Quasi-experimental Shift-share Research Designs

Kirill Borusyak — Harvard
Peter Hull — U Chicago
Xavier Jaravel — LSE*

June 4, 2018

Abstract

Many empirical studies, such as Bartik (1991) and Autor et al. (2013), leverage shift-share instruments by combining a set of aggregate shocks with measures of individual shock exposure. We derive a necessary and sufficient shock-level orthogonality condition for such instruments to identify causal effects. We then show that this condition holds when shocks are as-good-as-randomly assigned and the number of shocks grows large. Our quasi-experimental framework implies several tests of shift-share instrument validity. We also highlight potential inconsistency of shift-share IVs with many shocks – similar to that of two-stage least squares with many instruments – which we show can be solved by split-sample estimation.

*borusyak@gmail.com, hull@uchicago.edu, and x.jaravel@lse.ac.uk.

We thank Paul Goldsmith-Pinkham, Edward Glaeser, Larry Katz, and Isaac Sorkin for helpful comments. This draft is a work in progress and should not be cited without prior approval. An earlier version was circulated in May 2017 under the title *Consistency and Inference in Bartik Research Designs*.

1 Introduction

A large and growing number of empirical studies leverage shift-share instruments, which combine a set of aggregate shocks with measures of individual shock exposure. For example, following the seminal work of Bartik (1991), Autor et al. (2013, henceforth ADH) combine industry-specific shocks to Chinese import competition in non-U.S. countries with the industrial composition of U.S. labor markets to study the effect of Chinese imports on U.S. employment.[1] Despite the popularity of shift-share instrumental variable (IV) regressions, however, it is not always clear what assumptions generally underlie their validity – that is, when should researchers expect the constructed instrument to be uncorrelated with a structural error term?

To answer this question, we first show that the shift-share IV exclusion restriction is satisfied if and only if a simple orthogonality condition holds across the different "components" of treatment (industries, for ADH) receiving the aggregate shocks. Our condition requires shocks to be uncorrelated with a relevant unobservable attribute of each component – specifically, a component-level weighted average of the untreated potential outcomes for the "units" (local labor markets, for ADH) most exposed to this shock. In the ADH shift-share IV, orthogonality would be violated if the growth of competition with China, measured outside the U.S., is systematically different in industries that are concentrated in regions of the U.S. where employment is falling for other reasons. For example, industries with a higher growth of competition with China may also be differentially exposed to unobserved technological shocks, or these industries may happen to be more concentrated in regions that are affected by other employment shocks.

Motivated by this orthogonality condition, we next develop an intuitive set of sufficient assumptions for shift-share IV validity. The key assumption is that the aggregate shocks are as-good-as-randomly assigned to treatment components, as if arising from a natural experiment. However, quasi-random shock assignment alone is not enough. A distinctive feature of our approach is that while the shocks are assigned to treatment components, the IV regression in estimated in a sample of individual units. Thus for the law of large numbers to apply to shift-share quasi-experiments, the number of independently shocked components must increase with the sample. Moreover, even though the assumptions allow each individual unit to only be exposed to a small number of shocks, on average exposure must be sufficiently dispersed such that a small number of shocks do not dominate the sample. Under these conditions, we show that the shift-share instrument satisfies exclusion even when the exposure to shocks is endogenous.[2]

As with other designs, the assumption of quasi-experimental shock assignment may only be plausible conditional on a set of observed covariates. Unique to shift-share IVs, however, researchers may have access

[1] While most studies exploit variation in the instrument across locations, observations may also represent firms differentially exposed to shocks in foreign markets (Hummels et al., 2014) or product groups with different sets of consumers (Jaravel, 2017). Other influential and recent examples of shift-share IVs, spanning many fields of economics, include Blanchard and Katz (1992), Luttmer (2006), Card (2009), Saiz (2010), Kovak (2013), Nunn and Qian (2014), Nakamura and Steinsson (2014), Oberfield and Raval (2014), Greenstone et al. (2014), Diamond (2016), and Suárez and Zidar (2016).

[2] This approach to identification contrasts with the identification condition recently proposed by Goldsmith-Pinkham et al. (2018), who emphasize the exogeneity of exposure to shocks with respect to the structural error term. For instance, in the ADH setting our approach requires the shocks to Chinese import penetration across industries to be as good as random, while Goldsmith-Pinkham et al. (2018) requires employment shares across locations to be as good as random.

to both unit-level and component-level controls. We show how researchers may relax the quasi-experimental assumptions, allowing shocks to be conditionally as-good-as-randomly assigned, by including an exposure-weighted average of component-level observables. For example, when components are naturally clustered into larger groups, researchers may control for measures of local cluster exposure to accommodate endogenous cluster shocks or avoid bias from observing a small number of clusters. Given a researcher's stance on how shocks may have been quasi-experimentally assigned, we outline various validations of the shift-share IV design, such as component-level balance tests, pre-trend checks, and tests for auto- and intra-class correlation of shocks.

Regarding inference, we argue that researchers should account for the variance of both quasi-experimental shock assignment and usual individual-level sample variation in shift-share designs. We follow Adao et al. (2018) in showing how to correct conventional standard errors when shocks are assumed to be independently and identically distributed, and how to extend those corrections to situations when shocks are clustered and heteroskedastic.

Lastly, we note that the quasi-experimental framework raises new issues with shift-share estimation. In applications researchers often estimate the aggregate shocks in the sample, in which case the feasible shift-share instrument may be mechanically correlated with the structural error. As the number of shocks grows, this correlation need not vanish asymptotically, rendering the feasible IV estimator inconsistent even though the orthogonality condition holds. This issue is closely related to the classic bias of two-stage least squares with many instruments, which Angrist et al. (1999) show can be overcome with split-sample IV estimation. Correspondingly, we recommend that researchers use split-sample estimates of aggregate shocks when constructing shift-share instruments. Our analysis thus provides a formal justification for leave-one-out shock estimation and highlights a virtue of measuring shocks in a separate sample.[3]

Our quasi-experimental framework is not the only way to satisfy the shift-share IV exclusion restriction. In related work Goldsmith-Pinkham et al. (2018) propose a conceptually different view of shift-share designs, in which the validity of the instrument comes from the exogeneity of shock exposure rather than from the shocks themselves. The advantage of their approach is that it does not require a large number of shocks, nor for the shocks to be as-good-as-randomly assigned. In our view, however, the quasi-experimental framework better aligns with researchers' motivation and goals in many applications. For example, Hummels et al. (2014), who combine country-by-product supply shocks with lagged firm-specific composition of input sources, write:

> "While these shocks are exogenous to Danish firms, their impact varies markedly across firms [...]. That is, if only one Danish firm buys titanium hinges from Japan, *idiosyncratic shocks* to the supply or transport costs of those hinges affects just that one firm." [Emphasis added]

Similarly, as previously mentioned, Autor et al. (2013) construct an instrument for import penetration from China in U.S. regions using the industry growth of Chinese imports in other developed economies. This

[3]For example, in the setting of Bartik (1991), national employment growth shocks can be measured with a leave-one-out procedure for each industry. This bias is less of a concern for ADH because they measure shocks in a separate sample of non-U.S. economies.

research design attempts to "purge" Chinese industry import penetration from U.S.-specific factors, which is effectively an attempt to obtain quasi-random variation in the industry space.

Econometrically, our approach is closely related to that of Kolesar et al. (2015) who study consistency in IV designs with many invalid instruments. Identification in that setting follows when violations of individual instrument exclusion restrictions are uncorrelated with their first-stage effects. Here, per Goldsmith-Pinkham et al. (2018), the shock exposure measures can be thought of as a set of invalid instruments, while our key orthogonality condition requires their exclusion restriction violations to be uncorrelated with the aggregate shocks (instead of with their first-stage effects, as in Kolesar et al. (2015)).[4]

Our paper also contributes to the recent literature studying shift-share designs, including Jaeger et al. (2017) and Broxterman and Larson (2018). The former studies biases of shift-share IV due to local labor market dynamics, while the latter studies the empirical performance of different shift-share instrument constructions. More broadly, our paper adds to a growing econometric literature seeking to interpret, test, and relax the high-level assumptions of common applied research designs, including Goldsmith-Pinkham et al. (2018) for shift-share designs, Borusyak and Jaravel (2016) for event study designs, de Chaisemartin and D'Haultfoeuille (2018) for instrumented difference-in-difference designs, and Hull (2018) for mover designs.

The rest of the paper is organized as follows. Section 2 introduces the framework and derives the key orthogonality condition. Section 3 provides intuitive sufficient conditions on the quasi-experimental nature of the shocks that guarantee orthogonality, and proposes tests for these conditions. In Section 4 we highlight the potential for inconsistency when the aggregate shocks are estimated in the sample. Section 5 concludes. The Appendix builds an inferential framework for quasi-experimental shift-share designs, following Adao et al. (2018), and extends the quasi-experimental framework to panel IVs.

2 Shift-share Designs and Orthogonality

Suppose we observe an outcome y_ℓ and a treatment x_ℓ in an *iid* sample of size L. We assume treatment can be written as a weighted average of N components: $x_\ell = \sum_{n=1}^N s_{\ell n} x_{\ell n}$, where $\sum_{n=1}^N s_{\ell n} = 1$.[5] Here the exposure weights $s_{\ell n}$ are also observed, though we do not require the researcher to observe the treatment components $x_{\ell n}$ themselves. In the ADH design, y_ℓ and x_ℓ denote the growth rates of employment and Chinese import penetration in local labor market ℓ, respectively, while $s_{\ell n}$ is the employment share of industry n in location ℓ, measured in a base period. The $x_{\ell n}$ would then capture the local Chinese import penetration of each industry.[6]

[4] As we show below, Kolesar et al.'s application to examiner designs, as in Chetty et al. (2011), can be thought of as a special case of ours, in which the exposure measures are binary and the aggregate shocks are group-level treatment averages.

[5] Similar results are obtained if the sum of all shares varies across locations but is included as a control, as in ADH.

[6] Because ADH do not observe imports by region, they proxy local import penetration in each industry by the national average to construct x_ℓ. We abstract from this practical issue in our conceptual discussion of their paper.

We wish to estimate the causal effect of the treatment, assuming a linear constant-effect model

$$y_\ell = \gamma + x_\ell \beta + \epsilon_\ell, \tag{1}$$

where $\mathbb{E}[\epsilon_\ell] = 0$. Here ϵ_ℓ represents the (demeaned) untreated potential outcome of unit ℓ – that is, the outcome we would observe for ℓ if treatment were set to zero, relative to the population of units. We do not require realized treatment to be uncorrelated with ϵ_ℓ, in which case the causal parameter β would be consistently estimable by ordinary least squares (OLS). Instead, we construct a shift-share instrument by combining the observed exposure weights $s_{\ell n}$ with a set of N component-level shocks g_n:

$$z_\ell = \sum_{n=1}^{N} s_{\ell n} g_n. \tag{2}$$

Here z_ℓ can thus be thought of as a predicted shock for unit ℓ. For ADH, g_n is the growth of import penetration from China in eight developed non-U.S. economies in industry n, and the shift-share instrument z_ℓ is interpreted as a predicted "China shock" in the region. To start simply we treat the aggregate shocks as known, so that they do not need to be estimated; in Section 4 we return to the case where, as in Bartik (1991), the g_n are estimated in the shift-share IV sample.

We seek to establish conditions under which the instrumental variables (IV) estimator $\hat{\beta}$ that uses z_ℓ to instrument x_ℓ in equation (1) is consistent for the causal parameter β; that is, conditions under which $\hat{\beta} \xrightarrow{p} \beta$ as $L \to \infty$. As usual, IV consistency requires both instrument relevance (that z_ℓ and x_ℓ are asymptotically correlated) and validity (that z_ℓ is asymptotically uncorrelated with ϵ_ℓ). Since relevance can be inferred from the data, here we focus on validity. Our goal is to show that identification can come from properties of the component-level shocks g_n; a challenge is that the validity condition is expressed in terms of variation over observations ℓ, not over components n. However, using the fact that $\mathbb{E}[\epsilon_\ell] = 0$, we may rewrite

$$\begin{aligned} \operatorname{Cov}[z_\ell, \epsilon_\ell] &= \mathbb{E}\left[\sum_{n=1}^{N} s_{\ell n} g_n \epsilon_\ell\right] \\ &= \sum_{n=1}^{N} g_n \mathbb{E}[s_{\ell n} \epsilon_\ell] \\ &= \sum_{n=1}^{N} s_n g_n \phi_n, \end{aligned} \tag{3}$$

where $s_n \equiv \mathbb{E}[s_{\ell n}]$ measures the average exposure to treatment component n, and where $\phi_n \equiv \mathbb{E}[s_{\ell n} \epsilon_\ell] / \mathbb{E}[s_{\ell n}]$ is a weighted expectation of the untreated potential outcome ϵ_ℓ, with larger weights $s_{\ell n}$ given to observations where treatment is more exposed to component n. The validity of the shift-share design thus requires[7]

$$\sum_{n=1}^{N} s_n g_n \phi_n \to 0. \tag{4}$$

Equation (4) is our key orthogonality condition. The left-hand side represents a weighted covariance (with weights s_n) between two component-level variables: the shocks g_n and the relevant unobservables ϕ_n.[8] Thus in ADH, equation (4) characterizes the large-sample behavior of a weighted covariance between non-U.S. Chinese import penetration growth across different industries (g_n) and a weighted average of U.S. unobservables affecting employment growth in locations specializing in each industry (ϕ_n).

Equation (4) formalizes the sense in which it is useful in ADH to use non-U.S. Chinese import penetration growth across different industries to build the China shock instrument. Specifically, reverse causality would be a concern if U.S. Chinese import penetration growth rates were used, as China may gain market shares in certain U.S. industries precisely because these industries are not performing well (and, in particular, would have had low employment growth even absent Chinese competition). Using non-U.S. Chinese imports addresses such reverse causality, though only to the extent that the underlying performance of U.S. and non-U.S. industries is not correlated. In addition, equation (4) helps understand potential concerns about omitted variable bias, because of either unobserved industry shocks or regional shocks. For example, industries with a higher growth of competition with China may also be more exposed to certain technological shocks like automation; or these industries could happen to have larger employment shares in regions that are affected by other employment shocks, such as immigration.

Three points on the shift-share orthogonality condition are worth highlighting. First, note that the weight that the industry receives in the covariance (3) is the expected shock exposure s_n. For example, in the ADH setting s_n is the average employment share of industry n in the population. In practice researchers leveraging variation in employment shares sometimes weight shift-share IV regressions by total location employment (e.g., Card (2009) and ADH themselves). In this case s_n has a more intuitive interpretation: the lagged population size of industry n. Second, note that variation in component exposure across observations plays an important role in the orthogonality condition; if all units have the same exposure to a given shock n, $s_{\ell n} = \bar{s}_n$, then $\phi_n = \mathbb{E}[1 \cdot \epsilon_\ell] = 0$, and shock n will not contribute to (4).[9] Finally, note that if the exposure measures themselves are as-good-as-randomly assigned, i.e. if $\mathbb{E}[s_{\ell n}\epsilon_\ell] = 0$ for all n, then $\phi_n = 0$ for all n and equation (4) is satisfied for any set of shocks. This is the interpretation of shift-share IV proposed by

[7] For initial simplicity in this section we derive the validity condition given fixed sequences of the set of (g_n, s_n, ϕ_n). In the quasi-experimental framing it is more natural to imagine a hierarchical sampling design, in which sets are drawn from a larger population in each sample of size L. All expectations and covariances in this section should then be thought to be conditional on the component-level draws, with validity satisfied when $\sum_{n=1}^{N} s_n g_n \phi_n \xrightarrow{p} 0$. We formalize the hierarchical sampling logic in Section 3.

[8] The weighted covariance interpretation follows from observing that the weighted mean of ϕ_n is zero, $\sum_{n=1}^{N} s_n \phi_n = \sum_{n=1}^{N} \mathbb{E}[s_{\ell n}\epsilon_\ell] = \mathbb{E}[\epsilon_\ell] = 0$, and that the weights sum to one: $\sum_{n=1}^{N} s_n = \mathbb{E}\left[\sum_{n=1}^{N} s_{\ell n}\right] = \mathbb{E}[1] = 1$.

[9] More generally, the orthogonality condition is more sensitive to shocks with more variation in measured exposure. Indeed, we can also write $\text{Cov}[z_\ell, \epsilon_\ell]$ as a variance-weighted average $\sum_n \text{Var}[s_{\ell n}] \cdot g_n \phi_n^{\text{R}}$, where $\phi_n^{\text{R}} = \text{Cov}[s_{\ell n}, \epsilon_\ell] / \text{Var}[s_{\ell n}]$ is a regression-type metric of the association between $s_{\ell n}$ and ϵ_ℓ.

Goldsmith-Pinkham et al. (2018).

3 Quasi-experimental Shock Assignment

When might the orthogonality condition be satisfied, and how can it be validated? In this section we address these questions formally, by establishing a quasi-experimental framework in which the aggregate shocks are as-good-as-randomly assigned with respect to other relevant component characteristics. We also give guidance on which individual controls a researcher might include in a shift-share IV to weaken the central assumption of as-good-as-random shock assignment. As with other methods, different quasi-experimental designs imply different testable restrictions; to make best use of these tests, researchers thus should take a stand on the source of idiosyncratic shock variation.

To formalize the shift-share quasi-experiment, we first need to take a step back. Section 2 viewed component characteristics, including the aggregate shocks, as fixed given a sample of size L. We now imagine a hierarchical data-generating process, in which conditional on industry exposures $s_1, \ldots, s_N$ (which for simplicity we continue to take as fixed), the set of aggregate shocks g_n and relevant unobservables ϕ_n are drawn from some distribution. This assignment of shocks to components n constitutes the shift-share quasi-experiment: for example in ADH, we may imagine an unanticipated change to Chinese productivity due to a policy change that differentially affects non-U.S. import penetration growth g_n.[10]

In our baseline case, we consider two conditions on the component-level data-generating process:

A1. Shocks g_n are mean-independent of ϕ_n with the same mean for each n: $\mathbb{E}\left[g_n \mid \phi_n\right] = \mu$

A2. Shocks are mutually mean-independent, conditional on $\phi_1, \ldots, \phi_N$, and the Herfindahl index of component exposure converges to zero: $\sum_{n=1}^N s_n^2 \to 0$

Under these assumptions and weak regularity conditions, the law of large numbers then implies the orthogonality condition (4).[11]

Assumption A1 states that the aggregate shocks are quasi-randomly assigned, formalized by mean-independence of g_n with respect to the relevant unobservables ϕ_n. As usual, this sort of exogenous assignment cannot be directly tested. However, an indirect test is feasible when there are observable correlates of the structural error ϵ_ℓ, which we denote r_ℓ. For example, in the ADH setting the pre-period share of immigrants in the location r_ℓ is likely to predict immigrant inflow in the subsequent period and therefore affect employment growth through ϵ_ℓ. Substituting r_ℓ in place of ϵ_ℓ, the researcher can check orthogonality

[10] Note that for shift-share IV relevance the quasi-experiment must also affect $x_{\ell n}$ and thereby generate variation in individual treatment, which can be verified empirically. In the ADH setting, relevance means that changes in Chinese import penetration outside the U.S. predict import penetration in U.S. local labor markets, when weighted by location-specific employment weights.

[11] Formally, we assume that the support of the distribution of ϕ_n is bounded by $[-P, P]$ and $\text{Var}\left[g_n\right] < V$ for finite P and V. Then A1 implies $\mathbb{E}\left[\sum_{n=1}^N s_n \phi_n g_n\right] = \sum_{n=1}^N \mathbb{E}\left[s_n \phi_n\right] \mu = \mathbb{E}\left[\epsilon_\ell\right] \mu = 0$, while by A2 and the Cauchy-Schwartz inequality $\text{Var}\left[\sum_{n=1}^N s_n \phi_n g_n\right] = \mathbb{E}\left[\left(\sum_{n=1}^N s_n \phi_n g_n\right)^2\right] \le \sum_{n=1}^N \mathbb{E}\left[s_n^2 \phi_n^2\right] \text{Var}\left[g_n\right] \le PV \sum_{n=1}^N s_n^2 \to 0$. This guarantees L^2-convergence and thus weak convergence: $\sum_{n=1}^N s_n \phi_n g_n \xrightarrow{p} 0$.

either at the observation or component level. The former amounts to simply testing $\text{Cov}\left[z_\ell, r_\ell\right] = 0$, while the latter requires first measuring $\tilde{r}_n = \sum_{\ell=1}^{L} s_{\ell n} r_\ell / \sum_{\ell=1}^{L} s_{\ell n}$ and correlating it with the shocks.[12]

The second assumption A2 ensures that, given quasi-experimental assignment, a law of large numbers applies in large samples to weighted covariance (3). Since $\sum_{n=1}^{N} s_n^2 \geq 1/N$, the restriction on component exposure implies that the number of components grows with the sample, while the mutual mean-independence condition implies that additional shocks draw (3) closer to zero when A1 holds. The concentration condition is stated in terms of the Herfindahl index but this is not the only possibility; since $\sum_{n=1}^{N} s_n^2 \leq \sum_{n=1}^{N} \left(\max_m s_m\right) s_n = \max_m s_m$, it is sufficient that the largest component becomes vanishingly small as L grows.

Testing mean-independence is straightforward in a common case where components are grouped into bigger clusters $c\left(n\right) \in \left\{1, \ldots, C\right\}$: for example, detailed industries in ADH can be grouped into larger industrial sectors. Under A2 the intra-cluster correlation of shocks should be zero, which is easily verified. A modified version of A2 may be applied when aggregate shocks are correlated within clusters (but not across), requiring the number of clusters C to grow large with the cluster exposure concentration becoming small. When there are naturally-arising clusters, we recommend researchers compute the shock intra-cluster correlation coefficient and take a stance on how the shocks are assigned in the quasi-experiment.[13]

As with other quasi-experimental designs, researchers may wish to assume that A1 and A2 only hold conditionally on observed controls. For example, one could weaken the mean-independence restriction of A1 to allow the conditional mean of shocks to depend linearly on a vector of component-level observables q_n,

$$\mathbb{E}\left[g_n \mid \phi_n, q_n\right] = \mu + q_n' \tau \tag{5}$$

for some μ and τ, and similarly assume the mutual mean-independence condition of A2 holds for the residual $g_n^* = g_n - \mathbb{E}\left[g_n \mid \phi_n, q_n\right]$. In this case it is straightforward to use only the residual shock variation for identification by adding an exposure-weighted vector of component controls $\sum_{n=1}^{N} s_{\ell n} q_n$ to the IV specification. As shown in Appendix B, the resulting estimator is equivalent to a two-step procedure in which the residuals g_n^* are first estimated by a weighted regression of g_n on q_n and then used to construct a new shift-share instrument $\hat{z}_\ell^* = \sum_n s_{\ell n} \hat{g}_n^*$.

An intuitive application of this result is again found in the case of clustered components. Here the researcher may be more willing to assume that aggregate shocks are exogenously assigned within clusters, but that the cluster-average shock is endogenous. With q_n including a set of cluster indicators, the shift-share IV would then be valid with the researcher controlling for the individual exposure to each cluster (e.g. the local employment shares of each sector in ADH), $s_{\ell c} = \sum_{n=1}^{N} s_{\ell n} \mathbf{1}\left[c\left(n\right) = c\right]$. Garin and Silverio (2017)

[12] The two tests are equivalent. To implement them in the data under A2 one can use the asymptotic theory developed in Appendix A: e.g., regress r_ℓ on z_ℓ and use the appropriate standard errors. Importantly, and as made clear in the Appendix, inference for this test should account for the component-level source of quasi-random variation and is thus non-standard.

[13] The Herfindahl index can also be estimated from the sample, either at the component or cluster level, though judging whether it is sufficiently small in finite samples is beyond the scope of our paper.

and Jaravel (2017) follow this strategy. Interestingly, the same approach also relaxes A2. If shocks have a cluster-specific part, $g_n = \mu + q_{c(n)} + g_n^*$, and the number of clusters is small, then the law of large numbers can apply to g_n^* but not to g_n, even if q_c are as-good-as-randomly assigned and thus A1 holds. There are therefore two distinct reasons to control for cluster exposure: either to remove a non-random part of the shocks or to remove a random part that creates too much correlation across shocks.

In practice researchers often control for other individual observables w_ℓ. For example in their benchmark specifications, ADH control for a location's labor force demographics as measured in a previous period. Although our key orthogonality condition is stated at the component level, it is straightforward to justify such individual-level controls: they replace the relevant component-level unobservable ϕ_n in (4) with $\tilde{\phi}_n = \mathbb{E}\left[s_{\ell n}\tilde{\epsilon}_\ell\right]/\mathbb{E}\left[s_{\ell n}\right]$, where $\tilde{\epsilon}_\ell$ denotes the population residual from projecting the structural error ϵ_ℓ on w_ℓ. Thus ADH's specification allows growth rates to be correlated with an exposure-weighted average of location demographics, provided they are not correlated with other aggregated unobservables.

To conclude this section, we note that the quasi-experimental framework easily generalizes to a panel setting in which outcomes, treatment, and shocks are observed for the same units over T periods. This setup, which we formally develop in Appendix C, delivers three insights. First, when the analogs of assumptions A1 and A2 hold (implying that shocks are mean-independent across both units and periods), identification can come either from having many components N or from observing a small number of components over many periods T, provided shocks are quasi-experimentally assigned in each period. Second, researchers can relax the assumptions on the quasi-experiment by including unit and period fixed effects in the regression: aggregate shocks are then allowed to be correlated with time-invariant aggregated local unobservables (even if exposure varies over time) and may have time-varying means. Similarly, controls can be constructed to extract the idiosyncratic component from serially correlated shocks when their stochastic process belongs to a known parametric class, such as AR(1). Finally, the appendix shows how standard pre-trend tests can be used to validate the panel quasi-experimental design.

4 Many-Shock Inconsistency and Split-Sample Estimation

So far we have assumed that the researcher directly observes the set of shocks g_n. In practice, however, these shocks are typically estimated, sometimes within the same sample used for IV estimation. For example, Bartik (1991) uses national industry growth rates as shocks, which he estimates by the sample average growth of industries across observed locations, before estimating the labor supply equation in the same location sample.[14] We show in this section that with many shocks such two-step estimation may lead to inconsistency of the shift-share IV estimator, even when orthogonality holds. Intuitively, this bias arises from the fact that shock estimation error need not vanish asymptotically and may be systematically correlated

[14] Bartik (1991) develops a research design to estimate the impact of labor demand shocks on wages, i.e., the (inverse) labor supply elasticity. His insight is to use nation-wide industry employment growth rates as a proxy for "labor demand shocks," leveraging the idea that short run changes in employment at the national level are driven by labor demand, rather than labor supply.

with untreated potential outcomes.

Formally, suppose a researcher estimates each g_n from a weighted average of in-sample observations of a variable $g_{\ell n}$.[15] That is, given weights $\omega_{\ell n}$, she computes

$$\hat{g}_n = \sum_{\ell=1}^{L} \frac{\omega_{\ell n} g_{\ell n}}{\sum_{m=1}^{L} \omega_{mn}}. \tag{6}$$

For example in Bartik $g_{\ell n} = x_{\ell n}$, the local employment growth rate of industry n in location ℓ, and $\omega_{\ell n}$ denotes the local lagged industry employment. The researcher then uses $\hat{g}_n$ to form a feasible shift-share instrument

$$\begin{aligned} \hat{z}_\ell &= \sum_{n=1}^{N} s_{\ell n} \hat{g}_n \\ &= z_\ell + \sum_{n=1}^{N} s_{\ell n} (\hat{g}_n - g_n) \\ &= z_\ell + \psi_\ell, \end{aligned} \tag{7}$$

where we define

$$\psi_\ell = \sum_{n=1}^{N} s_{\ell n} \sum_{m=1}^{L} \frac{\omega_{mn}(g_{mn} - g_n)}{\sum_{k=1}^{L} \omega_{kn}} \tag{8}$$

as a weighted average (with weights $s_{\ell n}$) of component-level estimation error $\hat{g}_n - g_n$.

When the orthogonality condition for the infeasible shift-share instrument z_ℓ holds, validity of the feasible $\hat{z}_\ell$ requires an additional condition:

$$\text{Cov}\left[\psi_\ell, \epsilon_\ell\right] \to 0. \tag{9}$$

When might this condition fail? Intuitively as the number of aggregate shocks grows large we might expect the number of observations informing each g_n to remain small. The estimated shocks $\hat{g}_n$ may then be inconsistent and have a non-vanishing influence of the observations with a large exposure to g_n. If deviations $g_{\ell n} - g_n$ are systematically correlated with the structural error ϵ_ℓ, this will generate an asymptotic bias for the feasible shift-share IV estimator. This sort of bias is analogous to that of conventional two-stage least squares (2SLS) estimation with many instruments.

To make this link explicit, consider the special case of an "examiner," or "judge" shift-share design, in which each unit is exposed to only one shock, $s_{\ell n} \in \{0, 1\}$ and shocks are given by the group-level expectation

[15] For notational simplicity, in this section we return to the convention of treating the shocks g_n as fixed.

of treatment, $g_n = \mathbb{E}[x_\ell \mid s_{\ell n} = 1]$.[16] A researcher estimates the shocks by the sample group-level average,

$$\hat{g}_n = \sum_{\ell=1}^{L} \frac{s_{\ell n} x_\ell}{\sum_{m=1}^{L} s_{mn}}, \tag{10}$$

so that in terms of the above general formulation $g_{\ell n} = x_\ell$ and $\omega_{\ell n} = s_{\ell n}$. For simplicity we assume there are no additional controls w_ℓ. Decomposing treatment and the structural error into between- and within-group variation,

$$x_\ell = \sum_{n=1}^{N} g_n s_{\ell n} + \eta_\ell \tag{11}$$

$$\epsilon_\ell = \sum_{n=1}^{N} \gamma_n s_{\ell n} + \nu_\ell, \tag{12}$$

with $\mathbb{E}[\eta_\ell \mid s_{\ell 1}, \ldots, s_{\ell N}] = \mathbb{E}[\nu_\ell \mid s_{\ell 1}, \ldots, s_{\ell N}] = 0$, and assuming conditional homoskedasticity of (η_ℓ, ν_ℓ), it can then be shown that in an examiner design

$$\text{Cov}[\hat{z}_\ell, \epsilon_\ell] = \sum_{n=1}^{N} s_n g_n \gamma_n + \frac{N}{L}\sigma_{\eta\nu}, \tag{13}$$

where the first term represents $\text{Cov}[z_\ell, \epsilon_\ell]$ and the second represents $\text{Cov}[\psi_\ell, \epsilon_\ell]$, with $\sigma_{\eta\nu} = \text{Cov}[\eta_\ell, \nu_\ell]$.[17] Thus when our orthogonality condition is satisfied in the examiner case, the asymptotic bias of the feasible shift-share IV aligns with the original many-instrument 2SLS bias term of Nagar (1959), $\frac{N}{L}\sigma_{\eta\nu}$.

While stark, the examiner example highlights a general issue in shift-share designs when many poorly-estimated aggregate shocks are used to construct $\hat{z}_\ell$. For the asymptotic variance of the shift-share instrument to be non-degenerate, the exposure shares $s_{\ell n}$ must be sufficiently concentrated in a small number of components n. In many cases a valid shift-share design may thus resemble a "fuzzy" examiner design, in which each ℓ is "assigned" most of exposure to few n. In such cases researchers may worry about Nagar-style bias when the number of treatment components N is non-negligible relative to the sample L.

Fortunately, as with the conventional bias of 2SLS, this issue has a simple solution in the form of sample splitting (Angrist and Krueger (1995); Angrist et al. (1999)). Rather than using all observations to both estimate shocks and the shift-share IV coefficient, suppose the researcher randomly partitions the sample for

[16] Recent examples of examiner designs include Kling (2006), Chetty et al. (2011), Maestas et al. (2013), and Doyle et al. (2015). Notably, Kolesar et al. (2015) study the Chetty et al. (2011) design under the assumption that examiner groups are invalid instruments for treatment, leveraging an orthogonality similar to ours for shift-share designs. Our results in this section can thus be thought to generalize this analysis.

[17] Namely, under independent sampling $\text{Cov}[\psi_\ell, \epsilon_\ell] = \mathbb{E}\left[s_\ell'(s's)^{-1}s_\ell \mathbb{E}[\epsilon_\ell \eta_\ell \mid s]\right] = \mathbb{E}\left[s_\ell'(s's)^{-1}s_\ell \mathbb{E}[\nu_\ell \eta_\ell \mid s]\right]$, where s_ℓ is a vector collecting the $s_{\ell n}$ and s is an $N \times L$ matrix collecting the s_ℓ. Under homoskedasticity, this is $\mathbb{E}\left[s_\ell'(s's)^{-1}s_\ell\right]\sigma_{\eta\nu} = (N/L)\sigma_{\eta\nu}$. The first term of equation (13) follows from the fact that in the examiner case, $\phi_n = \mathbb{E}[s_{\ell n}\epsilon_\ell] / \mathbb{E}[s_{\ell n}] = \gamma_n$.

these two distinct purposes. At the extreme, we could imagine using leave-one-out estimates of the shocks

$$\tilde{g}_{\ell n} = \sum_{m \neq \ell} \frac{\omega_{mn} g_{mn}}{\sum_{k \neq \ell} \omega_{kn}} \tag{14}$$

to form a leave-one-out shift-share instrument $\tilde{z}_\ell = \sum_{n=1}^{N} s_{\ell n} \tilde{g}_{\ell n}$. Under independent sampling, we then have, by the law of iterated expectations,

$$\begin{aligned} \text{Cov}\left[\tilde{z}_\ell, \epsilon_\ell\right] &= \sum_{n=1}^{N} \mathbb{E}\left[s_{\ell n} \epsilon_\ell \mathbb{E}\left[\tilde{g}_{\ell n} \mid s_{\ell n}, \epsilon_\ell\right]\right] \\ &= \sum_{n=1}^{N} s_n \mathbb{E}\left[\tilde{g}_{\ell n}\right] \phi_n. \end{aligned}$$

Thus the validity condition for the feasible shift-share IV estimator is the same as the quasi-experimental orthogonality condition (4), with $\mathbb{E}\left[\tilde{g}_{\ell n}\right]$ replacing g_n. Of course when the leave-one-out shock estimator is unbiased, these conditions are exactly the same.

This discussion of many-shock bias provides a formal justification for leave-one-out estimation of aggregate shocks in shift-share designs, a practice that has become quite common (tracing back at least as far as Autor and Duggan (2003)), though often with little theoretical underpinning. The split-sample solution also highlights a virtue of shift-share designs in which the aggregate shocks are measured in a separate sample for other substantive reasons. For example, Autor et al. (2013) use Chinese import shocks measured outside the U.S. to instrument for import competition in U.S. regions. We recommend researchers use either form of sample splitting to minimize the risk of inconsistency in otherwise valid shift-share designs.

5 Conclusion

Shift-share instruments combine variation in the exposure to aggregate shocks with variation in the shocks *per se*. We provide a general framework for understanding the validity of these instruments, while focusing on the latter source of variation. We show that in general shift-share instruments are valid when the shocks are not systematically correlated with a weighted average effect of other unobserved factors in units which have high exposure to those shocks. Importantly, this correlation is not across observations (e.g., U.S. commuting zones in Autor et al., 2013) but across treatment components (industries). Technically, the orthogonality condition can be satisfied if either the exposure measures or the shocks are as-good-as-randomly assigned. We argue that quasi-experimental shock variation may be a more plausible assumption in applications, such as with the China shocks of Autor et al., 2013. The quasi-experimental approach assumes shocks are drawn as-good-as-randomly and independently across treatment components, perhaps conditional on observables, with the average exposure to any one component becoming small as the sample grows.

This analysis produces three practical recommendations for shift-share estimation. First, we argue that researchers should adopt the quasi-experimental mindset and take a stand on which variation in growth rates is plausibly random: for instance, within component clusters, across clusters, or both. This choice matters for whether exposure to clusters should be controlled for, and (as we show in the appendix) how to compute standard errors. Second, if potential correlates of untreated potential outcomes are observed, their correlation with the instrument may indicate design problems; we show how these sorts of balance tests may be run at the component level, in line with the quasi-experimental framework. Finally, we recommend researchers use leave-one-out or other types of split sample methods for estimating aggregate shocks when they are not directly observed, as the failure to do so with many components may cause the shift-share IV to be inconsistent. Each of these recommendations draws on intuitions that applied researchers are likely to have from other quasi-experimental settings, bringing shift-share IV estimators to familiar econometric ground.

A Asymptotic Theory

The quasi-experimental view on shift-share designs has implications for the inference on the IV estimator. While it is developed in detail by Adao et al. (2018), we also build it here for completeness and to help the reader implement the tests we proposed. Intuitively, the quasi-random assignment of g_n guarantees that the orthogonality condition holds in expectation yet not in any given sample of finite N. When $N \ll L$, the finite-sample correlation between g_n and ϕ_n may be much more important than the noise that is due to random sampling of units, which is accounted for by conventional standard errors.

Another way to understand this point is that units with similar exposure to observed shocks g_n may have similar exposure to some unobserved shocks as well. For example, in the Autor et al. (2013) application, the degree of import competition from China in the U.S. is measured using exports from China to third countries. This approach removes the impact of other industry shocks (e.g., U.S. supply shocks) from the instrument, but those shocks still have similar effects on locations with similar industry mix. More generally, exposure to unobserved shocks is one of the reasons why $s_{\ell n}$ shares may not be valid instruments (Kolesar et al., 2015). But when different units have similar exposure to the same shocks, their residuals will be correlated, invalidating conventional standard errors.

To characterize the asymptotic behavior of the 2SLS estimator $\hat{\beta}$, we exploit the quasi-experimental nature of g_n. We focus on the case where g_n is known and so $\hat{\beta}$ is consistent, abstracting from the issues of Section 4. Note that we can write

$$\begin{aligned}\hat{\beta} - \beta &= \frac{\tilde{z}'\epsilon}{\tilde{z}'x} \\ &= \frac{g'\tilde{S}'\epsilon}{\tilde{z}'x},\end{aligned} \tag{15}$$

where $\tilde{z}$ denotes the vector of residuals from projecting a set of L observations of z_ℓ on controls w_ℓ, and similarly $\tilde{S} = (\tilde{s}_{\ell n})$ is an $L \times N$ matrix of residuals from projecting the observed $s_{\ell n}$ on w_ℓ (which implies $\sum_n \tilde{s}_{\ell n} = 0$), for each n. Let $\xi = \frac{N}{L}\tilde{S}'g$, i.e. $\xi_n = \frac{N}{L}\sum_\ell \tilde{s}_{\ell n}\epsilon_\ell$. Then

$$\sqrt{N}(\hat{\beta} - \beta) = \frac{\xi' g/\sqrt{N}}{\tilde{z}'x/L}. \tag{16}$$

We assume that $\tilde{z}'x/L \xrightarrow{p} \pi \neq 0$, which is the standard instrument relevance condition. Consider the numerator. It is centered around zero ($\mathbb{E}[g'\xi] = 0$) because $\sum_n \xi_n = \sum_\ell \left(\sum_n \tilde{s}_{\ell n}\right)\epsilon_\ell = 0$. Moreover, by the *iid*ness of g_n, the terms $\xi_n g_n$ are uncorrelated with each other, and

$$\text{Var}\left[g'\xi/\sqrt{N}\right] = \sigma_g^2 \cdot \mathbb{E}\left[\frac{1}{N}\sum_n \xi_n^2\right],$$

where $\sigma_g^2 = \text{Var}[g_n]$ for any n. We will assume that $\frac{1}{N}\sum_n \xi_n^2 \xrightarrow{p} \sigma_\xi^2$. Then an appropriate CLT would imply

$$\sqrt{N}(\hat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V), \quad \text{where } V = \frac{\sigma_g^2 \sigma_\xi^2}{\pi^2}. \tag{17}$$

Moreover, σ_g^2 and σ_ξ^2 can be naturally estimated by the sample variance of g and $\hat{\xi}_n = \frac{N}{L}\sum_\ell \tilde{s}_{\ell n}\hat{\epsilon}_\ell$, respectively, where $\hat{\epsilon}_\ell = y_\ell - \hat{\beta}x_\ell - \hat{\mu}' w_\ell$.

This approach admits natural extensions to heteroskedasticity and clustering in the space of components n. If g_n is assumed to be independent of each other and mean-independent of ξ_n but potentially heteroskedastic,

$$\sqrt{N}(\hat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V_{\text{het}}), \quad \text{where } V_{\text{het}} = \frac{1}{\pi^2} \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_n \left((g_n - \mathbb{E}[g])\,\xi_n\right)^2.$$

This asymptotic variance can be consistently estimated by $\frac{1}{N}\sum_n \left((g_n - \bar{g})\,\hat{\xi}_n\right)^2 / \hat{\pi}^2$, where $\bar{g} = \frac{1}{N}\sum_n g_n$. Similarly, if there are C shock clusters denoted c such that g_n is independent across clusters and mean-independent of ξ_n,

$$\sqrt{N}(\hat{\beta} - \beta) \Rightarrow \mathcal{N}(0, V_{\text{clus}}), \quad \text{where } V_{\text{clus}} = \frac{1}{\pi^2} \operatorname*{plim}_{C\to\infty} \frac{1}{N}\sum_c \left(\sum_{n\in c} (g_n - \mathbb{E}[g])\,\xi_n\right)^2,$$

with a corresponding sample analog.

Example. To understand ξ_n better, consider the examiner design without controls, where each examiner n is randomly assigned to one of N pre-determined groups, denoted L_n, of equal size $K = L/N$. In that case, $\xi_n = \frac{1}{K}\sum_{\ell \in L_n} \epsilon_\ell - \frac{1}{L}\sum_\ell \epsilon_\ell$, so $\mathbb{E}[\xi_n] = \phi_n$ (recall $\phi_n = \mathbb{E}[s_{\ell n}\epsilon_\ell] / \mathbb{E}[s_{\ell n}]$) and $\text{Var}[\xi_n] = \frac{1}{K}\sigma_{\epsilon n}^2 + \frac{1}{L}\left(\sigma_\epsilon^2 - 2\sigma_{\epsilon n}^2\right)$,where $\sigma_{\epsilon n}^2 = \text{Var}[\epsilon_\ell \mid \ell \in L_n]$ and $\sigma_\epsilon^2 = \text{Var}[\epsilon_\ell]$. Therefore,

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_n \xi_n^2 = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_n \left(\frac{1}{K}\sigma_{\epsilon n}^2 + \phi_n^2\right).$$

The two terms in parentheses reflect two reasons why growth rates g_n may not be exactly orthogonal to ξ_n, which in this case equals the average error terms of locations specializing in industry n. First, since the sample of *units* is finite, ξ_n differs from its population average ϕ_n, which characterizes unobserved shocks the group is exposure to. When the number of units per examiner K is small, this estimation error is substantial. Second, with the finite number of *examiners*, quasi-experimentally drawn growth rates g_n may happen to be correlated with ϕ_n. For both reasons, the estimator converges at rate $\sqrt{N}$. The only exception is when $K \to \infty$ (i.e., the number of examiners stays constant or grows slowly) and $\phi_n = 0$ (i.e., examiner dummies are valid instruments). This is the situation considered by Goldsmith-Pinkham et al. (2018).

B Controlling for Component Observables

In this section we show that the resulting shift-share IV estimator which adds $\sum_n s_{\ell n} q_n$ as a control is equivalent to a two-step procedure in which the residuals g_n^* are first estimated by a matrix-weighted regression of g_n on q_n and then used to construct a new shift-share instrument $\hat{z}_\ell^* = \sum_n s_{\ell n} \hat{g}_n^*$.

Suppose a researcher estimates a cross-sectional shift-share IV regression that controls for an exposure-weighted vector of component-level controls, $\sum_{n=1}^{n} s_{\ell n} q_n$. By the Frisch-Waugh-Lovell theorem, the estimator can be written

$$\hat{\beta} = \frac{\sum_{\ell=1}^{L} \hat{z}_\ell^* y_\ell}{\sum_{\ell=1}^{L} \hat{z}_\ell^* x_\ell}, \tag{18}$$

where $\hat{z}_\ell^*$ is an OLS estimate of the residual of the auxiliary regression of the instrument on the controls:

$$z_\ell = \mu + \sum_{n=1}^{N} s_{\ell n} q_n' \tau + z_\ell^*. \tag{19}$$

Since $\sum_{n=1}^{N} s_{\ell n} = 1$, this regression can be written

$$\sum_{n=1}^{N} s_{\ell n} g_n = \sum_{n=1}^{N} s_{\ell n} \mu + \sum_{n=1}^{N} s_{\ell n} q_n' \tau + \sum_{n=1}^{N} s_{\ell n} z_\ell^*. \tag{20}$$

Let s be the $L \times N$ matrix collecting observations of $s_{\ell n}$, g be the $N \times 1$ vector stacking the g_n, and $\bar{q}$ be a matrix with rows $[1, q_n']$. Then we can write ordinary least square estimates of the parameters of (19) as

$$\begin{aligned} (\hat{\mu}, \hat{\tau})' &= \left((s\bar{q})' s\bar{q}\right)^{-1} (s\bar{q})' sg \\ &= \left(\bar{q}'(s's)\bar{q}\right)^{-1} \bar{q}' (s's)\, g, \end{aligned} \tag{21}$$

which is a matrix-weighted projection of g_n on q_n, with weight matrix $s's$. Thus

$$\begin{aligned} \hat{z}_\ell^* &= z_\ell - \hat{\mu} - \sum_{n=1}^{N} s_{\ell n} q_n' \hat{\tau} \\ &= \sum_{n=1}^{N} s_{\ell n} \left(g_n - \hat{\mu} - q_n' \hat{\tau}\right). \end{aligned} \tag{22}$$

This shows that $\hat{\beta}$ is equivalent to a shift-share IV regression coefficient that uses a modified aggregate shock $g_n - \hat{\mu} - q_n' \hat{\tau}$. This modified shock reflects the residual from the component-level projection (21).

C Shift-Share Instruments in Panels

In this section we consider a panel extension of the cross-sectional shift-share IV setting. We derive the orthogonality condition and quasi-experimental assumptions, paralleling Sections 2 and 3. We also show how these conditions could be relaxed by including unit and time fixed effects or by making assumptions on the stochastic process for aggregate shocks, and propose a simple pre-trend test.

Suppose we observe T repeated observations t of individual outcomes, treatment, exposure, and shocks over time. We continue with a constant effects model for outcomes and treatment, but now decompose the structural error term into a fixed location component and its residual: $\epsilon_{\ell t} = \alpha_\ell + \nu_{it}$ and

$$y_{\ell t} = \gamma + x_{\ell t}\beta + \alpha_\ell + \nu_{\ell t}, \tag{23}$$

where α_ℓ denotes the individual-specific mean of $\epsilon_{\ell t}$. Rather than estimating (23) by OLS, we construct a time-varying shift-share instrument as before

$$z_{\ell t} = \sum_{n=1}^{N} s_{\ell t n} g_{nt}, \tag{24}$$

where g_{nt} is now the shock to component n in time t. Note that this can be rewritten

$$z_{\ell t} = \sum_{n=1}^{N}\sum_{p=1}^{T} s_{\ell t n p} g_{np}, \tag{25}$$

where here $s_{\ell t n p} = s_{\ell t n}\mathbf{1}[n = p]$ denotes the exposure of location ℓ in time t to component n in time p, which is set to zero for $n \neq p$.

With this expanded share notation, the key orthogonality condition for a the validity of a fixed effects shift-share IV regression is

$$\begin{aligned}\text{Cov}\left[z_{\ell t}, \nu_{\ell t}\right] &= \sum_{n=1}^{N}\sum_{t=1}^{T} s_{nt} g_{nt} \phi_{nt} \\ &\to 0, \end{aligned} \tag{26}$$

where $s_{np} = \mathbb{E}\left[s_{\ell t n p}\right]$ and $\phi_{np} = \mathbb{E}\left[s_{\ell t n p}\nu_{\ell t}\right]/\mathbb{E}\left[s_{\ell t n p}\right]$. The right-hand side of the first line of (26) is again a weighted covariance of the aggregate shocks g_{nt}, now time-varying, and a time-varying measure of relevant unobservables ϕ_{nt}. Per above, these measures reflect a weighted average of only the time-varying component of structural residuals, $\nu_{\ell t}$, due to the inclusion of location fixed effect controls.

How do the quasi-experimental assumptions A1 and A2 map to the panel setting? Without further controls, the mean-independence condition becomes $\mathbb{E}\left[g_{nt} \mid \phi_{nt}\right] = \mu$, so that A1 is satisfied if the time varying shocks are mean-independent of the time-varying component unobservables. Importantly however shocks here need not be as-good-as-randomly assigned with respect to an exposure-weighted average of time-invariant heterogeneity, $\mathbb{E}\left[s_{\ell ntp}\alpha_\ell\right]/\mathbb{E}\left[s_{\ell tnp}\right]$. As before, A1 can be further weakened with the inclusion of component-by-time observables. In the panel setting, a natural choice is a set of time fixed effects; by the equivalence result in Appendix B, including time fixed effects in the shift-share IV regression allows the mean of the aggregate shocks to vary over periods.

In the panel setting, assumption A2 requires the time-varying shocks to be mutually mean-independent, conditional on $\left\{\{s_{nt}, \phi_{nt}\}_{n=1}^N\right\}_{t=1}^T$, and $\sum_{n=1}^N \sum_{t=1}^T s_{nt}^2 \to 0$. Note that in a balanced panel, $s_{np} = \mathbb{E}\left[s_{\ell tnp}\mathbf{1}[t=p]\right] = \mathbb{E}\left[s_{\ell tnp}\right]/T$. Thus the latter condition would be satisfied either when the number of periods T is fixed and $\max_{n,p} s_{np} \to 0$, or when N is fixed but $T \to \infty$; in long panels, shift-share IV may be consistent even with only a small number of components N.

The assumption of mutually mean-independent shocks in the panel setting rules out autoregressive shock processes: for example g_{nt} can not be conditionally correlated with $g_{n,t-1}$. One may imagine replacing assumption A2 to allow for strongly mixing or ergodic quasi-experimental shocks – we leave formalizing this approach for future work. Given a particular time series model for g_{nt}, however, one could easily use the earlier result on component-level controls to satisfy the current A2. For example, suppose the researcher assumes a first-order autoregressive process:

$$g_{nt} = \rho_0 + \rho_1 g_{n,t-1} + g_{nt}^*, \tag{27}$$

where the residuals g_{nt}^* are idiosyncratic. Then a researcher may choose to control for $\sum_n s_{\ell n,t} g_{n,t-1}$ in the panel IV specification.[18]

With assumption A2 holding either on the original shocks or their idiosyncratic residual, researchers can validate the panel identifying assumptions by testing for pre-trends. This entails correlating the outcome $y_{\ell t}$ with leads of the shift-share instrument, for example with $z_{\ell,t+1} = \sum_{n=1}^N s_{\ell n,t+1} g_{n,t+1}$. Under A1 and A2,

$$\text{Cov}\left[y_{\ell t}, z_{\ell,t+1}\right] = \sum_{n=1}^N g_{n,t+1}\text{Cov}\left[s_{\ell n,t+1}, y_{\ell t}\right] \to 0. \tag{28}$$

Validating this pre-trend condition is a special case of our test for assumption A1, where a lagged outcome is used as an observable proxy for the current-period error term.

[18] Jaeger et al. (2017) also consider autocorrelation-based inconsistency of panel shift-share IVs, arising from the dynamics of individual treatment response rather than of aggregate growth rates.

References

Adao, R., M. Kolesar, and E. Morales (2018): "Estimating Models with Spillovers: Formalizing Bartik Designs," Working Paper.

Angrist, J., G. Imbens, and A. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

Angrist, J. D. and A. B. Krueger (1995): "Split-Sample Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business and Economic Statistics*, 13, 225–235.

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

Autor, D. H. and M. G. Duggan (2003): "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics*, 118, 157–205.

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

Blanchard, O. J. and F. Katz (1992): "Regional Evolutions," *Brookings Papers on Economic Activity*, 1–75.

Borusyak, K. and X. Jaravel (2016): "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume," Working Paper.

Broxterman, D. A. and W. D. Larson (2018): "An Examination of Industry Mix Demand Indicators : The Bartik Instrument at Twenty-Five," Working Paper.

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

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, W. Schanzenbach, and D. Yagan (2011): "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 126, 1593–1660.

de Chaisemartin, C. and X. D'Haultfoeuille (2018): "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects," Working Paper.

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

Doyle, J., J. Graves, J. Gruber, and S. Kleiner (2015): "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy*, 123, 170–214.

Garin, A. and F. Silverio (2017): "How Does Firm Performance Affect Wages? Evidence from Idiosyncratic Export Shocks," Working Paper.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018): "Bartik Instruments: What, When, Why, and How," Working Paper.

Greenstone, M., A. Mas, and H.-L. Nguyen (2014): "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times," NBER Working Paper 20704.

Hull, P. (2018): "Estimating Treatment Effects in Mover Designs," Working Paper.

Hummels, D., R. Jorgensen, J. Munch, and C. Xiang (2014): "The Wage Effects of Offshoring: Evidence From Danish Matched Worker-Firm Data," *American Economic Review*, 104, 1597–1629.

Jaeger, D. A., J. Ruist, and J. Stuhler (2017): "Shift-Share Instruments and the Impact of Immigration," .

Jaravel, X. (2017): "The Unequal Gains from Product Innovations: Evidence from the US Retail Sector," Working Paper.

Kling, J. R. (2006): "Incarceration Length, Employment, and Earnings," *Ameri*, 96, 863–876.

Kolesar, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2015): "Identification and Inference With Many Invalid Instruments," *Journal of Business and Economic Statistics*, 33, 474–484.

Kovak, B. K. (2013): "Regional effects of trade reform: What is the correct measure of liberalization?" *American Economic Review*, 103, 1960–1976.

Luttmer, E. F. (2006): "Neighbors as negatives: relative earnings and well-being," *Quarterly Journal of Economics*, 120, 963–1002.

Maestas, N., K. J. Mullen, and A. Strand (2013): "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 103, 1797–1829.

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

Nakamura, E. and J. Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104, 753–792.

Nunn, N. and N. Qian (2014): "US Food Aid and Civil Conflict," *American Economic Review*, 104, 1630–1666.

Oberfield, E. and D. Raval (2014): "Micro Data and Macro Technology," .

Saiz, A. (2010): "The geographic determinants of housing supply," *Quarterly Journal of Economics*, 125, 1253–1296.

Suárez, J. C. and O. Zidar (2016): "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, 106, 2582–2624.