A Practical Guide to Shift-Share IV

Kirill Borusyak, Berkeley

Peter Hull, Brown

Xavier Jaravel, LSE*

March 2024

Abstract

A recent econometric literature shows two distinct paths for credible identification with shift-share instrumental variables: one leveraging many exogenous shocks and one leveraging exogenous shares. We present the core logic of these two paths and practical takeaways from this literature via simple "checklists" that applied researchers can use to follow either path. A variety of empirical settings are used for illustration.

*Contact: k.borusyak@berkeley.edu, peter_hull@brown.edu, and x.jaravel@lse.ac.uk.

1 Introduction

Many economic studies consider units that are differentially exposed to a common set of shocks. For example, a study of local labor supply might leverage the fact that regional labor markets are differentially exposed to a common set of labor demand shocks across industries. To identify causal effects or structural parameters in such settings, researchers often use shift-share instrumental variables (SSIVs, also called "Bartik instruments"), which sum or average the shocks (or "shifts") with heterogeneous exposure weights (or "shares"). A researcher might, for instance, construct a SSIV that averages some measured industry demand shocks (e.g. changes to federal subsidies) weighting by local industry employment shares. Such instruments are widely found in studies of labor, trade, macroeconomics, immigration, public economics, finance, and other topics.[1]

When do such instruments "work," and when might they "fail"? Answering these questions is challenging because SSIVs combine two distinct sources of variation—the shocks and the shares—and it can be unclear what properties of these objects are important or helpful for addressing endogeneity concerns. Intuitively, one might view the shocks as helpful because they represent changes to the system under study, which could perhaps be viewed as exogenous. However, unlike in conventional settings, these shocks operate at a different "level" (e.g. industries) than the unit of analysis (e.g. local labor markets). Are they still useful then? Moreover, the shares are usually predetermined (e.g., employment shares are measured in a pre-period). So how can their potential exogeneity or endogeneity be understood?

This article gives conceptual answers to such questions and provides practical guidance for constructing and using shift-share instruments, building on a recent econometric literature. The literature suggests two distinct paths to identification with SSIVs. One path, developed by Borusyak et al. (2022, henceforth BHJ) and Adao et al. (2019), leverages many exogenous shocks while making no assumption on the exogeneity of the shares. The second path, proposed by Goldsmith-Pinkham et al. (2020, henceforth GPSS), instead focuses on share exogeneity. Each of these two approaches has distinct practical implications regarding appropriate estimators, ways to conduct valid inference, and diagnostic tests.

Our practical guide begins in Section 2 with a discussion of broad motivations for using SSIVs and an overview of the core identification logic underlying the two paths. We discuss how BHJ-style identification "from the shocks" can be understood as leveraging a shock-level natural experiment, while GPSS-style identification "from the shares" can be viewed as pooling together multiple differences-in-differences designs which leverage differential shock exposure. Sections 3 and 4 then provide two checklists of practical steps that researchers can follow when implementing a shift-share design, considering BHJ and GPSS approaches in turn. We illustrate key concepts and practical steps with examples throughout, and several extensions are given in the appendix.

Appendix A answers other practical questions that seem to come up often when using SSIVs—

[1] See, e.g., Bartik (1991), Blanchard and Katz (1992), Autor et al. (2013), Hummels et al. (2014), Nakamura and Steinsson (2014), Jaravel (2019), Oberfield and Raval (2021), Card (2001, 2009), Saiz (2010), Diamond (2016), Greenstone et al. (2020), and Xu (2022). We discuss several of these applications in detail below.

such as how to interpret SSIV estimates as local averages of heterogeneous effects, how to handle multiple SSIVs and interaction terms, and whether a "leave-out" construction of the shocks is helpful. While the econometric content of this article largely follows BHJ and GPSS, we establish some new results along the way—namely, regarding SSIV regressions with multiple endogenous variables (Appendix B) and the role of share exogeneity in the canonical setting of Bartik (1991) (Appendix C). We take an applied perspective throughout; see Borusyak et al. (2023a) for a more technical review of the recent econometric literature on SSIV.

2 Shift-Share Basics

2.1 What are Shift-Share IVs and Where Do They Come From?

Shift-share instruments (SSIVs) have the form:

$$z_i = \sum_{k=1}^{K} \underbrace{s_{ik}}_{\text{Share}} \underbrace{g_k}_{\text{Shift}}, \tag{1}$$

where $(g_1, \ldots, g_K)$ is a set of shocks (a.k.a. shifts) that is common to all observations i and the $(s_{i1}, \ldots, s_{iK})$ are sets of exposure weights (a.k.a. shares) that vary across observations $i = 1, \ldots, N$. In many applications the shares sum to one for each observation, $\sum_{k=1}^{K} s_{ik} = 1$, such that z_i is a share-weighted average of the shocks.

Researchers use such z_i to instrument for an endogenous variable (or "treatment") x_i in a causal or structural model of the form:

$$y_i = \beta x_i + \gamma' w_i + \varepsilon_i, \tag{2}$$

where y_i is an outcome of interest, w_i is a vector of observed controls, ε_i is an unobserved error term. Here β is the parameter of interest.[2] "Reduced-form" shift-share regressions correspond to a special case where $x_i = z_i$ and β is the effect of the shift-share treatment. Outside of this case, z_i is used to address potential endogeneity of x_i: i.e., its possible correlation with ε_i.

Table 1 lists several prominent SSIV examples from a variety of empirical settings. We return to some of these examples to help illustrate our two checklists below. For now, the table shows the diversity of possible outcomes, treatments, shares, and shocks used in SSIV designs—for analyses at the levels of countries, regions, firms, products, and individuals.

Researchers might motivate SSIV constructions in different ways. One common motivation arises when the endogenous variable measures the growth of some economic variable over time, and can be decomposed into some start-of-period shares and over-time shifts. Suppose, for example, that $x_i = \frac{X_{i1} - X_{i0}}{X_{i0}}$ is the growth in employment X_{it} for local labor market i over two periods, $t = 0, 1$.

[2] We presume (2) is a correctly specified model, with a constant treatment effect β. We discuss an extension to heterogeneous effects in Section A. See Borusyak et al. (2023a) for further discussion of model misspecification, e.g. in the case where the γ coefficients are also heterogeneous.

Table 1: Shift-share IV Examples

Paper	i	Outcome y_i	Treatment x_i	k	SSIV	
					Share s_{ik}	Shock g_k
Bartik (1991)	Region	Δ Local wage	Δ Local employment	Industry	Employment_{ik}/ Employment_i	National growth of industry employment
Autor et al. (2013)	Region	Δ Local manufacturing employment	Δ Local imports from China	Industry	Employment_{ik}/ Employment_i	Δ imports from China in other countries
Hummels et al. (2014)	Worker	Wage	Imports of intermediate goods by employer	Product-by-country	Imports_{ik}/ Imports_i	Imports from k to other countries
Aghion et al. (2022)	Firm	Δ Firm employment	Δ Firm stock of automation technologies	Technology-by-country	Imports_{ik}/ Imports_i	Δ imports from k to other countries
Card (2009)	Region	Relative wage of migrants vs. natives	Relative employment of migrants vs. natives	Origin country	Migrant stock_{ik}/ Population_i	New migrants_k/ Migrant stock_k
Cai et al. (2015)	Individual	Takeup of insurance	% of friends selected for an information session*	Individual	Dummy(k is friend of i)/ # of friends i has	Dummy of information session
Jaravel (2019)	Product category	Inflation and innovation	Δ Quantity demanded	Socio-demographic group	Sales of i to group k/ Total sales of i	Population change
Greenstone et al. (2020)	Region	Δ Employment	Δ Credit	Bank	Credit market share of k	Estimated credit supply shock
Xu (2022)	Region	Δ Exports	Exposure to banking crisis*	Bank	Credit market share of k	Bankruptcy during banking crisis
Miguel and Kremer (2004)	Individual	Measures of health or education	Number of neighbors selected for deworming*	Individual	Dummy(k is friend of i)	Dummy of deworming treatment
Nunn and Qian (2014)	Country-by-year	Conflict	Quantity of food aid (wheat) from the U.S.	Year	Fraction of years with non-zero food aid	U.S. wheat production in previous year
Franklin et al. (2023)	Local labor market	Wage	Shift-share exposure to the intervention*	Residential neighborhood	Commuters_{ik}/ Employment_i	Dummy of public works intervention
Mohnen (2024)	Region	Δ Young labor market outcome	Retirement rate	Age group (within 45+)	Population_{ik}/ Population $45+_i$	National retirement rate at age k

Notes: We simplify many of the settings, suppressing the time dimension (except where it is central to the design), controls and fixed effects, interaction terms, log and other transformations of the outcome and treatment, etc. Asterisks (*) indicate OLS regressions, in which the treatment itself is the shift-share with shares s_{ik} and shocks g_k.

Regional employment can be decomposed across industries: $X_{it} = \sum_k X_{ikt}$ where X_{ikt} denotes the period-t employment of industry k in local labor market i. This leads to a decomposition of regional employment growth rates in terms of a set of period-0 industry employment shares and a set of local industry growth rate shifts:

$$x_i = \frac{\sum_k X_{ik1} - \sum_k X_{ik0}}{X_{i0}} = \sum_k \underbrace{\frac{X_{ik0}}{X_{i0}}}_{\text{Share}} \cdot \underbrace{\frac{X_{ik1} - X_{ik0}}{X_{ik0}}}_{\text{Local shift}}. \tag{3}$$

A researcher might construct a SSIV by choosing a set of common shocks g_k to replace the local shocks in such a decomposition. The researcher might keep the shares from the decomposition or also choose to replace them, e.g. with further lagged shares.[3]

To illustrate this motivation, consider a concrete example inspired by the canonical shift-share instrument of Bartik (1991) and Blanchard and Katz (1992). The goal in these papers is to estimate the (inverse) elasticity of regional labor supply β from employment growth x_i and wage growth y_i across regions i. Of course, in order to estimate a supply elasticity, we need an instrument that shifts labor demand. Decomposition (3) captures the idea that x_i averages local employment growth across different industries, $\frac{X_{ik1}-X_{ik0}}{X_{ik0}}$, using initial employment shares $s_{ik} = \frac{X_{ik0}}{X_{i0}}$ as weights. The local shifts reflect changes to both labor demand and labor supply. To isolate demand variation, we can form an instrument that keeps the local industry employment shares from the decomposition but introduces a set of common shocks. The shocks are meant to be predictive of the local industry growth rates while only capturing demand variation. Bartik (1991) and Blanchard and Katz (1992), for instance, proxy for aggregate demand shocks with national industry growth rates as g_k. One might also define g_k as specific industry demand shocks, such as the growth in government subsidies.

Instruments constructed in this way are likely to satisfy the usual IV relevance condition: z_i is likely correlated with x_i when the shocks and exposure shares in the instrument are correlated with, respectively, the shocks and shares in the treatment decomposition. Of course, for the instrument to be valid it must also satisfy an exogeneity condition: z_i must be uncorrelated with the error term ε_i. The plausibility of this condition will depend on the researcher's choice of shocks and shares, as we discuss below.

A decomposition of the treatment is not the only way to motivate a SSIV: in some cases, an instrument naturally takes a shift-share form. For instance, many network spillover studies (e.g., Cai et al. (2015)) use as an instrument the fraction of unit i's friends or neighbors (within some set of units $k = 1, \ldots, K$ which may or may not be the same as i) who have been selected for some intervention. This variable inherently has a shift-share structure. Indeed, denoting by g_k the dummy that k has been selected, we can write the share of friends who were selected for the intervention as

$$z_i = \sum_k \frac{\mathbf{1}\left[k \text{ is a friend of } i\right]}{\#\text{ of friends } i \text{ has}} \cdot g_k.$$

[3]When $x_i = \log X_{i1} - \log X_{i0}$, such that β can properly be interpreted as an elasticity, think of the decomposition in (3) as a first-order approximation.

Here the intervention dummies serve as shocks and friendship dummies (rescaled to sum to one for each i), serve as shares.

2.2 Why is There a Methodological Literature about SSIV Exogeneity?

As in all IV designs, it is important to make a case for instrument exogeneity: i.e., for z_i being uncorrelated with the error term ε_i. This case is typically made from contextual knowledge about the source of variation in the instrument. The key challenge with shift-share research designs is that there are two distinct sources of variation: the shocks and the shares. To argue convincingly that the SSIV is exogenous, one must explain what properties of the shocks and shares make z_i and ε_i uncorrelated, rather than simply stating the basic exogeneity restriction of $\mathbb{E}\left[z_i \varepsilon_i\right] = 0$.

Natural questions arise at this point. Do both the shocks and the shares have to be exogenous in some sense, or is the exogeneity of either of them sufficient? How should exogeneity of shocks and shares be understood anyway? The shocks do not vary across observations, so what does it mean for them to be correlated or uncorrelated with the error term? Moreover, the shares are predetermined and thus cannot suffer from reverse causality; what then are the main potential threats to their exogeneity?

A recent econometric literature has tackled these questions, showing there are two non-nested paths to ensuring exogeneity of the SSIV: one focusing on shock exogeneity (while using the shares to generate variation across observations) and the other focusing on share exogeneity (while using shocks to combine the shares). Each of these paths corresponds to a distinct set of sufficient identification conditions and practical steps for how to choose controls, perform statistical inference, and run diagnostic tests. We devote the rest of this section to introducing the core logic of these two identification strategies, before describing their practical implementation in Sections 3 and 4.

2.3 What is Identification From Many Exogenous Shocks?

One strategy to ensure that the SSIV z_i is exogenous is to have exogenous g_k: shocks that are as-if randomly assigned (in their dimension) and which only affect the outcome through the endogenous variable. In the above labor supply setting, for example, one might consider a lottery that randomly assigns subsidies g_k to each industry. Subsidies can be viewed as only affecting wages by shifting labor demand and do not have direct effects on labor supply, satisfying the relevant exclusion restriction.

Shock-based identification stems from a simple observation: if the shocks arise from a pure lottery, the shift-share instrument is also like a lottery outcome. This is true even if the shares that average the lottery draws are econometrically endogenous, in the sense that units with different shares may have systematically different unobservables. For instance, regions that specialize in high-tech industries may experience more immigration such that initial high-tech employment shares positively correlate with unobserved immigrant labor supply shocks in the error term. But as long as subsidies are assigned at random across industries, on average the places specializing in high-tech industries will have typical values of the instrument: a share-weighted average of random subsidies is itself as-good-as-random, and may be a valid instrument if it affects the outcome through the endogenous variable only. Thus, a SSIV research design based on experimental shocks puts no assumptions on the exogeneity of exposure shares.

While a lottery provides intuition for an idealized experiment, the necessary and sufficient condition for SSIV exogeneity is a weaker orthogonality condition at the shock level. The shocks g_k should be uncorrelated with the share-weighted average error term where averaging is across units: i.e., the error term of the units most exposed to that shock.[4] For instance, the regional exposure to industry subsidies (i.e., the SSIV) is uncorrelated with the regional labor supply shocks in the error if and only if the subsidy is not systematically different in industries which have higher employment shares in regions with high vs. low labor supply shocks. Violations of this orthogonality are thus the key threat to identification in the exogenous shocks approach.[5]

Another way to see the exogenous shocks logic is to view SSIV as a "translation device" for a set of as-good-as-random shocks to a different level of analysis. For instance, when industry subsidies are as-good-as-randomly assigned, one could imagine running an industry-level IV analysis which uses the subsidy g_k directly as an instrument for industry employment. Specifying the equation at the level of local labor markets may define a more interesting economic parameter, as it captures spillovers when workers move across industries in response to the subsidies. However, the key identification assumption is the same, with the SSIV translating the industry-level natural experiment to the local labor markets (i.e., SSIV $\sum_k s_{ik} g_k$ instrumenting for regional labor demand).

The "weighted average of lotteries" logic highlights two other requirements of the exogenous shocks approach. First, it requires many shocks since with only a few industries there will be no law of large numbers to ensure that the random subsidies satisfy the orthogonality condition described above.[6] BHJ show that a useful measure of the "effective" number of shocks is the inverse Herfindahl–Hirschman index of average exposure: $1/\sum_k s_k^2$, where $s_k = \frac{1}{N}\sum_i s_{ik}$. This index can be small even with a large number of shocks K if a few shocks dominate the SSIV in the sense of having large average shares.

Second, the shares have to add up to one such that the SSIV has an interpretation as a share-weighted *average* of shocks rather than a share-weighted *sum*. Otherwise, even if shocks are drawn fully at random, the instrument may systematically vary across units through the sum of shares. We discuss in Section 3 how the exogenous shocks approach extends in this "incomplete shares" case, which may naturally arise in some applications (e.g., if exogenous subsidies only arise for manufacturing industries but employment shares are measured relative to total employment in all industries).

[4] In a similar spirit, BHJ show a numerical equivalence of SSIV estimates and estimates from a certain shock-level IV specification which uses g_k as the instrument (see their Proposition 1).

[5] Formally, the relevant condition is $\mathbb{E}\left[\sum_k s_k g_k \bar{\varepsilon}_k\right] = 0$—the orthogonality of g_k and $\bar{\varepsilon}_k$ when weighted by s_k—where $\bar{\varepsilon}_k = \frac{\sum_i s_{ik}\varepsilon_i}{\sum_i s_{ik}}$ is the error term averaged across observations i with s_{ik} weights and $s_k = \frac{1}{N}\sum_i s_{ik}$ is the average share for k. BHJ show that this orthogonality is satisfied when the shocks are as-good-as-randomly assigned (in a particular mean-independence sense; see their Assumption 1). The shock-level orthogonality condition could also be satisfied in other ways, e.g. by parallel trend restrictions placed on $\bar{\varepsilon}_k$.

[6] One can think of this as a clustering problem: there are effectively only a few exogenous comparisons, regardless of how many regions are observed. The same logic suggests that even with many industries one needs to correct standard errors, as we discuss in Section 3.

2.4 What is Identification from Exogenous Shares?

A different strategy to ensure SSIV exogeneity is to have exogenous shares. This approach is easiest to see in the special case of only one non-zero shock, i.e. $g_k \neq 0$ for one $k = \kappa$ with $g_k = 0$ for all other k. In this case $z_i = \sum_{k \neq \kappa} s_{ik} \cdot 0 + s_{i\kappa} g_\kappa = s_{i\kappa} g_\kappa$ is perfectly collinear with $s_{i\kappa}$, so using this single exposure share as the instrument will produce numerically the same estimate as using the SSIV. In other words, the IV estimator using z_i effectively compares units with higher vs. lower exposure to the shock, such that IV exogeneity holds when $s_{i\kappa}$ is uncorrelated with the error term ε_i. GPSS extend this logic to the case when all shocks are non-zero, showing that SSIV estimates can generally be viewed as pooling together K "one-at-a-time" IV estimates each using one of the s_{ik} shares as instruments.

What does it mean for shares to be exogenous? One could imagine the set of s_{ik} being as-good-as-randomly assigned to units, as if drawn in a lottery, and satisfying an exclusion restriction that the shares affect the outcome only via the treatment of interest. Alternatively, when the outcome is measured in changes, one may interpret share exogeneity as a set of parallel trends conditions, similar to difference-in-differences (DiD) strategies. That is, for s_{ik} to be uncorrelated with ε_i one could assert that, if not for any change in the endogenous variables, outcomes would have trended similarly for units that were more vs. less exposed to shock k. Share exogeneity requires such parallel trends to hold for each k.

A concrete example inspired by Card (2009) helps illustrate this logic. Suppose we are interested in estimating the (inverse) elasticity of substitution between migrant vs. native workers in labor demand, β. The estimating equation relates changes in the relative wages of migrants vs. natives between two periods, y_i, to changes in the relative employment between these groups, x_i, across local labor markets. Suppose that between these periods we saw a sudden change in national migrant inflows from a particular origin country κ, such as the sudden inflow of Cuban immigrants following the Mariel Boatlift as studied by Card (1990). One might assume that regions which were more or less exposed to this inflow, as captured by the initial share of migrants from Cuba $s_{i\kappa}$, would have seen similar trends in migrant vs. native wages if not for any change in migrant vs. native employment: i.e., that $\mathbb{E}[\varepsilon_i \mid s_{i\kappa}] = 0$. In this case, the Cuban migrant share would be a valid instrument for identifying β. One could further imagine sudden changes in migrant inflows across many origin countries, to different extents. In this case, if a parallel trends condition holds with respect to each exposure share, a SSIV combining them with some g_k weights will also be a valid instrument.

The plausibility of share exogeneity depends on whether there are conceivably any unobserved shocks that affect the outcome via the same (or similar) shares as the ones used to construct the SSIV. Even if shares are drawn at random from a lottery, the presence of any such shocks would always lead to parallel trend violations (see BHJ for a formal version of this argument). The plausibility of share exogeneity is thus boosted by constructing the SSIV with shares which are "tailored" to the treatment of interest, in the sense of mediating only the shocks to x_i and not a broad set of shocks that might affect y_i.

For example, many papers use SSIV designs with shares reflecting local industrial composition to understand the regional impacts of specific industry shocks, such as import competition with China in Autor et al. (2013) or robotization in Acemoglu and Restrepo (2020). The industry employment shares are "generic," in that they could potentially measure an observation's exposure to other shocks (essentially, any industry shock), many of them unobserved. In studies using such shares, it would not be plausible to make a case for identification based on the exogeneity of shares (but one could consider the "shocks" view instead). According to the exogenous shares view of SSIV, Autor et al. (2013) and Acemoglu and Restrepo (2020) use essentially the same instruments (lagged employment shares) for different endogenous variables. In fact, there is no presumption that these exposure shares are more strongly related to import competition, robot adoption, or any other economic transformation. In contrast, studies like Card (2009) use exposure shares that are more tailored to the research question (immigration).

The role of the shocks is secondary with the exogenous shares strategy: GPSS show that the shocks affect the weights in their representation of SSIV as pooled one-at-a-time share-IV estimates, but they do not affect the identification of β so long as the shares are exogenous. The choice of g_k, however, may affect the power of the SSIV. Intuitively, the decomposition (3) suggests that a powerful instrument might use as the g_k the average of shifts x_{ik} across units (e.g. replacing the local growth rates of industry employment x_{ik} with the national ones g_k).[7]

3 SSIV with Exogenous Shocks

We now describe a list of practical steps for applying SSIV designs with many exogenous shocks. To start, we focus on the case with a single cross-section of outcomes y_i and treatments x_i capturing the change in variables over a period of a time. We extend this setup to consider multiple periods below.

We develop the exogenous-shocks checklist with the labor supply setting from above, where g_k represents as-good-as-randomly assigned federal subsidies to industries k. We discuss several real-world examples at the end of this section.

3.1 A Checklist for Applied Work

1. Motivate the shift-share strategy with a shock-level idealized experiment Compelling exogenous-shock SSIV strategies are motivated by a shock-level experimental ideal, which would address the key sources of endogeneity in equation (2). This experimental ideal can be developed in two sub-steps.

First, a researcher can present the main sources of potential bias when estimating equation (2) by ordinary least squares (OLS). Such a discussion is natural at the beginning of any IV analysis, as it spells out exactly what unobserved economic shocks we expect to enter ε_i and thus what

[7] Of course, one cannot use the x_{ik} as weights since they vary across observations and thus would not yield a combination of the share instruments.

unobservables we need to claim are uncorrelated with z_i. For example, when equation (2) represents an inverse labor supply equation, ε_i includes unobserved local labor supply shocks (e.g. immigration of foreign workers to region i). Since equilibrium employment x_i arises from both labor supply and labor demand shocks, it is generally correlated with ε_i, causing OLS estimates of β to be biased. To identify β, we need a demand-side instrument which is uncorrelated with local labor supply changes. Clearly describing the main sources of endogeneity can help readers appreciate the extent to which the proposed shift-share IV design addresses the limitations of OLS.

Second, a researcher can describe a hypothetical shock-level experiment which would generate shocks that are unrelated to the specified sources of bias while generating variation in the endogenous variable. For example, one can imagine assigning new federal subsidies at random across industries. Industries receiving larger subsidies are likely to expand their production and thus their demand for local workers, increasing x_i through the local shifts in decomposition (3). By virtue of random assignment), these subsidy shocks are unrelated to local labor supply conditions. A SSIV which replaces the local shifts with the subsidy shocks g_k in this decomposition would thus be valid.

The experimental ideal is useful in two ways. First, it clarifies exactly the type of shock-level variation one would want for identification. Second, it demonstrates the conceptual feasibility of leveraging many exogenous shocks—in principle, one could imagine randomizing federal subsidies across many industries and measuring employment and wages across local labor markets.

2. Approximate the experimental ideal with shock- and unit-level controls The next step is to describe the actual shift-share design used for the empirical analysis as approximating the idealized experiment from Step 1. This involves specifying any controls which are needed to bridge the gap between the two. For example, observed changes in subsidies across industries, g_k, could provide shock-level variation analogous to the randomized subsidies, up to some controls.

There could be two types of such controls, depending on whether shock-level or unit-level confounders motivate including them. For the former, one may consider shock-level observables q_k that both correlate with the g_k and can have a direct impact on the outcome of interest. For example, one might worry that subsidies are systematically larger in certain sectors (i.e., aggregate groups of industries, such as food manufacturing) and that these sectors differ in terms of labor supply shocks associated with them. In this case, one would like to control for sector effects in the SSIV specification. But how can this be done if such q_k vary at the industry level while the specification is estimated at the regional level? BHJ show that the answer is to control for $\sum_k s_{ik}q_k$: shift-share aggregates of the industry-level confounders, with the same exposure shares as in the construction of the instrument. When the q_k are indicators for sectors and the s_{ik} capture local industry employment shares, this amounts to controlling for regional employment shares of each sector. The SSIV specification will then only leverage the variation in g_k which is uncorrelated with the q_k: e.g., within-sector variation in subsidies.

The second type of controls arise from unit-level observables which are thought to correlate both with the error term ε_i and with the shift-share instrument For example, one might expect labor markets in the U.S. Rust Belt experience different unobserved local labor supply shocks vs. other parts of the country, and that industries more concentrated in the Rust Belt see systematically different subsidies. In this case, a straightforward solution is to include an indicator of the Rust Belt to the SSIV specification.[8]

Identifying the relevant set of potential confounders is, of course, a context-specific issue. Ideally, a researcher would make an *ex ante* case for the chosen controls from relevant institutional details. They can then run different *ex post* falsification tests or sensitivity checks to build empirical support for these controls, as we describe below. Beyond the controls needed to ensure instrument exogeneity, a researcher can also include additional controls which are uncorrelated with the instrument. These additional controls may remove variation from the error term and thus reduce noise in estimation. Again, it is good practice to be explicit *ex ante* which controls are included to address potential confounders and which are there to improve precision.

3. Include the "incomplete share" control In shift-share designs where the exposure shares s_{ik} do not add up to one — what BHJ label the "incomplete shares" case — a special control must be included: the sum of shares, $S_i = \sum_k s_{ik}$. To build intuition, recall that with "complete" shares (when $S_i = 1$), the shift-share instrument is a weighted *average* of the shocks. As discussed in Section 2.3, if the shocks arise from a pure lottery, the shift-share instrument is also like a lottery outcome, even if the shares are endogenous. This logic breaks down with incomplete shares, however, when z_i is a weighted *sum* of the shocks. Then, even with randomly assigned shocks—which have, say, a positive mean—an observation with a higher S_i would systematically get higher values of the SSIV. The SSIV is thus correlated with the sum of shares, which can in turn be correlated with the error, leading to bias. Controlling for S_i removes the problem, as units with the same S_i get different values of the SSIV only for random reasons.

One may notice that the incomplete share control is a special case of a share-aggregated control: $\sum_k s_{ik} q_k$ corresponding to a shock-level intercept $q_k = 1$. Like we always include intercepts in regressions, one should always include the sum of shares in SSIV regressions based on exogenous shocks.

4. Lag shares to the beginning of the natural experiment When constructing the SSIV, one needs to decide when to measure the shares. The decomposition in Section 2.1 suggests measuring them at the beginning of the period of interest but it is common in practice to lag them further. Is this practice justified?

In the exogenous shocks approach, it is best to measure the shocks at the beginning of the natural experiment that generates them. This avoids the situation where the shocks affect the shares, potentially generating bias.[9] At the same time, shares matter for instrument power, and lagging shares beyond what is necessary would typically lead to a weaker first-stage.

[8] An alternative way would be to consider industry-level controls as described above: e.g., the share of Rust Belt in the industry employment. The relative merits of these two approaches remain unexplored formally.

[9] Not every response of the shares to past shocks makes the shift-share IV endogenous: if this response is not related to the error terms, there is no problem. Footnote 32 in BHJ gives an example where the bias does arise.

What constitutes the beginning of the natural experiment? If there were no shocks correlated with g_k in the past, it is just the beginning of the period when the g_k are measured. However, if the shocks unfold over several periods in a serially correlated way, it is appropriate to lag the shares further, to the first of these periods — or alternatively extract unpredictable shock innovations and use them to construct the shift-share IV (see Section 3.2 for further discussion).[10]

5. Report descriptive statistics for shocks in addition to observations Empirical papers normally present the number of observations and summary statistics for the main variables. In shift-share analyses that leverage the variation in shocks it is important to present such descriptive statistics for shocks as well—in the same way as one would in a non-shift-share setting at the shock level. While the mean and standard deviation of z_i is useful to know, so are the mean and standard deviation of g_k.

One caveat is that, as we show below, each shock has an importance weight proportional to the exposure share of that shock for an average observation, $s_k = \frac{1}{N}\sum_i s_{ik}$. For example, when studying subsidy shocks across industries, the importance weights could correspond to the average industry employment share across local labor markets. Thus, it is natural to report descriptive statistics with those weights as well. For instance, the weighted version of the number of shocks is the "effective number of shocks"—the inverse of the Herfindahl index of shock importance weights, $1/\sum_k s_k^2$. When the effective number of shocks is small, a few shocks may drive the empirical analysis, potentially making the results noisy and unreliable. This is not specific to SSIVs: a similar issue can arise when running a weighted OLS regression, if some observations get disproportionately large weights.[11]

Descriptive analyses for the shocks need not be limited to their effective number and the distribution. For instance, one could also describe the distribution of the shocks after residualizing them on shock-level controls the research plans to include. Or one could plot the shocks on the map if they have a geographic dimension.

6. Implement falsification tests for shocks in addition to the instrument Like with every research design, it is useful to perform balance tests: that the variation believed to be exogenous is indeed not correlated with proxies for confounders. The special feature of shift-share designs is that this can be done in two ways: not only for z_i at the level of units, but also directly for g_k at the level of shocks.

Checking balance of the instrument at the unit level, z_i, is more standard. For instance, a standard pre-trend test involves regressing the lagged outcome on z_i while including the controls picked in advance (such as the incomplete share control). The only particularity of shift-share designs in this case is that standard errors should be computed appropriately, as we discuss in detail in the next step.

[10] Another problem that arises with serially correlated shocks is that past shocks may have direct dynamic effects on the current outcomes (see Jaeger et al. (2017)). Simply lagging the shares does not help with this problem; we return to it in Section 3.2, too.

[11] If shocks are correlated within certain clusters, the Herfindahl index can be computed at the level of such shock clusters, since having many correlated shocks may also not be enough for a reliable statistical analysis.

But since the identifying variation is at the shock level, it is also useful to check balance of shocks directly, with respect to shock-level variables that may proxy for unobservables. For example, in our running example with a change in industry subsidies, one could check whether the shock g_k correlates with variables reflecting labor supply factors, such as the composition of the workforce and the share of immigrants in the industry. This test is useful to assess whether changes in subsidies are systematically different for certain industries that would likely have been on different employment trends even absent changes in labor demand.

7. Produce the main estimates with corrected standard errors The final step is to obtain estimates and standard errors for the SSIV regression and check their robustness.

Valid inference with exogenous-shock SSIV designs requires a special "exposure robust" approach. Intuitively, inference must take into account that units with similar shares mechanically have correlated z_i and may also have correlated ε_i due to their common exposure to unobserved shocks. For example, regions specializing in the same industries will be affected by the same (potentially unobserved) industry shocks. Adao et al. (2019) show with a Monte Carlo simulation that this issue can be very serious in practice: in the setting of Autor et al. (2013), they find that conventional standard errors lead to rejections of the true null of no effect in placebo samples with randomly generated g_k around 50% of the time (rather than the targeted 5%).

Two solutions to this issue have been developed, both leveraging as-if random assignment of the shocks. First, Adao et al. (2019) provide a variance estimator which is asymptotically valid regardless of the correlation structure of the errors across observations, as long as the exogenous shocks are mutually uncorrelated or clustered in a known way (e.g., by group of industries). Second, BHJ show that one can simply run a particular shock-level IV regression which always produces an identical coefficient as $\hat{\beta}$ from the shift-share regression (2) but gives valid standard errors, since it is estimated at the same level at which the shocks are assigned. In this regression, the k-level outcome and treatment are certain transformations of the original outcome and treatment, shocks g_k directly serve as a single instrument, shock-level controls q_k are directly included as controls, and estimation is weighted by average shares $s_k = \frac{1}{N}\sum_i s_{ik}$.[12] The `ssaggregate` packages in Stata and R automate the transformation of the outcome and treatment for this regression. The shock-level regression offers the flexibility to accommodate various types of shock dependence: e.g., not only one-way clustering but also spatial clustering and serial correlation. The equivalent regression can also be used to produce exposure-robust first-stage F-statistics to judge the instrument strength; see BHJ for details.

After producing the main SSIV estimates, it can be instructive—like in any empirical study—to check their robustness to a variety of choices. For example, one may examine the stability of the SSIV estimate under alternative sets of controls which could correspond to different assumptions of conditional quasi-random shock assignment. Similarly, one may check that estimation with and without unit-level importance weights (e.g., population weights in a regional analysis) yields similar

[12] Specifically, the transformation of the outcome and treatment involves first residualizing them on the included i-level controls and then, for each k, averaging across observations with weights s_{ik}.

results.

3.2 What if Shocks are Observed over Multiple Periods in a Panel?

While the previous section considered a single cross-section, in many applications researchers have access to shocks g_{kt} happening in multiple periods t. In a panel of units i over periods $t = 1, \dots, T$, one may consider an IV specification

$$y_{it} = \beta x_{it} + \gamma' w_{it} + \varepsilon_{it},$$

with a SSIV $z_{it} = \sum_k s_{ikt} g_{kt}$ and some controls w_{it}. Here we indexed the shares s_{ikt} by the period when they are used, not when they are measured: for instance, s_{ikt} can be time-invariant, "fixed" in the earliest period.

This setting offers new possibilities. First, if a natural experiment generates exogenous shocks in several periods, "stacking" them provides more estimation power. Second, panels with *many* periods enable a new type of SSIVs: those leveraging purely time-series variation in shocks. In the simplest case, there may be just one shock in each period and heterogeneous unit exposure to this shock, such that $z_{it} = s_{it} g_t$ (where s_{it} is often time-invariant). Estimator consistency relies on observing many periods. The shock-level equivalent regression in this case is just a time-series regression, regardless of the number of units in the panel, and standard errors should correspondingly be clustered in a time-series way (e.g., by period). Below we provide a detailed illustration of these points in the setting of Nunn and Qian (2014).

Panel data also pose new challenges. First, the shocks may have different means in different periods. In conventional panel models, time-varying means are addressed by including period fixed effects (FEs), γ_t. Correspondingly, in shift-share designs, time-varying shock means are addressed by a share-weighted aggregate of period FEs: $\sum_k s_{ikt} \gamma_t$. With complete shares, i.e. when $\sum_k s_{ikt} = 1$, this control coincides with the period FEs. But in the incomplete shares case, the sum of shares control needs to be interacted with period FEs. We illustrate this point below in the setting of Autor et al. (2013).

Second, shocks can be serially correlated, in which case each period cannot be viewed as a separate natural experiment. Then, as mentioned in Step 4 above, the static specification in (2) suffers from an omitted variables bias problem when there are dynamic causal effects, i.e. if lagged shocks affect current outcomes (Jaeger et al., 2017). Intuitively, the estimated coefficient for the treatment in specification (2) is biased because it also includes the dynamic causal effect of past treatments. Moreover, if the shares can respond to past shocks which are correlated with contemporaneous shocks, the shares cannot be viewed as measured before the natural experiment in shocks began.

There are two solutions to the problems of serial correlation in shocks. One involves estimating richer specifications which include the relevant lagged treatments, as well as lagging the shares underlying the SSIV further. The shares need to be measured at a date before the sequence of serially correlated shocks began if such a date exists. Jaeger et al. (2017), for instance, show that migration rates by country of origin are very serially correlated since 1970s, but not correlated with those from earlier decades. Thus, year 1970 can be viewed as the beginning of the natural experiment in their setting.

An alternative solution is based on isolating the unpredictable component of the contemporaneous shocks before constructing the SSIV. For instance, if the shocks follow a first-order autoregressive process, one can control for the lagged shocks (by controlling for a share-aggregated version of them at the unit level). If the SSIV leverages the idiosyncratic component of shocks, the issues stemming from serial correlation disappear.[13] This approach only yields the contemporaneous effect but does not require a correct specification of the dynamic effects.

3.3 Examples

We now discuss three examples, which illustrate some of the key practical insights for exogenous-shock SSIVs in the contexts of a randomized experiment in shocks, quasi-experimental shock variation, and time series variation.

SSIV in a randomized trial Franklin et al. (2023) provide an example of leveraging a randomized control trial in a shift-share design to estimate the indirect impacts of a treatment. They study a large public works program offering employment at high wages to local low-income workers residing in specific neighborhoods in Addis Ababa, Ethiopia. The authors are interested in the impact of this program on private sector wages: by increasing labor supply for public works, the program can reduce labor supply for other activities and increase private wages. Identification relies on the randomized rollout of the program.

While the program is randomized at the level of residential neighborhoods k, it may have spillovers on wages in other neighborhoods (labor markets i) because workers can commute. Using data on the baseline-period commuting network, Franklin et al. (2023) build a measure of each labor market's exposure to the randomized roll-out: for each labor market, the shift-share treatment takes a weighted average of treatment status in each place of residence (the shocks) using as exposure weights the share of workers who commute from that place of residence (which sum to one). In this setting, if the shocks are simply randomly assigned, there is no need to introduce controls.

In reality, some residential neighborhoods k were ineligible for randomization. Thus, the total share of commuters from eligible areas is less than one, and controlling for this total is necessary.[14] With this control, and assuming that commuting shares correctly capture the structure of spillovers, the shift-share design identifies the causal impact of the program.

[13] There are different ways of extracting the idiosyncratic component of shocks. Instead of controlling for lagged shocks, another natural approach could be to control for the time-invariant component of shocks. Implementing this strategy is easy when time-invariant shares are used: then including *unit* fixed effects in the control vector w_{it} implicitly removes any shock-level confounders α_k, since the corresponding share-aggregated control $\sum_k s_{ikt}\alpha_k$ is time-invariant.

[14] An additional minor complication in Franklin et al. (2023) is that the experiment was stratified, such that eligible neighborhoods in 10 parts of the city had slightly different probabilities of being treated. This can be accounted for by viewing the treatment probability as a shock-level confounder.

SSIV without an experiment Autor et al. (2013, henceforth ADH) analyze the impact of local import competition on local employment. Simplifying details, the outcome is the employment change in a U.S. commuting zone (CZ) and the endogenous variable is the change in local exposure to import competition from China (measured as the growth of imports from China in dollars per U.S. worker). OLS may be biased, for example, if high productivity growth in China happens in industries (e.g., textiles) with systematically different productivity or demand trends in the U.S. or if U.S. consumers substitute to Chinese goods in industries where U.S. productivity is lagging.

In this setting, the idealized experiment would be to assign productivity shocks at random across manufacturing industries in China. These shocks would have different incidence across U.S. commuting zones given the pre-determined industrial composition of each area. That would motivate a shift-share IV with exposure governed by the shares s_{ik}, defined as the share of of industry k in total employment in local i in a base period, and the randomized productivity shocks in China. In the absence of such an experiment, ADH proxy for the productivity shocks by the observed growth of imports from China in industry k in eight high-income countries excluding the United States, g_k. Measuring imports those countries implies that demand and supply shocks that are idiosyncratic to the U.S. cannot bias the results.

An important feature of this setting is that the exposure shares do not sum to one, since only manufacturing industries are exposed to trade with China. Locations with a larger total share of employment in manufacturing are likely on different potential outcome trends, e.g. because of the secular decline in manufacturing (which can have many causes other than trade). To address this issue, it is necessary to control for the sum of exposure shares in each location.[15]

A further adjustment is called for because ADH conduct the analysis in a repeated cross-section over two ten-year periods, and the average shocks are different in the two periods. As discussed above, leveraging shock variation across industries within manufacturing and only within periods requires controlling for the interaction of the sum of exposure shares with period fixed effects. This control prevents the bias that would arise if the manufacturing sector as a whole (and regions specializing in manufacturing industries) declined differentially in the two periods for reasons unrelated to trade. BHJ show that this control is empirically relevant in the ADH setting.[16]

To assess the plausibility of the design, it is instructive to conduct shock-level and unit-level falsification tests. The shock level tests are meant to assess potential threats to identification at the shock level: it could be that China specializes in certain industries (e.g., low-skill industries) that could have been on different employment trends in the U.S. event absent trade shocks. To speak to this concern, one can correlate the shock g_k with industry-level variables reflecting the structure of employment and technologies, such as the employment shares of different types of

[15]Note that the appropriate control equals the total regional share of manufacturing employment in the period in which the *shares* are measured. Since ADH lag the shares by a decade relative to the period of the outcome and treatment, the incomplete share control should be lagged as well.

[16]If that bias did not actually arise, including the interacted control would not affect the coefficient. It can lead to some efficiency loss but the same is true when controlling for period fixed effects in standard (i.e., non-SSIV) OLS or IV regressions with multiple periods. Including period FEs and the sum of shares separately, without an interaction, would not address the bias.

workers (e.g., production workers), the labor share, the average wage at the firm, and investment in new technologies (e.g., computers) in a previous period. Using the data from ADH, BHJ find that the shocks are balanced across these dimensions.

Correlating the regional SSIV with potential CZ-level confounders is also instructive. Such a correlation would arise if China specializes in industries that are located in CZs with unusual observed characteristics, which can raise concerns they are on different potential employment trends, too. One can regress CZ-level outcomes—such as the lagged fraction of a CZ's population who is college-educated, foreign-born, female, working in routine occupations, and so on—on ADH's SSIV (controlling for the sum of shares interacted with period fixed effects). One can also implement a standard "pre-trend" test, with lagged CZ-level outcome on the left-hand side. ADH and BHJ find that most of these tests pass.

SSIV with time-series variation Nunn and Qian (2014) study the impact of U.S. food aid on conflicts in a long panel of recipient countries. Simple OLS, or even fixed-effect, estimates are subject to several potential biases: the presence of conflict may increase the demand for food aid; there might be many omitted variables—such as political and economic crises—affecting both conflict and food aid; or donors may decide to reduce food aid to countries engaged in conflict.

To resolve these issues, the authors leverage exogenous time variation in U.S. wheat production over time. Due to price stabilization policies requiring the U..S government to buy wheat from U.S. farmers at a set price, the U.S. government accumulates excess reserves in high production years, which is shipped to developing countries as food aid. The shift-share design leverages these time series shocks, using as exposure weights a country's likelihood of being a U.S. food aid recipient. Specifically, the quantity of wheat aid shipped from the U.S. to recipient i in year t is instrumented by $z_{it} = s_i g_t$, where g_t is the amount of U.S. wheat production in the previous year and s_i is the fraction of years that recipient country i receives a positive amount of U.S. food aid during the sample period, 1971–2006.

How can one follow our checklist in this context? For steps 1–2, the researcher would clarify whether all time series variation in wheat production is considered as-good-as-random. If, for instance, U.S. wheat production is correlated with key economic indicators such as oil prices, which can have a direct effect on conflict, the researcher would control for oil prices interacted with s_i. Indeed, this is one of the controls Nunn and Qian (2014) include. They also include other controls, such as dummies for six geographic regions of the world interacted with year dummies . For step 3, the "incomplete share" control here is simply s_i, the time-invariant exposure to U.S. aid, since each observation is exposed to only one shock; in the Nunn and Qian's regression it is absorbed by country fixed effects. For step 4, one would measure s_i before, rather than during, the sample period. For step 5, it would be useful to plot the time series of wheat prices, which serves as identifying variation. If wheat prices, for instance, exhibit strong serial correlation or a secular trend, it may be appropriate to analyze dynamic causal effects or extract an unpredictable component of the time series of U.S. wheat production. For step 6, one can check whether the time series of wheat production is correlated with potential confounders, such as the aforementioned oil prices. And at the country-by-year level, an IV regression with lagged conflict as the outcome would constitute a standard pre-trend test. Finally, for step 7, one can cluster standard errors at the level of identifying variation, i.e. by year (rather than by country, which is more conventional in panel regressions). Heteroskedasticity and autocorrelation-consistent (HAC) standard errors may be appropriate if the shocks (and errors) are serially correlated. Such standard errors can be conveniently obtained via the equivalent—purely time-series—IV regression.

4 SSIV with Exogenous Shares

We now provide a list of practical steps to determine whether and how to use the exogenous-share approach to SSIVs. We develop this checklist with the immigration setting from above, where s_{ik} represents the lagged share of immigrants from country k in region i. We discuss several applied examples at the end of this section.

4.1 A Checklist for Applied Work

1. Determine whether the exposure shares are potentially suitable instruments As before, the researcher can start by describing the main sources of treatment endogeneity and the reasons why the shares may be useful instruments to address the corresponding threats. While we illustrate how this can done with detailed empirical examples below, here we highlight two general guiding principles.

First, the instrument exogeneity argument can be strengthened by looking for shares which are "tailored" to the treatment. Recall from Section 2.4 that the shares cannot be exogenous instruments if they capture the exposure of the outcome to some unobserved shocks. This rules out cases where the shares are "generic," in the sense of capturing exposure to many shocks, while the treatment only captures one such mechanism (as with import competition in ADH). Conversely, it is conceivable (albeit not guaranteed) that the share of migrants from a certain origin only captures the region's exposure to migration shocks, making such shares potentially exogenous in our running migration example.

Second, the identification strategy can be strengthened by exploiting a source of variation in the initial shares that is more likely to satisfy exogeneity. For example, Terry et al. (2023) worry that the initial composition of migrants may be correlated with labor demand factors; they address this issue by replacing the shares with their component arising from a specific historical quasi-experiment. Simply lagging the shares can also sometimes help, although it does not by itself guarantee exogeneity.[17]

[17] Using lagged shares will typically weaken the first stage; it is therefore important to explain why it is plausible that lagging the shares reduces their covariance with the error term by more than it reduces the first stage, such that OLS bias is at least reduced by the SSIV. For example, in studies of the effect of immigration on local wages, lagging the shares to an earlier decade is helpful if two statements hold at the same time. First, labor demand factors that affected lagged migrant shares should have little correlation with the contemporaneous labor demand factors in the error term. Second, the relationship between lagged migrant shares and current immigration flows should be strong (e.g., because new migrants always tend to go to places with more previous migrants from the same origin).

We also note that the exogenous shares approach is appropriate if the researcher is in principle willing to use any of the shares, or any combination of them, as the instrument (provided the first stage is strong enough). For example, one may be willing to use initial immigrant shares from specific countries where a large emigration shock happened (e.g., Mexico or Cuba) as instruments directly. We discuss further below how such analyses can serve as a helpful robustness check.

2. Choose the necessary unit-level controls Even if the shares are "tailored," their exogeneity is a nontrivial assumption, like any parallel trends assumption. As usual, exogeneity can be relaxed by including control variables. For example, the researcher can control for certain sums of shares to leverage only conditional cross-sectional variation. In the migration setting, controlling for the total immigrant share would mean that the SSIV would leverage variation in the composition of migrants across locations, avoiding comparisons between regions with high and low migration intensity overall.

As in any empirical study, other controls that are uncorrelated with the shares may also be included to improve estimation efficiency.

3. Characterize which shares matter the most for the estimates When viewing the SSIV estimate as a pooled version of K one-at-a-time share-IV estimators, it can be important to understand whether a small subset of these instruments drive the results. If this is the case, the researcher can use those shares to explain how the identification strategy works and can focus on them when doing falsification tests that we describe below.

GPSS show how to measure the importance weight of each share instrument, which they refer to as "Rotemberg weights," referencing Rotemberg (1983). They are based on a decomposition of the SSIV estimator into a weighted sum of individual-share IV estimators with weights that add up to one, although some can be negative. The Rotemberg weights can thus be interpreted as measuring the sensitivity of the SSIV estimate to violations of exogeneity by each share instrument.[18] These weights are proportional to the shock g_k as well as the covariance between the share s_{ik} and the treatment x_i, after partialling out the included controls (see Proposition 3 in GPSS). The `bartik_weight` command in Stata and R provided by GPSS computes these weights.[19]

4. Check sensitivity to how share instruments are pooled When individual shares are exogenous instruments, there are infinitely many valid ways of pooling them. SSIV with the shocks originally chosen by the researcher provides one particular set of weights; some other options include IV estimators using one of the shares as the single instrument, or overidentified two-stage least

[18] We note that Rotemberg weights are only useful in the exogenous-share view on SSIV: if, as in the exogenous-shock view, the shares need not be exogenous in the first place, it does not matter how sensitive the SSIV estimator is to violations of share exogeneity. BHJ show, however, that very skewed Rotemberg weights may indicate that exposure-robust confidence intervals may be unreliable.

[19] Unfortunately, the Rotemberg weights are not unique when the shares add up to one. This is because the shares—and thus individual-share IV estimators—are perfectly multicollinear. Another way to see the issue is that adding a constant to all shocks does not change the SSIV estimator; yet, the Rotemberg weights change after this innocuous transformation, as they are proportional to the level of g_k. In this case, GPSS recommend choosing the Rotemberg weights that correspond to the demeaned shocks.

squares (2SLS) using all shares as instruments. When all shares are valid instruments and the treatment effects are homogeneous in the population, all of these methods should give the same estimate up to sampling noise. It can therefore be instructive to run the standard overidentification test for similarity of these K IVs (Wooldridge Ch. 6.2.2). A failure of this test could indicate that some of the shares are not exogenous, or that treatment effects are heterogeneous. In any case, it is instructive to examine how sensitive the headline estimate is to alternative ways of combining the share instruments. We discuss several graphical and statistical procedures that share this goal.

One graphical approach is a "visual IV" procedureAngrist and Pischke (2008, p.103), which involves plotting K reduced-form coefficients (from regressions of the outcome on one share one at a time and the included control variables) against the corresponding first-stage coefficients (from similar regressions of the treatment). When the coefficients concentrate around a ray from the origin, as would happen in large samples with valid share IVs and homogeneous effects, the choice of how the IVs are pooled will be unimportant. GPSS propose a different graphical procedure: a scatter plot of the K IV coefficients, which again use one share IV at a time, against their respective F-statistics. Here one hopes to see that all IV estimates are similar, especially those with a high F-statistic and large Rotemberg weight. With either approach, the signs and magnitudes of Rotemberg weights can be indicated with the shapes and sizes of the dots, respectively.

The researcher can also report how the headline estimate changes when pooling the share instruments in alternative ways. One approach is to examine whether the IV estimate changes when using only a few shares—e.g., those with the largest Rotemberg weights. Another approach is to keep all shares for higher precision but pool them in a different way. When the number of shares is small relative to the sample size 2SLS is the natural estimator to report; an efficient generalized method of moments (GMM) estimator is another option. With many shares, 2SLS suffers from bias but several estimators robust to "many weak instruments" are available instead: jackknife IV (JIVE; Angrist et al. (1999)), limited information maximum likelihood (LIML), heteroskedasticity-robust Fuller estimator (HFUL; Hausman et al. (2012)), and modified bias-corrected 2SLS (MB2SLS; Kolesar et al. (2015)).[20]

5. Implement falsification tests for individual shares in addition to the SSIV Like in any design, it is worth checking balance of the instrument on observable variables that may be expected to correlate with the error term. Different variables can serve for useful balance tests: pre-period changes in the outcome variable (corresponding to a pre-trends test), unit characteristics measured at the beginning of the period, or contemporaneous changes in placebo outcomes that are not expected to be causally affected by the treatment.

The special feature of the exogenous-share view on SSIV is that falsification tests are also useful to perform on individual shares, since each of them is assumed to be exogenous. To avoid the problems associated with testing many hypotheses, it is natural to focus on the subset of shares that are most important for the resulting estimate, as measured by the Rotemberg weights.

[20]The SSIV estimator also requires a similar bias correction when the shocks g_k are estimated from the sample, as in Bartik (1991). In Appendix A we discuss the leave-out SSIV estimator that helps in this scenario.

4.2 Examples

We now use two examples to illustrate the exogenous-share approach to SSIVs in the contexts of the labor market responses to migration and retirement rates.

Labor market effects of immigration We first consider the design of Card (2009, Table 6) and its re-analysis by GPSS. The goal is to estimate the (inverse) elasticity of substitution between immigrant workers and native workers in labor demand, i.e. the relationship between the log wage gap between immigrant and native workers and the ratio of immigrant to native hours worked. Simplifying details, the analysis considers a cross section of outcomes (in levels) in 2000 in 124 cities.

As with any demand equation, OLS estimates may be biased: a positive labor demand shock for migrants would draw more immigrants into a location and at the same time increase their wages relative to natives. An instrument is needed that captures the relative supply of migrant and native workers.

Card (2009) proposes to instrument the local ratio of immigrant to native hours with a SSIV, leveraging immigration patterns from 38 countries indexed by k. Here s_{ik} is the share of immigration group k in the population of city i in 1980; note that these shares add up to the initial migration share, rather than to one (and the initial migration rate is not controlled for). The shock g_k is the number of migrants in group k moving to the U.S. from 1990 to 2000, normalized by the national stock of migrants from k already in the U.S. in 1990.

This SSIV strategy alleviates some endogeneity concerns, as the shares are uncorrelated with some relative labor demand factors. Specifically, idiosyncratic regional labor demand shocks (which attract migrants to a particular location in the current period only) would be a problem for OLS but not for Card's SSIV, since the migrant shares are measured before these shocks are realized. In contrast, persistent regional labor demand factors (e.g., characteristics that always make immigrants more productive relative to native workers, such as the prevalence of certain languages like Spanish) would remain a problem for both OLS and the SSIV, since these factors impact the beginning-of-period migrant share while also entering contemporaneous labor demand in the error term.

Some of the potential limitations of Card's instrument can be addressed by simple adjustments to the empirical strategy. In particular, estimating the outcome equation in differences would alleviate concerns about time-invariant regional confounders. Moreover, controlling for the total initial share of migrants or renormalizing the shares relative to the initial stock of migrants would make the SSIV leverage the *composition* of migration origins. This would address labor demand shocks that affect all migrants equally.

Working with the original Card (2009) instrument, GPSS compute the Rotemberg weights to show which shocks matter most for the estimates. They show that Mexico receives half of the weight in the sample of high school equivalent workers. Thus, for these workers one can largely think of the research design as using the initial Mexican immigrant share as an instrument. This is consistent with the observation of Card (2009) that the SSIV is highly correlated with the initial fraction of

Mexican migrants. For college equivalent workers, GPSS document that the top country is the Philippines, receiving 15% of the total weight.

GPSS also report the results obtained with alternative estimators in the Card (2009) setting. They find that the point estimates remain very similar with the baseline shift-share instrument, 2SLS, LIML, and MB2SLS. Similarly, plotting the IV estimates using individual shares as instruments against their respective F-statistics, they find little variation in the estimates—especially for the strong instruments. These tests demonstrate the robustness of the baseline estimate to alternative ways of pooling the share instruments and suggest that treatment effect heterogeneity is limited in this setting.

The setting also lends itself to balance tests for individual share instruments. GPSS report that the 1980 Mexican immigrant share does not predict relative wages in 1980 or 1990, but does in 2000 (the year of analysis). While the patterns for Mexico are comforting, the 1980 share of immigrants from the Philippines correlates with the native-immigrant wage gap in all three periods. Other countries also feature statistically significant violations of balance, raising concerns about share exogeneity.

Labor market effects of retirement Mohnen (2024) studies the impact of the retirement of older generations on labor market outcomes for younger generations in the U.S., conducting the analysis at the CZ level. The author develops a shift-share strategy to address the key identification issue: local labor demand shifts may determine both retirement decisions for older workers and outcomes for younger workers.

The idea of the SSIV strategy is to leverage cross-CZ variation in age composition among older workers, which determine CZs' differential exposure to common shocks: national retirement rates by age. Specifically, s_{ik} is the local share of age group k in CZ i among the population aged 45 to 80 (such that shares sum to one in each CZ), and g_k is the national 10-year retirement rate by age group. CZs with a lower share of older workers will tend to experience fewer retirements, giving the instrument power. The author then regresses 10-year differences in labor market outcomes for the young (employment rate, share working in high-tech industries, etc.) on retirement rates over 10 years, using the shift-share instrument. The specification includes various start-of-period CZ controls (employment share of manufacturing and routine occupations, unemployment rate, etc.). The identification assumption is that, conditional on these characteristics, the age shares (among people above 45) are all valid instruments.

To better understand the source of variation, the author describes which age shares matter the most in driving the estimates. He reports Rotemberg weights, documenting that they are close to proportional to the 10-year national retirement rates by age group. To assess whether the results might depend on how the share instruments are pooled, the author reports alternative estimators: using a single share as an instrument (the initial share of the population age 52-59 as a fraction of the population above 45), or using all detailed age group shares as separate instruments with GMM. All three versions of the research design paint a similar picture: there is no significant impact of retirement on youth employment, but higher retirement rates lead to a rise in the share of young workers in high-skill jobs, as well as higher wages and higher job mobility for the youth. Finally, the author finds that the instruments pass an overidentification test, lending further support to the robustness of the design.

5 Conclusion

We have laid out two frameworks for shift-share research designs, which include sufficient conditions for IV validity, narratives for interpreting these conditions intuitively, falsification tests for the assumptions, and various practical recommendations.

To recap, with the many exogenous shocks approach, the researcher argues that the shocks arise as-if from a lottery, perhaps controlling for some shock-level and unit-level covariates. This approach is applicable if the researcher would be willing to use the shocks g_k directly as an instrument in a regression across k. It cannot apply when the number of shocks is small (or when a small number of shocks dominates the SSIV) or when the shocks are mechanically linked to the error terms (like with national industry growth rates in Bartik (1991)). Practically, balance tests for the shocks are helpful to falsify the assumptions, in addition to balance tests for the shift-share IV. To leverage shock variation correctly at the level of observations i, one should control for share-weighted aggregates of shock-level covariates, including the sum of shares. Confidence intervals should account for the exposure clustering of observations with similar shares.

With the exogenous shares approach, the researcher argues that regions with different exposure shares have parallel trends in the outcomes, perhaps controlling for some covariates. This approach is applicable if the researcher would be willing to use any of the individual shares as treatment intensity in a difference-in-differences analysis (abstracting from the possibly low power of such analyses). This approach cannot apply when the treatment is "specific" while the shares are "generic" (e.g., with an import competition treatment and local employment shares of different industries, as in Autor et al. (2013)). Practically, pre-trends and other balance tests on individual shares, especially those with high Rotemberg weights, are helpful to falsify the assumptions. Statistical and visual overidentification tests and alternative estimators that use the set of shares directly as instruments are useful, since SSIV is only one of many ways to combine variation in the shares.

How can one pick between the two approaches? As described above, in some settings one approach is a "non-starter": e.g., the exogenous shocks approach when K is too small or the exogenous shares approach when the treatment is specific while the shares are generic. In other settings, it may be productive to think through the potential bias and efficiency properties of the instruments each approach would suggest. For instance, when estimating the local demand elasticity for migrant labor, can a plausibly exogenous supply shock ("push factor") with a strong first stage be obtained? Or is it plausible that there are no national demand shocks for migrants of any origins—in which case an instrument in the spirit of Card (2009), with a likely stronger first stage, may be convincing enough? We hope our review will help researchers assess such tradeoffs.

References

Acemoglu, Daron and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

Adao, Rodrigo, Michal Kolesar, and Eduardo Morales, "Shift-Share Designs: Theory and Inference," *Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Aghion, Philippe, Celine Antonin, Simon Bunel, and Xavier Jaravel, "Modern Manufacturing Capital, Labor Demand, and Product Market Dynamics: Evidence from France," *Working Paper*, 2022.

Angrist, J D, Guido W. Imbens, and AB Krueger, "Jackknife instrumental variables estimation," *Journal of Applied Econometrics*, 1999, *67* (95), 57–67.

Angrist, Joshua D. and JS Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press, 2008.

Autor, David H. and Mark G. Duggan, "The rise in the disability rolls and the decline in unemployment," *Quarterly Journal of Economics*, 2003, *118* (1), 157–205.

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

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

Berman, Nicolas, Antoine Berthou, and Jerome Hericourt, "Export Dynamics and Sales at Home," *Journal of International Economics*, 2015, *96* (2), 298–310. Publisher: Elsevier B.V.

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

Bombardini, Matilde and Bingjing Li, "Trade , Pollution and Mortality in China," *Journal of International Economics*, 2020.

Borusyak, Kirill and Matan Kolerman-Shemer, "Regression Discontinuity Aggregation, with an Application to the Union Effects on Inequality," *Mimeo*, 2024.

_ and Peter Hull, "Non-Random Exposure to Exogenous Shocks: Theory and Applications," *Mimeo*, 2021.

_ and _ , "Non-Random Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

_ and _ , "Negative weights are no concern in design-based specifications," *Working Paper*, 2024.

—, —, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

—, —, and —, "Design-Based Identification with Formula Instruments: A Review," *Working Paper*, 2023.

—, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficienct Estimation," *Working Paper*, 2023.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet, "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, April 2015, *7* (2), 81–108.

Card, David, "The Impact of the Mariel Boatlift on the Miami Labor Market," *ILR Review*, 1990, *43* (2), 245–257.

—, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64. ISBN: 5927.

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

Clots-Figueras, Irma, "Women in politics," *Journal of Public Economics*, August 2011, *95* (7-8), 664–690.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum, "The rise of the east and the far east: German labor markets and trade integration," *Journal of the European Economic Association*, 2014, *12* (6), 1643–1675. ISBN: 9783863041267.

de Chaisemartin, Clement and Xavier D'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–2996. arXiv: 1803.08807v1.

— and Ziteng Lei, "More Robust Estimators for Instrumental-Variable Panel Designs, With An Application to the Effect of Imports from China on US Employment," *Working Paper*, 2023.

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

Franklin, Simon, Clement Imbert, Girum Abebe, and Carolina Mejia-Mantilla, "Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia," *Working Paper*, 2023.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen, "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and "Normal" Economic Times," *American Economic Journal: Economic Policy*, 2020, *12* (1), 200–225.

Hausman, Jerry A., Whitney K. Newey, Tiemen Woutersen, John C. Chao, and Norman R. Swanson, "Instrumental variable estimation with heteroskedasticity and many instruments: Instrumental variable estimation," *Quantitative Economics*, July 2012, *3* (2), 211–255.

Hummels, David, Rasmus Jorgensen, Jakob Munch, and Chong Xiang, "The Wage Effects of Offshoring: Evidence From Danish Matched Worker-Firm Data," *American Economic Review*, 2014, *104* (6), 1597–1629. arXiv: 1011.1669v3 ISBN: 9788578110796.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475. ISBN: 00129682.

Imbert, Clement, Marlon Seror, Yifan Zhang, and Yanos Zylberberg, "Migrants and Firms: Evidence from China," *American Economic Review*, 2022.

Jaeger, David A, Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," *Working Paper*, 2017.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidences from the US Retail Sector," *Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

Kolesar, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens, "Identification and Inference With Many Invalid Instruments," *Journal of Business and Economic Statistics*, 2015, *33* (4), 474–484. ISBN: 0735-0015.

Miguel, Edward and Michael Kremer, "Worms: Identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, 2004, *72* (1), 159–217.

Mohnen, Paul, "The Impact of the Retirement Slowdown on the U.S. Youth Labor Market," *Journal of Labor Economics*, 2024.

Nakamura, Emi and Jon Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 2014, *104* (3), 753–792.

Nunn, Nathan and Nancy Qian, "US Food Aid and Civil Conflict," *American Economic Review*, 2014, *104* (6), 1630–1666.

Oberfield, Ezra and Devesh Raval, "Micro Data and Macro Technology," *Econometrica*, 2021, *89* (2), 703–732.

Rotemberg, Julio, "Instrument variable estimation of misspecified models," *MIT Sloan Working Paper 1508-83*, 1983.

Saiz, Albert, "The geographic determinants of housing supply," *Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Terry, Stephen J, Thomas Chaney, and Konrad B Burchardi, "Immigration, Innovation, and Growth," *Working Paper*, 2023.

Xu, Chenzi, "Reshaping Global Trade: The Immediate and Long-Run Effects of Bank Failures," *Quarterly Journal of Economics*, 2022, *137* (4), 2107–2161.

A Other Practical Concerns with SSIVs

In this appendix we discuss some questions which come up frequently in SSIV designs. These eleven questions are: What kind of local average treatment effect does SSIV estimate? Do the "shares" have to really be shares, between zero and one? Should the shares be normalized to add up to one? What if the shares add up to one across observations, rather than shocks? Can the shocks be unit-specific? Can one take a shift-share average of shift-share IVs? What if a log, or another transformation, of a SSIV variable is used? Can one use multiple shift-share instruments? What about interaction terms in shift-share regressions? Should the instruments in Card (2009) and Bartik (1991), which measure the shocks as the national growth of some equilibrium outcome (industry employment or total migration by origin), be viewed through the lens of exogenous shocks or exogenous shares? And what is the role of leave-one-out construction of shocks?

1. What kind of LATE does SSIV estimate? BHJ show that exogenous-shock SSIVs identify a convex weighted average of heterogeneous treatment effects under natural extensions of the independence, exclusion, and monotonicity assumptions that Imbens and Angrist (1994) use to establish identification of local average treatment effects (LATEs). Specifically, BHJ assume the shocks are drawn randomly as in a lottery, only affect the outcome through the treatment, and only affect the treatment of each unit in one direction. The SSIV estimand then weights together the effects of units with such treatment responses ("compliers"), with more weight given to units that are more responsive. Borusyak and Hull (2024) extend this result to allow for conditional random assignment, showing that in this case the weights are also proportional to the conditional variance of the instrument.

The heterogeneous-effect interpretation of SSIVs under other identifying assumptions is less established. de Chaisemartin and Lei (2023) raise concerns of non-convex weighting when SSIVs are justified by parallel trend assumptions, with respect to either shares or shocks, adding to a large literature noting similar issues for popular two-way fixed effect specifications (e.g. de Chaisemartin and D'Haultfoeuille (2020) and Borusyak et al. (2023b)). Part of this issue is apparent in the GPSS Rotemberg weight decomposition since, as GPSS note, some weights may be negative.

2. Do the "shares" have to really be shares? No, they can be any exposure weights.

In most applications s_{ik} is non-negative and typically they are some initial shares; notably this is the case when the SSIV follows from the decomposition in Section 2.1. But econometric results go through when s_{ik} are any weights that measure the exposure of observation i's treatment to the shock g_k.

As an example, consider the Miguel and Kremer (2004) study of spillover effects of deworming. In their OLS specification, the key explanatory variable z_i is the number of student i's neighbors who have received a randomized deworming treatment. Upon inspection, one may notice that this is a shift-share variable: $z_i = \sum_{k=1}^{N} s_{ik} g_k$ where k indexes all students, s_{ik} is a dummy that equals one if students i and k are neighbors, and g_k is a dummy that student k has been selected for deworming. Here the exposure weights are not shares of anything: they take values of zero and one and their total is the number of neighbors student i has. There is no problem with this, as long as the sum of shares (i.e., the number of student's neighbors) is controlled for.

3. Should I normalize the shares to add up to one? No.

From our earlier discussion of how the incomplete shares case requires extra care (specifically, picking appropriate controls in the many exogenous shocks approach), one might conclude that this case is something to be avoided. This can be done by constructing the instrument using shares normalized to add up to one. For instance, while ADH define s_{ik} as employment shares of manufacturing industry k relative to total employment in labor market i, one could consider redefining the shares to have local manufacturing employment in the denominator.

Such a conclusion would be misguided, however. First consider IV regressions, where the treatment x_i is given by the economic question. Then the researcher needs to choose the best SSIV z_i, and in particular the shares, to maximize the first-stage power. Whether identification leverages exogenous shocks or exogenous shares, power is maximized when the shares reflect the relationship between the treatment and the shocks, e.g. following the treatment decomposition as in Section 2.1. For example, in the ADH setting, using the local manufacturing employment in the denominator would reduce power because the shift-share instrument would exhibit large variation even in areas where manufacturing is a low share of total employment and the treatment (import competition) is close to zero. Including appropriate controls is a better way to avoid OVB while retaining statistical power, compared to modifying the shares.

Second, consider OLS shift-share analyses, such as spillover regressions, where the researcher is deciding on the right-hand side variable $x_i = z_i$. This choice is about specifying the most plausible functional form for how the shocks affect the outcome, such that the coefficient is economically meaningful. Again, this is achieved by setting the shares to reflect the exposure of observations to the exogenous shocks. For example, the fraction of treated friends, as in Cai et al. (2015), is a SSIV with the shares adding up to one, while the number of treated friends, as in Miguel and Kremer (2004), is an incomplete shares example. Still, if the researcher believes that the outcome is determined by the *number* of treated friends, they should use that specification, and include appropriate controls to avoid OVB.

4. What if the shares add up to one across observations, rather than shocks? If that is the case, both the shocks and the shares need to be rewritten without necessarily changing the resulting shift-share IV. The practical steps in both exogenous shocks and exogenous shares approaches rely on the particular way of defining the shares and shocks in (1).

Sometimes SSIVs are written with shares that add up to one (or some other meaningful number) across observations. For instance, slightly simplifying, ADH write their instrument as

$$z_i = \sum_k \frac{Emp_{ik}}{Emp_k} \cdot \frac{\Delta Imp_k}{Emp_i}. \tag{4}$$

The logic is that the growth of national industry imports (in dollars), ΔImp_k, is apportioned to regions proportionally to initial employment shares in this industry across regions, Emp_{ik}/Emp_k. It thus may appear that Emp_{ik}/Emp_k are the shares, and either ΔImp_k or $\Delta Imp_k/Emp_i$ are the shocks. Card (2001, 2009) similarly introduces the enclave instrument for migration, by allocating migrants from a given origin to regions proportionally to where earlier migrants settled. While this intuition may be useful when coming up with the IV, it is not helpful for formal analyses.

It is straightforward to rewrite (4) in the standard way, by flipping the two denominators:

$$z_i = \sum_k \frac{Emp_{ik}}{Emp_i} \cdot \frac{\Delta Imp_k}{Emp_k}. \tag{5}$$

Here, unlike (4) the shares Emp_{ik}/Emp_i add up across k to an economically meaningful number — the total manufacturing share of the region — and the shocks are purely industry-level, $\Delta Imp_k/Emp_k$.

What issues would one face using (4) instead of (5)? The exogenous shocks approach cannot apply, as $\Delta Imp_k/Emp_i$ are not industry-level variables at all. Moreover, viewing imports in dollars, ΔImp_k, as the shocks would make as-good-as-random assignment of the shocks untenable, as larger industries of course get more imports on average. Measuring shocks in relative terms makes their as-if random assignment a more plausible assumption. In the exogenous shares approach, using Emp_{ik}/Emp_k as instruments is the same as using initial employment levels Emp_{ik}, since the share denominators in (4) do not vary across observations. Thus, variation in the local industry size is used instead of the local composition of industries that is usually intended in SSIV designs. Moreover, since $\Delta Imp_k/Emp_i$ varies across i, z_i cannot be viewed as pooling variation in Emp_{ik}/Emp_k. More practically, applying the checklists from Sections 3 and 4 to the wrong shares and shocks would lead to incorrect controls (e.g., incomplete share controls) and diagnostic tests (e.g., based on wrong Rotemberg weights).

5. Can the shocks be unit-specific? Yes.

While we introduced shift-share variables as combining heterogeneous shares with a common set of shocks, the econometric framework also nests settings where each unit is exposed to a distinct set of shocks. One can define k to index the shocks to all observations and redefine the shares such that the exposure of a unit to another unit's shock is zero.

A set of examples is considered by Borusyak and Kolerman-Shemer (2024) who study "regression discontinuity aggregation" designs in which a shift-share treatment aggregates policy discontinuities defined at smaller geographic units. For instance, Clots-Figueras (2011) estimates the effect of the fraction of women in state legislatures in India, using the fraction of women who won against a man in a close election as the IV. Although each state has a distinct set of constituencies, this instrument is a shift-share where each state has non-zero exposure only to its own constituencies' shocks.

6. Can I take a shift-share average of shift-share IVs? Yes, and the result is also a SSIV, with the same shocks but more complicated shares.

For instance, slightly simplifying the setting of Imbert et al. (2022), they first construct an agricultural income shock in rural locations o ("origins"), $z_o = \sum_k s_{ok} g_k$, from exogenous crop price shocks g_k weighted by the local composition of crops s_{ok}. Then they use it to construct an in-migration shock in urban locations ("destinations") as $z'_d = \sum_o s'_{do} z_o$ using the initial period's shares of migrants to d that come from o, s'_{od}. One can see that this instrument can be rewritten as $z'_d = \sum_k s''_{dk} g_k$ with compound shares $s''_{dk} = \sum_o s'_{do} s_{ok}$.

Representing the SSIV with the resulting shares and shocks, in one step, would then yield appropriate incomplete share and other share-aggregated controls, and correct standard errors.

7. What if I take logs of a SSIV? A log — or any other nonlinear transformation — of a SSIV is no longer a SSIV. This may or may not complicate IV exogeneity.

In the exogenous shares approach, which views the SSIV as a particular function of the shares (where the shocks serve as weights), a nonlinear function of a SSIV is just another function of the same shares. If all individual shares are exogenous instruments, i.e. $\mathbb{E}\left[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}\right] = 0$, then any function of them is exogenous, too.

On the contrary, shock exogeneity does not imply exogeneity of nonlinear transformations of the SSIV, such as taking the log; such transformations can lead to a new type of bias. To see this, imagine the shares add up to one and the exogenous shocks are assigned in a lottery with positive values. Then, regardless of how the shares are correlated with the error term, the share-weighted average of the lottery shocks $z_i = \sum_k s_{ik} g_k$ is not correlated with the error. That logic fails for $\log z_i$: because of Jensen's inequality, units with dispersed shares will on average have a higher $\log z_i$ than units with concentrated shares, potentially leading to OVB.[21]

There are two ways to avoid this bias. First, Borusyak and Hull (2023) propose a "recentering" adjustment to the nonlinear instrument, such as $\log z_i$, based on rerandomizing the shocks, e.g. by permuting them. Second, putting the log inside the sum, i.e. replacing $\log \sum_k s_{ik} g_k$ with $\sum_k s_{ik} \log g_k$, yields an actual SSIV with shares s_{ik} and shocks $\log g_k$.

For a concrete example, Berman et al. (2015) estimate the effects of log firm exports on the log of its domestic sales to measure scale economies or diseconomies. While our discussion so far has

[21] Similar issues arise with other transformations of SSIVs, e.g. using a dummy that a shift-share variable is in the lowest quartile of its distribution, as in Greenstone et al. (2020).

focused on outcomes and treatments measured as changes, consistent with the decomposition (3), Berman et al. (2015) perform the analysis in logs of levels, using a panel of firms and controlling for firm fixed effects. They instrument log exports with $z_{it} = \log \sum_k s_{ik} G_{kt}$, where k denotes product-by-country pairs, s_{ik} is the share of this pair in firm's exports (on average across periods), and G_{kt} is the total world exports of this product to this country. Leveraging exogeneity of G_{kt}, or the log-changes in G_{kt} over time, would require the corrections discussed above.[22]

8. What if I have multiple shift-share instruments? This is fine, both when multiple shift-share variables instrument for a single treatment and when multiple IVs are necessitated by multiple endogenous variables. One should just perform the relevant steps for each of the SSIVs, e.g. include incomplete share controls in the exogenous shocks approach and compute Rotemberg weights in the exogenous shares approach.

Getting exposure-robust standard errors may be more involved in this case. When the shares are the same but there are several sets of exogenous shocks, BHJ show how the shock-level equivalent IV regression extends in this case, yielding correct standard errors. Appendix B below extends this result by allowing for several SSIVs that use different shares and different shocks, as long as all shocks are defined at the same "level" k. We derive an equivalent shock-level representation of the estimator in terms of a set of moment conditions (but no longer as a simple IV). This equivalence result yields exposure-robust standard errors. A Stata example is available in our GitHub repository.

We give two examples. First, Dauth et al. (2014) consider the impacts of two import competition shocks in Germany, originating from the growth of China and from the accession of Eastern European countries into the European Union. Both are shift-share variables that combine the local employment shares of different industries with two national industry import competition shocks.

Second, including both direct and spillover effects of a certain treatment in the same specification can be viewed as using two shift-share variables with the same shocks but different shares. For instance, the right-hand side variables in Miguel and Kremer (2004) are the student i' own deworming dummy and the number of her dewormed friends. We explained above how their spillover treatment is a SSIV that uses deworming dummies as the shocks g_k and the patterns of friendship as exposure weights. Mechanically, one's own deworming status is also a shift-share with the exposure weight being one for $i = k$ and zero otherwise.

9. What if I have interaction terms in a shift-share regression? This is similar to having multiple SSIVs.

There can be two types of interaction terms in SSIV regressions. A more conventional one interacts z_i with some unit-level variable a_i. For instance, in the ADH context, one may be interested in understanding whether labor market responses to import competition vary by the share of college graduates in the region. This interaction can be written as a SSIV with the same shocks and different

[22] Borusyak and Hull (2021, footnote 82) show an additional problem with this IV: it implicitly uses the shares that are not the s_{ik} and does not capture the intended economic intuition. More appropriate constructions are either $\sum_k s_{ik} \log G_{kt}$ or $\log \sum_k s_{ik}(G_{kt}/G_{k0})$, where 0 is some base period.

exposure weights: $a_i z_i = \sum_k (a_i s_{ik}) g_k$.

The second type — albeit not exactly an interaction — aims to identify the heterogeneous responses to different groups of *shocks*. For instance, Bombardini and Li (2020) consider the health effects of two treatments: regional exposure to the national industry growth of exports in all industries and in pollution-intensive industries in particular. The former is a standard shift-share variable $z_i = \sum_k s_{ik} g_k$ while the latter can be written as $z'_i = \sum_k s_{ik}(b_k g_k)$ where b_k is industry's pollution intensity. This z'_i is a SSIV with shares s_{ik} and shocks $g_k b_k$.[23]

10. Can the instruments in Card (2009) and Bartik (1991) be valid without exogenous shares? No.

In Card (2009), the shocks measure national growth rates of migration from origin countries k. If the initial shares of migrants are exogenous, such construction of shocks provides a natural way to combine the share instruments (see Section 2.4). However, share exogeneity is a strong condition; can shock exogeneity provide a distinct path to identification, per Section 2.3?

Similarly, Bartik (1991) estimated the (inverse) elasticity of labor supply by instrumenting local employment changes with a shift-share instrument that combined local employment shares of different industries with the national growth rate of employment in each industry. This IV is valid if it is uncorrelated with supply factors in the error term. Interestingly, Blanchard and Katz (1992) focus on the shocks when introducing the Bartik (1991) instrument: *"This series will be valid for our purposes [of isolating a labor demand shock] as long as the national growth rates are not correlated with labor supply shocks in the state"* (p. 25). Can the properties of the shocks ensure the Bartik IV exogeneity without exogeneity (with respect to the labor supply shocks) of initial employment shares?

To explain why our answer is negative, consider the local labor market equilibrium in the Bartik (1991) setting. Local employment growth rates of each industry are confounded by the local supply shocks. One may hope that, at the national level, industry growth rates only reflect labor demand, while location-specific labor supply factors may average away. While this can in principle happen, it turns out that this requires local industry shares to be uncorrelated with the error terms: i.e., the shares to be exogenous. If the local shares s_{ik} of some industry k are endogenous — because, say, regions with a high share of that industry are particularly attractive to immigrants — the national growth of that industry will be affected by immigration and thus endogenous, too. In general, the shocks constructed as national averages of equilibrium objects that mechanically include the error term (or some components of it) can only be viewed as exogenous if the underlying shares are exogenous, too.[24]

[23] It can also be viewed as a SSIV with shares $s_{ik} b_k$ and shocks g_k. Both interpretations lead to the same practical conclusions, in different ways. For instance, with as-good-as-random g_k, one needs to control for $\sum_k s_{ik} b_k$. In the former interpretation this follows because the shocks $g_k b_k$ can be considered as-good-as-random only controlling for b_k (while the shares add up to one). In the latter interpretation this follows because the shares $s_{ik} b_k$ add up to $\sum_k s_{ik} b_k$ (while the shocks are already as-good-as-random)..

[24] Appendix C provides a simple formal argument. BHJ propose to view the shocks in Bartik (1991) and Card (2009) as approximately exogenous, e.g. national industry growth rates as mostly reflecting labor demand factors and incorporating labor supply factors to an innocuously small extent. However, per the above discussion, this requires

11. What is the role of leave-one-out construction of shocks? This is a useful way of mitigating bias when pooling variation from many share instruments.

In settings like Bartik (1991) where, as explained above, the shocks can be mechanically confounded by the errors, it is common sinceAutor and Duggan (2003) to use "leave-out" constructions of SSIVs: $z_i = \sum_k s_{ik} g_{k,-i}$, where $g_{k,-i}$ is, say, the industry growth rate in all regions except i (or perhaps except nearby regions, too).[25] BHJ show that using leave-one-out means to construct the national growth rates is useful to address the finite sample bias that can mechanically arise when using own-observation information. This approach is similar to how jackknife instrument variable estimators avoid bias of 2SLS in presence of many instruments (Angrist et al., 1999).

In practice, Autor and Duggan (2003) observed that including own region in shock construction made the IV first stage substantially stronger, raising concerns about the mechanical relationship. Other authors (e.g., GPSS) found that the leave-out correction is empirically minor when the measured shocks average over sufficiently many observations.

We finally note that the leave-out constructions of SSIVs are distinct from a practice of measuring the shocks from entirely different data, e.g. in different countries. ADH, for instance, instrument regional import penetration in the U.S. using industry shocks measured in other developed countries; Hummels et al. (2014) and Aghion et al. (2022) use similar approaches when instrumenting firm-level imports. Unlike leave-out constructions, here the shocks g_k are the same for all units in the sample. Moreover, the mechanical correlation between the error term and the shocks does not arise, such that the exogenous shocks approach can be applied under the appropriate assumptions (e.g., that import demand shocks in the U.S. and other developed economies are uncorrelated in the ADH context).[26]

B SSIV Regressions with Multiple Endogenous Variables

Consider a just-identified SSIV regression

$$y_i = \beta_1 x_{1i} + \cdots + \beta_R x_{Ri} + \gamma' w_i + \varepsilon_i \tag{6}$$

where $x_{1i}, \ldots, x_{Ri}$ are instrumented with a set of SSIVs $z_{1i}, \ldots, z_{Ri}$ for $z_{ri} = \sum_{k=1}^{K} s_{rik} g_{rk}$. Both the shares and the shocks can differ across r but we require the number of shocks K to be the same for all r. The vector w_i collects other controls, including the intercept. For any variable v_i, let $v_i^{\perp}$ denote the in-sample projection of v_i on w_i and $\tilde{v}_k^{(r)} = \frac{1}{N} \sum_i s_{rik} v_i^{\perp}$. We note that, if w_i includes $\sum_k s_{rik} q_{rk}$ for some vector q_{rk}, we have $\sum_k \tilde{v}_k^{(r)} q_{rk} = 0$.

the initial shares to be approximately exogenous, too.

[25]Strictly speaking, such z_i is not a SSIV, as defined by equation (1), since $g_{k,-i}$ has some variation across units.

[26]In settings like Hummels et al. (2014), the researcher may therefore entertain two options: to measure the shocks in a different country and follow the exogenous shocks approach, or to measure the shocks in the country of interest in a leave-out way and follow the exogenous shares approach.

The IV estimator $\hat{\beta}$ for $\beta = (\beta_1, \ldots, \beta_R)'$ in (6) satisfies a system of R equations:

$$\frac{1}{N}\sum_i \left(y_i^{\perp} - \sum_{j=1}^{R} \hat{\beta}_j x_{ji}^{\perp} \right) z_{ri} = 0.$$

Expanding z_{ri}, exchanging the order of summation, and combining terms yields

$$\sum_k \left(\tilde{y}_k^{(r)} - \sum_{j=1}^{R} \hat{\beta}_j \tilde{x}_{jk}^{(r)} \right) g_{rk} = 0.$$

Letting $\tilde{g}_{rk}$ be the projection of g_{rk} on q_{rk} weighted by $s_{rk} = \frac{1}{N}\sum_i s_{rik}$, we further have for $r = 1, \ldots, R$:

$$\sum_k \tilde{\psi}_k^{(r)} = 0, \quad \text{for } \tilde{\psi}_k^{(r)} = \left(\tilde{y}_k^{(r)} - \sum_{j=1}^{R} \hat{\beta}_j \tilde{x}_{jk}^{(r)} \right) \tilde{g}_{rk}$$

In matrix form, this can be rearranged as

$$\Omega \hat{\beta} = M$$

where $\Omega_{rj} = \sum_k \tilde{x}_{jk}^{(r)} \tilde{g}_{rk}$ and $M_r = \sum_k y_k^{(r)} \tilde{g}_{rk}$. Thus, $\hat{\beta} = \Omega^{-1} M$.[27] Moreover, since (6) implies $\tilde{y}_k^{(r)} - \sum_{j=1}^{R} \beta_j \tilde{x}_k^{(r)} = \tilde{\varepsilon}_k^{(r)}$ for true β, we also have

$$\hat{\beta} - \beta = \Omega^{-1} E \qquad \text{for } E_r = \sum\nolimits_k \varepsilon_k^{(r)} \tilde{g}_{rk}.$$

We assume that the appropriate relevance condition holds and suppose that vectors of shock residuals $\tilde{g}_k = (\tilde{g}_{1k}, \ldots, \tilde{g}_{Rk})'$ are asymptotically independent across shock clusters c. Letting $\tilde{\psi}_k = \left(\tilde{\psi}_k^{(1)}, \ldots \tilde{\psi}_k^{(R)} \right)'$, we then have:

$$\begin{aligned} \mathrm{Var}\left[\hat{\beta} - \beta\right] &\approx \Omega^{-1} \mathrm{Var}\left[E\right] (\Omega^{-1})' \\ &\approx \Omega^{-1} \left(\sum_c \left(\sum_{k \in c} \tilde{\psi}_k \right) \left(\sum_{k \in c} \tilde{\psi}_k \right)' \right) (\Omega^{-1})'. \end{aligned}$$

C Share and Shock Exogeneity in Bartik (1991)

This appendix illustrates that, when the shock construction mechanically includes the error term, the exogeneity of measured *shocks* generally requires *share* exogeneity. We show this in the context of the Bartik (1991) estimation of the local labor supply elasticity. A simple constant-elasticity model of equilibrium in local labor markets in BHJ (Appendix A.7) shows that the equilibrium

[27] Before proceeding to computing standard errors, one should verify that this formula at the shock level gives numerically the same answer as the initial IV regression.

employment growth of industry k in region i equals

$$x_{ik} = x_{ik}^* + \gamma \varepsilon_i,$$

where x_{ik}^* reflects a combination of labor demand shocks (to industry k, as well as other industries via equilibrium spillovers), ε_i is the local labor supply shock, and $\gamma \in (0, 1)$ depends on the labor demand and labor supply elasticities. In this model, workers can freely switch across industries within a region, and thus labor supply shocks are purely regional, with equal effects on employment of all industries.

Under this model, the national industry growth rate equals

$$g_k = \sum_i \frac{X_{ik0}}{X_{k0}} x_{ik} = \sum_i \frac{X_{ik0}}{X_{k0}} x_{ik}^* + \sum_i \frac{X_{ik0}}{X_{k0}} \varepsilon_i,$$

where X_{ik0} and X_{k0} are the initial local and national employment levels in the industry, respectively.

The contribution of labor supply shocks to the national industry growth rate vanishes — as would be required to view it as an exogenous labor demand shock — if and only if

$$0 = \sum_i \frac{X_{ik0}}{X_{k0}} \varepsilon_i = \left(\sum_i \frac{X_{i0}}{X_0} \cdot \frac{X_{ik0}}{X_{i0}} \varepsilon_i \right) / \left(X_{k0}/X_0 \right),$$

where X_{i0} and X_0 is the initial regional and national employment, overall.

In this expression, $\sum_i \frac{X_{i0}}{X_0} \cdot \frac{X_{ik0}}{X_{i0}} \varepsilon_i$ is the covariance between the local share of industry k, $\frac{X_{ik0}}{X_{i0}}$, and the error term ε_i weighted by the region's employment. Thus, it is equal to zero precisely when the condition for share exogeneity is satisfied, provided the regional analysis is weighted by initial employment, as is commonly done.