.71]$	$[20.70, 47.00]$	$[0.84, 1.44]$
AKM	$[20.50, 38.83]$	$[22.75, 44.96]$	$[0.86, 1.42]$
AKM0	$[17.02, 38.73]$	$[21.33, 47.29]$	$[0.89, 1.71]$
Panel B: ADH IV			
	-0.48	-0.72	1.49
Robust	$[-0.80, -0.16]$	$[-1.04, -0.39]$	$[0.79, 2.19]$
St-cluster	$[-0.78, -0.18]$	$[-0.93, -0.50]$	$[0.78, 2.21]$
AKM	$[-0.87, -0.09]$	$[-1.19, -0.24]$	$[0.63, 2.35]$
AKM0	$[-1.26, -0.12]$	$[-1.85, -0.35]$	$[0.89, 4.85]$

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 in square brackets. *Robust* is the Eicker-Huber-White standard error; *St-cluster* is the standard error that clusters of CZs in the same state; *AKM* is the standard error in eq. (29) with 3-digit SIC clusters; *AKM0* is the confidence interval with 3-digit SIC clusters described in the last sentence of Section 4.3.1.

United States, as in Autor, Dorn and Hanson (2013).[28]

Table 6 presents first-stage, reduced-form and IV estimates associated to the estimation of the parameter ϕ in eq. (35). Panels A and B report results using the Bartik (1991) instrumental variable and the ADH instrumental variable, respectively. In both cases, the estimates of ϕ are similar: 1.14 in Panel A and 1.49 in Panel B. In Panel A, our proposed CIs are wider than heteroskedasticity-robust CIs, but tighter than state-clustered CIs. For Panel B, the *AKM* and *AKM0* CIs are respectively 20% and 250% wider than those obtained with state-clustered standard errors. As discussed in Section 6.1, such differences are related to the shift-share component of the residuals. Our results suggest that the shift-share IV that exploits national employment growth as shifter absorbs the bulk of this component, leaving little correlation left for our inference procedures to correct. In contrast, the ADH shift-share IV absorbs a lower fraction of the shift-share component of the residuals, implying that our procedure has a larger impact on the length of the 95% confidence interval.

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

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

$$\Delta Y_{oit} = \beta \Delta ImmShare_{oit} + Z'_{oit}\delta + \epsilon_{oit}, \tag{36}$$

[28] For simplicity, in the case of Bartik (1991), we assume that the national employment growth itself is as good as randomly assigned, instead of thinking of it as a proxy for a randomly assigned national-level shock. The latter would create additional consistency and inference issues. See Goldsmith-Pinkham, Sorkin and Swift (2018) and Borusyak, Hull and Jaravel (2018) for a discussion of some of these consistency issues; the discussion in Section 4.3.2 may be extended to address the corresponding inference issues.

Table 7: Effect of immigration on occupations and commuting zones

	Change in log native employment	Change in avg. log weekly wage		
		All workers	High-Skill	Low-Skill
	(1)	(2)	(3)	(4)
Panel A: 2SLS Regression				
$\hat{\beta}$	-0.74	-0.07	0.14	-0.24
Robust	$[-1.05, -0.43]$	$[-0.22, 0.09]$	$[0.00, 0.29]$	$[-0.42, -0.06]$
St-cluster	$[-1.16, -0.31]$	$[-0.34, 0.20]$	$[-0.06, 0.35]$	$[-0.52, 0.04]$
AKM	$[-1.17, -0.31]$	$[-0.38, 0.25]$	$[-0.13, 0.42]$	$[-0.60, 0.12]$
AKM0	$[-1.49, -0.11]$	$[-0.39, 0.71]$	$[-0.12, 0.88]$	$[-0.71, 0.45]$
Panel B: OLS Reduced-Form Regression				
$\hat{\beta}$	-0.19	-0.02	0.04	-0.06
Robust	$[-0.27, -0.11]$	$[-0.05, 0.02]$	$[-0.00, 0.08]$	$[-0.10, -0.02]$
St-cluster	$[-0.33, -0.05]$	$[-0.08, 0.05]$	$[-0.03, 0.10]$	$[-0.11, -0.01]$
AKM	$[-0.37, -0.01]$	$[-0.09, 0.06]$	$[-0.05, 0.12]$	$[-0.13, 0.01]$
AKM0	$[-0.88, -0.02]$	$[-0.07, 0.39]$	$[-0.02, 0.47]$	$[-0.12, 0.27]$
Panel C: 2SLS First-Stage				
$\hat{\beta}$		0.26		
Robust		$[0.19, 0.32]$		
St-cluster		$[0.16, 0.36]$		
AKM		$[0.13, 0.38]$		
AKM0		$[0.13, 0.74]$		

Notes: $N = 108,300$ (722 CZs $\times$ 50 occupations $\times$ 3 time periods). Models are weighted by start of period occupation-region share of national population. All regressions include occupation and period dummies. 95% confidence intervals in square brackets. *Robust* is the Eicker-Huber-White standard error; *St-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.

where, for occupation o of CZ i between year t and $t-10$, ΔY_{oit} is the change in a labor market outcome for native workers and $\Delta ImmShare_{oit}$ is the change in the share of immigrants in total employment. In our application, Z_{oit} is a control vector that includes occupation and period fixed effects.

The OLS estimator of β in eq. (36) is likely to be inconsistent because the sorting decision of immigrants across occupations and regions is likely driven by the same demand shocks affecting the sorting decision of native workers. Since Card (2001), a large literature addresses these concerns exploiting shift-share instruments built using data on changes in the national stock of immigrants from different origin countries. Specifically, we use a version of this instrument that takes the form

$$\Delta X_{iot} \equiv \sum_j ImmShare_{oi1980,j} \frac{Imm_{j,t} - Imm_{j,t-10}}{Imm_{j,1980}}, \tag{37}$$

where, for occupation o in CZ i, $ImmShare_{oit_0,j}$ is the share of immigrants of origin country j in total employment in 1980, and $Imm_{j,t}$ is the total number of immigrants of origin j in the U.S. in year t.

We build a dataset for 50 occupations, 722 U.S. CZs, 57 origin countries, and three periods.[29] The

[29]See Appendix F of Burstein et al. (2018b) for a list of occupations and Table E.4 of Online Appendix E.2 for the list of

information on employment and average wages are from the Census Integrated Public Use Micro Samples for 1970-2000 and the American Community Survey for 2008-2012. For each variable, we construct separate measures for low-skilled workers, defined as those with at most a high school diploma, and high-skilled workers, defined as those with at least one year of college.

Table 7 reports the results. Column (1) indicates that an increase in the immigrant share is associated a decline in native employment across occupations and regions. For all inference methods, the estimated effect is significant at 5%. However, our proposed methods yield wider confidence intervals, which are as much as 123% wider than those obtained with commonly used inference methods. Columns (2)–(4) indicate that, for wage outcomes, the increase in the length of confidence intervals is even stronger: on average across these three columns, the AKM and $AKM0$ CIs are respectively 27% and 119% wider than that obtained with state-clustered standard errors.[30]

7 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 covariates. 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 sectoral shifters affecting the outcome. Such a correlation is ignored 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 the 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.

origin countries. We consider the periods 1980-1990, 1990-2000 and 2000-2012.

[30]Table E.5 in Online Appendix E.2 reports results of the estimation of eq. (36) with data aggregated at the CZ-level. In this case, we obtain similar qualitative results, but confidence intervals are slightly wider.

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*, 2018.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc J. Melitz, "The impact of exports on innovation: Theory and evidence," Technical Report 2018.

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, James E., "A Theoretical Foundation for the Gravity Equation," *The American Economic Review*, 1979, *69* (1), 106–116.

— **and Eric van Wincoop**, "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*, 2003, *93* (1), 170–192.

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.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Armington, Paul S., "A Theory of Demand for Products Distinguished by Place of Production," *Staff Papers*, 1969, *16* (1), 159–178.

Autor, David H. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, August 2013, *103* (5), 1553–97.

— **,** — **, 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 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," *University of Zurich mimeo*, 2017.

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.

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

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

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel, "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *University of California Los Angeles mimeo*, 2018.

—, 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.

— and John Dinardo, "Do Immigrant Inflows Lead to Native Outflows?," *American Economic Review*, 2000, *90* (2), 360–367.

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.*

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*, 2017, *forthcoming.*

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*, 2017.

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," *Princeton University mimeo*, 2015.

Hakobyan, Sushanik and John McLaren, "Looking for Local Labor Market Effects of NAFTA," *Review of Economics and Statistics*, 2016, *98* (4), 728–741.

Huber, Kilian, "Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties," *American Economic Review*, March 2018, *108* (3), 868–98.

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.

Jaeger, David A., Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," *CUNY University mimeo*, 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.

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*, 2017.

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.

Monras, Joan, "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," *CEMFI mimeo*, 2015.

Pierce, Justin R. and Peter K. Schott, "Trade Liberalization and Mortality: Evidence from US Counties," *Yale University mimeo*, 2017.

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

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 Microfoundation for stylized economic model in Section 3

Appendices A.1 and A.2 provide a microfoundation for the stylized economic model summarized in Section 3.1. Appendix A.3 performs an analysis analogous to that in Section 3.2 for the case of the microfoundation described in Appendices A.1 and A.2.

A.1 Environment

We consider a model with multiple sectors $s = 1, \ldots, S$ and multiple regions $i, j = 1, \ldots, J$. Regions are partitioned into countries indexed by c, and we denote the set of regions located in a country c by J_c. Region i has a population of M_i individuals who cannot move across regions.

Production. Each sector s in region i has a representative firm that produces a differentiated good. The quantity Q_{is} produced by sector s in region i is produced using labor with productivity A_{is},

$$Q_{is} = A_{is} L_{is}, \tag{A.1}$$

where L_{is} denotes the number of workers employed by the representative firm in this sector-region pair. Regions thus differ in terms of their sector-specific productivity A_{is}.

Preferences for consumption goods. Every individual has identical nested preferences over the sector- and region-specific differentiated goods. Specifically, we assume that individuals have Cobb-Douglas preferences over sectoral composite goods,

$$C_j = \prod_{s=1}^{S} \left(C_{js}\right)^{\gamma_s}, \tag{A.2}$$

where C_j is the utility level of a worker located in region j that obtains utility C_{js} from consuming goods in sector s, and C_{js} is a CES aggregator of the sector s goods produced in different regions:

$$C_{js} = \left[\sum_{i=1}^{J} \left(c_{ijs}\right)^{\frac{\sigma_s - 1}{\sigma_s}}\right]^{\frac{\sigma_s}{\sigma_s - 1}}, \qquad \sigma_s \in (1, \infty), \tag{A.3}$$

where c_{ijs} denotes the consumption in region j of the sector s good produced in region i. This preference structure has been previously used in Armington (1969), Anderson (1979) and multiple papers since (e.g. Anderson and van Wincoop, 2003; Arkolakis, Costinot and Rodríguez-Clare, 2012).

Preferences for sectors and non-employment. Workers have the choice of being employed in one of the sectors $s = 1, \ldots, S$ of the economy or opting for non-employment, which we index as $s = 0$. Conditional on being employed, all workers have identical homogeneous preferences over their sector of employment, but workers differ in their preferences for non-employment. Specifically, conditional on obtaining utility C_j from the consumption of goods, the utility of a worker ι living in region j is

$$U(\iota \mid C_j) = \begin{cases} u(\iota)C_j & \text{if employed in any sector } s = 1, \ldots, S, \\ C_j & \text{if not employed } (s = 0). \end{cases} \tag{A.4}$$

We assume that $u(\iota)$ is i.i.d. across individuals ι according to a Pareto distribution with scale parameter ν_i and shape parameter ϕ, so that the cumulative distribution function of $u(\iota)$ is given by

$$F_u(u) = 1 - \left(\frac{u}{\nu_i}\right)^{-\phi}, \qquad u \in [\nu_i, \infty), \qquad \phi > 1. \tag{A.5}$$

If a worker living in region j chooses to be employed, she will earn wage ω_j (as workers are indifferent about the sector of employment and can move freely across sectors, wages must be equalized across sectors in equilibrium). If a worker chooses to not be employed, she receives a benefit b_j.[31] We denote the total number of employed workers in region j by L_j, and the employment rate in j as $E_j \equiv L_j/M_j$.

Market structure. Goods and labor markets are perfectly competitive.

Trade costs. We assume that there are no trade costs, which implies that the equilibrium price of the good produced in a region is the same in every other region; i.e. $p_{ijs} = p_{is}$ for $j = 1, \ldots, J$. Thus, for every sector s there is a composite sectoral good that has identical price P_s in all regions; i.e.

$$(P_s)^{1-\sigma_s} = \sum_{s=1}^{S} (p_{is})^{1-\sigma_s}, \tag{A.6}$$

and the final good's price is $P = \prod_{s=1}^{S} (P_s)^{\gamma_s}$.

A.2 Equilibrium

We now characterize the equilibrium wage ω_j and total employment L_j of all regions $j = 1, \ldots, J$.

Consumption. We first solve the expenditure minimization problem of an individual residing in region j. Given the sector-level utility in eq. (A.3) and the condition that $p_{ijs} = p_{is}$ for $j = 1, \ldots, J$, all regions j have identical spending shares x_{is} on goods from region i, given by

$$x_{is} = \left(\frac{p_{is}}{P_s}\right)^{1-\sigma_s}. \tag{A.7}$$

[31] We assume that benefits are paid by a national government that imposes a flat tax χ_c on all income earned in country c. The budget constraint of the government is thus $\sum_{j \in J_c} \{\chi_c(\omega_j E_j + b_j(1-E_j))M_j\} = \sum_{j \in J_c} \{b_j(1-E_j)M_j\}$. Alternatively, we could think of the option $s = 0$ as home production and assume that workers that opt for home production in region j obtain b_j units of the final good, which they consume. This alternative model is isomorphic to that in the main text.

Labor supply. Every worker maximizes the utility function in eq. (A.4) in order to decide whether to be employed. Consequently, conditional on the wage ω_i and the non-employment benefit b_i, the employment rate in region i is $E_i = \Pr[u_i(\iota)\omega_i > b_i] = 1 - \Pr[u_i(\iota) < b_i/\omega_i]$. It therefore follows from eq. (A.5) that

$$L_i = v_i \omega_i^{\phi}, \qquad v_i \equiv M_i \left(\nu_i / b_i\right)^{\phi}. \tag{A.8}$$

Note that this labor supply equation is analogous to that in eq. (5).

Producer's problem. In perfect competition, firms must earn zero profits and, therefore,

$$p_{is} = \frac{\omega_i}{A_{is}}. \tag{A.9}$$

Goods market clearing. Given that labor is the only factor of production and firms earn no profits, the income of all individuals living in region i is $W_i \equiv \sum_s \omega_i L_{is}$, and world income is $W \equiv \sum_i W_i$. We normalize world income to one, $W = 1$. Given preferences in eq. (A.2), all individuals spend a share γ_s of their income on sector s, so that world demand for the differentiated good s produced in region i is $x_{is}\gamma_s$. Goods market clearing requires world demand for good s produced in region i to equal total revenue of the representative firm operating in sector s in region i, $\omega_i L_{is}$. Thus, using the expression in eq. (A.7), we obtain

$$L_{is} = (\omega_i)^{-\sigma_s} \left(A_{is} P_s\right)^{\sigma_s - 1} \gamma_s. \tag{A.10}$$

Note that this labor demand equation is analogous to that in eq. (3), with the region- and sector-specific demand shifter D_{is} defined as

$$D_{is} = \left(A_{is} P_s\right)^{\sigma_s - 1} \gamma_s.$$

If, without loss of generality, we split the region- and sector-specific productivity A_{is} into a sector component A_s and a residual $\tilde{A}_{is}$, $A_{is} = A_s \tilde{A}_{is}$, and we further consider P_s as our sectoral shock of interest, we can decompose D_{is} as in eq. (4), with

$$\chi_s = P_s, \tag{A.11}$$

$$\rho_s = \sigma_s - 1, \tag{A.12}$$

$$\mu_s = (A_s)^{\sigma_s - 1} \gamma_s, \tag{A.13}$$

$$\eta_{is} = (\tilde{A}_{is})^{\sigma_s - 1}. \tag{A.14}$$

Labor market clearing. Given the sector- and region-specific labor demand in eq. (A.10), total labor demand in region i is

$$L_i = \sum_{s=1}^{S} (\omega_i)^{-\sigma_s} \left(A_{is} P_s\right)^{\sigma_s - 1} \gamma_s. \tag{A.15}$$

Labor market clearing requires labor supply in eq. (A.8) to equal labor demand in eq. (A.15):

$$v_i(\omega_i)^{\phi} = \sum_{s=1}^{S} (\omega_i)^{-\sigma_s} \left(A_{is}P_s\right)^{\sigma_s - 1} \gamma_s. \tag{A.16}$$

Equilibrium. Given technology parameters $\{A_{is}\}_{i=1,s=1}^{J,S}$, preference parameters $\{(\sigma_s, \gamma_s)\}_{s=1}^{S}$, labor supply parameters $\{v_i\}_{i=1}^{J}$, and normalizing world income to equal 1, $W = 1$, we can use eqs. (A.6), (A.9) and (A.16) to solve for the equilibrium wage in every world region, $\{\omega_i\}_{i=1}^{J}$, the equilibrium price of every sector-region specific good $\{p_{is}\}_{i=1,s=1}^{J,S}$, and the sectoral price indices $\{P_s\}_{s=1}^{S}$. Given these equilibrium wages and sectoral price indices, we can use eq. (A.15) to solve for the equilibrium level of employment in every region, $\{L_i\}_{i=J}^{J}$.

A.3 Labor market impact of sectoral shocks

As in Section 3.2, we assume that, in every period, our model characterizes the labor market equilibrium in every region of the world economy. Across periods, we assume that the parameters $\{\sigma_s\}_{s=1}^{S}$ and ϕ are fixed and that all changes in the labor market outcomes $\{\omega_i, L_i\}_{i=1}^{J}$ are generated by changes in technology $\{A_{is}\}_{i=1,s=1}^{J,S}$, sectoral preferences $\{\gamma_s\}_{s=1}^{S}$ and labor supply parameters $\{v_i\}_{i=1}^{J}$.

We focus in this section on understanding how changes in these exogenous parameters affect the labor market equilibrium in a set of "small" regions whose share in world output is approximately zero for all sectors, i.e., $x_{is} \approx 0$ for $s = 1, \ldots, S$, with x_{is} defined in eq. (A.7). We assume that all small regions of interest belong to the same country c and that they correspond to the set N regions discussed in Section 3.1.

As illustrated in Online Appendix C.1, the "small region" assumption applied to all N regions of interest implies that the sectoral price index P_s of every sector s will not depend on the technology and labor supply parameters of these N regions; i.e., $\{P_s\}_{s=1}^{S}$ does not depend on $\{A_{is}\}_{s=1,i\in J_c}^{S}$ and $\{v_i\}_{i\in J_c}$. Thus, from the perspective of any one of these regions, changes in sectoral prices operate as exogenous shocks. Furthermore, as illustrated in eqs. (A.15) and (A.16), these sectoral prices mediate the impact of all foreign technology and labor supply shocks on the labor market equilibrium of every region in country c.

Consequently, across periods, our microfounded model implies that the changes in labor market outcomes in all N regions of country c, $\{\omega_i, L_i\}_{i\in J_c}$, are generated by changes in sectoral prices $\{P_s\}_{s=1}^{S}$, changes in an aggregate of all other sectoral shocks, $\{(A_s)^{\sigma_s - 1}\gamma_s\}_{s=1}^{S}$, changes in labor supply parameters of all regions in country c, $\{v_i\}_{i\in J_c}$, and changes in the sector- and region-specific technology parameters, $\{(\tilde{A}_{is})^{\sigma_s - 1}\}_{i\in J_c, s=1}^{S}$.

Isomorphism. Up to a first-order approximation around the initial equilibrium, eqs. (A.15) and (A.16) imply that

$$\hat{L}_i = \sum_{s=1}^{S} l_{is}^{0} \left[\beta_{is}\hat{P}_s + \lambda_i((\sigma_s - 1)\hat{A}_s + \hat{\gamma}_s) + \lambda_i((\sigma_s - 1)\hat{\tilde{A}}_{is})\right] + (1 - \lambda_i)\,\hat{v}_i, \tag{A.17}$$

with $\beta_{is} = (\sigma_s - 1)\lambda_i$ and λ_i defined as in Section 3.2; i.e. $\lambda_i \equiv \phi\left[\phi + \sum_s l_{is}^0 \sigma_s\right]^{-1}$. Given the equivalences in eqs. (A.11) to (A.14), the expression in eq. (A.17) is identical to that in eq. (8). Consequently, the environment described in Appendices A.1 and A.2 does indeed provide a microfoundation for the equilibrium relationship in eq. (8).

B 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. Before proving these results in Appendices B.3, B.3 and B.5, we collect some auxiliary Lemmata used in the proofs in Appendix B.2, and collect the assumptions that we impose on the DGP in Appendix B.1. Finally, Appendix B.6 discusses inference when the effects β_{is} are heterogeneous. Throughout this appendix, we use the following notation. 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 $A_S \leq CB_S$. We denote the σ-field generated by $(Y(0), B, W, U, \mathcal{Z})$ by $\mathcal{F}_0 = \sigma(Y(0), B, W, U, \mathcal{Z})$. Define $\overline{w}_{st} = \sum_{i=1}^N w_{is}w_{it}$, $\tilde{\mathcal{X}}_s = \mathcal{X}_s - \mathcal{Z}_s'\gamma$, and $\sigma_s^2 = \operatorname{var}(\mathcal{X}_s \mid \mathcal{F}_0) = \operatorname{var}(\mathcal{X}_s \mid \mathcal{Z}, W)$. Finally, let $n_s = \sum_{i=1}^N w_{is}$.

B.1 Assumptions

We first list and discuss the assumptions needed for the results in Section 4.1. We impose some regularity conditions on the DGP for $(Y(0), B, W, \mathcal{X})$ that generate the observed data (Y, X, W). Unless stated otherwise, all limits are taken as $S \to \infty$.

Assumption 1.

(i) $\{(Y(0), B, W, \mathcal{X}) \in \mathbb{R}^{N_S} \times \mathbb{R}^{N_S \times S} \times \mathbb{R}^{N_S \times S} \times \mathbb{R}^S\}_{S=1}^\infty$ is a triangular array of random variables with $N = N_S \to \infty$ as $S \to \infty$ that satisfies eq. (15), and $\sum_{i=1}^N E[Y_i(0)] = 0$. The observed data consists of the tuple (Y, X, W), with $Y_i = Y_i(\mathcal{X}_1, \dots, \mathcal{X}_S)$, such that eq. (9) holds.

(ii) Conditional on W, the shocks $\mathcal{X}_1, \dots, \mathcal{X}_S$ are mean zero, independent across s, with fourth moments that exist and are bounded uniformly over s.

(iii) $\frac{1}{N}\sum_{i=1}^N E[X_i^2 \mid W] = \frac{1}{N}\sum_{i=1}^N \sum_{s=1}^S \operatorname{var}(\mathcal{X}_s \mid W) w_{is}^2$ converges in probability to a strictly positive non-random limit.

Assumption 2.

(i) Conditional on W, the second moments of $Y_i(0)$ exist, and are bounded uniformly over i. The support of β_{is} is bounded.

(ii) $\max_s n_s/N \to 0$, where $n_s = \sum_{s=1}^S w_{is}$ denotes the total share of sector s.

By modeling the data as a triangular array, Assumption 1(i) allows the distribution of the data to change with the sample size.[32] The assumption that $Y_i(0)$ and $\mathcal{X}_s$ are mean zero is made to simplify

[32] 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 $\{(Y_{iS}(0), \beta_{isS}, w_{isS}, \mathcal{X}_{sS}) \colon i = 1, \dots, N_S, s = 1, \dots, S\}$.

the exposition in this section by allowing us to drop the intercept from the regression of Y_i on X_i, and is relaxed in Section 4.2. Assumption 1(iii) 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. The bounded support condition on β_{is} in part (i) of Assumption 2 is made to keep the proofs simple and can be relaxed.

For the estimator in eq. (11) to be asymptotically normal, we need to strengthen Assumption 1(ii) and Assumption 2 slightly:

Assumption 3.

(i) $\max_s n_s^2 / \sum_{t=1}^S n_t^2 \to 0$.

(ii) Conditional on W, the eighth moments of $\mathcal{X}_s$ are bounded uniformly over s, and the fourth moments of $Y_i(0)$ are bounded uniformly over i.

Part (i) ensures that the contribution of each sector to the asymptotic variance, which, according to the standard error formula below, is of the order $O(n_s^2)$, is asymptotically negligible. For instance, while the estimator $\hat{\beta}$ is consistent for β when the largest sector share is of the order $O(N/\sqrt{S})$ and the remaining sector shares are of the order $O(N/S)$, Assumption 3 rules this case out; $\hat{\beta}$ will not generally be asymptotically normal in this case due to failure of the Lindeberg condition.

Next, we state and discuss the assumptions needed for the results in Section 4.2. The next assumption generalizes Assumption 1 to allow for controls:

Assumption 4.

(i) $\{(Y(0), B, W, U, \mathcal{X}, \mathcal{Z}) \in \mathbb{R}^{N_S} \times \mathbb{R}^{N_S \times S} \times \mathbb{R}^{N_S \times S} \times \mathbb{R}^{N_S \times K} \times \mathbb{R}^{S} \times \mathbb{R}^{S \times K}\}_{S=1}^{\infty}$ is a triangular array of random variables with $N = N_S \to \infty$ as $S \to \infty$ that satisfies eqs. (21) and (22). The observed data consists of the tuple (Y, X, Z, W), with $Y_i = Y_i(\mathcal{X}_1, \ldots, \mathcal{X}_S)$, such that eqs. (9) and (20) hold.

(ii) Conditional on W and $\mathcal{Z}$, the shocks $\mathcal{X}_1, \ldots, \mathcal{X}_S$ are independent across s, with fourth moments that exist and are bounded uniformly over s.

(iii) $\frac{1}{N}\sum_{i=1}^N \sum_{s=1}^S \operatorname{var}(\mathcal{X}_s \mid W, \mathcal{Z}) w_{is}^2$ converges in probability to a strictly positive non-random limit, and $Z'Z/N$ converges in probability to a positive definite non-random limit.

(iv) Conditional on W, the second moments of U_i and $\mathcal{Z}_s$ exist and are bounded uniformly over i and s.

Parts (i), (ii) and (iii) are straightforward generalizations of parts (i), (ii) and (iii) of Assumption 1. Part (iv) imposes very mild restrictions on U and $\mathcal{Z}$.

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 5.

(i) Conditional on W, the fourth moments of $\mathcal{Z}_s$, and U_i exist and are bounded uniformly over s and i.

(ii) $\frac{N}{\sqrt{\sum_s n_s^2}}(\check{\delta} - \delta) = O_p(1)$

Part (i) strengthens Assumption 4(iv). Part (ii) is a high-level assumption that implies $\hat{\delta}$ converges at least as fast as $\hat{\beta}$; otherwise the error in estimating δ could dominate the asymptotic variance of β.

B.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 bounded uniformly over S and s. Then $E[|A_{Si}|^{\eta} \mid W]$ exists and is bounded uniformly over S and i.*

Proof. By Hölder's inequality,

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

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 bounded uniformly over S and i. Then $N^{-2}\sum_{s=1}^{S} E\big[(\sum_{i=1}^{N} w_{is}A_{Si})^2 \mid W\big] \to 0$, provided Assumption 2(ii) holds.*

Proof. By Cauchy-Schwarz inequality,

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

The result follows from the fact that $N^{-2}\sum_{s=1}^{S} n_s^2 \leq \max_s n_s/N$, which converges to zero by Assumption 2(ii). □

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 such that $E[A_{Si}^4 \mid W]$, $E[B_{Si}^4 \mid W]$, and $E[\mathcal{A}_{Ss}^2 \mid W]$ exist and are bounded uniformly over S, i and s. Then $(\sum_s n_s^2)^{-1}\sum_{i,j,s} w_{is}w_{js}A_{Si}B_{Sj}\mathcal{A}_{Ss} = O_p(1)$.*

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

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

The result then follows by Markov inequality and the dominated convergence theorem. □

B.3 Proof of Proposition 3

Let E_W denote expectation conditional on W. We first show that

$$\frac{1}{N}X'Z = \frac{1}{N}\sum_{i,s} w_{is}\mathcal{Z}_s'\gamma Z_i + o_p(1) \tag{B.1}$$

$$\frac{1}{N}X'X = \frac{1}{N}\sum_{s}\sigma_s^2\overline{w}_{ss} + \frac{1}{N}\sum_{s,t}\mathcal{Z}_s'\gamma\mathcal{Z}_t'\gamma\overline{w}_{st} + o_p(1) \tag{B.2}$$

$$\frac{1}{N}Z'Y = \frac{1}{N}\sum_{i} Z_iY_i(0) + \frac{1}{N}\sum_{i,t} Z_iw_{it}\mathcal{Z}_t'\gamma\beta_{it} + o_p(1) \tag{B.3}$$

$$\frac{1}{N}X'Y = \frac{1}{N}\sum_{i,s,t} w_{is}w_{it}(\mathcal{Z}_s'\gamma)(\mathcal{Z}_t'\gamma)\beta_{it} + \frac{1}{N}\sum_{i,s} w_{is}^2\sigma_s^2\beta_{is} + \frac{1}{N}\sum_{i,s} w_{is}(\mathcal{Z}_s'\gamma)Y_i(0) + o_p(1). \tag{B.4}$$

Consider (B.1). We have

$$\frac{1}{N}X'Z = \frac{1}{N}\sum_{s}\mathcal{X}_s\sum_{i} w_{is}Z_i = \frac{1}{N}\sum_{s}\tilde{\mathcal{X}}_s\sum_{i} w_{is}Z_i + \frac{1}{N}\sum_{i,s} w_{is}\mathcal{Z}_s'\gamma Z_i.$$

It therefore suffices to show that

$$\frac{1}{N}\sum_{s}\tilde{\mathcal{X}}_s\sum_{i} w_{is}Z_i = o_p(1). \tag{B.5}$$

The left-hand side has mean zero conditional on W, with the variance of the kth row given by

$$\operatorname{var}\left(\frac{1}{N}\sum_{i,s} w_{is}\tilde{\mathcal{X}}_sZ_{ik} \mid W\right) = \frac{1}{N^2}\sum_{s} E_W\sigma_s^2\left(\sum_{i} w_{is}Z_{ik}\right)^2 \preceq \frac{1}{N^2}\sum_{s} E_W\left(\sum_{i} w_{is}Z_{ik}\right)^2.$$

By Lemma 1, Assumption 4(iv), and the C_r-inequality, $E_W[Z_{ik}^2] = E_W[(\sum_s w_{is}\mathcal{Z}_{sk} + U_{ik})^2]$ is bounded, so that by Lemma 2, the right-hand side converges to zero. Equation (B.5) then follows by Markov inequality and the dominated convergence theorem.

Next, consider eq. (B.2). We have

$$\frac{1}{N}X'X = \frac{1}{N}\sum_{i,s,t}\mathcal{X}_s\mathcal{X}_tw_{is}w_{it} = \frac{2}{N}\sum_{s<t}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t\overline{w}_{st} + \frac{1}{N}\sum_{i,s}(\mathcal{X}_s^2 - E[\mathcal{X}_s^2 \mid \mathcal{Z}_s, W])w_{is}^2$$
$$+ \frac{2}{N}\sum_{s\neq t}\mathcal{Z}_s'\gamma\tilde{\mathcal{X}}_t\overline{w}_{st} + \frac{1}{N}\sum_{s}\sigma_s^2\overline{w}_{ss} + \frac{1}{N}\sum_{s,t}\mathcal{Z}_s'\gamma\mathcal{Z}_t'\gamma\overline{w}_{st}. \tag{B.6}$$

We will show that the first three summands are of the order $o_p(1)$. All three summands are mean zero since they are mean zero conditional on $\mathcal{F}_0$, so by Markov inequality and the dominated convergence theorem, it suffices to show that their variances, conditional on W, converge to zero. To that end,

$$\operatorname{var}\left(\frac{2}{N}\sum_{s<t}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t\overline{w}_{st} \mid W\right) = \frac{4}{N^2}\sum_{s<t} E_W[\sigma_s^2\sigma_t^2]\overline{w}_{st}^2 \preceq \frac{1}{N^2}\sum_{s,t}\overline{w}_{st}^2$$

$$\leq \frac{1}{N^2} \sum_{i,j,s} w_{is} w_{js} = \frac{1}{N^2} \sum_s n_s^2 \to 0. \quad \text{(B.7)}$$

where the last inequality follows from $\sum_s w_{is} w_{js} \leq \sum_s w_{is} = 1$, and the convergence to 0 follows by Assumption 2(ii). The variance of the second summand can be bounded by

$$\operatorname{var}\left(\frac{1}{N} \sum_{i,s} (\mathcal{X}_s^2 - E[\mathcal{X}_s^2 \mid \mathcal{Z}_s, W]) w_{is}^2 \mid W\right) \preceq \frac{1}{N^2} \sum_s \left(\sum_i w_{is}^2\right)^2 \leq \frac{1}{N^2} \sum_s n_s^2,$$

which converges to zero by Assumption 2(ii). Finally, variance of the third summand in eq. (B.6) can be bounded by

$$\operatorname{var}\left(\frac{2}{N} \sum_{i,s \neq t} \mathcal{Z}_s' \gamma \tilde{\mathcal{X}}_t w_{is} w_{it} \mid W\right) \leq \frac{4}{N^2} \sum_t E_W \sigma_t^2 \left(\sum_{s,i} |\mathcal{Z}_s' \gamma| w_{is} w_{it}\right)^2$$

$$\preceq \frac{1}{N^2} \sum_s E_W \left(\sum_i w_{is} \sum_t w_{it} |\mathcal{Z}_t' \gamma|\right)^2.$$

By Lemma 1, the second moment of $\sum_t w_{it} |\mathcal{Z}_t' \gamma|$ is bounded, so by Lemma 2, the right-hand side converges to zero.

Next, consider eq. (B.3). We can decompose

$$\frac{1}{N} Z' Y = \frac{1}{N} \sum_{i,s} Z_i w_{is} \tilde{\mathcal{X}}_s \beta_{is} + \frac{1}{N} \sum_i Z_i Y_i(0) + \frac{1}{N} \sum_{i,t} Z_i w_{it} \mathcal{Z}_t' \gamma \beta_{it}.$$

We will show that the first summand is $o_p(1)$. Since it has mean zero, by Markov inequality, it suffices to show that the variance of each row k conditional on W converges to zero. Now,

$$\operatorname{var}\left(\frac{1}{N} \sum_{i,t} Z_{ik} w_{it} \tilde{\mathcal{X}}_t \beta_{it} \mid W\right) = \frac{1}{N^2} \sum_s E_W \sigma_s^2 \left(\sum_i Z_{ik} w_{is} \beta_{is}\right)^2$$

$$\preceq \frac{1}{N^2} \sum_s E_W \left(\sum_i w_{is} |Z_{ik}|\right)^2 \to 0,$$

where the convergence follows by Lemma 2, since as observed above, $E_W[|Z_{ik}|^2]$ is bounded. Finally, consider eq. (B.4). Decompose

$$\frac{1}{N} \sum_i X_i Y_i = \frac{1}{N} \sum_s \tilde{\mathcal{X}}_s \sum_i w_{is} Y_i(0) + \frac{1}{N} \sum_{i,s<t} w_{is} w_{it} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t \beta_{it}$$
$$+ \frac{1}{N} \sum_{i,s>t} w_{is} w_{it} \tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t \beta_{it} + \frac{1}{N} \sum_{s \neq t} (\mathcal{Z}_s' \gamma) \tilde{\mathcal{X}}_t \sum_i w_{is} w_{it} \beta_{it}$$
$$+ \frac{1}{N} \sum_{s \neq t} \tilde{\mathcal{X}}_s (\mathcal{Z}_t' \gamma) \sum_i w_{is} w_{it} \beta_{it} + \frac{1}{N} \sum_{i,s} w_{is}^2 (\mathcal{X}_s^2 - E[\mathcal{X}_s^2 \mid \mathcal{Z}_s, W]) \beta_{is}$$

$$+\frac{1}{N}\sum_{i,s,t} w_{is}w_{it}(\mathcal{Z}_s'\gamma)(\mathcal{Z}_t'\gamma)\beta_{it}+\frac{1}{N}\sum_{i,s} w_{is}^2\sigma_s^2\beta_{is}+\frac{1}{N}\sum_{i,s} w_{is}(\mathcal{Z}_s'\gamma)Y_i(0).$$

We will show that all summands except for the last three are $o_p(1)$. Since they are all mean zero conditional on $\mathscr{F}_0$, it suffices to show that their variances conditional on W converge to zero. The variance of the first summand is bounded by

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

by Lemma 2. The variance of the second summand is bounded by

$$\operatorname{var}\left(\frac{1}{N}\sum_{i,s<t} w_{is}w_{it}\tilde{\mathcal{X}}_s\tilde{\mathcal{X}}_t\beta_{it}\mid W\right)=\frac{1}{N^2}\sum_{s<t}E_W\sigma_s\sigma_t\left(\sum_i w_{is}w_{it}\beta_{it}\right)^2\preceq\frac{1}{N^2}\sum_{s<t}\overline{w}_{st}^2\to 0,$$

where the convergence to zero follows by arguments analogous to those in (B.7). The variance of the third summand converges to zero by analogous arguments. Variance of the fourth summand satisfies

$$\begin{aligned}\operatorname{var}\left(\frac{1}{N}\sum_{s\neq t}(\mathcal{Z}_s'\gamma)\tilde{\mathcal{X}}_t\sum_i w_{is}w_{it}\beta_{it}\mid W\right) &\leq \frac{1}{N}\sum_s E_W\sigma_s^2\left(\sum_t |(\mathcal{Z}_t'\gamma)|\sum_i w_{is}w_{it}|\beta_{is}|\right)^2\\ &\preceq \frac{1}{N}\sum_s E_W\left(\sum_i w_{is}\sum_t w_{it}|(\mathcal{Z}_t'\gamma)|\right)^2,\end{aligned}$$

which converges to by Lemma 2, since by Lemma 1, the second moment of $\sum_t w_{it}|(\mathcal{Z}_t'\gamma)|$ is bounded. Variance of the fifth summand converges to zero by analogous arguments. Finally, variance of the sixth summand satisfies

$$\operatorname{var}\left(\frac{1}{N}\sum_{i,s} w_{is}^2(\mathcal{X}_s^2-E[\mathcal{X}_s^2\mid\mathcal{Z}_s,W])\beta_{is}\mid W\right)\preceq\frac{1}{N^2}\sum_s E_W\left(\sum_i w_{is}^2\beta_{is}\right)^2\preceq\frac{1}{N^2}\sum_s n_s^2\to 0,$$

which yields (B.4). We now use eqs. (B.1), (B.2), (B.3) and (B.4) to derive the result. Since $U_i'\gamma = 0$, Equation (B.1) implies $Z'X/N = Z'Z\gamma/N + o_p(1)$. Consequently, since by Assumption 4(iii), $(Z'Z/N)^{-1} = o_p(1)$,

$$\frac{1}{N}\ddot{X}'\ddot{X}=\frac{1}{N}X'X-\frac{1}{N}X'Z(Z'Z)^{-1}Z'X=\frac{1}{N}\sum_s\sigma_s^2\overline{w}_{ss}+o_p(1)=\frac{1}{N}\sum_{i,s}\pi_{is}+o_p(1),\tag{B.8}$$

and, since $Z'Y/N = o_p(1)$,

$$\frac{1}{N}\ddot{X}'Y=\frac{1}{N}X'Y-\gamma'\frac{1}{N}Z'Y+o_p(1)=\frac{1}{N}\sum_{i,s}\pi_{is}\beta_{is}+o_p(1).$$

Combining Assumption 4(iii) with the preceding two displays then yields the result.

B.4 Proof of Proposition 4

Let $r_N = 1/\sum_s n_s^2$, and let E_W denote expectation conditional on W. Note that $\gamma' U_i = 0$ implies $Z\gamma = W\mathcal{Z}\gamma$. Therefore, $\ddot{X}$ admits the decomposition

$$\ddot{X} = (I - Z(Z'Z)^{-1}Z')X = (I - Z(Z'Z)^{-1}Z')(X - Z\gamma) = (I - Z(Z'Z)^{-1}Z')W\tilde{\mathcal{X}}.$$

Using this decomposition, we obtain

$$\begin{aligned} r_N^{1/2}(\ddot{X}'\ddot{X})(\hat{\beta} - \beta) &= r_N^{1/2}\ddot{X}'(Y - X'\beta) = r_N^{1/2}\tilde{\mathcal{X}}'W'(Y - X\beta - Z\check{\delta}) \\ &= r_N^{1/2}\tilde{\mathcal{X}}'W'(Y - X\beta - Z\delta) - r_N^{1/2}\tilde{\mathcal{X}}'W'Z(\check{\delta} - \delta) \\ &= r_N^{1/2}\sum_{s,i}\tilde{\mathcal{X}}_s w_{is}\epsilon_i - \frac{\tilde{\mathcal{X}}'W'Z}{N}(r_N N^2)^{1/2}(\check{\delta} - \delta) = r_N^{1/2}\sum_{s,i}\tilde{\mathcal{X}}_s w_{is}\epsilon_i + o_p(1). \end{aligned}$$

where the last line follows by Assumption 5(ii) and (B.5). It follows from eq. (B.8) and Assumption 4(iii) that $(\ddot{X}'\ddot{X}/N)^{-1} = (1 + o_p(1))(N^{-1}\sum_{i,s}\pi_{is})^{-1}$, so 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

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

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

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

Then we can write $\epsilon_i = V_i + \sum_t w_{it}\tilde{\mathcal{X}}_t(\beta_{it} - \beta)$, and, using the fact that $0 = \sum_{i,s}\pi_{is}(\beta_{is} - \beta) = \sum_s \sigma_s^2 b_{ss}$, we can decompose

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

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 $r_N^2\sum_{s=1}^S E_W[\mathcal{Y}_s^4] \to 0$ 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

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

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

Note that since $\sum_s |\sum_t w_{it} \mathcal{Z}_t' \gamma(\beta_{it} - \beta)|^4 \preceq \sum_s |\sum_t w_{it} \mathcal{Z}_t' \gamma|^4$, it follows from Lemma 1, Assumption 3(ii), Assumption 5(i), 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

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

by Assumption 3(i), since $\sum_s n_s^4 \leq \max_s n_s^2 / r_N$. Second, since β_{is} is bounded by Assumption 2(i), $b_{ss} \preceq \sum_i w_{is}^2 \leq n_s$, so that

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

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

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

we have

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

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 $\operatorname{var}(D_j \mid W) = E_W[\operatorname{var}(D_j \mid \mathcal{F}_0)] \to 0$. 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 3(i). 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 the fact that since $\sum_s \overline{w}_{st} = n_t$ and $\overline{w}_{st} \leq n_s$,

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

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

B.5 Proof of Proposition 5

Let $\hat{\theta} = (\hat{\beta}, \hat{\delta}')'$, $\theta = (\beta, \delta)$, $M_i = (X_i, Z_i')'$, $r_N = 1/\sum_{s=1}^{S} n_s^2$, and let

$$\hat{\mathcal{V}}_N = r_N \sum_s \widehat{\mathcal{X}}_s \hat{R}_s^2.$$

Since $\mathcal{V}_N = r_N \sum_s \sigma_s^2 R_s^2$, we can decompose this estimator as

$$\hat{\mathcal{V}}_N = r_N \sum_s (\widehat{\mathcal{X}}_s^2 - \tilde{\mathcal{X}}_s^2)\hat{R}_s^2 + r_N \sum_s \tilde{\mathcal{X}}_s^2 (\hat{R}_s^2 - R_s^2) + r_N \sum_s (\tilde{\mathcal{X}}_s^2 - \sigma_s^2) R_s^2 + \mathcal{V}_N. \tag{B.9}$$

We'll show that the first three terms are $o_p(1)$. Since $\hat{\epsilon}_i = \epsilon_i + M_i'(\theta - \hat{\theta})$, with $\epsilon_i = Y_i(0) - Z_i'\delta$, we can decompose

$$\hat{R}_s^2 = \sum_{i,j} w_{is} w_{js} \hat{\epsilon}_i \hat{\epsilon}_j = R_s^2 + 2\sum_{i,j} w_{js} w_{is} M_i'(\theta - \hat{\theta})\epsilon_j + \sum_{i,j} w_{is} w_{js} M_i'(\theta - \hat{\theta}) M_j'(\theta - \hat{\theta}). \tag{B.10}$$

Therefore, the second term in eq. (B.9) satisfies

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

where the second line follows from Lemma 3. Second, the variance of the third term in eq. (B.9) can be bounded by

$$\operatorname{var}(r_N \sum_s (\mathcal{X}_s^2 - \sigma_s^2) R_s^2 \mid W) = r_N^2 \sum_s E[(\mathcal{X}_s^2 - \sigma_s^2)^2 R_s^4 \mid W] \preceq r_N^2 \sum_s E[R_s^4 \mid W] \preceq r_N^2 \sum_s n_s^4 \to 0$$

since $r_N^2 \sum_s n_s^4 \leq \max_s n_s^2 / \sum_t n_t^2 \to 0$ by Assumption 3(i). Since

$$E[r_N \sum_s (\mathcal{X}_s^2 - \sigma_s^2) R_s^2 \mid W] = E[r_N \sum_s E[(\mathcal{X}_s^2 - \sigma_s^2) \mid \mathcal{F}_0] R_s^2 \mid W] = 0,$$

it follows by Markov inequality and the dominated convergence theorem that $r_N \sum_s (\mathcal{X}_s^2 - \sigma_s^2) R_s^2 = o_p(1)$.

It remains to show that the first term in eq. (B.9) is $o_p(1)$. Let $\hat{\gamma} = (Z'Z)^{-1} Z'X$. Since $W\mathcal{X} = X$ and $Z = W\mathcal{Z} + U$, it follows that

$$\begin{aligned} \widehat{\mathcal{X}} &= (W'W)^{-1} W' \ddot{X} = (W'W)^{-1} W'(X - Z(Z'Z)^{-1} Z'X) = \mathcal{X} - (W'W)^{-1} W'Z(Z'Z)^{-1} Z'X \\ &= \mathcal{X} - (W'W)^{-1} W'Z(\hat{\gamma} - \gamma) - (W'W)^{-1} W'Z\gamma \\ &= \tilde{\mathcal{X}} - (W'W)^{-1} W'Z(\hat{\gamma} - \gamma) \\ &= \tilde{\mathcal{X}} - \mathcal{Z}(\hat{\gamma} - \gamma) - (W'W)^{-1} W'U(\hat{\gamma} - \gamma). \end{aligned}$$

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. Consequently,

$$\begin{aligned} r_N \sum_s (\widehat{\mathcal{X}}_s^2 - \tilde{\mathcal{X}}_s^2)\hat{R}_s^2 &= (\hat{\gamma}-\gamma)' r_N \sum_s \mathcal{Z}_s \hat{R}_s^2 - (\hat{\gamma}-\gamma)' r_N \sum_s \mathcal{U}_s \hat{R}_s^2 \\ &= (\hat{\gamma}-\gamma)' O_p(1) - (\hat{\gamma}-\gamma)' O_p(1) \\ &= o_p(1), \end{aligned}$$

where the second line follows by applying Lemma 3 after using the expansion in eq. (B.10), and the third line follows since by eq. (B.1) and Assumption 4(iii), $\hat{\gamma} = \gamma + o_p(1)$.

B.6 Inference under heterogeneous effects

For valid (but perhaps conservative) inference under heterogeneous effects, we need to ensure that that when $\beta_{is} \neq \beta$, eq. (28) holds with inequality, that is,

$$\frac{\sum_{s=1}^S \widehat{\mathcal{X}}_s^2 \hat{R}_s^2}{\sum_{s=1}^S n_s^2} \geq \mathcal{V}_N + o_p(1). \tag{B.11}$$

To discuss conditions under which this is the case, observe that the "middle-sandwdich" in the asymptotic variance sandwich formula, $\mathcal{V}_N$, as defined in Proposition 4, can be decomposed into three terms:

$$\begin{aligned} \mathcal{V}_N &= \frac{\operatorname{var}\left(\sum_s \tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0\right)}{\sum_{s=1}^S n_s^2} \\ &= \frac{\sum_s E[\tilde{\mathcal{X}}_s^2 R_s^2 \mid \mathcal{F}_0]}{\sum_{s=1}^S n_s^2} - \frac{\sum_s E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0]^2}{\sum_{s=1}^S n_s^2} + \frac{\sum_{s\neq t} E[(\tilde{\mathcal{X}}_s R_s - E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0])(\tilde{\mathcal{X}}_t R_t - E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0]) \mid \mathcal{F}_0]}{\sum_{s=1}^S n_s^2}. \end{aligned}$$

where, as before $R_s = \sum_s w_{is}\epsilon_i$, and $\epsilon_i = Y_i(0) - Z_i'\delta + \sum_s \mathcal{X}_s w_{is}(\beta_{is} - \beta)$. Under homogeneous effects, R_s is non-random conditional on $\mathcal{F}_0$, and the second and third term are equal to zero, since in this case $E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0] = E[\tilde{\mathcal{X}}_s \mid \mathcal{F}_0]R_s = 0$, and $E[\tilde{\mathcal{X}}_s R_s \tilde{\mathcal{X}}_t R_t \mid \mathcal{F}_0] = R_s R_t E[\tilde{\mathcal{X}}_s \tilde{\mathcal{X}}_t \mid \mathcal{F}_0] = 0$ if $s \neq t$. Therefore, only the first term remains, and the standard error estimator consistently estimates this term by Proposition 5.

It can be shown that the proposition remains valid under regularity conditions if the effects β_{is} are heterogeneous, so that to ensure valid inference under heterogeneous effects, one needs to ensure that the sum of the second and third term is weakly negative. This is the case under several different settings. We now discuss two of them.

First observe that since $E[\tilde{\mathcal{X}}_s R_s \mid \mathcal{F}_0] = E[\tilde{X}_s \sum_{t=1}^S \mathcal{X}_t w_{it}(\beta_{it} - \beta) \mid \mathcal{F}_0] = \sigma_s^2 w_{is}(\beta_{is} - \beta)$, the second term equals

$$-\frac{\sum_{s=1}^{S}\left(\sum_{i=1}^{N}\pi_{is}(\beta_{is}-\bar{\beta})\right)^{2}}{\sum_{s=1}^{S}n_s^2},$$

where $\pi_{is}=w_{is}^2\sigma_s^2$ as in the statement Proposition 3. The term is always negative, and it reflects the variability of the treatment effect. It makes the variance estimate that we propose conservative if the third term equals zero. This is analogous to the result that the robust standard error estimator is conservative in randomized trials, and that the cluster-robust standard error estimator is conservative in cluster-randomized trials (see, for example Imbens and Rubin, 2015, Chapter 6). The third term reflects the correlation between $\mathcal{X}_s R_s$ and $\mathcal{X}_t R_t$, and it has no analog in cluster-randomized trials. Indeed, the term can be written as

$$\frac{1}{\sum_s n_s^2}\sum_{s\neq t}\sigma_s^2\sigma_t^2\sum_{i,j}w_{is}w_{it}(\beta_{it}-\beta)w_{js}w_{jt}(\beta_{js}-\beta).$$

In the example with "concentrated sectors", which is the analog of the cluster-randomized setup if there are no covariates, the term is thus 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. Another sufficient condition for validity of inference is that β_{is} and β_{jt} are uncorrelated if $t\neq s$, in which case it follows from the display above that the third term converges to zero. Numerical work, not reported here, indicates that the correlation between β_{is} and β_{jt} needs to be quite high and depend on the shares w_{is} in order for the third term to dominate the second term. We therefore expect our inference to remain valid for empirically relevant distributions of the effects β_{is}.