Combining Quasi-Experimental Shocks with Exogenous Exposure: A General Framework

Kirill Borusyak
Princeton

Peter Hull
U Chicago & NBER*

February 19, 2019

Abstract

We study the properties of "shock-exposure instruments," constructed from a set of quasi-experimental shocks and endogenous measures of heterogeneous exposure. Validity of these instruments generally requires a simple but non-standard correction, derived from knowledge of counterfactual shocks that might have been realized. Such design knowledge can also be used for exact randomization inference and specification tests that are valid in finite samples. We further characterize the shock-exposure instruments that are asymptotically efficient. This framework has practical implications for the use of shift-share instruments, simulated eligibility instruments, model-implied instruments, and for other designs. We illustrate these implications in two applications.

*Contact: borusyak@princeton.edu and hull@uchicago.edu. We are grateful to Rodrigo Adão, Vishal Kamat, Michal Kolesár, Gabriel Kreindler, Eduardo Morales, and seminar participants and Princeton and U Chicago for helpful comments. We also thank Rafael Dix-Carneiro, Brian Kovak, and Yatang Lin for providing replication code and data.

1 Introduction

Researchers use instrumental variables (IVs) to leverage quasi-experimental variation in an endogenous treatment and estimate its causal effects. In simple settings each unit of observation receives its own as-good-as-randomly-assigned shock, which serves as the econometrician's instrument. Often, however, a natural experiment affects many observations jointly but to different extents: for example, improvements to transportation networks may differentially increase the connectivity of cities, changes in international trade policy may differentially affect the economic conditions of regions, or reforms to entitlement programs may differentially expand the eligibility of individuals. In these settings researchers might use a *shock-exposure instrument*, constructed as a function of the quasi-experimental shocks and a measure of heterogeneous exposure. Variation in such instruments across observations only arises from exposure differences, which need not be exogenous. This raises the question of whether shock-exposure instruments can be valid levers for causal inference.

We take a design-based approach to this question. We argue that researchers have strong grounds to claim quasi-experimental identification with shock-exposure IVs when they are able to specify sets of counterfactual shocks that were as likely to have occurred as the ones used to construct the instrument (or, more generally, a probability distribution over different potential shock realizations). One might, for example, believe that different railroad lines had an equal chance of being selected for construction from an initial transportation development plan, that observed tariff changes across different industries could as well have been exchanged, or that policy variation across states is as-good-as-randomly determined with respect to potential outcomes or trends. Our framework for shock-exposure instruments exploits this knowledge of the shock *assignment process*.

We first show that a simple design-based adjustment ensures the validity of shock-exposure instruments, which is not otherwise guaranteed. Intuitively, validity fails when observations with different unobservables receive systematically different values of the instrument. For shock-exposure instruments this may arise from non-random exposure: more centrally-located regions in a transportation network, for example, are likely to see systematically larger increases in predicted connectivity no matter which railroad lines are selected for construction. Knowledge of the shock assignment process, however, allows researchers to simulate the typical value of the instrument across different realizations of the quasi-experiment; we show that the deviation of a shock-exposure instrument from this *expected instrument* is always a valid IV. The *adjusted instrument* implicitly compares units with higher-than-expected values of the instrument to those that received lower-than-expected values. Its validity thus follows from the randomnessof shocks, though it continues to derive power from heterogeneous shock exposure.

We next show how design knowledge can be used to overcome challenges of inference with shock-exposure instruments. A natural experiment affecting many observations jointly may induce a complex and non-classical dependence structure, with observations "clustered" by their relative exposure to different shocks. Conventional asymptotic confidence intervals may then have incorrect coverage, confounding inference on estimated effects and empirical validations of the design. We adapt principles of randomization inference to overcome this challenge, developing a general procedure by which researchers can construct exact confidence intervals for second-stage effects and balance coefficients, with robustness to weak instruments. We complement this finite-sample approach with an asymptotic analysis of different shock-exposure constructions, characterizing the functions of quasi-experimental shocks and exposure measures that yield efficient IV estimates in large samples when an appropriate central limit theorem holds.

Our framework provides new insights for a number of empirical designs in economics, including shift-share (or "Bartik") instruments (Bartik, 1991; Blanchard and Katz, 1992; Card, 2001; Autor et al., 2013), simulated instruments for public health care eligibility (Currie and Gruber, 1996a,b; Cohodes et al., 2016), synthetic instruments for income tax changes and unemployment insurance eligibility (Cullen and Gruber, 2000; Gruber and Saez, 2002; East and Kuka, 2015), and model-implied instruments in international trade (Allen et al., 2018; Adão et al., 2019). Instruments for transportation infrastructure or frictions derived from models of economic geography, as in Donaldson (2018), Donaldson and Hornbeck (2016) and Allen et al. (2019), can also be studied within our framework, as can instruments generated by centralized school assignment mechanisms (Abdulkadiroglu et al., 2017). We discuss how in each setting the expected instrument correction and randomization inference approach relate to and depart from conventional methods, as well as suggest ways to improve the asymptotic power of the shock-exposure instruments.

Our analysis also relates to several econometric and statistical literatures. Sufficiency of the expected instrument correction follows similarly to that of the propensity score (Rosenbaum and Rubin, 1983) in simpler quasi-experimental designs. Our use of exact randomization inference also draws on a rich theoretical literature, dating back to Fisher (1935) and Pitman (1937) and recently deployed in quasi-experimental settings by, among others, Bertrand et al. (2004), Imbens and Rosenbaum (2005), Ho and Imai (2006), Abadie et al. (2010), Dell and Olken (2018), and Ganong and Jäger (2018). In particular we apply recent results from DiCiccio and Romano (2017) regarding the asymptotic robustness of permutation tests to certain violations of design assumptions, such as shock heteroskedasticity. Our derivation of asymptotically efficient shock-exposure instruments extends the analysis of optimal estimation with *iid* data and simpler designs in Chamberlain (1987).[1] As noted above our analysis also has implications for the econometrics of shift-share IV, contributing to a recent methodological literature that includes Goldsmith-Pinkham et al. (2018), Jaeger et al. (2017), Borusyak et al. (2018), and Adão et al. (2018). Lastly, our paper is related to the literature that takes a design approach to estimating causal effects under interference across units, as in Aronow and Samii (2017).[2]

[1] Adão et al. (2019) use the *iid* approach of Chamberlain (1987) to characterize efficient instruments in a model of spatial linkages and international trade. As discussed in Section 4, our characterization of efficient shock-exposure instruments allows for the complex dependence structure induced by shocks, as well as endogeneity of shock exposure.

[2] The Aronow and Samii (2017) setting pertains to estimating reduced-form effects of discretely-valued treatments, via

We illustrate the practical implications of our framework in two applications. We first show the necessity of the expected instrument correction by estimating the local employment effects of increased regional connectivity, leveraging a recent expansion of high-speed rail (HSR) in China (Lin, 2017). While panel regressions of urban employment growth on changes in regional market access suggest a large and statistically significant causal effect, we find this to be likely driven by systematic differences in a city's exposure to potential HSR quasi-experiments. Outcome pre-trends are strongly correlated with expected market access treatments, constructed under different assumptions on the HSR assignment process. Correspondingly, treatment effect estimates that correct for expected market access yield much smaller IV estimates and permutation-based 95% confidence intervals that do not exclude the null hypothesis of no effect.

Our second application contrasts the randomization inference approach with conventional asymptotics for shift-share instruments that exploit exogenous shocks. A concern raised by Goldsmith-Pinkham et al. (2018) is that the asymptotic framework of Borusyak et al. (2018) and Adão et al. (2018) may fail to accurately capture the distribution of such shift-share IV estimators when the number of shocks is small or a small subset of them are highly influential. We show in a reanalysis of Autor et al. (2013) and Dix-carneiro and Kovak (2018) that this concern may be warranted, reinforcing the importance of exact randomization inference. Permuting the 20 industry shocks in Dix-carneiro and Kovak (2018) leads to confidence intervals more than three times as wide as the asymptotic confidence intervals, including those which – as in Anderson and Rubin (1950)– impose the null for enhanced finite sample performance. In Autor et al. (2013), however, randomization-based and asymptotic confidence intervals tend to coincide, both with the original sample of 397 industry shocks and when aggregating industries to only 20 sectors. This aligns with the simulation results of Adão et al. (2018), which cautions against asymptotic shift-share inference when industry exposure shares are both small and skewed: one industry accounts for 50% of the average exposure across regions in Dix-carneiro and Kovak (2018), while shares are much more dispersed in Autor et al. (2013). Nevertheless, randomization-based inference remains valid in both settings, addressing the general concern of Goldsmith-Pinkham et al. (2018).

The remainder of the paper is organized as follows. The next section develops our general framework for shock-exposure instruments, showing how knowledge of the shock assignment process can be used to ensure validity. Section 3 then considers inference, outlining our design-based procedure for exact confidence intervals and characterizing asymptotically efficient shock-instrument constructions. We map the general framework to specific examples of shock-exposure instruments in Section 4,

inverse propensity score weighting, when the quasi-experiment individually assigns a treatment that potentially violates the Stable Unit Treatment Value Assumption. In contrast our approach accommodates IV estimation of the effects of arbitrary treatments, while imposing no restrictions on the nature of quasi-experimental shock assignment. Our papers also differ in our approach to inference: we obtain exact randomization-based inference assuming constant treatment effects, whereas they construct asymptotically conservative confidence intervals with unrestricted heterogeneity while assuming that only a small fraction of unit pairs are affected by the same shocks. As discussed in Section 3, inference with heterogeneous and interdependent effects is a common challenge in both settings.

discussing practical implications for shift-share instruments, simulated eligibility instruments, transportation instruments, and for other designs. Section 5 then presents our two applications, illustrating the importance of the expected instrument correction and our randomization inference procedure, and Section 6 concludes. All proofs of propositions, and details on the empirical replications, are included in the appendix.

2 Validity and Consistency of Shock-Exposure Instruments

2.1 Setting

We consider a setting in which a vector of outcomes y_ℓ and treatments x_ℓ are observed across L units. For simplicity we assume y_ℓ and x_ℓ are both scalar, though this is straightforward to generalize. Of interest is a causal parameter β, relating treatment to outcomes by

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

where ε_ℓ denotes a structural residual. Here and throughout our baseline analysis we assume the treatment effect of interest is constant. Appendix A.1 derives conditions under which the results of this section apply to models with nonlinear and heterogeneous causal response, and we discuss challenges of inference with heterogeneous effects in Section 3.

We seek to estimate β using quasi-experimental variation in an $N \times 1$ vector of shocks g. Throughout we assume this vector is as-good-as-randomly assigned with respect to the structural residual, perhaps conditional on an observed $K \times 1$ vector s:

$$g \perp\!\!\!\perp \varepsilon \mid s. \tag{2}$$

In simple quasi-experimental designs, $(y_\ell, x_\ell, \varepsilon_\ell, g_\ell, s_\ell)$ is drawn *iid* from a population (so $N = L$ and $K = L \times P$, where P denotes the dimension of s_ℓ) with $g_\ell \perp\!\!\!\perp \varepsilon_\ell \mid s_\ell$. The *iid* assumption guarantees that the shock assigned to unit ℓ has no impact on other units and that unobservables are also independent. Our more general treatment relaxes both of these assumptions, while maintaining the conditional independence of the shock vector from unobservables. In this sense variation in g is quasi-experimental, though it may affect many observations jointly.[3]

To leverage (2), we construct an instrument with values $z_\ell = f_\ell(g; s)$ given by the known functions f_ℓ, quasi-experimental shocks g, and observables s. Differences in $f_\ell(\cdot; s)$ across ℓ thus capture

[3] Note that equations (1) and (2) impose an implicit exclusion restriction: variation in shocks may only affect the outcome of unit ℓ via its own treatment status. Treatment thus serves as a sufficient statistic capturing the effect of all shocks. Such a restriction is typically necessary for valid IV estimation; in practice it may follow from a particular economic model (e.g. Adão et al. (2019)) or be relaxed by including multiple treatments in x_ℓ.

differences in the exposure of observation ℓ to the shocks.[4] For example in canonical shift-share IV $f_\ell(g;s) = \sum_n s_{\ell n} g_n$, where g_n often denotes the growth rate of an industry n and $s_{\ell n}$ is the employment share of that industry in each region ℓ (Borusyak et al., 2018). More generally the f_ℓ may be nonlinear functions of g that depend on features of the full sample of observations, as when z_ℓ denotes regional market access growth predicted, based on an economic model, from transportation upgrades g and a start-of-period equilibrium s.

We consider IV estimates of β obtained by instrumenting x_ℓ with z_ℓ in a regression of y_ℓ; this is

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell z_\ell y_\ell}{\frac{1}{L}\sum_\ell z_\ell x_\ell}, \tag{3}$$

where we have implicitly assumed the y_ℓ and x_ℓ have been de-meaned in the sample (below we will discuss the potential role for other included controls w_ℓ). Note that this framework allows $x_\ell = z_\ell$, in which case β reflects the causal effect of the instrument itself and $\hat{\beta}$ is an ordinary least squares estimate of β.

2.2 Adjusted Shock-Exposure Instruments

We say that the set of z_ℓ serves as a *valid* instrument for β when it is orthogonal to the structural residual: i.e. when $\mathbb{E}\left[z_\ell \varepsilon_\ell\right] = 0$ for each ℓ. We show in this section that validity of z_ℓ is not guaranteed; however, a simple adjustment for the instrument, derived from the knowledge of the quasi-experimental design, ensures the validity of any shock-exposure instrument.

Our interest in validity stems from the fact that, by (1),

$$\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell y_\ell\right] = \beta \cdot \mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right] + \frac{1}{L}\sum_\ell \mathbb{E}\left[z_\ell \varepsilon_\ell\right]. \tag{4}$$

Thus when in large samples the numerator and denominator of (3) are close to their expectation, $\hat{\beta}$ will be close to β when validity holds. We formalize this logic in the next subsection; here it is worth emphasizing that, absent the usual *iid* assumption, instrument validity is considered for each observation individually, with expectations taken over the distribution of complete vectors of shocks g and observables s.[5]

While g is itself as-good-as-randomly assigned, this need not imply validity of $f_\ell(g;s)$. Intuitively, non-random variation in s may also drive variation in z_ℓ which need not be orthogonal to ε_ℓ. In the shift-share example, if there are shocks only to manufacturing industries, which are positive on

[4] We express the dependence of z_ℓ on s explicitly, incorporating the case where s arises from some random process along with g (see, e.g., the discussion of shock permutation classes below). However all of our primary results hold with s treated as fixed, so it can be suppressed.

[5] This intuition for validity follows by the assumption of a constant causal effect β. We show in Appendix A.1 that more generally the ratio of expectations of the numerator and denominator of (3) identifies a weighted average of heterogeneous treatment effects under validity and an additional monotonicity condition; this parallels the canonical result of Imbens and Angrist (1994).

average, the instrument will be systematically higher in regions with a high manufacturing share, which may be correlated with unobservables (Borusyak et al., 2018). In the market access example, more centrally-located regions may systematically see faster growth in economic connectivity, while their unobservables may differ, too. In both cases validity will be violated, so that $\hat{\beta}$ need not provide a reliable estimate of β.

To overcome this challenge and isolate the quasi-experimental variation in shock-exposure instruments we define the *expected instrument*, given by

$$\mu_\ell(s) = \mathbb{E}\left[f_\ell(g; s) \mid s\right]. \tag{5}$$

This denotes the typical realization of z_ℓ, for each ℓ, conditional on the vector of observables s and across various realizations of the quasi-experiment. In simple *iid* quasi-experimental designs, when $z_\ell = g_\ell$ is binary and further the treatment and instrument are the same, $\mu_\ell(s)$ is equivalent to the well-known propensity score of Rosenbaum and Rubin (1983). Our expected instrument generalizes this concept to an instrumental variables framework with non-binary z_ℓ and dependence across observations.

As with propensity scores, knowledge of $\mu_\ell(s)$ derives from knowledge of the quasi-experimental design. For example one might know that g is exchangeable, in the sense that all permutations of shocks are equally likely to arise. Then

$$\mu_\ell(s) = \frac{1}{N!} \sum_{\pi \in \Pi} f_\ell(\pi(g); s), \tag{6}$$

where Π denotes the set of permutation operators $\pi(\cdot)$ on vectors of length N, and s includes the permutation class $\{\pi(g) \mid \pi(\cdot) \in \Pi\}$. More generally, knowledge of the conditional distribution of g given s is sufficient for μ_ℓ to be known, though knowledge of certain conditional moments may also suffice. With shift-share IV, for example, $\mu_\ell(s)$ is given by the first moment $\mathbb{E}[g \mid s]$.

Adjustment of any shock-exposure instrument by its expected instrument ensures validity. This follows simply from the law of iterated expectations, and from (2). We have, for each ℓ,

$$\begin{aligned}
\mathbb{E}\left[(z_\ell - \mu_\ell(s))\,\varepsilon_\ell\right] &= \mathbb{E}\left[\mathbb{E}\left[(f_\ell(g, s) - \mu_\ell(s))\varepsilon_\ell \mid s\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[f_\ell(g, s) - \mu_\ell(s) \mid s\right]\mathbb{E}\left[\varepsilon_\ell \mid s\right]\right] \\
&= 0.
\end{aligned} \tag{7}$$

Intuitively, by subtracting the expected instrument of observation ℓ across different realizations of g, one isolates the quasi-experimental variation in z_ℓ. Comparisons across values of the *adjusted instrument* $\tilde{z}_\ell = z_\ell - \mu_\ell(s)$ contrast observations shocked more than expected, given the design, with observations shocked less than expected. Variation in $\tilde{z}_\ell$ thus continues to arise from differences in each observation's exposure to g, though (unlike with z_ℓ) this variation is guaranteed to be not systematically related to ε_ℓ. We next establish conditions for the validity of $\tilde{z}_\ell$ to yield consistent estimates of the causal effect β.

2.3 Consistency of Adjusted Shock-Exposure IV

The deviation of the adjusted shock-exposure IV estimator $\tilde{\beta}$ from the true causal effect can be written

$$\tilde{\beta} - \beta = \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell}. \tag{8}$$

We consider the standard condition for consistency of $\tilde{\beta}$: $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell$ weakly converges to its expectation as $L \to \infty$, with a first stage $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell$ that is asymptotically non-zero.

Since the same set of shocks may jointly determine many $\tilde{z}_\ell$, a potentially complex correlation structure may arise across observations of $\tilde{z}_\ell \varepsilon_\ell$. This precludes the use of traditional weak laws of large numbers or their standard extensions (e.g., to clustering or ergodicity) for showing weak convergence of $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell$. To restrict those correlations, assumptions can be imposed either on the $\tilde{z}_\ell$ or on the ε_ℓ. We follow the design-based approach that typically views potential outcomes, captured by ε_ℓ in our constant effects framework, as unrestricted and even fixed (Imbens and Rubin, 2015); this approach allows *unobserved* factors to also affect observations jointly, in an unspecified manner.

We start with a high-level assumption on $\tilde{z}_\ell$ that guarantees consistency of the adjusted shock-exposure IV; we then establish lower-level sufficient conditions that are easier to verify in specific designs. Throughout, we maintain the following regularity condition on the existence of a non-zero asymptotic first stage and the conditional boundedness of the structural residual variance:

Assumption 1. $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \xrightarrow{p} M$ *for non-zero* M *and* $\mathbb{E}\left[\varepsilon_\ell^2 \mid s\right] < B_\varepsilon$ *for finite* B_ε*.*

For consistency, the dependence in $\tilde{z}_\ell$ across ℓ must be sufficiently limited, in which case $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell$ will be close to its expectation regardless of the properties of the residuals. For that to hold, observations must be well differentiated in terms of their exposure to g. Formally, we have the following result:

Proposition 1. *Suppose Assumption 1 holds. Then* $\tilde{\beta} \xrightarrow{p} \beta$ *if* $\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid s\right]\right|\right] \to 0$*.*

Here the key condition states that the average absolute cross-observation covariance of the adjusted instrument $\tilde{z}_\ell$ converges to zero as L grows.[6] Typically this would require the number of shocks N to grow with L, so that only a small fraction of observation pairs are most exposed to the same shocks.

Our two sufficient conditions for the average absolute covariance of $\tilde{z}_\ell$ to vanish are non-nested:

Corollary 1. *Suppose* $\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid s\right] \geq 0$ *almost surely for all* ℓ *and* m*, and Assumption 1 holds. Then* $\tilde{\beta} \xrightarrow{p} \beta$ *if* $\mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \mid s\right]\right] \to 0$*. Moreover, if* $f_\ell(g; s)$ *is weakly monotone in* g *for all* ℓ*, and*

[6] A straightforward reformulation requires the average absolute *correlation* to converge to zero, as long as $\mathbb{E}\left[\tilde{z}_\ell^2 \mid s\right]$ is also uniformly bounded.

components of g are jointly independent, $\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid s\right] \geq 0$ *almost surely.*

Corollary 2. *Suppose $G_\ell \subseteq \{1, \ldots, N\}$ is such that $f_\ell(\cdot; s)$ does not depend on g_n for any $n \notin G_\ell$ almost surely, and Assumption 1 holds. Then $\tilde{\beta} \xrightarrow{p} \beta$ if $\sum_{\ell,m} \mathbf{1}\left[G_\ell \cap G_m \neq \varnothing\right] = o(L^2)$, components of g are jointly independent, and $\mathbb{E}\left[\tilde{z}_\ell^2 \mid s\right]$ is uniformly bounded.*

The first corollary applies to the setting when all shocks affect all observations in the same direction, but to different extents. This holds, for example, for shift-share instruments with non-negative shares. More generally nonlinear $f_\ell(g; s)$ may also be monotone in the shock vector; for example each transportation infrastructure upgrade may weakly improve connectivity everywhere. In these cases the adjusted IV estimator is consistent when the average instrument $\frac{1}{L}\sum_\ell \tilde{z}_\ell$ weakly converges to its expectation (zero). For shift-share IV this extra condition requires the number of shocks to grow with L with the average exposure to each individual shock becoming vanishingly small, as in Borusyak et al. (2018) and Adão et al. (2018). The assumption of independent shocks can be weakened, for instance to allow for shocks that are independent across many clusters.

The second corollary follows Aronow and Samii (2017), in assuming that for most pairs of observations the two instruments $\tilde{z}_\ell$ and $\tilde{z}_m$ rely on non-overlapping sets of shocks g. This would be the case, for example, when each observation receives its own quasi-experimental shock, and $f_\ell(\cdot; s)$ only depends on ℓ's shock and those of its neighbors up to a fixed network distance.

Two points on the consistency of the adjusted IV estimator are worth highlighting. First, while Proposition 1 considers a regression with no controls, this is easily generalized. Namely the y_ℓ and x_ℓ may denote residuals from projecting observed outcomes and treatment on a control vector w_ℓ; by the Frisch-Waugh-Lovell theorem Proposition 1 then shows consistency of controlled IV regressions, provided (2) holds on the w_ℓ-adjusted ε_ℓ. In general observables determining the assignment of shocks should be included in s and thus incorporated in $\mu_\ell(s)$; controlling for other w_ℓ that are independent of g given s may improve finite-sample properties of the estimator but will not affect consistency.

Second, a consistent estimate of β can also be obtained by controlling linearly for the expected instrument, rather than by adjusting z_ℓ. This follows because

$$\operatorname{Cov}\left[z_\ell, \mu_\ell(s)\right] = \operatorname{Var}\left[\mu_\ell(s)\right] \tag{9}$$

for each ℓ, with $\mathbb{E}\left[\mu_\ell(s)\right] = \mathbb{E}\left[z_\ell\right]$. Thus the auxiliary regression of z_ℓ on $\mu_\ell(s)$ produces residuals $z_\ell - \hat{\gamma} - \hat{\lambda}\mu_\ell(s)$ with $\hat{\lambda} \xrightarrow{p} 1$ and $\hat{\gamma} \xrightarrow{p} 0$ under similar conditions as in Proposition 1. Again by the Frisch-Waugh-Lovell theorem, the IV estimator which uses the unadjusted z_ℓ and controls for $\mu_\ell(s)$ then effectively uses an instrument which asymptotically coincides with $\tilde{z}_\ell$.[7] This alternative approach can be particularly useful when the distribution of $g \mid s$ is not fully specified but $\mu_\ell(s)$ is known to

[7]Despite using the same instrument, the controlling estimator will be asymptotically more efficient when $\mu_\ell(s)$ predicts the residual. However, using the adjusted instrument and controlling for any other function of s that is correlated with the residual would do just the same.

be linear in a fixed number of parameters; then these parameters can be estimated jointly with β in large samples (see Section 4 for an example in the shift-share setting).

3 Finite-Sample Inference and Asymptotic Efficiency

3.1 Randomization Inference

We now show how knowledge of the quasi-experimental shock design can be used to construct exact tests and confidence intervals for β. Even in finite samples of both observations and shocks, where $\tilde{\beta}$ may be far from β and observations exhibit complex unobserved dependencies, such confidence intervals will have correct coverage due to the conditional independence of shocks and residuals. This approach follows a long tradition of randomization inference in statistics, adopted here to the shock-exposure IV setting.

We begin by considering a test of some null hypothesis $\beta = \beta_0$; for example, that outcomes are unaffected by treatment ($\beta_0 = 0$). Consider some statistic $T = \mathcal{T}(g, y - \beta_0 x, s)$, where g, y, and x collect the corresponding observations of shocks, outcomes, and treatments.[8] Under the null, $T = \mathcal{T}(g, \varepsilon, s)$ and the distribution of T conditional on ε and s is known from the shock assignment process, by virtue of (2). Simulating this distribution by drawing or permuting shocks thus reveals the conditional distribution of T under the null; if the observed value is in the tails of that distribution, we may reject β_0.

Formally, let g^* denote a random vector that, conditionally on s, is independent of (g, x, y) with the same distribution as g. Under exchangeability, for example, g^* may denote a random row-permutation of g. This yields a corresponding statistic $T^* = \mathcal{T}(g^*, y - \beta_0 x, s)$. For some scalar $\alpha \in (0,1)$, define $T^*_{\alpha/2} = \sup\left\{t \in \mathbb{R} \cup \{-\infty\} : \; Pr\left(T^* < t \mid y, x, s\right) \leq \frac{\alpha}{2}\right\}$ and, symmetrically, $T^*_{1-\alpha/2} = \inf\left\{t \in \mathbb{R} \cup \{+\infty\} : \; Pr\left(T^* > t \mid y, x, s\right) \leq \frac{\alpha}{2}\right\}$. These are computed from knowledge of the conditional distribution of g given s, under the null. For example under exchangeability they may be given by the observed distribution of $\mathcal{T}(\pi(g), y - \beta_0 x, s)$ across all row permutations π. We then have the following result:

Proposition 2. *Under the null $\beta = \beta_0$,*

$$Pr\left(T \in \left[T^*_{\alpha/2}, T^*_{1-\alpha/2}\right]\right) \geq 1 - \alpha, \tag{10}$$

with equality when T^* is continuously distributed.

Proposition 2 shows that a test of $\beta = \beta_0$ which rejects when $T \notin \left[T^*_{\alpha/2}, T^*_{1-\alpha/2}\right]$ has size of no more than α, and is thus valid.[9] In practice the quantiles of the T^* distribution can be approximated

[8] In principle $\mathcal{T}$ may be a stochastic function, though we abstract away from this possibility here.

[9] Note that when T^* is not continuously distributed it is still possible to construct a test of exact size by introducing randomness in $\mathcal{T}$ (e.g. Lehmann, 1986, p. 233).

using a large but finite number of draws from the assignment process, e.g. permutations (Lehmann and Romano, 2006, p. 636).

Different statistics $\mathcal{T}$ will lead to different tests of comparable size under the null, though they may differ in their power against alternative hypotheses. Given our focus on the IV estimator $\tilde{\beta}$ a natural choice for $\mathcal{T}$ is the sample covariance between the adjusted instrument and the implied error term:

$$\mathcal{T}\left(g, y-\beta_0 x, s\right) = \frac{1}{L}\sum_{\ell=1}^{L}\left(f_\ell\left(g; s\right) - \mu_\ell(s)\right)\left(y_\ell - \beta_0 x_\ell\right). \tag{11}$$

Imbens and Rosenbaum (2005) study randomization inference with such statistics in the canonical instrumental variables setting, finding its power to be comparable with conventional strong-instrument asymptotic inference in simulations. There, as here, tests based on (11) are still guaranteed to have correct size when the shock-exposure instrument is weak, in the sense of inducing a small first-stage coefficient.

In cases where such statistics are asymptotically normal (e.g., the shift-share example we discuss in Section 4), DiCiccio and Romano (2017) recommend "studentizing," by subtracting from $\mathcal{T}$ an estimate of its asymptotic mean and dividing by an estimate of its asymptotic standard deviation. This does not, of course, affect the finite-sample coverage of the test under β_0, but may yield additional robustness properties in large samples. For example tests based on a model for $g \mid s$ in which the g_n are *iid* may be studentized to be asymptotically valid when shocks are instead heteroskedastic. In principle studentizing can also provide valid asymptotic inference in the presence of treatment effect heterogeneity, though asymptotic variance estimates may be difficult to derive in this case (see, e.g., the discussion of shift-share inference in Adão et al. (2018)).

To construct confidence intervals for β with exact coverage, we may invert the test corresponding to a given $\mathcal{T}$; formally, we have the following:

Proposition 3. *Let CI denote the set of $\beta \in \mathbb{R}$ not rejected by a test constructed as in Proposition 2. Then $Pr\left(\beta \in CI\right) \geq 1-\alpha$, with equality if the distribution of T^* is continuous.*

In some settings the confidence interval CI may be infinite (on one or both sides) or empty, with the latter providing evidence against validity of the design (Imbens and Rosenbaum, 2005).[10]

Proposition 2 can also be used to validate the quasi-experimental model (1)–(2), both by conventional tests of covariate balance and no outcome pre-trends and by the observation in Section 2 that $\mathrm{Cov}\left[z_\ell, \mu_\ell(s)\right] = \mathrm{Var}\left[\mu_\ell(s)\right]$ and $\mathbb{E}\left[z_\ell\right] = \mathbb{E}\left[\mu_\ell(s)\right]$. For any observed variable r_ℓ such that $g \perp\!\!\!\perp r \mid s$, we may compare $\frac{1}{L}\sum_\ell \left(f_\ell(g; s) - \mu_\ell(s)\right) r_\ell$ to the simulated distribution of $\frac{1}{L}\sum_\ell \left(f_\ell(g^*; s) - \mu_\ell(s)\right) r_\ell$ to form a valid specification test. When $r_\ell = 1$ or $r_\ell = \mu_\ell(s)$ this tests $\mathbb{E}\left[z_\ell - \mu_\ell(s)\right] = 0$ or that the re-

[10] It is worth noting that in Propositions 2 and 3 we have not relied on the validity of the instrument. Intuitively, the bias in IV arising from endogeneity in exposure to shocks affects each T and T^* in the same way, which cancels out when the test is performed. Indeed, if the $\mu_\ell(s)$ adjustment is not done on the instrument, both T and the distribution of T^* would be shifted by the same quantity, $\frac{1}{L}\sum_\ell \mu_\ell(s)(y_\ell - \beta_0 x_\ell)$.

gression slope of z_ℓ on $\mu_\ell(s)$ is one, respectively, while setting r_ℓ to some predetermined characteristic or pre-trend yields balance tests.

3.2 Optimal Shock-Exposure Instruments

Along with the test statistic $\mathcal{T}$, one may choose between different sets of shock-exposure instrument constructions f_ℓ. This choice is also likely to affect the power of null hypothesis tests and lengths of associated confidence intervals. Here we characterize the f_ℓ that minimize the asymptotic variance of $\hat{\beta}$ and thus efficiently use the quasi-experimental variation in shocks. This extends the classic result of Chamberlain (1987) to the shock-exposure setting. Constructing these optimal instruments may not be feasible in practice, and typically requires an economic model for the dependence of treatment on shocks. Our characterization nevertheless provides guidance for shock-exposure instrument construction in general.

Formally, we have the following result:

Proposition 4. *Let $f_1, \dots, f_L$ be functions inducing the adjusted instrument $\tilde{z}$ and IV estimator $\tilde{\beta}$, where $K_L \sum_\ell \tilde{z}_\ell \varepsilon_\ell$ and $M_L \sum_\ell \tilde{z}_\ell x_\ell$ converge to a distribution and non-zero constant, respectively, for some sequences of positive K_L and M_L. Then, given appropriate continuity, the asymptotic variance of $\tilde{\beta}$ is minimized by functions inducing*

$$\tilde{z}^* = \mathbb{E}\left[\varepsilon\varepsilon' \mid s\right]^{-1} \left(\mathbb{E}\left[x \mid g, s\right] - \mathbb{E}\left[x \mid s\right]\right) \tag{12}$$

and equals

$$\operatorname{Var}\left[\frac{K_L}{M_L}\tilde{\beta}^*\right] = \frac{K_L^2}{M_L^2}\mathbb{E}\left[\left(\mathbb{E}\left[x \mid g, s\right] - \mathbb{E}\left[x \mid s\right]\right)' \mathbb{E}\left[\varepsilon\varepsilon' \mid s\right]^{-1} \left(\mathbb{E}\left[x \mid g, s\right] - \mathbb{E}\left[x \mid s\right]\right)\right]^{-1}. \tag{13}$$

The proof to Proposition 4 shows that since

$$\operatorname{Var}\left[(K_L/M_L)\tilde{\beta}\right] = \frac{K_L^2}{M_L^2}\frac{\operatorname{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2}(1 + o(1)), \tag{14}$$

the limiting re-scaled variance of $\tilde{\beta}$ is smallest for the $\tilde{z}$ that minimizes $\operatorname{Var}\left[\tilde{z}'\varepsilon\right]/\mathbb{E}\left[\tilde{z}'x\right]^2$, over all adjusted shock-exposure instruments which converge. Analogous to Chamberlain (1987) this is (12), which reweights a conditional expectation of treatment given shocks and s. A key difference in (12) is the subtraction of $\mathbb{E}\left[x \mid s\right]$, corresponding to the expected instrument adjustment. Another difference is that when $\mathbb{E}\left[\varepsilon \mid s\right] \neq 0$, the $\mathbb{E}\left[\varepsilon\varepsilon' \mid s\right]^{-1}$ reweighting is based not only on the residual variance but also the residual mean. If $(\varepsilon_\ell, g_\ell, x_\ell)$ are *iid* and s is constant, so no adjustment is necessary and $\mathbb{E}\left[\varepsilon \mid s\right] = 0$, the optimal shock-exposure instrument is $\tilde{z}_\ell^* = \mathbb{E}\left[x_\ell \mid g_\ell\right]/\operatorname{Var}\left[\varepsilon_\ell\right]$, as in Chamberlain (1987) with a fully-independent instrument g_ℓ.

Practical implementation of the optimal instrument typically requires a contextual model. The key ingredient in (12) is the best prediction for each observation's treatment based on all shocks and observables. With a large and *iid* sample this can potentially be estimated non-parametrically (e.g. Newey (1990)), though the curse of dimensionality may make it difficult in finite samples. However, when x_ℓ is jointly determined by all shocks g, observables s and some unobservable shocks u (i.e., $x_\ell = x_\ell(g, s, u)$ with $u \perp\!\!\!\perp g \mid s$ and $\mathbb{E}[u] = 0$), *a priori* restrictions will typically be needed on this mapping, such as those based on economic theory. The reweighting term $\mathbb{E}[\varepsilon\varepsilon' \mid s]^{-1}$ need not be implied even by such models.

While constructing the optimal shock-exposure instrument may be challenging, some instruments used in practice can be viewed as approximating $\tilde{z}^*$. Shift-share instruments have been derived from economic models as first-order approximations to how treatment changes in response to shocks g (Kovak, 2013; Adão et al., 2018), though as discussed below these instruments typically do not require the expected instrument adjustment. Moreover, some papers (e.g., Donaldson and Hornbeck (2016) and Tsivanidis (2017)) that do not impose linearity instead use a model to predict the counterfactual change in treatment in the scenario where g is the only shock, while all other exogenous variables are kept at their original equilibrium values. In terms of the above notation the instrument can therefore be written $x_\ell(g, s, 0)$ and viewed as a convenient approximation to $\mathbb{E}[x_\ell(g, s, u)]$: instead of averaging across all values of unobserved shocks, which would require additional knowledge of their distributions, x_ℓ is evaluated at the means of those shocks (see Berry et al. (1999), for the same idea in an entirely different context). These papers however do not perform the expected instrument adjustment, as required by Proposition 4.

To conclude, we make two observations about the role of controls. First, as in any regression, including controls correlated with the residual will tend to improve efficiency. This remains true here, even when using the optimal adjusted instrument corresponding to such controls: the $\mathbb{E}[\varepsilon\varepsilon' \mid s]$ matrix in (13) will tend to increase when functions of s are first partialled out. Second, if the ideal control $\mathbb{E}[\varepsilon \mid s]$ is known, the researcher can improve on (13). Replacing y with $\check{y} = y - \mathbb{E}[\varepsilon \mid s]$ and thus ε with $\check{\varepsilon} = \varepsilon - \mathbb{E}[\varepsilon \mid s]$ (while not adjusting x) implies $\mathbb{E}[\check{\varepsilon} \mid g, s] = 0$. In that case any function $f_\ell(g, s)$ is a valid instrument, and the expected instrument adjustment – which isolates the variation in g but reduces power – is not necessary. By analogy with Proposition 4, $\check{z} = \mathbb{E}[\check{\varepsilon}\check{\varepsilon}' \mid s]^{-1}\,\mathbb{E}[x \mid g, s]$ is then optimal, and generally differs from the optimal adjusted instrument. However, in practice $\mathbb{E}[\varepsilon \mid s]$ is not likely to be given by an economic model for treatment, making this approach infeasible.

4 Practical Implications

We now discuss implications of our framework for some common examples of shock-exposure instruments, as commonly deployed in the literature. These include shift-share or "Bartik" instruments, simulated eligibility instruments, and instruments derived from upgrades in transportation infrastructure, among others. Bringing a design-based approach to these settings reveals various ways researchers can ensure valid and powerful inference with straightforward adjustments to current methods.

4.1 Shift-Share Instruments

Shift-share instruments average a set of observed shocks with observation-specific shares. In a typical case, a regional instrument z_ℓ is constructed from industry shocks g_n by $z_\ell = \sum_n s_{\ell n} g_n$, where $s_{\ell n}$ measures the industry's share (of, say, employment) in the region. We assume that the shocks are quasi-randomly assigned, whereas the shares measuring regional shock exposure may be correlated with the structural residual, for example because unobserved industry shocks also affect regions via the same shares. Our discussion of shift-share instruments builds on and extends our earlier work (Borusyak et al., 2018, henceforth BHJ); to keep its presentation short, we refer the reader to BHJ for more details and examples of the setup.[11]

Viewing shift-share instruments as shock-exposure instruments delivers two main new insights. First, valid inference can be performed without the key assumption in BHJ and Adão et al. (2018) that industries are numerous and sufficiently dispersed in terms of the share in an average region. This assumption yields a tractable asymptotic approximation for shift-share IV, though as pointed out by Goldsmith-Pinkham et al. (2018) may limit its applicability. In contrast the randomization inference approach of this paper provides valid confidence intervals regardless of the number and concentration of shocks. Of course, if there are too few shocks or the concentration is too large, the quasi-experiment may not have enough power to reject interesting economic hypotheses. It is difficult to judge *a priori* what constitutes "too few" and "too large," however, whereas exact confidence intervals provide a valid data-driven answer.

Second, the power of the shift-share instrument can be improved by using economic theory, while preserving robustness to model misspecification. Kovak (2013) and Adão et al. (2018) justify shift-share instruments as first-order approximations of certain models of trade. Our results provide a framework for using nonlinear f_ℓ and thus avoiding such approximations. If the shocks are large (as they are in many applications) and the approximation is not precise, a feasible approximation for the optimal instrument from Section 3.2 may deliver a substantial improvement in power. These assumptions are already made (although possibly in a slightly weaker form) when the conventional shift-share instrument is used.

We demonstrate this point in Appendix **??**, by extending the model of Kovak (2013) to allow for imperfectly elastic regional labor supply. We then show that if a quasi-random industry demand

[11] The alternative approach of Goldsmith-Pinkham et al. (2018), in which shares are exogenous while shocks need not be, also fits within our shock-exposure instrument framework, with an exchangeable matrix of shares as g and the vector of shocks as s. The shock-exposure view however does not bring new insights in this case, unless the number of regions is small.

shock (e.g., from a trade liberalization) is observed, the model yields an intuitive *nonlinear* shock-exposure instrument for regional employment in the labor supply equation. We then characterize its power advantage compared to the shift-share approximation. Importantly, the adjusted shock-exposure instrument remains valid even if the economic model is misspecified: it only relies on the second-stage exclusion restriction and knowledge of the assignment process.

A limitation of the shock-exposure approach is that for conducting inference we assume full independence of the industry shocks from all shares and structural residuals, whereas for BHJ a weaker shock orthogonality assumption is sufficient. However, this limitation can be remedied by using robust randomization inference. Indeed, unlike general shock-exposure instruments, asymptotic standard errors are available for shift-share instruments which allow for heteroskedasticity or clustering in shocks(Adão et al., 2018). These can be used to studentize the test statistic when performing randomization inference, as described in Section 3.1. In this case, the resulting confidence intervals will be asymptotically robust to heteroskedasticity and clustering, while still being exact in finite samples without those deviations. That said, we find the performance of confidence intervals that studentize via heteroskedasticity-robust standard errors to be poor with few shocks (details are available upon request). In those applications below we instead use homoskedastic standard errors from the industry-level regression.

As a final note, we show how one result from BHJ can be derived and understood differently within the shock-exposure framework. BHJ show that when the regional shares do not add up to one, e.g. when only manufacturing industries participate in the quasi-experiment, one needs to control for this sum of shares in the IV regression. Indeed, if $\mathbb{E}\left[g_n \mid s\right] = \alpha$ for all industries n, the expected instrument is just a rescaled sum of regional shares: $\mu_\ell(s) = \mathbb{E}\left[\sum_n s_{\ell n} g_n \mid s\right] = \alpha \cdot \sum_n s_{\ell n}$. As we have shown in Section 2.3, adjusting a shock-exposure instrument is asymptotically equivalent to linearly controlling for it. More generally when the expected shock is a linear function of industry-level observables, $\mathbb{E}\left[g_n \mid q, s\right] = \tau' q_n$, the expected shift-share instrument equals $\mu_\ell(q,s) = \mathbb{E}\left[\sum_n s_{\ell n} g_n \mid q, s\right] = \tau' \sum_n s_{\ell n} q_n$, and it is thus sufficient to control for the vector of share-weighted observables $\sum_n s_{\ell n} q_n$, again as recommended by BHJ.

4.2 Simulated Eligibility Instruments

Our framework also brings insights to settings in which quasi-experimental policy variation differentially affects the eligibility of individuals for a government program, such as Medicaid or unemployment insurance (UI).[12] Currie and Gruber (1996a; 1996b, henceforth CG) famously construct "simulated instruments" for eligibility x_ℓ in regressions of individual outcomes, such as program takeup, health status or educational attainment. These instruments are designed to address the endogeneity of eligibility, which depends both on policies and on non-random individual characteristics that may be

[12] We thank Paul Goldsmith-Pinkham for bringing our attention to these designs.

correlated with potential outcomes. The CG solution is to leverage the plausible exogeneity of policies themselves, by constructing an instrument that only varies across an individual's state-of-residence.[13] Here we show that valid instruments that preserve individual-level variation can be constructed using the shock-exposure approach, potentially leading to substantial power gains.

The CG procedure simulates the average eligibility of a representative individual in the U.S., if she were to reside in each state and be subject to its policies; this "simulated instrument" is then assigned to individuals on the basis of their actual state-of-residence. In terms of our notation, $g = (g_1, \ldots, g_{50})'$, where g_n represents the eligibility policy of state n, and s collects observations of r_ℓ and w_ℓ, where r_ℓ indexes the state in which individual ℓ resides and w_ℓ are other observed demographics. With $h(g_n)$ denoting the simulated eligibility of the population under policy g_n, the CG instrument can then be written as a shock-exposure instrument: $z_\ell^{CG} = f_\ell(g; s) = h(g_{r_\ell})$.

When state policies are as-good-as-randomly assigned with respect to potential outcomes or trends ε_ℓ (i.e., the rows of g are exchangeable), the CG instrument is valid without further adjustment; in our framework the expected instrument $\mathbb{E}\left[f_\ell(g; s) \mid s\right] = \frac{1}{50} \sum_n h(g_n)$ is constant. Per Section 3, valid tests of $\beta = \beta_0$ can be constructed by comparing the z_ℓ^{CG} IV statistic (11) to the tails of its distribution under state policy permutations, although clustering standard errors by state is also likely to be asymptotically valid.

As in Section 3, more powerful tests and shorter confidence intervals may be obtained by replacing the CG instrument with a shock-exposure instrument that better predicts individual eligibility. In practice, all individual characteristics relevant to eligibility determination – such as income, family structure, or employment status – may be observed and contained in w_ℓ. In such cases $\mathbb{E}\left[x_\ell \mid g, s\right] = x_\ell$, or an individual's eligibility itself. Thus apart from the heteroskedasticity adjustment, the optimal instrument (12) is $\tilde{z}_\ell^* = x_\ell - \mathbb{E}\left[x_\ell \mid s\right]$, and the efficient IV procedure is equivalent to a regression of y_ℓ on z_ℓ^*. Our framework then motivates a regression of outcomes on eligibility, adjusted by or controlling for the individual's average eligibility over random reallocations of policies. This contrasts with the CG approach: while they apply ℓ's state policy to random individuals to construct the instrument, we apply random states' policies to ℓ to construct the control.

There are two additional considerations motivating the use of the CG instrument which can be incorporated in our approach. First, in some settings not all determinants of eligibility are observed in w_ℓ. For example Cohodes et al. (2016) study the long-term effects of Medicaid eligibility on children without observing its key determinant (parental income). They use a simulated eligibility instrument which assigns to each individual of a given age, race, and birth year the average eligibility among such individuals nationally if they were subject to her state-of-residence's policy. Our framework instead motivates instrumenting by the average eligibility of individuals with similar demographics actually

[13]For example, Currie and Gruber (1996a) write that their aim is "to achieve identification using only legislative variation in Medicaid policy" (p. 445). We interpret this as positing quasi-experimental variation in policies across states.

residing in each state, while adjusting for the average value of this instrument over permutations of state policies. By virtue of Proposition 4 this will again improve efficiency, to the extent the non-observed determinants of eligibility are systematically different across states.

Second, even when all eligibility determinants are observed, a researcher may not wish to include them in r_ℓ. This would be the case when, for example, parental income responds endogenously to the state policy, violating (2). Indeed, Currie and Gruber (1996a) discuss this as motivation for their original simulated instrument construction (p. 445). In such cases exogenous predictors of such determinants, such as parental income from before a state policy change, may be used instead. East and Kuka (2015) use a similar approach to augment simulated instrument construction in evaluating the effects of UI eligibility.

Our design-based approach suggests subtracting or controlling for the average true or predicted eligibility of an individual across permutations of state policies. Since this adjustment term is a function of the individual's observables (excluding state dummies), in principle one could instead flexibly control for w_ℓ. Such controls have an additional benefit of potentially predicting variation in the error term, improving asymptotic efficiency. Indeed, this approach is used by Cullen and Gruber (2000) and East and Kuka (2015) in their examinations of UI eligibility effects. However, Gruber (2003) finds this strategy difficult for Medicaid because several demographic variables have nonlinear effects on eligibility. Focusing on the (perhaps implicit) quasi-experiment in such investigations – that of as-good-as-random state-level policy variation – reveals the single control needed for valid causal inference, avoiding the curse-of-dimensionality when the space of observables is rich while still allowing for other flexible controls to reduce residual variance.

4.3 Transportation Instruments

A large body of research evaluates the impacts of improved transportation infrastructure, such as railroads and highways, on local population, wages, and rents (see Redding and Turner (2015) for a review). This literature has identified two important conceptual problems for identification: infrastructure upgrades may have effects that extend beyond their direct proximity and may be non-randomly distributed across space and time. Here we discuss how our framework sheds new light on the biases conventional methods may involve, as well as ways to overcome them.

Traditionally, infrastructure improvements were assumed to only affect the locations where they took place (e.g., Baum-Snow, 2007; Berger and Enflo, 2017; Donaldson, 2018). A more recent literature has recognized that infrastructure may affect other locations as well, violating the usual Stable Unit Treatment Value Assumption. The typical solution to this issue is to discipline the empirical framework with economic theory; standard gravity models of spatial equilibrium yield a sufficient statistic – "market access" – which summarizes the effects of all transportation networks on a given location (Redding and Venables, 2004; Donaldson and Hornbeck, 2016; Tsivanidis, 2017). These models imply the following log-linear structural equation (abstracting from some modeling details that may vary across papers), expressed in changes relative to an original equilibrium:

$$\Delta \log y_{\ell t} = \beta \Delta \log m_{\ell t} + \varepsilon_{\ell t}. \tag{15}$$

Here $y_{\ell t}$ is region ℓ's outcome (e.g., rents) in the t, $\varepsilon_{\ell t}$ captures productivity and other shocks, and

$$m_{\ell t} = \sum_k \tau_{\ell k t}^{-\theta} n_{kt} \tag{16}$$

defines market access in year t.[14] Market access is computed from measures of measures of market size across regions, $n_{\ell t}$ (e.g., population), and iceberg transportation costs between ℓ and other locations k, $\tau_{\ell k t}$. These costs are in turn determined by transportation infrastructure: $\tau_{\ell k t} = \tau_{\ell k t}\left(g_t, s_t\right)$, where we separate factors that faced a recent upgrade g_t (e.g., the railroad network) from other infrastructure factors s_t. We assume, consistent with the literature, that the mapping $\tau_{\ell k t}(\cdot)$ is known (given by, e.g., the speed of various transportation modes and how travel time maps into iceberg costs), as is the value of the elasticity θ.

We present two novel insights that the shock-exposure framework implies for these regression, one conceptual and one practical. First, many papers express concerns that estimating β by OLS may be biased, as transportation networks may be "selected endogenously"; this concern is present both when regressing $\Delta \ln y_{\ell t}$ on an indicator for a region's change in connectivity to the network and in market access regressions. The main proposed remedies, reviewed by Redding and Turner (2015), focus on the notion of "endogenously:" they use planned, historical, or minimum-cost routes that are not affected by recent trends in the outcome as instruments for the observed transportation network. However, we emphasize that "selected" is also important, for both consistency and inference. Researchers tend to consider "exogeneity" in the network while not discussing the set of counterfactual networks that could have taken place instead. It is unlikely that all regions had equal chances to get connected, and without an explicit stand on the counterfactual networks it is difficult to judge whether the places that were not connected form a valid control group.

It may appear difficult to imagine and specify counterfactual networks. One approach is to exploit knowledge of network upgrade plans, appealing to the quasi-randomness of which subset of plans materializes at all or by some time period. Donaldson (2018) and Berger and Enflo (2017) take a step in this direction, using planned but unbuilt railroads in a placebo exercise, as does Lin (2017) by exploiting engineering problems that slowed down construction of certain lines in one extension.[15]

[14] This expression is the approximation used by Donaldson and Hornbeck (2016) in empirical work, although the original recursive formula (p. 825) could instead be used. An advantage of this approximation in our framework is that it weakly increases in all transportation improvements, helping to establish consistency of shock-exposure IV as per Corollary 1. A prevalent further simplification sets $\theta = 1$, which corresponds to the Harris (1954) "market potential" measure.

[15] Dell and Olken (2018) provide an interesting example of this approach in a different context: they prepare 1000 counterfactual maps of sugar processing factories in 19th-century Java, taking into account factors that affected factory

Our framework, however, suggests a different use for upgrade plans than the typical approach to using them as instruments: rather than viewing planned network as exogenous, one may consider the built subset as exogenous, conditionally on the plan.

Second, we highlight an endogeneity problem of market access regressions that is present even if the network expansion is quasi-experimental, and propose a simple solution. Even when g_t is as-good-as-randomly determined (e.g., upgrades are equally likely across different railroad lines), $\Delta \ln m_{\ell t}$ is still affected by changes in market size of various locations $n_{\ell t}$, which are endogenous to unobserved productivity shocks in $\varepsilon_{\ell \ell}$. A natural solution is to construct the instrument $z_{\ell t} = \Delta \ln \hat{m}_{\ell t}$ as the predicted change in market access, following (16) with n_t (along with other transportation infrastructure factors s_t) held fixed at the original level (e.g. Tsivanidis, 2017). This $\Delta \ln \hat{m}_{\ell t}$ is thus a shock-exposure instrument which uses the original equilibrium as a measure of exposure and transportation upgrades as shocks. As in our general framework, however, fixing n_t and s_t is not sufficient for ensuring validity of this instrument. If the set of planned railroads is dense in some part of the country (e.g., the economic centers), market access will increase there no matter which railroads are selected. Since central places may be systematically different for reasons unrelated to the quasi-experiment (for example, due to changes in other transportation), this could cause $z_{\ell t}$ and $\varepsilon_{\ell t}$ to be correlated, creating a bias. As with any shock-exposure instrument, the expected instrument adjustment, made possible by formalizing the quasi-experiment, solves the problem. The adjusted instrument thus measures the growth of market access predicted by the observed network upgrades (holding other determinants fixed), relative to the average growth of MA across counterfactual upgrades.

Inference should also be adjusted appropriately in transportation regressions and IVs. Long railway lines, such as the Pacific Railroad in the 19th-century U.S. or Lanzhou-Xinjiang line in present-day China, may have large impacts on market access across large areas. Randomness in their construction route or timing thereby generates correlation in the instrument that may not be well captured by traditional spatially-clustered standard errors. Our design-based approach to inference again overcomes this issue, by simulating the distribution of statistics over different realizations of the transportation quasi-experiment.

We finally note that these ideas similarly apply to a related literature that uses transportation infrastructure changes to estimate gravity equation (e.g., Duranton et al., 2013). For source location ℓ and destination location k, a typical estimating equation expressed in time differences is

$$\Delta \log y_{\ell k} = \beta \Delta \log \tau_{\ell k} + \alpha_\ell + \beta_k + \varepsilon_{\ell k}, \tag{17}$$

where $y_{\ell k}$ denotes the flow of goods or people from ℓ and k. For Allen et al. (2019), for example, $\tau_{\ell k}$ is travel time between locations in the U.S. and Mexico, with a time change due to construction of

locations (e.g., proximity to a river and distance from each other).

a wall in some parts of the border between the countries. They find, however, that the travel time change is correlated with the distance between locations and other potential confounders. This is to be expected: regardless of which sections of the wall are built, places far away from the border will be affected less, as there would always be alternative routes between them unaffected by the wall. Allen et al. (2019) provide evidence that the estimates are not affected by this endogeneity. In general however taking an explicit stand on the counterfactual distribution of walls and performing our expected instrument adjustment is guaranteed to isolate quasi-experimental variation. Our design approach further allows for inference that takes into account that the same segment of the wall may simultaneously affect pairs of locations that are far away from each other.

4.4 Other Designs

We conclude this section by discussing implications for model-implied optimal instruments (Adão et al., 2019), as well as for instruments generated by partially-randomized centralized assignment mechanisms (Abdulkadiroglu et al., 2017). In both cases the shock-exposure framework relaxes certain restrictions, while suggesting a more flexible basis for inference.

Adão et al. (2019) develop a spatial general equilibrium model that implies tractable estimating equations for the structural parameters. Then they derive optimal instruments for the endogenous regressors, using the structure of the model. In constructing instruments, Adão et al. (2019) make two key assumptions. First, they assume that the structural residuals (e.g., unobserved supply-shifters) across U.S. local labor markets are mean-zero conditional on both the set of observed shocks g (e.g., China productivity growth across industries, similarly to Autor et al. (2013)) and the original equilibrium s; in our notation, $\mathbb{E}\left[\varepsilon_\ell \mid g, s\right] = 0$. Here s includes, for example, regional shares of different industries in a previous period, which are implicitly assumed to be valid instruments (as in Goldsmith-Pinkham et al. (2018)). Our framework clarifies that this can be replaced by a weaker assumption, $\mathbb{E}\left[\varepsilon_\ell \mid g, s\right] = \mathbb{E}\left[\varepsilon_\ell \mid s\right]$, requiring only exogenous changes in Chinese productivity in Adão et al. (2019). Permuting observed shocks across industries then yields the expected instrument adjustment for any $z_\ell = f_\ell(g, s)$, with Proposition 4 characterizing the optimal instrument among the adjusted IV estimators, which generally does not coincide with the one in Adão et al. (2019).

The second implicit assumption in Adão et al. (2019), used in their derivation of the optimal instrument, is that the structural residuals and instrument are independent across regions. This allows them to apply the results of Chamberlain (1987) directly. However such independence is not likely to hold, since each shock affects many of the U.S. regions both directly (as in Autor et al. (2013)) and indirectly (via spatial linkages within the U.S.). The proof technique we use in Proposition 4 is therefore more appropriate, as it allows for unrestricted dependence across observations. As usual, our randomization inference framework also remains valid in this case, while validity of the conventional asymptotics is not guaranteed.

Our approach also applies to settings in which instruments arise from randomizations embedded in centralized assignment mechanisms, such as deferred acceptance algorithms for school choice. For such mechanisms the indicator for assignment of a student ℓ to a given school or school type can be written in the shock-exposure form, $z_\ell = f_\ell(g; s)$. The set of student rankings over schools and administrative school priorities over students is given by s, while g contains a set of lottery numbers used to break ties among equal-priority students. Abdulkadiroglu et al. (2017) use market design theory to derive instrument propensity score formulas for deferred acceptance, conditional on which z_ℓ is a valid instrument. More generally they consider mechanisms satisfying the "equal treatment of equals" (ETE) property, in which students with the same preferences and administrative priorities face the same risk of assignment to each school. Abdulkadiroglu et al. (2017) show that with ETE it is sufficient to flexibly control for preferences and priorities when instrumenting with z_ℓ or, to overcome the curse of dimensionality, to control for the average assignment rate of a student across redrawings of g, holding s fixed. This simulated propensity score aligns with our expected instrument correction.

The shock-exposure view offers two further insights to this setting. First, the ETE assumption is not necessary: simulated propensity score can be obtained for any known centralized mechanism, even those that do not equalize assignment risk across students of the same type. The validity of instruments that correct for this score arises simply from the experimental variation in g. Second, our framework again gives a design-based inference approach that accounts for the inherent dependencies of school offers across students, and remains valid when student potential outcomes are not independent. The latter might arise when applicants with similar preferences and priorities are similar in other unobserved ways, or because of peer effects.

5 Applications

5.1 The Employment Effects of Chinese High-Speed Rail

In progress

5.2 Randomization Inference in Shift-Share Designs

In progress

5.3 Boosting Power of Simulated Eligibility Instruments

In progress

6 Conclusions

Natural experiments can have widespread and heterogeneous effects across many or all observations of treatment and outcomes. We develop a general framework to leverage this variation in conventional IV regressions, avoiding bias from the non-random exposure of observations to quasi-experimental shocks. The key ingredient to forming adjusted shock-exposure instruments is the researcher's stance on the shock assignment process, which specifies the conditional distribution of counterfactual shocks that may have as well occurred. This design-based approach allows for valid inference in finite samples of observations and shocks – following the classic statistical literature of randomization inference – though we also provide guidance for optimal shock-exposure instrument construction as in the asymptotic approach of Chamberlain (1987). Our discussion of shift-share, simulated eligibility, and transportation instruments illustrates practical implications of these general insights.

We conclude by noting two important limitations of our current analysis. First, while we characterize the shock-exposure IV estimand under general treatment effect heterogeneity in Appendix A.1, our approach to inference and efficiency is fundamentally based on an assumption of constant effects. That is, the randomization tests we discuss in Section 3.1 are guaranteed to have correct size in testing the strong null hypothesis that treatment has no effect on outcomes for any observations, but not the weaker hypothesis that the average effect is zero. In general exact inference for such null hypotheses is a difficult challenge, though recent progress has been made for ensuring robustness of randomization inference to weaker null hypotheses in much simpler settings (Chung and Romano, 2013). In more complex designs, such as shift-share IV, inference robust to heterogenous treatment effects is challenging even with the traditional asymptotic approach (Adão et al., 2018).

A second assumption we maintain throughout is *a priori* knowledge of the quasi-experimental design, or shock assignment process. It is enough in our setting to know the vector of shocks is exchangeable, such that different permutations of the observed vector (perhaps within observable groups) are equally likely to arise. As discussed in the shift-share example of Section 4, it is also enough to know the expected instrument is linear in observables. It is not, however, typically sufficient for shocks to be *iid* conditional on unobservables with an unknown distribution, or for shocks to be clustered by observable groups. In some settings shocks may themselves be estimated from observed data, inducing a complex distribution of the feasible instrument (Borusyak et al., 2018). Characterizing the properties of shock-exposure IVs that use estimated shock assignment processes appears a fruitful area of future study.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program," *Journal of the American Statistical Association*, 105, 493–505.

Abdulkadiroglu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017): "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 85, 1373–1432.

Adão, R., C. Arkolakis, and F. Esposito (2019): "Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence," *NBER Working Paper 25544*.

Adão, R., M. Kolesár, and E. Morales (2018): "Shift-Share Designs: Theory and Inference," *NBER Working Paper 24944*.

Allen, T., C. Arkolakis, and Y. Takahashi (2018): "Universal gravity," *Working Paper*.

Allen, T., C. Dobbin, and M. Morten (2019): "Border Walls," *Working Paper*.

Anderson, T. and H. Rubin (1950): "The Asymptotic Properties of Estimates of the Parameters of A Single Equation in A Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570–582.

Aronow, P. M. and C. Samii (2017): "Estimating average causal effects under general interference, with application to a social network experiment," *Annals of Applied Statistics*, 11, 1912–1947.

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

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

Baum-Snow, N. (2007): "Did highways cause suburbanization?" *Quarterly Journal of Economics*, 122, 775–805.

Berger, T. and K. Enflo (2017): "Locomotives of local growth: The short- and long-term impact of railroads in Sweden," *Journal of Urban Economics*, 98, 124–138.

Berry, S., J. Levinsohn, and A. Pakes (1999): "Voluntary Export Restraints on Automobiles : Evaluating a Trade Policy," *American Economic Review*, 89, 400–430.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

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

Borusyak, K., P. Hull, and X. Jaravel (2018): "Quasi-Experimental Shift-Share Research Designs," *NBER Working Paper 24997*.

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

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chung, E. and J. P. Romano (2013): "Exact and asymptotically robust permutation tests," *Annals of Statistics*, 41, 484–507.

Cohodes, S. R., D. S. Grossman, S. A. Kleiner, and M. F. Lovenheim (2016): "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions," *Journal of Human Resources*, 51, 727–759.

Cullen, J. B. and J. Gruber (2000): "Does Unemployment Insurance Crowd out Spousal Labor Supply?" *Journal of Labor Economics*, 18, 546–572.

Currie, J. and J. Gruber (1996a): "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *The Quarterly Journal of Economics*, 111, 431–466.

——— (1996b): "Saving Babies : The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *Journal of Political Economy*, 104, 1263–1296.

Dell, M. and B. Olken (2018): "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java," *Working Paper*.

DiCiccio, C. J. and J. P. Romano (2017): "Robust Permutation Tests For Correlation And Regression Coefficients," *Journal of the American Statistical Association*, 112, 1211–1220.

Dix-carneiro, R. and B. K. Kovak (2018): "Margins of Labor Market Adjustment to Trade," *Working Paper*.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

Donaldson, D. and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 799–858.

Duranton, G., P. M. Morrow, and M. A. Turner (2013): "Roads and trade: Evidence from the US," *Review of Economic Studies*, 81, 681–724.

East, C. N. and E. Kuka (2015): "Reexamining the consumption smoothing benefits of Unemployment Insurance," *Journal of Public Economics*, 132, 32–50.

Fisher, R. A. (1935): "The design of experiments," .

Ganong, P. and S. Jäger (2018): "A Permutation Test for the Regression Kink Design," *Journal of the American Statistical Association*, 113, 494–504.

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

Gruber, J. (2003): "Medicaid," in *Means-tested transfer programs in the United States*, University of Chicago Press, 15–78.

Gruber, J. and E. Saez (2002): "The elasticity of taxable income: Evidence and implications," *Journal of Public Economics*, 84, 1–32.

Harris, C. D. (1954): "The, Market as a Factor in the Localization of Industry in the United States," *Annals of the association of American geographers*, 44, 315–348.

Ho, D. E. and K. Imai (2006): "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election," *Journal of the American Statistical Association*, 101, 888–900.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467.

Imbens, G. W. and P. R. Rosenbaum (2005): "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168, 109–126.

Imbens, G. W. and D. B. Rubin (2015): *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

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

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

Lehmann, E. L. (1986): *Testing Statistical Hypotheses*, Springer texts in statistics, second edi ed.

Lehmann, E. L. and J. P. Romano (2006): *Testing statistical hypotheses*, Springer Science & Business Media.

Lin, Y. (2017): "Travel costs and urban specialization patterns: Evidence from China's high speed railway system," *Journal of Urban Economics*, 98, 98–123.

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

Pitman, E. (1937): "Significance Tests Which May be Applied to Samples From any Populations," *Journal of the Royal Statistical Society*, 4, 119–130.

Redding, S. J. and M. A. Turner (2015): "Transportation Costs and the Spatial Organization of Economic Activity," in *Handbook of regional and urban economics*, Elsevier, 1339–1398.

Redding, S. J. and A. J. Venables (2004): "Economic geography and international inequality," *Journal of International Economics*, 62, 53–82.

Rosenbaum, P. R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects Paul R. Rosenbaum, Donald B. Rubin," 70, 41–55.

Tsivanidis, N. (2017): "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Working Paper*, 1–59.

A Appendix

A.1 Heterogeneous Treatment Effects

In place of (1)-(2), suppose outcomes and treatment are given by $y_\ell(x_\ell, \upsilon)$ and $x_\ell(h_\ell(g), \eta)$, with $g \perp\!\!\!\perp (\upsilon, \eta) \mid s$. Suppose further that $x_\ell(h, \eta)$ is almost-surely monotone in h, and that the numerator and denominator of the adjusted IV estimator converge in probability. Then

$$\tilde{\beta} = \frac{\frac{1}{L}\sum_\ell \mathbb{E}\left[y_\ell(x_\ell(h_\ell(g), \eta), \upsilon)\left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right)\right]}{\frac{1}{L}\sum_\ell \mathbb{E}\left[x_\ell(h_\ell(g), \eta)\left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right)\right]} + o_p(1). \tag{18}$$

Suppose $y_\ell(x, \upsilon)$ and $x_\ell(h, \upsilon)$ are almost-surely continuous in x and h (an analogous result will follow when this fails) and let $\kappa_\ell(\upsilon, \eta) = \lim_{h \to \infty} y_\ell(x_\ell(h, \eta), \upsilon)$. Writing

$$y_\ell(x_\ell(h_\ell(g), \eta), \upsilon) = \kappa_\ell(\upsilon, \eta) + \int_{-\infty}^{h_\ell(g)} \frac{\partial}{\partial h} y_\ell(x_\ell(h, \eta), \upsilon) dh, \tag{19}$$

we have under appropriate regularity conditions

$$\begin{aligned}
&\mathbb{E}\left[y_\ell(x_\ell(h_\ell(g), \eta), \upsilon)\left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{h_\ell(g)} \frac{\partial}{\partial h} y_\ell(x_\ell(h, \eta), \upsilon) dh \left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right) \mid g, s\right]\right] \\
&= \mathbb{E}\left[\int_G \int_{-\infty}^{h_\ell(g)} \mathbb{E}\left[\frac{\partial}{\partial h} y_\ell(x_\ell(h, \eta), \upsilon) \mid s\right]\left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right) dh dF(g \mid s)\right] \\
&= \mathbb{E}\left[\int_{-\infty}^{\infty} \frac{\partial}{\partial h} y_\ell(x_\ell(h, \eta), \upsilon) \mu_\ell(h; s) dh\right],
\end{aligned}$$

where

$$\begin{aligned}
\mu_\ell(h; s) = &\left(\mathbb{E}\left[f_\ell(g; s) \mid h_\ell(g) \geq h, s\right] - \mathbb{E}\left[f_\ell(g; s) \mid h_\ell(g) < h, s\right]\right) \\
&\times Pr(h_\ell(g) \geq h \mid s)\left(1 - Pr(h_\ell(g) \geq h \mid s)\right).
\end{aligned} \tag{20}$$

Similarly,

$$\mathbb{E}\left[x_\ell(h_\ell(g), \eta)\left(f_\ell(g; s) - \mathbb{E}\left[f_\ell(g; s) \mid s\right]\right)\right] = \mathbb{E}\left[\int_{-\infty}^{\infty} \frac{\partial}{\partial h} x_\ell(h, \eta) \mu_\ell(h; s) dh\right]. \tag{21}$$

Thus

$$\hat{\beta} = \mathbb{E}\left[\sum_\ell \int_{-\infty}^{h_\ell(g)} \frac{\partial}{\partial x} y_\ell(x_\ell(h, \eta), \upsilon) \omega_\ell dh\right] + o_p(1), \tag{22}$$

where

$$\omega_\ell = \frac{\frac{\partial}{\partial h} x_\ell(h,\eta)\mu_\ell(h;s)}{\mathbb{E}\left[\sum_\ell \int_{-\infty}^{h_\ell(g)} \frac{\partial}{\partial h} x_\ell(h,\eta)\mu_\ell(h;s)dh\right]}. \tag{23}$$

When $\mu_\ell(h;s)$ is almost-surely non-negative, this shows the IV regression of y_ℓ on x_ℓ that instruments with $z_\ell - p_\ell$ captures a weighted average of heterogeneous treatment effects $\frac{\partial}{\partial x} y_\ell(x_\ell(h,\eta),\upsilon)$, averaged over the unobserved heterogeneity (υ,η), observations ℓ, and margins of instrument response h. Note that a sufficient condition for monotonicity of $\mu_\ell(h;s)$ is $\mathbb{E}\left[f_\ell(g;s) \mid h_\ell(g) \geq h, s\right]$ being almost-surely monotone in h. This would be satisfied if, for example, $f_\ell(g;s) = \mathbb{E}\left[x_\ell \mid g, s\right]$, as with the efficient instrument derived in Section (3).

A.2 Proof of Proposition 1

We have

$$\begin{aligned}\tilde{\beta} &= \beta + \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell} \\ &= \beta + \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{M}\left(1 + o_p(1)\right) \end{aligned} \tag{24}$$

since $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \xrightarrow{p} M$. Here $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right] = 0$; moreover by conditional independence of g and the Cauchy-Schwartz inequality

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right] &= \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right)^2\right] \\ &= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \varepsilon_\ell \varepsilon_m\right] \\ &= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid s\right]\mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid s\right]\right] \\ &\leq \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid s\right]\right| \sqrt{\mathbb{E}\left[\varepsilon_\ell^2 \mid s\right]\mathbb{E}\left[\varepsilon_m^2 \mid s\right]}\right] \\ &\leq B_\varepsilon \mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid s\right]\right|\right] \\ &\to 0, \end{aligned} \tag{25}$$

Thus $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell \xrightarrow{p} 0$, and by the continuous mapping theorem $\tilde{\beta} \xrightarrow{p} \beta$.

A.3 Proof of Proposition 2

In progress

A.4 Proof of Proposition 3

In progress

A.5 Proof of Proposition 4

Since

$$\operatorname{Var}\left[(K_L/M_L)\tilde{\beta}\right] = \frac{\operatorname{Var}\left[K_L\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[M_L\tilde{z}'x\right]^2}(1+o(1)) \tag{26}$$

the proof follows from showing $\operatorname{Var}\left[K_L\tilde{z}^{*\prime}\varepsilon\right]/\mathbb{E}\left[M_L\tilde{z}^{*\prime}x\right]^2 \leq \operatorname{Var}\left[K_L\tilde{z}'\varepsilon\right]/\mathbb{E}\left[M_L\tilde{z}'x\right]^2$ for any $\tilde{z}$. Note that

$$\begin{aligned}
\frac{\operatorname{Var}\left[K_L\tilde{z}^{*\prime}\varepsilon\right]}{\mathbb{E}\left[M_L\tilde{z}^{*\prime}x\right]^2} &= \frac{K_L^2}{M_L^2}\frac{\mathbb{E}\left[\mathbb{E}\left[\tilde{z}^{*\prime}\varepsilon\varepsilon' z^* \mid g,s\right]\right]}{\mathbb{E}\left[\mathbb{E}\left[\tilde{z}^{*\prime}x \mid g,s\right]\right]^2} \\
&= \frac{K_L^2}{M_L^2}\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]^{-1}
\end{aligned} \tag{27}$$

since $\mathbb{E}\left[\varepsilon\varepsilon'\mid g,s\right]=\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]$ and $\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\mathbb{E}\left[x\mid s\right]\right]=0$. Define

$$U(\tilde{z}) = \tilde{z}'\varepsilon - \mathbb{E}\left[\tilde{z}'x\right]\frac{\varepsilon'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)}{\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]} \tag{28}$$

and note that

$$\begin{aligned}
\mathbb{E}\left[U(\tilde{z})^2\right] &= \mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon'\tilde{z}\right] - 2\mathbb{E}\left[\tilde{z}'x\right]\frac{\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]}{\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]} \\
&\quad + \mathbb{E}\left[\tilde{z}'x\right]^2\frac{\mathbb{E}\left[\left(\varepsilon'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right)^2\right]}{\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]^2} \\
&= \mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon'\tilde{z}\right] - \frac{\mathbb{E}\left[\tilde{z}'x\right]^2}{\mathbb{E}\left[\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]^{-1}\left(\mathbb{E}\left[x\mid g,s\right]-\mathbb{E}\left[x\mid s\right]\right)\right]^2}
\end{aligned} \tag{29}$$

since $\mathbb{E}\left[\varepsilon\varepsilon'\mid g,s\right]=\mathbb{E}\left[\varepsilon\varepsilon'\mid s\right]$ and $\mathbb{E}\left[\tilde{z}'\mathbb{E}\left[x\mid s\right]\right]=0$. Thus

$$\begin{aligned}
\frac{\operatorname{Var}\left[K_L\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[M_L\tilde{z}'x\right]^2} - \frac{\operatorname{Var}\left[K_L\tilde{z}^{*\prime}\varepsilon\right]}{\mathbb{E}\left[M_L\tilde{z}^{*\prime}x\right]^2} &= \frac{K_L^2}{M_L^2}\frac{\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon'\tilde{z}\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} - \frac{\operatorname{Var}\left[K_L\tilde{z}^{*\prime}\varepsilon\right]}{\mathbb{E}\left[M_L\tilde{z}^{*\prime}x\right]^2} \\
&= \frac{K_L^2}{M_L^2}\frac{\mathbb{E}\left[U(\tilde{z})^2\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} \\
&\geq 0.
\end{aligned} \tag{30}$$