Non-Random Exposure to Exogenous Shocks: Theory and Applications

Kirill Borusyak
UCL and CEPR

Peter Hull
UChicago and NBER*

January 2021

Abstract

We develop new tools for estimating the causal effects of treatments or instruments that combine multiple sources of variation according to a known formula. Examples include treatments capturing spillovers in social and transportation networks, simulated instruments for policy eligibility, and shift-share instruments. We show how exogenous shocks to some, but not all, determinants of such variables can be leveraged while avoiding omitted variables bias. Our solution involves specifying counterfactual shocks that may as well have been realized and adjusting for a summary measure of non-randomness in shock exposure: the average treatment (or instrument) across such counterfactuals. We further show how to use shock counterfactuals for valid finite-sample inference, and characterize the valid instruments that are asymptotically efficient. We apply this framework to address bias when estimating employment effects of market access growth from Chinese high-speed rail construction, and to boost power when estimating coverage effects of expanded Medicaid eligibility.

*Contact: k.borusyak@ucl.ac.uk and hull@uchicago.edu. We are grateful to Rodrigo Adão, Gabriel Ahlfeldt, Nate Baum-Snow, Sophie Calder-Wang, Vasco Carvalho, Gabriel Chodorow-Reich, Dave Donaldson, Raffaella Giacomini, Paul Goldsmith-Pinkham, Richard Hornbeck, Kilian Huber, Xavier Jaravel, Tetsuya Kaji, Vishal Kamat, Michal Kolesár, Aureo de Paula, Andrés Rodríguez-Clare, Cyrus Samii, Ben Sommers, Chenzi Xu, and numerous seminar participants for helpful comments. We thank Molly Frean, Jonathan Gruber, and Ben Sommers for sharing code and data. Ruixue Li, Elise Parrish, and Steven Shi provided outstanding research assistance.

1 Introduction

Many questions in economics involve the causal effects of treatments z_ℓ which are computed from multiple sources of variation, according to a known formula. Consider four examples. First, when estimating spillovers from a randomized intervention, a typical z_ℓ counts the number of individual ℓ's neighbors who were selected for the intervention. This treatment combines variation in who was selected and variation in who neighbors whom. Second, in studies of transportation infrastructure effects, a common z_ℓ measures the growth of regional market access: a treatment determined both by the location and timing of transportation upgrades and by the spatial distribution of economic activity in a country. A third example is linear shift-share variables, $z_\ell = \sum_n w_{\ell n} g_n$, which may average a set of national industry shocks g_n with a set of local employment share weights $w_{\ell n}$. Finally, a z_ℓ capturing individual ℓ's eligibility for a public program, such as Medicaid, is jointly determined by the eligibility policy in ℓ's state and her household's demographics and income.[1]

This paper develops new tools for estimating the effects of such composite variables (or using them as instruments for other treatments) when some, but not all, of their determinants are generated by a true or natural experiment. In simpler settings with conventional experimentation, where z_ℓ is itself as-good-as-randomly assigned across observations, causal inference is possible without imposing potentially strong non-experimental restrictions on the unobservable determinants of an outcome, such as a parallel trends assumption. But it is not clear whether and how this useful property of randomization extends to settings where z_ℓ is determined jointly by a set of as-good-as-random "shocks" as well as other predetermined variables governing ℓ's "exposure" to these shocks. For instance, how can the estimation of market access effects leverage a natural experiment in the timing of different transportation upgrades when the other determinants of market access are non-random?

We first show how omitted variable bias (OVB) may confound conventional regression approaches with such z_ℓ. Bias arises from different observations receiving systematically different values of z_ℓ because of their non-random exposure to exogenous shocks. For example, even when transportation upgrades are randomly assigned to different places in the country, regions that are more central in the economic geography are likely to be closer to them and thus may see a larger growth in market access. Identification of market access effects then fails without an additional parallel trends assumption: that these more exposed regions do not differ in their relevant unobservables, such as changes in local productivity or amenities. Intuitively, randomizing transportation upgrades does not randomize the market access growth generated by them.

[1] Characteristic examples of these four settings include Miguel and Kremer (2004), Donaldson and Hornbeck (2016), Autor et al. (2013), and Currie and Gruber (1996a), respectively. We discuss many more examples below.

We then propose a general solution to this OVB challenge, based on the specification of counterfactual shocks that might as well have been realized. This approach views the observed shocks as one realization of some data-generating process—what we call the shock *assignment process*—which can be simulated to obtain counterfactuals. In a true experiment, the shock assignment process is given by the randomization protocol. Otherwise, in natural experiments, shock counterfactuals make explicit the experimental contrasts of interest, for instance by specifying permutations of the shocks that were as likely to have occurred.[2] For example, if the timing of comparable transportation upgrades is considered as-good-as-random, one might produce counterfactual upgrade maps by randomly exchanging the upgrades which happened earlier and later. Policy discontinuities, as commonly used in regression discontinuity designs, can similarly justify local permutations of shocks.

Valid shock counterfactuals can be used to avoid OVB with such z_ℓ, which we generically refer to as instruments, by measuring and appropriately adjusting for a single confounder: the *expected instrument*, μ_ℓ. To do so, a researcher draws counterfactual shocks from the assignment process, recomputes the instrument, and repeats many times. Then, for each observation ℓ, the instrument is averaged across these draws to obtain μ_ℓ. Finally, μ_ℓ is subtracted from z_ℓ to obtain the *recentered instrument* $\tilde{z}_\ell = z_\ell - \mu_\ell$. We show that using $\tilde{z}_\ell$ instead of z_ℓ as an instrument removes the bias from non-random shock exposure. Intuitively, observations only get high vs. low values of $\tilde{z}_\ell$ because of the realization of observed vs. counterfactual shocks, which is assumed to be by chance. For example, when μ_ℓ is constructed by permuting the timing of transportation upgrades, regressions that instrument with $\tilde{z}_\ell$ compare regions which received higher vs. lower market access growth because proximate lines were constructed early vs. late, and not because of the economic geography. Another solution, which leverages the same experimental comparisons, is to include μ_ℓ as a regression control while instrumenting by z_ℓ.[3]

In contrast, more familiar identification strategies—such as instrumenting directly by the exogenous shocks or controlling flexibly for shock exposure—are not appropriate in most settings we consider, where the shocks are assigned at a different "level" than observations and shock exposure is a complex object. In the market access example, upgrade shocks vary at the level of transportation lines, necessitating a mapping from them to regional observations via features of economic geography. This exposure cannot be non-parametrically controlled for, as each region's market access depends on the entire spatial distribution of economic activity. Similarly, controlling for the shares of all industries would remove all

[2] In this sense, shock counterfactuals formalize a natural experiment—what DiNardo (2008) defines as a "serendipitous randomized trial"—in terms of a particular randomization protocol. See Titiunik (2020) for alternative definitions.

[3] Controlling for μ_ℓ can be thought to combine recentering z_ℓ and taking out some residual variation in the outcome. Typically this makes controlling weakly more efficient in large samples, reflecting the precision gains that usually arise from including controls that are orthogonal to the instrument.

variation from a shift-share instrument. Expected instrument adjustment can be viewed as a systematic and transparent way to purge OVB from any mapping, via the appropriate function of exposure μ_ℓ, where conventional controls or fixed effects may fall short.

We next show how the specification of shock counterfactuals can also be used to overcome fundamental challenges with statistical inference. Realizations of z_ℓ are inherently dependent across observations because of their common exposure to the exogenous shocks. Such "exposure clustering" complicates asymptotic approaches to inference, which tend to rely on independence between most pairs of observations (for example, those from different clusters or separated by a geographic or network distance above some threshold, as in Conley (1999)). Our solution adapts principles of randomization inference (RI) via the specified shock counterfactuals. RI-based confidence intervals are exact under constant treatment effects, without any restrictions on the unobservables, and are robust to weak instruments (Imbens and Rosenbaum 2005). RI is particularly attractive for placebo and specification tests, where a constant effect of zero is a natural null.

We complement our framework for identification and finite-sample inference with an analysis of consistency and asymptotic efficiency. Recentered instruments yield consistent estimates and RI tests, regardless of the correlation structure of the unobservables, so long as the observed shocks induce sufficient cross-sectional variation in the instrument and treatment. Our characterization of asymptotically efficient instrument constructions extends the classical analysis of Chamberlain (1987). It involves finding the best predictor of the endogenous variable from the shocks and exposure, recentering it, and then adjusting for the structural residual's heteroskedasticity and dependence on shock exposure. While this instrument is typically infeasible, it can guide the construction of powerful and feasible recentered instruments.[4]

We apply this framework in two settings. First, we show how instrument recentering can help leverage variation in the timing of transportation upgrades and purge OVB when estimating the employment effects of market access (MA) growth due to new Chinese high-speed rail (Zheng and Kahn 2013; Lin 2017). Simple regressions of employment growth on MA growth suggest a large and statistically significant effect, which is only partially reduced by conventional geography-based controls. But this effect is eliminated when we adjust for expected MA growth, measured by permuting constructed HSR lines with similar ones that were planned but not built. The conventional estimates thus reflect the fact that employment grew in regions which were more exposed to high-speed rail upgrades, whether or not construction actually occurred. Importantly, our counterfactual shocks pass RI spec-

[4] Adão et al. (2020) also follow Chamberlain (1987) in characterizing efficient instruments in a setting with interdependence (specifically, in a model of spatial linkages). Our general characterization differs from theirs by allowing for a complex data dependence structure, induced by common shocks, as well as the endogeneity of shock exposure.

ification tests: recentering successfully eliminates the correlation between MA growth and predetermined geographic controls. We discuss how recentering relates to a long literature estimating transportation upgrade effects (e.g. Baum-Snow (2007), Donaldson and Hornbeck (2016), Donaldson (2018), Bartelme (2018), Ahlfeldt and Feddersen (2018), and Tsivanidis (2019)), contrasting the well-known challenge of strategically chosen transportation upgrades (Redding and Turner 2015) with the less discussed problem that regional exposure to exogenous upgrades may be unequal.

Second, we show how our framework helps improve the efficiency of Medicaid eligibility effect estimates when leveraging plausibly exogenous state-level variation in recent Affordable Care Act expansions (Frean et al. 2017; Leung and Mas 2018). A conventional "simulated instrument" approach isolates such variation by averaging over differences in individual exposure to policy shocks, such as family structure and income (e.g., Currie and Gruber (1996a, 1996b), Cohodes et al. (2016), Cullen and Gruber (2000) and Gruber and Saez (2002)). We show how incorporating non-random exposure variation, while appropriately recentering the instrument to isolate the same policy variation, improves the first-stage prediction of eligibility and yields 60–70% smaller standard errors.

We further discuss implications of our framework for other common z_ℓ: network spillover treatments, linear and nonlinear shift-share variables, model-implied instruments, instruments based on centralized school assignment mechanisms, "free-space" instruments for access to mass media, and variables leveraging weather shocks.[5] We provide a general formalization of OVB from non-random exposure in each of these settings, and a general solution, which have previously been given only in some special cases. For example, Borusyak et al. (2020) show how simple controls can address OVB when linear shift-share instruments combine exogenous industry shocks and non-random exposure shares. Relative to their paper, our framework also applies to nonlinear shift-share instruments—a class of more recent empirical strategies where the OVB problem is more challenging. Similarly, in the network setting, Aronow (2012) notes that the random selection of treated units does not imply the randomization of network proximity to them while Aronow and Samii (2017) propose a reweighting solution for when both such shocks and the z_ℓ are discrete (see also Gerber and Green (2012, p. 261)). Our general framework applies to a broader class of network settings by imposing no restrictions on the support of z_ℓ and shocks, with a convenient regression implementation.[6]

[5] Examples include Miguel and Kremer (2004), Acemoglu et al. (2015), Jaravel et al. (2018), and Carvalho et al. (2020) for network spillovers; Boustan et al. (2013), Berman et al. (2015), and Chodorow-Reich and Wieland (2020) for nonlinear shift-share variables; Adão et al. (2020) for model-implied instruments; Abdulkadiroglu et al. (2017, 2019) for school assignment; Olken (2009) and Yanagizawa-Drott (2014) for access to mass media; and Gomez et al. (2007) and Madestam et al. (2013) for weather shocks.

[6] Aronow et al. (2020) distinguish between methods to estimate spillover effects that allow all units to interact while imposing parametric structure (e.g., Manski (2013)) and those with unrestricted interactions

From an econometric perspective, the expected instrument can be seen as a generalization of the propensity score of Rosenbaum and Rubin (1983). Conventional propensity scores are defined in settings with randomly sampled data and a conditionally exogenous binary treatment. Earlier generalizations have considered binary instruments (e.g. Abadie (2003)) and non-binary treatments (e.g. Hirano and Imbens (2004)). Our setting accommodates these extensions but also allows for the kinds of interdependent data that naturally arise when exogenous shocks jointly affect the treatment of many observations.[7] Adjusting for the non-random shock exposure, as captured by the expected instrument, is relevant in such cases even when the shocks are unconditionally exogenous.[8]

Our use of randomization inference builds on a rich statistical literature dating back to Fisher (1935) and reviewed in Lehmann and Romano (2006, Ch. 15). RI was originally proposed for randomized control trials but has also been deployed in a range of non-experimental settings.[9] We apply RI to a broad class of settings where random or as-good-as-random shocks drive some but not all variation in a treatment or instrument, allowing for complex interdependencies across observations.

Broadly, this paper contributes to a growing literature on causal inference that focuses on the assignment process of observed exogenous shocks (e.g. Lee (2008), Athey and Imbens (2018), Shaikh and Toulis (2019), and De Chaisemartin and Behaghel (2018)). Our approach can be understood as combining a statistical model of how such shocks are drawn with an economic model of how the shocks affect an outcome (i.e. through some observed treatment). This approach contrasts with identification strategies that impose a statistical model for the residual determinants of the outcome, such as difference-in-difference strategies (e.g. De Chaisemartin and D'haultfœuille (2020) and Athey et al. (2018)) or fully-specified structural models. Unlike assumptions on the residuals, specifications of the shock assignment process come at no cost with true experiments, may be derived from institutional knowledge with natural experiments, and can be directly tested with any observational data.

among a small number of node pairs (e.g., Hudgens and Halloran (2008)). Like Aronow and Samii (2017), we advance the former approach.

[7]While we focus on regression-based estimators, we show that shock counterfactuals can also be used for inverse-probability weighting (as in Aronow and Samii (2017)) or in the two-step procedure of Hirano and Imbens (2004) (see Doudchenko et al. (2020) for an application of this idea in bipartite network experiments). Regression-based adjustment is more popular in applied research, avoids practical issues with propensity scores close to zero or one, and is natural for structural outcome models with constant treatment effects. With heterogeneous causal effects, recentered instrumental variable regressions identify a convex weighted average under an appropriate monotonicity condition (see Appendix C.1).

[8]Simulation-based recentering is reminiscent of Ellison and Glaeser (1997)'s "dartboard approach" to measuring spatial agglomeration. We correct biased estimates of causal effects, rather than descriptive statistics.

[9]See, e.g., Rosenbaum (1984), Rosenbaum (2002), Bertrand et al. (2004), Imbens and Rosenbaum (2005), Ho and Imai (2006), Abadie et al. (2010), Cattaneo et al. (2015), Dell and Olken (2018), Ganong and Jäger (2018), Canay and Kamat (2018), and Shaikh and Toulis (2019).

The remainder of this paper is organized as follows. The next section motivates our analysis with a stylized example of the OVB and inference challenges in market access regressions. Section 3 develops our general framework and results. Section 4 presents our two applications and discusses other practical implications. Section 5 concludes.

2 A Motivating Experimental Example

We begin with an idealized example that illustrates the key insights of this paper: when exogenous transportation shocks from a randomized control trial (RCT) are used to estimate the local effects of market access growth. Market access (MA) is a statistic which captures the average cost of transportation from a region ℓ to other regions of varying size (the exact formula is unimportant at this point). We consider a linear structural equation relating its growth, $\Delta \log MA_\ell$, to the growth of a regional outcome such as land value, $\Delta \log V_\ell$:

$$\Delta \log V_\ell = \beta \Delta \log MA_\ell + \varepsilon_\ell. \tag{1}$$

Here ε_ℓ captures unobserved shocks to local productivity and amenities occurring in region ℓ between two periods. This equation can be derived from standard models of economic geography (e.g. Redding and Venables (2004)), in which β is a structural elasticity. Equation (1) can also be interpreted as a reduced-form causal model, in which β captures the effect of interventions that affect MA but not the residuals. For these reasons equations like (1), as first proposed by Donaldson and Hornbeck (2016), have become increasingly popular in estimating the regional effects of transportation infrastructure upgrades (e.g. Bartelme (2018) and Tsivanidis (2019)).

We imagine estimating β by leveraging experimental shocks to market access. Specifically, we consider an RCT that changes transportation costs by randomly selecting for construction a set of new roads that connect different regions. We assume that the other determinants of MA are held fixed. New roads affect $\Delta \log MA_\ell$ for all regions (typically even those not directly connected by new roads) to different extents, according to the known market access formula. While we are not aware of actual experimental studies of MA, similar RCTs and natural experiments have been previously analyzed. For example, Gonzalez-Navarro and Quintana-Domeque (2016) study an RCT that paved streets in random neighborhoods across Mexico, while Volpe Martincus and Blyde (2013) exploit random road disruptions in various parts of Chile due to an earthquake.

At first glance, it may seem that the experimental variation in $\Delta \log MA_\ell$ is sufficient to estimate β by a simple linear regression. Since the new roads are selected at random, their construction is guaranteed to be exogenous: i.e., independent from all local productivity and amenity shocks in ε_ℓ. Exogenous transportation shocks are furthermore the only reason that $\Delta \log MA_\ell$ is not identically zero across regions, since market size and other determinants of transportation costs are held fixed in the RCT. When the linear model (1) is correctly specified, this observed variation in $\Delta \log MA_\ell$ fully captures the effects of the transportation shocks on the outcome $\Delta \log V_\ell$.

The first key insight of this paper is that even in this idealized experimental example, non-random exposure to exogenous transportation shocks can generate omitted variable bias in regression estimates of β. Intuitively, randomizing transportation upgrades does not randomize the MA growth generated by them. Even when new roads are placed randomly in space, some regions will tend to see systematically higher MA growth because of their position in the country's economic geography. This tendency can bias regression estimates of MA effects when unobserved productivity and amenity shocks differ systematically in different areas—a scenario allowed by the economic theory underlying equation (1). Formally, $\Delta \log MA_\ell$ and ε_ℓ need not be orthogonal, even though the transportation shocks underlying $\Delta \log MA_\ell$ are independent of ε_ℓ.

To see this OVB problem simply, consider a square island consisting of 64 equally-sized regions ℓ with no initial connectivity, such that initial MA is identical for all regions. Suppose new roads are constructed between regions completely at random: out of all potential roads connecting adjacent regions, the RCT selects half for construction. One such draw from this experiment is shown in Panel A of Figure 1, along with the resulting growth in MA.[10] Expectedly, regions that become connected by road tend to have higher $\Delta \log MA_\ell$. However, the figure reveals another tendency: many of the regions with high MA growth are in the center of the island. This concentration is not by chance. Panel B of Figure 1 shows that the average growth of MA in each region, simulated across 1,000 counterfactual road networks drawn randomly from the same assignment process (i.e. experimental protocol), is also higher in the center of the map. We label this statistic μ_ℓ, and it can be thought of as a region's "expected" MA prior to the realization of exogenous shocks. The spatial pattern of μ_ℓ indicates that more central regions are more exposed to the RCT: no matter where random roads are built, central regions are more likely to be closer to them and thus see a larger market access increase.

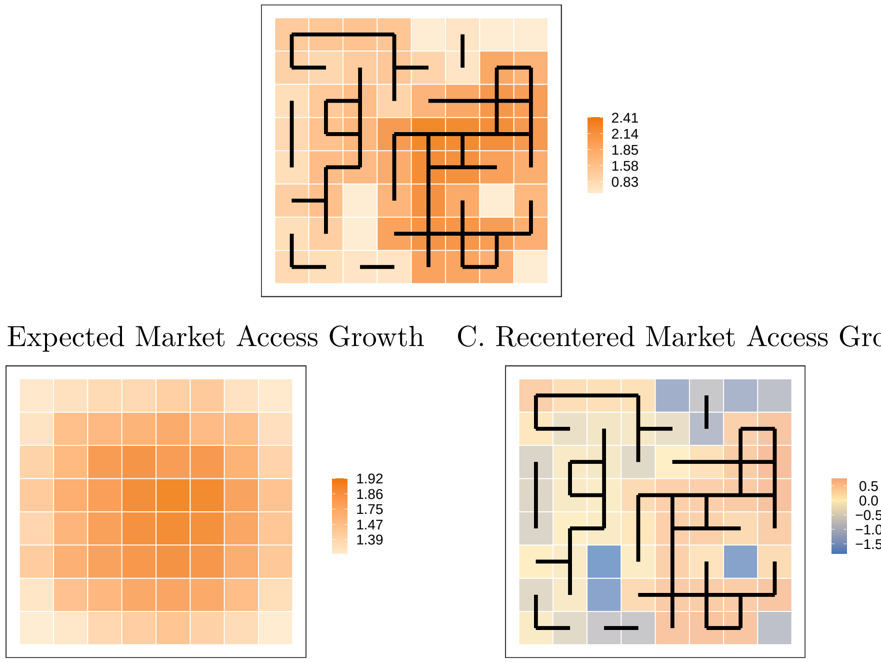
Systematic differences in shock exposure, as captured by μ_ℓ, can generate bias in ordinary least squares (OLS) estimates of β. The OLS estimates come from a comparison of outcome growth between regions with high and low MA growth, which tend to be regions with high and low μ_ℓ. Expected MA growth is predetermined with respect to the experimental shocks but may nevertheless cross-sectionally correlate with the residual (i.e.,

[10]Market access in period $t = 1, 2$ is here given by $MA_{\ell t} = \sum_k \tau_{\ell k t}^{-\theta} P_k$ where $\tau_{\ell k t}$ is a function of distance and connectivity in period t and P_k denotes region k's time-invariant market size (e.g., population). In this simplified example $P_k = 1$ is constant across regions, $\theta = 1$, and $\tau_{\ell k t} = 2^{0.1 d_{\ell k t}}$ where $d_{\ell k t}$ is the distance by road from ℓ to k in period t (or infinity if there is no path).

Figure 1: Market Access Growth in the Motivating Example

A. Line Construction and Market Access Growth

B. Expected Market Access Growth C. Recentered Market Access Growth



Notes: This figure illustrates the OVB problem and our recentering solution in the market access example. Panel A shows a random draw of the railroad construction experiment, with lines indicating connected regions and shading indicating corresponding market access growth (computed as described in the text). Panel B shows average MA growth over 1,000 such random draws. The shading in Panel C indicates the recentered MA measure which subtracts expected MA in Panel B from realized MA in Panel A, with the lines again indicating realized line construction.

be endogenous), biasing the regression comparisons. In the simple example of Figure 1, OVB arises when unobserved productivity and amenity shocks differ between the center and periphery of the map. For example, if rising sea levels reduce amenity values near the edges of the island then central regions will tend to see both higher MA growth and higher residuals, biasing OLS estimates of β upward.

The second insight of this paper is that this identification challenge has an intuitive but non-standard solution, based on the same knowledge of the shock assignment process that generated Panel B of Figure 1. In this experimental setting, one can simulate MA growth across counterfactual draws of the RCT to compute the expected MA growth μ_ℓ of each region ℓ. One can then construct a recentered measure, $\tilde{z}_\ell = \Delta \log MA_\ell - \mu_\ell$, which subtracts each region's expected MA growth from its observed MA growth. This $\tilde{z}_\ell$ is a component of MA growth that can be used as an instrument in equation (1).[11] Intuitively,

[11] Standard models of economic geography used to derive the MA statistic imply a constant elasticity β.

observations only get high vs. low values of $\tilde{z}_\ell$ because of the realization of observed vs. counterfactual shocks, which is by chance. In the example of Figure 1, observed MA growth is no longer concentrated in the center of the map after recentering by μ_ℓ (see Panel C). The same as-good-as-random variation can be leveraged by controlling for μ_ℓ in OLS estimation.

Simulating the road experiment and computing μ_ℓ can be seen as a systematic way to pick the appropriate function of geography that purges bias from non-random exposure. This function is not captured by conventional regression controls except in very special cases, such as in Figure 1 where μ_ℓ simply measures geographic centrality. In general, μ_ℓ depends intricately on the country's economic geography and the road assignment process; Appendix Figures A1 and A2 illustrate the potentially complex nature of μ_ℓ by considering non-uniform regional populations and road construction probabilities, respectively. Controlling for geography perfectly in such settings is of course not possible, as this would remove all variation in market access growth.

The third insight of this paper is that problems with statistical inference on β can also be overcome by simulating counterfactual transportation upgrades. The recentered MA instrument is inherently correlated across regions because of their common exposure to the experimental shocks. Such spatial dependence may generate challenges for the conventional Conley (1999) asymptotic approach to inference, which specifies a geographic distance threshold after which observations of $\tilde{z}_\ell \varepsilon_\ell$ are uncorrelated. For the asymptotic approximation to hold this threshold should be sufficiently small, which may be implausible with all regions exposed to all potential roads.[12] We show in the next section how classical methods of randomization inference (RI) can be applied to address such "exposure clustering."

In most settings, of course, transportation upgrades are not drawn randomly on a map with a known assignment process. In the next section we discuss how shock assignment processes may generally be specified, simulated, and validated in observational data where the exogeneity of shocks is *ex ante* plausible. In Section 4.1 we apply this approach to a specific MA setting and relate it to existing approaches to estimating transportation effects, with or without exogenous upgrades.

3 Identification, Inference, and Asymptotic Efficiency

We now develop a general econometric framework for settings with non-random exposure to exogenous shocks. We introduce the baseline setting, develop our approach to identification

In such cases it does not matter which component of variation is used to estimate β as long as there is no OVB. We show below that the recentering strategy more generally identifies an intuitive convex average of causal effects when they are thought to vary across regions.

[12] Random upgrades to long roads, for example, will tend to cause regions which are far apart in space to "cluster" by their common MA growth. If the unobserved shocks in ε_ℓ also tend to propagate widely then $\tilde{z}_\ell \varepsilon_\ell$ will tend to be correlated across long distances, invalidating spatially clustered standard errors.

based on counterfactual shocks, and discuss how such counterfactuals can be specified in Sections 3.1–3.3. We then show how shock counterfactuals can be used for finite-sample inference, characterize asymptotically most efficient recentered instruments, and summarize several extensions in Sections 3.4–3.6.

3.1 Setting

We suppose an outcome y_ℓ and treatment x_ℓ are observed for units $\ell = 1, \dots, L$. Of interest is a causal effect or structural parameter β, relating treatment to outcomes by

$$y_\ell = \beta x_\ell + \varepsilon_\ell, \tag{2}$$

where ε_ℓ denotes an unobserved residual. Initially we assume y_ℓ and x_ℓ are scalar and demeaned, and that the outcome model is linear with constant effects. We discuss extensions to heterogeneous causal effects, additional control variables, multiple treatments, and non-linear models in Section 3.6. Although we use a single index for observations, we note our framework accommodates repeated cross-sections and panel data where it is also relevant.

Importantly for the generality of our framework, we do not assume that the observations of (y_ℓ, x_ℓ) are independently or identically distributed (*iid*) as when arising from random sampling. This allows for complex dependencies across ℓ due to the common exposure to observed and unobserved shocks. The lack of random sampling is also consistent with settings where the L units represent a population—for example, all regions of a country—and conventional asymptotic frameworks are inappropriate (Abadie et al. 2020).

We suppose that to estimate β a researcher has constructed an instrument z_ℓ which incorporates variation from exogenous shocks, summarized by an $N \times 1$ observed vector g. However the instrument also incorporates additional predetermined variables which govern a unit's exposure to the shocks. Collecting these additional observables in the set w, we write the instrument as

$$z_\ell = f_\ell(g; w), \tag{3}$$

where $\{f_\ell(\cdot)\}_{\ell=1}^{L}$ is a set of known non-stochastic functions. In the previous motivating example, g contained information on transportation network upgrades and w summarized regional populations; the $f_\ell(\cdot)$ functions combined g and w to form market access growth for each region ℓ. As another example, linear shift-share instruments set $f_\ell(g; w) = \sum_{n=1}^{N} w_{\ell n} g_n$ where the $w_{\ell n}$ are non-negative exposure share weights. We note that our framework allows $x_\ell = z_\ell$, in which case β is the reduced-form causal effect of the instrument (as in the motivating example).

Equation (3) is very general, nesting many applied examples (as we discuss in Section 4). Any instrument that can be computed from a set of observed shocks g and other observed variables w can be described in this way.[13] Mapping the shocks into the instrument using some transformation $f_\ell(\cdot; w)$ is generally necessary, for example, when the shocks are defined at a different "level" than the unit of observation (e.g. industry shocks and regional data) or when shocks to one observation have spillover effects on others. In some cases, such as the market access and linear shift-share examples, the instrument specification may follow from a particular model for the treatment variable. For example, when $x_\ell = \tilde{f}_\ell(g, w, u)$ for a known $\tilde{f}_\ell(\cdot)$ and some (possibly unobserved) endogenous shocks u, an instrument may be specified as the treatment prediction that shuts down these shocks: $f_\ell(g, w) = \tilde{f}_\ell(g, w, 0)$. For now we take the choice of $f_\ell(\cdot)$ as given, addressing the question of which instrument constructions may be more desirable in Section 3.5.

Partitioning the determinants of z_ℓ into a set of shocks g and other variables w allows us to formalize the notion that some but not all sources of variation in the instrument are exogenous. In an RCT the exogeneity of shocks can naturally arise from the experimental intervention. With observational data, a researcher may appeal to an experimental ideal in which the shocks in g are as-good-as-randomly assigned given predetermined variables in w, which are not exogenous. For example, in shift-share designs it may be plausible that the industry-level shocks in g arise from a natural experiment but that local industrial composition w is endogenous (Borusyak et al. 2020).

We formalize shock exogeneity by the conditional independence of g from the residual vector $\varepsilon = \{\varepsilon_\ell\}_{\ell=1}^{L}$, given the other sources of instrument variation:

Assumption 1. *(Shock exogeneity):* $g \perp\!\!\!\perp \varepsilon \mid w$

This notion of shock exogeneity combines two conceptually distinct conditions. First, it imposes an exclusion restriction: that the realization of shocks only affects the outcome of each unit via its treatment x_ℓ and not through ε_ℓ. This condition may be violated when the structural equation (2) is misspecified; for example, when market access inadequately captures the local economic effects of new transportation.[14] Second, Assumption 1 requires the as-good-as-random assignment of shocks with respect to the unobserved outcome determinants ε. This condition is satisfied when the shocks are fully randomly assigned, as in an RCT: i.e., $g \perp\!\!\!\perp (\varepsilon, w)$. More generally, Assumption 1 allows w to contain variables that govern the shock assignment process. We discuss how such conditioning is useful for specifying shock counterfactuals in Section 3.3. The exclusion and as-good-as-random assignment assumptions are isolated in Appendix C.1, via a general potential outcomes model.

[13] Note that equation (3) does not contain a residual: it formalizes an algorithm for computing an instrument rather than characterizing an economic relationship.

[14] The shock exclusion restriction may follow from a particular economic model, as in Donaldson and Hornbeck (2016), or be relaxed by including multiple treatments in x_ℓ (e.g. allowing for both direct and spillover effects of the same shocks, as in Miguel and Kremer (2004)).

We consider identification of β from an instrumental variable (IV) regression of y_ℓ on x_ℓ, with z_ℓ as an instrument (with OLS obtained as a special case when $x_\ell = z_\ell$). Identification follows when z_ℓ is relevant to the treatment and orthogonal to the structural residual. In our *non-iid* setting, we formalize these conditions as $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right] \neq 0$ and

$$\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell y_\ell\right] - \beta \cdot \mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right] = 0, \tag{4}$$

which together imply that β is uniquely recoverable from the observable moments $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell y_\ell\right]$ and $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right]$. Here it is worth highlighting that full-data instrument orthogonality (4) combines two dimensions of variation: over the stochastic realizations of g, w, and ε, and across the cross-section of observations $\ell = 1, \dots, L$. In the *iid* case it reduces to the more familiar condition $\mathbb{E}[z_\ell \varepsilon_\ell] = 0$.

While our primary focus is on identification and finite-sample inference, some of our results consider the asymptotic properties of the IV estimator:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell z_\ell y_\ell}{\frac{1}{L}\sum_\ell z_\ell x_\ell}, \tag{5}$$

which is the solution to the sample analog of (4). We establish asymptotic properties by considering a sequence of data-generating processes, indexed by L, for the complete data (y, x, g, w). Consistency, for example, is defined as $\hat{\beta} \xrightarrow{p} \beta$ as $L \to \infty$, while asymptotic efficiency considers large-L approximations to the variance of $\hat{\beta}$. We emphasize that this asymptotic sequence should be viewed as a way to approximate the finite-sample distribution of the IV estimators, rather than as a description of the sampling process for the data.[15]

3.2 Identification and Instrument Recentering

Our first result formalizes the omitted variable bias problem: exogeneity of the shocks underlying z_ℓ is not generally enough for identification of β, even when they are fully randomly assigned. We then derive a simple but non-standard recentering of z_ℓ that purges OVB in this setting. We conclude this subsection with results on recentered IV consistency.

Identification under Assumption 1 fails when predetermined exposure to the natural experiment is endogenous. While this exposure variation is potentially high-dimensional, our first result shows that OVB is governed by a particular one-dimensional confounder: the expected instrument, μ_ℓ.

[15] This is similar to how Bekker (1994) studies IV regressions with many instruments. As he writes, "the [asymptotic] sequence is designed to make the asymptotic distribution fit the finite sample distribution better. It is completely irrelevant whether or not further sampling will lead to samples conforming to this sequence" (p. 658).

Lemma 1. *Under Assumption 1,*

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell} z_\ell \varepsilon_\ell\right] = \mathbb{E}\left[\frac{1}{L}\sum_{\ell} \mu_\ell \varepsilon_\ell\right], \tag{6}$$

where $\mu_\ell = \mathbb{E}\left[f_\ell(g;w) \mid w\right]$. Thus β is not identified by the instrument z_ℓ when μ_ℓ is endogenous, in the sense of $\mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \varepsilon_\ell\right] \neq 0$.

Proof. $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[f_\ell(g;w)\varepsilon_\ell \mid w\right]\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \mathbb{E}\left[\varepsilon_\ell \mid w\right]\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \varepsilon_\ell\right]$. The first and third equality follow from the law of iterated expectations, while the second equality follows by Assumption 1 and the definition of μ_ℓ. □

The expected instrument is defined as the average value of z_ℓ across different realizations of the shocks conditional on w. Lemma 1 shows that the exogeneity of shocks makes z_ℓ a valid instrument if and only if this μ_ℓ is orthogonal to the residual ε_ℓ. Absent further assumptions, adjustment for μ_ℓ is thus generally necessary to remove OVB. Note that adjustment is generally necessary even if the shocks are unconditionally as-good-as-randomly assigned, i.e. when $g \perp\!\!\!\perp (\varepsilon, w)$ in Assumption 1.

When shock exposure is endogenous but Assumption 1 holds, Lemma 1 suggests a simple but non-standard recentering of z_ℓ that identifies β. In fact, a weaker notion of shock exogeneity suffices.

Assumption 2. *(Weak shock exogeneity):*
(i) $\mathbb{E}\left[\varepsilon_\ell \mid g, w\right] = \mathbb{E}\left[\varepsilon_\ell \mid w\right]$ *almost surely for each ℓ.*
(ii) $\mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid g, w\right] = \mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid w\right]$ *almost surely for each ℓ and m.*

Such mean and covariance independence of the residuals from the shocks is implied by Assumption 1 and will also be sufficient for some of our later asymptotic results. Here we use the first condition to show that β is identified by a recentered instrument $\tilde{z}_\ell$, given a non-zero first-stage:

Proposition 1. *Suppose Assumption 2(i) holds and let $\tilde{z}_\ell = z_\ell - \mu_\ell$. Then*

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell} \tilde{z}_\ell \varepsilon_\ell\right] = 0, \tag{7}$$

such that β is identified by the instrument $\tilde{z}_\ell$ provided $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right] \neq 0$.

Proof. See Appendix B.1. □

A recentered IV regression compares units with a higher-than-expected value of z_ℓ, because of the realization of the shocks, to units affected less than expected. The validity of $\tilde{z}_\ell$ thus stems from the exogeneity of shocks (specifically, Assumption 2(i)), even though it continues to vary cross-sectionally due to heterogeneous shock exposure. First-stage relevance holds when the units with higher-than-expected values of z_ℓ have systematically different values of the treatment x_ℓ.[16]

A closely related regression-based solution to OVB is also implied by Lemma 1: including the expected instrument μ_ℓ as a control while using the original z_ℓ as an instrument. This regression yields the reduced-form and first-stage moments $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell y_\ell^\perp\right]$ and $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell^\perp\right]$, where $v_\ell^\perp$ denotes the residuals from a cross-sectional projection of v_ℓ on μ_ℓ. Appendix B.1 shows that these moments also identify β under Assumption 2(i). This result clarifies the role of conventional controls and fixed effects in purging OVB under our assumptions: shock exogeneity is sufficient to identify β without recentering z_ℓ or restricting unobservables only when the included controls absorb μ_ℓ.[17]

Given identification of β, one may be interested in consistency of the recentered IV estimator which instruments x_ℓ with $\tilde{z}_\ell$. Establishing consistency with our general asymptotic sequence is non-trivial, as we cannot rely on conventional sampling-based arguments for *iid* data. Instead, Proposition S2 in Appendix C.2 shows how consistency is achieved given an asymptotic first stage and weak mutual cross-sectional dependence of $\tilde{z}_\ell$. In line with our general approach, we make no restriction on the mutual dependence of residuals, imposing only a weak regularity condition on ε_ℓ. The substantive assumption on $\tilde{z}_\ell$ requires the recentered instrument construction to well-differentiate observations by their exposure to the exogenous shocks, yielding a law of large numbers that brings $\hat{\beta}$ close to β for large L. Lower-level conditions sufficient for this assumption are also given in Appendix C.2.

3.3 Specifying Shock Counterfactuals

Our solution to the OVB challenge involves measuring the expected instrument, which typically requires specifying counterfactual shocks that may well have occurred. Here we formalize this specification of counterfactual shocks and discuss general ways in which it may be accomplished. In Section 4 we discuss and illustrate specific approaches in the context of various applied settings.

Formally, we denote the shock assignment process by the conditional distribution of

[16] Whenever the shocks induce some variation in treatment, there exist $f_\ell(\cdot)$ constructions such that the corresponding recentered instrument satisfies the relevance condition. Formally, when $\text{Var}\left[\mathbb{E}\left[x_\ell \mid g, w\right] \mid w\right]$ is not almost-surely zero at least for some ℓ, the recentered instrument constructed as $\tilde{z}_\ell = \mathbb{E}\left[x_\ell \mid g, w\right] - \mathbb{E}\left[x_\ell \mid w\right]$ is relevant.

[17] In panel data with $z_{\ell t} = f_{\ell t}(g_t, w_t)$, for example, unit fixed effects generally purge OVB only when the expected instrument is time-invariant, which generally requires the $f_{\ell t}(\cdot)$ mapping, the value of w_t, and the distribution of g_t to be time-invariant. While plausible in some applications, these conditions (in particular, stationarity of the shock distribution) are quite restrictive. For instance, when roads tend to be built more than destroyed expected market access will tend to grow over time.

$g \mid w$, which we write as $G(g \mid w)$. When $G(\cdot)$ is known, the expected instrument $\mu_\ell = \int f_\ell(\gamma; w) dG(\gamma \mid w)$ can be computed and used to purge OVB. To emphasize such knowledge, we state it as an assumption:

Assumption 3. *(Known assignment process): $G(g \mid w)$ is known in the support of w.*

Specification of $G(\cdot)$ is most straightforward when the shocks are actually determined by a known randomization protocol, as in an RCT. Literal randomization of g given w implies both the exogeneity of shocks (i.e. Assumption 1, given shock exclusion) and Assumption 3. Policy discontinuities (as in regression discontinuity designs) also fit in this case, when viewed as generating a local RCT around the cutoffs (Lee 2008; Cattaneo et al. 2015).[18]

When randomization of shocks occurs naturally, scientific or institutional knowledge may yield $G(\cdot)$. For instance, when the locations of earthquake disruptions are viewed as exogenous shocks (e.g. Volpe Martincus and Blyde (2013) and Carvalho et al. (2020)), the probability distribution of counterfactual locations can be given by geological knowledge. Similarly, appropriate historical weather data may serve as counterfactuals for observed weather shocks (see Appendix D.8).

In observational data, specifying $G(g \mid w)$ makes explicit the features of shocks which are considered as-good-as-random (e.g. the placement vs. timing of transportation upgrades) and the corresponding experimental contrasts. For instance, the researcher may be willing to specify permutations of the g vector that were as likely to have occurred. To see how this satisfies Assumption 3, suppose that all permutations of g are equally likely to arise, as when the shocks g_n are *iid* across n. In this case $G(g \mid w)$ is known to be uniform when w is augmented by the permutation class $\Pi(g) = \{\pi(g) \mid \pi(\cdot) \in \Pi_N\}$, where Π_N denotes the set of permutation operators $\pi(\cdot)$ on vectors of length N (e.g. Lehmann and Romano 2006, p. 634). The marginal distribution of g_n (conditionally on other components of w) then need not be specified; the expected instrument is the average z_ℓ across all permutations of shocks, which serve as counterfactuals:

$$\mu_\ell = \frac{1}{N!} \sum_{\pi(\cdot) \in \Pi_N} f_\ell(\pi(g); w). \tag{8}$$

Such μ_ℓ are easy to compute (or approximate with a random set of permutations, when N is large).[19] This scenario highlights the potentially dual role of w: as a means of satisfying

[18] Assumption 3 requires specification of $G(\cdot \mid w)$ for all possible w. However, it is without loss to view w as a fixed object (i.e. part of $\{f_\ell(\cdot)\}_{\ell=1}^L$), in which case this is not restrictive. We allow w to be stochastic only for full generality and to make non-random exposure more explicit. With w viewed as stochastic, the support condition of Assumption 3 is still not restrictive when g arise from an RCT or satisfy conditional exchangeability, as discussed below.

[19] Approximating μ_ℓ is sufficient for identification because the recentered IV still identifies β in this case: i.e. $\mathbb{E}\left[\frac{1}{L}\sum_\ell \left(f_\ell(g, w) - f_\ell(\pi(g), w)\right)\varepsilon_\ell\right] = 0$ under Assumption 2(i), for any fixed or randomly drawn $\pi(\cdot)$.

exogeneity (Assumption 1) and as a way to simplify the specification of shock counterfactuals (Assumption 3).

Similar expected instrument calculations follow under weaker shock exchangeability conditions. When the g_n are *iid* within, but not across, a set of known clusters, Assumption 3 is satisfied when the class of within-cluster permutations is conditioned on and used to draw counterfactuals. Other symmetries in the joint shock distribution can also be used to construct valid counterfactuals, as we illustrate in a shift-share setting in Appendix D.4.

We emphasize that expected instrument adjustment generally requires some outside knowledge of $G\left(g \mid w\right)$, since μ_ℓ is typically not non-parametrically identified with non-*iid* data.[20] Nevertheless, as discussed below, our framework can apply with $G\left(g \mid w\right)$ specified up to a low-dimensional vector of unknown parameters—allowing, for example, parameterized heteroskedasticity of otherwise exchangeable shocks. We further note that in observational data it is imperative to corroborate an *ex ante* argument for Assumptions 1 and 3 by empirical tests. The next section shows that these assumptions yield testable implications and a natural testing procedure. Finally, we note that even incorrect specification of the shock assignment process may be useful as a robustness check: if Assumption 1 holds and there is already no OVB because the included controls perfectly capture either the endogenous features of exposure or the expected instrument, then controlling for any misspecified expected instrument $m_\ell(w)$ cannot change the estimand.[21]

3.4 Randomization Inference and Testing

Specification of the shock assignment process can be used to construct valid statistical tests and confidence intervals for β, following a long tradition of randomization inference (Fisher 1935). Under constant effects the RI approach guarantees correct coverage in finite samples, of both observations and shocks, even when the observations exhibit complex dependencies.[22] We focus on a particular type of RI tests which is tightly linked to the recentered IV estimator and which is expected to have favorable large-sample power. We then discuss how RI can be used to validate Assumptions 1 and 3, through exact falsification and specification tests.

[20] This is in contrast to conventional propensity score calculations with *iid* data. To see the difference, suppose $z_\ell = f(g, w_\ell)$, where $f(\cdot)$ is common across ℓ and w_ℓ is observation-specific, low-dimensional, and *iid*. Then there is no need to specify $G\left(g \mid w\right)$ explicitly: $\mu_\ell = \tilde{\mu}\left(w_\ell\right)$ is a common function of w_ℓ which can be flexibly estimated from observations of (z_ℓ, w_ℓ). This scenario, however, does not fit the majority of interesting cases of our setup.

[21] Formally, suppose either $\mathbb{E}\left[\check{z}_\ell \mid w\right] = 0$ or $\mathbb{E}\left[\check{\varepsilon}_\ell \mid w\right] = 0$ for each ℓ, where $\check{v}_\ell$ denotes the cross-sectional residualization of variable v_ℓ on some functions of w used as controls. Then $\mathbb{E}\left[\frac{1}{L}\sum_\ell \check{z}_\ell^\perp \check{\varepsilon}_\ell^\perp\right] = 0$, where here $v_\ell^\perp$ denotes the residuals from a cross-sectional projection of v_ℓ on $m_\ell(w)$. See Appendix C.6 for our framework extended to predetermined controls.

[22] Valid inference with heterogeneous effects and interdependent data is a difficult challenge, even in a more standard asymptotic approach (Adão et al. 2019).

In general, RI tests and confidence intervals for β are based on a scalar test statistic $T = \mathcal{T}(g, y - bx, w)$, where b is a candidate parameter value. Under the null hypothesis of $\beta = b$ and Assumption 1, the distribution of $T = \mathcal{T}(g, \varepsilon, w)$ conditional on ε and w is implied by the shock assignment process $G(g \mid w)$. One may simulate this distribution, by redrawing (e.g., permuting) the shocks in g and recomputing T. If the original value of T is far in the tails of the simulated distribution, one has grounds to reject the null that $\beta = b$. Appendix C.3 formalizes this logic and explains how inversion of such tests yields confidence interval for β by collecting all b that are not rejected. These intervals have correct size, both conditionally on (ε, w) and unconditionally. Valid RI confidence intervals can be obtained for any test statistic, although the the choice of $\mathcal{T}$ generally affects the power against alternative hypotheses.[23]

We address the practical issue of choosing a powerful randomization test statistic, and draw a tight link between $\mathcal{T}$ and the recentered IV estimator $\hat{\beta}$, by building on the theory of Hodges and Lehmann (1963). Specifically, we consider a $\mathcal{T}(g, y - bx, w)$ which $\hat{\beta}$ rationalizes as being typical under the null, in the following sense:

Lemma 2. *Let $T = \mathcal{T}(g, y - bx, w)$ and $T^* = \mathcal{T}(g^*, y - bx, w)$, where g^* is distributed according to $G(\cdot \mid w)$, independently of (g, x, y), conditionally of w. Define the Hodges-Lehmann estimator as the $b \in \mathbb{R}$ that solve $T = \mathbb{E}[T^* \mid y, x, w]$. Then the recentered IV estimator is the Hodges-Lehmann estimator associated with $T = \frac{1}{L}\sum_\ell (f_\ell(g, w) - \mu_\ell)(y_\ell - bx_\ell)$.*

Proof. See Appendix B.2. □

This result shows that the recentered IV estimator of β equates the sample covariance between the recentered instrument $\tilde{z}_\ell$ and implied residual $y_\ell - bx_\ell$ with the expectation of its randomization distribution (specifically, zero), satisfying our definition of a Hodges-Lehmann estimator.[24] Notably, the same randomization tests, confidence intervals, and Hodges-Lehmann estimators are obtained from the statistic based on the non-recentered instrument, $\frac{1}{L}\sum_\ell f_\ell(g, w)(y_\ell - bx_\ell)$.[25] In this sense, the RI approach performs the recentering needed for identification of β automatically.

Statistics chosen on the basis of Hodges-Lehmann estimators can inherit their power properties. While we are not aware of existing general results, Proposition S2 in Appendix C.2 shows that randomization tests of Lemma 2 are generally consistent, in the sense of

[23] There are no general results on the relative power of different RI statistics, But good power properties have been established in some special contexts (Lehmann and Romano 2006, Section 15.2.2).

[24] This definition follows Rosenbaum (2002) and Imbens and Rosenbaum (2005). The original definition in Hodges and Lehmann (1963) is the value of β that maximizes the p-value of the randomization test. For two-sided confidence interval this means equating T to its median, rather than its mean.

[25] This follows because recentering shifts both T and T^* by the same value, $\frac{1}{L}\sum_\ell \mu_\ell (y_\ell - bx_\ell)$, which does not depend on g. Appendix B.2 further shows that the μ_ℓ-controlled IV estimator is the Hodges-Lehmann estimator corresponding to the residualized covariance statistic $\frac{1}{L}\sum_\ell z_\ell \left(y_\ell^\perp - bx_\ell^\perp\right)$.

having power that asymptotically increases to one for any fixed alternative, under the conditions which make the recentered IV estimator consistent. This asymptotic result reinforces the tight connection between $\mathcal{T}$ and $\hat{\beta}$.[26] We note, however, that as in other settings (e.g. Abadie et al. 2010; Mackinnon and Webb 2020) the finite-sample validity of RI may be most useful when the conditions for consistency are *not* met, such as when there are few shocks with concentrated exposure. We discuss an example of such a setting in Appendix D.3 and illustrate good power of RI for shift-share instruments with few shocks in Appendix D.4.

Randomization inference can also be used to perform falsification tests on our key Assumptions 1 and 3. Recentering implies a testable prediction that $\tilde{z}_\ell$ is orthogonal to any variable r_ℓ satisfying $g \perp\!\!\!\perp r \mid w$, which holds for $r = \{r_\ell\}_{\ell=1}^L$ that are either functions of w or some other observables thought to be determined prior to (or independent of) the shocks g. To test this restriction, one may check that the sample covariance $\frac{1}{L}\sum_\ell \tilde{z}_\ell r_\ell$ is sufficiently close to zero by re-randomizing shocks and checking that T is not in the tails of its conditional-on-(w, r) distribution. Multiple falsification tests, based on a vector of predetermined variables R_ℓ, can be combined by an appropriate RI procedure, e.g. by taking T to be the sample sum of squared fitted values from regressing $\tilde{z}_\ell$ on R_ℓ.[27]

Falsification tests are useful in two ways. First, when r_ℓ is a lagged outcome or another variable thought to proxy for ε_ℓ, they provide an RI implementation of conventional placebo and covariate balance tests of Assumption 1. While the use of RI for inference on causal effects may be complicated by treatment effect heterogeneity, the sharp hypothesis of zero placebo effects is a natural null. Second, RI tests will generally have power to reject false specifications of the shock assignment process, i.e. violations of Assumption 3, even when r_ℓ does not proxy for ε_ℓ. For $r_\ell = 1$, for example (which is trivially conditionally independent of g), the test verifies that the sample mean of z_ℓ is typical for the realizations of the specified assignment process. Setting $r_\ell = \mu_\ell$ instead checks that the recentered instrument is not correlated with the expected instrument that it is supposed to remove.

3.5 Asymptotic Efficiency

While any instrument $f_\ell(g, w)$ can be made valid by appropriate recentering and used for valid randomization inference, the choice of instrument construction from the set of possible $\{f_\ell(\cdot)\}_{\ell=1}^L$ will generally matter for power. Proposition 2 in Appendix B.3 shows that the

[26] One might instead consider computing confidence intervals from the distribution of the recentered estimator itself with re-randomized shocks g^*. This idea fails in IV since the re-randomized instrument $f_\ell(g^*, w) - \mu_\ell$ has a true first-stage of zero. The distribution of reduced-form coefficients across re-randomized shocks is also not useful, except for testing $\beta = 0$, as that distribution is centered around zero rather than β.

[27] Formally, this $T = \tilde{z}' R \left(R'R\right)^{-1} R'\tilde{z}$ can be seen as a quadratic form of the vector-valued statistic $\frac{1}{L}\sum_\ell \tilde{z}_\ell R_\ell$, weighted by $\left(R'R\right)^{-1}$, where R is the matrix collecting R_ℓ and $\tilde{z}$ is the vector collecting $\tilde{z}_\ell$.

following instrument minimizes the asymptotic variance of recentered IV, under appropriate regularity conditions:

$$z^* = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1}\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right). \tag{9}$$

This characterization extends the classic result of Chamberlain (1987) to our setting in showing how exogenous shocks can be efficiently leveraged. Appendix C.4 further establishes that this z^* maximizes the local power of RI-based tests.

Constructing such optimal instruments may not be feasible in practice, and typically requires an economic model for both the dependence of treatment on shocks and the endogeneity of exposure: $\mathbb{E}\left[x \mid g, w\right]]$ and $\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]$, respectively. Our characterization nevertheless provides guidance for constructing recentered instruments, by showing what researchers may strive for when choosing between alternative IV estimators.

To build intuition for the optimal instrument, we establish the following Lemma:

Lemma 3. *Let* $\tilde{z} = \mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]$, $\psi = \mathbb{E}\left[\varepsilon \mid w\right]$, *and* $\Omega = \operatorname{Var}\left[\varepsilon \mid w\right]$. *Then*

$$z^* = \Omega^{-1}\left(\tilde{z} - \nu\rho\psi\right), \tag{10}$$

where $\rho\psi = \frac{\psi'\Omega^{-1}\tilde{z}}{\psi'\Omega^{-1}\psi}\psi$ *is the* Ω^{-1}*-weighted projection of* $\tilde{z}$ *on* ψ *and* $\nu = \frac{\psi'\Omega^{-1}\psi}{1+\psi'\Omega^{-1}\psi}$.

Proof. See Appendix B.3. □

Equation (10) permits an intuitive four-step description of the optimal instrument. First, one takes the best predictor of treatment given by the shocks and predetermined variables, $\mathbb{E}\left[x \mid g, w\right]$. Second, one recenters this predictor by $\mathbb{E}\left[x \mid w\right]$ to remove the potential OVB from non-random shock exposure, obtaining $\tilde{z}$. Third, one partially residualizes the recentered instrument on the predictable component of the residual, ψ.[28] Finally, one adjusts for the residual variance Ω, as in generalized least squares. While steps 1 and 4 follow the optimal instrument construction in Chamberlain (1987), steps 2 and 3 are new, stemming from the potential endogeneity of w.

Predicting treatment from shocks and exposure (step 1) is trivial when x_ℓ is a function of (g, w), since then $\mathbb{E}\left[x_\ell \mid g, w\right] = x_\ell$. Otherwise, powerful z_ℓ may be given by an economic

[28] This residualization is partial (i.e. $\nu \in [0, 1)$) for the same reason as why, in the conventional panel data context, the random effects estimator demeans the data within each unit only partially (e.g. Wooldridge 2002, p. 286). As with unit-specific means in the panel setting, ψ_ℓ is orthogonal to $\tilde{z}_\ell$ in expectation and so provides an additional moment for identifying β. We also note that if ψ is completely known, a more efficient but less robust instrument than (9) is available, which replaces y with $y - \psi$ and ε with $\varepsilon - \psi$ (without adjusting x) and uses the original z. Since $\mathbb{E}\left[\varepsilon - \psi \mid w\right] = 0$, instrument recentering that isolates variation in g but reduces power is unnecessary. However, this efficiency gain is obtained at the cost of losing robustness to misspecification of the residual model.

model for treatment.[29] Specifically, when $x_\ell = \tilde{f}_\ell(g, w, u)$ for a known $\tilde{f}_\ell(\cdot)$ and a set of unobserved shocks u, a reasonable stand-in for $\mathbb{E}[x_\ell \mid g, w]$ may be obtained by $\tilde{f}_\ell(g, w, 0)$; that is, a treatment prediction which shuts down the role of unobserved shocks. This approach has been taken, for example, by Bartelme (2018) in the market access setting (see also Berry et al. (1999) for the same idea in an entirely different context). Instrument recentering is then generally necessary to isolate exogenous variation in shocks (step 2).

The third and fourth steps in Lemma 3 may be more difficult to implement as they require models of unobservables rather than the observed treatment. Practically, Step 3 calls to control for predetermined variables which may be correlated with the residual, as including these controls may approximate the projection of $\tilde{z}$ on ψ. By the logic of Proposition 1 such controls are orthogonal to $\tilde{z}$ in expectation and will not weaken the first stage, but their inclusion will generally improve efficiency by reducing residual variance. Step 4 is a more standard correction for heteroskedasticity and mutual correlation of residuals. We expect that performing the more feasible steps 1 and 2 alone will typically improve power, although there is no guarantee (see Appendix A.3 for a counterexample discussed in the context of the application in Section 4.2).

3.6 Extensions

Appendices C.1 and C.5–C.8 extend our basic identification and inference results in several ways. Appendix C.1 first shows that in presence of treatment effect heterogeneity the recentered IV estimator captures a convexly weighted average of causal effects under an appropriate monotonicity condition, extending the classic result of Imbens and Angrist (1994) to this general setting. For example, in reduced-form models of the form $y_\ell = \beta_\ell z_\ell + \varepsilon_\ell$ the heterogeneous effects β_ℓ are weighted by the conditional variance of $\tilde{z}_\ell \mid w$ across counterfactual shocks. This appendix further shows how a particular rescaling of the recentered instrument—with a factor given by the shock assignment process—can identify local average treatment effects in the traditional setting of a binary treatment and instrument, and how the approach of Hirano and Imbens (2004) can also be adapted.

Appendix C.5 shows how recentered IVs can be constructed, and RI applied, when the shock assignment process is only partially specified. We allow for a vector of unknown parameters of $G(\cdot)$ which may govern, for example, how shocks vary systematically with observables. Appendix C.6 shows how predetermined observables can be included as regression controls to reduce residual variation and potentially increase power. Appendix C.7 discusses identification and inference with multiple treatments or instruments. Finally, Appendix C.8 extends the framework to nonlinear outcome models.

[29] Obtaining $\mathbb{E}[x_\ell \mid g, w]$ without a treatment model is challenging in our general non-*iid* setup, in contrast to other settings where the first stage can be non-parametrically estimated (e.g. Newey (1990)).

4 Practical Implications and Applications

We now present two empirical applications showing how our theoretic framework can be used to avoid OVB and improve efficiency in practice. Specifically, we estimate the market access effects of Chinese high-speed rail and the insurance coverage effects of Medicaid expansions. In both applications we contrast recentered IV estimation with existing approaches. We conclude this section by summarizing practical implications for other empirical settings.

4.1 Effects of Transportation Infrastructure

We first apply our framework to estimate the effect of market access growth on Chinese regional employment growth over 2007–2016, leveraging the recent construction of high-speed rail (HSR). We show how counterfactual HSR shocks can be specified, and how correcting for expected market access growth can help purge OVB. We then discuss how our approach to estimating transportation infrastructure effects relates to existing methods.

The recent construction of Chinese HSR has produced a network longer than in all other countries combined (Lawrence et al. 2019). The network mostly consists of dedicated passenger lines and has developed rapidly since 2007. Construction was started by the Medium- and Long-Term Railway Plan in 2004; this plan was later expanded in 2008, as part of the stimulus package during the financial crisis, and again in 2016. Construction objectives included freeing up capacity on the low-speed rail network and supporting economic development by improving regional connectivity (Lawrence et al. 2019; Ma 2011). While affordable fares make HSR popular for different purposes, business travel is an important component of rail traffic, ranging between 28% and 62%, depending on the line (Ollivier et al. 2014; Lawrence et al. 2019). The role of HSR may also extend beyond directly connected regions, as passengers frequently transfer between HSR and traditional lines (and between intersecting HSR lines). An early analysis by Zheng and Kahn (2013) finds positive effects of HSR on housing prices, while Lin (2017) similarly finds positive effects on regional employment.

We analyze HSR-induced market access effects for 340 sub-province-level administrative divisions in mainland China. We follow Potlogea and Cheng (2017) in referring to these units as prefectures: although most are officially called "prefecture-level cities," they typically include multiple urban areas. We measure market access in 2007 and 2016 by combining data on the development of the HSR network and each prefecture's location and population (as measured in the 2000 census). A total of 83 HSR lines opened between these years, with the first in 2008; a further 66 lines (which we refer to as "planned") were completed or under construction as of April 2019.[30] We compute a simple market access measure in

[30] We define a line by a contiguous set of inter-prefecture HSR links that were proposed together and

each prefecture ℓ and year t based on the formula in Zheng and Kahn (2013): $MA_{\ell t} = \sum_k \exp(-0.02\tau_{\ell kt}) \cdot P_{k,2000}$. The summation is over all prefectures (including $k = \ell$), $P_{k,2000}$ denotes the predetermined population of prefecture k, and $\tau_{\ell kt}$ denotes predicted travel time between regions ℓ and k in year t (in minutes). Travel time predictions are based on the operational speed of each HSR line as well as geographic distance, which proxies for the travel time by car or a low-speed train. We relate MA growth, $z_\ell = \log MA_{\ell,2016} - \log MA_{\ell,2007}$, to the corresponding growth in prefecture's urban employment y_ℓ from the Chinese City Statistical Yearbooks. This yields a set of 274 prefectures with non-missing outcome data; see Appendix A.1 for details on the sample construction and market access measure. Panel A of Figure 2 shows the Chinese HSR network as of the end of 2016, along with the implied growth of market access relative to 2007.

Column 1 of Table 1, Panel A, reports the coefficient from a simple regression of employment growth on MA growth; Appendix Figure A3 visualizes this relationship.[31] The estimated elasticity of 0.23 is large. Given an average MA growth of 0.54 log points, it implies a 12.4% employment growth attributable to the HSR for an average prefecture—almost half of the 26.6% average employment growth over this period. The estimate is also highly statistically significant using the spatially-clustered standard errors of Conley (1999), echoing the findings of Lin (2017) (while not being directly comparable due to our use of later years and a different specification).

Panel A of Figure 2, however, gives immediate reason for caution against interpreting the OLS coefficient as causal. Prefectures with high MA growth, which serve as the effective treatment group, tend to be clustered in the main economic areas in the southeast of the country where HSR lines are concentrated. Areas near major cities, such as Shanghai and Beijing, also tend to see high MA growth as they are connected by the HSR network. A comparison between these prefectures and the economic periphery may be confounded by the effects of unobserved policies, both contemporaneous and historical, that differentially affected the economic center.

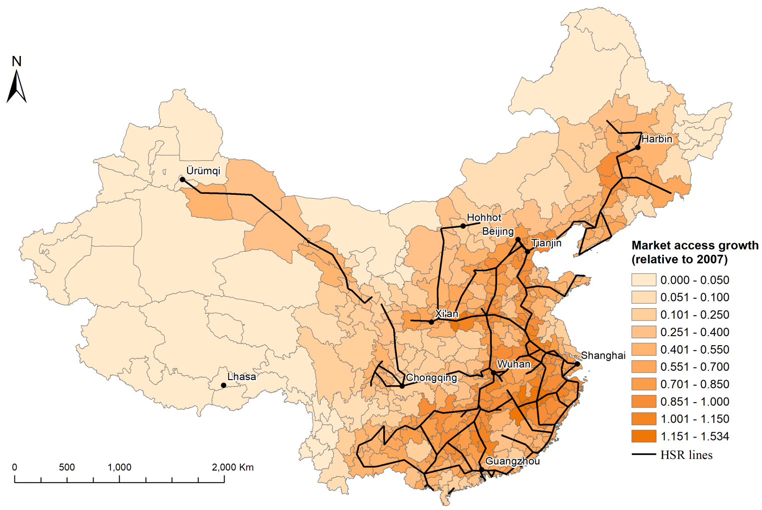
We quantify the systematic nature of spatial variation in MA growth in Column 1 of Table 2, by regressing it on a prefecture's distance to Beijing, its latitude, and its longitude. These simple predictors capture over 80% of the variation in MA growth, as measured by the regression's R^2. The high significance suggests an OVB concern: for causal interpretation of the Table 1 regression, one would need to assume that all unobserved determinants

opened simultaneously. One experimental HSR line between Qinhuangdao and Shenyang opened in 2003. We include it in our market access measure but focus on the majority of HSR-induced changes in the network over 2007–2016.

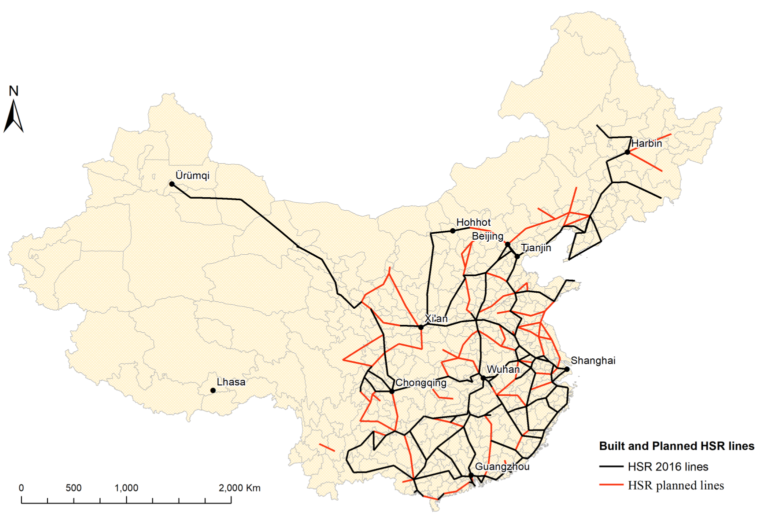
[31] This regression can be viewed as a reduced form of a hypothetical IV regression, in which the treatment variable is a measure of market access that accounts for changing population. We focus on the reduced form here because of data constraints: we observe annual population for all 340 prefectures only in the Census year of 2000. We discuss the potential roles of controls below.

Figure 2: Chinese High Speed Rail and Market Access Growth, 2007-2016

A. Completed Lines and Market Access Growth



B. All Completed and Planned Lines



Notes: Panel A shows the completed China high-speed rail network by the end of 2016, with shading indicating MA growth relative to 2007. Panel B shows the network of all HSR lines, including those planned but not yet completed as of 2016 (in red).

Table 1: Employment Effects of Market Access: Unadjusted and Recentered Estimates

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
Panel A. No Controls			
Market Access Growth	0.232	0.081	0.069
	(0.075)	(0.098)	(0.094)
		[-0.315, 0.328]	[-0.209, 0.331]
Expected Market Access Growth			0.318
			(0.095)
Panel B. With Geography Controls			
Market Access Growth	0.132	0.055	0.045
	(0.064)	(0.089)	(0.092)
		[-0.144, 0.278]	[-0.154, 0.281]
Expected Market Access Growth			0.213
			(0.073)
Recentered	No	Yes	Yes
Prefectures	274	274	274

Notes: This table reports coefficients from regressions of employment growth on MA growth in Chinese prefectures from 2007–2016. MA growth is unadjusted in Column 1. In Column 2 this treatment is instrumented by MA growth recentered by permuting the opening status of built and planned HSR lines with the same number of cross-prefecture links. Column 3 instead estimates an OLS regression with recentered MA growth as treatment and controlling for expected MA growth given by the same HSR counterfactuals. The regressions in Panel B control for distance to Beijing, latitude, and longitude. Standard errors which allow for linearly decaying spatial correlation (up to a bandwidth of 500km) are reported in parentheses. 95% RI confidence intervals based on the HSR counterfactuals are reported in brackets.

of employment growth (such as local productivity shocks) are uncorrelated with such geographic features. While one could of course control for the specific geographic variables from Table 2 (as we explore below), controlling perfectly for prefecture geography is impossible without removing all variation in z_ℓ.

Our solution is to view certain features of the HSR network as realizations of a natural experiment. By specifying a set of counterfactual HSR networks, which we corroborate with appropriate falsification tests, we can compute the appropriate function of geography μ_ℓ to remove the systematic variation in MA growth. The recentered regression leverages contrasts between actual and counterfactual realizations of the HSR assignment process, and not other cross-sectional variation.

Our specification of counterfactual upgrades exploits the heterogeneous timing of HSR construction. Specifically we permute the 2016 completion status of the built and planned

Table 2: Regressions of Market Access Growth on Measures of Economic Geography

	Unadjusted	Recentered		
	(1)	(2)	(3)	(4)
Distance to Beijing	−0.292	0.069		0.089
	(0.063)	(0.040)		(0.045)
Latitude/100	−3.323	−0.325		−0.156
	(0.648)	(0.277)		(0.320)
Longitude/100	1.329	0.473		0.425
	(0.460)	(0.239)		(0.242)
Expected Market Access Growth			0.027	0.056
			(0.056)	(0.066)
Constant	0.536	0.014	0.014	0.014
	(0.030)	(0.018)	(0.020)	(0.018)
Joint RI p-value		0.489	0.807	0.536
R^2	0.823	0.079	0.007	0.082
Prefectures	274	274	274	274

Notes: This table reports coefficients from regressing the unadjusted and recentered MA growth of Chinese prefectures (2007–2016) on predetermined geographic controls. Recentering is done by permuting the opening status of built and planned lines with the same number of cross-prefecture links. All regressors are measured for the prefecture's main city and demeaned such that the constant in each regression captures the average outcome value. Distance from Beijing is measured in 1,000km. Standard errors which allow for linearly decaying spatial correlation (up to a bandwidth of 500km) are reported in parentheses. Joint RI p-values are based on the 999 HSR counterfactuals and the sum-of-square fitted values statistic, as described in footnote 27.

lines, assuming that the timing of line completion is conditionally as-good-as-random. Panel B of Figure 2 compares the built and planned lines which form our counterfactuals. Planned lines tend to be concentrated in the same areas of China as built lines, reinforcing the fact that (unlike in our motivating example in Section 2) construction is not uniformly distributed in space. Although planned lines are of similar length, they tend to connect more regions: the average number of cross-prefecture "links" is 3.31 and 2.45 for built and planned lines, respectively, with a statistically significant difference ($p = 0.029$). To account for this difference we construct counterfactual upgrades by permuting 2016 completion status only among lines with the same number of links. This procedure generates counterfactual HSR maps that are visually similar to the actual 2016 network (see Appendix Figure A4 for an illustrative example) and which isolate more plausibly exogenous variation. For example the main Beijing to Shanghai HSR line, which has the greatest number of links, is always included in the counterfactuals.

Columns 2–4 of Table 2 validate this specification of the HSR assignment process by the test described in Section 3.4. Column 2 shows that recentering according to this specification (based on 999 counterfactual maps) successfully removes systematic geographic variation in market access. Specifically, it regresses the resulting recentered MA growth on a constant and the same controls as in Column 1 (distance to Beijing, latitude, and longitude). The regression coefficients and R^2 fall dramatically relative to Column 1, while a permutation-based p-value for their joint significance (based on the regression's sum-of-squares, as suggested in footnote 27) is 0.49. Columns 3 and 4 further show that recentered MA growth is uncorrelated with the expected instrument. These results are consistent with correct specification of counterfactuals (i.e. we cannot reject Assumption 3), though we note they do not provide direct support for the exogeneity of HSR construction to the unobserved determinants of employment (Assumption 1).[32]

Figure 3 plots expected and recentered MA growth (μ_ℓ and $\tilde{z}_\ell$) given by the permutations of built and planned lines. The effect of recentering is apparent by contrasting the dark- and light-shaded regions in Panel A of Figure 2 (indicating high and low MA growth) with the solid and striped regions in Panel B of Figure 3 (indicating high and low recentered MA growth). The recentered $\tilde{z}_\ell$ no longer places western prefectures in the effective control group, as their MA growth is as low as expected, and therefore $\tilde{z}_\ell \approx 0$. Similarly, some prefectures in the east (such as Tianjin) are no longer in the effective treatment group, as they saw an expectedly large increase in MA. At the same time, recentering provides a justification for retaining other regional contrasts. Hohhot, for example, expected higher MA growth than Harbin due to the planned connection to Beijing. This line was still under construction in 2016, however, resulting in lower MA growth in Hohhot than Harbin.

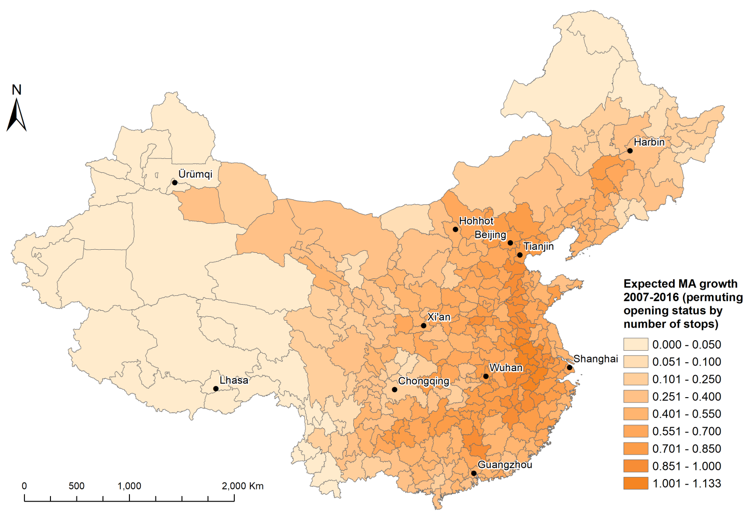
Column 2 of Table 1, Panel A, shows that instrumenting MA growth with its recentered measure reduces the estimated employment elasticity substantially, from 0.23 to 0.08. Controlling for expected MA growth yields a similar corrected estimate of 0.07 in Column 3. Neither of the two adjusted estimates is statistically distinguishable from zero according to either Conley (1999) spatial-clustered standard errors or permutation-based inference based on Lemma 2 (which yields a wider confidence interval in this setting). The difference between the unadjusted and adjusted estimates is explained by the fact that employment growth is strongly predicted by expected MA growth: in Column 3 we find a large coefficient on μ_ℓ, of 0.32.[33] This means employment grew faster in prefectures that were more highly

[32]While our specification tests pass for the 2007–2016 long difference, and are robust to using long differences ending in 2014 or 2015, we have verified in unreported results that the same assignment process is rejected in specifications which focus on earlier years of HSR development, when the network is much less dense and it is more difficult find good experimental contrasts. Focusing on the long difference also alleviates concerns of dynamic employment adjustments.

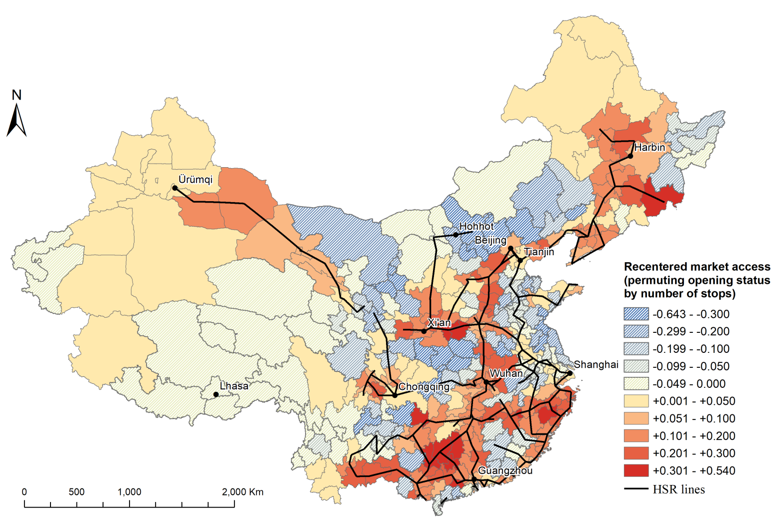
[33]Appendix Figure A5 visualizes these findings. We use recentered (rather than unadjusted) MA growth as the treatment in Column 3 of Table 1. This does not change the estimate of β, but makes the coefficient on the expected instrument more interpretable: the Column 1 estimate is then a weighted average of the two Column 3 coefficients.

Figure 3: Expected and Recentered Market Access Growth from Chinese HSR

A. Expected Market Access Growth



B. Recentered Market Access Growth



Notes: Panel A shows the variation in expected 2007–16 MA growth across Chinese prefectures, computed from 999 counterfactuals that permute the opening status of built and planned lines with the same number of cross-prefecture links. Panel B plots the variation in corresponding recentered MA growth: the difference between the MA growth shown in Panel A of Figure 2 and expected MA growth. The HSR network as of 2016 is also shown in this panel.

exposed to new HSR construction, whether or not the nearby lines have been built yet.[34]

Panel B of Table 1 shows that the geographic controls from Table 2 do not isolate the same variation as expected MA growth adjustment. Including these controls in the unadjusted regression of Column 1 yields a smaller but still economically and statistically significant estimate, of 0.13. In contrast, Columns 2 and 3 show that the finding of no significant MA effect after adjusting for μ_ℓ is robust to including these conventional controls. The μ_ℓ adjustment alone appears sufficient to remove the geographic dependence of MA, as Table 2 also showed.

Additional robustness checks are given in Appendix Tables A1 and A2. First, we show that the role of the expected instrument adjustment is virtually unchanged with two modifications to the market access regression often found in the literature. Specifically, we use a leave-one-out MA measure (e.g. Donaldson and Hornbeck 2016) and drop influential prefectures, which we define as province capitals, from the sample (e.g. Chandra and Thompson 2000). Second, we find similar results when replacing the MA treatment with a simpler measure of prefecture's connectivity to the HSR network (e.g. Faber 2014; Donaldson 2018). Third, we explore sensitivity to adding province fixed effects, which here bring the unadjusted coefficient on MA growth closer to zero while again confirming the robustness of the adjusted estimates. Finally, we show that recentering eliminates the effects on other measurements of employment growth, but not on rail passenger traffic (providing a useful reality check).

While our primary interest is to illustrate the above methodology, we note that there are several possible explanations for the substantive finding of a small employment effect of MA after recentering. Unlike other networks used for trading goods, the Chinese HSR network operates passenger trains. Its scope for directly affecting production is therefore smaller, although it could still facilitate cross-regional business relationships. In addition, the employment effects of growing market access could be positive for some regions but negative for others, as easier commuting between regions relocates employers. We leave analyses of such mechanisms and heterogeneity for future study.

In Appendix D.1 we discuss how the idea of recentering market access relates to the literature estimating the effects of transportation infrastructure upgrades on regional and bilateral outcomes, which remains challenging despite a long history in economics (Redding and Turner 2015). We first contrast the well-known challenge of strategically chosen transportation upgrades with the less discussed problem that regional exposure to exogenous

[34] Appendix Figure A6 shows that the μ_ℓ-adjusted estimates capture an average treatment effect corresponding to a diverse and reasonably representative set of regions, alleviating concerns that the difference relative to Column 1 of Table 1 may be compositional. Specifically, we plot the weights that the recentered IV implicitly places on each prefecture in the sample if the effects β_ℓ are linearly heterogeneous (see Corollary S1 in Appendix C.1).

upgrades may be unequal. We then explain how common strategies to address the former issue (e.g. by leveraging historical routes or inconsequential places) can be incorporated in our framework, at least in principle. At the same time, we highlight that recentering may still be needed to address the latter issue. We further discuss how some of the existing approaches naturally yield specifications of counterfactual networks (e.g. the placebos in Donaldson (2018) and Ahlfeldt and Feddersen (2018)) and summarize the conceptual and practical advantages of our approach relative to employing more conventional controls. We finally emphasize that even when a convincing specification of counterfactuals is challenging to obtain, *any* specification can yield a useful robustness check on these alternative identification strategies (see footnote 21).

4.2 Effects of Policy Eligibility

We next show how our framework can be used to construct more efficient instruments when estimating the effects of policy eligibility, relative to the commonly employed simulated instrument approach of Currie and Gruber (1996a, 1996b). Validity of our instruments relies on the same policy exogeneity assumptions, but power is increased by incorporating predictive endogenous variation in policy exposure and applying appropriate recentering. We first describe the general approach, drawing on the optimal IV results of Section 3.5. We then illustrate the power gains in an application estimating the take-up and crowd-out effects of Medicaid eligibility.

General Approach Suppose that β captures the causal effect of eligibility $x_\ell \in \{0, 1\}$ of individual ℓ for a public program (such as Medicaid or unemployment insurance) on some outcome y_ℓ (such as program takeup, health status, or educational attainment). Eligibility is a function—possibly a complicated one—of regional (e.g. state-level) government policy and individual characteristics such as income and family structure. We suppose the variation in state policies can plausibly be viewed as exogenous, while the individual characteristics are potentially correlated with the residual.

In such settings Currie and Gruber (1996a, 1996b; henceforth CG) propose the use of simulated instruments to isolate the exogenous policy variation.[35] The CG procedure measures the generosity of each state's policy as the average eligibility of a simulated nationally representative sample of individuals, if they were to reside in that state. We write this generosity as $h(g_n)$, indexing states by $n = 1, \ldots, 50$ and denoting policies by g_n. Each individual ℓ is then assigned the generosity of policy in their state of residence s_ℓ as the

[35]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 exogenous variation in policies across states.

instrument, $z_\ell^{CG} = h(g_{s_\ell})$. Since the CG instrument is a function of state policies only, it is valid when these policies arise in a natural experiment (as formalized below).[36]

The Section 3 framework suggests a different and likely more powerful approach, which leverages the same natural policy experiment. Indeed, eligibility can be written as $x_\ell = f(g_{s_\ell}; v_\ell)$, where v_ℓ denotes the demographics relevant to the policies (which we assume are observed by the econometrician and determined prior to the policies) and $f(\cdot)$ is a known mapping. It can therefore be perfectly predicted from the policies g and observables ($s = \{s_\ell\}$ and $v = \{v_\ell\}$) and, once recentered, coincides with the inner term of the optimal instrument in equation (9). To make recentering feasible, we formalize the natural experiment by assuming exchangeability of the policies g_n across states, conditional on (s, v) and error terms ε. This implies both Assumption 1 and Assumption 3, as the distribution of g_n conditional on $w = \{s, v, \Pi(g)\}$ is uniform across the 50 values of g_n and therefore known.[37]

With this formalization, one can purge OVB from an OLS regression of y_ℓ on x_ℓ by measuring each individual's expected eligibility over the possible policy counterfactuals, $\mu_\ell = \mathbb{E}\left[x_\ell \mid w\right] = \frac{1}{50}\sum_n f(g_n; v_\ell)$, and either instrumenting x_ℓ by $\tilde{z}_\ell = x_\ell - \mu_\ell$ or controlling for μ_ℓ. This procedure contrasts with the simulation in CG's approach: rather than applying ℓ's state policy to random individuals in order to construct an instrument for x_ℓ, our approach applies random state policies to individual ℓ in order to construct a control μ_ℓ.

The power gains with $\tilde{z}_\ell$ relative to z_ℓ^{CG} arise from a better first-stage prediction of x_ℓ. This can be understood by considering individuals who have the same eligibility under every state's policy, such that $x_\ell = \mu_\ell$. The presence of such individuals weakens the CG first stage, since their treatment status is unaffected by variation in z_ℓ^{CG}. The recentered IV estimator effectively removes these inframarginal individuals, for whom $\tilde{z}_\ell = 0$.

In Appendix D.2 we extend these insights by showing how more efficient instruments can be constructed when some individual determinants of eligibility are unobserved (as in Cohodes et al. (2016)) or endogenously respond to the state policies (as in East and Kuka (2015)). The results similarly apply in settings where only some policy variation is exogenous, as our application next illustrates. We further discuss in Appendix D.2 the advantages of our recentered IV relative to controlling for individual characteristics flexibly, as is common in the related literature on the eligibility effects of unemployment insurance (e.g., Cullen and Gruber 2000).

[36] It is straightforward to verify that under the assumptions of the natural experiment, the expected instrument corresponding to z_ℓ^{CG} is constant across individuals, and therefore there is no need to recenter.

[37] Like in Section 3.3, $\Pi(g)$ denotes the permutation class of g. Other specifications of counterfactuals, such as permutations within clusters, are similarly allowed. We illustrate such an approach in the application. We also note that statistical inference in this setting is straightforward with both z_ℓ^{CG} and $\tilde{z}_\ell$: when eligibility policies are *iid* across states, conventional state-clustered standard errors suffice.

Application We illustrate our approach by estimating the insurance coverage effects of a partial expansion of Medicaid eligibility in 2014. Medicaid is the largest U.S. health insurance program, covering around 29 million poor, non-disabled adults. One of the goals of the 2009 Affordable Care Act (ACA) was to extend Medicaid eligibility to all U.S. citizens and legal residents earning below 138% of the federal poverty line (FPL), replacing older eligibility rules that were mostly stricter and varied widely across U.S. states. The constitutionality of such an expansion was challenged (broadly along partisan lines), leading to a 2012 Supreme Court decision that left expansion to the discretion of individual state governors (NFIB v. Sebelius, 567 U.S. 519). In January 2014, when the ACA generally took effect, the federal Medicaid expansion was implemented by only 19 among the 43 states that had not expanded under the ACA or had a universal 138% FPL cutoff in prior years. The divide was partially along the party line: a minority (8 out of 30) of states with Republican governors but a majority (11 out of 13) of states with Democratic governors expanded eligibility. We refer to the former 19 states as having expanded Medicaid under the ACA, with the remaining 24 labeled as non-expansion states. Exact Medicaid eligibility criteria continued to have some variation across states in 2014, with some expansion states raising eligibility beyond the ACA's 138% FPL threshold and some non-expansion states partially raising eligibility though not fully to the ACA threshold.[38]

Applying our framework to this setting requires explicitly specifying counterfactual 2014 Medicaid expansions. Our baseline assumption is that a state's decision to expand is exchangeable within the sets of Republican and Democratic-governed states, while allowing states with different-party governors to have different propensities to expand. Thus, all scenarios in which some 8 Republican and some 11 Democratic states expanded are viewed as valid counterfactuals. This view of the 2014 expansions, as arising from a natural experiment, conforms with some earlier analyses (e.g. Ghosh et al. (2019), Black et al. (2019)).[39] We consider alternative assumptions on the expansion assignment process in robustness checks below.

We apply the framework using data from the 2013 and 2014 American Community Surveys on a representative 1% sample of non-disabled U.S. adults (ages 21-64) residing in the 43 states eligible for expansion in 2014. This repeated cross-section includes information

[38]We follow Frean et al. (2017) in using the Kaiser Family Foundation State Action database to determine which states adopted Medicaid expansions in each year; see https://web.archive.org/web/20150110162937/https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/. States which expanded coverage under the ACA or which had a universal 138% FPL cutoff prior to 2014 (and which are excluded from our analysis) are California, Connecticut, Massachusetts, Minnesota, New Jersey, Washington, and Vermont, plus the District of Columbia.

[39]Other analyses that do not reference natural experiments explicitly describe the expansions as "exogenous" and leverage difference-in-differences specifications comparing the outcome trends of individuals in expansion and non-expansion states before and after 2014 (e.g. Hu et al. (2018), Averett et al. (2019)).

on insurance coverage (by Medicaid, ACA marketplaces, and employer-sponsored plans), household income and other demographics determining Medicaid eligibility, such as employment status and family structure. We combine this main estimation sample with data from 2012 for falsification exercises; Appendix A.2 describes the sample construction in detail.

We estimate take-up and crowd-out effects from second-stage specifications of the form

$$y_{\ell t} = \beta x_{\ell t} + \alpha_{s_{\ell t}} + \tau_t + c_{\ell t}' \gamma + \varepsilon_{\ell t}, \tag{11}$$

where ℓ indexes individuals and t indexes years (either 2013 or 2014). The outcome $y_{\ell t}$ is an indicator for a particular type of health insurance coverage (e.g. Medicaid or private insurance), and the treatment $x_{\ell t}$ is an indicator for Medicaid eligibility under the year-t eligibility rules of ℓ's state of residence $s_{\ell t}$. We include state and year fixed effects α_s and τ_t and time-varying controls $c_{\ell t}$, discussed below. Recognizing that eligibility is likely endogenous, we instrument it with two alternative IVs.

We construct the simulated eligibility instrument $z_{\ell t}^{CG}$ consistently with our stance that only a state's decision to expand Medicaid in 2014 is exogenous (and not, for example, its prior level of generosity). As a function of policy variation only, the CG instrument is in this case equivalent (in the sense of producing the same estimates) to $g_{s_{\ell t}} \cdot \mathbf{1}[t = 2014]$, the simple interaction of residing in an expansion state ($g_{s_{\ell t}} = 1$, where $s_{\ell t}$ is individual ℓ's state of residence) with the 2014 indicator. We nevertheless construct $z_{\ell t}^{CG}$ by a simulation that follows the original logic of Currie and Gruber (1996a); see Appendix A.2 for details. We include in the control vector $c_{\ell t}$ an indicator for residing in a Republican-governed state, interacted with year, to match our assumption of conditional exogeneity of expansion decisions within each governor's party.

The alternative recentered IV also leverages conditionally exogenous variation in state Medicaid expansion decisions while further incorporating individual heterogeneity to better predict Medicaid eligibility. We construct eligibility predictions $z_{\ell t} = f(g_{s_{\ell t}}; v_{\ell t})$ by including in $v_{\ell t}$ all individual demographics that affect eligibility (household income, parental and employment status) as well as the precise eligibility rules of the individual's state in 2013, as they are also viewed as non-random. This construction allows for a perfect prediction of $z_{\ell t} = x_{\ell t}$ in 2013; in 2014 we predict eligibility from state expansion decisions and prior eligibility policy (again see Appendix A.2 for details).

The expected instrument which corresponds to this $z_{\ell t}$ is obtained by permuting expansion decisions within Republican- and Democratic-governed states. It defines a sample of "non-exposed" individuals whose demographics and state of residence make them always or never eligible for Medicaid in 2014 regardless of the expansion decision, and a set of "exposed" individuals for whom the expansion shock is relevant. Per the discussion above,

Table 3: Medicaid Eligibility Effects: First-Stage Estimates

	(1)	(2)	(3)
Simulated IV	0.851	0.032	
	(0.113)	(0.140)	
	[0.567,1.115]	[-0.254,0.503]	
Recentered IV		0.817	0.972
		(0.171)	(0.015)
		[0.397,1.162]	[0.941,1.014]
Partial R^2	0.022	0.113	0.894
Exposed Sample	N	N	Y
States	43	43	43
Individuals	2,397,313	2,397,313	421,042

Notes: This table reports coefficients from first-stage regressions of Medicaid eligibility on the two instruments described in the text: a simulated eligibility instrument and a recentered prediction of Medicaid eligibility. Columns 1 and 2 estimate regressions in the full sample of individuals in 2013–14, while Column 3 restricts to the sample of individuals in both years whose individual characteristics make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year. State-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets. R^2 statistics partial out the fixed effects and controls.

we drop non-exposed individuals from the 2014 sample and, in keeping with the difference-in-difference structure, also drop individuals in 2013 whose individual characteristics would make them non-exposed in 2014. The remaining variation in $\mu_{\ell t}$ is absorbed by the year-interacted state party indicator in $c_{\ell t}$, making recentering unnecessary.

Table 3 shows that the recentered IV is much more predictive of actual Medicaid eligibility $x_{\ell t}$ than the simulated instrument. Column 1 regresses $x_{\ell t}$ on the simulated instrument $z^{CG}_{\ell t}$, controlling for state and year fixed effects and the year-interacted state party control. The partial R^2 in this first-stage regression instrument is quite small, at 2.2 percent. Adding the recentered IV in Column 2 increases the partial R^2 dramatically, to 11.3 percent. Moreover, the coefficient on the simulated instrument falls from 0.85 to an insignificant 0.03.[40] In Column 3 we restrict estimation to the "exposed" sample of individuals whose demographics and state of residence make them marginal for the potential expansion of Medicaid eligibility in 2014. Here we find that a one percentage point increase in the recentered IV predicts a 0.97 percentage point increase in actual Medicaid eligibility. This first-stage coefficient

[40] We use state-clustered standard errors but, to address finite-sample concerns with a relatively small number of state clusters, also report confidence intervals by a wild score bootstrap as suggested by Kline and Santos (2012). This computationally efficient approach requires inverting bootstrapped test statistics, which generally makes confidence intervals asymmetric around the IV point estimate.

would equal one if changes in eligibility were only driven by state expansion decisions; the fact that it is close to one reflects the relatively small role of other policy changes. We also find a high partial R^2 in this specification, of 89.4 percent, reflecting the fact that we have removed individuals whose eligibility is unaffected by state expansion decisions.

Table 4 shows that estimates of Medicaid take-up and private insurance crowd-out effects are correspondingly much more precise when obtained with the recentered IV. Associated standard errors and confidence interval lengths fall by around 60-70 percent when we replace $z_{\ell t}^{CG}$ with $z_{\ell t}$ and restrict estimation to the exposed sample. In Columns 1 and 2 of Panel A we obtain a recentered IV confidence interval of [0.05,0.09] for the take-up effect, relative to a much wider simulated instrument confidence interval of [0.08,0.22]. For private insurance crowd-out, the respective confidence intervals in Columns 3 and 4 are [-0.04,-0.01] and [-0.10,0.01]. Thus we can only reject the null hypothesis of no crowd-out with 95% confidence when using the recentered IV. These two columns include both the conventional crowd-out margin of employer-sponsored insurance as well as the novel form of private marketplace insurance introduced by the ACA. In Columns 5 and 6 we focus on crowd-out of employer-sponsored plans. Neither the simulated nor recentered IV yields statistically significant estimates at the 95% level, though the latter is again much more precise.

In economic terms, the recentered IV estimates in Panel A of Table 4 suggest a total private insurance crowd-out rate of 30%, with a 7.2 percentage point increase in overall coverage offset by a 2.3 percentage point decrease in private insurance coverage. This overall effect is similar to the 42% crowd-out that Leung and Mas (2018) find in applying a difference-in-differences specification to the 2014 Medicaid expansion. However, we find no evidence for crowd-out from employer-sponsored insurance plans even with our more powerful recentered IV. Instead, our estimates suggest the crowd-out arises entirely from direct-purchase private insurance via new ACA marketplaces. This aligns with the finding of Frean et al. (2017), who exploit multiple sources of ACA-induced policy variation in a simulated instrument design (see also Kaestner et al. (2017) and Maclean and Saloner (2019)), and contrasts with earlier settings (e.g. Cutler and Gruber 1996).[41]

Panel B of Table 4 shows that these substantive findings and power gains are not driven by the relatively simple regression specification. Adding flexible controls for the individual characteristics which drive exposure to different policies (deciles of household income, interacted with indicators for parental and work status and year) in $c_{\ell t}$ leaves both the point estimates and the difference between simulated IV and recentered IV standard errors and confidence interval lengths unchanged.

We further analyze the robustness of this analysis in Appendix A.3. First, we validate

[41] See Guth et al. (2020) for a review of the literature on ACA expansion effects, which suggests that more widespread eligibility increased access to and utilization of care, led to local economic gains, and improved health outcomes.

Table 4: Medicaid Eligibility Effects: Simulated and Recentered IV Estimates

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV (1)	Recentered IV (2)	Simulated IV (3)	Recentered IV (4)	Simulated IV (5)	Recentered IV (6)
Panel A. Baseline Controls						
Eligibility	0.132	0.072	−0.048	−0.023	0.009	−0.009
	(0.028)	(0.010)	(0.023)	(0.007)	(0.014)	(0.005)
	[0.080,0.218]	[0.051,0.094]	[-0.109,0.010]	[-0.039,-0.008]	[-0.035,0.053]	[-0.021,0.004]
Panel B. With Demographics $\times$ *Post*						
Eligibility	0.135	0.073	−0.050	−0.024	0.003	−0.008
	(0.029)	(0.010)	(0.022)	(0.007)	(0.013)	(0.005)
	[0.082,0.223]	[0.051,0.096]	[-0.114,-0.002]	[-0.041,-0.008]	[-0.038,0.036]	[-0.020,0.005]
Exposed Sample	N	Y	N	Y	N	Y
States	43	43	43	43	43	43
Individuals	2,397,313	421,042	2,397,313	421,042	2,397,313	421,042

Notes: This table reports coefficients from IV regressions of different measures of health insurance coverage on Medicaid eligibility, instrumented by one of the two IVs described in the text: a simulated eligibility instrument and a recentered prediction of Medicaid eligibility. Columns 1, 3, and 5 estimate regressions in the full sample of individuals in 2013–2014, while Columns 2, 4, and 6 restrict to the sample of individuals whose individual characteristics make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year; the regressions in Panel B additionally control for deciles of household income, interacted with indicators for parental and work status and year. Conventional state-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

our assumption of expansion exogeneity with a placebo test that replaces 2014 outcomes with a comparable cross-section from 2012. Although with the increased precision from the recentered IV we are able to reject the null hypothesis of no pre-trends, Appendix Table A3 shows that the magnitude of the placebo coefficient is small (around 0.01–0.02) regardless of the outcome and the instrument we use. Second, we relax the key exogeneity assumption by allowing a state's decision to expand to depend not only on the political party of their governor, but also on the state's median household income and previous rate of Medicaid coverage. Appendix Table A4 shows that the estimated effects of eligibility remain very similar across specifications. Third, we explore robustness to another implementation of our approach: namely, using the recentered $z_{\ell t}$ as the instrument without restricting to the exposed sample. Appendix Table A5 shows that this approach only yields power gains when the additional demographic controls (those from Panel B of of Table 4) or an indicator for being in the exposed sample interacted with year are included in $c_{\ell t}$. We discuss the reason for this in Appendix A.3 by relating it to our general efficiency theory of Section 3.5.

Finally, we confirm large and uniform power gains from using the recentered IV in a Monte Carlo study based on our baseline estimates. In this controlled environment the true causal effect and the shock assignment process are known, allowing us to verify that recentered IV estimator is both close to unbiased and significantly more efficient than the traditional simulated instrument approach. We find, for example, that the minimum detectable effects of simulated IV (the smallest null hypotheses which are rejected by a 0.05-size test with probability 0.8) are roughly three times larger than those of the recentered IV (see Appendix Figures A7 and A8).

4.3 Other Settings

Our framework bears practical lessons for a range of other common z_ℓ: network spillover treatments, linear and nonlinear shift-share instruments, model-implied instruments, instruments from centralized school assignment mechanisms, "free-space" instruments for mass media access, and weather instruments. Here we map these settings to the general framework. In Appendices D.3–D.8 we detail how expected instrument recentering and RI can be used to relax various assumptions often imposed with such z_ℓ.

In spillover regressions, the units ℓ represent nodes in a network (of people, firms, regions, etc.) and g captures exogenous shocks assigned to the same or other nodes in an RCT or a natural experiment. Spillover treatments can then count the number of ℓ's shocked neighbors (perhaps with importance weights), check whether this number exceeds a certain threshold, or measure the network distance to the nearest treated node (e.g. Miguel and Kremer (2004), Jaravel et al. (2018), and Carvalho et al. (2020)). All such treatments are co-determined by g and the network adjacency matrix w (often nonlinearly in g), and thus may require recentering to leverage the exogenous variation.[42]

Shift-share instruments (SSIVs), often constructed for regions ℓ, combine lagged local shares $w_\ell = \{w_{\ell n}\}_{n=1}^N$ of, for instance, employment across a set of industries $n = 1, \ldots, N$ with a set of national shocks $g = \{g_n\}_{n=1}^N$. In many applications the shocks are considered exogenous, perhaps conditionally on a vector of industry-level controls q_n, while the shares may be endogenous. Borusyak et al. (2020) show how for linear SSIVs, of the form $z_\ell = \sum_n w_{\ell n} g_n$, OVB from non-random exposure is removed by controlling for $\sum_n w_{\ell n} q_n$ provided $\mathbb{E}[g_n \mid q_n]$ is linear. In the language of the present paper, such controls absorb the expected instrument. Relative to their paper, our framework nests nonlinear SSIVs $z_\ell = f(g, w_\ell)$—a recent and growing class of instruments which has not previously been formalized, and for which the OVB problem is more challenging.[43] Examples of nonlinear SSIVs include predicted local Gini indices based on national shocks to the income distribution (Boustan et al. 2013), predicted labor reallocation indices based on national industry shocks (Chodorow-Reich and Wieland 2020), and "SSIV in logs" such as predicted firm-level log-exports (Berman et al. 2015).

Adão et al. (2020) propose "model-implied optimal instruments" as a way to estimate parameters of general equilibrium models with spatial spillovers. They leverage the responses of endogenous variables y_ℓ to a set of exogenous shifters g_n in a log-linear approximation around some initial equilibrium w. Since such responses are functions of the non-random w (e.g. they depend on the initial local industry composition and migration shares), recentering is generally required to just leverage the exogenous variation in g_n. Linearity of the model in g makes recentering straightforward, even when the responses are nonlinear in parameters. In Appendix D.5 we discuss how our efficiency result extends that of Adão et al. (2020) to account for recentering and the non-*iid*ness of data in spatial equilibrium.

When estimating the causal effects of enrollment in certain groups of schools (e.g., charter schools), researchers increasingly leverage partially randomized school assignment mechanisms (Abdulkadiroglu et al. 2017; 2019). For example, they may instrument enrollment with centralized assignment to a charter school, $z_\ell \in \{0, 1\}$. This z_ℓ is a function of the non-random inputs to the assignment mechanism captured by w (e.g. the submitted preferences and priorities of all students) and a set of exogenous inputs g, such as random numbers used to break ties among students with equal priority. Abdulkadiroglu et al. (2017; 2019) derive analytical propensity scores μ_ℓ for some deferred acceptance mechanisms, which are valid in large samples, and discuss how in some cases re-randomizing the exogenous shocks

[42] In a related class of applications, the treatment of interest is node centrality, affected by the shocks g that change network edges (e.g. Campante and Yanagizawa-Drott (2018)). The market access regressions of Section 4.1 can also be understood this way.

[43] In Appendix D.4 we show that this problem can also be solved without fully specifying counterfactual shocks by using a first-order approximation to z_ℓ (which is a linear SSIV), at a likely efficiency cost.

can also be used in finite samples. Our framework nests the latter solution while not being limited to binary z_ℓ or to specific assignment mechanisms.

The literature on the effects of access to mass media (e.g. to radio or television) points out that the local quality of reception $z_\ell = f_\ell(g, w)$ is co-determined by the location of transmitters w and the country topography g (e.g. mountain ranges) that can inhibit transmission (e.g. Olken 2009; Yanagizawa-Drott 2014). Viewing w as endogenous, some papers control for the "free-space" measure of access that assumes away any transmission barriers. While this control is similar in spirit to μ_ℓ, it is not identical: our framework shows that controlling for the average quality of reception under realistic counterfactual topographies may be more appropriate.

Our final example is when spatial variation in weather is used as an instrument, e.g. with rainfall instrumenting for the election turnout (e.g. Gomez et al. 2007; Madestam et al. 2013). Causal identification may be threatened by the fact that local weather z_ℓ is co-determined by exogenous day-specific factors g and local climate w_ℓ, which may be correlated with unobservables. Moreover, statistical inference is difficult because all determinants of weather are heavily spatially correlated (Lind 2019). Recentering and permutation inference based on historic weather maps (e.g. from the same day of other years) may therefore be attractive solutions, and some versions of it have been applied in the literature (see Appendix D.8).

5 Conclusion

Many studies in economics construct treatments or instruments that combine multiple sources of variation, according to a known formula. We develop a general econometric framework for such settings when some, but not all, of such variation is exogenous. Except in special cases, endogenous exposure to exogenous shocks generates bias in conventional regression estimators, and the interdependencies inherent in such settings invalidate standard modes of statistical inference. We show how these identification and inference problems can be solved by specifying an assignment process for the exogenous shocks: namely, a set of counterfactual shocks that might as well have been realized.

This general framework has concrete implications for a large number of settings. We illustrate the usefulness of specifying counterfactuals for new railroad construction when leveraging this variation to estimate market access effects. Estimates of the effects of high-speed rail on local employment in China are reduced to a statistical zero when adjusting for a region's expected market access growth. We further show how our framework can be used to construct instruments which may be a more powerful alternative to simulated eligibility IVs. Estimates of Medicaid take-up and crowd-out effects are more than three times as precise when obtained by an instrument incorporating both as-good-as-random policy variation and non-random individual exposure, without a need for stronger identification assumptions. We discuss practical implications for other settings, including spillover regressions, linear and nonlinear shift-share IV regressions, structural estimation with model-implied IVs, and estimation of the effects of centralized school assignment, access to mass media, and weather.

The key challenge of applying our framework, absent true randomization, is in specifying plausible shock counterfactuals. In the paper we illustrate how this can be accomplished in a variety of settings by finding exchangeable features of the shocks. For example, permuting the timing of railroad upgrades within observably similar groups may yield a plausible set of counterfactuals for gauging the potential for OVB. We also show how such specifications can be tested. We consider some partly-specified shock assignment processes in Appendix C.5; future research may yield more flexible approaches.

In our view, specifying shock counterfactuals has inherent value in observational studies that claim to leverage a natural experiment, understood as a serendipitous randomized trial (DiNardo 2008). A virtue of randomized trials is that valid causal inference can be conducted without imposing non-experimental assumptions on the unobservables. In the settings we consider, this property is only maintained when an expected instrument adjustment is performed, which generally requires an explicit shock assignment process. Methods that instead rely on properties of the unobservables, such as by a parallel trends assumption, are instead referred to as quasi-experimental by DiNardo (2008).[44] Generalizing our framework to combine specifications of shock counterfactuals with plausible restrictions on the residual appears a fruitful area for future work and may yield new "double-robust" identification results, in a sense similar to that of Arkhangelsky and Imbens (2019).

References

Abadie, Alberto. 2003. *Semiparametric instrumental variable estimation of treatment response models,* 113:231–263.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2020. "Sampling-based vs. Design-based Uncertainty in Regression Analysis." *Econometrica* 88:265–296.

Abadie, Alberto, Alexis Diamond, and Jens 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, Atila, Joshua D Angrist, Yusuke Narita, and Parag A. Pathak. 2019. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Working Paper.*

[44] An instructive example can be found in the setting of Redding and Sturm (2008) who study the effects of German reunification—an event for which parallel trends may plausibly hold while no obvious counterfactuals exist, and thus a natural experiment may not be a fitting term.

Abdulkadiroglu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica* 85:1373–1432.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105:2364–2409.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2020. "General Equilibrium Indirect Effects in Space: Theory and Measurement." *Working Paper.*

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134:1949–2010.

Ahlfeldt, Gabriel M., and Arne Feddersen. 2018. "From periphery to core: Measuring agglomeration effects using high-speed rail." *Journal of Economic Geography* 18:355–390.

Arkhangelsky, Dmitry, and Guido W. Imbens. 2019. "Double-Robust Identification for Causal Panel Data Models."

Aronow, Peter M. 2012. "A General Method for Detecting Interference Between Units in Randomized Experiments." *Sociological Methods and Research* 40:3–16.

Aronow, Peter M., Dean Eckles, Cyrus Samii, and Stephanie Zonszein. 2020. "Spillover Effects in Experimental Data." *arXiv preprint.*

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

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido W. Imbens, and Khashayar Khosravi. 2018. "Matrix Completion Methods for Causal Panel Data Models." *NBER Working Paper 25132.*

Athey, Susan, and Guido W. Imbens. 2018. "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption." *Working Paper.*

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

Averett, Susan L, Julie K Smith, and Yang Wang. 2019. "Medicaid expansion and opioid deaths." *Health Economics* 28:1491–1496.

Bartelme, Dominick. 2018. "Trade Costs and Economic Geography: Evidence from the U.S." *Working Paper.*

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

Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62:657.

Berman, Nicolas, Antoine Berthou, and Jérôme Héricourt. 2015. "Export Dynamics and Sales at Home." *Journal of International Economics* 96:298–310.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review* 89:400–430.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119:249–275.

Black, Bernard, Alex Hollingworth, Leticia Nunes, and Kosali Simon. 2019. "The Effect of Health Insurance on Mortality: Power Analysis and What We Can Learn from the Affordable Care Act Coverage Expansions." *NBER Working Paper 25568.*

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2020. "Quasi-Experimental Shift-Share Research Designs." *NBER Working Paper 24997.*

Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric M. Zolt. 2013. "The effect of rising income inequality on taxation and public expenditures: Evidence from U.S. Municipalities and school districts, 1970-2000." *Review of Economics and Statistics* 95:1291–1302.

Campante, Filipe, and David Yanagizawa-Drott. 2018. "Long-Range Growth: Economic Development in the Global Network of Air Links." *Quarterly Journal of Economics* 133:1395–1458.

Canay, Ivan A., and Vishal Kamat. 2018. "Approximate permutation tests and induced order statistics in the regression discontinuity design." *Review of Economic Studies* 85:1577–1608.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2020. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." *Working paper.*

Cattaneo, Matias D., Brigham R. Frandsen, and Rocío Titiunik. 2015. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." *Journal of Causal Inference* 3:1–24.

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

Chandra, Amitabh, and Eric Thompson. 2000. "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system." *Regional Science and Urban Economics* 30:457–490.

Chodorow-Reich, Gabriel, and Johannes Wieland. 2020. "Secular Labor Reallocation and Business Cycles." *Journal of Political Economy.*

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

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92:1–45.

Cullen, Julie Berry, and Jonathan Gruber. 2000. "Does Unemployment Insurance Crowd out Spousal Labor Supply?" *Journal of Labor Economics* 18:546–572.

Currie, Janet, and Jonathan 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.

Cutler, David M., and Jonathan Gruber. 1996. "Does Public Insurance Crowd out Private Insurance?" *The Quarterly Journal of Economics* 111:391–430.

De Chaisemartin, Clement, and Luc Behaghel. 2018. “Estimating the Effect of Treatments Allocated by Randomized Waiting Lists.” *Working Paper.*

De Chaisemartin, Clement, and Xavier D’haultfœuille. 2020. “Two-way fixed effects estimators with heterogeneous treatment effects.” *American Economic Review* 110:2964–2996.

Dell, Melissa, and Benjamin Olken. 2018. “The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java.” *Working Paper.*

DiNardo, John. 2008. “Natural Experiments and Quasi-Natural Experiments.” In *The New Palgrave Dictionary of Economics.* London: Palgrave Macmillan.

Donaldson, Dave. 2018. “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” *American Economic Review* 108:899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. “Railroads and American Economic Growth: A "Market Access" Approach.” *Quarterly Journal of Economics* 131:799–858.

Doudchenko, Nick, Minzhengxiong Zhang, Evgeni Drynkin, Edoardo Airoldi, Vahab Mirrokni, and Jean Pouget-Abadie. 2020. “Causal Inference with Bipartite Designs”: 1–35.

East, Chloe N., and Elira Kuka. 2015. “Reexamining the consumption smoothing benefits of Unemployment Insurance.” *Journal of Public Economics* 132:32–50.

Ellison, Glenn, and EL Glaeser. 1997. “Geographic concentration in US manufacturing industries: a dartboard approach.” *Journal of Political Economy* 105:889–927.

Faber, Benjamin. 2014. “Trade Integration, Market Size , and Industrialization: Evidence from China’ s National Trunk Highway System.” *Review of Economic Studies* 81:1046–1070.

Fisher, Ronald Aylmer. 1935. *The design of experiments.* Oliver & Boyd.

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers. 2017. “Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act.” *Journal of Health Economics* 53:72–86.

Ganong, Peter, and Simon Jäger. 2018. “A Permutation Test for the Regression Kink Design.” *Journal of the American Statistical Association* 113:494–504.

Gerber, Alan S, and Donald P Green. 2012. *Field experiments: Design, analysis, and interpretation.* W. W. Norton & Company.

Ghosh, Ausmita, Kosali Simon, and Benjamin D Sommers. 2019. “The Effect of Health Insurance on Prescription Drug Use Among Low-Income Adults : Evidence from Recent Medicaid Expansions.” *Journal of Health Economics* 63:64–80.

Gomez, Brad T., Thomas G. Hansford, and George A. Krause. 2007. “The Republicans should pray for rain: Weather, turnout, and voting in U.S. presidential elections.” *Journal of Politics* 69:649–663.

Gonzalez-Navarro, Marco, and Climent Quintana-Domeque. 2016. “Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects.” *Review of Economics and Statistics* 98:254–267.

Gruber, Jon, and Emmanuel Saez. 2002. “The elasticity of taxable income: Evidence and implications.” *Journal of Public Economics* 84:1–32.

Guth, Madeline, Rachel Garfield, and Robin Rudowitz. 2020. “The effects of medicaid expansion under the ACA: Updated Findings from a Literature Review.” *Kaiser Family Foundation* 37:944–950.

Hirano, Keisuke, and Guido W. Imbens. 2004. "The Propensity Score with Continuous Treatments." Chap. Chapter 7 in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family,* 73–84.

Ho, Daniel E., and Kosuke 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.

Hodges, J.L. Jr., and Erich L Lehmann. 1963. "Estimates of Location Based on Rank Tests." *The Annals of Mathematical Statistics* 34:598–611.

Hu, Luojia, Robert Kaestner, Bhashkar Mazumder, Sarah Miller, and Ashley Wong. 2018. "The effect of the affordable care act Medicaid expansions on fi nancial wellbeing." *Journal of Public Economics* 163:99–112.

Hudgens, Michael G., and M. Elizabeth Halloran. 2008. "Toward causal inference with interference." *Journal of the American Statistical Association* 103:832–842.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62:467.

Imbens, Guido W., and Paul 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 (January): 109–126.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-Specific Capital and Innovation." *American Economic Review* 108:1034–1073.

Kaestner, Robert, Bowen Garrett, Jiajia Chen, Anuj Gangopadhyaya, and Caitlyn Fleming. 2017. "Effects of ACA Medicaid Expansions on Health Insurance Coverage and Labor Supply." *Journal of Policy Analysis and Management* 36:608–642.

Kline, Patrick, and Andres Santos. 2012. "A Score Based Approach to Wild Bootstrap Inference." *Journal of Econometric Methods* 1:23–41.

Lawrence, Martha, Richard Bullock, and Ziming Liu. 2019. *China's High-Speed Rail Development.* Washington, D.C.: World Bank.

Lee, David S. 2008. "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics* 142:675–697.

Lehmann, Erich L, and Joseph P Romano. 2006. *Testing statistical hypotheses.* Springer Science & Business Media.

Leung, Pauline, and Alexandre Mas. 2018. "Employment Effects of the ACA Medicaid Expansions." *Industrial Relations* 57:206–234.

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

Lind, Jo Thori. 2019. "Spurious weather effects." *Journal of Regional Science* 59:322–354.

Ma, Damien. 2011. "China's Long, Bumpy Road to High-Speed Rail." *The Altantic.*

Mackinnon, James G., and Matthew D. Webb. 2020. "Randomization Inference for Difference-in-Differences with Few Treated Clusters." *Journal of Econometrics.*

Maclean, Johanna Catherine, and Brendan Saloner. 2019. "The Effect of Public Insurance Expansions on Substance Use Disorder Treatment: Evidence from the Affordable Care Act." *Journal of Policy Analysis and Management* 38:366–393.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do Political Protests Matter? Evidence from the Tea Party Movement." *Quarterly Journal of Economics* 128:1633–1685.

Manski, Charles F. 2013. "Identification of treatment response with social interactions." *Econometrics Journal* 16:1–23.

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

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

Olken, Benjamin A. 2009. "Do television and radio destroy social capital? Evidence from indonesian villages." *American Economic Journal: Applied Economics* 1:1–33.

Ollivier, Gerald, Richard Bullock, Ying Jin, and Nanyan Zhou. 2014. "High-Speed Railways in China: A Look at Traffic." *China Transport Topics:* 1–12.

Potlogea, Andrei, and Wenya Cheng. 2017. "Trade Liberalization and Economic Development: Evidence from China's WTO Accession." *Working paper.*

Redding, Stephen, and Daniel Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review* 98:1766–1797.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of regional and urban economics,* 1339–1398. Elsevier.

Redding, Stephen J., and Anthony J. Venables. 2004. "Economic geography and international inequality." *Journal of International Economics* 62:53–82.

Rosenbaum, Paul R, and Donald 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.

Rosenbaum, Paul R. 1984. "Conditional permutation tests and the propensity score in observational studies." *Journal of the American Statistical Association* 79:565–574.

———. 2002. "Covariance adjustment in randomized experiments and observational studies." *Statistical Science* 17:286–327.

Shaikh, Azeem, and Panagiotis Toulis. 2019. "Randomization Tests in Observational Studies with Staggered Adoption of Treatment." *Working Paper.*

Titiunik, Rocío. 2020. "Natural Experiments." *arXiv preprint.*

Tsivanidis, Nick. 2019. "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." *Working Paper.*

Volpe Martincus, Christian, and Juan Blyde. 2013. "Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment." *Journal of International Economics* 90:148–161.

Wooldridge, Jeffrey M. 2002. "Econometric analysis of cross section and panel data MIT press." *Cambridge, MA.*

Yanagizawa-Drott, David. 2014. "Propaganda and conflict: Evidence from the Rwandan genocide." *Quarterly Journal of Economics* 129:1947–1994.

Zheng, Siqi, and Matthew E. Kahn. 2013. "China's bullet trains facilitate market integration and mitigate the cost of megacity growth." *Proceedings of the National Academy of Sciences of the United States of America* 110:1248–1253.

Online Appendix to "Non-Random Exposure to Exogenous Shocks: Theory and Applications"

Kirill Borusyak, UCL and CEPR
Peter Hull, UChicago and NBER

A Empirical Appendix

A.1 Data for Section 4.1

Our analysis of market access effects uses data on 340 prefectures of mainland China. This excludes the islands of Hainan and Taiwan and the special administrative regions of Hong Kong and Macau. At the same time this includes six sub-prefecture-level cities (e.g. Shihezi) that do not belong to any prefecture. We use United Nations shapefiles to geocode each prefecture by the location of its main city (or, in a few cases, by the prefecture centroid).[45]

We use a variety of sources to assemble a comprehensive database of the HSR network in 2016 as well as the lines planned (and in many cases under construction) as of April 2019 but not opened yet by the end of 2016. Our starting points are Map 1.2 of Lawrence et al. (2019), China Railway Yearbooks, and the replication files of Lin (2017). We cross-check network links across these sources and use Internet resources such as Wikipedia and Baidu Baike to confirm and fill in missing information. Our database includes various types of HSR lines, including the National HSR Grid (4+4 and 8+8) and high-speed intercity railways. However, we only consider newly built HSR lines, excluding traditional lines upgraded to higher speeds. We do not put further restrictions on the class of trains (e.g. to G- and D-classes only) or specify an explicit minimum speed. The operating speed therefore ranges between 160 and 380kph, although the majority of lines are at 250kph. For each line we collect the date of its official opening (if it has opened), the actual or planned operating speed, and the list of prefecture stops. When different sections of the same line opened in a staggered way, we classify each section as a separate line for the purposes of constructing our 999 counterfactuals, following the definition of a line in footnote 30. We include only one contiguous stop per prefecture and drop lines that do not cross prefecture borders.

To measure the 2007–2016 market access growth according to the formula given on p. 22, we compute travel time between all pairs of cities k and ℓ as of the ends of those years for both the actual and counterfactual networks. Travel time combines traditional modes of

[45] The shapefiles are obtained from https://data.humdata.org/dataset/province-and-prefecture-capitals-of-china and https://data.humdata.org/dataset/china-administrative-boundaries, accessed on April 4, 2020.

transportation (car or low-speed train) with HSR, where available. We allow for unlimited changes between different HSR lines and between HSR and traditional modes without a layover penalty, as HSR trains tend to operate frequently and traditional modes also involve downtime. Following the existing literature, we proxy for travel time by traditional modes by the straight-line distance, and specify the speed of $100 = 120/1.2$kph, where 120kph is their typical speed and the 1.2 adjustment for actual routes that are longer than a straight line. For two prefectures connected by an HSR line, we compute the distance along the line as the sum of straight-line distances between adjacent prefectures on the line. We use the operating speed of each line divided by an adjustment factor of 1.3 to capture the fact that the average speed is lower than the nominal speed we record. Computing market access further requires the population of each of the 340 prefectures from the 2000 population Census, which we obtain from the CityPopulation.de website.[46] For the robustness checks in Table A1, Panels A and D, we also compute the leave-one-out measure of market access, $MA_{\ell t}^{LOO} = \sum_{k \neq \ell} \exp\left(-0.02\tau_{\ell k t}\right) \cdot P_{k,2000}$, and define a simple binary indicator that a prefecture has at least one HSR stop by the end of year t under the actual or counterfactual network map.

We measure prefecture employment in the 2008–2017 China City Yearbooks.[47] Each yearbook covers the previous year (so our data cover 2007–2016). The yearbooks provide most variables for two spatial definitions: the entire prefecture and the "urban district" (Shixiaqu), which is the main urban area of the prefecture; we use the former in the main analysis but also collect the latter for the robustness analysis. The employment variables we describe below measure urban employment, but they are still measured both for the main urban district as well as the for the entire prefecture which may include other urban areas. The main data in the yearbooks are reported at the prefecture level but some urban district variables are also provided for county-level cities—a finer administrative division. We use county-level city data to complete some missing data in the prefecture-level variables where possible; this however does not apply to our main variable as it is not for urban districts.

Total urban employment data come from two Yearbook chapters: "People's Living Conditions and Social Security" and "Population, Labor Force, and Land Resources." The economic difference between them is not entirely clear. We use the former one, labeled "The Average Number of Staff and Workers", as its whole-prefecture version has by far the lowest number of strong year-to-year deviations which may indicate data quality issues. The other variable, "Persons Employed in Various Units at Year End", is used for robustness

[46] https://www.citypopulation.de/php/china-admin.php, accessed on November 20, 2018.

[47] Data for 2008-2015, excluding 2009 and 2011, are from http://oversea.cnki.net.proxy.uchicago.edu/kns55/default.aspx (accessed on January 23, 2019 via a University of Chicago portal). Data from 2009, 2011, 2016, and 2017, are from http://tongji.oversea.cnki.net/chn/navi/HomePage.aspx?id=N2018050234&name=YZGCA (accessed January 23, 2019). We checked that these sources agree in years where both are available.

checks in Appendix Table A2, together with the urban district versions of both variables. In that table we further use a measure of total rail ridership originating in the prefecture that includes HSR, traditional intercity, and intracity lines. Since it is only available until 2014, for this analysis we use the 2007–2014 change in ridership (instead of 2007–2016).

We finally apply a data cleaning procedure to all outcome variables used in the analysis. We first mark a prefecture-year observation as a one-off jump, and replace it with a missing value, if (*i*) the variable changes by more than twice in either direction relative to the previous non-missing value for the prefecture, (*ii*) it is followed by a change in the opposite direction that is at least 75% as large (in terms of log-changes), and (*iii*) the previous value has not been marked as a jump. We then mark an observation as a sustained change if condition (*i*) is satisfied but (*ii*) is not. We view the outcome change between 2007 and 2016 as valid only if neither 2007 nor 2016 are marked as jumps and there are no sustained changes in any year in between. For the main outcome variable this reduces the sample from 282 to the final set of 274 prefectures, but for other outcomes the sample reduction is more substantial.

A.2 Data for Section 4.2

Our application to simulated eligibility instruments uses a repeated cross-section of annual data from the American Community Survey (ACS; Ruggles et al. 2020). Our baseline estimation uses a representative 1% sample of individuals from 2013 and 2014 and we use the analogous 1% sample from 2012 to explore pre-trends. We restrict the sample to non-disabled adults (aged 21-64) residing in one of the 43 states eligible for Medicaid expansion under the ACA. To define this sample of states we follow Frean et al. (2017) in excluding "early expansion" states which had previously expanded Medicaid before 2013, as well as Massachusetts and Vermont who had previously made all adults with household income less than 138% FPL eligible. We also follow Frean et al. (2017) in designating 19 of these states as having expanded under the ACA in 2014, with 24 not expanding.[48]

In each year, we classify an individual as insured under Medicaid when she is covered by Medicaid or an equivalent government-assistance program, excluding Medicare and Veterans' Administration (VA) insurance. We classify an individual as having private insurance when she is covered by a plan purchased through an employer or union or when she purchases this private coverage directly. We further separate individuals covered employer-sponsored insurance and having private insurance that they purchased directly.

[48] Frean et al. (2017) study coverage effects over 2014-2015, designating 24 states as having expanded during this time, 21 states as having not expanded, and 6 states as expanding early. We use their classification system as of 2014, when only 19 of their 24 states have expanded, and additionally exclude two states (Massachusetts and Vermont) where the 2013 eligibility policy already made individuals with a household income of less than 138% FPL eligible for Medicaid.

Our simulated eligibility instrument is constructed by simulating the average Medicaid eligibility of a representative 10% sample of our analysis data under different state policies. Namely we use the representative sample to simulate two shares: that of individuals who would be eligible had their state expanded eligibility in 2014 to everyone under 138% of FPL, and that of individuals who would be eligible if their state kept 2013 policy intact. We assign the former share (24.5%) to all individuals in 2014 residing in expansion states and the latter share (11.6%) to individuals in 2014 residing in non-expansion states. For individuals in 2013, where there is no as-good-as-random variation, we fix z_ℓ^{CG} at 7.1%: the national share of eligible individuals under 2013 policies.

Our recentered IV is constructed by predicting the actual Medicaid eligibility of each individual. In 2013 we use actual 2013 eligibility policies, again following Frean et al. (2017). In 2014 we predict eligibility by combining information on the 2013 policies and a state's decision to expand. An individual is eligible for Medicaid in 2014 if either she was eligible under the 2013 policies of her state (whether or not the state expanded eligibility) or if her household income is below 138% FPL and her state expanded eligibility under the ACA. To compute the expected instrument we identify individuals who would have been eligible in 2014 if their state expanded but not otherwise (the "Exposed Sample"). Outside of this sample the expected instrument in 2014 is simply the individual's actual 2014 eligibility, while inside this sample the expected instrument is the fraction of states which expanded conditional on the governor's party. The 2013 expected eligibility IV is actual 2013 eligibility. Political party affiliation of state governors is determined as of December 2013,[49] and in all regressions we control for an indicator for state party affiliation (interacted with year indicators). In robustness checks we control for other time-interacted state characteristics: a state's 2012 median income or share insured under Medicaid (both from the ACS).

A.3 Robustness Checks for Section 4.2

This appendix describes additional analyses of recentered IV power gains in the simulated eligibility instrument application.

Pre-Trend Tests First, we estimate pre-trends corresponding to in each of the outcomes and specifications of Table 4 by exchanging the cross-section of individuals in 2014 with an equivalent cross-section in 2012. We continue to construct the endogenous variable and instrument as an individual's Medicaid eligibility in 2013 and 2014 for comparability, and also keep all controls unchanged.

Appendix Table A3 shows that we obtain relatively small pre-trend estimates across all specifications, with similar coefficients obtained by conventional simulated IV (odd columns)

[49] https://en.wikipedia.org/w/index.php?title=List_of_United_States_governors&oldid=587575534

and recentered IV (even columns). The same efficiency gains we document in Table 4 are found here, with significantly smaller 95% confidence intervals for the recentered IV (again obtained by a wild score bootstrap) which exclude zero for the take-up and crowd-out outcomes. We find no significant pre-trends in the employer-sponsored insurance outcome in Columns 5 and 6.

Alternative Assignment Processes Second, we explore the robustness of our estimates to alternative assumptions on the shock assignment process. Specifically we allow a state's decision to expand Medicaid coverage to depend not only on the party of its governor (as in our baseline specification) but additionally on the state's 2012 median income and 2012 level of Medicaid coverage. We accomplish this by including a quadratic in these three state characteristics (including their interaction), interacted with year indicators, in the control vector $c_{\ell t}$. This allows the expected instrument to depend flexibly on these characteristics in the exposed sample. Appendix Table A4 shows that we obtain virtually identical estimates, standard errors, and 95% confidence intervals.

Alternative IV Implementations Third, we apply alternative IV estimators implied by our framework. Recall that in the even-numbered columns of Table 4 we restrict the sample to individuals whose individual characteristics make them exposed to the expansion natural experiment in 2014. A different approach is to recenter the IV $z_{\ell t}$ by (or control for) the expected instrument $\mu_{\ell t}$, while keeping the full sample of individuals. Appendix Table A5 reports estimates from this approach for the three outcomes of interest. Panel B, which includes demographic controls, again finds much narrower confidence intervals relative to the simulated eligibility instrument. However, excluding these controls in Panel A yields an intriguing pattern: confidence intervals for the recentered IV are much wider than those of the simulated instrument.

In this section we explain how a combination of two factors generates the discrepancy between panels A and B of Appendix Table A5. First, the regression residuals are strongly correlated with the indicator for an individual being exposed to the expansion experiment, which is not controlled for in this regression. Second, exogenous shocks are assigned at the level of states, which include both exposed and non-exposed individuals. This discussion reveals why the problem does not arise when focusing on the non-exposed sample or when appropriate controls are included. We further relate this problem to Step 3 of the optimal instrument construction in Section 3.5.

For clarity of the theoretical discussion, we simplify the setup. First, we suppose that a single 2014 cross-section is available and state fixed effects are not included; we correspondingly drop the t subscript throughout. Second, we assume states only change eligibility as prescribed by their expansion decision, i.e. $x_{\ell t} = z_{\ell t}$. Finally, we assume that state decisions to expand are independent with a known propensity $\mathbb{E}[g_n \mid w]$ (e.g., as a function of the state governor's party). Thus, the recentered expansion indicator $\tilde{g}_n = g_n - \mathbb{E}[g_n \mid w]$ can be computed without permutations.[50]

Under these additional assumptions, using the recentered CG instrument is equivalent to using the recentered expansion indicator: $\tilde{z}_\ell^{CG} = \tilde{g}_{s_{\ell t}}$. The recentered IV only differs by setting $\tilde{z}_{\ell t}^{CG}$ to zero for the non-exposed sample: $\tilde{z}_\ell = z_\ell - \mathbb{E}[z_\ell \mid w] = f_\ell \tilde{g}_{s_\ell}$, where f_ℓ is an indicator for individual ℓ being in the exposed group. With $x_\ell = z_\ell$, the first stage can be written $x_\ell = \mu_\ell + f_\ell \tilde{g}_{s_\ell}$, where the expected instrument μ_ℓ equals 0 for individuals who are not eligible regardless of g_{s_ℓ}, 1 for those always eligible, and $\mathbb{E}[g_{s_\ell} \mid w]$ otherwise.

We now consider the variances of the two estimators, approximated as in the proof of Proposition 2: $\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG}\varepsilon_\ell^\perp\right] / \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG} x_\ell^\perp\right]^2$ and $\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\varepsilon_\ell^\perp\right] / \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp\right]^2$, respectively, where $\perp$ denotes the in-sample projection residual on the control variables (including a constant). We focus our attention on the numerators of these expressions because the first-stage covariances in the denominator are asymptotically equivalent (and equal in finite samples without controls).[51] For simplicity of exposition we also consider an individual's state of residence s_ℓ as fixed. Letting $L_n = \sum_\ell \mathbf{1}[s_\ell = n]$ denotes the (fixed) number of individuals in each state n, it can then be shown that

$$\frac{\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\varepsilon_\ell^\perp\right]}{\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG}\varepsilon_\ell^\perp\right]} = \frac{\sum_n \left(\frac{L_n}{L}\right)^2 \operatorname{Var}[\tilde{g}_n]\, \mathbb{E}\left[e_{SEIV,n}^2\right]}{\sum_n \left(\frac{L_n}{L}\right)^2 \operatorname{Var}[\tilde{g}_n]\, \mathbb{E}\left[e_{CG,n}^2\right]}, \tag{12}$$

where $e_{SEIV,n} = \frac{1}{L_n}\sum_{\ell:\, s_\ell = n} \varepsilon_\ell^\perp f_\ell$ is the sum of residuals of *exposed* individuals in state n (normalized by L_n), while $e_{CG,n} = \sum_{\ell:\, s_\ell = n} \varepsilon_\ell^\perp$ averages over *all* observations in the state.[52]

Equation (12) shows that the recentered IV delivers power gains relative to the simulated instrument approach whenever the normalized sum of residuals is closer to zero for a typical state, in the mean-squared sense, when restricting to exposed individuals. The restricted sum has fewer summands, working in favor of the recentered IV. If the expansion shocks were assigned at the individual level, without state clustering, this would guarantee that

[50] Formally, we assume that w does not include $\Pi(g)$. Under this assumption, $\tilde{g}_n$ is independent across states conditionally on w, simplifying the analysis.

[51] Namely, since f_ℓ is binary, $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG} x_\ell\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{g}_{s_\ell}(\mu_\ell + f_\ell \tilde{g}_{s_\ell})\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell \tilde{g}_{s_\ell}^2\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell \tilde{g}_{s_\ell}(\mu_\ell + f_\ell \tilde{g}_{s_\ell})\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right]$. With conventional controls this equality holds asymptotically, since the difference between x_ℓ and $x_\ell^\perp$ is uncorrelated with $\tilde{z}_\ell$.

[52] Namely $\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\varepsilon_\ell^\perp\right] = \sum_n \left(\frac{L_n}{L}\right)^2 \mathbb{E}\left[\left(\frac{1}{L_n}\sum_{\ell:\, s_\ell = n} \tilde{z}_\ell\varepsilon_\ell^\perp\right)^2\right] = \sum_n \left(\frac{L_n}{L}\right)^2 \mathbb{E}\left[\tilde{g}_n^2 \cdot \left(\frac{1}{L_n}\sum_{\ell:\, s_\ell = n} f_\ell\varepsilon_\ell^\perp\right)^2\right] = \sum_n \left(\frac{L_n}{L}\right)^2 \operatorname{Var}[\tilde{g}_n]\, \mathbb{E}\left[e_{SEIV,n}^2\right]$, since $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\varepsilon_\ell^\perp\right] = 0$ by Proposition 1, and similarly for $\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell^{CG}\varepsilon_\ell^\perp\right]$.

the recentered IV is more efficient (since $e_{SEIV,n} = e_{CG,n} = \varepsilon_\ell^\perp$ in that case).

However, this simplified example shows that the recentered IV is likely to deliver a power loss when the shocks g_n are clustered and $\varepsilon_\ell^\perp$ is strongly correlated with the indicator of exposed sample f_ℓ (i.e., exposed individuals have systematically different residuals, and f_ℓ is not controlled for). To see this simply, suppose $\mathbb{E}\left[\varepsilon_\ell^\perp \mid f_\ell = 1, w\right] = \alpha \neq 0$ for all ℓ. In this scenario $e_{SEIV,n}$ is not mean-zero, even on average across states, which potentially yields a high mean-squared residual:

$$\begin{aligned}\mathbb{E}\left[e_{SEIV,n}\right] = \mathbb{E}\left[\mathbb{E}\left[e_{SEIV,n} \mid w\right]\right] &= \mathbb{E}\left[\frac{1}{L_n}\sum_{\ell:\ s_\ell = n} \mathbb{E}\left[\varepsilon_\ell^\perp f_\ell \mid w\right]\right] \\ &= \mathbb{E}\left[\frac{1}{L_n}\sum_{\ell:\ s_\ell = n} \mathbb{E}\left[\varepsilon_\ell^\perp \mid f_\ell = 1, w\right] f_\ell\right] = \alpha \cdot \mathbb{E}\left[\frac{\sum_{\ell:\ s_\ell = n} f_\ell}{L_n}\right] \neq 0. \qquad (13)\end{aligned}$$

The simulated instrument, which does not condition on $f_\ell = 1$, does not suffer from this problem since $\varepsilon_\ell^\perp$ is mean-zero in the sample. Another interpretation of this problem is that in this case the sums of residuals over the exposed and non-exposed individuals of a given state will tend to have opposite signs, increasing efficiency of the Currie-Gruber instrument that uses both subsamples.

The predictions of this discussion are borne out in the data. In Panel C of Table A5 we verify that the confidence interval of recentered IV become dramatically narrowed with a single control of f_ℓ (interacted with the 2014 dummy appropriately for difference-in-differences).[53] Moreover, demographic controls in Panel B of Table A5 capture most of the variation in f_ℓ, delivering similar results. Our recentered IV specifications in the main text, by restricting the sample to the exposed individuals, effectively control for state dummies interacted with f_ℓ and achieve the best efficiency properties.

We note that here controlling for the exposed sample indicator is closely related to our third step in constructing the optimal recentered IV, discussed in Section 3.5: this control plays the role of the predetermined predictors of the residual, ψ. Our application therefore highlights that in general there is no guarantee of an efficiency gain from improving the first stage with a recentered IV (i.e., performing Steps 1 and 2) if Steps 3 and 4 are not feasible.

Monte Carlo Simulation Finally, we verify large and pervasive power gains from using the recentered IV in a Monte Carlo study, in which the true causal effect and the shock assignment process are known. We draw 999 counterfactual state expansion decisions by choosing random sets of 8 Republican- and 11 Democratic-controlled states as expansion states and use these shocks to compute counterfactual instruments $\tilde{z}_{\ell t}^{CG}$ and $\tilde{z}_{\ell t}$. We do not

[53] The efficiency of the IV that controls for $\mu_{\ell t}$ is lower because this control is not interacted with f_ℓ.

specify a model of the first stage (i.e., which exact policies states would have implemented if they randomly changed their decision to adopt the ACA Medicaid expansion), instead imagining that states either expand to 138% FPL or keep their 2013 policy. We therefore use $\tilde{x}_{\ell t} = \tilde{z}_{\ell t}$ as the endogenous variable. Finally, for the Medicaid take-up and ESI crowd-out outcomes we take the second-stage residuals $\varepsilon^*_{\ell t}$ from Columns 2 and 6 of Table 4, panel A. These outcomes are unrelated to the endogenous variable by design, corresponding to the true causal effect of zero for all individuals, while keeping the correlation structure from the actual data. With these generated data, we re-estimate equation (11) with the fixed effects and controls as in our baseline implementation in Panel A of Table 4. By design, both sets of estimates should be centered at the true effects of zero, while we expect the recentered IV procedure to systematically reject a larger set of alternative hypotheses.

Figure A7 first shows the simulated distribution of simulated and recentered IV estimates from this exercise. Both estimators are approximately unbiased, with both distributions in both panels centered around the true effects of zero. However, consistent with the dramatically shorter confidence intervals in Table 4, the distribution of recentered IV coefficients is dramatically tighter around this mean. The estimate standard deviation falls from 0.014 to 0.006 as we move from the simulated IV to recentered IV in Panel A, with a larger decline from 0.020 to 0.007 in Panel B. With minimal bias, these correspond to simulated root mean-squared error reductions of 58.5% and 66.5% with the recentered IV, respectively.

Figure A8 shows that these reductions in estimator variance translate to increased rejection rates of false null hypotheses for both outcomes, while also suggesting the wild bootstrap 95% confidence intervals in Table 4 have approximately correct size. Away from the true null hypothesis of zero the recentered IV power curve is much more steeply sloping, with uniformly higher rejection rates. With the Medicaid take-up outcome, for example, the recentered IV is found to reject coefficients outside the range of $[-0.018, 0.017]$ with probability of at least 0.8, while the simulated IV only has such high power outside a nearly three times as long range, of $[-0.042, 0.056]$. For the ESI crowd-out outcome this contrast in minimum detectable effects is even starker, at $[-0.022, 0.018]$ for the recentered IV versus $[-0.073, 0.051]$ for the simulated IV.

B Proposition Proofs

B.1 Proof of Proposition 1

For the recentered IV regression,

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\varepsilon_{\ell}\right]=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\mathbb{E}\left[\varepsilon_{\ell}\mid g,w\right]\right]=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\mathbb{E}\left[\varepsilon_{\ell}\mid w\right]\right]$$
$$=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\mathbb{E}\left[\tilde{z}_{\ell}\mid w\right]\mathbb{E}\left[\varepsilon_{\ell}\mid w\right]\right]=0. \tag{14}$$

The first and third equalities follow from the law of iterated expectations. The second equality follows from Assumption 2(i), and the final equality follows from the fact that $\mathbb{E}\left[\tilde{z}_{\ell}\mid w\right]=0$.

The alternative approach that regression-adjusts by μ_{ℓ} while using the unadjusted z_{ℓ} as an instrument identifies β when

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\varepsilon_{\ell}^{\perp}\right]=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}y_{\ell}^{\perp}\right]-\beta\cdot\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}x_{\ell}^{\perp}\right]=0, \tag{15}$$

by the Frisch-Waugh-Lovell theorem. Here $\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\varepsilon_{\ell}^{\perp}\right]=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}(z_{\ell}-\mu_{\ell})\varepsilon_{\ell}^{\perp}\right]$ since $\frac{1}{L}\sum_{\ell}\mu_{\ell}\varepsilon_{\ell}^{\perp}=0$ by construction. Moreover, in matrix form,

$$\mathbb{E}\left[\varepsilon^{\perp}\mid g,w\right]=(I-P_{\mu})\,\mathbb{E}\left[\varepsilon\mid g,w\right]=(I-P_{\mu})\,\mathbb{E}\left[\varepsilon\mid w\right]=\mathbb{E}\left[\varepsilon^{\perp}\mid w\right], \tag{16}$$

where P_{μ} denotes the sample projection matrix for μ_{ℓ} and a constant (which is fixed conditional on w). Following the same steps as before, we thus have

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_{\ell}\varepsilon_{\ell}^{\perp}\right]=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}(z_{\ell}-\mu_{\ell})\mathbb{E}\left[\varepsilon_{\ell}^{\perp}\mid g,w\right]\right]$$
$$=\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\mathbb{E}\left[\tilde{z}_{\ell}\mid w\right]\mathbb{E}\left[\varepsilon_{\ell}^{\perp}\mid w\right]\right]=0, \tag{17}$$

showing that the alternative μ_{ℓ}-controlled regression also identifies β.

B.2 Proof of Lemma 2

The Hodges-Lehmann estimator of interest solves:

$$\frac{1}{L}\sum_{\ell} f_{\ell}\left(g,w\right)\left(y_{\ell}-bx_{\ell}\right) = \mathbb{E}\left[\frac{1}{L}\sum_{\ell} f_{\ell}\left(g^{*},w\right)\left(y_{\ell}-bx_{\ell}\right) \mid w,y,x\right]$$
$$= \frac{1}{L}\sum_{\ell}\mu_{\ell}\left(y_{\ell}-bx_{\ell}\right), \tag{18}$$

since $g^{*}\sim G\left(\cdot\mid w\right)\mid\left(y,x,w\right)$. This linear equation has a unique solution:

$$\hat{\beta} = \frac{\frac{1}{L}\sum_{\ell}\left(f_{\ell}(g,w)-\mu_{\ell}\right)y_{\ell}}{\frac{1}{L}\sum_{\ell}\left(f_{\ell}(g,w)-\mu_{\ell}\right)x_{\ell}}, \tag{19}$$

which coincides with the recentered IV estimator.

For the statistic that uses the μ_{ℓ}-residualized outcome and treatment we similarly have

$$\frac{1}{L}\sum_{\ell} f_{\ell}\left(g,w\right)\left(y_{\ell}^{\perp}-bx_{\ell}^{\perp}\right) = \mathbb{E}\left[\frac{1}{L}\sum_{\ell} f_{\ell}\left(g^{*},w\right)\left(y_{\ell}^{\perp}-bx_{\ell}^{\perp}\right) \mid w,y,x\right]$$
$$= \frac{1}{L}\sum_{\ell}\mu_{\ell}\left(y_{\ell}^{\perp}-bx_{\ell}^{\perp}\right). \tag{20}$$

The resulting estimator $\frac{\frac{1}{L}\sum_{\ell}(f_{\ell}(g,w)-\mu_{\ell})y_{\ell}^{\perp}}{\frac{1}{L}\sum_{\ell}(f_{\ell}(g,w)-\mu_{\ell})x_{\ell}^{\perp}} = \frac{\frac{1}{L}\sum_{\ell}f_{\ell}(g,w)y_{\ell}^{\perp}}{\frac{1}{L}\sum_{\ell}f_{\ell}(g,w)x_{\ell}^{\perp}}$ equals the recentered IV estimator with the instrument z_{ℓ} and controlling for μ_{ℓ}, as in the Appendix B.1 proof.

B.3 Proof of Asymptotic Efficiency and Lemma 3

Here we formalize the efficiency result discussed in Section 3.5 and then prove it together with Lemma 3.

Asymptotic Efficiency. We first define some useful asymptotic concepts. For a non-random sequence $r_L\to\infty$, we say that an estimator $\tilde{\beta}$ converges to β at rate r_L when $r_L\left(\tilde{\beta}-\beta\right)$ converges to a non-degenerate distribution with zero mean and variance $V>0$ as $L\to\infty$. We refer to V as the asymptotic variance of $\tilde{\beta}$, and say that the convergence rate r_L^{*} is faster than r_L when $\lim_{L\to\infty}\frac{r_L}{r_L^{*}}=0$.[54] We consider IV estimators of the form $\tilde{\beta}=\frac{1}{L}\tilde{z}'y/\frac{1}{L}\tilde{z}'x$ where $\tilde{z}=f\left(g,w\right)$ for an $L\times 1$ vector of functions f such that $\mathbb{E}\left[\tilde{z}\mid w\right]=0$; the last condition requires that $\tilde{z}$ is a recentered instrument. We say that $\tilde{\beta}$ is "regular" if

[54]In general, the asymptotic variance concept is useful when the limiting distribution of $\tilde{\beta}$ is normal. However, it can be considered more broadly; in particular, a researcher with a quadratic loss function will generally value reductions in V outside the normal case. We therefore do not restrict the shape of the asymptotic distribution until this is required in Proposition S4.

(i) it converges to β at some rate r_L, *(ii)* it has an asymptotic first stage, i.e. $\frac{1}{L}\tilde{z}'x \xrightarrow{p} M$ for some $M \neq 0$, and *(iii)* the sequences of $\frac{1}{L}\tilde{z}'x$ and $\left(r_L \frac{1}{L}\tilde{z}'\varepsilon\right)^2$ are uniformly integrable. These definitions yield the following result, which characterizes the efficient instrument under weak conditions.[55] In Appendix C.4 we establish further conditions under which $\tilde{z}^*$ maximizes the asymptotic power of our preferred RI test by minimizing the asymptotic variance of the associated IV estimator.

Proposition 2. *Suppose Assumption 2 holds and $\mathbb{E}[\varepsilon\varepsilon' \mid w]$ is almost-surely invertible. Consider the recentered z^* defined by equation (9). Then if the associated estimator $\beta^* = \frac{1}{L}z^{*\prime}y/\frac{1}{L}z^{*\prime}x$ is regular, it has the smallest asymptotic variance of all regular recentered IV estimators $\tilde{\beta}$: there is no $\tilde{\beta}$ that converges at a rate faster than that of β^*, and any $\tilde{\beta}$ converging at the same rate has an asymptotic variance at least as large as that of β^*.*

Proof. Consider some recentered IV $\tilde{z}$ associated with a regular estimator $\tilde{\beta}$ that converges at rate $\tilde{r}_L$ to an asymptotic distribution $\tilde{\mathcal{D}}$ with variance $\tilde{V}$. Uniform integrability of $\frac{1}{L}\tilde{z}'x$ implies that $\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right] \to M$. Then, by the continuous mapping theorem,

$$\tilde{r}_L \frac{\frac{1}{L}\tilde{z}'\varepsilon}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]} = \tilde{r}_L\left(\tilde{\beta} - \beta\right) \cdot \frac{\frac{1}{L}\tilde{z}'x}{M} \cdot \frac{M}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]} \Rightarrow \tilde{\mathcal{D}}, \tag{21}$$

as $r_L\left(\tilde{\beta} - \beta\right) \Rightarrow \tilde{\mathcal{D}}$, $\frac{\frac{1}{L}\tilde{z}'x}{M} \xrightarrow{p} 1$, and $\frac{M}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]} \to 1$. Furthermore, by the uniform integrability of $\left(r_L \frac{1}{L}\tilde{z}'\varepsilon\right)^2$,

$$\operatorname{Var}\left[\tilde{r}_L \frac{\frac{1}{L}\tilde{z}'\varepsilon}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]}\right] = \tilde{r}_L^2 \frac{\operatorname{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2} \to \tilde{V}. \tag{22}$$

The same argument applies to β^*:

$$\operatorname{Var}\left[r_L^* \frac{\frac{1}{L}z^{*\prime}\varepsilon}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]}\right] = {r_L^*}^2 \frac{\operatorname{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2} \to V^*, \tag{23}$$

where r_L^* and V^* denote its convergence rate and asymptotic variance, respectively. Combining the two statements yields

$$\frac{\tilde{r}_L^2/\tilde{V}}{{r_L^*}^2/V^*} \cdot \frac{\operatorname{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]/\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2}{\operatorname{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]/\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2} \to 1. \tag{24}$$

□

[55] If $\mathbb{E}[\varepsilon\varepsilon' \mid w]$ were not invertible the unobservables would be unusually dependent, in that there would exist a function $c(w)$ satisfying $c(w)'\varepsilon = 0$ and revealing β almost-surely provided $c(w)'x \neq 0$.

We prove below that

$$\frac{\operatorname{Var}\left[\frac{1}{L}\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}\tilde{z}'x\right]^2} \geq \frac{\operatorname{Var}\left[\frac{1}{L}z^{*\prime}\varepsilon\right]}{\mathbb{E}\left[\frac{1}{L}z^{*\prime}x\right]^2}. \tag{25}$$

whenever the denominators on both sides are not equal to zero (which holds for large enough L, since both $\tilde{z}$ and z^* have asymptotic first-stages). This concludes the proof, since (24) and (25) jointly imply that

$$\limsup_{L\to\infty} \frac{\tilde{r}_L^2/\tilde{V}}{{r_L^*}^2/V^*} \leq 1. \tag{26}$$

This, in turn, implies that $\lim_{L\to\infty} \frac{\tilde{r}_L}{r_L^*} \neq \infty$ and, if $\tilde{r}_L = r_L^*$, then $\tilde{V} \geq V^*$.

To establish (25), note that by the law of iterated expectations and Assumption 2,

$$\begin{aligned}\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon' z^*\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{z}'\varepsilon\varepsilon' z^* \mid g, w\right]\right] = \mathbb{E}\left[\tilde{z}'\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right] \\ &= \mathbb{E}\left[\tilde{z}'\mathbb{E}\left[x \mid g, w\right]\right] = \mathbb{E}\left[\tilde{z}'x\right],\end{aligned}$$

where the third line uses the fact that $\mathbb{E}\left[\tilde{z}'\mathbb{E}\left[x \mid w\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{z}' \mid w\right]\mathbb{E}\left[x \mid w\right]\right] = 0$ since $\mathbb{E}\left[\tilde{z}' \mid w\right] = 0$, and the fourth line follows because $\tilde{z}$ is non-stochastic given g and w. For $\tilde{z} = z^*$, this shows that

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

It also shows that with

$$U = \frac{\tilde{z}'\varepsilon}{\mathbb{E}\left[\tilde{z}'x\right]} - \frac{(z^*)'\varepsilon.}{\mathbb{E}\left[(z^*)'x\right]} \tag{28}$$

we have

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

implying equation (25).

Proof of Lemma 3 By the law of total variance, $\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right] = \Omega + \psi\psi'$. Since $\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]$ is almost-surely invertible, Ω is also invertible since $\psi\psi'$ has a rank of one (assuming $L > 1$). By the Sherman-Morrison formula in linear algebra,

$$\left(\Omega + \psi\psi'\right)^{-1} = \Omega^{-1} - \Omega^{-1}\psi\frac{\psi'\Omega^{-1}}{1 + \psi'\Omega^{-1}\psi}. \tag{29}$$

Thus, as claimed,

$$z^* = \left(\Omega + \psi\psi'\right)^{-1}\tilde{z} = \Omega^{-1}\left(\tilde{z} - \frac{\psi'\Omega^{-1}\tilde{z}}{1 + \psi'\Omega^{-1}\psi}\psi\right) = \Omega^{-1}\left(\tilde{z} - \rho\nu\psi\right).$$

References

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers. 2017. "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act." *Journal of Health Economics* 53:72–86.

Lawrence, Martha, Richard Bullock, and Ziming Liu. 2019. *China's High-Speed Rail Development.* Washington, D.C.: World Bank.

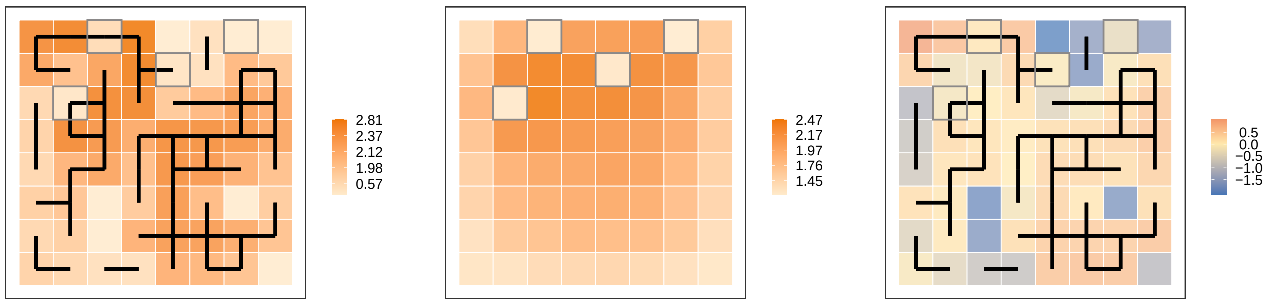
Lin, Yatang. 2017. "Travel costs and urban specialization patterns: Evidence from China's high speed railway system." *Journal of Urban Economics* 98:98–123.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. *IPUMS USA: Version 10.0.* Minneapolis, MN.

Appendix Figures and Tables

Figure A1: Market Access Growth with Unequal Population

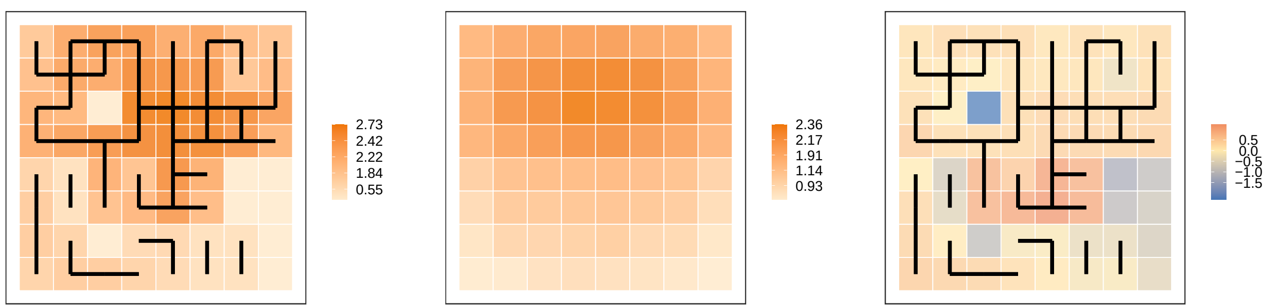
A. Line Construction and MA Growth | B. Expected MA Growth | C. Recentered MA Growth



Notes: This figure parallels Figure 1, except with the population of four highlighted regions ten times larger than of all others.

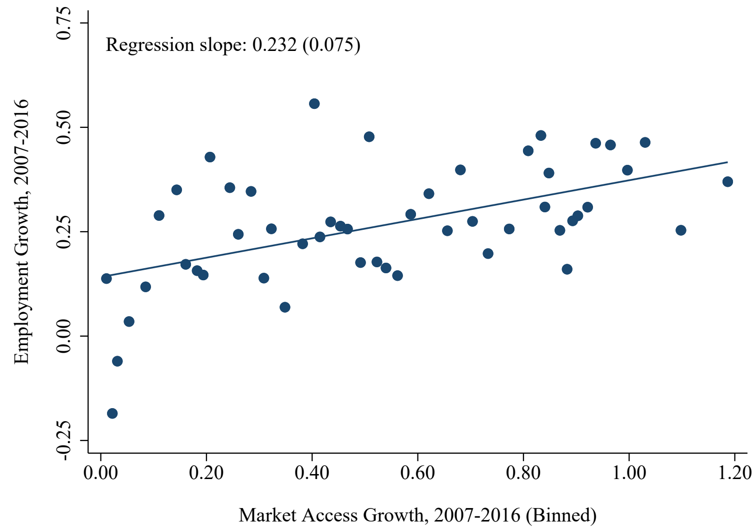
Figure A2: Market Access Growth with Non-Uniform Line Density

A. Line Construction and MA Growth | B. Expected MA Growth | C. Recentered MA Growth



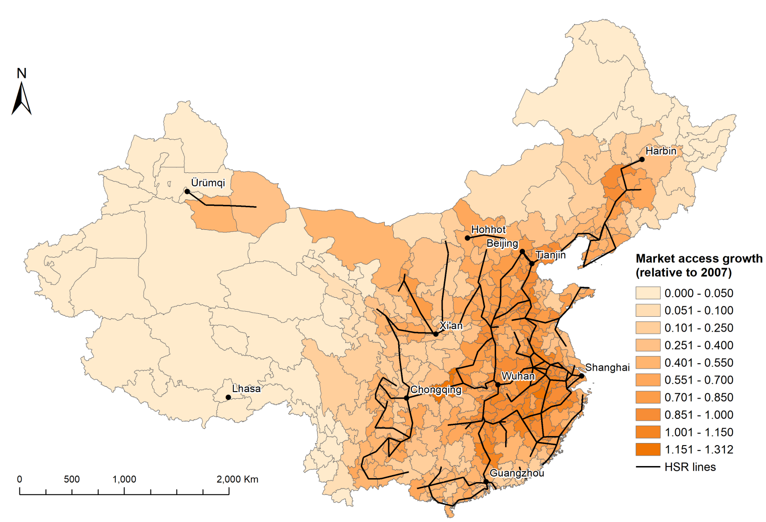
Notes: This figure parallels Figure 1, except assuming that pairs of adjacent regions in the northern half of the island are connected with probability 2/3, and others are connected with probability 1/3.

Figure A3: Employment Growth and Market Access Growth



Notes: This figure shows a binned scatterplot of employment growth against MA growth across 274 prefectures in China in 2007–16. Fifty bins of approximately equal size are shown. The regression line of best fit is also indicated, along with the coefficient and spatial-clustered standard error.

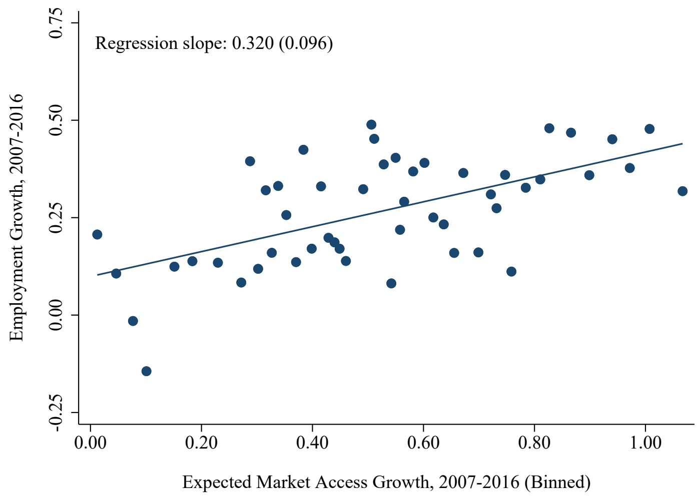
Figure A4: Simulated HSR Lines and Market Access Growth



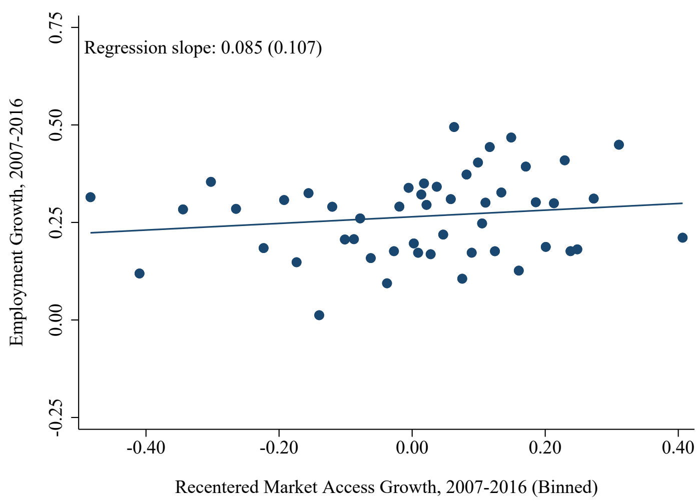
Notes: This figure shows an example map of simulated Chinese HSR lines and market access growth over 2007–2016, obtained by permuting the opening status of built and planned lines with the same number of cross-prefecture links as described in Section 4.1.

Figure A5: Employment Growth and Expected/Recentered Market Access Growth

A. Expected Market Access Growth

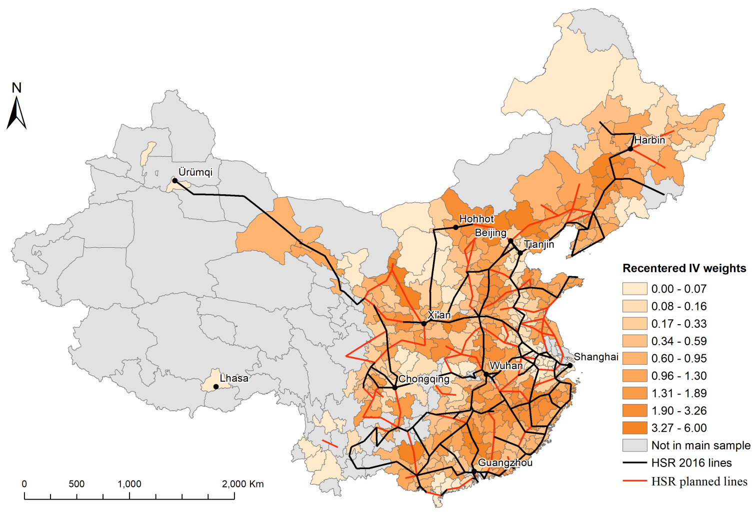


B. Recentered Market Access Growth



Notes: This figure shows binned scatterplots of employment growth against the expected and recentered MA growth across 274 prefectures in China in 2007–16. Expected and recentered MA are constructed by permuting the opening status of built and planned lines with the same number of cross-prefecture links. Fifty bins of approximately equal size are shown. Regression lines of best fit are indicated along with coefficients and spatial-clustered standard errors.

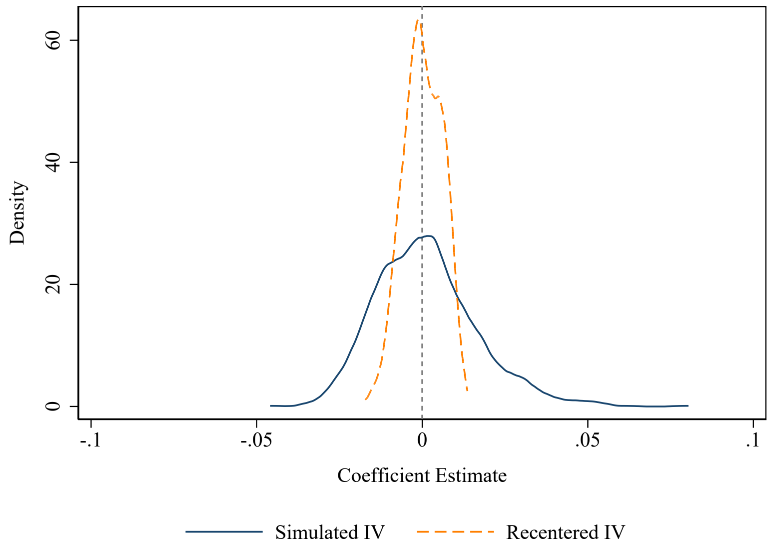
Figure A6: Recentered Market Access Growth Regression Weights



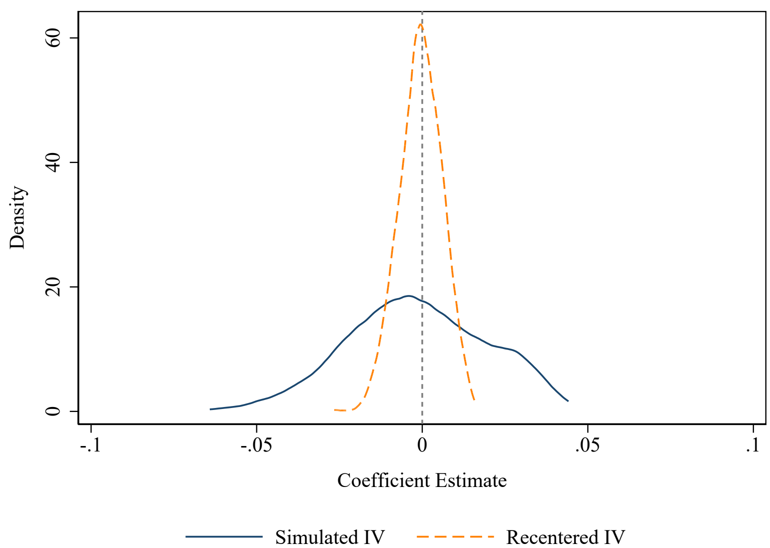
Notes: This figure shows the implied weights the recentered market access growth IV regression (of employment growth on observed market access growth) puts on different Chinese prefectures. Weights are given by the conditional variance of $\tilde{z}_\ell \mid w$ (see Corollary S1 in Appendix C.1), computed by permuting the opening status of built and planned lines with the same number of cross-prefecture links as described in Section 4.1. The weights are normalized to average to one in the sample. Prefectures in gray are missing employment growth data.

Figure A7: Medicaid Eligibility Effects: Simulated Distributions of Simulated and Recentered IV Estimators

A. Has Medicaid



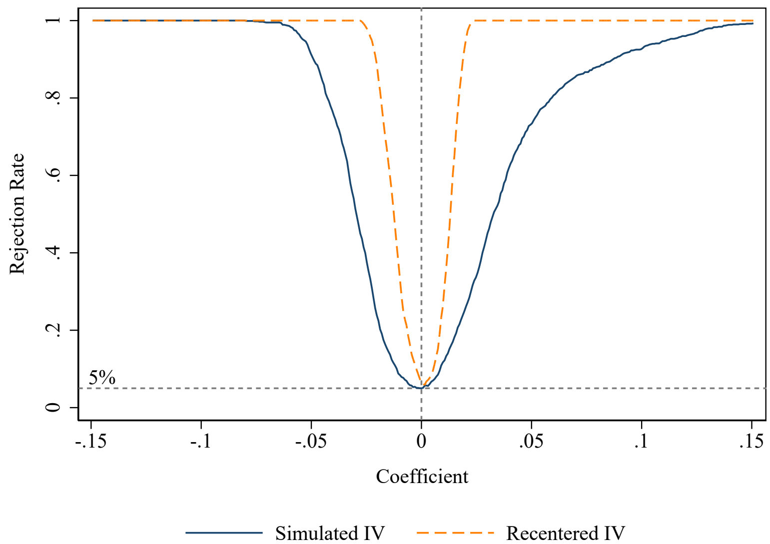
B. Has Employer-Sponsored Insurance



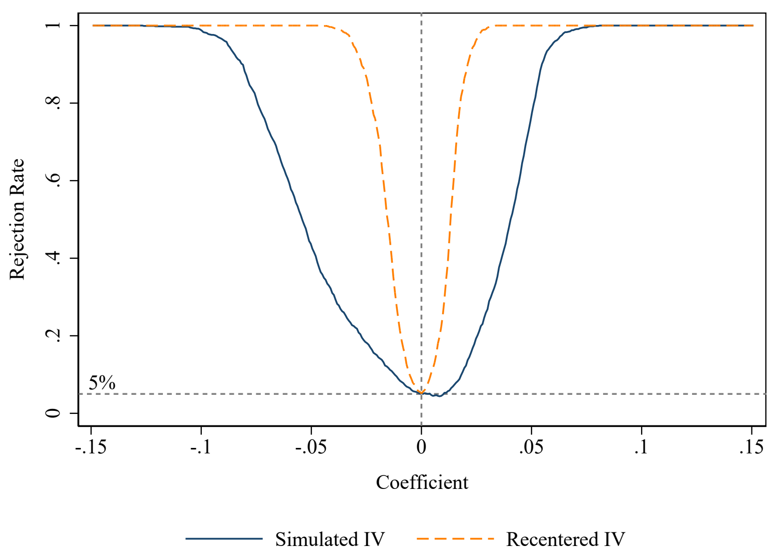
Notes: This figure plots the simulated distributions of IV coefficients from regressions of different measures of health insurance coverage on Medicaid eligibility, instrumented by one of two IVs described in the text: a simulated eligibility instrument and a recentered prediction of Medicaid eligibility. See Appendix A.3 for a description of the data-generating process and instruments. The true effect of zero in both panels is indicated by the dashed vertical line.

Figure A8: Medicaid Eligibility Effects: Simulated Size and Power of Simulated and Recentered IVs

A. Has Medicaid



B. Has Employer-Sponsored Insurance



Notes: This figure plots the simulated rejection rates of IV procedures regressing different measures of health insurance coverage on Medicaid eligibility, instrumented by one of two IVs described in the text: a simulated eligibility instrument and a recentered prediction of Medicare eligibility. See Appendix A.3 for a description of the data-generating process and instruments. Rejection rates are for nominal 5%-level tests of each coefficient based on wild score bootstraps, clustered by state. The true effect of zero in both panels is indicated by the dashed vertical line. The nominal 5% level of the tests is indicated by the dashed horizontal lines.

Table A1: Effects of Market Access: Alternative Specifications

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
Panel A. Using Leave-One-Out Market Access (N=274)			
Market Access Growth	0.229	0.081	0.070
	(0.078)	(0.104)	(0.103)
		[-0.360, 0.357]	[-0.124,216]
Expected Market Access Growth			0.207
			(0.118)
Panel B. Dropping Province Capitals (N=247)			
Market Access Growth	0.215	0.068	0.060
	(0.078)	(0.104)	(0.099)
		[-0.303, 0.321]	[-0.202, 0.320]
Expected Market Access Growth			0.303
			(0.097)
Panel C. Using HSR Connectivity (N=274)			
Connectivity Growth	0.155	0.051	0.049
	(0.049)	(0.057)	(0.056)
		[-0.037, 0.149]	[-0.041, 0.145]
Expected Connectivity Growth			0.257
			(0.071)
Panel D. Adding Province Fixed Effects (N=268)			
Market Access Growth	0.108	0.099	0.097
	(0.046)	(0.070)	(0.079)
		[-0.014, 0.268]	[-0.018, 0.270]
Expected Market Access Growth			0.121
			(0.071)
Recentered	No	Yes	Yes

Notes: This table reports coefficients from alternative specifications of the regressions in Table 1. Panel A uses leave-one-out MA growth as an instrument for full-sample MA growth (see Appendix A.1 for variable definitions). In Column 1 this instrument is unadjusted; in Column 2 it is recentered by permuting the opening status of built and planned lines with the same number of cross-prefecture links. Column 3 instead controls for expected leave-one-out MA growth (given by the same HSR counterfactuals) while instrumenting with unadjusted leave-one-out MA growth. Panels B and D use the OLS and IV specifications from Table 1, Panel A, but either dropping province capitals or including province fixed effects. In Panel D we drop singleton provinces (including province-level cities such as Beijing and Tianjin). Panel C replaces MA growth with the 2007–2016 change in the indicator for whether a prefecture is connected by high-speed rail (in the endogenous variable, instrument, and the Column 3 control). Standard errors which allow for linearly decaying spatial correlation (up to a bandwidth of 500km) are reported in parentheses. 95% RI confidence intervals based on the HSR counterfactuals are reported in brackets, and the sample size N is reported next to the panel titles.

Table A2: Effects of Market Access on Additional Outcomes

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
Panel A. Average Number of Employed Staff and Workers, Urban District (N=262)			
Market Access Growth	0.179	0.039	0.021
	(0.080)	(0.113)	(0.113)
		[-0.380, 0.272]	[-0.292, 0.273]
Expected Market Access Growth			0.268
			(0.097)
Panel B. Persons Employed in Various Units at Year End, Whole City (N=267)			
Market Access Growth	0.198	0.098	0.090
	(0.096)	(0.103)	(0.105)
		[-0.334, 0.391]	[-0.238, 0.386]
Expected Market Access Growth			0.255
			(0.122)
Panel C. Persons Employed in Various Units at Year End, Urban District (N=263)			
Market Access Growth	0.169	0.030	0.014
	(0.084)	(0.110)	(0.108)
		[-0.412, 0.286]	[-0.332, 0.290]
Expected Market Access Growth			0.256
			(0.107)
Panel D. Railway Passenger Traffic, Whole City (N=191)			
Market Access Growth	0.366	0.282	0.270
	(0.104)	(0.171)	(0.185)
		[-0.207, 0.771]	[-0.157, 0.773]
Expected Market Access Growth			0.421
			(0.139)
Recentered	No	Yes	Yes

Notes: This table reports coefficients from regressing different measures of employment growth and rail ridership on MA growth in Chinese prefectures. Panels A, B, and C use employment growth from 2007–2016, while Panel D uses rail ridership growth from 2007–2014 (see Appendix A.1 for variable definitions). The specifications parallel those of Table 1. Standard errors which allow for linearly decaying spatial correlation (up to a bandwidth of 500km) are reported in parentheses. 95% confidence intervals based on the same HSR assignment process are reported in brackets, and the sample size N is reported next to the panel titles.

Table A3: Medicaid Eligibility Pre-Trends, Simulated and Recentered IV Estimates

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV	Recentered IV	Simulated IV	Recentered IV	Simulated IV	Recentered IV
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Baseline Controls						
Eligibility	-0.022	-0.020	0.015	0.011	0.011	0.007
	(0.009)	(0.004)	(0.017)	(0.004)	(0.017)	(0.005)
	[-0.042,0.009]	[-0.028,-0.008]	[-0.021,0.071]	[0.003,0.020]	[-0.026,0.059]	[-0.005,0.020]
Panel B. With Demographics $\times$ Post						
Eligibility	-0.023	-0.020	0.019	0.014	0.016	0.011
	(0.010)	(0.004)	(0.014)	(0.004)	(0.016)	(0.005)
	[-0.040,0.012]	[-0.027,-0.009]	[-0.022,0.056]	[0.005,0.022]	[-0.029,0.049]	[-0.002,0.022]
Exposed Sample	N	Y	N	Y	N	Y
States	43	43	43	43	43	43
Individuals	2,400,142	425,112	2,400,142	425,112	2,400,142	425,112

Notes: This table reports coefficients from IV regressions of different measures of health insurance coverage in 2012–13 on 2014 Medicaid eligibility, instrumented by one of the two IVs described in the text: a simulated eligibility instrument and a recentered prediction of Medicaid eligibility. Columns 1, 3, and 5 estimate regressions in the full sample of individuals in 2012 or 2013, while Columns 2, 4, and 6 restrict to the sample of individuals whose individual characteristics make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year; the regressions in Panel B additionally control for deciles of household income, interacted with indicators for parental and work status and year. Conventional state-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Table A4: Recentered IV Estimates of Medicaid Eligibility Effects, Alternative Assignment Processes

	Has Medicaid	Has Private Insurance	Has Employer-Sponsored Insurance
	(1)	(2)	(3)
Panel A. Expansion Based on State Governor's Party and Median Income			
Eligibility	0.077	−0.018	−0.005
	(0.011)	(0.008)	(0.006)
	[0.053,0.092]	[-0.042,0.002]	[-0.019,0.011]
Panel B. Based on Governor's Party, Income, and Baseline Medicaid Coverage			
Eligibility	0.076	−0.023	−0.009
	(0.011)	(0.007)	(0.005)
	[0.054,0.102]	[-0.040,-0.008]	[-0.020,0.003]
Exposed Sample	Y	Y	Y
States	43	43	43
Individuals	421,042	421,042	421,042

Notes: This table reports coefficients from IV regressions of different measures of health insurance coverage on Medicaid eligibility, instrumented by a recentered prediction of Medicaid eligibility. Estimation is restricted to the sample of individuals whose individual characteristics make them exposed to the partial ACA Medicaid expansion in 2014. All regressions control for state and year fixed effects, an indicator for Republican-governed states interacted with year, and 2012 state median income interacted with year; the regressions in Panel B additionally control for 2012 state Medicaid coverage rates interacted with year. Conventional state-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Table A5: Recentered IV Estimates of Medicaid Eligibility Effects, Including Non-Exposed Individuals

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Recentered (1)	Controlled (2)	Recentered (3)	Controlled (4)	Recentered (5)	Controlled (6)
Panel A. Baseline Controls						
Eligibility	0.032	0.071	0.193	0.098	0.208	0.110
	(0.085)	(0.044)	(0.290)	(0.168)	(0.301)	(0.173)
	[-0.441,0.148]	[-0.088,0.140]	[-0.223,1.805]	[-0.170,0.675]	[-0.205,2.023]	[-0.174,0.745]
Panel B. With Demographics × Post						
Eligibility	0.116	0.114	−0.029	−0.029	−0.018	−0.018
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.014)
	[0.092,0.151]	[0.082,0.147]	[-0.051,0.002]	[-0.053,0.012]	[-0.040,0.013]	[-0.041,0.022]
Panel C. With Exposed Sample Indicator × Post						
Eligibility	0.094	0.093	−0.012	−0.011	−0.005	−0.004
	(0.011)	(0.023)	(0.015)	(0.043)	(0.017)	(0.045)
	[0.065,0.119]	[0.002,0.129]	[-0.037,0.034]	[-0.070,0.167]	[-0.034,0.048]	[-0.070,0.189]
Exposed Sample	N	N	N	N	N	N
States	43	43	43	43	43	43
Individuals	2,397,313	2,397,313	2,397,313	2,397,313	2,397,313	2,397,313

Notes: This table reports coefficients from IV regressions of different measures of health insurance coverage on Medicaid eligibility, instrumented by different predictions of Medicaid eligibility. Regressions are estimated in the full sample of individuals in 2013–14. Columns 1, 3, and 5 use a recentered instrument while Columns 2, 4, and 6 do not recenter but control for expected Medicaid eligibility. All regressions control for state and year fixed effects and an indicator for Republican-governed states interacted with year. The regressions in Panel B additionally control for deciles of household income, interacted with indicators for parental and work status and year. The regressions in Panel C instead add controls for an individual having characteristics that make them exposed to the partial ACA Medicaid expansion in 2014. Conventional state-clustered standard errors are reported in parentheses; 95% confidence intervals, obtained by a wild score bootstrap, are reported in brackets.

Supplementary Material (Not For Publication) to "Non-Random Exposure to Exogenous Shocks: Theory and Applications"

Kirill Borusyak, UCL and CEPR
Peter Hull, UChicago and NBER

Contents

C Supplementary Theory

C.1 Potential Outcomes and Heterogeneous Treatment Effects

This appendix recasts our key assumptions in a general potential outcomes framework and extends classic results on IV identification in the presence of heterogeneous treatment effects (e.g., Imbens and Angrist 1994) to our setting. We first derive an appropriate "first-stage monotonicity" condition under which recentered IV regressions estimate a convex average of heterogeneous effects. We then show how more conventional weighted averages can be obtained in our framework.

We define potential treatments and outcomes given the instrument components g and w as $x_\ell = x_\ell(g, w, u)$ and $y_\ell = y_\ell(g, w, e, u)$, where u and e capture sources of unobserved first- and second-stage heterogeneity. We do not require that the functions $x_\ell(\cdot)$ and $y_\ell(\cdot)$ are known. An instructive example is given by a reduced-form model with linear treatment effect heterogeneity: i.e. $x_\ell = z_\ell$ and $y_\ell = \beta_\ell x_\ell + \varepsilon_\ell$ where $u = 0$ and e collects the $(\beta_\ell, \varepsilon_\ell)$.

We first use this general model to formalize exclusion and shock exogeneity restrictions:

Assumption S1. *(Exclusion): $y_\ell(g, w, e, u) = y_\ell(x_\ell(g, w, u), w, e)$ almost-surely.*

Assumption S2. *(Independence): $g \perp\!\!\!\perp e \mid w$.*

The exclusion restriction requires shocks to only affect the outcome through their relationship with the treatment. Under this condition, the independence assumption requires that shock assignment is conditionally as-good-as-random; this condition is equivalent to assuming $g \perp\!\!\!\perp y_\ell(\gamma, w, e)$ jointly across ℓ and γ. We do not assume $g \perp\!\!\!\perp u \mid w$, allowing the first-stage relationship to be non-causal as discussed more below.

We characterize the recentered IV estimand in terms of the general marginal effects as $\beta_\ell(x, w, e) = \frac{\partial}{\partial x} y_\ell(x, w, e)$. For notational simplicity we assumed that x_ℓ is continuous and that $y_\ell(x, w, e)$ is differentiable in x, though below and in Appendix E.1 we show that it is straightforward to allow for discrete treatments. We also assume bounded support of x_ℓ to avoid integrability issues, but the result can be generalized to unbounded supports under appropriate regularity conditions. We then have the following result:

Proposition S1. *Suppose Assumptions 1 and S1 hold, $Pr(x_\ell \geq x \mid z_\ell = z, e, w)$ is weakly increasing in z for each x almost-surely over (e, w), and the support of x_ℓ is bounded by some $\left[\underline{\chi}, \bar{\chi}\right]$. Then recentered IV identifies is*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{\underline{\chi}}^{\bar{\chi}} \beta_\ell(\chi, w, e)\omega_\ell(\chi, w, e) d\chi\right], \tag{S1}$$

where $\omega_\ell(\chi, w, e)$ are convex weights (i.e., $\omega_\ell(\chi, w, e) \geq 0$ almost-surely and $\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{\underline{\chi}}^{\bar{\chi}} \omega_\ell(\chi, w, e) d\chi\right] = 1$) that are proportional to the conditional-on-(w, e) covariance of z_ℓ and $\mathbf{1}[x_\ell \geq \chi]$.

Proof. See Appendix E.1. $\square$

Proposition S1 imposes a first-stage monotonicity condition: that x_ℓ is stochastically increasing in z_ℓ conditional on e and w. This condition is substantially more general than conventional ones (e.g., Angrist et al. 2000). Conventional monotonicity specifies a causal and monotone relationship between the treatment and the instrument: i.e., $x_\ell = x_\ell(z_\ell, \eta)$ with $z \perp\!\!\!\perp \eta$ and $\frac{\partial}{\partial z} x_\ell(z, \eta) \geq 0$ almost-surely. This is sufficient for our stochastic monotonicity (with η included in the list of unobservables e, which is without loss of generality). However, our condition also applies to settings where the shocks g affect many observations of z_ℓ and x_ℓ jointly and differentially, such that a causal first stage does not exist. For example, in the linear shift share case of $z_\ell = \sum_n w_{\ell n} g_n$, we may suppose that the shares underlying z_ℓ are partially misspecified, such that $x_\ell = \sum_n \pi_{\ell n} g_n + \eta_\ell$ for unobserved $(\pi, \eta) \perp\!\!\!\perp g \mid w$ (but exclusion still holds). Proposition S1 shows that the recentered IV regression remains causal in this case provided x_ℓ is stochastically increasing in z_ℓ conditional on e and w. This holds, for example, when the $w_{\ell n}$ and $\pi_{\ell n}$ are almost-surely non-negative and the g_n are mutually independent; Proposition S1 can thus be seen to generalize a monotonicity condition for shift-share IV established by Borusyak et al. (2020).

To gain intuition for the weights in Proposition S1, we consider two special cases. First, we consider the reduced-form model with linear treatment effect heterogeneity:

Corollary S1. *Suppose $x_\ell = z_\ell$, $y_\ell = \beta_\ell x_\ell + \varepsilon_\ell$, and Assumptions 1 and S1 hold with e collecting the $(\beta_\ell, \varepsilon_\ell)$. Then the estimand of the recentered regression is*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) z_\ell\right]} = \frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell \beta_\ell \sigma_\ell^2\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \sigma_\ell^2\right]}, \tag{S2}$$

where $\sigma_\ell^2 = \mathrm{Var}\left[\tilde{z}_\ell \mid w\right]$.

Proof. See Appendix E.1. $\square$

This result shows that the recentered IV identifies a variance-weighted average of treatment effects. We compute these weights for the empirical application in Section 4.1.

Second, we consider the IV model with a binary treatment and instrument. Here $y_\ell = y_\ell(0)(1-x_\ell)+y_\ell(1)x_\ell$ with the potential outcomes $y_\ell(0)$ and $y_\ell(1)$ defining the heterogeneous treatment effects as $\beta_\ell = y_\ell(1) - y_\ell(0)$. We further adopt a causal first stage relationship, writing $x_\ell = x_\ell(0)(1 - z_\ell) + x_\ell(1) z_\ell$. A version of Proposition S1 adapted to this setting shows how the recentered IV estimand differs from the average treatment effect (ATE), in the $x_\ell = z_\ell$ case, or the local average treatment effect (LATE):

Corollary S2. *Suppose Assumptions S1 and 1 holds, that x_ℓ and z_ℓ are binary, and that $x_\ell(1) \geq x_\ell(0)$ almost surely. Then the estimand of the recentered IV is.*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\beta_\ell \mid x_\ell(1) > x_\ell(0), w\right]\left(\frac{p_\ell \sigma_\ell^2}{\mathbb{E}\left[\frac{1}{L}\sum_\ell p_\ell \sigma_\ell^2\right]}\right)\right], \tag{S3}$$

where $p_\ell = Pr(x_\ell(1) > x_\ell(0) \mid w)$ and $\sigma_\ell^2 = \mathrm{Var}\left[\tilde{z}_\ell \mid w\right]$.

Proof. See Appendix E.1. □

This result shows that in this case the recentered IV identifies a weighted average of conditional treatment effects for "compliers" (defined by $x_\ell(1) > x_\ell(0)$), with weights again given by the conditional variance of the instrument.

It is immediate from both corollaries that a more conventional causal estimand, such as the ATE or LATE in the second case, is obtained by rescaling the instrument by σ_ℓ^2: i.e. instrumenting by $(z_\ell - \mu_\ell)/\sigma_\ell^2$. This approach requires $\sigma_\ell^2 > 0$ almost surely, which with binary instruments can be understood as an overlap condition–that $Pr(z_\ell = 1 \mid w) \in (0, 1)$.[56] Rescaling is typically no more difficult than recentering, since σ_ℓ^2 is also given by the shock assignment process (i.e., Assumption 3). It may, however, significantly increase the variance of the estimator.

In the case of reduced-form studies with continuous $x_\ell = z_\ell$ and arbitrary heterogeneity of treatment effects, one can further adopt the two-step procedure of Hirano and Imbens (2004) to identify the general dose-response function $\rho(\zeta) = \frac{1}{L}\sum_\ell \mathbb{E}\left[y_\ell\left(\zeta, w, e\right)\right]$. Identification can be established under Assumptions 1 and S1 as $\rho(\zeta) = \frac{1}{L}\sum_\ell \mathbb{E}\left[\beta_\ell(\zeta, r_\ell(\zeta; w))\right]$ where $\beta_\ell(\zeta, r) = \mathbb{E}\left[y_\ell\left(\zeta, w, e\right) \mid r_\ell(\zeta; w) = r\right]$ for the generalized propensity score $r_\ell(\cdot; w)$, defined as the conditional density of z_ℓ given w. This density is also given by the shock assignment process (i.e., Assumption 3). However, applying this identification argument can be challenging, as it requires estimating a series of conditional expectations $\beta_\ell(z, r)$ across potentially non-*iid* $\ell = 1, \ldots, L$. This also requires an overlap condition (that z_ℓ is non-degenerate given w) and as with simpler propensity-score-based methods the variance of such estimators may be high relative to a simpler recentered regression approach.

We finally note that IV inference may be challenging when treatment effects vary. For testing the so-called "sharp null" of $\beta_\ell(x, w, e) = 0$, almost surely, the randomization-based

[56]Note that in the conventional reduced-form treatment effects setting, where $x_\ell = z_\ell$, the recentered and reweighted instrument coincides with a conventional inverse-propensity score weight (Horvitz and Thompson (1952); Hirano et al. (2003)): $\frac{z_\ell - \mu_\ell}{\sigma_{z,\ell}^2} = \frac{x_\ell - Pr(x_\ell = 1 \mid w)}{Pr(x_\ell = 1 \mid w)(1 - Pr(x_\ell = 1 \mid w))}$, since here $\mu_\ell = Pr(x_\ell = 1 \mid w)$ has the interpretation of a treatment propensity score (Rosenbaum and Rubin (1983)).

tests in Section 3.4 still apply but may reject under the "weak null" of no average effect (i.e. that the estimand in Proposition S1 is zero). Inverting RI tests to form confidence intervals for β is also no longer sensible with heterogeneous effects. This issue is not specific to RI, as asymptotic inference may also be challenging in this case. For example in the linear shift-share setting, Adão et al. (2019) derive conservative asymptotic variance estimators only for a reduced-form estimator, under strong conditions. Aronow and Samii (2017) similarly construct conservative asymptotic variance estimators in the network interference setting. We view generalizing these approaches as a potentially fruitful area for future research.

C.2 Consistency of Recentered IVs

This appendix establishes conditions under which the recentered IV estimator and associated RI tests are consistent. We give a high-level condition regarding the cross-sectional variation in the instrument conditional on w, then provide lower-level sufficient conditions, and finally consider the case when w includes the permutation class of shocks $\Pi(g)$.

We study consistency of a recentered IV estimator,

$$\tilde{\beta} = \beta + \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell}, \tag{S4}$$

by considering a sequence of data-generating processes implicitly indexed by L. As usual, $\tilde{\beta} \xrightarrow{p} \beta$ as $L \to \infty$ provided $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell$ and $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell$ weakly converge to zero and a non-zero constant, respectively. We focus here on the former exclusion restriction, maintaining a general condition of instrument relevance:

Assumption S3. *(Relevance):* $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \xrightarrow{p} M \neq 0$.

In practice, the relevance of a given recentered instrument may be tested by extending the RI procedures in the previous section. That is, to test that z_ℓ has no first-stage effect on x_ℓ (for any ℓ) one may leverage knowledge of the shock assignment process to construct randomization-based rejection regions for statistics involving z_ℓ and x_ℓ.

The potentially complex correlation structure across observations of $\tilde{z}_\ell \varepsilon_\ell$ precludes the use of traditional weak laws of large numbers or standard extensions to show that $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell \xrightarrow{p} 0$. To restrict those correlations, assumptions can be imposed on either the $\tilde{z}_\ell$, the ε_ℓ, or both. In the recentered IV approach, which draws on substantial knowledge of the shock process (e.g. Assumption 3), it is natural to make further assumptions on the observed $\tilde{z}_\ell$. In doing so, we impose only a weak regularity condition on the unobserved ε_ℓ:

Assumption S4. *(Regularity):* $\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right] \leq B$ *for finite* B.

We start by establishing recentered IV consistency under a high-level condition that limits mutual dependence of $\tilde{z}_\ell$; we then establish lower-level sufficient conditions that are easier to verify in specific designs. The high-level condition intuitively states that observations are well-differentiated, in terms of their exposure to the shocks g through the recentered instrument:

Assumption S5. *(Weak IV dependence):* $\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0$.

Given this assumption, we may show the consistency of both the recentered IV estimator and its associated RI test:

Proposition S2.
(i) *Suppose Assumptions 2, 3, and S3-S5 hold. Then* $\tilde{\beta} \xrightarrow{p} \beta$.
(ii) *Suppose Assumptions 1, 3, and S3-S5 hold with* $\mathbb{E}\left[x_\ell^2 \mid w\right]$ *and* $\mathbb{E}\left[x_\ell \varepsilon_\ell \mid w\right]$ *uniformly bounded. Then the randomization test of Proposition S3 with* $T = \frac{1}{L}\sum_\ell f_\ell(g, w)(y_\ell - b x_\ell)$ *is consistent, i.e. for any* $b \neq \beta$ *we have* $Pr\left(T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \to 1$.

Proof. See Appendix E.2. □

The key condition of weak IV dependence states that the average absolute value of mutual covariances of the recentered instrument $\tilde{z}_\ell$ converges to zero as L grows. 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 in g. When this condition holds, Proposition S2 shows that $\tilde{\beta}$ is consistent even when unobserved shocks affect observations jointly (through ε_ℓ), in an unspecified manner.[57] Proposition S2 applies to recentered IV; Appendix C.6 extends it to μ_ℓ-controlled regressions (see Proposition S6(v)).

Our two sufficient conditions for Assumption S5 are non-nested:

Lemma S1.
(i) *Suppose* $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost-surely for all* (ℓ, m) *and Assumption 1 holds. Then Assumption S5 holds if* $\text{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\right] \to 0$*. Moreover, if* $f_\ell(g; w)$ *is weakly monotone in* g *for all* ℓ *and components of* g *are jointly independent conditionally on* w*, then* $\text{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost-surely.*
(ii) *Suppose* $G_\ell \subseteq \{1, \ldots, N\}$ *is such that* $f_\ell(\cdot; w)$ *does not depend on* g_n *for any* $n \notin G_\ell$ *almost-surely. Then Assumption S5 holds if* $\frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \varnothing\right] \to 0$*, the components of* g *are jointly independent conditionally on* w*, and* $\mathbb{E}\left[\tilde{z}_\ell^2 \mid w\right]$ *is uniformly bounded.*

Proof. See Appendix E.2. □

[57] We note that the recentering of z_ℓ is key for this result: the non-recentered IV estimator may not converge to β even when z_ℓ is valid in the sense of $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0$. For instance, suppose observations with systematically high z_ℓ (i.e., high μ_ℓ) are similarly exposed to an unobserved aggregate shock in ε_ℓ: the variance in $\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell$ due to this shock may not vanish, even in large samples when Assumption S5 holds. This problem does not arise for recentered IV which does not systematically vary across observations. See Lee and Ogburn (2019) for a related discussion.

The first condition 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 exposure weights. More generally nonlinear $f_\ell(\cdot)$ may also be monotone in the shock vector; for example each transportation infrastructure upgrade may weakly improve market access everywhere. In these cases, the recentered IV estimator is consistent when the first-stage covariance converges to a non-zero constant M and the average instrument $\frac{1}{L}\sum_\ell \tilde{z}_\ell$ converges to its expectation of zero in the l_2 norm. For linear 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. (2020) and Adão et al. (2019). The assumption of independent shocks can be weakened, for instance to allow for shocks that are independent across many clusters. The second condition in Lemma S1 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 random shock, and $f_\ell(\cdot)$ only depends on ℓ's shock and those of its neighbors up to a fixed network distance.

We note that Assumption S5 and Lemma S1 may be difficult to apply when w includes a permutation class $\Pi(g)$, as in our market access application. Even if the shocks are *iid* conditionally on other components of w, they can be dependent conditionally on $\Pi(g)$ (negatively correlated, in the scalar g_n case). An extension of Proposition S2 that applies in this case is available on request.

C.3 Randomization Inference

We begin the discussion of RI by considering a test of some null hypothesis $\beta = b$. With $b = 0$, for example, we test that outcomes y_ℓ are unaffected by treatment x_ℓ. We consider a scalar test statistic $T = \mathcal{T}(g, y - bx, w)$, where y and x are $L \times 1$ vectors collecting the outcome and treatment observations. When $b = \beta$, $T = \mathcal{T}(g, \varepsilon, w)$, and under Assumption 1 the distribution of this T conditional on ε and w is given by the shock assignment process $G(g \mid w)$. We may simulate this distribution under Assumption 3, by redrawing (e.g., permuting) the shocks in g and recomputing T (sometimes this distribution can be known analytically). If the original value of T is far in the tails of the simulated distribution, we then have grounds to reject the null that $\beta = b$.

Formally, we have the following result on hypothesis testing:

Proposition S3. *Suppose Assumptions 1 and 3 hold, let $\alpha \in (0,1)$, and for some $b \in \mathbb{R}$ and scalar-valued $\mathcal{T}(\cdot)$ let $T = \mathcal{T}(g, y - bx, w)$ and $T^* = \mathcal{T}(g^*, y - bx, w)$, where g^* is distributed according to $G(\cdot \mid w)$, independently of (g, x, y) conditionally on w. Under the* *null of $\beta = b$,*

$$Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \geq 1 - \alpha, \tag{S5}$$

where the acceptance region is constructed for a given b as

$$T_{\alpha/2} = \max\left\{t \in \mathbb{R}\colon \; Pr(T^* < t \mid y, x, w) \leq \frac{\alpha}{2}\right\} \tag{S6}$$

$$T_{1-\alpha/2} = \min\left\{t \in \mathbb{R}\colon \; Pr(T^* > t \mid y, x, w) \leq \frac{\alpha}{2}\right\}. \tag{S7}$$

Equation (S5) further holds with equality when $T^* \mid (y, x, w)$ is continuously distributed under the null.

Proof. See Appendix E.3. □

This result shows that when shocks are as-good-as-randomly assigned, a test of $\beta = b$ which rejects when $T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]$ has size of exactly α in finite samples provided the test statistic is conditionally continuously distributed under the null. When this distribution is not continuous, the test is still guaranteed to be conservative with a rejection rate of no greater than α.[58] The lower- and upper-bounds of the test region, $T_{\alpha/2}$ and $T_{1-\alpha/2}$, are given by the shock assignment process (Assumption 3) and represent the lower- and upper $\frac{\alpha}{2}$th percentile tails of the known conditional distribution of T^*. With exchangeable shocks, for example, $T_{\alpha/2}$ and $T_{1-\alpha/2}$ are given by the tails of the permutation distribution of $\mathcal{T}\left(g^*, y - bx, w\right)$ where $g^* = \pi(g)$ for random permutations $\pi(\cdot) \in \Pi$, holding (y, x, w) fixed. These tails can be computed from all permutations or from a random sample (Lehmann and Romano 2006, p. 636).[59] We note that while the previous intuition for such a testing procedure conditioned on ε and w, Proposition S3 establishes correct unconditional coverage of the test. This follows by the law of iterated expectations: the unconditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right)$ is the expectation, across realizations of ε and w, of the controlled conditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right] \mid \varepsilon, w\right)$.[60]

[58] $T \mid w$ will be discretely distributed when $g \mid w$ is discrete, such as when the support of $g \mid w$ represents some set of permutations of g. It is straightforward to show that in such cases one can construct a test of exact size by introducing randomness in $\mathcal{T}(\cdot)$; see, e.g., Lehmann (1986, p. 233).

[59] When a random sample of permutations is used, the realized g (i.e. identity permutation) should be added to this sample. The test then remains exact, or slightly conservative because of discreteness (Lehmann and Romano 2006, p. 636; Hemerik and Goeman 2018). In contrast to identification (see footnote 19) randomness of the chosen permutations is important here: non-random permutation sets do not generally guarantee valid inference (e.g. Southworth et al. 2009).

[60] It is instructive to highlight how exactly the knowledge of the shock assignment process matters in Proposition S3. Suppose that g is incorrectly assumed to be exchangeable, i.e. a uniform distribution is imposed over the $N!$ elements of g's permutation class. By construction, the test is guaranteed to reject the true β in some set of at most $\alpha \cdot N!$ permutations regardless of the true assignment process. However, unless

It follows from Proposition S3 that one can construct confidence intervals for β with correct coverage in finite samples under Assumptions 1 and 3. Formally, we have the following result:

Corollary S3. *Suppose Assumptions 1 and 3 hold and let CI denote the set of $b \in \mathbb{R}$ that are not rejected by the test in Proposition S3. Then $Pr(\beta \in CI) \geq 1 - \alpha$, with equality if $T^* \mid (y, x, w)$ is continuously distributed under the null.*

Proof. Follows from Proposition S3 by the standard logic of test inversion. □

In some settings, the confidence interval (or confidence set) CI obtained from inverting randomization tests may be infinite on one or both sides or empty, with the last possibility providing evidence against correct specification (Imbens and Rosenbaum 2005).

C.4 RI Efficiency

This appendix establishes conditions under which the optimal recentered instrument $\tilde{z}^*$ maximizes the asymptotic power of our preferred RI test by minimizing the asymptotic variance of the associated IV estimator. Consider the class of regular recentered IV estimators $\tilde{\beta} = \frac{1}{L}\tilde{z}'y / \frac{1}{L}\tilde{z}'x$, for $\tilde{z} = f(g)$ satisfying $\mathbb{E}[\tilde{z}] = 0$ (suppressing dependence on w throughout), that converges at rate r_L and has asymptotic first-stage M and asymptotic variance V. The asymptotic variance of $\frac{1}{L}\tilde{z}'\varepsilon$ is thus $\tilde{V} = M^2 V$. Consider the following:

Assumption S6. *Let $T(g^*, \varepsilon) = r_L \frac{1}{L} f(g^*)' \varepsilon$. For g_1^* and g_2^* distributed according to $G(\cdot)$, with g_1^*, g_2^*, and ε mutually independent, $(T(g_1^*, \varepsilon), T(g_2^*, \varepsilon)) \xrightarrow{d} \left(\sqrt{\tilde{V}} Z_1, \sqrt{\tilde{V}} Z_2\right)$, where Z_1 and Z_2 are independent standard normal variables.*

This assumption requires that $T = r_L \frac{1}{L}\tilde{z}'\varepsilon$ is (i) asymptotically normal and (ii) asymptotically independent of $T(g^*, \varepsilon)$ when g and g^* are independent. The latter part rules out cases where mutual correlation in the residuals is so strong that the randomization distribution of T depends on a particular realization of ε. From these conditions we have the following proposition:

Proposition S4. *Suppose the assumptions of Proposition 2 hold, along with Assumption S6. Fix $\alpha \in (0, 1)$ and $\delta \neq 0$. Then the limiting power of an RI test of size α based on $T(g, y - b_L x) = r_L \frac{1}{L} f(g^*)' (y - b_L x)$, against a sequence of local alternatives $b_L = \beta - \delta / r_L$, is a decreasing function of only the recentered IV estimator's asymptotic variance, V.*

Proof. See Appendix E.4. □

the true conditional distribution of g is uniform, the probability of the realized shocks g being in the true rejection set need not be α, leading to size distortions.

C.5 Assignment Processes with Unknown Parameters

This appendix considers the case where the shock assignment process is known up to a finite-dimensional vector of parameters θ. For example, instead of assuming that each railroad line in a transportation plan has an equal chance of being opened by a given date, a researcher may model the probability of line completion as a logistic function of the line length with an unknown coefficient θ. Similarly, instead of assuming that some industry shocks (e.g., to productivity) are fully exchangeable one may allow for parameterized heteroskedasticity: larger industries, for example, may have less dispersed shocks than small industries. We propose a plug-in estimator for the structural parameter β in which θ is estimated and used for recentering. We then adapt the Berger and Boos (1994) approach to inference with nuisance parameters to build conservative finite-sample confidence intervals.

We consider extensions of Assumption 3 where the distribution of $g \mid w$ is given by by a known function $G(g; w, \theta)$ of unknown θ. For example, one may assume conditionally independent binary shocks g_n with $Pr(g_n = 1 \mid w, \theta) = \Lambda(r_n'\theta)$ for a $K \times 1$ vector of shock-level observables r_n (including a constant) included in w, where $\Lambda(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$ is the logistic function. In this class of models, θ can be estimated from (g, w) by maximum likelihood (MLE), which is consistent under standard conditions, although other estimators may also be available. Given an estimate $\hat{\theta}$ a recentered IV instrument $\hat{z}_\ell = z_\ell - \mu_\ell\left(\hat{\theta}, w\right)$ can be measured, for $\mu_\ell(\theta_0, w) = \mathbb{E}_{\theta_0}[z_\ell \mid w] \equiv \int f_\ell(g, w) dG(g; w, \theta_0)$. We establish the conditions for large-sample consistency for this plug-in estimator for β below.

Valid, but likely quite conservative confidence intervals for β in such cases can be obtained by a simple extension of the previous randomization inference procedure. Given a value of θ, the randomization test for $\beta = b$ of Proposition S3 applies. Thus using the maximum p-value of this test across all possible values of θ yields a conservative test for β (with a corresponding confidence interval).[61] However, these confidence intervals are likely to be quite wide: even if the observed g is very informative about the precise value of θ, this test still searches through values very far from $\hat{\theta}$.

We propose an alternative two-step approach following Berger and Boos (1994) that is likely to be much less conservative but still valid (see Ding et al. (2016) for another application of this idea to RI). In the first step, a confidence interval CI_θ for θ with coverage $1 - \gamma$ is constructed for some $\gamma \in (0, \alpha)$; Berger and Boos (1994) recommend $\gamma = 0.001$. Such tests are easy to build since the distribution of g is fully specified given θ; thus an exact RI-based confidence interval for θ can be constructed from any statistic $S = \mathcal{S}(g; w, \theta_0)$ by rerandomizing g according to $G(\cdot; w, \theta_0)$. As usual, the choice of S determines the power of the test and the length of the confidence interval. We propose a statistic that corresponds

[61] An equivalent view on this procedure is to test joint hypotheses $\beta = b$ and $\theta = \theta_0$ using the test of Proposition S3 and then project the resulting confidence interval on the space of β.

to the score test, $S=\frac{\partial}{\partial\theta}\log G(g;w,\theta_0)$, since the Hodges-Lehmann estimator induced by it is the MLE.[62] For vector-valued θ, S can be converted to a scalar Lagrange Multiplier (LM) statistic $S'\mathbb{E}_{\theta_0}\left[SS'\mid w\right]^{-1}S$; a value θ_0 is rejected if the LM statistic is in the right tail of its distribution. In the second step, the maximum p-value of the Proposition S3 test is taken across $\theta_0\in CI_\theta$ only—a much smaller set in large samples than the entire parameter set used in the more conservative procedure. The p-value of the Berger and Boos (1994) test is the obtained maximum plus γ. A value of β is therefore rejected at significance level α if it is rejected under all $\theta_0\in CI_\theta$ with significance $\alpha-\gamma$.

The following proposition establishes the conditions for the plug-in estimator consistency and derives an exact confidence interval for θ using the Berger and Boos (1994) approach.

Proposition S5.
(i) *Suppose Assumptions 1 holds, $\hat\theta$ is consistent for θ, and $\mu_\ell(\theta_0,w)$ is almost-surely differentiable with respect to θ_0 in a convex parameter space Θ and with a bounded gradient $\frac{\partial\mu_\ell}{\partial\theta}$. Then when Assumptions S3-S5 hold at the true value of θ, and the sequences $\frac{1}{L}\sum_\ell|x_\ell|$ and $\frac{1}{L}\sum_\ell|\varepsilon_\ell|$ are bounded in probability, the plug-in recentered IV estimator with instrument $\hat z_\ell$ is consistent.*
(ii) *Suppose Assumption 1 holds. Let $p_\beta(\beta;\theta_0)$ be the p-value of the randomization test of Proposition S3 for a given value of θ and let CI_θ denote a confidence interval for θ such that $Pr\left(\theta\in CI_\theta\right)\geq 1-\gamma$ for $\gamma<\alpha$. Construct $CI_\beta=\{b\in\mathbb{R}:\ \max_{\theta_0\in CI_\theta}p_\beta\left(\beta,\theta_0\right)+\gamma>\alpha\}$. Then CI_β is conservative for β, i.e. $Pr\left(\beta\in CI_\beta\right)\geq 1-\alpha$.*

Proof. See Appendix E.5 for part (i). Part (ii) follows from Berger and Boos (1994). □

Five remarks are due. First, while the Berger and Boos (1994) test is conservative in finite samples only when CI_θ is, using an asymptotic confidence interval for θ will generally yield an asymptotically conservative interval for β. This simplifies computation: constructing the conventional Wald confidence interval for the MLE estimator of θ is much easier than inverting the score-based randomization test. Second, in some cases even simpler RI confidence intervals for β which plug in the estimate of $\hat\theta$ as if it was known are asymptotically correct (Shaikh and Toulis 2019), although general conditions for this are unknown. Third, as discussed in Berger and Boos (1994), in some cases the nuisance parameter θ can be eliminated by using sufficient statistics which also yields a simpler exact confidence interval. In the above binary shocks example, if r_n captures a saturated set of dummy variables then elements of g are exchangeable within the clusters corresponding to them and it is not necessary to know or estimate θ. [63] Fourth, for a consistent $\hat\theta$, including

[62] This follows because $\frac{\partial}{\partial\theta}\log G(g;w,\hat\theta_{\mathrm{MLE}})=0=\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log G\left(g^*;w,\theta\right)\right]$ for the MLE estimator $\hat\theta_{\mathrm{MLE}}$ and g^* randomly drawn from G.

[63] Rosenbaum (1984) shows how this idea can be extended in the logit model with arbitrary discrete

$\mu_\ell\left(\hat{\theta}, w\right)$ as a linear control (with an additional coefficient in front of it) may produce a consistent estimator of β, as long as the slope of the auxiliary regression of z_ℓ on $\mu_\ell\left(\hat{\theta}, w\right)$ converges. This is because $\text{Cov}\left[z_\ell, \mu_\ell\left(\theta, w\right)\right] = \text{Var}\left[\mu_\ell\left(\theta, w\right)\right]$ by definition of $\mu_\ell\left(\theta, w\right)$, such that the slope coefficient will converge to one and the regression will asymptotically use the recentered $\tilde{z}_\ell$ as an instrument (by the Frisch-Waugh-Lovell theorem).[64]

Finally, a closely related way to incorporate θ is by assuming that some one-to-one transformation of shocks $\tilde{g} = h\left(g; w, \theta\right)$ has a known nuisance parameter-free distribution conditionally on w (with w that may itself depend on θ, such as when it includes permutation classes of $\tilde{g}$). An intuitive case is when $\tilde{g}_n = (g_n - \rho_n(\theta, w))/\sigma_n(\theta, w)$ is exchangeable, after recentering and rescaling shocks according to a parametric model; here the conditional distribution of $\tilde{g}$ over its permutation class is uniform. Again, RI yields exact permutation-based confidence intervals for θ as well as corresponding Hodges-Lehmann estimators $\hat{\theta}$, and the Berger and Boos (1994) approach yields a conservative confidence interval for β. We discuss the choice of powerful randomization statistics next.

Suppose first that $\tilde{g}_n = g_n - \rho_n\left(\theta, w\right)$ is exchangeable across n. Here the expression for the mean $\rho_n\left(\theta, w\right)$ does *not* include an unknown constant because a constant is redundant: $\tilde{g}_n$ is exchangeable if and only if $\tilde{g}_n - \zeta$ is exchangeable for constant ζ. To estimate θ, one may consider the nonlinear least squares estimator of θ from a model $g_n = \zeta + \rho_n\left(\theta, w\right) + u_n$, which is consistent as N grows under standard assumptions given conditionally mutually independent u_n. It is then straightforward to verify that this is the Hodges-Lehmann estimator corresponding to the RI statistic $T_\theta = \frac{1}{N}\sum_n \tilde{g}_n \frac{\partial \rho_n}{\partial \theta}$. Therefore, one may use this statistic to construct an exact confidence interval for θ. In the second step, the expected instrument given θ is constructed by the following simulation: $\tilde{g}_n$ are randomly permuted to get $\tilde{g}_n^*$ and $g_n^* = \rho_n\left(\theta, w\right) + \tilde{g}_n^*$ is then used in constructing $z_\ell^* = f_\ell\left(\left(g_n^*\right)_{n=1}^N, w\right)$.

The second case is heteroskedasticity, and for simplicity we assume that shocks are known to have a constant mean. One may therefore be willing to assume that $\tilde{g}_n = g_n/\sigma_n\left(\theta, w\right)$ is exchangeable; in this case a multiplicative constant is redundant in the formulation of the shock conditional variance, $\zeta\sigma_n^2(\theta, w)$. As usual, a variety of RI statistics can be used, and one reasonable choice is $T_\theta = \frac{1}{N}\sum_n \tilde{g}_n^2 \sigma_n^2 \frac{\partial \sigma_n^2}{\partial \theta}$ as it induces the Hodges-Lehmann estimator that corresponds to the moment of nonlinear least squares estimation for the model $g_n^2 = \zeta^2\sigma_n^2\left(\theta, w\right) + u_n$.[65] With an estimate of θ, recentering is performed by permuting $\tilde{g}_n^*$ and simulating $g_n = \tilde{g}_n^*\sigma_n\left(\hat{\theta}, w\right)$, and the Berger and Boos (1994) confidence

observables r_n. He exploits the property of logit that, regardless of θ, $G\left(g \mid w\right)$ is the same for any binary vector g that yields the same vector $\sum_n g_n r_n$.

[64] At the same time, including $\mu_\ell\left(\theta, w\right)$ as a nonlinear control and jointly estimating (β, θ) will not generally work because there is no appropriate Frisch-Waugh-Lovell theorem for nonlinear IV.

[65] To be precise, the Hodges-Lehmann estimator solves $\sum_n \left(g_n^2 - \zeta^2\sigma_n^2\right)\frac{\partial \sigma_n^2}{\partial \theta} = 0$ for $\zeta^2 = \frac{1}{N}\sum_n g_n^2/\sigma_n^2$. This estimator is consistent for θ when u_n are conditionally mutually independent and under standard regularity conditions.

interval for β is obtained similarly.

C.6 Efficiency Controls

This appendix considers the case where a researcher wishes to include an $R \times 1$ vector of predetermined controls a_ℓ (which includes a constant) that absorb some of residual variation in y_ℓ to increase the efficiency of estimating β. Here we show, following Rosenbaum (2002), that our recentered IV estimation and RI results generalize directly to this case. This section also justifies the approach proposed in Section 3.2 of controlling for μ_ℓ instead of recentering the instrument by it. We abstract away from the assignment process parameters θ for clarity but those can be straightforwardly incorporated.

The following result extends Propositions 1, S3, 2, and S2(i):

Proposition S6. *Suppose $g \perp\!\!\!\perp (a, \varepsilon) \mid w$ where a collects the $a_\ell = (a_{\ell 1}, \ldots, a_{\ell r})$. Let $v_\ell^\perp$ denote the sample projection of a variable v_ℓ on a_ℓ: i.e., $v_\ell^\perp = v_\ell - \hat{\alpha}_v' a_\ell$ for $\hat{\alpha}_v = \left(\frac{1}{L}\sum_\ell a_\ell a_\ell'\right)^{-} \frac{1}{L}\sum_\ell a_\ell v_\ell$ and $(\cdot)^-$ denoting a generalized inverse of a matrix. Then:*
(i) *β is identified by $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell^\perp\right] / \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp\right]$, assuming $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp\right] \neq 0$;*
(ii) *The randomization test based on the statistic $T = \frac{1}{L}\sum_\ell z_\ell \left(y_\ell^\perp - b x_\ell^\perp\right)$ is valid;*
(iii) *The Hodges-Lehmann estimator induced by this RI statistic is the recentered IV estimator of y_ℓ on x_ℓ instrumented by $\tilde{z}_\ell$ and with the a_ℓ controls, $\tilde{\beta}_\perp = \frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell^\perp / \frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp$;*
(iv) *Recentering the instrument does not affect the estimator when μ_ℓ is included in a_ℓ.*
(v) *$\tilde{\beta}_\perp \xrightarrow{p} \beta$ if Assumptions 3 and S3–S5 hold, $\mathbb{E}\left[a_{\ell r}^2 \mid w\right] \leq B_a$ almost surely for all ℓ and $r = 1, \ldots, R$, $\frac{1}{L}\sum_\ell a_\ell a_\ell'$ is almost surely invertible (such that $\hat{\alpha}_v$ is unique), $\hat{\alpha}_x = O_p(1)$, and $\hat{\alpha}_\varepsilon = O_p(1)$.*

Proof: See Appendix E.6.

The independence condition of the proposition is automatically satisfied when a is non-random conditionally on w. The first two results of the proposition exploit the fact that $\varepsilon^\perp$ is constructed from ε and a, both conditionally independent of g. The third result directly follows Rosenbaum's (2002) result on covariate adjustment in randomization inference. It is a consequence of the Frisch-Waugh-Lovell theorem: an IV estimator with controls can be represented as the bivariate IV estimator for y_ℓ and x_ℓ residualized on the controls but with the original instrument $\tilde{z}_\ell$. The fourth result restates the fact that recentering by μ_ℓ is not necessary when y_ℓ and x_ℓ have been residualized on it. The final result provides regularity conditions for estimator consistency.

C.7 Multiple Treatments and Instruments

This appendix considers the case when the outcome equation includes several endogenous variables. For example, in network spillover regressions of Section D.3 the researcher may specify both a direct effect of the shock to the unit and the spillover effect from other units. We show that the main results of the paper apply in that case: instrument recentering restores instrument validity, and randomization inference yields a joint confidence interval for the coefficient vector. We then discuss several special cases where separate confidence intervals may be obtained for individual coefficients, more efficiently than by projecting the joint interval. Notably, one set of shocks arising from the natural experiment is generally sufficient to identify multiple causal effects, as long as endogenous variables differ in their exposure to the same shocks. Finally, we discuss how most of the results in our framework generalize to the overidentified case, with multiple instruments.

Consider a just-identified IV estimator of a constant-effect regression

$$y_\ell = \boldsymbol{\beta}'\boldsymbol{x}_\ell + \varepsilon_\ell, \tag{S8}$$

where $\boldsymbol{x}_\ell = (x_{1\ell}, \ldots, x_{M\ell})'$ is an $M \times 1$ vector of endogenous variables ("treatments") instrumented by a vector of instruments $\boldsymbol{z}_\ell = (z_{1\ell}, \ldots, z_{M\ell})'$ for $z_{m\ell} = f_{m\ell}(g, w)$, $m = 1, \ldots, M$. A constant term and other "efficiency" controls are allowed as in Appendix C.6, and are assumed to have been partialled out. For each m we define the expected instrument $\mu_{m\ell} = \mathbb{E}\left[z_{m\ell} \mid w\right]$ and the recentered instrument $\tilde{z}_{m\ell} = z_{m\ell} - \mu_{m\ell}$ collected into vectors $\boldsymbol{\mu}_\ell$ and $\tilde{\boldsymbol{z}}_\ell$, respectively.[66]

Lemma 1 and Proposition 1 extend trivially to this setup, establishing identification of β provided the first-stage matrix $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{\boldsymbol{z}}_\ell \boldsymbol{x}_\ell'\right]$ is of full rank. Interestingly, only one set of exogenous shocks g is generally sufficient to satisfy the rank condition and identify multiple coefficients when different treatments have different exposure to the same shocks. For example, when $z_{1\ell} = g_\ell$ is the random treatment status of network node ℓ, $z_{2\ell}$ is the average treatment of ℓ's neighbors, and a reduced-form regression is considered ($\boldsymbol{x}_\ell = \boldsymbol{z}_\ell$), $\tilde{z}_{1\ell}$ and $\tilde{z}_{2\ell}$ have independent variation identifying both effects.

Now consider randomization inference. As before, the distribution of any scalar or vector-valued statistic $\mathcal{T}\left(g, y - \boldsymbol{x}\boldsymbol{b}, w\right)$, where $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L)'$, is known conditional on w and ε under the null of $\boldsymbol{\beta} = \boldsymbol{b}$, which can be used to construct valid tests and confidence intervals for $\boldsymbol{\beta}$ as in Proposition S3. The only complication here is that the natural choice of test statistic that extends Proposition 2, $T = \frac{1}{L}\sum_\ell \tilde{\boldsymbol{z}}_\ell'(y_\ell - \boldsymbol{b}'\boldsymbol{x}_\ell)$, is vector-valued and

[66] Unlike the single endogenous variable case, constant effects are important for identification here. For instance, in the reduced-form model of $y_\ell = \beta_{1\ell}g_\ell + \beta_{2\ell}w_\ell g_\ell + \varepsilon_\ell$ with arbitrarily heterogeneous effects $\beta_{1\ell}$, the parameter β_2 is not identified as it is impossible to isolate the effect of $w_\ell g_\ell$ from the effect of g_ℓ that varies in a way correlated with w_ℓ.

requires to pick a rejection region in $\mathbb{R}^M$. A natural approach is to map T into the scalar LM statistic $T_{LM} = TV(\boldsymbol{b})^{-1}T$, where $V(\boldsymbol{b})$ is the randomization variance matrix of T that imposes the null $\boldsymbol{\beta} = \boldsymbol{b}$. That is, V is computed by re-randomizing g according to $G(g \mid w)$ while holding $\varepsilon_\ell = y_\ell - \boldsymbol{b}'\boldsymbol{x}_\ell$ fixed. The null $\boldsymbol{\beta} = \boldsymbol{b}$ is then rejected when T_{LM} exceeds its $1-\alpha$ randomization quantile, which is equivalent to T being outside a particular ellipsoid in $\mathbb{R}^M$ centered at zero. This test follows the Hodges-Lehmann approach: the RI p-value is maximized at the recentered IV estimator at which $T = 0$ and thus $T_{LM} = 0$.[67] The exact joint confidence interval for $\boldsymbol{\beta}$ is constructed by inverting this test, as usual.

One problem with applying classical randomization inference in this extension is that it yields joint confidence intervals for the multiple coefficients in $\boldsymbol{\beta}$, rather than separate confidence intervals for each β_m. This is because only sharp nulls of $\beta = b$ can be tested, and not partial nulls of $\beta_m = b_m$. One may of course take a projection of the joint confidence interval on each component: i.e. reject $\beta_m = b_m$ when it is rejected for all values of b_{-m}. However, the implied intervals for individual coefficients can be very conservative or even infinite. The problem is particularly important when $M > 2$ and thus the joint interval cannot be easily visualized. We therefore describe several approaches how more powerful confidence intervals for individual coefficients can be constructed in special cases. For notational simplicity, we suppose $M = 2$ and that we are interested in inference on β_1.

A first approach to marginal confidence interval construction applies when one of the endogenous variables can be isolated by an appropriate randomization test. For a simple example, suppose an interacted outcome equation is specified, with $x_{2i} = x_{1i}r_i$ for some some predetermined variable r_i satisfying $Pr\left(r_i = 0\right) > 0$ (e.g. a dummy variable). In the subsample with $r_i = 0$ the second term vanishes, and a confidence interval for β_1 is obtained by standard RI. Following Aronow (2012), one can also consider a more elaborate situation in which the reduced-form spillover effect β_1 of some exogenous shock is of interest. Here $x_{1\ell}$ may be, for instance, the average treatment of ℓ's neighbors on a network g, but a direct effect β_2 of ℓ's own randomly assigned shock, $x_{2\ell} = g_\ell$, is also allowed for. Then the following procedure can be used: fix some subset of units $\bar{L} \subset \{1, \dots, L\}$ and condition the distribution of g on the observed shocks to the units in $\bar{L}$, $\bar{g} = (g_\ell)_{\ell \in \bar{L}}$. Then $y_\ell - \beta_1 x_{1\ell}$ is independent of g conditionally on $\bar{g}$ because the direct effects are the same across these realizations of g. Yet, there is conditional variation in $x_{1\ell}$ arising from shocks to other units $\ell \notin \bar{L}$ which allows for identification of β_1 and randomization tests on this coefficient.

Second, if a confidence interval for one coefficient is obtained (e.g. in one of the situations discussed above), the approach of Berger and Boos (1994) and Ding et al. (2016) can be used to get conservative marginal confidence intervals for the remaining coefficient. Specifically,

[67] One may notice that for $M = 1$ this test slightly differs from the one in Section 3.4 as it is based on T^2 rather than T. The two tests coincide when the randomization distribution of T is symmetric around its mean of zero.

let CI_2 be an exact interval for β_2 with coverage $1-\gamma$ for some $\gamma \in (0, \alpha)$, e.g. $\gamma = 0.001$. Then $\beta_1 = b_1$ is rejected if $\boldsymbol{\beta} = (b_1, b_2)$ is rejected by the RI-based LM test for every $b_2 \in CI_2$ at significance level $\alpha - \gamma$. In other words, the p-value of the test for β_1 is the maximum p-value of the joint test across $b_2 \in CI_2$, plus γ. When $\gamma \to 0$, CI_2 becomes uninformative, and this procedure converges to the projection of the joint confidence interval. However, for a given γ CI_2 may be narrow in large samples, and taking the maximum across $b_2 \in CI_2$ rather than the entire real line may result in a much more powerful test for β_1.

We also conjecture that the following asymptotic approach may apply in many applications, though we leave a formal analysis of this approach to future research. One may expect under certain regularity conditions that some central limit theorem applies to $\frac{1}{\sqrt{L}} \sum_\ell \tilde{\boldsymbol{z}}_\ell \left(y_\ell - \boldsymbol{\beta}' \boldsymbol{x}_\ell\right)$, such that it converges to a jointly normal distribution. Moreover, under some conditions, this unconditional distribution may be well approximated by the randomization distribution across g only (see Lehmann (1986), Theorem 15.2.3). Furthermore, estimating this distribution at a consistent estimate $\hat{\beta}$ instead of β may be asymptotically innocuous (see Shaikh and Toulis (2019)). In such cases, $\hat{\boldsymbol{\beta}}$ is asymptotically normal and an asymptotically valid confidence interval for each coefficient separately can easily be obtained by delta method, using the randomization variance matrix of $\frac{1}{\sqrt{L}} \sum_\ell \tilde{\boldsymbol{z}}_\ell \left(y_\ell - \hat{\boldsymbol{\beta}}' \boldsymbol{x}_\ell\right)$ as an estimate of the asymptotic variance of $\frac{1}{\sqrt{L}} \sum_\ell \tilde{\boldsymbol{z}}_\ell \varepsilon_\ell$.

We finally note that while this discussion has focused on the just-identified case, some aspects easily generalize to the case where M instruments are used for $J < M$ endogenous treatments (including where $J = 1$). The identification results (Lemma 1 and Proposition 1) for the recentered IV and RI-based LM tests extend to that case without modification. One difference is in the Hodges-Lehmann estimator corresponding to this LM test, i.e. the value of $\boldsymbol{b}$ which minimizes $TV(\boldsymbol{b})^{-1}T$. While in the just-identified case this is the recentered IV estimator, with overidentification it is more similar to a continuously updating general method of moments estimator, since the variance matrix is also a function of $\boldsymbol{b}$.

C.8 Nonlinear Outcome Models

This appendix considers settings where the parameter of interest is specified in terms of a nonlinear model, $y_\ell = m_\ell(x; \beta) + \varepsilon_\ell$,where $\{m_\ell(\cdot)\}_{\ell=1}^L$ is a set of known functions and x includes an unrestricted set of observables. We show that our results on identification, inference, and asymptotic efficiency generalize naturally to this setup.

For ease of exposition we assume the parameter β is one-dimensional, as in the main text; extensions to the multidimensional case are given by integrating the insights in Appendix C.7. We continue to assume an instrument of $z_\ell = f_\ell(g; w)$. IV identification of β typically requires instrument recentering, as in the linear case. When Assumption 2(i) holds, it is immediate that

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\left(y_{\ell}-m_{\ell}(x;\beta)\right)\right]=0, \tag{S9}$$

and identification follows when β uniquely solves this moment condition. In particular, local identification (uniqueness in a neighborhood of β) follows when $\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\frac{\partial}{\partial\beta}m_{\ell}(x;\beta)\right]$ is non-zero. As in Lemma 1, identification fails absent instrument recentering, unless the expected instrument $\mu_{\ell}=\mathbb{E}\left[f_{\ell}(g;w)\mid w\right]$ is orthogonal to the structural residual ε_{ℓ} in the sense of $\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\mu_{\ell}\varepsilon_{\ell}\right]=0$.

Valid finite-sample inference on β is similarly obtained as in the linear case. The test statistic which induces as a Hodges-Lehmann estimator the solution to the sample analog of (S9) is

$$T=\frac{1}{L}\sum_{\ell}\left(f_{\ell}(g,w)-\mu_{\ell}\right)\left(y_{\ell}-m_{\ell}(x;b)\right), \tag{S10}$$

which can be used to form randomization tests and confidence intervals from specified counterfactual shocks.

Derivation of the efficient instrument also follows similarly. As in Proposition 2, we consider the class of recentered instruments yielding IV estimators that converge at some rate. Given analogous regularity conditions, it is straightforward to verify that the asymptotically variance-minimizing instrument in this class is

$$z^{*}=\mathbb{E}\left[\varepsilon\varepsilon'\mid w\right]^{-1}\left(\mathbb{E}\left[\frac{\partial}{\partial\beta}m(x;\beta)\mid g,w\right]-\mathbb{E}\left[\frac{\partial}{\partial\beta}m(x;\beta)\mid w\right]\right), \tag{S11}$$

where we write $m(x;b)$ as the collection of $m_{\ell}(x;b)$. This nests equation (9) in the linear case, where $\frac{\partial}{\partial\beta}m(x;\beta)=x$. Outside of this case, the optimal instrument generally depends on β. A two-step optimal instrument could then be obtained by applying a first-step estimate of β to this formula, given its consistency and additional regularity conditions.

We next show that under additional conditions the variance-minimizing instrument z^{*} also maximizes the local power of the associated RI test. The theoretical argument closely follows Lehmann and Romano (2006, Section 5.2.2), except without assuming that the distribution of shocks is the permutation distribution or that the data are *iid*. The proposition requires asymptotic normality of the estimator, for which we do not have low-level conditions because of the generality of our framework but which should hold in typical applications. For simplicity we here treat w as non-stochastic, but the argument generalizes since the characterization of efficient recentered IV in Proposition 2 applies conditionally on w.

D Further Practical Implications

Here we discuss the practical implications of our Section 3 framework for several applied literatures, relating our approach to identification and inference to the previously employed strategies. Specifically, Appendices D.1 and D.2 consider estimating the effects of transportation upgrades and policy eligibility, extending the discussions of Sections 4.1 and 4.2, respectively. The following subsections elaborate on the summaries in Section 3.6 on network spillover effects (Appendix D.3), linear and nonlinear shift-share instruments (Appendix D.4), instruments implied by spatial equilibrium models (Appendix D.5), instruments derived from partial randomness in centralized assignment mechanisms (Appendix D.6), "free-space" instruments for mass media access (Appendix D.7), and weather-based instruments (Appendix D.8).

D.1 Effects of Transportation Upgrades

Our theoretical results provide a new general approach for estimating the effects of transportation infrastructure upgrades. Traditionally, these studies specified as treatment an indicator that region ℓ is connected to the network (e.g. Chandra and Thompson 2000; Michaels 2008) or a measure of local connection intensity (e.g. Baum-Snow 2007; Duranton and Turner 2012). Modern approaches often use more elaborate model-based market access measures (Donaldson and Hornbeck 2016), in recognition of the fact that infrastructure upgrades can impact regions not directly connected to the network. While many analyses of transportation shocks study local outcomes of individual regions, some estimate the effects on bilateral outcomes, such as trade or migration between pairs of regions or firms (Allen et al. 2019; Volpe Martincus and Blyde 2013).

Our framework formalizes three distinct challenges with identifying such effects. The first is strategic placement of infrastructure upgrades in anticipation of regional productivity or amenity growth, a concern that is well-recognized in the literature (Redding and Turner 2015). When viewing upgrades as our shocks g, strategic placement can be formalized by a dependence of g on ε which violates shock exogeneity (i.e. Assumption 1). The transportation effects literature has devised several remedies to this challenge, in particular by excluding major cities or other regions which directly affect the placement of infrastructure (the "inconsequential place approach") or by using planned or historical routes to instrument for the constructed railways. Our framework can accommodate these solutions by either limiting attention to inconsequential places (which changes the sample, and therefore ε) or by viewing planned or historical routes as the exogenous shocks (which changes g).[68]

[68]Other weakened versions of Assumption 1 are possible too, e.g. that ε_ℓ is conditionally independent of g_n for all potential lines n that do not cross region ℓ; the market access instrument should then be

When no assumption of upgrade exogeneity seems *ex ante* plausible, one can study market access effects by leveraging the exogeneity of its other determinants. For example, Bartelme (2018) estimates market access effects by leveraging exogenous shifters to market size and not transportation upgrades. Our framework also applies to this strategy, with market size shifters collected in g. It is generally difficult to obtain causal estimates without any exogeneity assumption placed on any shock to market access. We therefore suppose that Assumption 1 holds for some g and ε, with the other determinants of market access collected in w. For concreteness we suppose g captures some features of transportation upgrades, as the alternative assumption of Bartelme (2018) is less standard.

A second challenge that is less discussed in the transportation upgrade literature is that a cross-sectional correlation between regional connectivity and unobservables is likely to arise from their common dependence on regional geography. Even when upgrades are exogenous in the sense of Assumption 1, they are unlikely to be uniformly assigned across regions. For example, if upgrades are concentrated in more economically developed areas (as in our HSR application above), which differ in their unobservables, OVB may arise. Formally, $\mathbb{E}[\varepsilon_\ell \mid w]$ and expected network connectivity may covary across regions ℓ.

A further third challenge arises with measures like market access, which respond to non-local upgrade shocks and are defined by a non-linear formula. As discussed in Section 2, this can lead to complex variation in expected market access growth even when exogenous upgrades are uniformly distributed in space. This challenge highlights a novel tradeoff between traditional connectivity regressions and the modern approach of Donaldson and Hornbeck (2016): although market access regressions may better capture the non-local effects of transportation upgrades (and, as such, satisfy the exclusion restriction in Assumption 1), they may also lead to more intricate OVB challenges.

These two new challenges are not specific to connectivity and market access regressions. For example Allen et al. (2019) study how migration between locations in the U.S. and Mexico depends on a measure of difficulty of traveling between them, as affected by the wall constructed in some parts of the border between the two countries. Even when new wall placement is not strategic, it is unlikely to be uniformly distributed along a border (the second challenge). The effects of wall upgrades are furthermore non-local, complicating OVB (the third challenge). One may expect, for instance, that wall-induced changes in travel difficulty are correlated with the distance between locations (and thus potentially with the error term): regardless of which sections of the wall are built, places far away from the border will be affected less. This is because, when traveling between them, it is easier to substitute away from newly blocked routes.

constructed excluding those lines (Lin 2017). We also note that these approaches may not suffice to remove all statistical dependence between g and ε, for instance if productivity growth is correlated between major and "inconsequential" regions and historical routes affect current productivity in unobserved ways.

Both of the new challenges raised by our framework are solved by specifying and simulating counterfactual transportation network upgrades, which can be achieved in several ways. A first approach, which we illustrated above, is to use predetermined upgrade plans and appeal to the randomness of which subset of the plan materializes as of some date. This approach contrasts with the typical use of upgrade plans in the literature, as instruments themselves (Redding and Turner 2015), which can relax Assumption 1 but does not help with specifying shock counterfactuals.[69] Our strategy for specifying counterfactuals is instead similar to the use of accepted and rejected plans for new plant construction in Greenstone et al. (2010), but in the transportation setting. 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 a robustness check.

A second approach to specifying counterfactuals is to model upgrades in terms of their engineering, economic or political requirements and find alternative upgrades that satisfy these criteria. One of the placebo analyses by Ahlfeldt and Feddersen (2018) follows this logic: they note that the new railway line connected two major cities in Germany at distance around 160km and had three intermediate stops. They then construct 1,000 random placebo lines that satisfy the same description. One could also obtain engineering estimates of viable alternative routes for lines connecting major cities, augmenting the inconsequential places approach with valid counterfactuals.[70]

A third approach leverages known discontinuities in the policies determining network links, as in a simpler regression discontinuity analysis. Campante and Yanagizawa-Drott (2018), for example, note that cities just under 6,000 miles apart are distinctly more likely to have direct air links, relative to cities just above that threshold. A local randomization view of such discontinuities (e.g., Cattaneo et al. (2015)) might motivate counterfactual networks that perturb links around the threshold.

Finally, the assignment process for the shocks may sometimes be given by external (e.g. institutional or scientific) knowledge. For example, Volpe Martincus and Blyde (2013) leverage an earthquake that blocked a large number of roads in Chile to estimate the effect of infrastructure downgrades on trade patterns. Geological knowledge could be used in this case to specify the disruption locations of counterfactual earthquakes and construct the

[69] Historical routes can be used in place of the plan if the researcher is willing to assume that some of them decayed for random reasons while others continued to be in use. This again contrasts with the typical use of historical routes (e.g. Duranton et al. (2013)). We note that if historical routes or plans are instead used to satisfy Assumption 1 then counterfactual routes and plans need to be simulated to apply our approach.

[70] These approaches can be viewed as special cases of specifying a stochastic network formation model and simulating it to construct counterfactuals. Other models of network formation (e.g., Chandrasekhar and Jackson 2014; Acemoglu and Azar 2020) could similarly be used to study causal effects of network centrality in appropriate economic contexts (see footnote 72). This would seem a novel use of such models.

expected instrument.

Without specifying counterfactuals (or even assuming upgrade exogeneity) one may remove some sources of OVB by including control variables, such as province fixed effects, geographic coordinates, and pre-period characteristics of the local economy. The unobservables ε_ℓ which are not captured by those controls should then be assumed orthogonal to the treatment variable. The case for such orthogonality can, however, be challenging to make for market access (or similar treatments) constructed from many sources of variation. It may be untenable to assume that *all* of the treatment determinants are orthogonal to the remaining variation in ε_ℓ, while the weaker assumption that such orthogonality holds despite the endogeneity of some determinants of treatment appears hard to justify *a priori*.

Relative to a conventional controlling approach, our framework for estimating transportation upgrade effects offers two important advantages. First, by clarifying an experimental ideal it makes the argument for market access exogeneity explicit and transparent and allows for a substantive debate on whether the institutional features of the setting make such argument *ex ante* plausible. Second, it yields additional tests: while pre-trend and balance tests are useful with any identification approach, our framework offers new specification tests (as illustrated in the HSR setting above), with a new mode of inference. Our approach also yields new robustness checks for a conventional controlling strategy. As footnote 21 explains, if market access effects are identified without recentering because the included controls perfectly capture either the expected instrument or the endogenous features of shock exposure, then the estimates should be robust to further controlling for expected market access growth constructed with some reasonable counterfactuals.

D.2 Effects of Policy Eligibility

We next discuss extensions to our baseline approach to policy eligibility instruments, presented in Section 4.2. We further discuss advantages of our recentered IV relative to a controlling strategy used in the literature to estimate effects of unemployment insurance.

We first note that the recentered IV approach may generate power gains over conventional simulated instruments when not all determinants of eligibility are observed and included in v_ℓ. Cohodes et al. (2016), for example, use a simulated instrument to study the long-term effects of Medicaid eligibility on children without observing a key eligibility determinant (parental income). Their instrument assigns to each individual ℓ the average eligibility of a nationally representative sample of individuals with the same observed demographics (age, race, and birth year), if they were subject to the policy in ℓ's state of residence (see Currie and Gruber (2001) for a similar approach). This instrument is a function of the state policy and observed demographics only and overlooks useful variation in the state of residence which is likely correlated with the error term but predictive of unobserved parental income. Our IV framework therefore suggests one might use the average eligibility of individuals with similar demographics who are residing in the same state as the instrument, while adjusting for its average value over permutations of state policies.

This approach is also useful when all eligibility determinants are observed, but a researcher does not wish to include them in v_ℓ. This would be the case when, for example, parental income responds endogenously to the state policy, violating Assumption 1. Indeed, Currie and Gruber (1996) discuss this as one of the motivations for their original simulated instrument construction. In such cases predictors of such determinants that cannot respond to the natural experiment, such as parental income from before a state policy change, may be instead used to boost asymptotic power. East and Kuka (2015) use a similar approach to augment simulated instrument construction in evaluating the effects of unemployment insurance eligibility.

Our framework also yields insights to an alternative approach in the related literature on the eligibility effects of unemployment insurance (e.g., Cullen and Gruber 2000; East and Kuka 2015). This approach regresses outcomes on true or predicted eligibility while flexibly controlling for individual characteristics v_ℓ. When policies are exchangeable across states, this approach is also justified within the Section 3 framework since the expected instrument is a function of v_ℓ. Flexible controls for individual characteristics have an additional benefit of potentially predicting variation in the error term, thus improving asymptotic efficiency. However, this approach is vulnerable to a curse of dimensionality; indeed, Gruber (2003) finds this strategy difficult to implement for Medicaid, where individual characteristics can have complex nonlinear effects on eligibility. Our approach reveals the single expected instrument control needed for valid causal inference under the same exogeneity assumption.

D.3 Network Spillovers

This appendix discusses the implications of our framework for identification and inference in spillover regressions. In such settings the units ℓ represent nodes in a network (of people, firms, regions, etc.), and g captures shocks that are as-good-as-randomly assigned to them.[71] The target parameter β denotes the causal effect of a node-specific treatment x_ℓ which captures the spillovers from g at node ℓ. For example, one may be interested in the effects of an inventor's death on the future productivity of her co-inventors (Jaravel et al. 2018), having a direct supplier or supplier's supplier hit by a natural disaster on a firm's growth (Carvalho et al. 2020), or having more "dewormed" students at neighboring schools on

[71] It is unimportant that the shocks are assigned to and the outcomes are observed for *all* nodes. Our framework applies directly when, for instance, the shocks and the observations correspond to non-overlapping subsets of nodes (e.g. Doudchenko et al. (2020)).

a student's test scores (Miguel and Kremer 2004).[72] In bringing our framework to such settings we maintain the assumption that the spillover treatment has been well-specified, in that it captures all relevant channels by which the shocks affect a given node.[73] Spillover regressions are typically reduced-form, such that x_ℓ is a function of the predetermined network structure in w and the shocks g and $z_\ell = x_\ell$, but we also allow for IV regressions where treatment is affected by both g and other shocks.

Such spillover regressions may suffer from OVB when different nodes face systematically different exposure to the exogenous shocks because of their network position. This exposure, summarized by the expected instrument, depends on how the shocks are assigned and how z_ℓ is constructed (e.g., whether it is nonlinear in the shocks). We first illustrate how these factors determine whether the OVB can arise and whether it can be solved by conventional regression controls.

For concreteness we center this discussion of OVB around a stylized version of Miguel and Kremer (2004, henceforth MK), where $z_\ell = x_\ell$ counts the number of dewormed students at schools within a fixed radius from student ℓ's school (excluding itself) and y_ℓ is some health outcome. The deworming shocks are generated by an RCT which as-good-as-randomly selected half of all schools for deworming. Thus z_ℓ can be written as a linear function of which students are experimentally dewormed (collected in a binary vector g) with coefficients determined by the pre-existing network of students and schools (summarized in w). To simplify the discussion we abstract from other features of the actual MK setting, including the fact that they estimate direct effects of deworming together with the spillover effect; our framework is extended to such specifications with multiple treatments in Appendix C.7.

The OVB issue is easy to see in this setting: students who live in denser areas will have systematically more dewormed neighbors, and dense areas may also have different unobserved determinants of health. MK address this threat by controlling for the total number of eligible students in neighboring schools, n_ℓ. This is indeed what our approach recommends, provided all students have an equal chance of being dewormed (a probability we denote by q). The expected number of dewormed neighbors $\mu_\ell = \mathbb{E}\left[z_\ell \mid w\right] = qn_\ell$ is then proportional to n_ℓ, so including this control purges OVB.[74]

We now consider four deviations from the benchmark setup, in which conventional regression controls are no longer sufficient to span the expected spillover treatment and

[72] One can also consider settings in which x_ℓ measures node ℓ's network centrality itself, after a shock to the network links. The market access measure we discuss in Sections 2 and 4.1 can be viewed as an example, capturing a region's centrality in a transportation network.

[73] This exclusion restriction may follow from a particular model for peer effects, as in Manski (1993) and Manski (2013), or a general equilibrium gravity model in the market access case. See Angrist (2014) for a discussion of potential biases from misspecifying social interaction models in simpler settings.

[74] The n_ℓ control would not be necessary if spillovers were instead specified as driven by the fraction of dewormed neighbors. Under simple randomization all students would then have the same expected spillover treatment, $\mu_\ell = q$.

purge OVB. First, there may be a more complex shock assignment process. For example, stratified random assignment which makes deworming more likely for some students or schools (depending on, for example, student gender or school size) would make μ_ℓ no longer proportional to n_ℓ.

Second, spillover effects may involve network weights. Suppose MK had instead specified the spillover treatment as $z_\ell = \sum_{n \in \mathcal{N}_\ell} w_{\ell n} g_n$, where $\mathcal{N}_\ell$ is the set of ℓ's neighbors and $w_{\ell n}$ measures the strength of the spillover from neighbor n's deworming status g_n, for instance determined by how frequently ℓ and n interact.[75] Then even in simple experimental designs, the expected instrument equals $\mu_\ell = \sum_{n \in \mathcal{N}_\ell} w_{\ell n}$ which need not be collinear with n_ℓ: a person who interacts with neighbors more will tend to be more affected by the deworming experiment even conditionally on the number of neighbors. An example of this issue can be found in Acemoglu et al. (2015) where ℓ and n are municipalities in Colombia and $w_{\ell n}$ are inverse distance weights; here μ_ℓ reflects the geographic centrality of the region.

Third, spillover treatments may be nonlinear in the network shocks. If, for example, MK had studied the effects of having at least one treated neighbor, the appropriate μ_ℓ control would have been the student-specific probability of this event. Under simple randomization of deworming, this μ_ℓ is a nonlinear function of n_ℓ and OVB may be purged by flexibly controlling for n_ℓ. However this is no longer the case with school-level randomization or more complex experimental protocols.

Finally, spillovers may arise from shocks across the entire network, and not just from immediate neighbors. Suppose MK had instead studied the effect of deworming spillovers given by the geographic distance from the nearest dewormed school (which may be large for some schools). This non-local specification of network spillovers makes μ_ℓ inherently more complex: while in expectation the distance is smaller in dense areas, there is no simple measure that fully captures the expected distance to the nearest dewormed school. This example is inspired by Carvalho et al. (2020) who study the effects of network distance between firm ℓ and the nearest firm located in the geographic area of a natural disaster (specifically, the 2011 Tōhoku earthquake) in the firm-to-firm supplier network.[76]

In RCT settings like MK's the more intricate OVB challenges raised in each of these scenarios are easily solved. A researcher can simply compute the expected spillover treatment by redrawing shocks according to the randomization protocol, and appropriate adjust

[75] One can view such z_ℓ as a special case of a shift-share instrument further discussed in Section D.4 below. The benchmark MK case is obtained with $w_{\ell n} \equiv \mathbf{1}[n \in \mathcal{N}_\ell]$.

[76] Among firms with a given number of direct suppliers (a covariate Carvalho et al. (2020) control for), the probability of having at least one supplier's supplier in the randomly assigned earthquake zone (i.e. at distance of two) still depends on how connected its suppliers are. Nonlinearity of the network distance as a function of shocks makes this probability also depend on the correlation structure of the shocks. Consider a firm for which the set of second-degree suppliers is not very large but is geographically dispersed. This firm is much more likely to be affected by a shock hitting random geographic areas (like an earthquake) than a shock hitting firms regardless of their geography.

for it. With natural experiments like that of Carvalho et al. (2020) institutional or scientific knowledge can help specify counterfactual shocks, for instance by drawing on appropriate geological models of earthquake probabilities across regions. With observational data counterfactual shocks can be specified by the partial exchangeability of shocks, perhaps conditional on node-level observables. The approach of Jaravel et al. (2018), for example, in which deceased and non-deceased co-inventors are matched based on age and other characteristics, can be viewed as leveraging such an exchangeability assumption.

Besides providing a solution to the OVB problem, our framework helps address well-known challenges of inference in network regressions. A conventional approach in the literature is to assume that $z_\ell \varepsilon_\ell$ is uncorrelated beyond a small geographical or network distance (Conley (1999); see Acemoglu et al. (2015) for a network example). This may work well when z_ℓ captures local spillovers. However, when both observed and unobserved shocks propagate further through the network conventional standard errors may be distorted. As usual, randomization inference is valid regardless of the correlation structure of unobserved shocks, relying only on the correlation structure of the recentered instrument that is implied by knowledge of the shock assignment process.

An even more challenging problem that is solved by RI is when few shocks are observed in the data, such as the single earthquake in Carvalho et al. (2020). A small number of shocks makes the asymptotic approach inapplicable: even absent spillovers, it is generally impossible to consistently estimate the effect of earthquakes if only one region is treated. The lack of consistency, however, does not preclude informative inference provided shock counterfactuals are specified. For example, if the true effect is zero, it is unlikely that unobserved shocks hit exactly the same region where the earthquake randomly happened. RI-based confidence intervals capture this idea formally.

D.4 Linear and Nonlinear Shift-Share Instruments

We next consider instruments of the form $z_\ell = f(g, w_\ell)$, where w_ℓ is a vector capturing the exposure of observation ℓ to the set of shocks g (usually of the same dimensionality N). Conventional shift-share instrument variables (SSIVs) set $f(\cdot)$ to take an exposure-weighted average of the shocks: typically a regional instrument z_ℓ is constructed from a set of industry shocks g_n as $z_\ell = \sum_n w_{\ell n} g_n$, where $w_{\ell n}$ measures the industry's share (of, say, employment) in the region (e.g., Bartik (1991) and Autor et al. (2013)). These instruments are often employed when the treatment variable can similarly be represented as a share-weighted average $x_\ell = \sum_n w_{\ell n} \tilde{x}_{\ell n}$, where component $\tilde{x}_{\ell n} = g_n + u_{\ell n}$ includes a potentially endogenous $u_{\ell n}$. More recently, the SSIV approach has inspired a number of nonlinear instrument constructions for treatments that combine exogenous shocks with local shares in more complex ways (e.g. Boustan et al. 2013; Berman et al. 2015; Chodorow-Reich and

Wieland 2020).

We first connect the Section 3 framework to the earlier work of Borusyak et al. (2020) in showing that simple regression controls are typically enough to purge OVB when using linear shift-share instruments, without specifying counterfactual shocks. We then present two sets of novel insights. First, we formalize the class of nonlinear SSIVs and propose an approach to identification with them. Explicit specification of counterfactual shocks is typically necessary for nonlinear SSIVs to be valid, but linear approximations to such instruments may obviate this need at an efficiency cost. Second, asymptotic inference is generally challenging in the nonlinear SSIV setting but valid randomization inference based on shock counterfactuals is straightforward. Moreover, we find in Monte Carlo simulations that in the linear SSIV case RI may serve as a useful complement to the existing asymptotic approach of Adão et al. (2019). Throughout this analysis we follow Borusyak et al. (2020) in assuming the shocks g_n are as-good-as-randomly assigned while the exposure shares $w_{\ell n}$ may be endogenous (for example, if unobserved industry shocks affect regions via the same shares).[77]

OVB with Linear SSIVs Borusyak et al. (2020) establish the validity of linear SSIVs under the assumption of quasi-random shock assignment, formalized as equal conditional expectations of g_n across n, and that the exposure shares sum to one across n for each observation. When the second assumption fails (what they label the "incomplete shares case"), they show how OVB is purged by controlling for the sum of shares. They further show how simple quasi-random shock assignment may be relaxed to allow the conditional expectation of g_n to depend on shock-level observables, by controlling for share-weighted averages of these potential confounders.

In our general framework, these insights can be seen to follow from the linearity of the expected shift-share instrument, since

$$\mu_\ell = \mathbb{E}\left[z_\ell \mid w\right] = \sum_n w_{\ell n} \mathbb{E}\left[g_n \mid w\right] \tag{S12}$$

is an exposure-weighted average of the expected shocks $\mathbb{E}\left[g_n \mid w\right]$. When the expected shock is constant ($\mathbb{E}\left[g_n \mid w\right] = \alpha$ for some α) and the exposure shares sum to one ($\sum_n w_{\ell n} = 1$ for all ℓ) the expected instrument is also constant ($\mu_\ell = \alpha$) and no correction to the linear SSIV is needed to avoid OVB. In the incomplete shares case where $W_\ell = \sum_n w_{\ell n}$ varies, it is enough to linearly control for W_ℓ since $\mu_\ell = \alpha W_\ell$. Under the weaker assumption of

[77] Our framework also nests the alternative linear SSIV framework of Goldsmith-Pinkham et al. (2020) in which the shares are exogenous but shocks need not be. With shocks considered non-stochastic and with *iid* data, as in Goldsmith-Pinkham et al. (2020), OVB does not arise as all observations are similarly exposed to the exogenous shares.

conditional random shock assignment, i.e. $\mathbb{E}[g_n \mid w] = q_n'\alpha$ for some vector of observables q_n, it is furthermore enough to control for the share-weighted sums of confounders, $Q_\ell = \sum_n s_{\ell n} q_n$, as they absorb $\mu_\ell = Q_\ell'\alpha$. The Section 3 framework also implies an alternative solution to this general case: instead of controlling for Q_ℓ one can recenter the *shocks* by their conditional expectation. This is because the recentered instrument is also a linear SSIV:

$$\tilde{z}_\ell = z_\ell - \mu_\ell = \sum_n w_{\ell n}\tilde{g}_n, \tag{S13}$$

where $\tilde{g}_n = g_n - \mathbb{E}[g_n \mid w]$. SSIV approaches which first residualize shocks on observables (e.g., Greenstone et al. (2020)) may be interpreted as following this recentering logic.

It is worth highlighting that the linearity of SSIV relaxes Assumption 3: the conditional expectation $\mathbb{E}[g_n \mid w]$ is the only moment of the shock assignment process that needs to be specified. Unmodeled higher moments, such as shock heteroskedasticity and clustering, do not pose problems for instrument recentering or controlling, as they might in the general case of nonlinear instrument constructions.

Nonlinear SSIVs The SSIV logic has recently been extended to cases where the treatment variable can be represented as a nonlinear function of some predetermined exposure shares w_ℓ and potentially endogenous "shifts," i.e. $x_\ell = f(\tilde{x}_\ell, w_\ell)$ for $\tilde{x}_\ell = (\tilde{x}_{\ell n})_{n=1}^N$. Chodorow-Reich and Wieland (2020), for example, study the effects of a regional labor reallocation index x_ℓ measuring the dispersion of local industry growth rates $\tilde{x}_{\ell n}$ with initial industry employment shares $w_{\ell n}$ as weights. If the researcher observes an exogenous shifter g_n for $\tilde{x}_{\ell n}$, that is not observation-specific (i.e., measured nationally or in other countries), an intuitive instrument can be constructed as $z_\ell = f(g, w_\ell)$, predicting x_ℓ via exogenous shocks and predetermined exposure. We believe this general formulation of such nonlinear SSIVs is novel, while nesting several applied examples, including the predicted change in a regional Gini coefficient in Boustan et al. (2013), the predicted share of migrants in Basso and Peri (2015), and the predicted foreign demand instrument of Berman et al. (2015).

Nonlinearity of $f(\cdot)$ generically leads to a challenging OVB problem. Even with fully exchangeable shocks and "complete shares," where linear SSIV is valid without correction, the expected instrument μ_ℓ is a complex function of w_ℓ. Moreover, unlike in the linear case, second and higher moments of the shock assignment process may be relevant to the expected instrument and must generally be specified for the appropriate recentering. Shock heteroskedasticity and clustering, in particular, are potential problems for nonlinear SSIV.

Our framework yields two solutions for OVB with nonlinear SSIVs, which give a likely tradeoff between efficiency and robustness. A researcher may recenter a nonlinear z_ℓ given a specification of shock counterfactuals. Alternatively, she may take a first-order approximation of z_ℓ around some fixed vector of shocks g^0 (e.g., $g^0 = 0$) to obtain a linear SSIV of $\check{z}_\ell = \check{w}_\ell' g$, for $\check{w}_\ell = \frac{\partial f}{\partial g}\left(g^0, w_\ell\right)$. As before, the linear instrument is valid when the shocks have a common mean and $\sum_n \check{w}_{\ell n}$ is controlled for (or, more generally, when the means of shocks depend linearly on some q_n and $\sum_n \check{w}_{\ell n}' q_n$ is controlled for). As an approximation to z_ℓ the linear SSIV is likely to predict x_ℓ less well and thus be less efficient. On the other hand, its validity depends on correct specification of fewer moments of the shock assignment process making it more robust. Such linear approximations have indeed motivated SSIV instruments in the context of some economic models (e.g., Kovak (2013), Adão et al. (2019), and Adão et al. (2020)) but the same logic applies generally, e.g. to predicted labor reallocation indices and Gini coefficients.

To make concrete the potential for OVB with nonlinear SSIVs, and our two solutions, we take a stylized example of a popular instrument which could be called a "SSIV in logs" (e.g., Berman et al. 2015; Berthou et al. 2019; Costa et al. 2019). Suppose $x_\ell = \log \frac{X_{\ell 1}}{X_{\ell 0}}$ denotes the growth rate of some regional variable which can be represented as a sum of industry components, $X_{\ell t} = \sum_n \tilde{X}_{\ell n t}$. For example, $X_{\ell t}$ may denote the total demand for a regional output that aggregates demand across industries n. Then x_ℓ can be rewritten as a nonlinear function of initial shares $w_{\ell n} = \frac{\tilde{X}_{\ell n 0}}{X_{\ell 0}}$ and regional growth rates $\tilde{x}_{\ell n} = \frac{\tilde{X}_{\ell n 1}}{\tilde{X}_{\ell n 0}}$, as $x_\ell = \log \sum_n w_{\ell n} \tilde{x}_{\ell n}$. Suppose that a researcher suspects endogeneity in regional growth rates but observes an industry characteristic G_{nt} with plausibly exogenous growth rates $g_n = \frac{G_{n1}}{G_{n0}}$ that predict the $\tilde{x}_{\ell n}$. This motivates a nonlinear SSIV:

$$z_\ell = \log \sum_n w_{\ell n} g_n. \tag{S14}$$

Although exogeneity of g_n makes $\sum_n w_{\ell n} g_n$ a valid instrument (after controlling for $\sum_n w_{\ell n}$), the log transformation in equation (S14) generally introduces OVB. This is because the log function is concave, so the expected instrument $\mu_\ell = \mathbb{E}\left[z_\ell \mid w\right]$ is systematically higher for regions where $\sum_n w_{\ell n} g_n$ has a lower variance. In particular, regions with more diversified economies (i.e., with $w_{\ell n}$ more dispersed across n) will tend to have systematically higher z_ℓ, while they may also have systematically different unobservables. For example, diversification may make the local economy more resilient to unobserved shocks, which would generate an upward bias in an IV estimator which takes some measure of regional economic growth as an outcome.[78]

Both of our solutions to OVB are quite intuitive for SSIVs in logs. One could recenter

[78] There is a further problem with the use of (S14) in practice. In panel specifications with unit and period fixed effects the instrument is commonly specified as $z_{\ell t} = \log \sum_n w_{\ell n} G_{nt}$ (e.g., Berman et al. 2015). In first differences this corresponds to $\Delta z_\ell = \log \sum_n \tilde{w}_{\ell n} g_n$ where shares are reweighted, $\tilde{w}_{\ell n} = \frac{w_{\ell n} G_{n0}}{\sum_{n'} w_{\ell n'} G_{n'0}}$. These shares do not align with the economic intuition behind the instrument construction and are likely to lead to a weaker first-stage. The correct construction is $z_{\ell t} = \log \sum_n w_{\ell n} \frac{G_{nt}}{G_{n0}}$.

equation (S14) by the appropriate measure of diversification μ_ℓ, which generally requires specifying and simulating counterfactual shocks. Alternatively one may take a log-linear approximation around $g_n = 1$, which here yields an intuitive linear SSIV: $\check{z}_\ell = \sum_n w_{\ell n} \log g_n$. Removing OVB from this instrument requires only specifying (and appropriately controlling for) the expected $\log g_n$.[79]

Inference and Monte Carlo Simulations We finally discuss how our framework brings new tools to SSIV inference. We are not aware of any general asymptotic theory for the nonlinear case (unless restrictive independence assumptions are placed on the residuals), making RI an attractive approach to inference. For the linear case we argue that RI may serve as a useful complement to the asymptotic theory of Adão et al. (2019).[80]

The choice between RI and asymptotic approaches involves tradeoffs. An advantage of RI is its validity even with relatively few or concentrated shocks, when the asymptotics approximation may not be accurate. At the same time, RI requires specification of the shock assignment process, rather than its first moment only, and thus assumptions of homoskedasticity (or a known parametric form of heteroskedasticity, as in Appendix C.5), distribution symmetry, or similar conditions which are not required for asymptotic exposure-robust inference. On other dimensions the two approaches are hard to compare in general: they both require constant treatment effects (and their sensitivity to treatment effect heterogeneity is not known), and their power may differ.[81]

We therefore examine the power and robustness properties of RI in a Monte Carlo simulation. Our simulation is based on the influential SSIV study by Autor et al. (2013), who estimate the effects of import competition with China on U.S. local labor markets. The simulation process follows Borusyak et al. (2020) in redrawing import competition shocks according to a wild bootstrap (to preserve shock heteroskedasticity), holding fixed the exposure shares and estimated structural residuals (see the data description at the end of this subsection for details). We consider two asymptotic approaches to SSIV inference:

[79] More robust linear SSIVs may more generally be obtained by taking the nonlinear transformation before averaging. Derenoncourt (2019), for example, addresses a skewed distribution of shocks by taking the sample percentiles of a shift-share instrument, yielding a nonlinear SSIV. Taking instead an exposure-weighted average of shock percentiles yields a linear SSIV that is valid without recentering, provided these percentiles are as-good-as-randomly assigned and the weights sum to one.

[80] Adão et al. (2019) consider an asymptotic sequence with many uncorrelated (or weakly correlated) shocks and sufficiently dispersed exposure. Linearity of the instrument is indispensable for their results. This is clear from the equivalence result in Borusyak et al. (2020), which shows that the Adão et al. (2019) standard errors can be obtained from conventional calculations after transforming the data to the level of shocks. For nonlinear SSIV no such equivalence generally exists, making it difficult to draw on conventional asymptotic theory.

[81] Adão et al. (2019) show that when treatment effects vary their standard errors are asymptotically conservative, but this result only applies to reduced-form SSIV regressions under certain conditions on the exposure shares. We know of no such guarantees in the IV case.

the conventional "exposure-robust" standard errors of Adão et al. (2019) (obtained via the equivalent shock-level regression in Borusyak et al. (2020)) and a version designed for better finite-sample coverage by imposing the null hypothesis. We contrast these approaches with an RI procedure based on randomly flipping the signs of the simulated shocks, which leverages a known symmetry of the shock distribution shape.

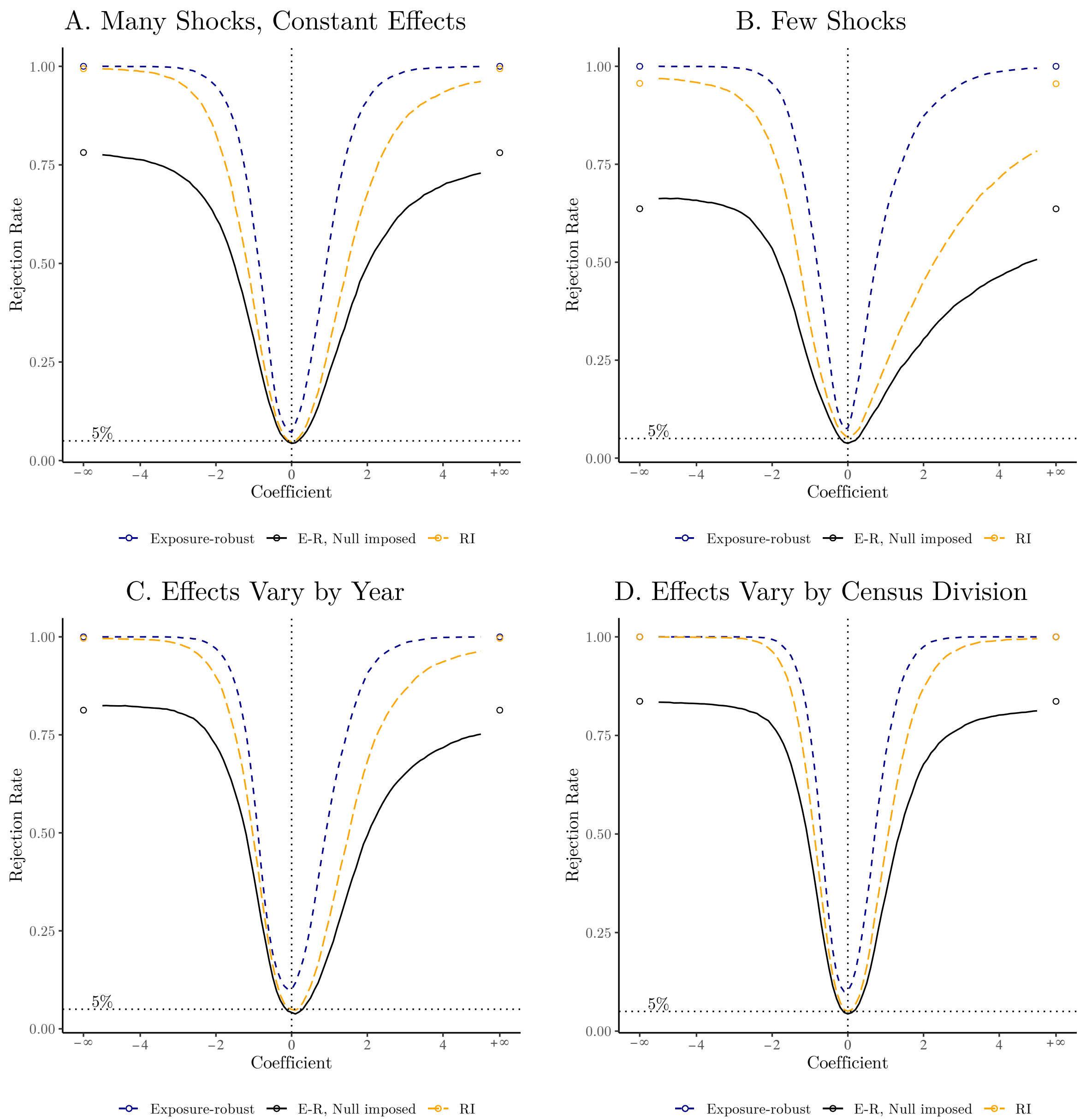
Panel A of Figure S1 presents a simulated power curve for the main Autor et al. (2013) data-generating process with 794 industry-by-period shocks (for 397 manufacturing SIC4 industries in two periods). We plot simulated rejection rates for each of the three modes of inference across both a true causal effect of $\beta = 0$ and a range of alternative hypotheses (normalized such that the SSIV estimate in the data has $\beta = 1$). Consistent with Borusyak et al. (2020) we find that conventional exposure-robust standard errors yield a mild over-rejection of the nominal 5% level test (see their Table C6, rows (b)), while the other two procedures exhibit no significant size distortions. Despite correct coverage, power of the null-imposed asymptotic inference procedure is low, failing to reject arbitrarily large values of β (i.e. yielding infinite 95% confidence intervals) over 20% of the time. In contrast the RI procedure has power similar to that of conventional exposure-robust asymptotics, if a bit smaller.

In Panel B of Figure S1 we consider a modified data-generating process with fewer shocks by aggregating manufacturing industries to their SIC2 groups (yielding 20 shocks in each period), again following Borusyak et al. (2020). The size of all three methods is similar to that in Panel A, with the overrejection of conventional exposure-robust standard errors increasing slightly. Here the power of RI is asymmetric: similar to that of conventional exposure-robust inference for negative values of β_0, but weaker for positive values. Still, it is again uniformly much stronger than that of the null-imposed procedure which (approximately) shares the RI property of correct size.

Finally, Panels C and D of Figure S1 explore robustness of the three methods to treatment effect heterogeneity, either across the two periods (Panel C) or nine geographic regions (Census divisions, Panel D), with the original 794 shocks. The data-generating process here is based on the heterogeneous effects estimated in the Autor et al. (2013) data and demeaned in such a way that sets the median SSIV estimate to zero to make coverage well-defined (see the data description below for details). Although there is no theoretical guarantee for either RI nor the asymptotic approximations in this case, the results are surprisingly similar to those in Panel A: RI coverage is correct and its power is close to that of exposure-robust inference, while substantially exceeding that of the null-imposed version.

Data for the Monte Carlo Simulation Our simulations above are based on a data-generating process that Borusyak et al. (2020) develop for the setting of Autor et al. (2013).

Figure S1: Simulated Size and Power of Alternative Shift-Share IV Inference Procedures



Notes: This figure plots simulated power curves for different shift-share IV inference procedures. The baseline data-generating process in Panel A comes from Autor et al. (2013), as described in Appendix D.4. In Panel B we reduce the number of industry shocks in each period from 397 to 20. In Panels C and D we specify a data-generating process with heterogeneous effects by period or Census division. Exposure-robust tests are obtained from the equivalent shock-level regressions of Borusyak et al. (2020). Hollow circles indicate the power at $\beta = \pm\infty$, approximated by $\beta = \pm 1,000$.

The baseline process, used in Panel A of Figure S1, is calibrated to the IV estimates in Column 3 of Table 4 in Borusyak et al. (2020) with a second and first stage of

$$y_{\ell t} = \beta x_{\ell t} + \gamma' r_{\ell t} + \varepsilon_{\ell t}, \tag{S15}$$

$$x_{\ell t} = \pi z_{\ell t} + \rho' r_{\ell t} + u_{\ell t}. \tag{S16}$$

The outcome $y_{\ell t}$ corresponds to the change in manufacturing employment as a fraction of the working-age population in U.S. commuting zone ℓ in decade t (either 1990-2000 or 2000-2007), the treatment $x_{\ell t}$ is a measure of regional import competition with China, and the shift-share instrument $z_{\ell t} = \sum_n s_{\ell nt} g_{nt}$ is constructed by combining the industry-level growth of China imports in eight developed economies g_{nt} with lagged regional employment weights of different industries $s_{\ell nt}$. The vector $r_{\ell t}$ includes the sum of lagged employment shares, interacted with period indicators, and other pre-treatment controls as described in Borusyak et al. (2020). The sum-of-share controls linearly span the expected instrument when the industry shocks have a common mean in each period, and without loss we demean g_{nt} by period. There are a total of 1,444 observations (722 commuting zones in two periods) and estimation is weighted by the start-of-period population of the commuting zone.

Each draw of the baseline simulation generates 1,444 new observations of $(y_{\ell t}, x_{\ell t}, z_{\ell t})$ by holding fixed the employment shares, pre-treatment controls, and estimated coefficients and residuals of equations (S15) and (S16) but redrawing the industry shocks g_{nt}. We generate new shocks from a wild bootstrap of $g^*_{nt} = g_{nt}\nu^*_{nt}$ by multiplying the original year-demeaned shocks by a standard normal ν^*_{nt}. This process preserves the heteroskedasticity of the shocks, and corresponds to the process in row (b) of Table C6 in Borusyak et al. (2020).

In Panel B of Figure S1 we modify the baseline process to reduce the number of shocks in each period, from $N = 397$ manufacturing SIC industries to 20 two-digit industries. This modification corresponds to the process in row (g) of Table C6 in Borusyak et al. (2020). We aggregate imports from China to the U.S. and either developed economies as well as the number of U.S. workers by manufacturing industry to construct the new g_{nt}, $z_{\ell t}$, and $x_{\ell t}$, as described in Appendix A.10 of Borusyak et al. (2020), holding fixed other variables. We then redraw shocks again by a wild bootstrap,

In Panels C and D of Figure S1 we modify the baseline process to add treatment effect heterogeneity, by period and Census division. We use the original number of shocks but instead estimate versions of equations (S15) and (S16) which interact both $x_{\ell t}$ and $z_{\ell t}$ with period or division fixed effects. In Panel C the estimated second-stage effects are -0.491 for the 1990s and -0.225 for the 2000s, replicating Table C3 of Borusyak et al. (2020). In Panel D the estimated effects vary between -0.609 for the East North Central Census division and -0.135 for the West North Central division. We then generate data as before, with a wild bootstrap for shocks.

In each panel we simulate power curves for three inference procedures: the "exposure-robust" asymptotic approach of Adão et al. (2019), this approach with the null hypothesis imposed, and randomization inference. We implement the two asymptotic tests by the equivalent industry-level regressions described in Borusyak et al. (2020). RI is based on the test statistic $\sum_{\ell t} z_{\ell t}(y_{\ell t}^{\perp} - b x_{\ell t}^{\perp})$ which residualizes on the control vector and leverages the known symmetry of g_{nt}^{*} around zero to specify counterfactual shocks as $\check{g}_{nt} = g_{nt}^{*}\xi_{nt}$ where ξ_{nt} equals 1 or -1 with equal probability. We normalize the true value of β to zero in each simulation of Panels A and B; for Panels C and D we normalize the heterogeneous true effects by subtracting a constant in such a way that the median of the second-stage coefficients across simulations is zero.

D.5 Model-Implied Instruments

Adão et al. (2020) develop a spatial general equilibrium model and propose a novel way to identify its parameters. They assume that a researcher observes shifters g to some model's primitives, such as trade costs. They then propose log-linearizing the model around the initial equilibrium to derive an estimating equation for the changes in some regional outcome y_ℓ, of

$$y_\ell = \sum_n m_{\ell n}(w, \beta)\, g_n + \varepsilon_\ell, \tag{S17}$$

where $m_{\ell n}(w, \beta)$ represents the general-equilibrium elasticity of y_ℓ with respect to g_n as a function of predetermined equilibrium variables in w, and ε_ℓ captures unobserved shocks. Adão et al. (2020) treat w as non-stochastic and impose exogeneity of the observed shifters by assuming[82]

$$\mathbb{E}[\varepsilon \mid g] = 0. \tag{S18}$$

They then directly apply the classic efficiency result of Chamberlain (1987) to derive an optimal "model-implied" generalized method of moments estimator of β.

Our framework clarifies and permits a relaxation of an assumption that predetermined equilibrium is exogenous, which Adão et al. (2020) implicitly make in this setting. The identification assumption (S18) is stronger than it might appear because the predetermined equilibrium variables in w are implicitly conditioned on, and therefore treated as exogenous. To see this assumption clearly, note that (S18) implies $\mathbb{E}[\varepsilon_\ell] = 0$ by the law of iterated expectations. While innocuous under *iid*-sampling, this condition is strong in the interdependent economy, as it requires the unobserved shocks to be on average the same for each region ℓ,

[82] More precisely, they use the property of their model that g_n enter (S17) via particular shift-share averages $\eta = \{\eta_\ell\}_{\ell=1}^{L}$, and assume $\mathbb{E}[\varepsilon \mid \eta] = 0$. This detail does not affect our discussion of this setting.

regardless of the local characteristics in w like industry composition or migration shares.[83] Our framework allows for a weaker assumption of $\mathbb{E}[\varepsilon \mid g] = \mathbb{E}[\varepsilon]$ (Assumption 2(i)) without restricting $\mathbb{E}[\varepsilon_\ell] = 0$, and therefore allows the predetermined equilibrium to be endogenous; identification then follows from appropriate recentering. Since the treatment in (S17) is linear in shocks, recentering simply requires recentering g_n by their conditional expectation (as in Section D.4; see also Appendix C.8 for a discussion of how the Section 3 framework is extended to nonlinear equations). Our generalization of Chamberlain (1987) (similarly extended in Appendix C.8) further shows how optimal recentered instruments can be constructed in this setting to allow for both endogeneity of the predetermined equilibrium as well as potential non-*iid*ness of the unobserved error ε_ℓ. As always, randomization inference can be used to account for unspecified error dependence.

D.6 Centralized School Assignment Instruments

Our approach also applies to settings in which instruments arise from centralized assignment mechanisms, such as those used to allocate public school seats to students. Abdulkadiroglu et al. (2017) first show how centralized school assignments can be used as valid instruments when generated by mechanisms satisfying the "equal treatment of equals" property, in which students with the same school preferences and administrative priorities face the same assignment propensity. They use market design theory to derive large-sample approximations to this assignment risk in deferred acceptance mechanisms and further show how such assignment propensity scores can be simulated by redrawing the randomized lottery numbers which break ties between students with the same preferences and priorities. Abdulkadiroglu et al. (2019) extend this framework to deferred acceptance mechanisms with discontinuities in assignment rules (e.g. over scores from a school entrance exam) by showing how large-sample assignment risk can again be computed.

Our framework nests this setting by writing indicators for assignment of student ℓ to a given school (or any other function of centralized assignments) as $z_\ell = f_\ell(g; w)$, where $(g; w)$ partitions the inputs of a given assignment mechanism. The shock vector g might for example contain the set of tie-breaking lottery numbers in a stochastic deferred acceptance mechanism, with w containing the set of students' rankings over schools and administrative school priorities. The discontinuities in Abdulkadiroglu et al. (2019) may be accommodated by a local randomization approach. Our expected instrument μ_ℓ would then coincide with the assignment propensity scores defined in this literature.

Our analysis offers two new insights to this setting. First, expected assignments may be generated by simulating any mechanism with a random or locally random component

[83] This point is more clearly seen in our framework, which views w as potentially stochastic and writes (S18) as $\mathbb{E}[\varepsilon \mid g, w] = 0$, which in turn implies $\mathbb{E}[\varepsilon \mid w] = 0$.

g, whether or not it satisfies equal treatment of equals or has the deferred acceptance structure. Examples include top trading cycle mechanisms with single or multiple tie-breakers or mechanisms that incorporate affirmative action constraints (e.g. Angrist et al. 2019). The validity of assignment instruments that recenter by or control for these expected instruments arises simply from the exogenous variation in g. Second, valid finite-sample tests and confidence intervals can also be obtained by simulating the mechanism. This RI approach accounts for the inherent dependencies of school offers across students. It remains valid when potential outcomes are not independent, as when applicants with similar school preferences and priorities are similar in other unobserved ways, or are affected by common unobserved shocks.

D.7 Mass Media Access Instruments

Our approach further applies to a literature estimating the effects of access to mass media. For example, Olken (2009) studies the effects of television and radio access on social capital in Indonesia, while Yanagizawa-Drott (2014) estimates the impact of radio-based propaganda on violence in the 1994 Rwandan Genocide (see also Enikolopov et al. (2011), Della Vigna et al. (2014), and Wang (2020)). Papers in this literature recognize that local media access is a treatment that combines variation in the location of television or radio transmitters and in local topographic features (such as mountain ranges) that can inhibit transmission. Viewing the latter, but not the former, variation as plausibly exogenous, Olken (2009) controls for a measure of "free-space signal strength" that ignores local topography, while Yanagizawa-Drott (2014) controls for a quadratic in the distance to radio transmitters (noting that the power density of radio signals decreases in squared distance). In our framework, where media access can be written $f_\ell(g; w)$ with g denoting topographical features and w denoting transmitter location, the free-space signal strength control can be written $f_\ell(0; w)$ and may be absorbed by the Yanagizawa-Drott (2014) controls.

Our approach to isolating exogenous topographical features would be to instead control for $\mu_\ell = \mathbb{E}\left[f_\ell(g; w) \mid w\right]$, for some (perhaps uniform) permutation of such features across the map. This μ_ℓ will generally differ from both control strategies in the literature and may better remove transmitter location-driven variation. For instance, imagine that in free space signal quality decays by distance-squared, but in a hilly terrain the decay is faster. Since the counterfactuals underlying μ_ℓ involve other hilly terrains, our control will automatically rely on the appropriate distance elasticity, when the free-space controls would not.[84] Naturally, the superiority of the μ_ℓ control relies on the similarity between the actual and proposed

[84] To illustrate this point a bit more poetically, Yanagizawa-Drott (2014) motivates this identification strategy by pointing out that Rwanda is called "The Land of the Thousand Hills." Our approach suggests that media access in The Land of Zero Hills (the free-space measure) may not capture the relevant geographic confounders as well as the average media access in the Lands of the Thousand Random Hills would.

counterfactual hills—an assumption that we formalize by viewing them as equally likely, and that gives empirical content to viewing geographic features as exogenous.

D.8 Weather Instruments

Finally, our approach applies to empirical designs leveraging spatial variation in weather, such as rainfall on the days of elections and other political events. It is standard in this literature to measure "normal weather" in each location as the average of historical data and then subtract it from the actual weather (e.g. Gomez et al. 2007) or control for it (e.g. Madestam et al. 2013). These two approaches to the OVB problem directly parallel our general solutions, given an assumption of stationary weather. Our framework implies that this assumption is not important for identification: any meteorological model that yields a weather distribution for the event days could similarly be used for recentering. Further, randomization inference is natural in this setup. While Lind (2019) has shown that conventional modes of inference have severely distorted coverage because of the spatial correlation in both weather and residuals, Cooperman (2017) has addressed this problem by permutation tests based on historical weather maps. Applying similar permutations to obtain confidence intervals for the actual estimates would be in line with our general RI framework.

E Supplementary Proofs

E.1 Proof of Proposition S1 and Corollaries

Proposition S1 We write $y_\ell = y_\ell(\underline{\chi}, w, e) + \int_{\underline{\chi}}^{x_\ell} \beta_\ell(\chi, w, e) d\chi$. Note that $\mathbb{E}\left[\tilde{z}_\ell y_\ell(\underline{\chi}, w, e)\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell y_\ell(\underline{\chi}, w, e) \mid w, e\right]\right] = 0$ by the law of iterated expectations and Assumption 1. Thus,

$$\begin{aligned}
\mathbb{E}\left[\tilde{z}_\ell y_\ell\right] &= \mathbb{E}\left[\tilde{z}_\ell \int_{\underline{\chi}}^{x_\ell} \beta_\ell(\chi, w, e) d\chi\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\int_{\underline{\chi}}^{x_\ell} \beta_\ell(\chi, w, e) \tilde{z}_\ell d\chi \mid e, w\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\int_{\underline{\chi}}^{\bar{\chi}} \beta_\ell(\chi, w, e) \tilde{z}_\ell \mathbf{1}\left[x_\ell \geq \chi\right] d\chi \mid e, w\right]\right] \\
&= \mathbb{E}\left[\int_{\underline{\chi}}^{\bar{\chi}} \beta_\ell\left(\chi, w, e\right) \phi_\ell(\chi, w, e) d\chi\right] \qquad \text{(S19)}
\end{aligned}$$

where, since $\mathbb{E}\left[\tilde{z}_\ell \mid e, w\right] = 0$ by Assumption 1,

$$\begin{aligned}\phi_\ell(\chi, w, e) &= \mathbb{E}\left[\tilde{z}_\ell \mathbf{1}\left[x_\ell \geq \chi\right] \mid e, w\right] \\ &= \mathrm{Cov}\left[\tilde{z}_\ell, \mathbf{1}\left[x_\ell \geq \chi\right] \mid e, w\right].\end{aligned} \tag{S20}$$

By similar steps we can write $\mathbb{E}\left[\tilde{z}_\ell x_\ell\right] = \mathbb{E}\left[\int_{\underline{\chi}}^{\bar{\chi}} \phi_\ell(\chi, w, e) d\chi\right]$. Note that

$$\phi_\ell(\chi, w, e) = \mathrm{Cov}\left[\tilde{z}_\ell, Pr\left(x_\ell \geq \chi \mid z_\ell, e, w\right) \mid e, w\right], \tag{S21}$$

again by the law of iterated expectations. Thus when $Pr\left(x_\ell \geq \chi \mid z_\ell = z, e, w\right)$ is weakly increasing in z for each χ almost-surely, $\phi_\ell(\chi, w, e) \geq 0$ almost-surely and

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \beta_\ell\left(\chi, w, e\right) \omega_\ell(\chi, w, e) d\chi\right], \tag{S22}$$

where

$$\omega_\ell(\chi, w, e) = \frac{\phi_\ell(\chi, w, e)}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \phi_\ell(\chi, w, e) d\chi\right]} \geq \text{ a.s.} \tag{S23}$$

gives a weighting scheme with $\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \omega_\ell(\chi, w, e) d\chi\right] = 1$.

Corollary S1 In the reduced-form linear heterogeneity case, where $\beta_\ell = \beta_\ell\left(w, e\right)$, the proof of Proposition S1 simplifies considerably:

$$\begin{aligned}\mathbb{E}\left[\tilde{z}_\ell y_\ell\right] &= \mathbb{E}\left[\tilde{z}_\ell\left(\beta_\ell z_\ell + \varepsilon_\ell\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell\left(\beta_\ell z_\ell + \varepsilon_\ell\right) \mid e, w\right]\right] \\ &= \mathbb{E}\left[\beta_\ell \mathrm{Var}\left[\tilde{z}_\ell \mid w\right]\right]\end{aligned}$$

by the law of iterated expectations and using Assumption 1. Similarly, $\mathbb{E}\left[\tilde{z}_\ell z_\ell\right] = \mathbb{E}\left[\mathrm{Var}\left[\tilde{z}_\ell\right]\right]$, yielding the desired result.[85]

[85] The same result can be obtained by directly applying Proposition S1 and verifying that $\int_{\underline{\chi}}^{\bar{\chi}} \phi_\ell(\chi, w, e) d\chi = \mathrm{Var}\left[\tilde{z}_\ell \mid w\right]$ in this case.

Corollary S2 Here

$$\begin{aligned} y_\ell &= y_\ell(0) + \beta_\ell x_\ell \\ &= y_\ell(0) + \beta_\ell x_\ell(0) + \beta_\ell \left(x_\ell(1) - x_\ell(0)\right) z_\ell \\ &= y_\ell(0) + \beta_\ell x_\ell(0) + \beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \left(\tilde{z}_\ell + \mu_\ell\right) \end{aligned} \tag{S24}$$

and

$$\begin{aligned} &\mathbb{E}\left[\tilde{z}_\ell \left(y_\ell(0) + \beta_\ell x_\ell(0) + \beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \mu_\ell\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \left(y_\ell(0) + \beta_\ell x_\ell(0) + \beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \mu_\ell\right) \mid w\right]\right] = 0, \end{aligned} \tag{S25}$$

by the law of iterated expectations and Assumption 1. Thus,

$$\begin{aligned} \mathbb{E}\left[\tilde{z}_\ell y_\ell\right] &= \mathbb{E}\left[\beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \tilde{z}_\ell^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \tilde{z}_\ell^2 \mid w\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell \left(x_\ell(1) - x_\ell(0)\right) \mid w\right] \mathbb{E}\left[\tilde{z}_\ell^2 \mid w\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\beta_\ell \mid x_\ell(1) > x_\ell(0), w\right] p_\ell \sigma_\ell^2\right] \end{aligned} \tag{S26}$$

where the second equality again uses the law of expectations, the third equality follows by Assumption 1, and the fourth equality follows by definition of σ_ℓ^2 and when p_ℓ is almost-surely non-negative. Similar steps show that $\mathbb{E}\left[\tilde{z}_\ell x_\ell\right] = \mathbb{E}\left[p_\ell \sigma_\ell^2\right]$, so

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\beta_\ell \mid x_\ell(1) > x_\ell(0), w\right] \tilde{\omega}_\ell\right] \tag{S27}$$

where

$$\tilde{\omega}_\ell = \frac{p_\ell \sigma_\ell^2}{\frac{1}{L}\sum_\ell \mathbb{E}\left[p_\ell \sigma_\ell^2\right]}. \tag{S28}$$

E.2 Proof of Proposition S2

Proof of $\hat{\beta}$ consistency. We have

$$\begin{aligned} \tilde{\beta} - \beta &= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell} \\ &= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{M} \left(1 + o_p(1)\right) \end{aligned} \tag{S29}$$

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

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right] = \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right)^2\right] &= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \varepsilon_\ell \varepsilon_m\right] \\
&= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid w\right]\right] \\
&\leq \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\right|\sqrt{\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right]\mathbb{E}\left[\varepsilon_m^2 \mid w\right]}\right] \\
&\leq B\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0 \qquad \text{(S30)}
\end{aligned}$$

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

Proof of RI test consistency. Assumption 1 is stronger than the shock exogeneity assumptions of part (i), hence $\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell \xrightarrow{p} 0$. Note that

$$\begin{aligned}
T = \frac{1}{L}\sum_\ell \tilde{z}_\ell (y_\ell - b x_\ell) &= \frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell + (\beta - b)\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \\
&\xrightarrow{p} (\beta - b)\, M \neq 0. \qquad \text{(S31)}
\end{aligned}$$

For the test to be consistent it is then enough that $\frac{1}{L}\sum_\ell \tilde{z}^*_\ell (y_\ell - b x_\ell) \xrightarrow{p} 0$ for $\tilde{z}^*_\ell = f_\ell(g^*, w) - \mu_\ell$. For any b,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}^*_\ell (y_\ell - b x_\ell)\right] &= \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\tilde{z}^*_\ell (y_\ell - b x_\ell) \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[\tilde{z}^*_\ell \mid w\right]\mathbb{E}\left[y_\ell - b x_\ell \mid w_\ell\right]\right] = 0 \qquad \text{(S32)}
\end{aligned}$$

by the definition of $\tilde{z}^*_\ell$ and the law of iterated expectations. Furthermore,

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}^*_\ell(y_\ell - bx_\ell)\right] &= \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}^*_\ell\left(y_\ell - bx_\ell\right)\right)^2\right] \\
&= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\mathbb{E}\left[\tilde{z}^*_\ell \tilde{z}^*_m \mid w\right]\mathbb{E}\left[\left(y_\ell - bx_\ell\right)\left(y_m - bx_m\right) \mid w\right]\right] \\
&\le \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}^*_\ell \tilde{z}^*_m \mid w\right]\right|\sqrt{\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right]\mathbb{E}\left[\left(y_m - bx_m\right)^2 \mid w\right]}\right] \\
&\le C(b)\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0, \qquad \text{(S33)}
\end{aligned}$$

where $C(b)$ is such that $\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right] \le C(b)$ uniformly across w and ℓ, and the last line follows because the distributions of z^* and z are the same conditionally on w. The $C(b)$ bound can be constructed using the bounds for $\mathbb{E}\left[x_\ell \varepsilon_\ell \mid w\right]$ and $\mathbb{E}\left[x^2_\ell \mid w\right]$ from

$$\begin{aligned}
\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right] &= \mathbb{E}\left[\varepsilon^2_\ell + 2\left(\beta - b\right)x_\ell\varepsilon_\ell + \left(\beta - b\right)^2 x^2_\ell \mid w\right] \\
&\le B + 2\left|\beta - b\right| \cdot \left|\mathbb{E}\left[x_\ell\varepsilon_\ell \mid w\right]\right| + \left(\beta - b\right)^2\mathbb{E}\left[x^2_\ell \mid w\right]. \qquad \text{(S34)}
\end{aligned}$$

Proof of Lemma S1(i). For the first statement of the lemma, we have

$$\begin{aligned}
\frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] &= \sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right] \\
&= \mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \mid w\right]\right] \\
&= \operatorname{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\right] \to 0, \qquad \text{(S35)}
\end{aligned}$$

where the first line uses $\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \ge 0$ a.s., the second line rearranges the terms, and the third line follows by the law of total variance because $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \mid w\right] = 0$.

For the second statement, we first establish two general lemmas.

Lemma S1. *If $h\colon \mathbb{R}^N \to \mathbb{R}$ is monotone and random variables $g_1, \dots, g_N$ are independent, then for any $k \in \{1, \dots, N-1\}$ the conditional expectation $\mathbb{E}\left[h\left(g_1, \dots, g_N \mid g_1, \dots g_k\right)\right]$ is monotone.*

Proof: Denote the cumulative distribution function of g_n by $G_n(\cdot)$ and consider the $N \times 1$ vector $g' = (g'_1, \dots, g'_k, g_{k+1}, \dots, g_N)$, with $g'_n \ge g_n$ for $n \le k$. Then

$h(g') \geq h(g)$ by monotonicity. Therefore,

$$\begin{aligned}\mathbb{E}\left[h\left(g' \mid g_1, \ldots g_k\right)\right] &= \int \cdots \int h\left(g'\right) dG_{k+1}(g_{k+1}) \ldots dG_N(g_N) \\ &\geq \int \cdots \int h\left(g\right) dG_{k+1}(g_{k+1}) \ldots dG_N(g_N) \\ &= \mathbb{E}\left[h(g \mid g_1, \ldots g_k)\right], \end{aligned} \tag{S36}$$

as required.

Lemma S2. *For any monotone $h_1, h_2 \colon \mathbb{R}^N \to \mathbb{R}$, $\operatorname{Cov}\left[h_1(g), h_2(g)\right] \geq 0$ for $g = (g_1, \ldots, g_n)$ with independent components.*

Proof: For $N = 1$ this is well known. For $N > 1$ we prove that by induction. Suppose it is true for $N - 1$. Then by the law of total covariance

$$\operatorname{Cov}\left[h_1(g), h_2(g)\right] = \mathbb{E}\left[\operatorname{Cov}\left[h_1(g), h_2(g) \mid g_1\right]\right] + \operatorname{Cov}\left[\mathbb{E}\left[h_1(g) \mid g_1\right], \mathbb{E}\left[h_2(g) \mid g_1\right]\right]. \tag{S37}$$

The first term is the expectation of a covariance between two monotone functions of $N-1$ variables, where monotonicity follows by Lemma S1. The second term, again by Lemma S1, is a covariance of two monotone functions of random scalars. Thus both of the terms are non-negative.

Applying Lemma S2 to $\tilde{z}_\ell = f_\ell(g, w) - \mu_\ell(w)$ and $\tilde{z}_m = f_m(g, w) - \mu_m(w)$ and conditioning on w everywhere, we obtain the second result of Lemma S1(i).

Proof of Lemma S1(ii). Suppose $\mathbb{E}\left[\tilde{z}_\ell^2 \mid w\right] \leq B_Z$ a.s. for all ℓ. For ℓ and m such that $\mathbf{1}\left[G_\ell \cap G_m = \emptyset\right]$, $\tilde{z}_\ell \perp\!\!\!\perp \tilde{z}_m \mid w$ because f_ℓ and f_m are functions of two non-overlapping subvectors of g, the components of which are conditionally independent. Thus $\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] = 0$ a.s. for such (ℓ, m) pairs. We therefore obtain

$$\begin{aligned}\frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] &= \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right]\mathbb{E}\left[\sum_{\ell,m}\left|\operatorname{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \\ &\leq \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right]\mathbb{E}\left[\sqrt{\operatorname{Var}\left[\tilde{z}_\ell \mid w\right]\operatorname{Var}\left[\tilde{z}_m \mid w\right]}\right] \\ &\leq B_Z \cdot \frac{1}{L^2}\sum_{\ell,m}\mathbf{1}\left[G_\ell \cap G_m \neq \emptyset\right] \to 0. \end{aligned} \tag{S38}$$

E.3 Proof of Proposition S3

Suppose the null $\beta = b$ holds. The acceptance region $R = \left[T_{\alpha/2}, T_{1-\alpha/2}\right]$ is non-stochastic conditionally on (ε, w). Thus

$$Pr\left(T^* \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid y, x, w\right) \geq 1-\alpha \tag{S39}$$

by construction, with equality if $T^* \mid (\varepsilon, w)$ is continuous.

By Assumption 1, the distribution $g \mid (\varepsilon, w)$ is the same as $g \mid w$, which in turn is the same as the distribution of $g^* \mid (\varepsilon, w)$ as $g^* \perp\!\!\!\perp \varepsilon \mid w$. Therefore, conditionally on (ε, w), T and T^* have the same distribution, yielding

$$Pr\left(T \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid \varepsilon, w\right) \geq 1-\alpha. \tag{S40}$$

E.4 Proof of Proposition S4

Let $\hat{R}\left(t, e\right) = \int \mathbf{1}\left[T\left(\gamma, e\right) \leq t\right] dG\left(\gamma\right)$ denote the re-randomization distribution of the normalized RI test statistic. We first prove that when testing the correct null, i.e. for $e = \varepsilon$, this cdf converges in probability to $\Phi\left(t/\sqrt{\tilde{V}}\right)$ for each t, where $\Phi\left(\cdot\right)$ is the cdf of the standard normal distribution. By Assumption S6 and the Law of Iterated Expectations

$$\mathbb{E}\left[\hat{R}\left(t, \varepsilon\right)\right] = Pr\left(T\left(g^*, \varepsilon\right) \leq t\right) \to \Phi\left(t/\sqrt{\tilde{V}}\right). \tag{S41}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\left[\hat{R}\left(t, \varepsilon\right)^2\right] &= \mathbb{E}\left[\int\int \mathbf{1}\left[T\left(\gamma_1, \varepsilon\right) \leq t\right] \mathbf{1}\left[T\left(\gamma_2, \varepsilon\right) \leq t\right] dG\left(\gamma_1\right) dG\left(\gamma_2\right)\right] \\
&= Pr\left(T\left(g_1^*, \varepsilon\right) \leq t,\ T\left(g_2^*, \varepsilon\right) \leq t\right) \\
&\to \Phi^2\left(t/\sqrt{\tilde{V}}\right),
\end{aligned} \tag{S42}$$

where the last line again uses Assumption S6. Thus $\text{Var}\left[\hat{R}\left(t, \varepsilon\right)\right] = \mathbb{E}\left[\hat{R}^2\left(t, \varepsilon\right)\right] - \mathbb{E}\left[\hat{R}\left(t, \varepsilon\right)\right]^2 \to 0$, showing that $\hat{R}\left(t, \varepsilon\right) \xrightarrow{p} \Phi\left(t/\sqrt{\tilde{V}}\right)$.

Since the normal distribution is continuous, convergence of the re-randomization cdf implies convergence in probability of the RI critical values $T_{\alpha/2}$ and $T_{1-\alpha/2}$ by Lemma 11.2.1(ii) of Lehmann and Romano (2006): $T_{\alpha/2} \xrightarrow{p} \sqrt{\tilde{V}}\Phi^{-1}(\alpha/2)$ and $T_{1-\alpha/2} \xrightarrow{p} \sqrt{\tilde{V}}\Phi^{-1}(1-\alpha/2)$ where $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function.

Now consider the RI procedure for testing the local alternative b_L. The randomization test is based on the statistic

$$
\begin{aligned}
T\left(g^{*}, y-b_{L} x\right) & =r_{L} \frac{1}{L} f\left(g^{*}\right)^{\prime}\left(\varepsilon+x \cdot \delta / r_{L}\right) \\
& =r_{L} T\left(g^{*}, \varepsilon\right)+\delta \frac{1}{L} f\left(g^{*}\right)^{\prime} x .
\end{aligned} \tag{S43}
$$

While the first term converges to a distribution as before, the second term converges to zero in probability under the assumptions of Proposition S2(ii) (see equation (S33)). Thus, by contiguity, the RI critical values are asymptotically the same and converge in probability to $\sqrt{\tilde{V}}\Phi^{-1}(\alpha/2)$ and $\sqrt{\tilde{V}}\Phi^{-1}(1-\alpha/2)$. In contrast, the asymptotic distribution of the test statistic is shifted by δM:

$$
\begin{aligned}
T\left(g, y-b_{L} x\right) & =r_{L} \frac{1}{L} z^{\prime}\left(\varepsilon+x \cdot \delta / r_{L}\right) \\
& =r_{L} T(g, \varepsilon)+\delta \frac{1}{L} f(g)^{\prime} x \\
& \xrightarrow{d} \mathcal{N}\left(\delta M, \sqrt{\tilde{V}}\right) .
\end{aligned} \tag{S44}
$$

Therefore, with Z denote a standard normal variable, the limiting power of the RI test equals

$$
\begin{aligned}
& Pr\left(\delta M+\sqrt{\tilde{V}} \cdot Z<\sqrt{\tilde{V}} \Phi^{-1}(\alpha / 2)\right)+Pr\left(\delta M+\sqrt{\tilde{V}} \cdot Z>\sqrt{\tilde{V}} \Phi^{-1}(1-\alpha / 2)\right) \\
& =Pr\left(Z<\Phi^{-1}(\alpha / 2)-\delta M / \sqrt{\tilde{V}}\right)+Pr\left(-Z<\delta M / \sqrt{\tilde{V}}-\Phi^{-1}(1-\alpha / 2)\right) \\
& =\Phi\left(\Phi^{-1}(\alpha / 2)-\delta / \sqrt{V}\right)+\Phi\left(\Phi^{-1}(\alpha / 2)+\delta / \sqrt{V}\right),
\end{aligned} \tag{S45}
$$

by symmetry of $\Phi(\cdot)$. Differentiating (S45) by V yields

$$
-\frac{1}{2} V^{-3/2} \cdot \delta\left[\Phi^{\prime}\left(\Phi^{-1}(\alpha / 2)+\delta / \sqrt{V}\right)-\Phi^{\prime}\left(\Phi^{-1}(\alpha / 2)-\delta / \sqrt{V}\right)\right] . \tag{S46}
$$

It is clear that this derivative is negative, since the standard normal density $\Phi'(\cdot)$ is an even function that increases towards zero, and $\Phi^{-1}(\alpha/2)+\delta/\sqrt{V}$ is closer to zero than $\Phi^{-1}(\alpha/2)-\delta/\sqrt{V}$ if and only if $\delta>0$, since $\Phi^{-1}(\alpha/2)<0$. This concludes the proof.

E.5 Proof of Proposition S5

By the mean value theorem, $\mu_\ell\left(\hat{\theta}, w\right)-\mu_\ell(\theta, w)=\frac{\partial \mu_\ell}{\partial \theta}\left(\theta^*, w\right)^{\prime}\left(\hat{\theta}-\theta\right)$ for some $\theta^* \in \Theta$ and with $\frac{\partial \mu_\ell}{\partial \theta}$ component-wise bounded by a scalar B_μ. Thus, for any variable v_ℓ satisfying

$\frac{1}{L}\sum_\ell |v_\ell| = O_p(1)$,

$$\begin{aligned}\left|\frac{1}{L}\sum_\ell v_\ell\left(\mu_\ell\left(\hat{\theta},w\right)-\mu_\ell\left(\theta,w\right)\right)\right| &\le \frac{1}{L}\sum_\ell \left|v_\ell\left(\mu_\ell\left(\hat{\theta},w\right)-\mu_\ell\left(\theta,w\right)\right)\right| \\ &= \frac{1}{L}\sum_\ell \left|v_\ell \frac{\partial \mu_\ell}{\partial\theta}\left(\theta^*,w\right)'\left(\hat{\theta}-\theta\right)\right| \\ &\le \left(\frac{1}{L}\sum_\ell |v_\ell|\right) B_\mu \left\|\hat{\theta}-\theta\right\|_1 \xrightarrow{p} 0. \end{aligned} \tag{S47}$$

Therefore, with $\hat{z}_\ell = z_\ell - \mu_\ell\left(\hat{\theta},w\right)$,

$$\frac{1}{L}\sum_\ell \hat{z}_\ell x_\ell = \frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell - \frac{1}{L}\sum_\ell x_\ell\left(\mu_\ell\left(\hat{\theta},w\right)-\mu_\ell\left(\theta,w\right)\right) \xrightarrow{p} M \neq 0 \tag{S48}$$

and

$$\frac{1}{L}\sum_\ell \hat{z}_\ell \varepsilon_\ell = \frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell - \frac{1}{L}\sum_\ell \varepsilon_\ell\left(\mu_\ell\left(\hat{\theta},w\right)-\mu_\ell\left(\theta,w\right)\right) \xrightarrow{p} 0, \tag{S49}$$

where the first line uses Assumption S3 and stochastic boundedness of $\frac{1}{L}\sum_\ell |x_\ell|$, and the second line follows from Proposition S2 and stochastic boundedness of $\frac{1}{L}\sum_\ell |\varepsilon_\ell|$. Together equations (S48) and (S49) show consistency of the plug-in recentered estimator $\sum_\ell \hat{z}_\ell y_\ell / \sum_\ell \hat{z}_\ell x_\ell$.

E.6 Proof of Proposition S6

For part (i) observe that $g \perp\!\!\!\perp \varepsilon^\perp \mid w$ because $g \perp\!\!\!\perp (a,\varepsilon) \mid w$. Therefore, $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell^\perp\right] = 0$ by the law of iterated expectations, yielding identification. (A proof under a weaker exogeneity assumption $\mathbb{E}\left[\varepsilon_\ell \mid g,a,w\right] = \mathbb{E}\left[\varepsilon_\ell \mid a,w\right]$ can be constructed along the lines of Proposition 1, see equation (16)). Part (ii) follows because under the null the distribution of $g \mid \varepsilon^\perp, w$ is the same as $g \mid w$, by independence established in part (i). Part (iii) is analogous to the proof of Proposition 2 for the μ_ℓ-controlled regression (Appendix B.2). Part (iv) follows from the fact that for any variable v_ℓ, $\frac{1}{L}\sum_\ell z_\ell v_\ell^\perp = \frac{1}{L}\sum_\ell \tilde{z}_\ell v_\ell^\perp$ because $\frac{1}{L}\sum_\ell \mu_\ell v_\ell^\perp = 0$ by the properties of projection. Finally, for part (v) we write $\tilde{\beta}_\perp - \beta = \frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell / \frac{1}{L}\sum_\ell x_\ell^\perp \tilde{z}_\ell$. We first show that the numerator converges to zero in probability. We have:

$$\frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell = \frac{1}{L}\sum_\ell \varepsilon_\ell \tilde{z}_\ell - \hat{\alpha}_\varepsilon'\left(\frac{1}{L}\sum_\ell a_\ell \tilde{z}_\ell\right). \tag{S50}$$

By Proposition S2(i), $\frac{1}{L}\sum_\ell \varepsilon_\ell \tilde{z}_\ell = o_p(1)$. Moreover, using $\mathbb{E}\left[a_{\ell r}^2 \mid w\right] \le B_a$, $g \perp\!\!\!\perp a \mid w$, and Assumption S5 and applying the proof of Proposition S2(i) with $a_{\ell r}$ in place of ε_ℓ yields $\frac{1}{L}\sum_\ell a_{\ell r}\tilde{z}_\ell = o_p(1)$ for each $r = 1,\ldots,R$. Since $\hat{\alpha}_\varepsilon = O_p(1)$, we have $\frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell = o_p(1)$.

A similar argument implies that the first stage of $\tilde{\beta}_\perp$ converges to $M \neq 0$:

$$\frac{1}{L}\sum_\ell x_\ell^\perp \tilde{z}_\ell = \frac{1}{L}\sum_\ell x_\ell \tilde{z}_\ell - \hat{\alpha}_x' \left(\frac{1}{L}\sum_\ell a_\ell \tilde{z}_\ell\right), \tag{S51}$$

where $\frac{1}{L}\sum_\ell x_\ell \tilde{z}_\ell = M + o_p(1)$ by Assumption S3 and $\hat{\alpha}_x = O_p(1)$. Therefore, $\tilde{\beta}_\perp \xrightarrow{p} \beta$.

References

Abdulkadiroglu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A. Pathak. 2019. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Working Paper.*

———. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica* 85:1373–1432.

Acemoglu, Daron, and Pablo D. Azar. 2020. "Endogenous Production Networks." *Econometrica* 88:33–82.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105:2364–2409.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2020. "General Equilibrium Indirect Effects in Space: Theory and Measurement." *Working Paper.*

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134:1949–2010.

Ahlfeldt, Gabriel M., and Arne Feddersen. 2018. "From periphery to core: Measuring agglomeration effects using high-speed rail." *Journal of Economic Geography* 18:355–390.

Allen, Treb, Caue Dobbin, and Melanie Morten. 2019. "Border Walls." *Working Paper.*

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30:98–108.

Angrist, Joshua D, Kathryn Graddy, and Guido W. Imbens. 2000. "The Interpretation of in Instrumental Variables Estimators Equations an Simultaneous Models to with the Application Demand for Fish." *Review of Economic Studies* 67:499–527.

Angrist, Joshua D, Parag A. Pathak, and Roman Andres Zarate. 2019. "Choice and Consequence: Assessing Mismatch at Chicago Exam Schools." *Working Paper.*

Aronow, Peter M. 2012. "A General Method for Detecting Interference Between Units in Randomized Experiments." *Sociological Methods and Research* 40:3–16.

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

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

Bartelme, Dominick. 2018. "Trade Costs and Economic Geography: Evidence from the U.S." *Working Paper.*

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

Basso, Gaetano, and Giovanni Peri. 2015. "The Association between Immigration and Labor Market Outcomes in the United States." *Working Paper.*

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

Berger, Roger L., and Dennis D. Boos. 1994. "P values maximized over a confidence set for the nuisance parameter." *Journal of the American Statistical Association* 89:1012–1016.

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

Berman, Nicolas, Antoine Berthou, and Jérôme Héricourt. 2015. "Export Dynamics and Sales at Home." *Journal of International Economics* 96:298–310.

Berthou, Antoine, John Jong-hyun Chung, Kalina Manova, and Charlotte Sandoz Dit Bragard. 2019. "Productivity, (Mis)allocation and Trade." *Working Paper.*

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2020. "Quasi-Experimental Shift-Share Research Designs." *NBER Working Paper 24997.*

Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric M. Zolt. 2013. "The effect of rising income inequality on taxation and public expenditures: Evidence from U.S. Municipalities and school districts, 1970-2000." *Review of Economics and Statistics* 95:1291–1302.

Campante, Filipe, and David Yanagizawa-Drott. 2018. "Long-Range Growth: Economic Development in the Global Network of Air Links." *Quarterly Journal of Economics* 133:1395–1458.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2020. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." *Working paper.*

Cattaneo, Matias D., Brigham R. Frandsen, and Rocío Titiunik. 2015. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." *Journal of Causal Inference* 3:1–24.

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

Chandra, Amitabh, and Eric Thompson. 2000. "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system." *Regional Science and Urban Economics* 30:457–490.

Chandrasekhar, Arun G, and Matthew O Jackson. 2014. "Tractable and Consistent Random Graph Models." *NBER Working Paper 20276.*

Chodorow-Reich, Gabriel, and Johannes Wieland. 2020. "Secular Labor Reallocation and Business Cycles." *Journal of Political Economy.*

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

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92:1–45.

Cooperman, Alicia Dailey. 2017. “Randomization inference with rainfall data: Using historical weather patterns for variance estimation.” *Political Analysis* 25:277–288.

Costa, Rui, Swati Dhingra, and Stephen Machin. 2019. “Trade and Worker Deskilling.” *Working Paper.*

Cullen, Julie Berry, and Jonathan Gruber. 2000. “Does Unemployment Insurance Crowd out Spousal Labor Supply?” *Journal of Labor Economics* 18:546–572.

Currie, Janet, and Jonathan Gruber. 1996. “Health Insurance Eligibility, Utilization of Medical Care, and Child Health.” *The Quarterly Journal of Economics* 111:431–466.

———. 2001. “Public health insurance and medical treatment : the equalizing impact of the Medicaid expansions.” *Journal of Public Economics* 82:63–89.

Della Vigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya. 2014. “Cross-border media and nationalism: Evidence from serbian radio in croatia.” *American Economic Journal: Applied Economics* 6:103–132.

Derenoncourt, Ellora. 2019. “Can you move to opportunity ? Evidence from the Great Migration.” *Working Paper.*

Ding, Peng, Avi Feller, and Luke Miratrix. 2016. “Randomization inference for treatment effect variation.” *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 78:655–671.

Donaldson, Dave. 2018. “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” *American Economic Review* 108:899–934.

Donaldson, Dave, and Richard Hornbeck. 2016. “Railroads and American Economic Growth: A "Market Access" Approach.” *Quarterly Journal of Economics* 131:799–858.

Doudchenko, Nick, Minzhengxiong Zhang, Evgeni Drynkin, Edoardo Airoldi, Vahab Mirrokni, and Jean Pouget-Abadie. 2020. “Causal Inference with Bipartite Designs”: 1–35.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner. 2013. “Roads and trade: Evidence from the US.” *Review of Economic Studies* 81:681–724.

Duranton, Gilles, and Matthew A Turner. 2012. “Urban Growth and Transportation.” *Review of Economic Studies* 79:1407–1440.

East, Chloe N., and Elira Kuka. 2015. “Reexamining the consumption smoothing benefits of Unemployment Insurance.” *Journal of Public Economics* 132:32–50.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. “Media and political persuasion: Evidence from Russia.” *American Economic Review* 101:3253–3285.

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

Gomez, Brad T., Thomas G. Hansford, and George A. Krause. 2007. “The Republicans should pray for rain: Weather, turnout, and voting in U.S. presidential elections.” *Journal of Politics* 69:649–663.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings.” *Journal of Political Economy* 118:536–598.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2020. “Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great

Recession and "Normal" Economic Times." *American Economic Journal: Economic Policy* 12:200–225.

Gruber, Jonathan. 2003. "Medicaid." In *Means-tested transfer programs in the United States,* 15–78. University of Chicago Press.

Hemerik, Jesse, and Jelle Goeman. 2018. "Exact testing with random permutations." *Test* 27:811–825.

Hirano, Keisuke, and Guido W. Imbens. 2004. "The Propensity Score with Continuous Treatments." Chap. Chapter 7 in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family,* 73–84.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71:1161–1189.

Horvitz, D.G., and D.J. Thompson. 1952. "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association* 47:663–685.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62:467.

Imbens, Guido W., and Paul 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 (January): 109–126.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-Specific Capital and Innovation." *American Economic Review* 108:1034–1073.

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

Lee, Youjin, and Elizabeth L. Ogburn. 2019. "Network Dependence and Confounding by Network Structure Lead to Invalid Inference": 1–29.

Lehmann, Erich L. 1986. *Testing Statistical Hypotheses.* Second edi. Springer texts in statistics.

Lehmann, Erich L, and Joseph P Romano. 2006. *Testing statistical hypotheses.* Springer Science & Business Media.

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

Lind, Jo Thori. 2019. "Spurious weather effects." *Journal of Regional Science* 59:322–354.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do Political Protests Matter? Evidence from the Tea Party Movement." *Quarterly Journal of Economics* 128:1633–1685.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60:531–542.

———. 2013. "Identification of treatment response with social interactions." *Econometrics Journal* 16:1–23.

Michaels, Guy. 2008. "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System." *Review of Economics and Statistics* 90:683–701.

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

Olken, Benjamin A. 2009. "Do television and radio destroy social capital? Evidence from indonesian villages." *American Economic Journal: Applied Economics* 1:1–33.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of regional and urban economics,* 1339–1398. Elsevier.

Rosenbaum, Paul R, and Donald 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.

Rosenbaum, Paul R. 1984. "Conditional permutation tests and the propensity score in observational studies." *Journal of the American Statistical Association* 79:565–574.

———. 2002. "Covariance adjustment in randomized experiments and observational studies." *Statistical Science* 17:286–327.

Shaikh, Azeem, and Panagiotis Toulis. 2019. "Randomization Tests in Observational Studies with Staggered Adoption of Treatment." *Working Paper.*

Southworth, Lucinda K., Stuart K. Kim, and Art B. Owen. 2009. "Properties of balanced permutations." *Journal of Computational Biology* 16:625–638.

Volpe Martincus, Christian, and Juan Blyde. 2013. "Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment." *Journal of International Economics* 90:148–161.

Wang, Tianyi. 2020. "Media, Pulpit, and Populist Persuasion: Evidence from Father Coughlin." *SSRN Electronic Journal.*

Yanagizawa-Drott, David. 2014. "Propaganda and conflict: Evidence from the Rwandan genocide." *Quarterly Journal of Economics* 129:1947–1994.