Non-Random Exposure to Exogenous Shocks

Kirill Borusyak
UCL and CEPR

Peter Hull
Brown and NBER*

December 2021

Abstract

We develop a new approach to 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 or transportation networks and simulated instruments for policy eligibility. 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 shock counterfactuals. We use this approach to address bias when estimating employment effects of market access growth from Chinese high-speed rail construction.

*Contact: k.borusyak@ucl.ac.uk and peter_hull@brown.edu.

We are grateful to Rodrigo Adão, Gabriel Ahlfeldt, Dmitry Arkhangelsky, Nate Baum-Snow, Sophie Calder-Wang, Vasco Carvalho, Gabriel Chodorow-Reich, Dave Donaldson, Raffaella Giacomini, Paul Goldsmith-Pinkham, Richard Hornbeck, Kilian Huber, Guido Imbens, Xavier Jaravel, Tetsuya Kaji, Vishal Kamat, Michal Kolesár, Whitney Newey, Aureo de Paula, Andrés Rodríguez-Clare, Cyrus Samii, Jesse Shapiro, Ben Sommers, Chenzi Xu, and numerous seminar participants for helpful comments. Ruixue Li, Elise Parrish, and Steven Shi provided outstanding research assistance.

1 Introduction

Many questions in economics involve the causal effects of treatments x_i which are computed from multiple sources of variation, according to a known formula. Consider three examples. First, when estimating spillovers from a randomized intervention, x_i might count the number of individual i'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 x_i measures the growth of regional market access: a treatment computed from the location and timing of transportation upgrades and the spatial distribution of economic activity in a country. A third example is an x_i capturing individual i's eligibility for a public program, such as Medicaid, which is jointly determined by the eligibility policy in i's state and her household's demographics and income.[1]

This paper develops a new approach to estimating the effects of such composite variables when some, but not all, of their determinants are generated by a true or natural experiment. We ask, for example, how one can estimate market access effects by leveraging the timing of new railroad line construction as exogenous shocks, when the other determinants of market access (such as the pre-determined location of large markets and planned lines) are non-random.

We first show that omitted variable bias (OVB) may confound conventional regression approaches in such settings. Bias arises from different observations receiving systematically different values of x_i because of their individual non-random "exposure" to the exogenous shocks. For example, even when construction is delayed for a random set of lines, regions that are economically or geographically more central will tend to see a larger growth in market access because they are closer to a typical potential line. Regression identification of market access effects then fails without an additional assumption on the exogeneity of economic geography: that more exposed (e.g., central) 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.

Our solution to the OVB challenge is based on the specification of counterfactual exogenous 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 which the re-

[1] Examples of these three settings include Miguel and Kremer (2004), Donaldson and Hornbeck (2016), and Currie and Gruber (1996), respectively. Our working paper (Borusyak and Hull 2021) discusses other common treatments and instruments nested in our framework: linear and nonlinear shift-share variables, model-implied optimal instruments, instruments based on centralized school assignment mechanisms, "free-space" instruments for access to mass media, and variables leveraging weather shocks.

searcher wishes to leverage, for instance by specifying permutations of the shocks that were as likely to have occurred. For example, if line construction delays are considered as-good-as-random, one might produce counterfactual network maps by randomly exchanging the lines which were completed earlier and later.

Valid shock counterfactuals can be used to avoid OVB by a "recentering" procedure which involves measuring and appropriately adjusting for a single confounder: the *expected treatment*, μ_i. To do so, a researcher draws counterfactual shocks from the assignment process and recomputes the instrument many times. Then, for each observation i, the treatment is averaged across these many draws to obtain μ_i. Finally, μ_i is subtracted from x_i to obtain the *recentered treatment* $\tilde{x}_i = x_i - \mu_i$. We show that using $\tilde{x}_i$ as an instrument for x_i removes the bias from non-random shock exposure. Intuitively, observations only get high vs. low values of $\tilde{x}_i$ because the observed shocks were drawn instead of the counterfactuals, which is assumed to be by chance. For example, when μ_i is constructed by permuting the timing of new line construction, regressions that instrument with $\tilde{x}_i$ 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, closely related, solution to OVB is to include μ_i as a control in the regression of an outcome on x_i; this can be viewed as recentering x_i while also removing some residual variation in the outcome.[2]

This approach to causal inference with composite variables, in which some determinants are labeled as exogenous and characterized by an assignment process, can be seen as formalizing the natural experiment of interest and bringing composite variables to familiar econometric territory.[3] The conditions we impose on the exogenous shocks are similar to those which might be used if the shocks were directly used as treatments: e.g., if shocks to the timing of railroad line upgrades were used in a regression of outcomes defined at the "level" of those lines. Recentering ensures identification from the natural experiment, even when the regression is estimated at a different level (e.g., across regions instead of lines).

Our general identification framework further allows the treatment to have endogenous or unobserved determinants. In this case one may construct candidate composite instruments for x_i based on its exogenous and predetermined components. The same OVB problem arises in this instrumental variable (IV) case, and it can again be solved by recentering the candidate instrument by its expectation over the shock assignment process. Controlling for

[2]While recentering is the key step that removes OVB, removing residual variation is likely to increase the efficiency of estimation in large samples. We give practical recommendations for each adjustment in the paper's conclusion.

[3]Our approach is "design-based," in that identification is achieved by specifying the assignment process of some observed shocks (see, e.g., Lee (2008), Athey and Imbens (2018), Shaikh and Toulis (2019), and De Chaisemartin and Behaghel (2018)). It contrasts with other identification strategies that instead model the residual determinants of the outcome, such as difference-in-difference strategies (e.g. Chaisemartin and D'Haultfœuille (2020) and Athey et al. (2021)) or fully-specified structural models.

the expected instrument is again another solution.

We establish several attractive properties of the recentering approach, beyond our primary identification results. First, shock counterfactuals can be used for exact finite-sample inference and specification tests via randomization inference (RI). Second, consistency of the estimates and RI tests follows regardless of the correlation structure of unobservables, provided the exogenous shocks induce sufficient cross-sectional variation in the instrument and treatment. Finally, while our RI and consistency results rely on an assumption of constant treatment effects, identification with recentered IV estimators generalizes to heterogeneous effects under a natural first-stage monotonicity condition.

We apply this framework to estimate the employment effects of market access (MA) growth due to new high-speed rail in China. We show how recentering can help leverage variation in the timing of transportation upgrades to purge OVB. 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 unadjusted 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.

Econometrically, expected treatment and instrument adjustment is similar to propensity score methods for removing OVB (Rosenbaum and Rubin 1983), with two key differences. First, we propose using the structure of composite x_i to compute expected treatments/instruments from more primitive assumptions on the assignment process for exogenous shocks. This approach is similar to how Borusyak et al. (2021) and Aronow and Samii (2017) address OVB when using linear shift-share instruments and network treatments, respectively. It differs from conventional methods of directly estimating propensity scores; such methods are typically infeasible in the settings we consider because the exposure to exogenous shocks is high-dimensional. Second, our regression-based adjustment differs from conventional approaches of weighting by or matching on propensity scores.[4] Regression adjustment is more popular in applied research, avoids practical issues of limited overlap (due to, e.g., propensity scores that are close to zero or one), does not require to treatments or instruments to be binary, and is natural for estimating constant structural parameters or convex averages of heterogeneous treatment effects.

The remainder of this paper is organized as follows. The next section motivates our analysis with three examples related to network spillovers, market access effects, and Medicaid eligibility effects. Section 3 develops our general framework and results. Section 4

[4]A notable exception of a recentering-type regression adjustment in the traditional propensity scores setting is the E-estimator of Robins et al. (1992).

presents our application, and Section 5 concludes. Additional results and extensions are given in an earlier working paper, Borusyak and Hull (2021, henceforth BH).

2 Motivating Examples

We develop three stylized examples, inspired respectively by the settings of Miguel and Kremer (2004), Donaldson and Hornbeck (2016), and Currie and Gruber (1996), to illustrate the main insights of this paper. In each example we consider estimating the parameter β of a causal or structural model which relates an outcome y_i to a treatment x_i,

$$y_i = \beta x_i + \varepsilon_i, \tag{1}$$

for a set of units $i = 1, \dots, N$ with an unobserved error ε_i. The common feature of the examples is that x_i is computed from multiple sources of variation by a known formula.

Example 1: Network spillovers. Suppose y_i is student i's educational achievement and x_i counts the number of i's neighbors who have been dewormed in an intervention:

$$x_i = \sum_{k=1}^{N} Neighbor_{ik} Dewormed_k. \tag{2}$$

Here $Dewormed_k \in \{0, 1\}$ is an indicator for student k being selected for the deworming intervention and $Neighbor_{ik} \in \{0, 1\}$ indicates that i and k are neighbors (i.e., connected by an observed network link). The error term ε_i captures i's educational outcome when none of her neighbors are dewormed.

Example 2: Market access. Suppose y_i is the growth of land values in region i between two dates $t \in \{0, 1\}$ and $x_i = \log MA_{i1} - \log MA_{i0}$ is the growth of regional market access (MA) due to improvements to the cross-country railroad network. Market access is computed as

$$MA_{it} = \sum_{j=1}^{N} \frac{Pop_j}{\tau\left(Network_t, Loc_i, Loc_j\right)}, \tag{3}$$

following standard models of economic geography (e.g. Redding and Venables (2004)). Here Pop_j is the time-invariant population of region j, $Network_t$ is the set of railway lines and other types of transit which comprise the transportation network in operation at time t, Loc_j is the location of region j on the map, and $\tau(\cdot)$ is a function giving the travel time between regions i and j. The error term ε_i captures location i's land value growth in the absence of market access growth, due to some regional amenity and productivity shocks.

Example 3: Medicaid eligibility. Suppose y_i is individual i's health outcome and

$x_i \in \{0,1\}$ indicates her eligibility for Medicaid. Let $IncDem_i$ be a vector of individual income and demographics, $State_i \in \{1,\ldots,50\}$ index i's state of residence, and $Policy_k$ be state k's eligibility policy: i.e. the set of income and demographic groups eligible for Medicaid in that state. Then

$$x_i = \mathbf{1}\left[IncDem_i \in Policy_{State_i}\right]. \tag{4}$$

The error term ε_i captures individual i's outcome when she is ineligible for Medicaid.

To estimate β in each example, we consider an actual or natural experiment that manipulates some of the determinants of x_i. Formally, we partition the variables from which $x = (x_1,\ldots,x_N)$ is computed into two groups: a set of shocks g and a set of predetermined variables w. The shocks $g = (g_1,\ldots,g_K)$ are assumed to be exogenous, i.e. independent of the errors $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_N)$. Shock exogeneity combines two conceptually distinct assumptions—that g is as-good-as-randomly assigned, and that this assignment only affects the outcome of each unit i via its treatment x_i (an exclusion restriction). The shocks can be assigned at a different "level" than the observations, with $K \neq N$. The remaining variables w have an arbitrary structure and govern the mapping from the exogenous shocks to each unit's treatment, i.e. the observation's "exposure" to the shocks. We assume that w is determined prior to the (natural) experiment and is unaffected by the shocks.

Example 1 (cont.) Suppose deworming is assigned in a randomized control trial (RCT) and βx_i fully captures its spillover effects. Then $g = (Dewormed_k)_{k=1}^K$ collects the exogenous shocks, for $K = N$.[5] The remaining determinants of the spillover treatments of all units, $w = (Neighbor_{ik})_{i,k=1}^N$, are fixed in the experiment.

Example 2 (cont.) Suppose the timing of new railroads is exogenous. Specifically, suppose that among K lines planned to be constructed at $t = 1$ some are randomly delayed by unexpected engineering problems (unrelated to the trends in regional land values). Suppose also the model of economic geography is correctly specified, so $\beta \log MA_{it}$ fully captures the effects of transportation upgrades. Then $g = (Open_k)_{k=1}^K$ collects the exogenous shocks, where $Open_k$ is an indicator for whether planned line k faces no delays. Assuming no other changes to the network at $t = 1$, we can partition the determinants of MA growth into g and $w = \left(\{Loc_i, Pop_i\}_{i=1}^N, Network_0\right)$ as $Network_1$ is fully determined by $Network_0$ and the set of newly opened lines.

Example 3 (cont.) Suppose Medicaid policies across the $K = 50$ states are exogenous, i.e. are chosen irrespective of the potential health outcomes and affect individual outcomes only via Medicaid eligibility. Then $g = (Policy_k)_{k=1}^K$ collects the exogenous

[5] For simplicity here we assume away any direct effects of deworming; see Section 3.5 on the extension to multiple treatments.

shocks, with the other determinants of eligibility collected in $w = (IncDem_i, state_i)_{i=1}^N$.

The first point of this paper is that ordinary least squares (OLS) estimation of β can suffer from OVB, despite the experimental variation underlying x_i.[6] The OVB problem arises because some units receive systematically higher values of x_i than others, as a consequence of their non-random exposure to the shocks. This systematic variation may be cross-sectionally correlated with the errors ε_i, generating bias in OLS estimation of equation (1).

Example 1 (cont.) Even when deworming is randomly assigned to students, those with more neighbors (e.g., because they live in dense urban areas) will tend to have more dewormed neighbors and therefore be more exposed to the deworming intervention. Urban areas may have different educational outcomes for reasons unrelated to deworming, generating OVB.

Example 2 (cont.) Even when the opening status of lines is as-good-as-randomly assigned, regions in the economic and geographic center of the country will tend to see more market access growth than peripheral regions as the former are closer to a typical potential line.[7] Central regions may face different amenity and productivity shocks, generating OVB.

Example 3 (cont.) Even when Medicaid policies are as-good-as-randomly assigned to states, poorer individuals with certain demographics will tend to see higher rates of eligibility. Poor individuals may face different health shocks, generating OVB.

Our second insight is that this OVB problem has a conceptually simple solution, which follows from viewing the set of realized g as one draw from a shock "assignment process" and considering what counterfactual sets of exogenous shocks could have as likely been drawn. The specification of such counterfactuals allows one to measure and remove the systematic component of variation in the treatment which drives OVB. Specifically, the researcher recomputes the treatment x_i of each unit i across many counterfactual sets of shocks and takes their average to measure the "expected treatment," μ_i. We show that this μ_i, which is co-determined by the exposure of x_i to the shocks g and the shock assignment process, is the sole confounder in equation (1). OVB can then be purged by "recentering" the treatment, i.e. instrumenting x_i with $\tilde{x}_i = x_i - \mu_i$ in equation (1), or by simply adding

[6] Such OVB may arise even if (as in Examples 1 and 2) variation in the treatment "results" from the experimental shocks, in the sense that $x_1 = \cdots = x_N = 0$ whenever $g_1 = \cdots = g_K = 0$.

[7] Our working paper (BH, Section 2) illustrates this phenomenon with a simple example of a square map with no period-0 network. Even when the population is equal in all regions and period-1 lines connect any adjacent regions with equal probability, MA growth is systematically higher in the center of the map. When the set of planned lines or regional populations are not uniform across the map, the variation in MA growth becomes more intricate: e.g. it is systematically higher in economic centers, except for populous regions where the large local market makes connections less important.

μ_i as a control in OLS estimation. The key to removing bias with this approach is thus to credibly specify and average over shock counterfactuals—a task which is trivial in true experiments and which otherwise formalizes the natural experiment of interest.

> **Example 1 (cont.)** With deworming assigned in an RCT, the shock assignment process is given by the known randomization protocol. If, say, each student has a 30% chance of being dewormed, the expected number of i's dewormed neighbors over repeated draws of deworming shocks μ_i is 0.3 times their number of neighbors $\sum_{k=1}^{K} Neighbor_{ik}$. OVB is thus purged by controlling for the number of neighbors, or by using the recentered number of dewormed neighbors $\tilde{x}_i = x_i - \mu_i$ to instrument for x_i. With either adjustment, the regression will only compare students who had more neighbors dewormed than expected (given the network) to those with fewer than expected dewormed neighbors.
>
> **Example 2 (cont.)** The as-good-as-random assignment of opening status can be formalized by each planned line facing an equal and independent chance of opening. Then, if $\sum_{k=1}^{K} Open_k = K_1$ railway lines open by $t = 1$, every counterfactual network in which K_1 lines from the plan opened was as likely to have occurred. One can thus compute expected MA growth μ_i as the average MA growth of region i across these counterfactuals (or a random subset of them). Recentering by or controlling for this μ_i ensures that the regressions only compare regions which saw higher MA growth than expected (given pre-existing economic geography and the plan) to those which saw less-than-expected MA growth.
>
> **Example 3 (cont.)** The as-good-as-random assignment of Medicaid policies can be formalized by each state randomly drawing from a pool of potential policies, such that every permutation of the realized policies was equally likely to have occurred. Averaging individual i's eligibility across these permutations yields an expected eligibility μ_i which equals the share of states in which she would be eligible. Our solution is to instrument actual eligibility with recentered eligibility $x_i - \mu_i$, or control for μ_i in an OLS regression. Either approach would, for example, effectively remove from the sample "always-eligible" or "never-eligible" individuals (with $x_i = \mu_i = 1$ or $x_i = \mu_i = 0$) whose income and demographics make them unaffected by policy variation.

The recentering solution generally dominates more conventional ones, such as instrumenting directly by the shocks or controlling for the other determinants of x_i. Instrumenting with the shocks is infeasible when the shocks are assigned at a different level than the units, and generally discards variation in treatment due to w. Controlling for an observation's non-random shock exposure flexibly is typically infeasible, because such exposure is high-dimensional. Conversely, low-dimensional controls are only guaranteed to purge OVB (absent additional non-experimental restrictions on the error term) when they linearly span μ_i—which is difficult to establish except when μ_i is known and recentering is feasible. If either the assignment process or shock exposure mapping is complex, μ_i is unlikely to be a simple function of observed characteristics.

Example 1 (cont.) Using student i's own deworming status as an instrument is infeasible as it does not predict the number of dewormed neighbors; leveraging the non-random network adjacency matrix is necessary. Controlling for the entire row of the adjacency matrix (which characterizes student's exposure) is also infeasible, as it would absorb all cross-sectional variation in the treatment. Miguel and Kremer (2004) follow a different controlling strategy, by including the number of i's neighbors in the regression. Under completely random assignment of deworming this control is proportional to μ_i and thus purges OVB. However, such simple controls would not linearly span μ_i with more complex randomization protocols, such as with two tiers (by school, then by student) or stratification (e.g., with girls dewormed with a known higher probability). Simple controls are also generally insufficient with more complex specifications of spillovers.[8]

Example 2 (cont.) Railroad timing shocks vary at the level of lines, so it is infeasible to use them as instruments for regional market access without incorporating some non-random features of economic geography. Controlling perfectly for these features is also infeasible, as each region's market access depends on the entire spatial distribution of economic activity. Simple sets of controls, such as polynomials in the latitude and longitude of a region, need not linearly span μ_i given the complexity of x_i, and thus are not guaranteed to purge OVB.

Example 3 (cont.) Currie and Gruber (1996) propose instrumenting individual eligibility with a measure of the overall policy generosity of her state—a so-called "simulated instrument." Such instruments are simple functions of $Policy_k$ for all individuals in state k and are thus exogenous and relevant under random policy assignment. However they discard relevant within-state variation in i's income and demographics and are thus likely to yield a less powerful first-stage prediction of x_i than recentered eligibility.[9]

[8] An example is given by Carvalho et al. (2021), where i is a Japanese firm and x_i is the distance in the firm-to-firm supply network from i to the nearest firm located in the area hit by an earthquake. Unlike the number of treated neighbors, this spillover treatment is a nonlinear function of the earthquake shock dummies. The earthquake assignment process is also more complex, exhibiting spatial correlation. Our recentering approach still applies naturally in cases like this.

[9] With completely random policy assignment, flexibly controlling for $IncDem_i$ may purge OVB as this is the only source of variation in μ_i. However, even in this setting the relevant demographics in $IncDem_i$ and their interactions can be high-dimensional, as discussed by Gruber (2003). This problem is exacerbated under more complex assignment processes, e.g. if policies can be viewed as random only within some groups of states, in which case group indicators and their interactions with the demographics would also have to be

We conclude this section by noting that the OVB problem and recentering solution both extend to the case with an arbitrary endogenous x_i and a candidate instrument z_i which is constructed from exogenous shocks and other variables by a known formula. This approach is natural when the treatment can be represented as a function of exogenous shocks g, predetermined variables w, and endogenous (and possibly unobserved) variables u: i.e. when $x_i = h_i(g, w, u)$ for a known $h_i(\cdot)$. An intuitive candidate instrument for x_i is the prediction of x_i in the scenario when the u shocks are ignored: $z_i = h_i(g, w, 0)$. Our framework shows that these candidate instruments are generally invalid, again because of the non-random exposure of z_i to g. Yet OVB can again be purged by measuring the expected instrument μ_i—the average z_i across counterfactual g—and either instrumenting x_i with the recentered IV $\tilde{z}_i = z_i - \mu_i$ or controlling for μ_i while instrumenting with z_i.

Example 2 (cont.) Suppose population sizes also change between $t = 0$ and $t = 1$, and the observable changes $u = (Pop_{i1} - Pop_{i0})_{i=1}^{N}$ are not exogenous (e.g. they respond to house price shocks in ε). Then one can consider instrumenting the observed change in MA by a predicted change in MA which keeps population sizes fixed at the predetermined $t = 0$ levels. Without recentering, this IV regression may suffer from the same OVB as the OLS regression discussed above. OVB is now avoided by recentering the MA prediction via counterfactual railroad networks.

Example 3 (cont.) Suppose one is interested in the effects of Medicaid takeup, instead of eligibility. Takeup is the product of eligibility and $1 - NeverTaker_i$, where $NeverTaker_i$ indicates that individual i would decline Medicaid if eligible and $u = (NeverTaker_i)_{i=1}^{N}$ is unobserved. Under the appropriate exclusion restriction one can consider instrumenting takeup with eligibility; our recentering strategy then again removes OVB from non-random variation in policy exposure.

3 Theory

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 and estimation based on the recentering procedure, and discuss how this recentering can be performed by specifying counterfactual shocks in Sections 3.1–3.3. We then discuss how shock counterfactuals can be used for finite-sample inference and summarize several extensions in Sections 3.4 and 3.5.

---

included. Recentering extends naturally and avoids the curse of dimensionality.

3.1 Setting

We consider identification of β in the causal or structural model (1) with scalar and demeaned y_i and x_i. Below we discuss extensions to heterogeneous causal effects, nonlinear models, multiple treatments, and additional control variables. 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_i, x_i) are independently or identically distributed (*iid*) as when arising from random sampling. This allows for complex dependencies across the units due to their common exposure to observed and potentially unobserved shocks. It is also consistent with settings where the N units represent a population—for example, all regions of a country—and conventional random sampling assumptions are inappropriate (Abadie et al. 2020).[10]

We suppose that to identify β a researcher has constructed a candidate instrument

$$z_i = f_i(g; w), \tag{5}$$

where $\{f_i(\cdot)\}_{i=1}^N$ is a set of known non-stochastic functions, g is a $K \times 1$ vector of shocks, and w is a set of predetermined variables. Equation (5) is very general: any z_i that can be computed from a set of observed data, according to a known formula, can be described in this way.[11] It also allows $x_i = z_i$, in which case β is the effect of a composite treatment.

We assume that the shocks g are exogenous, which we formalize by their conditional independence from the vector of errors given the other sources of instrument variation:

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

As noted in Section 2, this notion of shock exogeneity combines two conceptually distinct conditions. First, it imposes an exclusion restriction, reflecting an economic model of how g can affect y. Second, it requires as-good-as-random shock assignment. This latter condition is satisfied when the shocks are fully randomly assigned, as in an RCT (i.e., $g \perp\!\!\!\perp (\varepsilon, w)$), but also allows w to contain variables that govern the shock assignment process.[12] Importantly, Assumption 1 allows $\mathbb{E}[\varepsilon_i \mid w]$ to vary arbitrarily across i; this reflects the lack of non-experimental assumptions, such as parallel trends, constraining the error in equation (1).[13]

[10]Formally, we assume $\{x_i, \varepsilon_i\}_{i=1}^N$ and the g and w variables introduced below are all drawn from some joint distribution which is unrestricted at this point.

[11]Note that equation (5) does not contain a residual: it formalizes an algorithm for computing an instrument rather than characterizing an economic relationship.

[12]The exclusion and as-good-as-random assignment assumptions are isolated in Appendix C.1 of BH, via a general potential outcomes model.

[13]Our identification results hold under the weaker conditional mean independence assumption of $\mathbb{E}[\varepsilon \mid g, w] = \mathbb{E}[\varepsilon \mid w]$. This assumption can be understood as defining a partially linear model, as in Robinson (1988): $y_i = \beta x_i + \psi_i(w) + \tilde{\varepsilon}_i$ where $\psi_i(w) = \mathbb{E}[\varepsilon_i \mid w]$ and $\mathbb{E}[\tilde{\varepsilon}_i \mid g, w] = 0$ for $\tilde{\varepsilon}_i = \varepsilon_i - \psi_i(w)$. A

We consider identification of β by an instrumental variable (IV) regression of y_i on x_i instrumenting with z_i. Identification follows when z_i is relevant to the treatment and orthogonal to the structural residual. In our non-*iid* setting, we formalize these conditions as $\mathbb{E}\left[\frac{1}{N}\sum_i z_i x_i\right] \neq 0$ and $\mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right] = 0$, which together imply that β is uniquely recoverable from the observable moments $\mathbb{E}\left[\frac{1}{N}\sum_i z_i x_i\right]$ and $\mathbb{E}\left[\frac{1}{N}\sum_i z_i y_i\right] = \beta\mathbb{E}\left[\frac{1}{N}\sum_i z_i x_i\right] + \mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right]$.[14]

3.2 OVB and Instrument Recentering

We define the *expected instrument* $\mu_i = \mathbb{E}\left[f_i(g;w) \mid w\right]$ as the average value of z_i across different realizations of the shocks, conditional on w. Our first result shows that IV identification fails when predetermined exposure to the natural experiment is endogenous, and that this failure is entirely governed by the relationship between μ_i and the error ε_i. Formally, under Assumption 1 the IV moment condition need not be satisfied: $\mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right] \neq 0$ in general. Rather, the moment condition violation satisfies:

$$\mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right] = \mathbb{E}\left[\frac{1}{N}\sum_i \mu_i \varepsilon_i\right]. \tag{6}$$

This result follows from the law of iterated expectations: $\mathbb{E}\left[z_i\varepsilon_i\right] = \mathbb{E}\left[\mathbb{E}\left[f_i(g;w)\varepsilon_i \mid w\right]\right] = \mathbb{E}\left[\mu_i \mathbb{E}\left[\varepsilon_i \mid w\right]\right] = \mathbb{E}\left[\mu_i\varepsilon_i\right]$ for all i, where the second equality uses Assumption 1 and the definition of μ_i.

The central role of μ_i in governing OVB immediately suggests a "recentering" solution: even though OVB results from potentially high-dimensional variation in units' exposure to shocks, adjustment for the one-dimensional confounder μ_i is sufficient for identification. We adjust z_i by defining the *recentered instrument* $\tilde{z}_i = z_i - \mu_i$. By equation (6), the IV moment condition holds for this instrument:

$$\mathbb{E}\left[\frac{1}{N}\sum_i \tilde{z}_i \varepsilon_i\right] = \mathbb{E}\left[\frac{1}{N}\sum_i z_i \varepsilon_i\right] - \mathbb{E}\left[\frac{1}{N}\sum_i \mu_i \varepsilon_i\right] = 0. \tag{7}$$

Thus, if $\tilde{z}_i$ is also relevant, IV estimation of (1) which instruments x_i with $\tilde{z}_i$ (instead of the initial z_i) identifies β.[15]

A closely related solution, also suggested by equation (6), is to include the expected instrument μ_i as a control in specification (1) while using the original z_i as the instrument.

difference from Robinson (1988) arises because we do not assume *iid* data; for instance, we do not assume $\psi_i(w) \equiv \psi(w_i)$ for *iid* w_i.

[14] Here it is worth highlighting that the orthogonality condition $\mathbb{E}\left[\frac{1}{N}\sum_i z_i\varepsilon_i\right] = 0$ combines two dimensions of variation: over the stochastic realizations of g, w, and ε, and across the cross-section of observations $i = 1, \ldots, N$. In the *iid* case it reduces to the more familiar condition $\mathbb{E}\left[z_i\varepsilon_i\right] = 0$.

[15] There exist $f_i(\cdot)$ constructions that yield a relevant recentered instrument whenever the shocks induce some variation in treatment. Formally, when $\mathrm{Var}\left[\mathbb{E}\left[x_i \mid g, w\right] \mid w\right]$ is not almost-surely zero at least for some i, the recentered instrument constructed as $\tilde{z}_i = \mathbb{E}\left[x_i \mid g, w\right] - \mathbb{E}\left[x_i \mid w\right]$ is relevant.

Controlling for μ_i can be thought of as recentering z_i while also removing the residual variation in y_i which is cross-sectionally correlated with μ_i.[16] As usual, removing this residual variation may generate precision gains in large samples; similar gains may arise from including (a fixed number of) any predetermined controls in a recentered IV regression.[17]

Equation (6) further shows that adjusting for μ_i is generally *necessary* for identification, absent additional restrictions on the unobserved error. Conventional controls and fixed effects are only guaranteed to purge OVB when they linearly span μ_i—a condition that is difficult to verify except when recentering is also feasible.[18] Data-driven procedures for selecting appropriate controls (e.g. Belloni et al. (2013) and Chernozhukov et al. (2018)) will also fail unless the set of candidate controls sparsely spans both μ_i and $\mathbb{E}\left[\varepsilon_i \mid w\right]$, which is again difficult to establish in most settings we consider.

While our primary focus is on identification, we now briefly discuss the asymptotic properties of the recentered IV estimator $\hat{\beta} = \sum_i \tilde{z}_i y_i / \sum_i \tilde{z}_i x_i$, which is the solution to the sample analog of (7). Leaving details to Section 3 of BH, we consider a sequence of data-generating processes, indexed by N, for the complete data (y, x, g, w). We show that $\hat{\beta}$ is consistent, i.e. $\hat{\beta} \xrightarrow{p} \beta$ as $N \to \infty$, under weak mutual cross-sectional dependence of $\tilde{z}_i$ and an asymptotic first stage (Proposition S2 in BH). In line with our general approach, this result does not rely on conventional sampling-based arguments that may be inappropriate with non-*iid* data. We also make no restriction on the mutual dependence of residuals, imposing only a weak regularity condition on ε_i. At a high-level, the substantive assumption on $\tilde{z}_i$ requires the recentered instrument construction to well-differentiate observations by their exposure to the exogenous shocks; BH also provide lower-level sufficient conditions.

Adjustments based on μ_i, as the sole confounder of z_i, are similar to more conventional propensity score methods. There are three key differences, concerning the setting, adjustment method, and computation of μ_i. First, propensity score methods have mostly been applied to binary treatments, starting from Rosenbaum and Rubin (1983). While generalizations to binary instruments (e.g. Abadie (2003)) and non-binary treatments (e.g. Hirano and Imbens (2004)) have been proposed, our setting allows for arbitrary domains of treatments or instruments. Second, the propensity score literature has mostly used non-regression adjustment methods, such as matching or binning (Abadie and Imbens 2016;

[16] Formally, this regression yields the reduced-form and first-stage moments $\mathbb{E}\left[\frac{1}{N}\sum_i z_i y_i^{\perp}\right]$ and $\mathbb{E}\left[\frac{1}{N}\sum_i z_i x_i^{\perp}\right]$, where $v_i^{\perp}$ denotes the residuals from a cross-sectional projection of v_i on μ_i. We show in BH that these moments also identify β under Assumption 1 (Appendix B.1).

[17] Appendix C.9 of BH shows that controlling for μ_i always reduces asymptotic variance of the estimator when $z_i \mid w$ is homoskedastic, while also giving a counterexample under heteroskedasticity.

[18] In panel data with $z_{it} = f_{it}(g_t, w_t)$, for example, unit fixed effects generally purge OVB only when the expected instrument is time-invariant, which generally requires the $f_{it}(\cdot)$ 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) can be quite restrictive. For instance, when new railroad lines tend to be built more than destroyed expected market access will tend to grow over time.

King and Nielsen 2019). A notable exception is the E-estimator of Robins et al. (1992), which similarly leverages linearity of an outcome model like (1) to recenter by a scalar variable. Third, and most importantly, propensity scores are usually estimated from the data by relating the treatment to a vector of observation-specific confounders. This approach is generally not feasible because exposure to exogenous shocks is high-dimensional: for instance, as noted in Section 2, the expected market access of any region i depends on the entire economic geography of the country. We therefore take a different approach to computing μ_i, which we turn to next.

3.3 Computing the Expected Instrument via Shock Counterfactuals

We propose computing the expected instrument by specifying an assignment process for the shocks, drawing many sets of counterfactual shocks from this process, recomputing the instrument each time, and averaging it across the counterfactuals. Here we formalize this approach, discuss general ways in which counterfactual shocks can be specified, and highlight the advantages of our approach over alternatives.

We define the shock assignment process as the conditional distribution of $g \mid w$, denoted $G\left(g \mid w\right)$. When $G\left(\cdot\right)$ is known, the expected instrument $\mu_i = \int f_i(\gamma; w) dG(\gamma \mid w)$ can be computed and either used to recenter z_i or added as a regression control. To emphasize the importance of a known shock assignment process, we write it as an assumption:

Assumption 2. *(Known assignment process): $G\left(g \mid w\right)$ is known in the support of w.*

This assumption is unrestrictive when the shocks are determined by a known randomization protocol, as in an RCT or with policy randomizations (such as tie-breaking lottery numbers in centralized assignment mechanisms; Abdulkadiroglu et al. (2017)). The assignment process may also be given by scientific knowledge when the shocks are randomized naturally, such as when g captures weather or seismic shocks governed by meteorological or geological processes (e.g., Carvalho et al. (2021) and Madestam et al. (2013)). Policy discontinuities (as in regression discontinuity designs) can also yield a known $G\left(\cdot\right)$ when viewed as generating local randomization around known cutoffs (Lee 2008; Cattaneo et al. 2015).[19]

In observational data, where the distribution of shocks is unknown, Assumption 2 can be satisfied by specifying some permutations of shocks that were as likely to have occurred. For instance, if one is willing to assume the shocks g_k are *iid* across k, it follows that all permutations of g are equally likely to arise. In this case $G(g \mid w)$ is known to be uniform when w is augmented by the permutation class $\Pi(g) = \{\pi(g) \mid \pi\left(\cdot\right) \in \Pi_K\}$, where Π_K

[19] Assumption 2 requires specification of $G(\cdot \mid w)$ in the entire domain of w. However, our identification results hold when w is viewed as non-stochastic 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.

denotes the set of permutation operators $\pi(\cdot)$ on vectors of length K (e.g. Lehmann and Romano 2006, p. 634). The distribution of each g_k (conditionally on other components of w) then needs not be specified; the expected instrument is the average z_i across all permutations of shocks, which serve as counterfactuals:

$$\mu_i = \frac{1}{K!} \sum_{\pi(\cdot)\in\Pi_K} f_i(\pi(g); w). \tag{8}$$

Such μ_i are easy to compute (or approximate with a random set of permutations).[20]

Similar expected instrument calculations follow under weaker shock exchangeability conditions, such as when the g_k are *iid* within, but not across, a set of known clusters and the class of within-cluster permutations is used to draw counterfactuals. We illustrate this approach in Section 4. In BH we discuss how our framework can also apply with $G(g \mid w)$ specified up to a low-dimensional vector of consistently estimable parameters (Appendix C.5); we also show how Assumption 2 can derive from an economic model (e.g. of transportation network formation) with stochastic shocks or from symmetries of the joint shock distribution (Appendices D.1 and D.2)

We note that even when $G(\cdot)$ is challenging to specify, a possibly incorrect specification can be useful as a sensitivity check. Specifically, if Assumption 1 holds and there is already no OVB because the included regression controls perfectly capture either the endogenous features of exposure or the expected instrument, then controlling for any candidate expected instrument $m_i(w)$ cannot introduce bias. In this case the researcher may safely control for one or several $m_i(w)$ based on some guesses of the assignment process.[21] More generally, researchers may achieve additional robustness by controlling for multiple candidate $m_i(w)$ based on multiple shock assignment process guesses; only one such guess needs to be right to purge OVB.

3.4 Randomization Inference and Testing

In some applications of our framework, natural assumptions on the mutual independence of $\tilde{z}_i$ or ε_i across observations can make conventional (e.g. clustered) asymptotic inference valid. Generally, however, the common exposure of observations to observed and unobserved shocks generates complex dependencies across observations making conventional asymptotic

[20]Approximating μ_i is sufficient for identification because the recentered IV still identifies β in this case: i.e. $\mathbb{E}\left[\frac{1}{N}\sum_i \left(f_i(g,w) - f_i(\pi(g),w)\right)\varepsilon_i\right] = 0$ under Assumption 1, for any fixed or randomly drawn $\pi(\cdot)$.

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

analysis inapplicable.[22] In such cases, in may be attractive to construct confidence intervals for the constant effect β and tests for Assumptions 1 and 2 based on the specification of the shock assignment process, following a long tradition of randomization inference (RI; Fisher 1935). The RI approach guarantees correct coverage in finite samples, of both observations and shocks.[23] We focus on a particular type of RI test which is tightly linked to the recentered IV estimator $\hat{\beta}$.

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 the g shocks 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. 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 (see Appendix C.3 of BH for details).

We propose addressing the practical issue of choosing a randomization test statistic by picking a T that is tightly linked to the recentered IV estimator, building on the theory of Hodges and Lehmann (1963) and Rosenbaum (2002). Specifically, we consider the sample covariance of the recentered instrument and implied residual: $T = \frac{1}{N}\sum_i (f_i(g,w) - \mu_i) \cdot (y_i - bx_i)$. Lemma 2 of BH shows that $\hat{\beta}$ is a Hodges-Lehmann estimator corresponding to this T, meaning that $\hat{\beta}$ equates T with its expectation across counterfactual shocks (specifically, zero).[24] This connection makes RI tests and confidence intervals based on T inherit the consistency of $\hat{\beta}$: the test power asymptotically increases to one for any fixed alternative $b \neq \beta$, under additional regularity conditions (Proposition S2 of BH).[25]

Randomization inference can also be used to perform falsification tests on our key Assumptions 1 and 2. Recentering implies a testable prediction that $\tilde{z}_i$ is orthogonal to any variable $r = (r_i)_{i=1}^N$ satisfying $g \perp\!\!\!\perp r \mid w$, which holds for any function of w or 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 $\frac{1}{N}\sum_i \tilde{z}_i r_i$ is sufficiently close to zero

[22]Adão et al. (2019) derive non-standard asymptotic inference in one such setting: when z_i is a linear shift-share variable, i.e. with $f_i(g,w) = \sum_k w_{ik} g_k$.

[23]Specifically, RI guarantees the validity of tests for the model parameter β, which can be interpreted as a constant treatment effect. Valid inference with heterogeneous effects in the kind of interdependent data we study is a difficult challenge, even with an asymptotic approach (Adão et al. 2019).

[24]With additional predetermined controls included in the regression (e.g. μ_i), the same property is satisfied by the residualized statistic $\frac{1}{N}\sum_i \tilde{z}_i \left(y_i^\perp - bx_i^\perp\right)$, where here $v_i^\perp$ denotes the residuals from a cross-sectional projection of v_i on the included controls.

[25]RI confidence intervals based on this statistic are still obtained by test inversion, and not from the distribution of the recentered estimator itself across counterfactual shocks g^*. The latter idea fails in IV since the re-randomized instrument $f_i(g^*, w) - \mu_i$ has a true first-stage of zero. The distribution of reduced-form coefficients across counterfactual shocks is also not useful, except for testing $\beta = 0$, as that distribution is centered around zero rather than β.

by drawing counterfactual 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_i, can be combined by an appropriate RI procedure, e.g. by taking T to be the sample sum of squared fitted values from regressing $\tilde{z}_i$ on R_i.[26]

Falsification tests are useful in two ways. First, when r_i is a lagged outcome or another variable thought to proxy for ε_i, 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 2, even when r_i does not proxy for ε_i. For $r_i = 1$, for example (which is trivially conditionally independent of g), the test verifies that the sample mean of z_i is typical for the realizations of the specified assignment process. Setting $r_i = \mu_i$ instead checks that the recentered instrument is not correlated with the expected instrument that it is supposed to remove.

3.5 Extensions

While we analyze the constant-effect model (1), identification by μ_i-adjusted regressions extends to settings with heterogeneous treatment effects. Namely, Appendix C.1 of BH shows that the recentered IV estimator generally identifies a convexly weighted average of heterogeneous effects under an appropriate monotonicity condition, extending the classic result of Imbens and Angrist (1994). The weights are proportional to the conditional variance of $\tilde{z}_i \mid w$ across counterfactual shocks, σ_i^2. These σ_i^2, like μ_i, are given by the shock assignment process (Assumption 2) and therefore can be computed by the researcher. Moreover, they can be used to identify more conventional weighted average effects. For example, in reduced-form models of the form $y_i = \beta_i z_i + \varepsilon_i$ a recentered and rescaled IV $(z_i - \mu_i)/\sigma_i^2$ identifies the average effect $\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \beta_i\right]$. Similarly, in IV settings with binary x_i and z_i, the same rescaled instrument identifies the local average treatment effect of Imbens and Angrist (1994).[27]

Further extensions to our framework are given in the appendix of BH. 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.

[26] Formally, this $T = \tilde{z}' R\left(R'R\right)^{-1} R'\tilde{z}$ can be seen as a quadratic form of the vector-valued statistic $\frac{1}{N}\sum_i \tilde{z}_i R_i$, weighted by $\left(R'R\right)^{-1}$, where R is the matrix collecting R_i and $\tilde{z}$ is the vector collecting $\tilde{z}_i$.

[27] We note that this heterogeneous effects extension applies to identification but not randomization-based confidence intervals which, as noted above, require a sharp null hypothesis $\beta_i = b$ for all i.

4 Application: Effects of Transportation Infrastructure

We now present an empirical application showing how our theoretic framework can be used to avoid OVB in practice. Specifically, we 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.

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.[28] 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 multiple 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, referred to as prefectures.[29] We measure market access growth between 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 were completed or under construction as of April 2019.[30] We compute a simple market access measure in each prefecture i and year t based on the formula in Zheng and Kahn (2013): $MA_{it} = \sum_j \exp(-0.02\tau_{ijt}) \cdot Pop_{j,2000}$, where $Pop_{j,2000}$ denotes the predetermined population of prefecture j and τ_{ijt} denotes predicted travel time between regions i and j 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, $x_i = \log MA_{i,2016} - \log MA_{i,2007}$, to the corresponding growth in prefecture's urban employment y_i from Chinese City Statistical Yearbooks. This yields a set of 275 prefectures with non-missing outcome data; see Appendix A for details on the sample construction and MA measure. Panel A of Figure 1 shows the Chinese HSR network

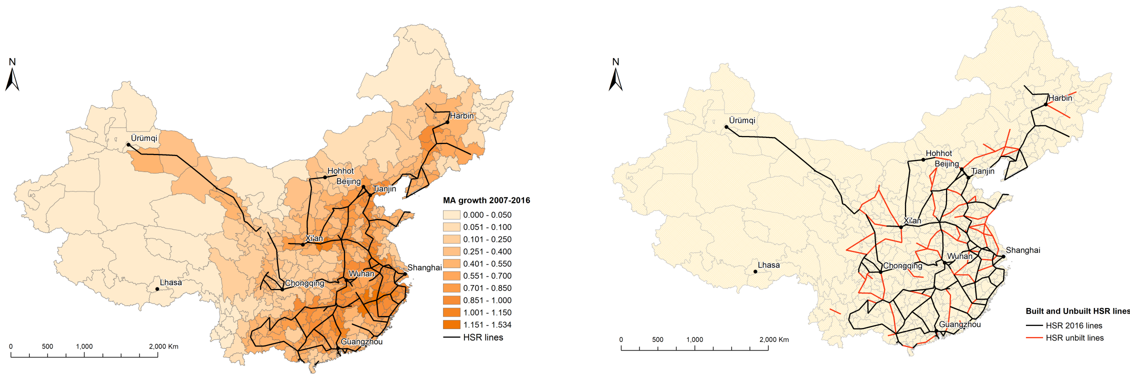
[28] Construction was started by the Medium- and Long-Term Railway Plan in 2004; this plan was later expanded in 2008 and again in 2016.

[29] Most prefectures are officially called "prefecture-level cities," but typically include multiple urban areas.

[30] We define a line by a contiguous set of inter-prefecture HSR links that were proposed together and opened simultaneously. One pilot HSR line between Qinhuangdao and Shenyang opened in 2003. We include it in our market access measure but focus on the bulk of HSR growth over 2007–2016.

Figure 1: Chinese High Speed Rail and Market Access Growth, 2007-2016

A. Completed Lines and MA Growth B. All Planned Lines



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

as of the end of 2016, along with the implied MA growth of relative to 2007.

Column 1 of Table 1, Panel A, reports the coefficient from a regression of employment growth on MA growth.[31] The estimated elasticity of 0.23 is large. With an average MA growth of 0.54 log points, it implies a 12.4% employment growth attributable to HSR for an average prefecture—almost half of the 26.6% average employment growth. The estimate is also highly statistically significant using Conley (1999) spatially-clustered standard errors.

Panel A of Figure 1, however, gives reason for caution against causally interpreting the OLS coefficient. 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 and large markets are concentrated. 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, latitude, and longitude. These predictors capture over 80% of the variation in MA growth (as measured by the regression's R^2), reinforcing the OVB concern: for a causal interpretation of the Table 1 regression, one would need to assume that all unobserved determinants of employment growth (e.g. local productivity shocks) are uncorrelated with these geographic features. While one could of course control for the specific geographic variables from Table 2 (as we explore below), controlling perfectly for geography is impossible without removing all variation in x_i.

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

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	0.231	0.081	0.067
	(0.075)	(0.098)	(0.094)
		[-0.280, 0.330]	[-0.187, 0.339]
Expected Market Access Growth			0.319
			(0.095)
Panel B. With Geography Controls			
Market Access Growth	0.133	0.053	0.043
	(0.064)	(0.089)	(0.092)
		[-0.132, 0.281]	[-0.154, 0.284]
Expected Market Access Growth			0.216
			(0.072)
Recentered	No	Yes	Yes
Prefectures	275	275	275

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 unbuilt 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.

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 we can compute the appropriate function of geography μ_i which removes the systematic variation in MA growth.

Our specification of counterfactuals exploits the heterogeneous timing of HSR construction. Specifically, we permute the 2016 completion status of the built and unbuilt (but planned) lines, assuming that the timing of line completion is conditionally as-good-as-random. Panel B of Figure 1 compares the built and unbuilt lines which form our counterfactuals. Unbuilt lines tend to be concentrated in the same areas of China as built lines, reinforcing the fact that construction is not uniformly distributed in space. Moreover, built lines tend to connect more regions: the average number of cross-prefecture "links" is 3.19 and 2.48 for built and unbuilt lines, respectively, with a statistically significant difference ($p = 0.048$). To account for this difference we construct counterfactual upgrades by permuting the 2016 completion status only among lines with the same number of links. For

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

	Unadjusted	Recentered		
	(1)	(2)	(3)	(4)
Distance to Beijing	−0.291	0.066		0.087
	(0.062)	(0.039)		(0.045)
Latitude/100	−3.324	−0.324		−0.147
	(0.646)	(0.274)		(0.315)
Longitude/100	1.321	0.455		0.404
	(0.458)	(0.234)		(0.236)
Expected Market Access Growth			0.030	0.059
			(0.056)	(0.067)
Constant	0.536	0.017	0.017	0.017
	(0.029)	(0.018)	(0.020)	(0.018)
Joint RI p-value		0.510	0.715	0.558
R^2	0.824	0.077	0.010	0.080
Prefectures	275	275	275	275

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 unbuilt 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. Distance to 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 26.

example the main Beijing to Shanghai HSR line, which has the greatest number of links, is always included in the counterfactuals. This procedure generates 999 counterfactual HSR maps that are visually similar to the actual 2016 network; Appendix Figure A1 gives an illustrative example.

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 this recentering successfully removes the systematic geographic variation in market access. Specifically, it regresses recentered MA growth on a constant and the same geographic controls as in Column 1. 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 26) is 0.51. Columns 3 and 4 further show that recentered MA growth is uncorrelated with expected MA growth.[32]

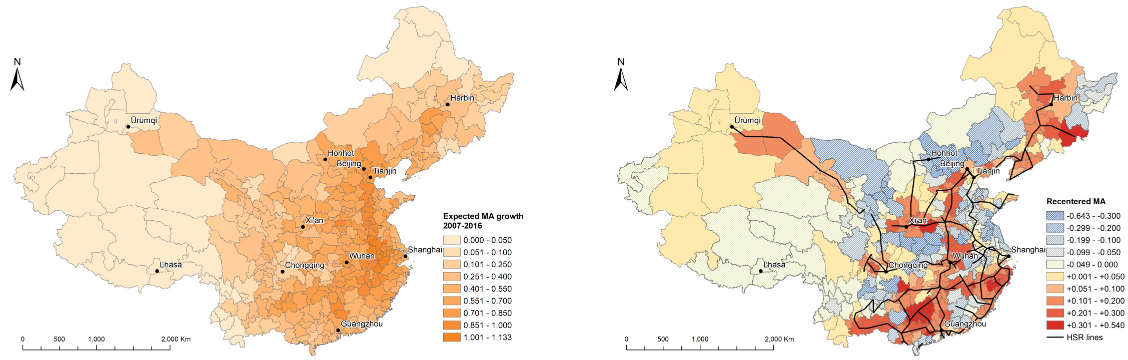
Figure 2 plots expected and recentered MA growth given by the permutations of built and unbuilt lines. The effect of recentering is apparent by contrasting the solid and striped

[32] These results are consistent with correct specification of counterfactuals (i.e. we cannot reject Assumption 2), though we note they do not provide direct support for the exogeneity of HSR construction to the unobserved determinants of employment (Assumption 1).

Figure 2: 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 HSR counterfactuals that permute the opening status of built and unbuilt 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 1 and expected MA growth. The HSR network as of 2016 is also shown in this panel.

regions in Panel B of Figure 2 (indicating high and low recentered MA growth) with the dark- and light-shaded regions in Panel A of Figure 1 (indicating high and low MA growth). The recentered treatment no longer places western prefectures in the effective control group, as their MA growth is as low as expected. 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 recentered MA growth reduces the estimated employment elasticity substantially, from 0.23 to 0.08. Controlling for expected MA growth yields a similar 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 (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_i, of 0.32, meaning that employment grew faster in prefectures that were more highly exposed to potential HSR construction, whether or not the nearby lines were built yet.

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 coefficient of 0.13. In contrast, Columns 2 and 3 show that the finding of no significant MA effect after adjusting for μ_i is robust to including geographic controls. The μ_i adjustment alone appears sufficient to remove the geographic dependence of MA, as Table 2 also showed.[33]

While our primary interest is to illustrate the recentering approach, we note that there are several possible explanations for the substantive finding of a small employment effect of MA. Unlike other transportation networks used for trading goods, the Chinese HSR network primarily 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 BH we discuss how market access recentering relates to other approaches in the long literature estimating transportation infrastructure upgrade effects (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 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 emphasize that even when it is challenging to obtain a convincing specification of counterfactuals, *any* specification can yield a robustness check on these alternative strategies (see footnote 21).

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 approach to causal inference when some, but not all, of this variation is exogenous. Non-random exposure to exogenous shocks can bias conventional regression estimators, but this problem 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. Averaging the treatment or

[33] In BH we provide additional robustness checks, adjusting the definitions of MA and outcome variables, using a binary measure of connectivity to the HSR network, including province fixed effects, dropping influential prefectures, and examining the role of treatment effect heterogeneity.

instrument over this assignment process yields a single confounder μ_i which can be adjusted for to achieve identification. The specification of counterfactuals also yields a natural form of valid finite-sample inference.

In practice, researchers face a choice in how to use μ_i in a regression analysis: recentering by it or controlling for it. When the assignment process is given by a true randomization protocol, as in a RCT, we recommend researchers recenter first to purge OVB. Then *any* predetermined controls (i.e. functions of exposure) can be included to remove variation in the error term and likely increase estimation efficiency. While μ_i is one possible control, which automatically recenters the treatment or instrument, it need not be the best choice in terms of predicting the residual variation. Our recommendation is different in natural experiments where assumptions must be placed on the assignment process. Then controlling for candidate μ_i instead of recentering can have a valuable "double-robustness" property. Researchers can compute and control for several candidate μ_i based on different assignment processes, such that OVB is purged if at least one of the processes is specified correctly (or if there is no OVB to begin with).

We conclude by noting that our framework bears practical lessons for a range of common treatments and instruments, well beyond the market access measure in our empirical application. In our working paper (Borusyak and Hull 2021), we discuss and illustrate some of these implications for policy eligibility treatments, 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. We expect other settings may also benefit from explicit specification of shock counterfactuals and appropriate adjustment for non-random shock exposure.

A Data Appendix

Our analysis of market access effects uses data on 340 prefectures of mainland China. This excludes the islands of Hainan and Taiwan and the special administrative regions of Hong Kong and Macau, but includes six sub-prefecture-level cities (e.g. Shihezi) that do not belong to any prefecture. We use United Nations shapefiles to geocode each prefecture by the location of its main city (or, in a few cases, by the prefecture centroid).[34]

We use a variety of sources to assemble a comprehensive database of the HSR network in 2016 as well as the lines planned (and in many cases under construction) as of April 2019 but not opened yet by the end of 2016. Our starting points are Map 1.2 of Lawrence et al. (2019), China Railway Yearbooks, and the replication files of Lin (2017). We cross-check

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

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

We compute travel time τ_{ijt} between all pairs of prefectures i and j as of the ends of 2007 and 2016 for both the actual and counterfactual networks. Travel time combines traditional modes of transportation (car or low-speed train) with HSR, where available. We allow for unlimited changes between different HSR lines and between HSR and traditional modes without a layover penalty, as HSR trains tend to operate frequently and traditional modes also involve downtime. Following the existing literature, we proxy for travel time by traditional modes by the straight-line distance, and specify the speed of 100 = 120/1.2kph, where 120kph is their typical speed and the 1.2 adjustment for actual routes that are longer than a straight line. For two prefectures connected by an HSR line, we compute the distance along the line as the sum of straight-line distances between adjacent prefectures on the line. We use the operating speed of each line divided by an adjustment factor of 1.3 to capture the fact that the average speed is lower than the nominal speed we record. Computing MA further requires the population of each of the 340 prefectures from the 2000 population Census, which we obtain from the CityPopulation.de website.[35]

We measure prefecture employment in the 2008–2017 China City Yearbooks.[36] Each yearbook covers the previous year (so our data cover 2007–2016). While the yearbooks provide several employment variables, we use "The Average Number of Staff and Workers" (from the "People's Living Conditions and Social Security" chapter), as measured in the entire prefecture and not just the main urban core. This employment series has by far the lowest number of strong year-to-year deviations which may indicate data quality issues.

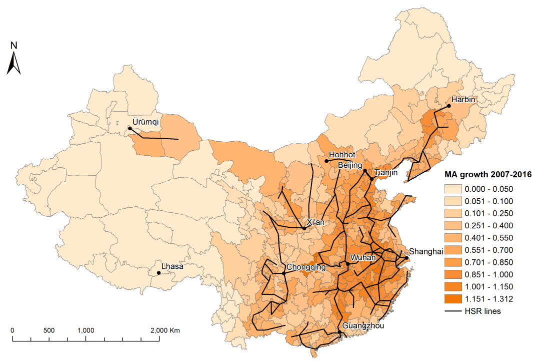
We finally apply a data cleaning procedure to the outcome variable. We first mark a

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

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

prefecture-year observation as exhibiting a "structural break" if (*i*) the outcome changes by more than twice in either direction relative to the previous non-missing value for the prefecture, (*ii*) it is not followed by a change in the opposite direction that is between 3/4 and 4/3 as large in terms of log-changes (which we view as a one-off jump and ignore), and (*iii*) the previous change does not satisfy (i). We view the outcome change between 2007 and 2016 as valid only if there are no structural breaks in any year in between. This reduces the sample from 283 to the final set of 275 prefectures.

Figure A1: Simulated HSR Lines and Market Access Growth



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

References

Abadie, Alberto. 2003. "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics* 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, and Guido W. Imbens. 2016. "Matching on the Estimated Propensity Score." *Econometrica* 84:781–807.

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.

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.

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. 2021. "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.*

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2013. "Inference on treatment effects after selection among high-dimensional controls." *Review of Economic Studies* 81:608–650.

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2021. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies.*

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." *Quarterly Journal of Economics* 136:1255–1321.

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.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110:2964–2996.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/debiased machine learning for treatment and structural parameters." *Econometrics Journal* 21:C1–C68.

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

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

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

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.

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

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

Hirano, Keisuke, and Guido W. Imbens. 2004. "The Propensity Score with Continuous Treatments." Chap. 7 in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives,* 73–84.

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

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

King, Gary, and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis* 27:435–454.

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.

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

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

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.

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

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

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.

Robins, James M, Steven D Mark, and Whitney K Newey. 1992. "Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders." *Biometrics* 48:479–495.

Robinson, P.M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 56:931–954.

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. 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.*

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.