NON-RANDOM EXPOSURE TO EXOGENOUS SHOCKS:
THEORY AND APPLICATIONS

Kirill Borusyak
Peter Hull

Non-Random Exposure to Exogenous Shocks: Theory and Applications
Kirill Borusyak and Peter Hull
NBER Working Paper No. 27845
September 2020, Revised January 2021
JEL No. C21,C26,F14,I13,R40

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.

Kirill Borusyak
Drayton House
30 Gordon Street
London
WC1H 0AX
United Kingdom
k.borusyak@ucl.ac.uk

Peter Hull
University of Chicago
5757 South University Avenue
Chicago, IL 60637
and NBER
hull@uchicago.edu

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
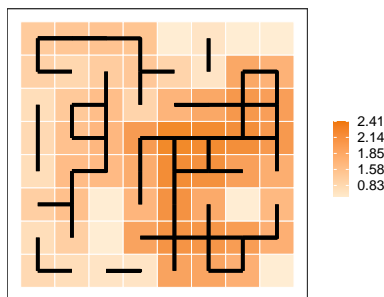
B. Expected Market Access Growth   C. Recentered Market Access Growth
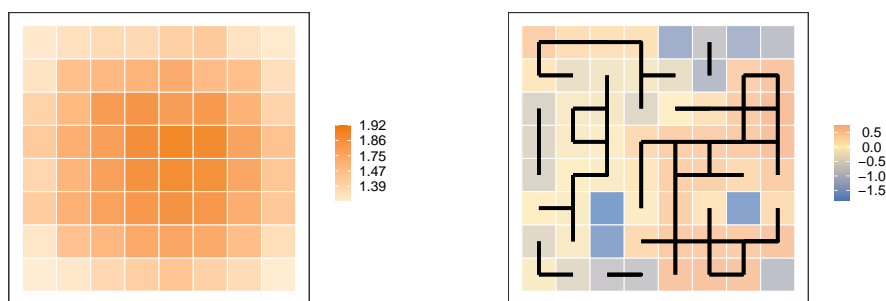
Notes: This figure illustrates the OVB problem and our recentering solution in the market access example. Panel A shows a random draw of the railroad construction experiment, with lines indicating connected regions and shading indicating corresponding market access growth (computed as described in the 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 $\ell = 1, \ldots, L$. 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 de-meaned, 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 $\ell = 1, \ldots, L$. In the *iid* case it reduces to the more familiar condition $\mathbb{E}\left[z_\ell \varepsilon_\ell\right] = 0$.

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. $\qquad\square$

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. $\qquad\square$

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 $\mathrm{Var}\left[\mathbb{E}\left[x_\ell \mid g, w\right] \mid w\right]$ 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 $G\left(g \mid w\right)$. When $G\left(\cdot\right)$ 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): $G\left(g \mid w\right)$ 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 $\Pi(g) = \{\pi(g) \mid \pi\left(\cdot\right) \in \Pi_N\}$, 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 $\{f_\ell\left(\cdot\right)\}_{\ell=1}^L$), 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 $\pi\left(\cdot\right)$.

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 $G(g \mid w)$, since μ_ℓ is typically not non-parametrically identified with non-*iid* data.[20] Nevertheless, as discussed below, our framework can apply with $G(g \mid w)$ 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 $G(g \mid w)$ explicitly: $\mu_\ell = \tilde{\mu}(w_\ell)$ 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 $\mathbb{E}[\check{z}_\ell \mid w] = 0$ or $\mathbb{E}[\check{\varepsilon}_\ell \mid w] = 0$ 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 $\frac{1}{L} \sum_\ell z_\ell (y_\ell^\perp - bx_\ell^\perp)$.

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 $T = \tilde{z}' R (R'R)^{-1} R' \tilde{z}$ can be seen as a quadratic form of the vector-valued statistic $\frac{1}{L} \sum_\ell \tilde{z}_\ell R_\ell$, weighted by $(R'R)^{-1}$, 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* $\Omega = \mathrm{Var}\left[\varepsilon \mid w\right]$. *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): $MA_{\ell t} = \sum_k \exp\left(-0.02\tau_{\ell kt}\right) \cdot P_{k,2000}$. 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
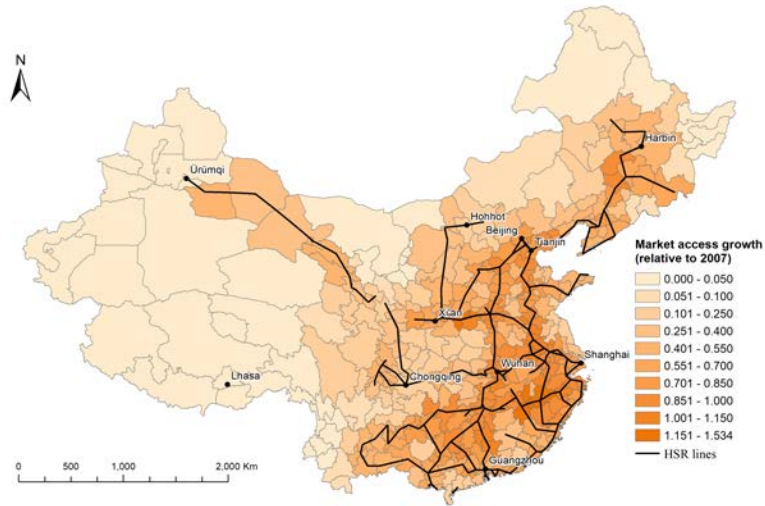
B. All Completed and Planned Lines

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
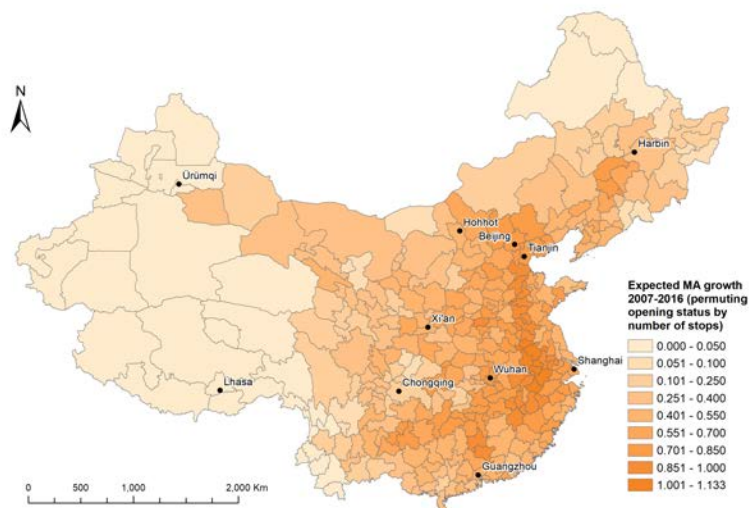
B. Recentered Market Access Growth
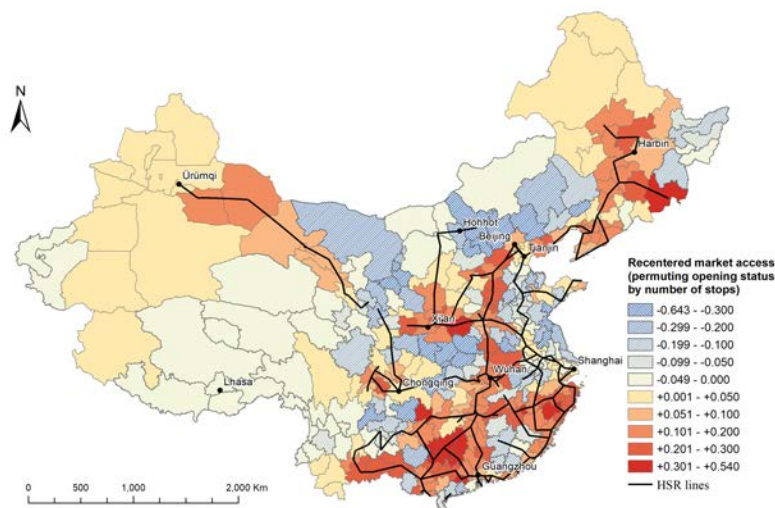
Notes: Panel A shows the variation in expected 2007–16 MA growth across Chinese prefectures, computed from 999 counterfactuals that permute the opening status of built and planned lines with the same number of cross-prefecture links. Panel B plots the variation in corresponding recentered MA growth: 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.

27

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, $\mu_\ell = \mathbb{E}[x_\ell \mid w] = \frac{1}{50} \sum_n f(g_n; v_\ell)$, 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 $z_{\ell t}^{CG}$, 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 × 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.