

# Quasi-Experimental Shift-Share Research Designs

Kirill Borusyak

Peter Hull

Xavier Jaravel

UCL and CEPR

UChicago and NBER

LSE and CEPR\*

February 2021

## Abstract

Many studies use shift-share (or “Bartik”) instruments, which average a set of shocks with exposure share weights. We provide a new econometric framework for shift-share instrumental variable (SSIV) regressions in which identification follows from the quasi-random assignment of shocks, while exposure shares are allowed to be endogenous. The framework is motivated by an equivalence result: the orthogonality between a shift-share instrument and an unobserved residual can be represented as the orthogonality between the underlying shocks and a shock-level unobservable. SSIV regression coefficients can similarly be obtained from an equivalent shock-level regression, motivating shock-level conditions for their consistency. We discuss and illustrate several practical insights of this framework in the setting of Autor et al. (2013), estimating the effect of Chinese import competition on manufacturing employment across U.S. commuting zones.

---

\*Contact: k.borusyak@ucl.ac.uk, hull@uchicago.edu, and x.jaravel@lse.ac.uk. We are grateful to Rodrigo Adão, Joshua Angrist, David Autor, Moya Chin, Andy Garin, Ed Glaeser, Paul Goldsmith-Pinkham, Larry Katz, Michal Kolesár, Gabriel Kreindler, Jack Liebersohn, Eduardo Morales, Jack Mountjoy, Jörn-Steffen Pischke, Brendan Price, Isaac Sorkin, Jann Spiess, Itzhak Tzachi Raz, various seminar participants, and five anonymous referees for helpful comments. We thank David Autor, David Dorn, and Gordon Hanson, as well as Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift, for providing replication code and data.

# 1 Introduction

A large and growing number of empirical studies use shift-share instruments: weighted averages of a common set of shocks, with weights reflecting heterogeneous shock exposure. In many settings, such as those of Bartik (1991), Blanchard and Katz (1992) and Autor et al. (2013), a regional instrument is constructed from shocks to industries with local industry employment shares measuring the shock exposure. In other settings, researchers may combine shocks across countries, socio-demographic groups, or foreign markets to instrument for treatments at the regional, individual, or firm level.<sup>1</sup>

The claim for instrument validity in shift-share instrumental variable (SSIV) regressions must rely on some assumptions about the shocks, exposure shares, or both. This paper develops a novel framework for understanding such regressions as leveraging exogenous variation in the shocks, allowing the variation in exposure shares to be endogenous. Our approach is motivated by an equivalence result: the orthogonality between a shift-share instrument and an unobserved residual can be represented as the orthogonality between the underlying shocks and a shock-level unobservable. Given a first stage, it follows that the instrument identifies a parameter of interest if and only if the shocks are uncorrelated with this unobservable, which captures the average unobserved determinants of the original outcome among observations most exposed to a given shock. SSIV regression coefficients can similarly be obtained from an equivalent IV regression estimated at the level of shocks. In this regression the outcome and treatment variables are first averaged, using exposure shares as weights, to obtain shock-level aggregates. The shocks then directly instrument for the aggregated treatment. Importantly, these equivalence results only rely on the structure of the shift-share instrument and thus apply to outcomes and treatments that are not typically computed at the level of shocks.

We use these equivalence results to derive two conditions sufficient for SSIV consistency. First, we assume shocks are as-good-as-randomly assigned as if arising from a natural experiment. This is enough for the shift-share instrument to be valid: i.e. for the shocks to be uncorrelated with the relevant unobservables in expectation. Second, we assume that a shock-level law of large numbers applies—that the instrument incorporates many sufficiently independent shocks, each with sufficiently small average exposure. Instrument relevance further holds when individual units are mostly exposed to only a small number of shocks, provided those shocks affect treatment. Our two quasi-experimental conditions are similar to ones imposed in other settings where the underlying shocks are directly used as instruments, bringing SSIV to familiar econometric territory.<sup>2</sup>

---

<sup>1</sup>Observations in shift-share designs may, for example, represent regions impacted by immigration shocks from different countries (Card 2001; Peri et al. 2016), firms differentially exposed to foreign market shocks (Hummels et al. 2014; Berman et al. 2015), product groups purchased by different types of consumers (Jaravel 2019), groups of individuals facing different national income trends (Boustan et al. 2013), or countries differentially exposed to the U.S. food aid supply shocks (Nunn and Qian 2014). We present a taxonomy of existing shift-share designs, and how they relate to our framework, in Section 6.1.

<sup>2</sup>For example, Acemoglu et al. (2016) study the impact of import competition from China on U.S. industry employment using industry (i.e. shock-level) regressions with shocks constructed similarly to those underlying the regional shift-share instrument used in Autor et al. (2013). Our framework shows that both studies can rely on similar econometric assumptions, though the economic interpretations of the estimates differ.

We extend our quasi-experimental approach to settings where shocks are as-good-as-randomly assigned only conditionally on shock-level observables, to SSIVs with exposure shares that do not add up to a constant for each observation, and to panel data. For conditional random assignment, we show that quasi-experimental shock variation can be isolated with regression controls that have a shift-share structure. Namely, it is enough to control for an exposure-weighted sum of the relevant shock-level confounders. Relatedly, in SSIVs with “incomplete shares,” where the sum of exposure shares varies across observations, it is important to control for the sum of exposure shares as the exposure-weighted sum of a constant. In panel data, we show that the SSIV estimator can be consistent both with many shocks per period and with many periods. We also show that unit fixed effects only isolate variation in shocks over time when exposure shares are time-invariant. In other extensions we show how SSIV with multiple endogenous variables can be viewed quasi-experimentally and how multiple sets of quasi-random shocks can be combined with new overidentified shock-level IV procedures.

Our framework also bears practical tools for SSIV inference and testing. Adão et al. (2019) show that conventional standard errors in SSIV regressions may be invalid because observations which similar exposure shares are likely to have correlated residuals. They are also the first to propose a solution to this inference problem in a framework based on ours, with identifying variation in shocks. We present a convenient alternative based on our equivalence result: estimating SSIV coefficients at the level of identifying variation (shocks) can yield asymptotically valid standard errors. The validity of this solution requires an additional assumption on the structure of the included controls (producing standard errors that are conservative otherwise). However, it offers several practical features: it can be implemented with standard statistical software, extended to various forms of shock dependence (e.g. autocorrelation), and computed in some settings where the estimator of Adão et al. (2019) fails (e.g. when there are more shocks than observations). Appropriate measures of first-stage relevance and valid falsification tests of shock exogeneity can also be obtained with conventional shock-level procedures. Monte-Carlo simulations confirm the accuracy of our asymptotic approximations in moderately-sized samples of shocks, and that the finite-sample properties of SSIV are similar to those of conventional shock-level IV regressions which use the same shocks as instruments.

We illustrate the practical insights from our framework in the setting of Autor et al. (2013), who estimate the effect of increased Chinese import penetration on manufacturing employment across U.S. commuting zones. We find supporting evidence for the interpretation of their SSIV as leveraging quasi-random variation in industry-specific Chinese import shocks. This application uses a new Stata package, *ssaggregate*, which we have developed to help practitioners implement the appropriate shock-level analyses.<sup>3</sup>

Our quasi-experimental approach is not the only framework for SSIV identification and consis-

---

<sup>3</sup>This Stata package creates the shock-level aggregates used in the equivalent regression. Users can install this package with the command `ssc install ssaggregate`. See the associated help file and this paper’s replication archive at <https://github.com/borussyak/shift-share> for more details.

tency. In related work, Goldsmith-Pinkham et al. (2020) formalize a different approach based on the exogeneity of the exposure shares, imposing no explicit assumption of shock exogeneity. This framework is motivated by a different equivalence result: the SSIV coefficient also coincides with a generalized method of moments estimator, with exposure shares as multiple excluded instruments. Though exposure exogeneity is a sufficient condition for SSIV identification (and, as such, implies our shock-level orthogonality condition), we focus on plausible conditions under which it is not necessary.

We delineate two cases where identification via exogenous shocks is attractive. In the first case, the shift-share instrument is based on a set of shocks which can itself be thought of as an instrument. Consider the Autor et al. (2013; hereafter ADH) shift-share instrument, which combines industry-specific changes in Chinese import competition (the shocks) with local exposure given by the lagged industrial composition of U.S. regions (the exposure shares). In such a setting, exogeneity of industry employment shares is difficult to justify *a priori* since unobserved industry shocks (e.g., automation or innovation trends) are likely to affect regional outcomes through the same mixture of exposure shares. Our approach, in contrast, allows researchers to specify a set of shocks that are plausibly uncorrelated with such unobserved factors. Consistent with this general principle, ADH attempt to purge their industry shocks from U.S.-specific confounders by measuring Chinese import growth outside of the United States. Similarly, Hummels et al. (2014) combine country-by-product changes in transportation costs to Denmark (as shocks) with lagged firm-specific composition of intermediate inputs and their sources (as shares). They argue these shocks are “idiosyncratic,” which our approach formalizes as “independent from relevant country-by-product unobservables.” Other recent examples of where our approach may naturally apply are found, for example, in finance (Xu 2019), the immigration literature (Peri et al. 2016), and studies of innovation (Stuen et al. 2012).

In the second case, a researcher can think of quasi-experimental shocks which are not observed directly but are instead estimated in-sample in an initial step, potentially introducing biases. In the canonical estimation of regional labor supply elasticities by Bartik (1991), for example, the shocks are measured as national industry growth rates. Such growth captures national industry labor demand shocks, which one may be willing to assume are as-good-as-randomly assigned across industries; however, industry growth rates also depend on unobserved regional labor supply shocks. We show that our framework can still apply to such settings by casting the industry employment growth rates as noisy estimates of latent quasi-experimental demand shocks and establishing conditions to ensure the supply-driven estimation error is asymptotically ignorable. These conditions are weaker if the latent shocks are estimated as leave-one-out averages. Although leave-one-out shift-share IV estimates do not have a convenient shock-level representation, we provide evidence that in the Bartik (1991) setting this leave-out adjustment is unimportant.

Formally, our approach to SSIV relates to the analysis of IV estimators with many invalid instruments by Kolesar et al. (2015). Consistency in that setting follows when violations of individual instru-

ment exclusion restrictions are uncorrelated with their first-stage effects. For quasi-experimental SSIV, the exposure shares can be thought of as a set of invalid instruments (per the Goldsmith-Pinkham et al. (2020) interpretation), and our orthogonality condition requires their exclusion restriction violations to be uncorrelated with the shocks. Despite this formal similarity, we argue that shift-share identification is better understood through the quasi-random assignment of a single instrument (shocks), rather than through a large set of invalid instruments (exposure shares) that nevertheless produce a consistent estimate. This view is reinforced by our equivalence results, yields a natural shock-level identification condition, and suggests new validations and extensions of SSIV.

Our analysis also relates to other recent methodological studies of shift-share designs, including those of Jaeger et al. (2018) and Broxterman and Larson (2018). The former highlights biases of SSIV due to endogenous local labor market dynamics, and we show how their solution can be implemented in our setting. The latter studies the empirical performance of different shift-share instrument constructions. As discussed above we also draw on the inferential framework of Adão et al. (2019), who derive valid standard errors in shift-share designs with a large number of idiosyncratic shocks. More broadly, our paper adds to a growing literature studying the causal interpretation of common research designs, including work by Borusyak and Jaravel (2017), Goodman-Bacon (2018), Sun and Abraham (Forthcoming), and Callaway and Sant’Anna (Forthcoming) for event study designs; Chaisemartin and D’Haultfoeuille (2019) for two-way fixed effects regressions; Słoczyński (Forthcoming) for regressions with other controls; and Hull (2018) for mover designs.

The remainder of this paper is organized as follows. Section 2 introduces the environment, derives our equivalence results, and motivates our approach to SSIV identification and consistency. Section 3 establishes the baseline quasi-experimental assumptions and Section 4 derives various extensions. Section 5 discusses shock-level procedures for valid SSIV inference and testing. Section 6 summarizes the types of empirical settings where our framework may be applied and illustrates its practical implications in the ADH setting. Section 7 concludes.

## 2 Setting and Motivation

We begin by presenting the SSIV setting and motivating our approach to identification and consistency with two equivalence results. We first show that population orthogonality of the shift-share instrument can be recast at the shock level, motivating identification by exogenous shocks when exposure shares are endogenous. We then derive a similar shock-level equivalence result for the SSIV estimator, motivating its consistency with many as-good-as-randomly assigned shocks.

## 2.1 The Shift-Share IV Setting

We observe an outcome  $y_\ell$ , treatment  $x_\ell$ , control vector  $w_\ell$  (which includes a constant) and shift-share instrument  $z_\ell$  for a set of observations  $\ell = 1, \dots, L$ . We also observe a set of regression weights  $e_\ell > 0$  with  $\sum_\ell e_\ell = 1$  ( $e_\ell = \frac{1}{L}$  covers the unweighted case). The instrument can be written as

$$z_\ell = \sum_n s_{\ell n} g_n, \quad (1)$$

for a set of observed shocks  $g_n$ ,  $n = 1, \dots, N$ , and a set of observed shares  $s_{\ell n} \geq 0$  defining the exposure of each observation  $\ell$  to each shock  $n$ . Initially we assume the sum of these exposure weights is constant across observations, i.e. that  $\sum_n s_{\ell n} = 1$ ; we relax this assumption in Section 4.2.<sup>4</sup> Although our focus is on shift-share IV, we note that the setup nests shift-share reduced-form regressions, of  $y_\ell$  on  $z_\ell$  and  $w_\ell$ , when  $x_\ell = z_\ell$ .

We seek to estimate the causal effect or structural parameter  $\beta$  in a linear model of

$$y_\ell = \beta x_\ell + w'_\ell \gamma + \varepsilon_\ell, \quad (2)$$

where the residual  $\varepsilon_\ell$  is defined to be orthogonal with the control vector  $w_\ell$ .<sup>5</sup> For example, we might be interested in estimating a classic model of labor supply which relates observations of log wage growth  $y_\ell$  and log employment growth  $x_\ell$  across local labor markets  $\ell$  by an inverse labor supply elasticity  $\beta$ . The residual  $\varepsilon_\ell$  in equation (2) would then contain all local labor supply shocks, such as those arising from demographic, human capital, or migration changes, that are not systematically related to the observed controls in  $w_\ell$ .<sup>6</sup> To estimate  $\beta$  we require an instrument capturing variation in local labor demand.

We consider a  $z_\ell$  based on the introduction of new import tariffs  $g_n$  across different industries  $n$ , with  $s_{\ell n}$  denoting location  $\ell$ 's lagged shares of industry employment.<sup>7</sup> In estimating  $\beta$  we may weight observations by the overall lagged regional employment,  $e_\ell$ . We return to this labor supply example at several points to ground the following theoretic discussion.

<sup>4</sup>Note that the shares are defined relative to the total across components  $n$ . In practice shift-share instruments are sometimes presented differently, with the shares defined relative to the total across observations  $\ell$  (see footnote 39 for an example in the Autor et al. (2013) setting). We recommend that researchers follow the representation in (1) to apply our theoretical results.

<sup>5</sup>Formally, given a linear causal or structural model of  $y_\ell = \beta x_\ell + \varepsilon_\ell$  we define  $\gamma = \mathbb{E} [\sum_\ell e_\ell w_\ell w'_\ell]^{-1} \mathbb{E} [\sum_\ell e_\ell w_\ell \varepsilon_\ell]$  and  $\varepsilon_\ell = \varepsilon_\ell - w'_\ell \gamma$  as the residual from this population projection, satisfying  $\mathbb{E} [\sum_\ell e_\ell w_\ell \varepsilon_\ell] = 0$ . Defining a unique  $\gamma$  requires an implicit maintained assumption that  $\mathbb{E} [\sum_\ell e_\ell w_\ell w'_\ell]$  is of full rank, which holds when there is no perfect collinearity in the control vector. We consider models with heterogeneous treatment effects in Appendix A.1; see footnote 16 for a summary.

<sup>6</sup>While this simple labor supply equation is only well-defined under certain assumptions (for instance, it rules out wage bargaining and profit sharing between firms and workers), it is a standard modeling tool. We note that it is inconsequential whether wages or employment are on the right-hand side of the second stage regression; we choose wage growth as the outcome following the tradition of Bartik (1991).

<sup>7</sup>Import tariffs affect import prices, consumer demand for domestic products, and in turn labor demand. Recent studies illustrate that one can obtain quasi-random identifying variation in import tariffs in practice. Changes in import tariffs across industries have been used for identification in both industry-level analyses (e.g. Fajgelbaum et al. (2020)) and shift-share analyses of regional outcomes (e.g. Kovak (2013)).

It is worth highlighting that in studying this setting we do not impose a typical assumption of independent and identically-distributed (*iid*) data  $\{e_\ell, z_\ell, w_\ell, x_\ell, \varepsilon_\ell\}$ , as might arise from random sampling of potential observations. Relaxing the usual *iid* assumption is required for us to treat the  $g_n$  as random variables, which generate dependencies of the instrument (1) across observations exposed to the same random shocks. The non-*iid* setting further allows for unobserved common shocks, which may generate dependencies in the residual  $\varepsilon_\ell$ .

Given this non-*iid* setting, we consider IV identification of  $\beta$  by the full-data moment condition

$$\mathbb{E} \left[ \sum_{\ell} e_{\ell} z_{\ell} \varepsilon_{\ell} \right] = 0. \quad (3)$$

This condition captures the orthogonality of the shift-share instrument with the second-stage residual, in expectation over realizations of  $\{e_\ell, z_\ell, \varepsilon_\ell\}$  for all  $\ell = 1, \dots, L$ . When such orthogonality holds the  $\beta$  parameter is identified: i.e., uniquely recoverable from observable moments, provided the instrument has a first stage.<sup>8</sup> The full-data orthogonality condition generalizes the conventional condition of  $\mathbb{E}[z_\ell \varepsilon_\ell] = 0$ , which might be considered in an *iid* setting with fixed  $e_\ell$ .

The moment condition (3) yields a natural estimator of  $\beta$ : the coefficient on  $x_\ell$  in an IV regression of  $y_\ell$  which instruments by  $z_\ell$ , controls for  $w_\ell$ , and weights by  $e_\ell$ . By the Frisch-Waugh-Lovell theorem, this SSIV estimator can be represented as a bivariate IV regression of outcome and treatment residuals, or as the ratio of  $e_\ell$ -weighted sample covariances between the instrument and the residualized outcome and treatment:

$$\hat{\beta} = \frac{\sum_{\ell} e_{\ell} z_{\ell} y_{\ell}^{\perp}}{\sum_{\ell} e_{\ell} z_{\ell} x_{\ell}^{\perp}}, \quad (4)$$

where  $v_{\ell}^{\perp}$  denotes the residual from an  $e_{\ell}$ -weighted sample projection of a variable  $v_{\ell}$  on the control vector  $w_{\ell}$ . Note that by the properties of residualization, it is enough to residualize  $y_{\ell}$  and  $x_{\ell}$  without also residualizing the shift-share instrument  $z_{\ell}$ .

In our non-*iid* setting, we study consistency and other asymptotic properties of  $\hat{\beta}$  by considering a sequence of data-generating processes, indexed by  $L$ , for the complete data  $\{e_{\ell}, s_{\ell n}, g_n, w_{\ell}, x_{\ell}, \varepsilon_{\ell}\}$ , for  $\ell = 1, \dots, L$ ,  $n = 1 \dots, N$ , and  $N = N(L)$ . Consistency, for example, is defined as  $\hat{\beta} \xrightarrow{P} \beta$  as  $L \rightarrow \infty$  along this sequence. We do not employ conventional sampling-based asymptotic sequences (and corresponding laws of large numbers) as these are generally inappropriate in a non-*iid* setting where both  $z_{\ell}$  and  $\varepsilon_{\ell}$  may exhibit non-standard mutual dependencies. It is worth emphasizing that any assumptions on the data-generating sequence are useful only to approximate the finite-sample distribution of the SSIV estimator, not to define an actual process for realizations of the data. For example, we will consider below a sequence in which the number of shocks  $N$  grows with  $L$ , recognizing that in reality shift-share instruments are constructed from a fixed set of shocks (e.g. tariffs across all

---

<sup>8</sup>Formally, when equation (3) holds the moment condition  $m(b, c) \equiv \mathbb{E}[\sum_{\ell} e_{\ell}(z_{\ell}, w'_{\ell})(y_{\ell} - bx_{\ell} - w'_{\ell}c)] = 0$  has a unique solution of  $(\beta, \gamma)$ , provided  $\mathbb{E}[\sum_{\ell} e_{\ell}(z_{\ell}, w'_{\ell})(x_{\ell}, w'_{\ell})]$  is of full-rank.

industries) along with a fixed number of observations (e.g. all local labor markets). The assumption of growing  $N$  should here be interpreted as a way to capture the presence of a large number of shocks in a given set of observations, such that the asymptotic sequence provides a good approximation to the observed data.<sup>9</sup>

## 2.2 A Shock-Level Orthogonality Condition

We first build intuition for our approach to satisfying the IV moment condition by showing that the structure of the shift-share instrument allows equation (3) to be rewritten as condition on the orthogonality of shocks  $g_n$ . Namely, by exchanging the order of summation across  $\ell$  and  $n$ , we obtain

$$\mathbb{E} \left[ \sum_{\ell} e_{\ell} z_{\ell} \varepsilon_{\ell} \right] = \mathbb{E} \left[ \sum_{\ell} e_{\ell} \sum_n s_{\ell n} g_n \varepsilon_{\ell} \right] = \mathbb{E} \left[ \sum_n s_n g_n \bar{\varepsilon}_n \right] = 0, \quad (5)$$

where we define  $s_n = \sum_{\ell} e_{\ell} s_{\ell n}$  and  $\bar{\varepsilon}_n = \frac{\sum_{\ell} e_{\ell} s_{\ell n} \varepsilon_{\ell}}{\sum_{\ell} e_{\ell} s_{\ell n}}$ . Just as the left-hand side of this expression captures the orthogonality of the instrument  $z_{\ell}$  with the residual  $\varepsilon_{\ell}$  when weighted by  $e_{\ell}$ , the right-hand side captures the orthogonality of shocks  $g_n$  and  $\bar{\varepsilon}_n$  when weighted by  $s_n$ . Since these two expressions are equivalent, equation (5) shows that such shock orthogonality is necessary and sufficient condition for the orthogonality of the shift-share instrument. As with  $e_{\ell}$ , the shock-level weights are also non-negative and sum to one, since  $\sum_n s_n = \sum_{\ell} e_{\ell} (\sum_n s_{\ell n}) = 1$ . The shock-level unobservables  $\bar{\varepsilon}_n$  represent exposure-weighted averages of the residuals  $\varepsilon_{\ell}$ .

The labor supply example is useful for unpacking this first equivalence result. When  $s_{\ell n}$  are lagged employment shares and  $e_{\ell}$  are similarly lagged regional employment weights, the  $s_n$  weights are proportional to the lagged industry employment.<sup>10</sup> Moreover, with  $\varepsilon_{\ell}$  capturing unmeasured supply shocks,  $\bar{\varepsilon}_n$  is the average unobserved supply shock among regions  $\ell$  that are the most specialized in industry  $n$ , in terms of their lagged employment  $e_{\ell} s_{\ell n}$ . Equation (5) then shows that for the shift-share instrument  $z_{\ell}$  to identify the labor supply elasticity  $\beta$ , the industry demand shocks  $g_n$  must be orthogonal with these industry-level unobservables when weighted by industry size. For example, the industries which experience a rise in import tariffs should not face systematically different unobserved labor supply conditions (e.g., migration patterns) in their primary markets.

Shock orthogonality is a necessary condition for SSIV identification and is satisfied when, as in the preferred interpretation of Goldsmith-Pinkham et al. (2020), the exposure shares are exogenous, the data are *iid*, and the shocks are considered non-random.<sup>11</sup> In practice, however, this approach to SSIV identification may be untenable in many settings. In our labor supply example, the Goldsmith-Pinkham et al. (2020) approach to identification requires the (lagged) local employment share of each

<sup>9</sup>This is similar to how Bekker (1994) uses a non-standard asymptotic sequence to analyze IV estimators with many instruments: “The sequence is designed to make the asymptotic distribution fit the finite sample distribution better. It is completely irrelevant whether or not further sampling will lead to samples conforming to this sequence” (p. 658).

<sup>10</sup>Without regression weights (i.e.  $e_{\ell} = \frac{1}{L}$ ),  $s_n$  is instead the average employment share of industry  $n$  across locations.

<sup>11</sup>Formally, in this framework  $\mathbb{E}[e_{\ell} s_{\ell n} \varepsilon_{\ell}] = 0$  for each  $(\ell, n)$ , so  $\mathbb{E}[\sum_n s_n g_n \bar{\varepsilon}_n] = \sum_n g_n \sum_{\ell} \mathbb{E}[e_{\ell} s_{\ell n} \varepsilon_{\ell}] = 0$ .

industry to be a valid instrument in the labor supply equation, i.e. uncorrelated with all unobserved labor supply shocks. This assumption is unlikely to hold: changes in foreign immigration, for example, are a type of local labor supply shock which is likely related to the local industry composition (e.g., new immigrants may prefer to settle in areas with larger clusters of specific industries, such as high-tech, even conditionally on the prevailing wage). Formally, whenever the second-stage error term has a component with the shift-share structure,  $\sum_n s_{\ell n} \nu_n$  for unobserved shocks  $\nu_n$ , then the exposure shares will be mechanically endogenous even if the  $\nu_n$  and  $g_n$  are uncorrelated (see Appendix A.2 for a proof).<sup>12</sup>

When shares are endogenous, equation (5) suggests that identification may instead follow from the exogeneity of shocks. We formalize this approach in Section 3.1, by specifying a quasi-experimental design in which the  $g_n$  are as-good-as-randomly assigned with respect to the other terms in the expression. We show how this simple exogeneity can be relaxed with controls in Section 3.2.

### 2.3 Estimator Equivalence

We next build intuition for our approach to SSIV consistency by showing that the estimate  $\hat{\beta}$  is equivalently obtained as the coefficient from a non-standard shock-level IV procedure, in which  $g_n$  directly serves as the instrument. This equivalence result suggests that the large-sample properties of  $\hat{\beta}$  can be derived from a law of large numbers for the equivalent shock-level regression. An attractive feature of this approach is that it does not rely on an assumption of *iid* observations, which can be untenable in the presence of observed and unobserved  $n$ -level shocks. We instead place assumptions on the assignment of the equivalent IV regression’s instrument  $g_n$ , similar to a more standard analysis of a randomized treatment in an experimental settings (Abadie et al. 2019).

Formally, we have the following equivalence result:

**Proposition 1** The SSIV estimator  $\hat{\beta}$  equals the second-stage coefficient from a  $s_n$ -weighted shock-level IV regression that uses the shocks  $g_n$  as the instrument in estimating

$$\bar{y}_n^\perp = \alpha + \beta \bar{x}_n^\perp + \bar{\varepsilon}_n^\perp, \quad (6)$$

where  $\bar{v}_n = \frac{\sum_\ell e_\ell s_{\ell n} v_\ell}{\sum_\ell e_\ell s_{\ell n}}$  denotes an exposure-weighted average of variable  $v_\ell$ .

PROOF By definition of  $z_\ell$ ,

$$\hat{\beta} = \frac{\sum_\ell e_\ell (\sum_n s_{\ell n} g_n) y_\ell^\perp}{\sum_\ell e_\ell (\sum_n s_{\ell n} g_n) x_\ell^\perp} = \frac{\sum_n g_n (\sum_\ell e_\ell s_{\ell n} y_\ell^\perp)}{\sum_n g_n (\sum_\ell e_\ell s_{\ell n} x_\ell^\perp)} = \frac{\sum_n s_n g_n \bar{y}_n^\perp}{\sum_n s_n g_n \bar{x}_n^\perp}. \quad (7)$$

<sup>12</sup>In the immigration example,  $\nu_n$  is positive in high-tech industries and negative in industries that do not attract immigrants. Note that the same argument applies to migration flows within the U.S., which can similarly make local labor supply shocks related to local industry composition. Lagging local employment shares does not alleviate these threats to identification in general.

Furthermore  $\sum_n s_n \bar{y}_n^\perp = \sum_\ell e_\ell (\sum_n s_{\ell n}) y_\ell^\perp = \sum_\ell e_\ell y_\ell^\perp = 0$ , since  $y_\ell^\perp$  is an  $e_\ell$ -weighted regression residual and  $\sum_n s_{\ell n} = 1$ . This and an analogous equality for  $\bar{x}_n^\perp$  imply that (7) is a ratio of  $s_n$ -weighted covariances, of  $\bar{y}_n^\perp$  and  $\bar{x}_n^\perp$  with  $g_n$ . Hence it is obtained from the specified IV regression.

As with equation (5), Proposition 1 exploits the structure of the instrument to exchange orders of summation in the expression for the SSIV estimator (4). This exchange shows that SSIV estimates can also be thought to arise from variation across shocks, rather than across observations. The equivalent IV regression uses the shocks  $g_n$  directly as the instrument and shock-level aggregates of the original (residualized) outcome and treatment,  $\bar{y}_n^\perp$  and  $\bar{x}_n^\perp$ . Specifically,  $\bar{y}_n^\perp$  reflects the average residualized outcome of the observations most exposed to the  $n$ th shock, while  $\bar{x}_n^\perp$  is the same weighted average of residualized treatment. The regression is weighted by  $s_n$ , representing each shock’s average exposure across the observations.<sup>13</sup>

The fact that shift-share estimates can be equivalently obtained by a shock-level IV procedure suggests a new approach to establishing their consistency. Generally, IV regressions of the form of (7) will be consistent when the instrument (here,  $g_n$ ) is as-good-as randomly assigned, there is a large the number of observations (here,  $N$ ), the importance weights are sufficiently dispersed (here, that the  $s_n$  are not too skewed), and there is an asymptotic first stage. Consistency is then guaranteed regardless of the correlation structure of the residuals  $\bar{\varepsilon}_n^\perp$ , and thus in the primitive residuals  $\varepsilon_\ell$  and exposure shares  $s_{\ell n}$ . We formalize this approach below.

Before proceeding, it is worth emphasizing that while the shock-level IV regression from Proposition 1 motivates our approach to *identification* of  $\beta$ , it does not affect the *interpretation* of the coefficient as measuring an  $\ell$ -level relationship. The shock-level equation (6), in which the outcome and treatment are unconventional shock-level objects, does not have independent economic content. For example, in the labor supply setting  $\bar{y}_n$  is not industry  $n$ ’s wage growth; rather, it measures the average wage growth in regions where industry  $n$  employs the most workers. Thus, while  $\hat{\beta}$  can be computed at the industry level it estimates the elasticity of regional, rather than industry, labor supply, and could, for example, capture the kinds of local spillovers that a regression of industry wages on industry employment cannot.<sup>14</sup> Furthermore, Proposition 1 holds even for the outcomes and treatments which cannot be naturally computed at the shock level, e.g. when  $n$  indexes industries and  $y_\ell$  measures labor force non-participation, as in Autor et al. (2013).

<sup>13</sup>In the special case of reduced-form shift-share regressions, Proposition 1 shows that the equivalent shock-level procedure is still an IV regression, of  $\bar{y}_n^\perp$  on the transformed shift-share instrument  $\bar{z}_n^\perp$ , again instrumented by  $g_n$  and weighted by  $s_n$ .

<sup>14</sup>In Appendix A.3 we develop a stylized model to illustrate how the SSIV coefficient can differ from a “native” shock-level IV coefficient in the presence of local spillovers or treatment effect heterogeneity, though both parameters may be of interest. Intuitively, in the labor supply case one may estimate a low regional elasticity but a high elasticity of industry labor supply if, for example, migration is constrained but workers are mobile across industries within a region.

### 3 A Quasi-Experimental SSIV Framework

We now show how SSIV identification and consistency can be satisfied by a quasi-experiment in which shocks are as-good-as-randomly assigned, mutually uncorrelated, large in number, and sufficiently dispersed in terms of their average exposure. Instrument relevance generally holds in such settings when the exposure of individual observations tends to be concentrated in a small number of shocks, and when those shocks affect treatment. We then show how this framework is naturally generalized to settings in which shocks are only conditionally quasi-randomly assigned or exhibit some forms of mutual dependence, such as clustering.

#### 3.1 Quasi-Randomly Assigned and Mutually Uncorrelated Shocks

Our approach to SSIV consistency is based on a thought experiment in which the shocks  $g_n$  are as-good-as-randomly assigned conditional on the shock-level unobservables  $\bar{\varepsilon}_n$  and exposure weights  $s_n$ . As motivated above, placing assumptions on this assignment process (rather than on the sampling properties of observations) has two key advantages. First, we do not rely on conventional assumptions of independent or clustered data which are generally inconsistent with the shift-share data structure when the shocks are considered random variables. Second, in conditioning on  $\bar{\varepsilon} = \{\bar{\varepsilon}_n\}_n$  and  $s = \{s_n\}_n$  we place no restrictions on the dependence between the  $s_{\ell n}$  and  $\varepsilon_\ell$ , allowing shock exposure to be endogenous. We first show that such endogeneity need not pose problems for SSIV identification:

**Proposition 2** The SSIV moment condition (3) is satisfied by the following condition:

**Assumption 1** (*Quasi-random shock assignment*):  $\mathbb{E}[g_n | \bar{\varepsilon}, s] = \mu$ , for all  $n$ .

PROOF By equation (5) and the law of iterated expectations,  $\mathbb{E}[\sum_\ell e_\ell z_\ell \varepsilon_\ell] = \mathbb{E}[\sum_n s_n g_n \bar{\varepsilon}_n] = \mu \cdot \mathbb{E}[\sum_n s_n \bar{\varepsilon}_n]$  under Assumption 1. Furthermore, since  $\sum_\ell s_{\ell n} = 1$  and  $\mathbb{E}[\sum_\ell e_\ell \varepsilon_\ell] = 0$  by construction,  $\mathbb{E}[\sum_n s_n \bar{\varepsilon}_n] = \mathbb{E}[(\sum_n s_{\ell n}) (\sum_\ell e_\ell \varepsilon_\ell)] = 0$ .

Proposition 2 shows that the shift-share instrument is valid, in that the IV moment condition (3) holds, when the underlying shocks are as-good-as-randomly assigned: each  $g_n$  has the same mean  $\mu$ , regardless of the realizations of the relevant unobservables  $\bar{\varepsilon}$  (and average exposures  $s$ ). In the labor supply example this assumption would mean that import tariffs should not have been chosen strategically, based on labor supply trends, or in a way that is correlated with such trends.<sup>15</sup>

It follows from Proposition 2 that  $\beta$  is identified by Assumption 1 provided the instrument is relevant.<sup>16</sup> In practice, the existence of a non-zero first stage can be inferred from the data; we discuss

<sup>15</sup>For example, if labor supply trends differ between regions specializing in manufacturing vs. services, the import tariffs should apply to both types of sectors. We discuss in Section 4.2 how to apply our framework in the case where import tariffs only apply to a subsector of the economy, e.g. in manufacturing only.

<sup>16</sup>Appendix A.1 shows how SSIV identifies a convex average of heterogeneous treatment effects (varying potentially across both  $\ell$  and  $n$ ) under a stronger notion of as-good-as-random shock assignment and a first-stage monotonicity condition. This can be seen as generalizing both the IV identification result of Angrist et al. (2000) to shift-share instruments, as well as the reduced-form shift-share identification result in Adão et al. (2019).

appropriate inferential techniques in Section 5. To illustrate how such instrument relevance might hold with quasi-experimental shocks, we consider a simple first-stage model. Consider a setting without controls ( $w_\ell = 1$ ) and where treatment is a share-weighted average of shock-specific components:  $x_\ell = \sum_n s_{\ell n} x_{\ell n}$ , where  $x_{\ell n} = \pi_{\ell n} g_n + \eta_{\ell n}$  with  $\pi_{\ell n} \geq \bar{\pi}$  almost surely for some fixed  $\bar{\pi} > 0$ . In line with Assumption 1, suppose that the shocks are independent mean-zero, given the full set of exposure shares  $s_{\ell n}$  and regression weights  $e_\ell$  and the full set of  $\pi_{\ell n}$  and  $\eta_{\ell n}$ , with variances that are bounded below by some fixed  $\bar{\sigma}_g^2 > 0$ . Then the instrument first stage is positive:

$$\begin{aligned} \mathbb{E} \left[ \sum_\ell e_\ell z_\ell x_\ell \right] &= \mathbb{E} \left[ \sum_\ell e_\ell \left( \sum_n s_{\ell n} g_n \right) \left( \sum_n s_{\ell n} (\pi_{\ell n} g_n + \eta_{\ell n}) \right) \right] \\ &\geq \bar{\pi} \bar{\sigma}_g^2 \mathbb{E} \left[ \sum_\ell e_\ell \sum_n s_{\ell n}^2 \right] > 0. \end{aligned} \quad (8)$$

Given identification, SSIV consistency follows from an appropriate law of large numbers. Motivated by the estimator equivalence in Section 2.3, we consider settings in which the effective sample size of the shock-level IV regression (6) is large and the observations of the effective instrument (shocks) are mutually uncorrelated:

**Assumption 2** (*Many uncorrelated shocks*):  $\mathbb{E} [\sum_n s_n^2] \rightarrow 0$  and  $\text{Cov} [g_n, g_{n'} \mid \bar{\varepsilon}, s] = 0$  for all  $(n, n')$  with  $n' \neq n$ .

The first part of Assumption 2 states that the expected Herfindahl index of average shock exposure,  $\mathbb{E} [\sum_n s_n^2]$ , converges to zero as  $L \rightarrow \infty$ . This condition implies that the number of observed shocks grows with the sample (since  $\sum_n s_n^2 \geq 1/N$ ), and can be interpreted as requiring a large effective sample for the equivalent shock-level IV regression. An equivalent condition is that the largest importance weight in this regression,  $s_n$ , becomes vanishingly small.<sup>17</sup> The second part of Assumption 2 states that the shocks are mutually uncorrelated given the unobservables and  $s_n$ . Both of these conditions, while novel for SSIV, would be standard assumptions to establish the consistency of a conventional shock-level IV estimator with  $g_n$  as the instrument and  $s_n$  weights.

Assumptions 1 and 2 are the baseline assumptions of our quasi-experimental framework. Given a standard relevance condition and additional regularity conditions listed in Appendix B.1, they are sufficient to establish SSIV consistency:<sup>18</sup>

**Proposition 3** Suppose Assumptions 1 and 2 hold,  $\sum_\ell e_\ell z_\ell x_\ell \xrightarrow{P} \pi$  with  $\pi \neq 0$ , and Assumptions

<sup>17</sup>Goldsmith-Pinkham et al. (2020) propose a different measure of the importance of a given  $n$ , termed “Rotemberg weights.” In Appendix A.4 we show the formal connection between  $s_n$  and these weights, and that the latter do not carry the sensitivity-to-misspecification interpretation as they do in the exogenous shares view of Goldsmith-Pinkham et al. (2020). Instead, the Rotemberg weight of shock  $n$  measures the leverage of  $n$  in the equivalent shock-level IV regression from Proposition 1. Shocks may have large leverage either because of large  $s_n$ , as would be captured by the Herfindahl index, or because the shocks have a heavy-tailed distribution which is allowed by Assumption 2.

<sup>18</sup>One high-level condition used in Proposition 3 (Assumption B2) is that the control coefficient  $\gamma$  is consistently estimated by its sample analog,  $\hat{\gamma} = (\sum_\ell e_\ell w_\ell w_\ell')^{-1} \sum_\ell e_\ell w_\ell \epsilon_\ell$  (see footnote 5). We discuss sufficient conditions for this assumption in Appendix A.5.

B1-B2 hold. Then  $\hat{\beta} \xrightarrow{p} \beta$ .

PROOF See Appendix B.1.

As before, the relevance condition merits further discussion. In our simple first-stage model,  $\sum_{\ell} e_{\ell} z_{\ell} x_{\ell}^{\perp}$  converges to  $\mathbb{E} [\sum_{\ell} e_{\ell} z_{\ell} x_{\ell}]$  under appropriate regularity conditions, which is bounded above zero by a term proportional to  $\mathbb{E} [\sum_{\ell} e_{\ell} \sum_n s_{\ell n}^2]$ . Thus, in this case, SSIV relevance holds when the  $e_{\ell}$ -weighted average of local exposure Herfindahl indices  $\sum_n s_{\ell n}^2$  across observations does not vanish in expectation. In our running labor supply example, where  $x_{\ell n}$  is industry-by-region employment growth, SSIV relevance generally arises from individual regions  $\ell$  tending to specialize in a small number of industries  $n$ , provided import tariffs have a non-vanishing effect on local industry employment.<sup>19</sup> Compare this to the Herfindahl condition in Assumption 2, which instead states that the *average* shares of industries across locations become small. Both conditions may simultaneously hold when most regions specialize in a small number of industries, differentially across a large number of industries.<sup>20</sup>

### 3.2 Conditional Shock Assignment and Weak Shock Dependence

Proposition 3 establishes SSIV consistency when shocks have the same expectation across  $n$  and are mutually uncorrelated, but both requirements are straightforward to relax. We next provide extensions that allow the shock expectation to depend on observables and for weak mutual dependence (such as clustering or serial correlation) of the residual shock variation.

We first relax Assumptions 1 and 2 to only hold conditionally on a vector of shock-level observables  $q_n$  (that includes a constant). For example, it may be more plausible that shocks are as-good-as-randomly assigned within a set of observed clusters  $c(n) \in \{1, \dots, C\}$  with non-random cluster-average shocks, in which case  $q_n$  collects  $C - 1$  cluster dummies and a constant. In the labor supply example, this may allow import tariffs to vary systematically across groups of industries with similar labor intensity, but be as-good-as-random within each of those groups. In general, with  $q = \{q_n\}_n$ , we consider the following weakened version of Assumption 1:

**Assumption 3** (*Conditional quasi-random shock assignment*):  $\mathbb{E} [g_n \mid \bar{e}, q, s] = q_n' \mu$ , for all  $n$ .

Similarly, we consider a weakened version of Assumption 2 which imposes mutual uncorrelatedness on the residual  $\tilde{g}_n = g_n - q_n' \mu$ :

<sup>19</sup>Note that this precludes consideration of an asymptotic sequence where  $L$  remains finite as  $N$  grows. With  $L$  (and also  $e_1, \dots, e_L$ ) fixed, Assumption 2 implies  $\sum_{\ell} e_{\ell}^2 \mathbb{E} [\sum_n s_{\ell n}^2] \rightarrow 0$  and thus  $\text{Var} [z_{\ell}] = \text{Var} [\sum_n s_{\ell n} g_n] \rightarrow 0$  for each  $\ell$  if  $\text{Var} [g_n]$  is bounded. If the instrument has asymptotically no variation it cannot have a first stage, unless the  $\pi_{\ell n}$  grow without bound. This result also highlights the role of picking the shares which reflect the impact of  $g_n$  on  $x_{\ell}$ . Here when the shares are misspecified, i.e. when the treatment is constructed from different shares  $\tilde{s}_{\ell n}$  as  $x_{\ell} = \sum_n \tilde{s}_{\ell n} x_{\ell n}$ , the first-stage is bounded by a term proportional to  $\mathbb{E} [\sum_{\ell} e_{\ell} \sum_n s_{\ell n} \tilde{s}_{\ell n}]$ , which can be arbitrarily small even if  $\mathbb{E} [\sum_{\ell} e_{\ell} \sum_n \tilde{s}_{\ell n}^2] \neq 0$ .

<sup>20</sup>As an extreme example, suppose each region specializes on one industry only:  $s_{\ell n} = \mathbf{1} [n = n(\ell)]$  for some  $n(\ell)$ . Then the average local concentration index  $\sum_{\ell} e_{\ell} \sum_n s_{\ell n}^2$  equals one, while Assumption 2 holds when national industry composition is asymptotically dispersed: for example, when  $e_{\ell} = 1/L$  and  $n(\ell)$  is drawn *iid* across regions and uniformly over  $1, \dots, N$ .

**Assumption 4** (*Many uncorrelated shock residuals*):  $\mathbb{E} [\sum_n s_n^2] \rightarrow 0$  and  $\text{Cov} [\tilde{g}_n, \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s] = 0$  for all  $(n, n')$  with  $n' \neq n$ .

In the shock cluster example, Assumption 4 applies with a small number of clusters, each with its own random effect, as in that case a law of large numbers may apply to the within-cluster residuals  $\tilde{g}_n$  but not the original shocks  $g_n$ .

By a simple extension of the proof to Proposition 3, the SSIV estimator is consistent when these conditions replace Assumptions 1 and 2 and the residual shift-share instrument  $\tilde{z}_\ell = \sum_n s_{\ell n} \tilde{g}_n$  replaces  $z_\ell$ . While this instrument is infeasible, since  $\mu$  is unknown, the following result shows that SSIV regressions that control for the exposure-weighted vector of shock-level controls,  $\tilde{w}_\ell = \sum_n s_{\ell n} q_n$ , provide a feasible implementation:

**Proposition 4** Suppose Assumptions 3 and 4 hold,  $\sum_\ell e_\ell z_\ell x_\ell^\perp \xrightarrow{P} \pi$  with  $\pi \neq 0$ , and Assumptions B1-B2 hold. Then  $\hat{\beta} \xrightarrow{P} \beta$  provided  $\tilde{w}_\ell$  is included in  $w_\ell$ .

PROOF See Appendix B.1.

This result highlights a special role of controls with a shift-share structure (i.e.  $\sum_n s_{\ell n} q_n$ ): besides removing confounding variation from the residual (as any  $w_\ell$  would do), they can also be viewed as removing such variation directly from the shocks (i.e. implicitly using  $\tilde{g}_n$  in place of  $g_n$ ). In particular, Proposition 4 shows that controlling for each observation's individual exposure to each cluster  $\sum_n s_{\ell n} \mathbf{1}[c(n) = c]$  isolates the within-cluster variation in shocks. This allows for a thought experiment in which shocks are drawn quasi-randomly only within observed clusters, but not across clusters with potentially different shock means. Note that Proposition 3 is obtained as a special case of Proposition 4, which sets  $q_n = 1$ .

Even conditional on observables, mutual shock uncorrelatedness may be undesirably strong. It is, however, straightforward to further relax this assumption to allow for shock assignment processes with weak mutual dependence, such as further clustering or autocorrelation. In Appendix B.1 we prove extensions of Proposition 4 which replace Assumption 4 with one of the following alternatives:

**Assumption 5** (*Many uncorrelated shock clusters*): There exists a partition of shocks into clusters  $c(n)$  such that  $\mathbb{E} [\sum_c s_c^2] \rightarrow 0$  for  $s_c = \sum_{n: c(n)=c} s_n$  and  $\text{Cov} [\tilde{g}_n, \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s] = 0$  for all  $(n, n')$  with  $c(n) \neq c(n')$ ;

**Assumption 6** (*Many weakly correlated shocks*): For some sequence of numbers  $B_L \geq 0$  and a fixed function  $f(\cdot)$  satisfying  $\sum_{r=1}^\infty f(r) < \infty$ ,  $B_L \mathbb{E} [\sum_n s_n^2] \rightarrow 0$  and  $|\text{Cov} [\tilde{g}_n, \tilde{g}_{n'} \mid \bar{\varepsilon}, q, s]| \leq B_L \cdot f(|n' - n|)$  for all  $(n, n')$ .

Assumption 5 relaxes Assumption 4 by allowing shock residuals to be grouped within mutually mean-independent clusters  $c(n)$ , while placing no restriction on their within-cluster correlation. At the same

time, the Herfindahl index assumption of Assumption 4 is strengthened to hold for industry clusters, with  $s_c$  denoting the average exposure of cluster  $c$ . Assumption 6 takes a different approach, allowing all nearby shock residuals to be mutually correlated provided their covariance is bounded by a function  $B_L \cdot f(|n' - n|)$ . This accommodates, for example, the case of first-order autoregressive time series with the covariance bound declining at a geometric rate, i.e.  $f(r) = \delta^r$  for  $\delta \in [0, 1)$  and constant  $B_L$ . With  $B_L$  growing, stronger dependence of nearby shocks is also allowed (see Appendix B.1).

## 4 Extensions

We now present several other extensions of our quasi-experimental framework. Section 4.1 discusses how our framework applies when the shocks are estimated within the sample, as in the canonical Bartik (1991) study. Section 4.2 explains the need for additional controls when the sum of exposure shares vary across observations. Section 4.3 considers shift-share identification with panel data. Finally, Section 4.4 extends the framework to allow for multiple treatments and shift-share instruments.

### 4.1 Shift-Share Designs with Estimated Shocks

In some shift-share designs, the shocks are equilibrium objects that can be difficult to view as being quasi-randomly assigned. For example, in the canonical Bartik (1991) estimation of the regional labor supply elasticity, the shocks are national industry employment growth rates. Such growth reflects labor demand shifters, which one may be willing to assume are as-good-as-randomly assigned across industries. However industry growth also aggregates regional labor supply shocks that directly enter the residual  $\varepsilon_\ell$ . Here we show how the quasi-experimental SSIV framework can still apply in such cases, by viewing the  $g_n$  as noisy estimates of some latent true shocks  $g_n^*$  (labor demand shifters, in the Bartik (1991) example) that satisfy Assumption 1. We establish the conditions on estimation noise (aggregated labor supply shocks, in Bartik (1991)) such that a feasible shift-share instrument estimator, perhaps involving a leave-one-out correction as in Autor and Duggan (2003), is asymptotically valid.

We leave a more general treatment of this issue to Appendix A.6 and for concreteness focus on the Bartik (1991) example. The industry growth rates  $g_n$  can be written as weighted averages of the growth of each industry in each region:  $g_n = \sum_\ell \omega_{\ell n} g_{\ell n}$ , where the weights  $\omega_{\ell n}$  are the lagged shares of industry employment located in region  $\ell$ , with  $\sum_\ell \omega_{\ell n} = 1$  for each  $n$ . In a standard model of regional labor markets,  $g_{\ell n}$  includes (to first-order approximation) an industry labor demand shock  $g_n^*$  and a term that is proportional to the regional supply shock  $\varepsilon_\ell$ .<sup>21</sup> We suppose that the demand shocks are as-good-as-randomly assigned across industries, such that the infeasible SSIV estimator which uses

<sup>21</sup>Appendix A.7 presents such a model, showing that  $g_{\ell n}$  also depends on the regional average of  $g_n^*$  (via local general equilibrium effects) and on idiosyncratic region-specific demand shocks. Both of these are uncorrelated with the error term in the model and thus do not lead to violations of Assumption 1; we abstract away from this detail here.

$z_\ell^* = \sum_n s_{\ell n} g_n^*$  as an instrument satisfies our quasi-experimental framework. The asymptotic bias of the feasible SSIV estimator which uses  $z_\ell = \sum_n s_{\ell n} g_n$  then depends on the large-sample covariance between the labor supply shocks  $\varepsilon_\ell$  and an aggregate of the supply shock “estimation error,”

$$\psi_\ell = z_\ell - z_\ell^* \propto \sum_n s_{\ell n} \sum_{\ell'} \omega_{\ell' n} \varepsilon_{\ell'}. \quad (9)$$

Two insights follow from considering the bias term  $\sum_\ell e_\ell \psi_\ell \varepsilon_\ell$ . First, part of the covariance between  $\psi_\ell$  and  $\varepsilon_\ell$  is mechanical, since  $\varepsilon_\ell$  enters  $\psi_\ell$ . In fact, if supply shocks are spatially uncorrelated this is the only source of bias from using  $z_\ell$  rather than  $z_\ell^*$  as an instrument. This motivates the use of a leave-one-out (LOO) shock estimator,  $g_{n,-\ell} = \sum_{\ell' \neq \ell} \omega_{\ell' n} g_{\ell' n} / \sum_{\ell' \neq \ell} \omega_{\ell' n}$ , and the feasible instrument  $z_\ell^{LOO} = \sum_n s_{\ell n} g_{n,-\ell}$  to remove this mechanical covariance.<sup>22</sup> Conversely, if the regional supply shocks  $\varepsilon_\ell$  are spatially correlated a LOO adjustment may not be sufficient to eliminate mechanical bias in the feasible SSIV instrument, though more restrictive split-sample methods (e.g. those estimating shocks from distant regions) may suffice.

Second, in settings where there are many regions contributing to each shock estimate even the mechanical part of  $\sum_\ell e_\ell \psi_\ell \varepsilon_\ell$  may be ignorable, such that the conventional non-LOO shift-share instrument  $z_\ell$  (which, unlike  $z_\ell^{LOO}$ , has a convenient shock-level representation per Proposition 1) is asymptotically valid when  $z_\ell^{LOO}$  is.<sup>23</sup> In Appendix A.6, we derive a heuristic for this case, under the assumption of spatially-independent supply shocks. In a special case when each region is specialized in a single industry and there are no importance weights, the key condition is  $L/N \rightarrow \infty$ , or that the average number of regions specializing in the typical industry is large. With incomplete specialization or weights, the corresponding condition requires the typical industry to be located in a much larger number of regions than the number of industries that a typical region specializes in.

To illustrate the preceding points in the data, Appendix A.6 replicates the setting of Bartik (1991) with and without a LOO estimator, using data from Goldsmith-Pinkham et al. (2020). We find that in practice the LOO correction does not matter for the SSIV estimate, consistent with the findings of Goldsmith-Pinkham et al. (2020) and Adão et al. (2019), and especially so when the regression is estimated without regional employment weights. Our framework provides an explanation for this: the heuristic statistic we derive is much larger without importance weights. These findings imply that if, in the canonical Bartik (1991) setting, one is willing to assume quasi-random assignment of the underlying industry demand shocks and that the regional supply shocks are spatially-uncorrelated, one can interpret the uncorrected SSIV estimator as leveraging demand variation in large samples, as some of the literature has done (e.g. Suárez Serrato and Zidar (2016)).

<sup>22</sup>This problem of mechanical bias is similar to that of two-stage least squares with many instruments (Bound et al. 1995), and the solution is similar to the jackknife instrumental variable estimate approach of Angrist et al. (1999).

<sup>23</sup>Adão et al. (2019) derive the corrected standard errors for LOO SSIV and find that they are in practice very close to the non-LOO ones, in which case the SSIV standard errors we derive in the next section are approximately valid even when the LOO correction is used.

## 4.2 SSIVs with Incomplete Shares

While we have so far assumed the sum of exposure shares is constant, in practice this  $S_\ell = \sum_n s_{\ell n}$  may vary across observations  $\ell$ . For example, in the labor supply setting, the quasi-experiment in tariffs may only cover manufacturing industries, while the lagged manufacturing employment shares of  $s_{\ell n}$  may be measured relative to total employment in region  $\ell$ . In this case  $S_\ell$  equals the lagged total share of manufacturing employment in region  $\ell$ . The Autor et al. (2013) setting is another example of this scenario, as we discuss below.<sup>24</sup>

Our framework highlights a potential problem with such “incomplete share” settings: even if Assumptions 1 and 2 hold, the SSIV estimator will generally leverage non-experimental variation in  $S_\ell$  in addition to quasi-experimental variation in shocks. To see this formally, note that one can always return to the complete shares setting by rewriting the shift-share instrument with the “missing” (e.g., non-manufacturing) shock included:  $z_\ell = s_{\ell 0}g_0 + \sum_{n>0} s_{\ell n}g_n$ , where  $g_0 = 0$  and  $s_{\ell 0} = 1 - S_\ell$ , yielding  $\sum_{n=0}^N s_{\ell n} = 1$  for all  $\ell$ . The previous quasi-experimental framework then applies to this expanded set of shocks  $g_0, \dots, g_N$ . But since  $g_0 = 0$ , Proposition 3 requires in this case that  $\mathbb{E}[g_n | s, \bar{\varepsilon}] = 0$  for  $n > 0$  as well; that is, that the expected shock to each manufacturing industry is the same as the “missing” non-manufacturing shock of zero. Otherwise, even when manufacturing shocks are random, regions with higher manufacturing shares  $S_\ell$  will tend to have systematically different values of the instrument  $z_\ell$ , leading to bias when these regions also have different unobservables.<sup>25</sup>

Cast in this way, the incomplete shares issue has a natural solution via Assumption 3: to control for the sum of exposure shares. Formally, one can allow the missing and non-missing shocks to have different means by conditioning on the indicator  $\mathbf{1}[n > 0]$  in the  $q_n$  vector. By Proposition 4, the SSIV estimator allows for such conditional quasi-random assignment when the control vector  $w_\ell$  contains the exposure-weighted average of  $\mathbf{1}[n > 0]$ , which here is  $\sum_{n=0}^N s_{\ell n} \mathbf{1}[n > 0] = S_\ell$ .<sup>26</sup> Thus, in the labor supply example, quasi-experimental variation in manufacturing shocks is isolated provided one controls for a region’s lagged manufacturing share  $S_\ell$ . More generally, the control  $\sum_{n=1}^N s_{\ell n} q_n$  which, per Proposition 4, allows the shock mean to depend on observables  $q_n$  for  $n > 0$  changes the interpretation in the incomplete shares case: it is a exposure-weighted sum, rather than average.

<sup>24</sup>We note that this scenario applies to quasi-experiments in which shocks are impossible (*ex ante*) for some industries. In contrast, if all industries were equally likely to receive tariffs but only some did *ex post*, the set of  $n$  should include all industries, with  $S_\ell = 1$ , and zero tariffs captured by  $g_n = 0$  for some  $n$ . In such a case, however, it is unlikely that all manufacturing industries receive the tariffs by chance when no non-manufacturing industries do.

<sup>25</sup>Formally, if Assumptions 1 and 2 hold for all  $n > 0$  we have from the proof to Proposition 3 that  $\sum_\ell e_\ell z_\ell \varepsilon_\ell = \sum_{n=0}^N s_n g_n \bar{\varepsilon}_n = \mathbb{E} \left[ \sum_{n=0}^N s_n (g_n - \mu) \bar{\varepsilon}_n \right] + o_p(1) = -\mu \mathbb{E}[s_0 \bar{\varepsilon}_0] + o_p(1)$ . If  $\mu \neq 0$  and the missing industry share is large ( $s_0 \not\rightarrow 0$ ) this can only converge to zero when  $\mathbb{E}[s_0 \bar{\varepsilon}_0] = \mathbb{E}[\sum_\ell e_\ell s_{\ell 0} \varepsilon_\ell]$  does, i.e. when  $S_\ell$  is exogenous.

<sup>26</sup>By effectively “dummying out” the missing industry, SSIV regressions that control for  $S_\ell$  further require a weaker Herfindahl condition:  $\mathbb{E} \left[ \sum_{n=1}^N s_n^2 \right] \rightarrow 0$ , allowing the non-manufacturing industry share  $s_0$  to stay large.

### 4.3 Panel Data

In practice, SSIV regressions are often estimated with panel data, where the outcome  $y_{\ell t}$ , treatment  $x_{\ell t}$ , exposure shares  $s_{\ell nt}$ , and shocks  $g_{nt}$  are additionally indexed by time periods  $t = 1, \dots, T$ .<sup>27</sup> In such settings a time-varying instrument  $z_{\ell t} = \sum_n s_{\ell nt} g_{nt}$  is used, and the controls  $w_{\ell t}$  may include unit- or period-specific fixed effects.

It is straightforward to apply the preceding quasi-experimental framework to the panel case with a simple relabeling:  $\tilde{\ell} = (\ell, t)$  for the  $LT$  observations and  $\tilde{n} = (n, \tau)$  for the  $NT$  shocks (where  $\tau$  also indexes time periods). With exposure shares redefined as  $\tilde{s}_{\tilde{\ell}\tilde{n}} = s_{\ell nt} \mathbf{1}[t = \tau]$  (i.e. by definition zero for  $t \neq \tau$ ), the instrument can be rewritten as  $z_{\tilde{\ell}} = \sum_{\tilde{n}} \tilde{s}_{\tilde{\ell}\tilde{n}} g_{\tilde{n}}$ , mirroring the cross-sectional case.

With this relabeling, standard intuitions for panel consistency readily translate into shift-share designs. In short panels or repeated cross-sections (i.e. with fixed  $T$ ) the SSIV estimator can be consistent if  $L, N \rightarrow \infty$  and the cross-sectional conditions of Proposition 4 hold.<sup>28</sup> Alternatively, consistency of the estimator can follow from a long time series of shocks ( $T \rightarrow \infty$ ) that have weak serial dependence, even if  $L$  and  $N$  are small. This case accommodates, in particular, shift-share designs that leverage purely time-series shocks ( $N = 1, T \rightarrow \infty$ ), as in Nunn and Qian (2014).<sup>29</sup>

One subtlety of panel SSIV regressions concerns the role of fixed effect (FE) controls. Unit fixed effects play a dual role in conventional panels (with exogenous shocks varying at the same level as the observations): they purge both time-invariant unobservables ( $\frac{1}{T} \sum_{\tau} \varepsilon_{\ell\tau}$ ) from the residual and the time-invariant component of the shocks ( $\frac{1}{T} \sum_{\tau} g_{n\tau}$ ). While the first role directly extends to the FEs of cross-sectional units  $\ell$  in the shift-share case, the second role only does when exposure shares are fixed across periods, i.e. when  $s_{\ell nt} \equiv s_{\ell n0}$ .<sup>30</sup> Similarly, while period FEs always purge period-specific unobservables ( $\frac{1}{L} \sum_{\ell} \varepsilon_{\ell t}$ ), in SSIV designs they only isolate within-period shock variation when the exposure shares add up to one. With incomplete shares, in contrast, period FEs need to be interacted with the sum of exposure shares  $S_{\ell}$ .<sup>31</sup>

Finally, we note that while fixing exposure shares may have the advantage of isolating cleaner time-varying shock variation, it may also have an efficiency cost: lagging the shares by many periods

<sup>27</sup>Exposure shares are typically lagged and sometimes fixed in a pre-period. Our subscript  $t$  notation indicates that these shares are used to construct the instrument for period  $t$ , not that they are measured in that period.

<sup>28</sup>Arbitrary serial correlation of shocks can be allowed here via Assumption 5, with  $n$  defining a cluster. One complication arises when an increasing number of unit fixed effects are included in the control vector, violating Assumption B2 in Proposition 4. In Appendix A.8 we show how this incidental parameter problem can be solved by imposing shock exogeneity with respect to demeaned residuals.

<sup>29</sup>Nunn and Qian (2014) estimate the impact of U.S. food aid on civil conflict, using variation in U.S. wheat production (a single “shock” per period) over a long time horizon ( $T = 36$  years), interacted with a country’s tendency to receive US food aid (the “exposure shares” for  $L = 125$  countries). Our approach may also be appropriate in settings where  $LT$  and  $NT$  are large despite moderate  $N$  and  $T$ . Berman et al. (2017), for example, leverage price changes for  $N = 14$  minerals over  $T = 14$  years in a very large cross-section of spatial cells.

<sup>30</sup>Shift-share IV settings with panel data and time-invariant shares include, for example, Berman et al. (2015), Berman et al. (2017), Hummels et al. (2014), Imbert et al. (2019), and Nunn and Qian (2014).

<sup>31</sup>Both results follow from Proposition 4. The exposure-weighted sums of shock-level unit FEs, which isolate the time-varying component of  $g_{nt}$ , are only absorbed by observation-level FEs when the exposure shares are time-invariant. Similarly, if the  $q_n$  include shock-level period FEs, the corresponding exposure-weighted sums equal  $\sum_{\tilde{n}} \tilde{s}_{\tilde{\ell}\tilde{n}} \mathbf{1}[\tau = \tilde{t}] = S_{\ell} \mathbf{1}[t = \tilde{t}]$ , which simplifies to period FEs when  $S_{\ell} = 1$ .

is likely to make the shift-share instrument less predictive of treatment (see Appendix A.9 for a formal argument). If a researcher wants to update the shares to maximize the first stage, but also isolate the shock variation over time (which is not achieved by controlling for unit  $\ell$  fixed effects with time-varying shares), she may instead use the first-differenced specification. That is, estimate  $\Delta y_{\ell t} = \beta \Delta x_{\ell t} + \gamma' \Delta w_{\ell t} + \Delta \varepsilon_{\ell t}$ , instrumenting  $\Delta x_{\ell t}$  with  $z_{\ell t, FD} = \sum_n s_{\ell n, t-1} \Delta g_{nt}$ , where  $\Delta$  is the first-differencing operator for both observations and shocks. This strategy has been employed, for example, by Autor et al. (2013) as we discuss in Section 6.2.<sup>32</sup>

#### 4.4 Multiple Shocks and Treatments

In some shift-share designs one may be interested in leveraging several shift-share instruments, corresponding to multiple sets of shocks satisfying Assumptions 3 and 4. For example while Autor et al. (2013) construct an instrument from average Chinese import growth across eight non-U.S. countries, in principle the industry shocks from each individual country may be each thought to be as-good-as-randomly assigned. Jaeger et al. (2018) instrument two treatments—the current and lagged immigration rates—with two shift-share instruments. Bombardini and Li (2019) estimate the reduced-form effects of two shift-share variables: the regional growth of all exports and the regional growth of exports in pollution-intensive sectors. In Appendix A.10 we show that our quasi-experimental framework extends to these settings, in which the exposure shares used to construct the instruments are the same but the shocks differ. The key insight is that SSIV regressions with multiple instruments—with and without multiple endogenous variables—again have an equivalent representation as particular shock-level IV estimators, although the equivalence result is more complex under overidentification.

### 5 Shock-Level Inference and Testing

A shock-level view also brings new insights to SSIV inference and testing. In this section we first show how a problem with conventional SSIV inference, first studied by Adão et al. (2019), has a convenient solution based on our shock-level equivalence result. We then discuss how other novel shock-level procedures can be used to assess first-stage relevance and to implement valid falsification tests of shock exogeneity. Lastly, we summarize a variety of Monte-Carlo simulations illustrating the finite-sample properties of SSIV.

---

<sup>32</sup>There is another argument for fixing the shares in a pre-period that arises when the current shares are affected by lagged shocks in a way that is correlated with unobservables  $\varepsilon_{\ell t}$ . In the labor supply example, suppose local labor markets vary in flexibility, with stronger reallocation of employment to industries with larger increases in import tariffs in flexible markets. If import tariffs are random but persistent, industries with growing tariffs will be increasingly concentrated in regions with flexible labor markets and Assumption 1 will be violated if such flexibility is correlated with  $\varepsilon_{\ell t}$ . This concern is not specific to panel data, but may also arise in cross-sections or repeated cross-sections.

## 5.1 Exposure-Robust Standard Errors

As with consistency, SSIV inference is complicated by the fact that the observed shocks  $g_n$  and any unobserved shocks  $\nu_n$  induce dependencies in the instrument  $z_\ell$  and residual  $\varepsilon_\ell$  across observations with similar exposure shares. This problem can be understood as an extension of the standard clustering concern (Moulton 1986), in which the instrument and residual are correlated across observations within predetermined clusters, with the additional complication that in SSIV every pair of observations with overlapping shares may have correlated  $(z_\ell, \varepsilon_\ell)$ . Adão et al. (2019) develop a novel approach to conducting valid inference in presence of exposure-based clustering, building on our quasi-experimental framework for identification.

Our equivalence result in Section 2.3 motivates a convenient alternative approach to valid SSIV inference. By estimating SSIV coefficients with an equivalent shock-level IV regression, one directly obtains valid (“exposure-robust”) standard errors under the assumptions in Adão et al. (2019) and an additional condition on the controls that we discuss below.<sup>33</sup>

**Proposition 5** Consider  $s_n$ -weighted IV estimation of the second stage equation

$$\bar{y}_n^\perp = \alpha + \beta \bar{x}_n^\perp + q_n' \delta + \bar{\varepsilon}_n^\perp \quad (10)$$

where  $\tilde{w}_\ell = \sum_n s_{\ell n} q_n$  is included in the control vector  $w_\ell$  used to compute  $\bar{y}_n^\perp$  and  $\bar{x}_n^\perp$ , and  $\bar{x}_n^\perp$  is instrumented by  $g_n$ . The IV estimate of  $\beta$  is numerically equivalent to the SSIV estimate  $\hat{\beta}$ . Furthermore, when Assumptions B3–B6 in Appendix B.2 hold and  $\sum_\ell e_\ell x_\ell^\perp z_\ell \xrightarrow{p} \pi$  for  $\pi \neq 0$ , the conventional heteroskedasticity-robust standard error for  $\hat{\beta}$  yields asymptotically-valid confidence intervals for  $\beta$ .

PROOF See Appendix B.2.

Equation (10) extends the previous shock-level estimating equation (6) by including a vector of controls  $q_n$  which, as in Proposition 4, are included in the SSIV control vector  $w_\ell$  as exposure-weighted averages. The first result in Proposition 5 is that the addition of these controls does not alter the coefficient equivalence established in Proposition 1. The second result states conditions, which strengthen those of Proposition 4, under which conventional shock-level standard errors from estimation of (10) yield valid asymptotic inference on  $\beta$ .<sup>34</sup>

While our previous results do not restrict the structure of the control vector  $w_\ell$ , Proposition 5 (specifically, Assumption B4) allows for only two types of controls. All sources of shock-level confounding have to be captured by controls with a shift-share structure (i.e.  $\sum_n s_{\ell n} q_n$ , as in Proposition

<sup>33</sup>This solution generalizes a well-known approach to addressing conventional group clustering (Angrist and Pischke 2008, p. 313): by estimating a regression at the level of as-good-as-random variation (here, shocks) one avoids inferential biases due to clustering (here, by shock exposure).

<sup>34</sup>Appendix B.2 also establishes two related results regarding specifications without controls and an alternative inference procedure to improve finite-sample performance.

4); the other controls should not be asymptotically correlated with the instrument although they may increase the asymptotic efficiency of the estimator. While valid shift-share inference with general control vectors remains an open problem, we show in Appendix B.2 that the standard errors from Proposition 5 are asymptotically conservative under a much weaker assumption, which allows for controls of the form  $\sum_n s_{\ell n} p_n + u_\ell$ , where  $p_n$  are unobserved confounders and  $u_\ell$  is noise.<sup>35</sup>

Our shock-level approach to estimating exposure-robust standard errors offers three practical features. First, it can be performed with standard statistical software packages given a simple initial transformation of the data (i.e. to obtain  $\bar{y}_n^\perp$ ,  $\bar{x}_n^\perp$ , and  $s_n$ ), for which we have released a Stata package *ssaggregate* (see footnote 3). Second, it is readily extended to settings where shocks are clustered or autoregressive, as in Assumptions 5 and 6 respectively. Conventional cluster-robust or heteroskedastic-and-autocorrelation-consistent (HAC) standard error calculations applied to equation (10) are then valid. Third, the shock-level inference approach works when  $N > L$  or when some exposure shares are collinear.<sup>36</sup>

## 5.2 Falsification and Relevance Tests

Our Proposition 5 also provides a practical way to perform valid regression-based tests of shock orthogonality (i.e. falsification tests) and first-stage relevance. As a falsification test of Assumption 3, one may regress an observed proxy  $r_\ell$  for the unobserved residual  $\varepsilon_\ell$  on the instrument  $z_\ell$  (controlling for  $w_\ell$ ). Examples of such  $r_\ell$  include baseline characteristics realized prior to the shocks or lagged observations of the outcome (yielding a so-called “pre-trend” test). To make the magnitude of the placebo coefficient more interpretable, one may also regress  $r_\ell$  on  $x_\ell$  while instrumenting with  $z_\ell$ . To address the exposure clustering problem that may arise in these regressions, the shock-level regression of Proposition 5 can be used, yielding valid inference on these coefficients. If a researcher instead starts from a shock-level confounder  $r_n$ , they can construct its observation-level average  $r_\ell = \sum_n s_{\ell n} r_n$  and proceed similarly; a simpler test regresses  $r_n$  on  $g_n$  directly (weighting by  $s_n$  and controlling for  $q_n$ ).<sup>37</sup>

Similarly, Proposition 5 yields a convenient way to test first-stage relevance in the OLS regression of  $x_\ell$  on  $z_\ell$  and  $w_\ell$ . We note that the equivalent shock-level regression is IV, not OLS (see footnote 13). The first stage  $F$ -statistic, which is a common heuristic for relevance, is then obtained as a squared  $t$ -statistic.<sup>38</sup>

<sup>35</sup>Adão et al. (2019) provide asymptotically valid standard errors in a special case of this weaker assumption: when the average variance of the noise  $u_\ell$  is asymptotically small for all controls that are necessary for identification. Our standard errors remain asymptotically conservative in this case.

<sup>36</sup>This is not possible with the standard error calculation of Adão et al. (2019) because their procedure involves projecting  $z_\ell^\perp$  on the vector of shares in order to account for the shock-level confounders underlying the approximate shift-share controls. This issue can be empirically relevant: for instance, employment shares of some industries are collinear in the Autor et al. (2013) setting.

<sup>37</sup>While pre-trend and other  $\ell$ -level balance tests are also useful in the alternative Goldsmith-Pinkham et al. (2020) framework (albeit with a different approach to inference), this shock-level test is specific to our approach to identification. We emphasize that all tests discussed here are meant to falsify the quasi-random shock assignment assumption made *a priori*, and not to test the two frameworks against each other.

<sup>38</sup>We generalize this result to the case of multiple shift-share instruments in Appendix A.10 by detailing the appropriate construction of the “effective” first-stage  $F$ -statistic of Montiel Olea and Pflueger (2013), again based on an equivalent

### 5.3 Monte-Carlo Simulations

Though the exposure-robust standard errors obtained from estimating equation (10) are asymptotically valid, it is useful to verify that they offer appropriate coverage with a finite number of observations and shocks. In Appendix A.11 we provide Monte-Carlo simulations confirming that the finite-sample performance of the equivalent regression (10) is comparable to that of more conventional shock-level IV regressions, in which the outcome and instrument are not aggregated from a common set of  $y_\ell$  and  $x_\ell$ . The asymptotic approximation performs well even with a Herfindahl concentration index  $\sum_n s_n^2$  of 1/20 (which can be compared to a shock-level regression with 20 equal-sized industries); the conventional rule of thumb for detecting weak instruments based on the appropriate constructed first-stage  $F$ -statistic applies equally well to SSIV estimators. These results indicate that a researcher who is comfortable with the finite-sample performance of a shock-level analysis with some set of  $g_n$  should also be comfortable using such shocks in SSIV, provided there is sufficient variation in exposure shares to yield a strong SSIV first stage.

## 6 Shift-Share IV in Practice

We now summarize and illustrate the practical implications of our econometric framework. We first characterize the kinds of empirical settings to which the foregoing framework may be applied. We then apply the framework to the influential setting of Autor et al. (2013).

### 6.1 A Taxonomy of SSIV Settings

Our framework can be applied to various empirical settings. To characterize these settings, we distinguish between three cases of SSIVs employed in the literature.

In the first case, the shift-share instrument is based on a set of shocks which can itself be thought of as an instrument. For example, the  $g_n$  which enter  $z_\ell$  might correspond to a set of observed growth rates that could be plausibly thought of as being randomly assigned to a large number of industries. Our framework shows how the shift-share instrument maps these shocks to the level of observed outcomes and treatments (e.g., geographic regions). A researcher who is comfortable with the identification conditions and finite-sample performance of an industry-level analysis based on  $g_n$  should generally also be comfortable applying our framework, provided there is sufficient variation in the exposure shares and treatment to yield a strong first stage. Autor et al. (2013) and the corresponding industry-level analysis conducted by Acemoglu et al. (2016) give a prime example of this case, as we show below.

Empirical settings covered by this first case belong to various fields in economics, with outcomes and shocks defined at levels ( $\ell$  and  $n$ , respectively) different than regions and industries. In international shock-level IV regression.

trade, Hummels et al. (2014) estimate the wage effects of offshoring across Danish importing firms  $\ell$ . They leverage a shift-share instrument for offshoring based on shocks to export supply by type of intermediate inputs and origin country; titanium hinges from Japan is an example of an  $n$ . While they translate these shocks to the firm level by using the lagged composition of firm imports as the shares, one could imagine an analysis of Danish imports at the input-by-country level directly that would leverage the same supply shocks. In finance, Xu (2019) examines the long-term effects of financial shocks on exports across countries  $\ell$ . Her shift-share instrument is based on a disruption that affected some but not all London-based banks  $n$  in 1866, with country-specific exposure shares measuring pre-1866 market shares of those banks in each country. In line with considering bank shocks as-good-as-randomly assigned, she reports that affected and unaffected banks were balanced on various observable characteristics. In the immigration literature, Peri et al. (2016) estimate the effect of immigrant STEM workers on the labor market outcomes of natives across U.S. cities  $\ell$ . They exploit variation in the supply of STEM workers across migration origin countries  $n$  and over time that arises from plausibly exogenous shifts to national H1-B policy. Similarly, in a literature on innovation, Stuen et al. (2012) leverage education policy shocks in foreign countries as a supply shock to U.S. doctoral programs.

In the second case, a researcher does not directly observe a large set of quasi-experimental shocks, but can still conceive of an underlying set of  $g_n$  which if observed would be a useful instrument. Constructing the instrument then requires an initial step where these shocks are estimated in-sample, potentially introducing mechanical biases. In the canonical setting of Bartik (1991) and Blanchard and Katz (1992), for example, a local labor demand instrument is sought, with the ideal  $g_n$  measuring an aggregate change in industry labor demand that may be assumed orthogonal to local labor supply shocks. Aggregate demand changes are however not directly observed and must be estimated from national industry employment growth (often using leave-out corrections, as in Autor and Duggan (2003) and Diamond (2016)). We have discussed how our framework generalizes to this more involved setting in Section 4.1, showing the additional assumptions required for the estimation error to be asymptotically ignorable. While this case differs from the first in terms of the instrument construction, the underlying logic of our framework still applies. This case also covers instruments in the immigration literature, as in Card (2001) and Card (2009), where latent shocks to out-migration from foreign countries can be thought to be as-good-as-randomly assigned but are estimated from aggregate immigration flows in the U.S.

The third case is conceptually distinct, in that the  $g_n$  underlying the (perhaps idealized) instrument cannot be naturally viewed as an instrument itself. This could either be because it is not plausible that these shocks are as-good-as-randomly assigned, even conditionally on shock-level observables, or because there are too few shocks. Identification in this case may instead follow from exogeneity of the exposure shares, as suggested by Goldsmith-Pinkham et al. (2020).

Share exogeneity may be a more plausible approach in the third case when the exposure shares are “tailored” to the specific economic question, and to the particular endogenous variable included in the model. In this case, the scenario considered in Section 2.2—that there are unobserved shocks  $\nu_n$  which enter  $\varepsilon_\ell$  through the shares—may be less of a concern. Mohnen (2019), for example, uses the age profile of older workers in local labor markets as the exposure shares of a shift-share instrument for the change in the local elderly employment rate in the following decade. He argues, based on economic intuition, that these tailored shares are uncorrelated with unobserved trends in youth employment rates. This argument notably does not require one to specify the age-specific shocks  $g_n$ , which only affect power of the instrument (in fact, the shocks are dispensed with altogether in robustness checks that directly instrument with the shares). Similarly, Algan et al. (2017) use the lagged share of the construction sector in the regional economy as an instrument for unemployment growth during the Great Recession, arguing that it does not predict changes in voting outcomes in other ways. With a single industry the identification assumption reduces to that of conventional difference-in-differences with continuous treatment intensity and our framework cannot be applied.

In contrast, our framework may be more appropriate in settings where shocks are tailored to a specific question while the shares are “generic,” in that they could conceivably measure an observation’s exposure to multiple shocks (both observed and unobserved). Both Autor et al. (2013) and Acemoglu and Restrepo (Forthcoming), for example, build shift-share instruments with similar lagged employment shares but different shocks—rising trade with China and the adoption of industrial robots, respectively. According to the Goldsmith-Pinkham et al. (2020) view, these papers use essentially the same instruments (lagged employment shares) for different endogenous variables (growth of import competition and growth of robot adoption), and are therefore mutually inconsistent. Our framework helps reconcile these identification strategies, provided the variation in each set of shocks can be described as arising from a natural experiment. In principle, shares and shocks may simultaneously provide valid identifying variation, but in practice it would seem unlikely for both sources of variation to be *a priori* plausible in the same setting.

This discussion highlights that plausibility of our framework, as with the alternative framework of Goldsmith-Pinkham et al. (2020), depends on the details of the SSIV application. We encourage practitioners to use our framework only after establishing an *a priori* argument for the plausibility of exogenous shocks. Various diagnostics on the extent of shock variation and falsification of this assumption may then be conducted to assess *ex post* the plausibility of exogenous shocks. We next illustrate this approach in the Autor et al. (2013) setting.

## 6.2 Application to Autor, Dorn, and Hanson (2013)

Our application to Autor et al. (2013, henceforth ADH) aims to illustrate our theoretical framework only, and not to reassess their substantive findings. In line with this goal, we first describe how the

ADH instrument could be thought to leverage quasi-experimental shocks and discuss potential threats to this identification strategy. We then illustrate the tools and lessons that follow from our framework, demonstrating steps that researchers can emulate in their own SSIV applications. Specifically, we analyze the distribution of shocks to assess the plausibility of Assumption 4 (many conditionally uncorrelated shocks), use balance tests to corroborate the plausibility of Assumption 3 (conditional quasi-random shock assignment), use equivalent shock-level IV regressions to obtain exposure-robust inference, and analyze the sensitivity of the results to the inclusion of different shock-level controls. This analysis shows how our quasi-experimental framework can help understand the identifying variation in the ADH SSIV design.

### 6.2.1 Setting and Intuition for Identification

ADH use a shift-share IV to estimate the causal effect of rising import penetration from China on U.S. local labor markets. They do so with a repeated cross section of 722 commuting zones  $\ell$  and 397 four-digit SIC manufacturing industries  $n$  over two periods  $t$ , 1990-2000 and 2000-2007. In these years U.S. commuting zones were exposed to a dramatic rise in import penetration from China, a historic change in trade patterns commonly referred to as the “China shock.” Variation in exposure to this change across commuting zones results from the fact that different areas were initially specialized in different industries which saw different changes in the aggregate U.S. growth of Chinese imports. ADH combine import changes across industries in eight comparable developed economies (as shocks) with lagged industry employment (as exposure shares) to construct their shift-share instrument.

To illustrate our framework in this setting we focus on ADH’s primary outcome of the change in total manufacturing employment as a percentage of working-age population during period  $t$  in location  $\ell$ , which we write as  $y_{\ell t}$ . The treatment variable  $x_{\ell t}$  measures local exposure to the growth of imports from China in \$1,000 per worker. The vector of controls  $w_{\ell t}$ , which comes from the preferred specification of ADH (Column 6 of their Table 3), contains start-of-period measures of labor force demographics, period fixed effects, Census region fixed effects, and the start-of-period total manufacturing share to which we return below. The shift-share instrument is  $z_{\ell t} = \sum_n s_{\ell nt} g_{nt}$ , where  $s_{\ell nt}$  is the share of manufacturing industry  $n$  in total employment in location  $\ell$  (measured a decade before each period  $t$  begins) and  $g_{nt}$  is industry  $n$ ’s growth of imports from China in the eight comparable economies over period  $t$  (also expressed in \$1,000 per U.S. worker).<sup>39</sup> Importantly, the

<sup>39</sup>To be precise, local exposure to the growth of imports from China is constructed for period  $t$  as  $x_{\ell t} = \sum_n s_{\ell nt}^{\text{current}} g_{nt}^{\text{US}}$ . Here  $g_{nt}^{\text{US}} = \frac{\Delta M_{nt}^{\text{US}}}{E_{nt}^{\text{current}}}$  is the growth of U.S. imports from China in thousands of dollars ( $\Delta M_{nt}^{\text{US}}$ ) divided by the industry employment in the U.S. at the beginning of the current period ( $E_{nt}^{\text{current}}$ ) and  $s_{\ell nt}^{\text{current}}$  are local employment shares, also measured at the beginning of the period. The instrument, in contrast, is constructed as  $z_{\ell t} = \sum_n s_{\ell nt} g_{nt}$  with  $g_{nt} = \frac{\Delta M_{nt}^{\text{8 countries}}}{E_{nt}}$ , where  $\Delta M_{nt}^{\text{8 countries}}$  measures the growth of imports from China in eight comparable economies (in thousands of U.S. dollars) and both local employment shares  $s_{\ell nt}$  and U.S. employment  $E_{nt}$  are lagged by 10 years. The eight countries are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland. Note that Autor et al. (2013) express the same instrument differently, based on employment shares relative to the industry total, rather than the regional total. Our way of writing  $z_{\ell t}$  aims to clearly separate the exposure shares from the

sum of lagged manufacturing shares across industries ( $S_{\ell t} = \sum_n s_{\ell nt}$ ) is not constant across locations and periods, placing the ADH instrument in the “incomplete shares” class discussed in Section 4.2. All regressions are weighted by  $e_{\ell t}$ , which measures the start-of-period population of the commuting zone, and all variables are measured in ten-year equivalents.

To see how the ADH instrument can be viewed as leveraging quasi-experimental shocks, consider an idealized experiment generating random variation in the growth of imports from China across industries. One could imagine, for example, random variation in industry-specific productivities in China affecting import growth both in the U.S. and in comparable economies. This would yield a set of observed productivity changes  $g_{nt}$  which would plausibly satisfy our Assumption 1. Assumption 2 would further hold when the productivity shocks are idiosyncratic across many industries, with small average exposure to each shock across commuting zones. Weaker versions of this experimental ideal, in which productivity shocks can be partly predicted by industry observables and are only weakly dependent across industries, are accommodated by the extensions in Section 3.2. For example, in ADH’s repeated cross section one might invoke Assumption 3 in allowing the average shock to vary across periods, in recognition that the 1990s and 2000s were very different trade environments, as China joined the World Trade Organization in 2001. Here  $q_{nt}$  would indicate periods.

ADH’s approach can be seen as approximating this idealized experiment by using observed changes in trade patterns between China and a group of developed countries outside the United States. Trade between the U.S. and China depends on changes in U.S. supply and demand conditions, which may have direct effects on employment dynamics in U.S. regions. In contrast, variation in the ADH  $g_{nt}$  reflects only Chinese productivity shocks and the various supply and demand shocks in the non-U.S. developed countries. In this way, the ADH strategy can be understood as eliminating bias from shocks that are specific to the United States.

This discussion gives an *a priori* justification for thinking of the ADH instrument as leveraging quasi-experimental shocks within the two periods. Nonetheless, since the ADH shocks are not truly randomized, one may still worry that they are confounded by other unobserved characteristics. For example, China’s factor endowment may imply that it specializes in low-skill industries, which could have been on different employment trends in the U.S. even absent increased trade with China. Similarly, one can imagine a common component of import growth in the U.S. and the group of comparable developed economies due to correlated technological shocks in those countries, which may have a direct effect on U.S. labor markets. Given these potential concerns, it will be important to assess the plausibility of Assumption 3 in this setting by conducting within-period falsification tests of the kind we describe in Section 5.2. It will also be important to assess whether there is sufficient variation in the ADH shocks for Assumption 4 to hold.

Before applying these tests, it is worth highlighting that the assumption of exogenous exposure 

---

 industry shocks, highlighting the shift-share structure of the instrument.

shares, as discussed by Goldsmith-Pinkham et al. (2020), is likely to be *a priori* implausible in the ADH setting. As indicated in Section 2.2, any unobserved shocks  $\nu_{nt}$  invalidate the share exogeneity assumption if they enter the error  $\varepsilon_{\ell t}$  in a manner which is correlated with the shares. Because ADH use generic manufacturing employment shares to instrument for a specific treatment variable, the possibility of other industry shocks entering  $\varepsilon_{\ell t}$  looms large. These unobserved shocks could take many forms, for example heterogeneous speeds of automation, secular changes in consumer demand, or changes in factor prices which differentially affect industries based on their skill intensity. This is in contrast to our Assumption 3 which allows for any of these unobserved shocks as long as they are uncorrelated with  $g_{nt}$  across industries, conditionally on observables  $q_{nt}$ .<sup>40</sup>

With a plausible justification of our framework in hand, we next illustrate its application.

### 6.2.2 Properties of Industry Shocks and Exposure Shares

Our quasi-experimental view of the ADH research design places particular emphasis on the variation in Chinese import growth rates  $g_{nt}$  and their average exposure  $s_{nt}$  across industries and periods. With few or insufficiently-variable shocks, or highly concentrated shocks exposure, the large- $N$  asymptotic approximation developed in Section 3 is unlikely to be a useful tool for characterizing the finite-sample behavior of the SSIV estimator. We thus first summarize the distribution of  $g_{nt}$ , as well as the industry-level weights from our equivalence result,  $s_{nt} \propto \sum_{\ell} e_{\ell t} s_{\ell nt}$  (normalized to add up to one in the entire sample).<sup>41</sup>

In summarizing the industry-level variation it is instructive to recall that the ADH instrument is constructed with “incomplete” manufacturing shares. Per the discussion in Section 4.2, this means that absent any regression controls the SSIV estimator uses variation not only in manufacturing industry shocks but also implicitly the variation in the 10-year lagged total manufacturing share  $S_{\ell t}$  across commuting zones and periods. In practice, ADH control for the start-of-period manufacturing share, which is highly—though not perfectly—correlated with  $S_{\ell t}$ . We thus summarize the ADH shocks both with and without the “missing” shock  $g_{0t} = 0$ , which here represents the lack of a “China shock” in service (i.e. non-manufacturing) industries. Given that trade with China was very different in the 1990s and 2000s, we focus on the within-period variation in manufacturing shocks.

Table 1 reports summary statistics for the ADH shocks  $g_{nt}$  computed with importance weights  $s_{nt}$ , and characterizes these weights. Column 1 includes the “missing” service industry shock of zero in each period. It is evident that with this shock the distribution of  $g_{nt}$  is unusual: for example, its interquartile range is zero. This is because the service industry accounts for a large fraction of total employment ( $s_{0t}$  is 71.9% of the period total in the 1990s and 79.5% in the 2000s). As a result

<sup>40</sup>This assumption, which allows one to isolate import competition from other industry shocks, is standard in similar industry-level analyses (e.g. Acemoglu et al. 2016) and can be tested with falsification tests, as we do in Section 6.2.3.

<sup>41</sup>Note that  $s_{nt}$  would be proportional to lagged industry employment if the ADH regression weights  $e_{\ell t}$  were lagged regional employment. ADH however use a slightly different  $e_{\ell t}$ : the start-of-period commuting zone population.

we see a high concentration of industry exposure as measured by the inverse of its Herfindahl index (HHI),  $1/\sum_{n,t} s_{nt}^2$ , which corresponds to the effective sample size of our equivalent regression and plays a key role in Assumption 2. With the “missing” shock included, the effective sample size is only 3.5. For an HHI computed at the level of three-digit industry codes  $\sum_c s_c^2$ , where  $s_c$  aggregates exposure across the two periods and industries within the same 3-digit group  $c$ , it is even lower, at 1.7. This suggests even less industry-level variation is available when shocks are allowed to be serially correlated or clustered by groups. Furthermore, the mean of manufacturing shocks is significantly different from the zero shock of the missing service industry.<sup>42</sup> Together, these analyses suggest that the service industry should be excluded from the identifying variation, because it is likely to violate both Assumption 1 ( $\mathbb{E}[g_{nt} | \bar{\varepsilon}, s] \neq g_{0t} = 0$ ) and Assumption 2 ( $\sum_{n,t} s_{nt}^2$  is not close to zero).

Column 2 of Table 1 therefore summarizes the sample with the service industry excluded. The distribution of shocks is now much more regular, with an average of 7.4, a standard deviation of 20.9 and an interquartile range of 6.6. The inverse HHI of the  $s_{nt}$  is also relatively high: 191.6 across industry-by-period cells and 58.4 when exposure is aggregated by SIC3 group. The largest shock weights in this column are only 3.4% across industry-by-periods and 6.5% across SIC3 groups. This suggests a sizable degree of variation at the industry level, consistent with Assumption 2. In general, we recommend that researchers report the inverse of the HHI of shock-level average exposure as a simple way of describing their effective sample size. A first-stage  $F$ -statistic, which we discuss appropriate computation of in Section 5.2, will provide a formal test of the power of the shock variation.

Finally, column 3 of Table 1 summarizes the distribution of within-period manufacturing shocks, which would be leveraged by an assumption of conditional quasi-experimental assignment (Assumption 3). The column confirms that even conditional on period there is sizable residual shock variation. The standard deviation and interquartile range of shock residuals (obtained from regressing shocks on period fixed effects with  $s_{nt}$  weights) are only mildly smaller than in Column 2, despite the higher mean shock in the later period, at 12.6 versus 3.6.

Besides the condition on the effective sample size, Assumption 2 (and its clustered version in Assumption 5) requires the shocks to be sufficiently mutually uncorrelated. To assess the plausibility of this assumption and choose the appropriate level of clustering for exposure-robust standard errors, we next analyze the correlation patterns of shocks across manufacturing industries using available industry classifications and the time dimension of the pooled cross section. In particular, we compute intra-class correlation coefficients (ICCs) of shocks within different industry groups, as one might do to correct for conventional clustering parametrically (e.g. Angrist and Pischke (2008, p. 312)).<sup>43</sup>

<sup>42</sup>The weighted mean of manufacturing shocks is 7.4, with a standard error clustered at the 3-digit SIC level (as in our analysis below) of 1.3.

<sup>43</sup>Note that similar ICC calculations could be implemented in a setting that directly regresses industry outcomes on industry shocks, such as Acemoglu et al. (2016). Mutual correlation in the instrument is a generic concern that is not specific to shift-share designs, although one that is rarely tested for. Getting the correlation structure in shocks right is especially important for inference in our framework, since the outcome and treatment in the industry-level regression ( $\bar{y}_{nt}^\perp$  and  $\bar{x}_{nt}^\perp$ ) are by construction correlated across industries.

These ICCs come from a random effects model, providing a hierarchical decomposition of residual within-period shock variation:

$$g_{nt} = \mu_t + a_{\text{ten}(n),t} + b_{\text{sic2}(n),t} + c_{\text{sic3}(n),t} + d_n + e_{nt}, \quad (11)$$

where  $\mu_t$  are period fixed effects;  $a_{\text{ten}(n),t}$ ,  $b_{\text{sic2}(n),t}$ , and  $c_{\text{sic3}(n),t}$  denote time-varying (and possibly auto-correlated) random effects generated by the ten industry groups in Acemoglu et al. (2016), 20 groups identified by SIC2 codes, and 136 groups corresponding to SIC3 codes, respectively; and  $d_n$  is a time-invariant industry random effect (across our 397 four-digit SIC industries). Following convention, we estimate equation (11) as a hierarchical linear model by maximum likelihood, assuming Gaussian residual components.<sup>44</sup>

Table 2 reports estimated ICCs from equation (11), summarizing the share of the overall shock residual variance due to each random effect. These reveal moderate clustering of shock residuals at the industry and SIC3 level (with ICCs of 0.169 and 0.073, respectively). At the same time, there is less evidence for clustering of shocks at a higher SIC2 level and particularly by ten cluster groups (ICCs of 0.047 and 0.016, respectively, with standard errors of comparable magnitude). This supports the assumption that shocks are mean-independent across SIC3 clusters, so it will be sufficient to cluster standard errors at the level of SIC3 groups, as Acemoglu et al. (2016) do in their conventional industry-level IV regressions. The inverse HHI estimates in Table 1 indicate that at this level of shock clustering there is still an adequate effective sample size.

### 6.2.3 Falsification Tests

We next implement falsification tests of ADH shock orthogonality, which provide a way of assessing the plausibility of Assumption 3. Following Section 5.2, we do this in two ways, both different from conventional falsification tests sometimes run in SSIV settings. First, we regress potential proxies for the unobserved residual (i.e., any unobserved industry labor demand or labor supply shock) on the instrument  $z_\ell$  but use exposure-robust inference that takes into account the inherent dependencies of the data. Second, we regress potential industry-level confounders directly on the shocks (while again clustering by SIC3). While this second type of falsification tests would be standard in industry-level analyses, such as Acemoglu et al. (2016), it has rarely been used to assess the plausibility of SSIV designs (with Xu (2019), mentioned above, being a rare exception).

Choosing the set of potential confounders for these exercises is a context-specific issue, which should be justified separately in every application. To discipline our illustrative exercise, we use the industry-level production controls in Acemoglu et al. (2016) and the regional controls in ADH. Consistent with our *a priori* view of the quasi-experiment, we maintain only the period fixed effects as

<sup>44</sup>In particular we estimate an unweighted mixed-effects regression using Stata's *mixed* command, imposing an exchangeable variance matrix for  $(a_{\text{ten}(n),1}, a_{\text{ten}(n),2})$ ,  $(b_{\text{sic2}(n),1}, b_{\text{sic2}(n),2})$ , and  $(c_{\text{sic3}(n),1}, c_{\text{sic3}(n),2})$ .

controls when evaluating balance on these other observables. For the industry-level balance test this amounts to regressing each potential confounder on the manufacturing shocks (normalized to have a unit variance) and period fixed effects, weighting by average industry employment shares. Regional balance coefficients are obtained by regressing each potential confounder on the shift-share instrument (normalized to have a unit variance) and the share-weighted average of period effects (i.e., the period-interacted sum-of-shares), since ADH is a setting with incomplete shares. To obtain exposure-robust standard errors, we implement these regressions at the shock level, as discussed above.

Panel A of Table 3 reports the results of our industry-level balance tests. The five Acemoglu et al. (2016) production controls are an industry’s share of production workers in employment in 1991, the ratio of its capital to value-added in 1991, its log real wages in 1991, the share of its investment devoted to computers in 1990, and the share of its high-tech equipment in total investment in 1990.<sup>45</sup> Broadly, these variables reflect the structure of employment and technology across industries. If the ADH shocks are as-good-as-randomly assigned to industries within periods, we expect them to not predict these predetermined variables. Panel A shows that there is indeed no statistically significant correlation within periods, consistent with Assumption 3.

Panel B of Table 3 reports the results of our regional balance tests. The five ADH controls are the fraction of a commuting zone’s population who is college-educated, the fraction of its population who is foreign-born, the fraction of its workers who are female, its fraction of employment in routine occupations, and the average offshorability index of its occupations. Broadly, these variables reflect the composition of a region’s workforce. We again find no statistically significant relationships between these variables and the shift-share instrument within periods, except for the foreign-born population fraction. Locations exposed to a large ADH trade shock tend to have a higher fraction of immigrants, suggesting that they may be subject to different labor supply dynamics. We explore the importance of this imbalance for the SSIV coefficient estimate in sensitivity tests below.

Finally, the last two rows of the same panel conduct a regional “pre-trends” analysis. We regress the pre-trend variables from ADH—manufacturing employment growth in the 1970s and 1980s—on the shift-share instrument, using the same specification as in the previous rows. We find no relationship between the shift-share instrument and manufacturing employment growth in the 1980s, but there is a positive statistically significant relationship with manufacturing employment growth in the 1970s. Both findings are similar to those from ADH’s pre-trend analysis.

Overall, we fail to reject imbalance in ten out of the twelve potential confounders at conventional levels of statistical significance. How to proceed when some balance tests fail is a general issue in quasi-experimental analyses and has to be decided in the context of an application. One might view the balance failures as sufficient evidence against Assumption 3 to seek alternative shocks or more appropriate shock-level controls. Alternatively, one may argue that the observed imbalances

---

<sup>45</sup>The last two controls are missing for five out of 397 industries. We impute the missing values by the medians in the SIC3 industry group or, when not available, in the SIC2 group.

are unlikely to invalidate the research design. ADH, for example, note that the positive relationship they find between the shift-share instrument and manufacturing employment in the 1970s occurs in the distant past, while the insignificant relationship in the 1980s demonstrates that the relationship between rising China trade exposure and declining manufacturing employment was absent in the decade immediately prior to China’s rise. Similarly the imbalance of the foreign-born share that we observe need not generate a bias in the estimate if it is not strongly correlated with the second-stage residual. To gauge this potential for omitted variable bias one can include such variables as controls in the SSIV specification and check sensitivity of the coefficient; we report results of this exercise next.

#### 6.2.4 Main Estimates and Sensitivity Analyses

We next estimate the effects of import competition on local labor market outcomes, leveraging within-period exogeneity of the industry shocks  $g_{nt}$ . We then check sensitivity of results to inclusion of the Autor et al. (2013) regional controls and Acemoglu et al. (2016) industry-level controls.

Table 4 reports SSIV coefficients from regressing regional manufacturing employment growth in the U.S. on the growth of import competition from China, instrumented by predicted Chinese import growth.<sup>46</sup> Per the results in Section 5.1, we estimate these coefficients with equivalent industry-level regressions in order to obtain valid exposure-robust standard errors. Consistent with the above analysis of shock ICCs, we cluster standard errors at the SIC3 level. We also report first-stage  $F$ -statistics with corresponding exposure-robust inference. As discussed in Section 5.2, these come from industry-level IV regressions of the aggregated treatment and instrument (i.e.  $\bar{x}_{nt}^\perp$  on  $\bar{z}_{nt}^\perp$ ), instrumented with shocks and weighting by  $s_{nt}$ . The  $F$ -statistics are well above the conventional threshold of ten in all columns of the table.

Column 1 first replicates column 6 of Table 3 in Autor et al. (2013) by including in  $w_{lt}$  period fixed effects, Census division fixed effects, start-of-period conditions (% college educated, % foreign-born, % employment among women, % employment in routine occupations, and the average offshorability index), and the start-of-period manufacturing share. The point estimate is -0.596, with a corrected standard error of 0.114.<sup>47</sup>

As noted, the ADH specification in column 1 does not include the lagged manufacturing share control  $S_{lt}$ , which is necessary to solve the incomplete shares issue in Section 4.2, though it does include a highly correlated control (start-of-period manufacturing share). In column 2 of Table 4

<sup>46</sup>Appendix Table C1 reports estimates for other outcomes in ADH: growth rates of unemployment, labor force non-participation, and average wages, corresponding to columns 3 and 4 of Table 5 and column 1 of Table 6 in ADH.

<sup>47</sup>Appendix Table C2 implements three alternative methods for conducting inference in Table 4, reporting conventional state-clustered standard errors as in ADH (which are not exposure-robust), the Adão et al. (2019) standard errors (which are asymptotically equivalent to ours but differ in finite samples), and null-imposed confidence intervals obtained from shock-level Lagrange multiplier tests (which may have better finite-sample properties). Consistent with the theoretical discussion in Appendix B.2, the conventional standard errors are generally too low, while the Adão et al. (2019) standard errors are slightly larger than those from Table 4 in most columns. Imposing the null widens the confidence interval more substantially, by 30–50%, although more so on the left end, suggesting that much larger effects are not rejected by the data. This last finding is consistent with Adão et al. (2019), except that we use the equivalent industry-level regression to compute the null-imposed confidence interval.

we isolate within-manufacturing variation in shocks by replacing the latter sum-of-share control with the former. The SSIV point estimate remains almost unchanged, at -0.489 (with a standard error of 0.100). Here exposure-robust standard errors are obtained from an industry-level regression that drops the implicit service sector shock of  $g_{0t} = 0$ .

Isolating the within-period variation in manufacturing shocks requires further controls in the incomplete shares case, as discussed in Section 4.3. Specifically, column 3 controls for lagged manufacturing shares interacted with period indicators, which are the share-weighted sums of period effects in  $q_{nt}$ . This is equivalent to the use of period fixed effects in the industry-level analysis of Acemoglu et al. (2016). With these controls the SSIV point estimate is -0.267 with an exposure-robust standard error of 0.099.<sup>48</sup> While the coefficient remains statistically and economically significant, it is smaller in magnitude than the estimates in columns 1 and 2. The difference stems from the fact that 2000–07 saw both a faster growth in imports from China (e.g., due to its entry to the WTO) and a faster decline in U.S. manufacturing. The earlier columns attribute the faster manufacturing decline to increased trade with China, while the specification in Column 3 controls for any unobserved shocks specific to the manufacturing sector overall in the 2000s (e.g., any demand or supply shock affecting the manufacturing sector, which could include automation, innovation, falling consumer demand due to income effects, etc.). Conventional industry-level IV regressions control for such unobserved shocks with period fixed effects, as in Table 3, column 1 of Acemoglu et al. (2016).<sup>49</sup> The translation of their industry specification into the regional setup of ADH requires interacting the lagged manufacturing shares with period indicators, a simple but important insight of our framework.

Column 4 implements a simple sensitivity test to assess the stability of the results when the controls from ADH are omitted. This test is motivated by the result of the balance test in panel B of Table 3, which indicated that the shift-share instrument was correlated with the share of foreign born population. It is therefore instructive to see whether the headline regression coefficient is sensitive to the inclusion of this and other controls. In fact, we find that the results remain very similar without controls, with a point estimate of -0.314 and an exposure-robust standard error of 0.134. We proceed by keeping the ADH controls for the remainder of the analysis.

Further columns of Table 4 parallel the specifications of Acemoglu et al. (2016, Table 3) that

---

<sup>48</sup>Appendix Figure C1 reports binned scatter plots that illustrate the first-stage and reduced-form industry-level relationships corresponding to the column 3 specification. This estimate can be interpreted as a weighted average of two period-specific shift-share IV coefficients. Column 1 of Appendix Table C3 shows the underlying estimates, from a just-identified IV regression where both treatment and the instrument are interacted with period indicators (as well as the manufacturing share control, as in column 3), with exposure-robust standard errors obtained by the equivalent industry-level regression discussed in Section 5. The estimated effect of increased Chinese import competition is negative in both periods (-0.491 and -0.225). Other columns repeat the analysis for other outcomes.

<sup>49</sup>In principle, China could have affected the path of the U.S. manufacturing sector as a whole, and thus the variation in the average China shock across periods may be informative about the effects of interest. However, because of the multiplicity of shocks that may affect the manufacturing sector as a whole in a given period, this variation cannot be viewed as a quasi-experimental source of variation for the impact of trade with China on employment and other outcomes. This is why industry-level studies of the China shock use period fixed effects, possibly reducing power but substantially improving robustness of the estimates. In the ADH application the estimation power is not actually reduced, as the Table 4 column 3 standard error is even slightly smaller than that in the previous columns.

include further industry-level controls. This illustrates how our framework makes it straightforward to introduce more detailed industry-level controls in SSIV, which are commonly used in industry-level studies of the China shock. The validity of these estimates relies on weaker versions of conditional random assignment (Assumption 3), and robustness of the coefficients is therefore reassuring. Specifically, Acemoglu et al. (2016) control for fixed effects of ten broad industry groups (one-digit manufacturing sectors) in column 2 of their Table 3. By Proposition 4, we can exploit shock variation within these industry groups in the SSIV design by controlling for the lagged shares of exposure to these industry groups (and including fixed effects of these groups in the equivalent industry regressions for correct inference, per Section 5.1). The resulting point estimate in column 5 of Table 4 remains very similar to that of column 3, at -0.310 with a standard error of 0.134.

Column 6 instead parallels the specification of Acemoglu et al. (2016) that includes production controls, which we used for the balance tests in Panel A of Table 3. This is done by controlling for the regional share-weighted sums of those controls. The results remain virtually unchanged, with a regression coefficient of -0.293 and an exposure-robust standard error of 0.125.

Finally, column 7 instead introduces industry fixed effects, again following Acemoglu et al. (2016). This specification is more ambitious because it isolates changes in trade with China within each four-digit SIC industry, across the two periods. To translate the industry fixed effects to the location-level setup, we control for the lagged location-specific share of exposure to each industry.<sup>50</sup> The magnitude of the regression coefficient increases, to -0.432, with an exposure-robust standard error of 0.205. Broadly, these results demonstrate the stability of the SSIV regression coefficient under alternative sets of controls, corresponding to different assumptions of conditional quasi-random shock assignment.

The appendix reports estimates from additional specifications. Appendix Table C4 includes additional controls corresponding to other specifications of Acemoglu et al. (2016), Table 3: for example, controlling for observed changes in employment in the pre-periods or combining multiple sets of controls. The regression coefficients remain stable across all specifications. Appendix Table C5 instead shows robustness of the coefficients to using overidentified SSIV procedures (leveraging variation in eight country-specific Chinese import growth, instead of the ADH total), illustrating the theoretical results of Section 4.4. The table also reports a  $p$ -value for the shock-level overidentification test of 0.142, providing further support to the identification assumptions.

### 6.2.5 Discussion

Taken together, the sensitivity, falsification, and overidentification exercises suggest that the ADH approach can be reasonably viewed as leveraging exogenous shock variation via our framework. This is notably in contrast to the analysis of Goldsmith-Pinkham et al. (2020), who find the ADH exposure

---

<sup>50</sup>If the shares used to construct the instrument were time-invariant, a more conventional and intuitive way to exploit over-time variation in the shocks would be by including the regional fixed effects in the regression, as Section 4.3 explained. In the ADH setting where the shares vary over time, they need to be controlled for directly.

shares to be implausible instruments via different balance and overidentification tests. This contrast should perhaps come as no surprise. As mentioned, the exogeneity of industry employment shares is an *ex ante* implausible research design, because it is invalidated by any unobserved labor demand or supply shocks across industries (which we view as an inherent feature of the economy).

In contrast, our approach relies on the exogeneity of the specific ADH trade shocks, allowing for endogenous exposure shares. With this view, the potential confounders are a more specific set of unobserved industry shocks (namely, unobserved shocks that would correlate with the ADH shocks), rather than any unobserved shocks. In principle, the conditions for shock orthogonality could still fail because of these specific unobserved shocks. In practice, our balance tests indicate that there is little evidence to suggest that the ADH shocks are confounded.

Our ADH application therefore illustrates two points. First, the assumptions of our framework are plausible, both *ex ante* and *ex post*, in an influential empirical setting, where an alternative SSIV framework is inapplicable. Second, our framework helps researchers translate shock-level identifying assumptions to appropriate SSIV regression controls, falsify those assumptions with appropriate balance tests, and perform correct inference.

## 7 Conclusion

Shift-share instruments combine a common set of observed shocks with variation in shock exposure. In this paper, we provide a quasi-experimental framework for the validity of such instruments based on identifying variation in the shocks, allowing the exposure shares to be endogenous. Our framework revolves around novel equivalence results: the orthogonality between a shift-share instrument and an unobserved residual can be represented as the orthogonality between the underlying shocks and a shock-level unobservable, and SSIV regression coefficients can be obtained from a transformed shock-level regression with shocks directly used as an instrument. Shift-share instruments are therefore valid when shocks are idiosyncratic with respect to an exposure-weighted average of the unobserved factors determining the outcome variable, and yield consistent IV estimates when the number of shocks is large and they are sufficiently dispersed in terms of their average exposure.

Through various extensions and illustrations, we show how our quasi-experimental SSIV framework can guide empirical work in practice. By controlling for exposure-weighted averages of shock-level confounders, researchers can isolate more plausibly exogenous variation in shocks, such as over time or within narrow industry groups. By estimating SSIV coefficients, placebo regressions, and first stage  $F$ -statistics at the level of shocks, researchers can conveniently perform exposure-robust inference that accounts for the inherent non-standard clustering of observations with common shock exposure. Our shock-level analysis also raises new concerns: SSIV designs with few or insufficiently dispersed shocks may have effectively small samples, despite there being many underlying observations, and

instruments constructed from exposure shares that do not add up to a constant require appropriate controls in order to isolate quasi-random shock variation. We illustrate these practical implications in an application to the influential study of Autor et al. (2013).

In sum, our analysis formalizes the claim that SSIV identification and consistency may arise from the exogeneity of shocks, while providing new guidance for SSIV estimation and inference that may be applied across a number of economic fields, including international trade, labor economics, urban economics, macroeconomics, and public finance. Our shock-level assumptions connect SSIV in these settings to conventional shock-level IV estimation, bringing shift-share instruments to more familiar econometric territory and facilitating the assessment of SSIV credibility in practice.

**Data Availability Statement:** The data underlying this article are publicly available on Zenodo, at <http://doi.org/10.5281/zenodo.4619197>.

## Figures and Tables

Table 1: Shock Summary Statistics in the Autor et al. (2013) Setting

	(1)	(2)	(3)
Mean	1.79	7.37	0
Standard deviation	10.79	20.92	20.44
Interquartile range	0	6.61	6.11
<u>Specification</u>			
Excluding service industries		✓	✓
Residualizing on period FE			✓
<u>Effective sample size (1/HHI of <math>s_{nt}</math> weights)</u>			
Across industries and periods	3.5	191.6	191.6
Across SIC3 groups	1.7	58.4	58.4
<u>Largest <math>s_{nt}</math> weight</u>			
Across industries and periods	0.398	0.035	0.035
Across SIC3 groups	0.757	0.066	0.066
<u>Observation counts</u>			
# of industry-period shocks	796	794	794
# of industries	398	397	397
# of SIC3 groups	137	136	136

Notes: This table summarizes the distribution of China import shocks  $g_{nt}$  across industries  $n$  and periods  $t$  in the Autor et al. (2013) application. Shocks are measured as the total flow of imports from China in eight developed economics outside of the United States. All statistics are weighted by the average industry exposure shares  $s_{nt}$ ; shares are measured from lagged manufacturing employment, as described in Section 6.2.1. Column 1 includes the non-manufacturing industry aggregate in each period with a shock of 0, while columns 2 and 3 restrict the sample to manufacturing industries. Column 3 residualizes manufacturing shocks on period indicators. We report the effective sample size (the inverse renormalized Herfindahl index of the  $s_{nt}$  weights, as described in Section 6.2.2) with and without the non-manufacturing industry, at the industry-by-period level and at the level of SIC3 groups (aggregated across periods), along with the largest  $s_{nt}$ .

Table 2: Shock Intra-Class Correlations in the Autor et al. (2013) Setting

	Estimate	SE
	(1)	(2)
<u>Shock ICCs</u>		
10 sectors	0.016	(0.022)
SIC2	0.047	(0.052)
SIC3	0.073	(0.057)
Industry	0.169	(0.047)
<u>Period means</u>		
1990s	4.65	(1.38)
2000s	16.87	(3.34)
# of industry-periods	794	

Notes: This table reports intra-class correlation coefficients for the Autor et al. (2013) manufacturing shocks, estimated from the hierarchical model described in Section 6.2.2. Estimates come from a maximum likelihood procedure with an exchangeable covariance structure for each industry and sector random effect and with period fixed effects. Robust standard errors are reported in parentheses.

Table 3: Shock Balance Tests in the Autor et al. (2013) Setting

## Panel A: Industry-Level Balance

Balance variable	Coef.	SE
Production workers' share of employment, 1991	-0.011	(0.012)
Ratio of capital to value-added, 1991	-0.007	(0.019)
Log real wage (2007 USD), 1991	-0.005	(0.022)
Computer investment as share of total, 1990	0.750	(0.465)
High-tech equipment as share of total investment, 1990	0.532	(0.296)
# of industry-periods		794

## Panel B: Regional Balance

Balance variable	Coef.	SE
Start-of-period % of college-educated population	0.915	(1.196)
Start-of-period % of foreign-born population	2.920	(0.952)
Start-of-period % of employment among women	-0.159	(0.521)
Start-of-period % of employment in routine occupations	-0.302	(0.272)
Start-of-period average offshorability index of occupations	0.087	(0.075)
Manufacturing employment growth, 1970s	0.543	(0.227)
Manufacturing employment growth, 1980s	0.055	(0.187)
# of region-periods		1,444

Notes: Panel A of this table reports coefficients from regressions of the industry-level covariates in Acemoglu et al. (2016) on the Autor et al. (2013) shocks, controlling for period indicators and weighting by average industry exposure shares. Standard errors are reported in parentheses and allow for clustering at the level of three-digit SIC codes. Panel B reports coefficients from regressions of commuting zone-level covariates and pre-trends from Autor et al. (2013) on the shift-share instrument, controlling for period indicators interacted with the lagged manufacturing share. Balance variables (the first five rows of this panel) vary across the two periods, while pre-trends (the last two rows) do not. SIC3-clustered exposure-robust standard errors are reported in parentheses and obtained from equivalent industry-level IV regressions as described in Section 6.2.3. Independent variables in both panels are normalized to have a variance of one in the sample.

Table 4: Shift-Share IV Estimates of the Effect of Chinese Imports on Manufacturing Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Coefficient	-0.596	-0.489	-0.267	-0.314	-0.310	-0.290	-0.432
	(0.114)	(0.100)	(0.099)	(0.107)	(0.134)	(0.129)	(0.205)
<u>Regional controls</u>							
Autor et al. (2013) controls	✓	✓	✓		✓	✓	✓
Start-of-period mfg. share	✓						
Lagged mfg. share		✓	✓	✓	✓	✓	✓
Period-specific lagged mfg. share			✓	✓	✓	✓	✓
Lagged 10-sector shares					✓		✓
Local Acemoglu et al. (2016) controls						✓	
Lagged industry shares							✓
SSIV first stage $F$ -stat.	185.6	166.7	123.6	272.4	64.6	63.3	27.6
# of region-periods	1,444	1,444	1,444	1,444	1,444	1,444	1,444
# of industry-periods	796	794	794	794	794	794	794

Notes: This table reports shift-share IV coefficients from regressions of regional manufacturing employment growth in the U.S. on the growth of import competition from China, instrumented with predicted China import growth as described in Section 6.2.1. Column 1 replicates column 6 of Table 3 in Autor et al. (2013) by controlling for period fixed effects, Census division fixed effects, start-of-period conditions (% college educated, % foreign-born, % employment among women, % employment in routine occupations, and the average offshorability index), and the start-of-period manufacturing share. Column 2 replaces the start-of-period manufacturing shares control with the lagged manufacturing shares underlying the instrument, while column 3 interacts this control with period indicators. Column 4 removes the Census division fixed effects and start-of-period covariates. Columns 5–7 instead add exposure-weighted sums of industry controls from Acemoglu et al. (2016): indicators of 10 industry sectors (column 5), production controls (column 6), and indicators of 397 industries (column 7). Production controls are: employment share of production workers, ratio of capital to value-added, log real wage (all measured in 1991); and computer investment as share of total and high-tech equipment as share of total employment (both measured in 1990). Exposure-robust standard errors (reported in parentheses) and first-stage  $F$ -statistics are obtained from equivalent industry-level IV regressions, as described in the text, allowing for clustering of shocks at the level of three-digit SIC codes. For commuting zone controls which have a shift-share structure (all controls starting with the lagged manufacturing share), we include the corresponding  $q_{nt}$  controls in the industry-level IV regression. The sample in columns 2–7 includes 722 locations (commuting zones) and 397 industries, each observed in two periods; the estimate in column 1 implicitly includes an additional two observations for the non-manufacturing industry with a shock of zero in each period.

## References

- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2019. "Sampling-based vs. Design-based Uncertainty in Regression Analysis." *Working Paper*.
- Acemoglu, Daron, David H. Autor, David Dorn, Gordon H. Hanson, and Brendard Price. 2016. "Import Competition and the Great U.S. Employment Sag of the 2000s." *Journal of Labor Economics* 34:S141–S198.
- Acemoglu, Daron, and Pascual Restrepo. Forthcoming. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy*.
- Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *The Quarterly Journal of Economics* 134:1949–2010.
- Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. 2017. "The European Trust Crisis and the Rise of Populism." *Brookings Papers on Economic Activity* 2017:309–400.
- Angrist, Joshua D., Kate Graddy, and Guido W. Imbens. 2000. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *The Review of Economic Studies* 67:499–527.
- Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14:57–67.
- Angrist, Joshua D., and JS Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103:2121–2168.
- Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *The Quarterly Journal of Economics* 118:157–205.
- Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.
- Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62:657.
- Berman, Nicolas, Antoine Berthou, and Jérôme Héricourt. 2015. "Export Dynamics and Sales at Home." *Journal of International Economics* 96:298–310.
- Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig. 2017. "This mine is mine! How minerals fuel conflicts in Africa." *American Economic Review* 107:1564–1610.
- Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, 1–75.
- Bombardini, Matilde, and Bingjing Li. 2019. "Trade, Pollution and Mortality in China." *Working Paper*.

- Borusyak, Kirill, and Xavier Jaravel.** 2017. “Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume.” Working Paper.
- Bound, John, David A. Jaeger, and Regina M. Baker.** 1995. “Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak.” *Journal of the American Statistical Association* 90:443–450.
- Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric M. Zolt.** 2013. “The Effect of Rising Income Inequality on Taxation and Public Expenditures: Evidence from U.S. Municipalities and School Districts, 1970-2000.” *The Review of Economics and Statistics* 95:1291–1302.
- Broxterman, Daniel A, and William D Larson.** 2018. “An Examination of Industry Mix Demand Indicators : The Bartik Instrument at Twenty-Five.” Working Paper.
- Callaway, Brantly, and Pedro H. C. Sant’Anna.** Forthcoming. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics*.
- Card, David.** 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” *Journal of Labor Economics* 19:22–64.
- . 2009. “Immigration and Inequality.” *American Economic Review* 99:1–21.
- Chaisemartin, Clément de, and Xavier D’Haultfoeuille.** 2019. “Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects.” NBER Working Paper No. 25904.
- Diamond, Rebecca.** 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review* 106:479–524.
- Fajgelbaum, Pablo D., Pinelopi K. Goldberg, Patrick J. Kennedy, and Amit K. Khandelwal.** 2020. “The Return to Protectionism.” *Quarterly Journal of Economics* 135:1–55.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift.** 2020. “Bartik Instruments: What, When, Why, and How.” *American Economic Review* 110:2586–2624.
- Goodman-Bacon, Andrew.** 2018. “Difference-in-Differences with Variation in Treatment Timing.” NBER Working Paper 25018.
- Hull, Peter.** 2018. “Estimating Treatment Effects in Mover Designs.” Working Paper.
- Hummels, David, Rasmus Jorgensen, Jakob Munch, and Chong Xiang.** 2014. “The Wage Effects of Offshoring: Evidence From Danish Matched Worker-Firm Data.” *American Economic Review* 104:1597–1629.
- Imbert, Clement, Marlon Seror, Yifan Zhang, and Yanos Zylberberg.** 2019. “Migrants and Firms: Evidence from China.” Working Paper.
- Jaeger, David A, Joakim Ruist, and Jan Stuhler.** 2018. “Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration.” Working Paper.
- Jaravel, Xavier.** 2019. “The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector.” *The Quarterly Journal of Economics* 134:715–783.

- Kolesar, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens.** 2015. "Identification and Inference With Many Invalid Instruments." *Journal of Business and Economic Statistics* 33:474–484.
- Kovak, Brian K.** 2013. "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" *American Economic Review* 103:1960–1976.
- Mohnen, Paul.** 2019. "The Impact of the Retirement Slowdown on the U.S. Youth Labor Market." *Working Paper*.
- Montiel Olea, Jose Luis, and Carolin Pflueger.** 2013. "A robust test for weak instruments." *Journal of Business and Economic Statistics* 31:358–369.
- Moulton, Brent R.** 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32:385–397.
- Nunn, Nathan, and Nancy Qian.** 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104:1630–1666.
- Peri, Giovanni, Kevin Shih, and Chad Sparber.** 2016. "STEM workers, H-1B Visas, and productivity in US cities." *Journal of Labor Economics* 49:277–307.
- Sloczyński, Tymon.** Forthcoming. "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights." *Review of Economic Statistics*.
- Stuen, Eric T., Ahmed Mushfiq Mobarak, and Keith E. Maskus.** 2012. "Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in U.S. Doctoral Programmes." *The Economic Journal* 122:1143–1176.
- Suárez Serrato, Juan Carlos, and Owen Zidar.** 2016. "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms." *American Economic Review* 106:2582–2624.
- Sun, Liyang, and Sarah Abraham.** Forthcoming. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*.
- Xu, Chenzi.** 2019. "Reshaping Global Trade: The Immediate and Long-Run Effects of Bank Failures." *Working Paper*.