Instruments Combining Quasi-Random Shocks with Non-Random Exposure: Theory and Applications

Kirill Borusyak
UCL

Peter Hull
U Chicago and NBER*

March 2020

Abstract

We study the properties of "shock-exposure instruments": functions of quasi-experimental shocks and pre-determined but endogenous measures of heterogeneous shock exposure. Examples include shift-share instruments, instruments capturing spillovers in social and transportation networks, simulated policy eligibility instruments, and model-implied instruments. We show that the validity of these instruments generally requires a simple but non-standard recentering, derived from knowledge of counterfactual shocks that might well have been realized. This recentering benchmarks each observation of the instrument against its expectation across counterfactual shocks, holding fixed each observation's shock exposure. Knowledge of the shock assignment process can also be used for exact randomization inference and specification tests that are valid in finite samples. We further establish conditions for large-sample consistency and characterize the shock-exposure instruments that are asymptotically efficient. We illustrate the practical implications of our framework in several applications.

PRELIMINARY AND INCOMPLETE:
PLEASE DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION

*Contact: k.borusyak@ucl.ac.uk and hull@uchicago.edu. We are grateful to Rodrigo Adão, Raffaella Giacomini, Vishal Kamat, Michal Kolesár, Gabriel Kreindler, Eduardo Morales, and seminar participants at Princeton, U Chicago, UCL, and LSE for helpful comments. We also thank Yatang Lin for providing replication code and data. Ruixue Li and Elise Parrish provided excellent research assistance.

1 Introduction

Researchers use instrumental variables to leverage quasi-experimental variation in an endogenous treatment and estimate its causal effects. In traditional settings, each observation receives its own as-good-as-randomly-assigned shock which can be used directly as an instrument. Often, however, quasi-experimental shocks affect the treatment status of many observations jointly but to different extents. For example, improvements to a transportation network may differentially increase the market access of many cities, changes in international trade policy may differentially affect the economic conditions of many regions, and reforms to an entitlement program may differentially expand the eligibility of many individuals. In these settings, researchers may construct an instrument—what we call a *shock-exposure instrument* (SEIV)—which combines the quasi-experimental shocks with pre-determined measures of heterogeneous shock exposure. Typically the assumption of quasi-random shock variation is viewed as sufficient for such instruments to be valid levers for causal inference, perhaps conditional on a set of observed regression controls.

This paper establishes problems with and solutions for identification and inference with SEIVs, using a novel econometric framework that encompasses common research designs across many fields. We first show that SEIV estimates may suffer from omitted variables bias (OVB), even when the shocks underlying the instrument are as-good-as-randomly assigned. The OVB problem arises when different observations receive systematically different values of the instrument because of their predetermined exposure to quasi-experimental shocks. Even when the placement of new railroad construction or the timing of line upgrades is random, for example, cities in the center of the transportation network may get systematically larger increases in market access because they tend to be closer to any given line. The validity of shock-exposure instruments then fails if more central cities are also systematically different in terms of the unobservable shocks affecting the outcome. This source of OVB need not be captured by standard regression controls

We propose a general solution to OVB with shock-exposure instruments that involves specifying and simulating the shock *assignment process*; that is, redrawing counterfactual shocks that were as likely to have occurred as the realized ones. One might, for example, assume that different railroad lines had an equal chance of being selected for upgrade from an initial transportation development plan and use this assumption to simulate counterfactual upgrade schemes. Similarly, one could assume that observed tariff changes across different industries and policy variation across states could as well have been exchanged. By measuring, separately for each observation, the average value of the SEIV across possible shocks—what we call the *expected instrument*—the researcher can recenter the shock-exposure instrument to purge OVB. The recentered instrument compares units with higher-than-expected values of the instrument, because of the realized shocks, to those with lower-than-expected values. By construction, this instrument is not systematically higher for any predetermined set of observations with potentially different unobservables and is thus valid. Including the expected

instrument as a regression control also solves the OVB problem, by extracting the same exogenous variation leveraged by the recentered instrument.

Taking a stand on the shock assignment process is generally necessary for SEIVs to rely just on quasi-experimental variation in shocks and not also on an assumption of exogenous shock exposure. Absent such a stance, causally interpreting SEIV estimates requires an assumption that unobserved outcome shocks are uncorrelated with the expected instrument—an assumption that is difficult to assess precisely because the expected instrument is unspecified. Causal identification with the recentered instrument instead follows solely from the quasi-randomness of shocks, though it continues to derive power from heterogeneous shock exposure.

We next show how the challenge of statistical inference with interdependent (e.g. spatial or networked) SEIVs can also be overcome by specifying a shock assignment process. This problem arises when both observed and unobserved shocks affect many observations jointly, inducing complex data dependencies where observations are "fuzzily clustered" by their common shock exposure. In such cases, conventional approaches to asymptotic inference that rely on sufficient independence between most pairs of observations (for example, those from different clusters or located at geographic or network distance above some threshold (Conley, 1999)) can fail. Our solution is to adapt principles of Fisher's randomization inference (RI) to the SEIV setting with specified counterfactual shocks. RI-based tests and confidence intervals for second-stage effects and balance coefficients are exact, for any correlation structure in the residuals, and robust to weak instruments (Imbens and Rosenbaum, 2005). We address the practical issue of choosing a powerful RI test statistic for SEIV by using the observation that each statistic induces a Hodges-Lehmann estimator which rationalizes the observed statistic as typical under its null (Hodges and Lehmann, 1963; Rosenbaum, 2002). We pick the statistic that induces the recentered SEIV estimator, tightly linking our approaches to estimation and inference.

We complement this finite-sample inferential approach with an asymptotic analysis, establishing sufficient conditions for consistency of the recentered estimator and associated RI tests as well as characterizing the large-sample efficient constructions of recentered SEIV instruments. We show that recentered SEIVs yield consistent estimates and powerful RI tests when the shocks induce sufficiently-uncorrelated variation in the instrument, regardless of the correlation structure of the unobservables. Our derivation of asymptotically efficient SEIVs extends the canonical Chamberlain (1987) analysis of optimal estimation to *non-iid* data and a non-standard conditional moment restriction.[1]

We extend our basic analysis of SEIVs in several ways. First, we show how SEIVs can be used when the shock assignment process is only partially specified, allowing for a vector of unknown parameters which govern, for example, how shocks vary systematically with observables. Valid (though conservative) finite-sample confidence intervals can be obtained in this case by a two-step RI proce-

[1] Adão et al. (2019a) use the *iid* approach of Chamberlain (1987) to characterize efficient instruments in a model of spatial linkages and international trade. As discussed in Section 5.6, our characterization of efficient shock-exposure instruments allows for the complex dependence structure induced by shocks, as well as endogeneity of shock exposure.

dure (Berger and Boos, 1994). Second, we discuss how predetermined observables can be included as SEIV regression controls to reduce variation in the residuals in a straightforward way. Third, we discuss SEIV regressions with multiple instruments or treatments. Finally, we show that the recentered SEIV estimator captures a convex average of heterogeneous treatment effects under an appropriate monotonicity assumption, extending the classic Imbens and Angrist (1994) result to this setting.

Our general framework provides concrete insights for a large number of empirical designs. First, we discuss the OVB and inference problems associated with instruments based on network spillovers and upgrades, including transportation and other spatial networks (Carvalho et al., 2016; Jaravel et al., 2018; Donaldson and Hornbeck, 2016; Donaldson, 2018; Allen et al., 2019). In models of economic geography, for example, instruments for market access may combine quasi-experimental variation in transportation upgrades with predetermined but non-random variation in network centrality. We illustrate the role of SEIV recentering in such settings by an application that estimates the local employment effects of increased regional connectivity from a recent expansion of high-speed rail (HSR) in China (Lin, 2017). While panel regressions of urban employment growth on changes in regional market access suggest a large and statistically significant causal effect, we find this to be likely driven by systematic differences in a city's exposure to potential HSR quasi-experiments. Outcome pre-trends are strongly correlated with expected market access growth, constructed under a simple assumption of exchangeable HSR line opening dates as well as a more realistic assumption of idiosyncratic construction timing. Correspondingly, treatment effect estimates that correct for expected market access yield much smaller IV estimates and permutation-based confidence intervals that do not exclude the null hypothesis of no effect.

Second, we show how our SEIV framework can be used to boost precision in settings with a simulated policy eligibility instrument (Currie and Gruber, 1996a,b; Cohodes et al., 2016; Cullen and Gruber, 2000; Gruber and Saez, 2002; East and Kuka, 2015). Traditional simulated eligibility instruments isolate quasi-experimental policy variation by removing much of the individual heterogeneity in policy exposure, constructing an instrument that only varies across U.S. states or broad groups. We show how endogenous exposure can be efficiently leveraged in SEIVs that are nevertheless valid. We illustrate the precision gains from such additional variation in an application that estimates Medicaid take-up and crowd-out effects from recent Affordable Care Act expansions. SEIV estimates have at least 30-40% smaller standard errors than more conventional simulated instrument estimates, and Monte Carlo simulations show dramatic power improvements.

Third, we show how our framework yields new methods for inference with linear shift-share (or "Bartik") instruments (Bartik, 1991; Blanchard and Katz, 1992; Card, 2001; Autor et al., 2013), in which the SEIV is an exposure-weighted average of shocks.[2] Randomization inference based on *iid* or exchangeable shocks can serve as a useful complement to existing asymptotic approaches, which rely

[2]For recent methodological literature on shift-share instruments see Adão et al. (2019b), Borusyak et al. (2019), Goldsmith-Pinkham et al. (2019), and Jaeger et al. (2017).

on observing a large number of shocks (Adão et al., 2019b; Borusyak et al., 2019). In Monte-Carlo simulations based on the "China shock" setting of Autor et al. (2013), we find that when the number of shocks is small or their distribution has heavy tails (which are both common scenarios in practice), RI dominates the asymptotic approaches both in terms of size (which is correct for RI by construction but distorted in the existing approaches) and power. RI also dominates alternative procedures in Adão et al. (2019b) and Borusyak et al. (2019) which improve finite-sample performance by imposing the null hypothesis: while the size distortion of the resulting asymptotic confidence intervals is much smaller, their power is substantially worse than that of RI confidence intervals.

Finally, we discuss practical implications for nonlinear shift-share instruments (Berman et al., 2015; Boustan et al., 2013; Derenoncourt, 2019; Chodorow-Reich et al., 2019), model-implied instruments in international trade (Allen et al., 2018; Adão et al., 2019a), and instruments generated by partially-randomized school assignment mechanisms (Abdulkadiroglu et al., 2017, 2019). We show how in each setting appropriate instrument recentering and randomization inference relate to and depart from conventional methods, and how the shock assignment process may be given by appropriate models or institutional knowledge.

OVB from combining random shocks with non-random exposure has been previously noted in several special cases. For example, in the context of testing for spillover effects Aronow (2012) notes that "simple randomization of treatment to individuals does not imply simple randomization of proximity to treated units." In the linear shift-share IV setting, Borusyak et al. (2019) show that instrument validity does not follow from exogeneity of industry shocks when the shares used to construct the instrument do not add up to the same constant for each observation. Our paper develops a general framework to analyze such OVB and a general recentering solution. Formally, this solution can be viewed as a generalization of conventional propensity score adjustment in simpler quasi-experimental designs (Rosenbaum and Rubin, 1983).

Our use of randomization inference also draws on a rich statistical literature, dating back to Fisher (1935); see Lehmann and Romano (2006, Ch. 15) for a modern treatment. Our analysis of SEIVs particularly relates to the theoretical literature studying causal inference in settings with interference across units, as in Aronow and Samii (2017). In practice, RI has recently deployed in quasi-experimental settings by, among others, Bertrand et al. (2004), Imbens and Rosenbaum (2005), Ho and Imai (2006), Abadie et al. (2010), Dell and Olken (2018), and Ganong and Jäger (2018), Shaikh and Toulis (2019), and in non-regression network contexts by Nyblom et al. (2003).

The remainder of this paper is organized as follows. The next section motivates our analysis with an intuitive example of the OVB problem and our solution. Section 3 develops our general framework and formal results on SEIV identification and inference, while Section 4 establishes large-sample results on consistency and asymptotic efficiency. Extensions to the basic framework are given in the appendix. We discuss and illustrate the practical implications of our framework in Section 5.

Section 6 concludes.

2 A Motivating Example

We begin with a stylized presentation of the key ideas from our general conceptual framework in a setting inspired by Donaldson and Hornbeck (2016). Consider estimation of the effect of local growth in market access, $\Delta \log MA_\ell$, on the log growth of a regional outcome $\Delta \log V_\ell$, such as the value of agricultural land. We assume a linear structural model, of

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

In this equation, which may be derived from standard models of economic geography with symmetric trade costs (e.g. Redding and Venables, 2004), ε_ℓ is a productivity shock in region ℓ occurring between two periods and MA_ℓ is a statistic that depends on the sizes of all markets and the relative cost of transportation to all markets from ℓ (the exact market access formula is unimportant at this point). The coefficient β has a structural interpretation and can be used to evaluate policies which affect market access but not regional productivity.

We suppose there is a change in relative transportation costs from the construction of new railroad lines connecting certain regions ℓ. We abstract from strategic placement of new lines and assume that the choice of line location is exogenous, in the sense of being independent of all productivity shocks ε_ℓ. This assumption suggests one might leverage the railroad "natural experiment" to estimate β. A researcher may, however, recognize that quasi-random variation in railroad construction is not sufficient for consistent OLS estimation of equation (1): the dependence of market access on the local market size, which is plausibly affected by productivity shocks ε_ℓ, potentially induces a correlation between ε_ℓ and $\Delta \log MA_\ell$. As a result, she may instrument $\Delta \log MA_\ell$ with the predicted growth of market access that would arise just from the quasi-experimental changes to the transportation network, holding market size fixed at a predetermined level. Let z_ℓ denote this constructed instrument.

The construction of this shock-exposure instrument is intuitively appealing: z_ℓ is different from zero only because of the construction of new railroad lines, which are selected randomly with respect to the productivity shocks ε_ℓ. The first insight of this paper is, however, that the resulting cross-sectional variation in z_ℓ is *not* exogenous, because of non-random exposure to the railroad natural experiment. Even when new lines are placed randomly in space, some regions ℓ will tend to see systematically higher growth of predicted market access in a way that may be correlated with their own productivity shocks. This induces a bias in conventional instrumental variables (IV) estimates of β that cannot be eliminated by standard control variables.

To see this result most clearly, consider a square country consisting of 64 equally-sized regions with no initial connectivity, such that there are no preexisting differences in market access across ℓ.

Suppose new railroad lines are placed on the map exogenously, in a very strict sense: out of all pairs of adjacent regions, half are connected at random. One such draw from the railroad quasi-experiment is shown in Panel A of Figure 1, together with the resulting growth in market access.[3] Expectedly, regions that have been connected tend to have higher $\Delta \log MA_\ell$. However, the figure reveals another tendency: many of the regions with high market access are in the center of the country. It turns out that this concentration is not by chance. Panel B of Figure 1 shows that the average growth of market access in each region across 1000 random draws from the railroad assignment process, a statistic we label μ_ℓ, is also higher in the center of the map. This indicates that more central regions are more exposed to the railroad quasi-experiment: no matter where random lines are built, more central regions are more likely to be close to them.

Variation in the expected instrument μ_ℓ generates a potential source of omitted variables bias (OVB) for an IV estimate of β. The IV estimate comes from a comparison of outcomes between regions with high and low values of the instrument, which will tend to be regions with high and low μ_ℓ. Since unobserved productivity shocks are also likely to differ between the center of the map and the periphery, OVB is likely to arise from such a comparison. For example, if productivity growth is on average higher in more central regions (a scenario that is not precluded by exogeneity of line placement), the IV estimate of β will tend to be biased upward.

This simple example shows that the exogeneity of transportation network upgrades is not generally sufficient for the validity of a market access instrument, which combines as-good-as-random shocks (line placement) with shock exposure (region location and other relevant features of the country's geography). Shock exposure is predetermined but need not be exogenous: when productivity shocks depend on the same geographic features as expected market access growth, geography is endogenous in the sense of generating OVB. Although in the example of Figure 1 such OVB is a simple function of an observable (geographic centrality), for more complex settings and shock assignment processes the variation in μ_ℓ need not be captured by standard low-dimensional controls. Controlling for regional geography perfectly, of course, is not feasible as this would remove all variation in z_ℓ.[4]

The second insight of this paper is that this OVB problem has a very intuitive solution, based on the same knowledge of the railroad assignment process that generated Panel B of Figure 1. Given such knowledge, one can simulate values of z_ℓ across counterfactual draws from the shock quasi-experiment to compute μ_ℓ, separately for each region ℓ. One can then construct a *recentered* instrument $\tilde{z}_\ell = z_\ell - \mu_\ell$, values of which are shown in Panel C of Figure 1 along with the actual realization of railroad

[3] Specifically, market access in period $t = 1, 2$ is given by $MA_{\ell t} = \sum_k \tau_{\ell kt}^{-\theta} P_{kt}$ where $\tau_{\ell kt}$ is a function of distance and connectivity and P_{kt} denotes region k's population in period t. In this simple example $P_{kt} = 1$ is constant across regions and periods, $\theta = 1$, and $\tau_{\ell kt} = 2^{0.1 d_{\ell kt}}$ where $d_{\ell kt}$ is the minimum number of railroad links required to get from ℓ to k at time t (or infinity if there is no path). See Section 5.1 for a full discussion of this statistic.

[4] The OVB is also not solved by using panel data before and after the upgrade: in that case μ_ℓ varies over time and is not captured by region fixed effects. There are exceptional cases in which the problem cannot arise, as μ_ℓ does not vary across regions. For example, if the country fills the entire globe and the distribution of lines is perfectly symmetric, all regions are equally likely to see market access growth. The reader is invited to let us know if they encounter such an application.

lines. When used to estimate β, this recentered instrument compares regions with market access growth that is higher and lower than that which was expected given each region's exposure to the quasi-experiment. By construction, $\tilde{z}_\ell$ thus has no tendency to be higher or lower in any predetermined group of regions. We show in the next section how this recentering generally ensures the validity of shock-exposure instruments.

The third insight of this paper regards valid statistical inference on β with such $\tilde{z}_\ell$, which are likely to be correlated across different observations ℓ. A conventional approach to account for spatial dependence is Conley (1999) standard errors. This method specifies a geographic distance threshold, after which $\tilde{z}_\ell \varepsilon_\ell$ is assumed to be mutually uncorrelated; for the asymptotic approximation to hold this threshold should be sufficiently small. The existence of such a threshold can be a strong assumption in the SEIV setting, however. In the market access example, all regions are likely to be exposed to quasi-random transportation network upgrades. For example, long railroad lines will tend to cause regions which are far apart in space to "cluster," in terms of market access growth. Moreover, unobserved shocks will also tend to propagate through the transportation network. We show later in the paper how classical methods of randomization inference (RI) can be applied to address this issue of "exposure clustering." RI confidence intervals for β have valid coverage in finite samples (which is especially useful when the data represent the entire country, without any sampling), regardless of the unobserved distribution of productivity shocks. The procedure derives from the knowledge of the railroad assignment process, and is again based on simulating the railroad quasi-experiment.

The challenge of valid causal and statistical inference with shock-exposure instruments thus reduces to the challenge of specifying the shock assignment process. In practice, of course, railroad lines are not randomly drawn on a map. For correct recentering of z_ℓ and valid inference on β, without strong assumptions on the distribution of unobserved productivity shocks ε_ℓ, a researcher must take an explicit stand on the true shock assignment—specifically, on the set of counterfactual network maps that were as likely as the one that was realized. We return to the question of how such counterfactual railroad maps may be specified in Section 5.1 with a general discussion and empirical application.

3 SEIV Identification and Inference

3.1 Setting

We now consider a general setting in which an outcome y_ℓ and treatment x_ℓ are observed across L units, indexed by ℓ. Of interest is a causal parameter β which relates treatment to outcomes by

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

where ε_ℓ denotes a structural residual. Here we assume y_ℓ and x_ℓ are both scalar and demeaned, and that the effect of interest is linear; in Appendices A.1–A.3 we consider extensions to cases with multiple treatments, additional control variables, and heterogeneous causal effects. Importantly, we do not assume that the units are independently or identically distributed as if arising from random sampling. As motivated above, this will allow for complex unobserved dependencies across ℓ which are inherent to the SEIV approach. The lack of random sampling is also consistent with settings where the L units represent a population—for example, all regions of a country—where conventional asymptotic frameworks are inappropriate (Abadie et al., 2019).

We suppose a researcher has specified an instrument z_ℓ to estimate β. The intuitive argument for using this instrument is based on the quasi-experimental assignment of some $N \times 1$ vector shocks g which enter the construction of z_ℓ. This construction may also depend on other observed variables collected in the set w. We therefore write the shock-exposure instrument as

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

where $\{f_\ell(\cdot)\}_{\ell=1}^{L}$ is a fixed set of known functions mapping g and w to each observation ℓ. In the previous motivating section, g contained information on transportation network upgrades and w summarized the preexisting network structure and regional populations. The $f_\ell(\cdot)$ functions combined g and w to form predicted market access growth for each city 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 nonnegative exposure share weights. More generally, $f_\ell(\cdot)$ may be a complex nonlinear function predicting the treatment from g and w; allowing $f_\ell(\cdot)$ to flexibly vary across ℓ covers all instruments that are measurable given (g, w).[5] For now we take the researcher's choice of $f_\ell(\cdot)$ as given, addressing the question of which functions of (g, w) may be more desirable in Section 4.2. We also note that our framework allows $x_\ell = z_\ell$, in which case β reflects the causal effect of the instrument itself.

Partitioning the variables from which z_ℓ is constructed into shocks g and relevant exposure variables w allows us to formalize the notion of exogeneity in shock-exposure instruments. Researchers may be willing to assume that some, but not all, sources of variation in z_ℓ are as-good-as-randomly assigned with respect to the structural residual. For example, in shift-share designs it may be plausible that shocks, but not shares, are unrelated to the regression residual. We formalize as-good-as-random assignment by the conditional independence of g from the residual vector $\varepsilon = \{\varepsilon_\ell\}_{\ell=1}^{L}$, given the other sources of SEIV variation collected in w:

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

This condition is satisfied when shocks are fully randomly assigned, as in the stylized example in

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

Section 2: i.e. $g \perp\!\!\!\perp (\varepsilon, w)$. More generally, we will allow w to contain variables affecting the realization of shocks such that g is only conditionally exogenous.

Our notion of exogeneity in SEIVs generalizes that of conventional quasi-experimental designs, in which each unit ℓ is assigned its own exogenous shock with $g_\ell \perp\!\!\!\perp \varepsilon_\ell \mid w_\ell$ for some predetermined control vector w_ℓ. Under *iid* sampling of $(y_\ell, x_\ell, g_\ell, w_\ell)$, the shock assigned to unit ℓ is then guaranteed to have no relationship with other units, satisfying the usual Stable Unit Treatment Value Assumption (Imbens and Rubin, 2015). The more general Assumption 1 allows for the set of quasi-experimental shocks g, which need not vary at the same "level" as the outcome and treatment (i.e. $L \neq N$), to affect the treatment of many non-*iid* observations jointly. At the same time, we retain an implicit exclusion restriction central to IV estimation of β: exogenous variation in the shocks only affects the outcome of each unit via its own treatment status x_ℓ and not also through ε_ℓ.[6]

The conventional use for such a z_ℓ is as an instrument in a regression of y_ℓ on x_ℓ. This leads to the SEIV estimate

$$\hat{\beta} \equiv \frac{\frac{1}{L}\sum_\ell z_\ell y_\ell}{\frac{1}{L}\sum_\ell z_\ell x_\ell} = \beta + \frac{\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell z_\ell x_\ell}, \tag{4}$$

As usual, $\hat{\beta}$ is consistent when the instrument z_ℓ is correlated with the treatment x_ℓ in large samples while being asymptotically uncorrelated with the structural residual ε_ℓ, such that the second term of (4) converges to zero and $\hat{\beta} \xrightarrow{p} \beta$ as $L \to \infty$.

We discuss SEIV consistency in Section 4 and initially focus on IV-based identification of β. The key condition for the conventional SEIV regression to identify β is that the shock-exposure instrument z_ℓ is not systematically correlated with the residual ε_ℓ in the cross-section:

$$\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0. \tag{5}$$

Given this, β is identified by the reduced-form and first-stage moments $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell y_\ell\right]$ and $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell x_\ell\right]$ provided the latter is non-zero. Here it is worth highlighting that in our non-*iid* setting the appropriate notion of instrument exogeneity 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, where $\mathbb{E}[z_\ell \varepsilon_\ell] = \mathbb{E}[z_m \varepsilon_m]$ for $\ell \neq m$, equation (5) reduces to the more familiar $\mathbb{E}[z_\ell \varepsilon_\ell] = 0$.

[6] The IV 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, as we discuss in Appendix A.2. Some cases of exclusion violation may nevertheless yield interpretable estimates of β, as with heterogeneous treatment effects discussed in Appendix A.3 (see also Appendix A.1 in Borusyak et al. (2019)).

3.2 Expected and Recentered Shock-Exposure Instruments

Our first result shows that conventional SEIV regressions may suffer from omitted variables bias (OVB) even when the shocks are as-good-as-randomly assigned. Formally, we show that Assumption 1 is not sufficient for equation (5) to hold, such that β need not be identified by a conventional SEIV regression. The source of OVB in this case is the non-random variation in shock exposure, given by w and the $f_\ell(\cdot)$ mappings. While this exposure variation is potentially high-dimensional, our result shows that OVB is governed by a particular one-dimensional confounder μ_ℓ: what we call the *expected instrument*.

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 a conventional SEIV regression of y_ℓ on x_ℓ instrumented by z_ℓ when $\mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \varepsilon_\ell\right] \neq 0$.

Proof. We have $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mathbb{E}\left[f_\ell(g;w)\varepsilon_\ell \mid w\right]\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \mathbb{E}\left[\varepsilon_\ell \mid w\right]\right] = \mathbb{E}\left[\frac{1}{L}\sum_\ell \mu_\ell \varepsilon_\ell\right]$. 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. □

As with the motivating market access example in Section 2, the expected instrument μ_ℓ captures the average value of z_ℓ across different realizations of quasi-random shocks g, given w. Lemma 1 shows that z_ℓ is a valid instrument in the sense of identifying β if and only if μ_ℓ is not systematically correlated with the residual ε_ℓ, where as in Equation (3) this notion of exogeneity is over both the stochastic realizations of (g, w, ε) and across observations in the cross-section.

As the sole relevant confounder, μ_ℓ can be thought to generalize the well-known propensity score of Rosenbaum and Rubin (1983). Propensity scores are typically defined for binary and conditionally unconfounded treatments x_ℓ in settings with *iid* sampling. In such cases μ_ℓ can be written as a function $\tilde{\mu}(w_\ell) = Pr(x_\ell = 1 \mid w_\ell)$ of *iid* control vectors w_ℓ, which fully specifies the conditional distribution of x_ℓ. This μ_ℓ can, in principle, be non-parametrically estimated from observations of x_ℓ and w_ℓ. In our generalization μ_ℓ instead captures the conditional mean (rather than the distribution) of an instrument (rather than a treatment). Lemma 1 shows that only the conditional mean is relevant for OVB, as in the binary case, since IV is a linear estimator. This conditional mean, however, cannot generally be estimated non-parametrically estimated since the z_ℓ are co-determined by common $(g, w)$and the mapping of w to μ_ℓ may be different for each ℓ.

An immediate implication of Lemma 1 is that β is identified by the expected instrument μ_ℓ when Assumption 1 holds. In fact, again since IV is a linear estimator, identification follows under a weaker notion of shock exogeneity. Consider:

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

Such mean and covariance independence of shocks is implied by Assumption 1 and will also be sufficient for some of our later results. Here we use the first condition (mean independence) to show that recentering any shock-exposure instrument by its expected instrument ensures its validity:

Proposition 1. *Suppose Assumption 2(i) holds and let* $\tilde{z}_\ell = z_\ell - \mu_\ell$ *be a recentered instrument. 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 recentered SEIV regression of* y_ℓ *on* x_ℓ *instrumented by* $\tilde{z}_\ell$*, provided the first stage* $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right]$ *is non-zero.*

Proof. See Appendix C.1. □

As in the motivating railroad example, the recentered SEIV regression compares units with a higher-than-expected value of z_ℓ, because of the realized shocks, to units affected less than expected. Validity of $\tilde{z}_\ell$ thus stems from the exogeneity of shocks (specifically, Assumption 2(i)), even though the it continues to vary cross-sectionally due to heterogeneous shock exposure.

Lemma 1 also implies a regression-based solution to OVB: 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 sample projection of v_ℓ on μ_ℓ. Appendix C.1 shows that these moments also identify β under Assumption 2(i).

Implementing either solution to OVB requires measuring μ_ℓ. The expected instrument is generally derived from knowledge of the shock *assignment process*: the conditional quasi-experimental distribution of $g \mid w$. In Section 5, we discuss particular examples of how such knowledge might arise in different economic settings. For the remaining general discussion, we adopt the following general condition:

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

It is immediate from Proposition 1 that the recentered instrument $\tilde{z}_\ell$ which identifies β is observed under Assumption 3, since $\mu_\ell = \int f_\ell(g; w) dG(g \mid w)$ is observed.

Though the assumption of a known shock distribution may appear daunting, we note that Assumption 3 can be satisfied by the non-parametric and intuitive condition of *exchangeability* given an appropriate specification of w. Suppose, for example, that one assumes the shocks g_n are *iid* across n, such that all permutations of g are equally likely to arise. In this case $G(g \mid w)$ is known to be

uniform when w includes the permutation class $\Pi(g) = \{\pi(g) \mid \pi(\cdot) \in \Pi_N\}$, where Π_N denotes the set of permutation operators $\pi(\cdot)$ on vectors of length N (e.g. Lehmann and Romano, 2006, p. 634). The marginal distribution of g_n (conditionally on other components of w) need not be specified in this case; the expected instrument is the average z_ℓ across all permutations of shocks,

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

which is straightforward to compute (or approximate with a random set of permutations, when N is large).[7] Similar expected instrument calculations follow under weaker shock exchangeability conditions, such as when the g_n are *iid* within, but not across, a set of known clusters. In this case Assumption 3 is satisfied with w containing a restricted (within-cluster) permutation class and the μ_ℓ calculation appropriately modified. Appendix A.5 discusses how weaker assumptions on the shock assignment process, allowing for shock heteroskedasticity or unknown distributional parameters, can also be accommodated.

Another subtlety with Assumption 3 is that it requires the conditional distribution of g to be known in the full support of w. In some cases this is not restrictive, such as with the exchangeability assumption discussed above. It is also not restrictive if $g \perp\!\!\!\perp w$, as when w contains predetermined variables and g arises from a truly random process or natural event (such as an earthquake, as discussed in Section 5.5), since then $G(g \mid w)$ does not depend on w. However, in other cases specifying the distribution of $g \mid w$ for counterfactual w may be challenging. In a transportation network setting where upgrades g are drawn randomly from a predetermined plan w, it may be infeasible to specify the distribution of upgrades given *any* possible plan, or even define the set of such plans, i.e. the support of w. In these situations our framework still directly applies viewing w as non-stochastic. The only cost of this is in the interpretation of confidence intervals and other inferential objects discussed below.

3.3 Randomization Inference and Testing

We next show how knowledge of the quasi-experimental shock assignment process can be used to construct tests and confidence intervals for β, following a long tradition of randomization inference in statistics. This inference approach guarantees correct coverage in finite samples of both observations and shocks, even when the observations exhibit complex unobserved dependencies. We also show how SEIV randomization inference can be used to validate the quasi-experimental assumptions, through exact specification and falsification tests.

We begin by considering a test of some null hypothesis $\beta = b$. With $b = 0$, for example, we test that outcomes y_ℓ are unaffected by treatment x_ℓ. We consider a scalar test statistic $T = \mathcal{T}(g, y - bx, w)$,

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

where y and x are $L \times 1$ vectors collecting the outcome and treatment observations. When $b = \beta$, $T = \mathcal{T}(g, \varepsilon, w)$, and under Assumption 1 the distribution of this T conditional on ε and w is given by the shock assignment process $G(g \mid w)$. We know or may simulate this distribution under Assumption 3, by redrawing (or, in a special case, permuting) the shocks in g and recomputing T. If the original value of T is sufficiently far in the tails of the simulated distribution, we then have grounds to reject the null that $\beta = b$.

Formally, we have the following result on hypothesis testing:

Proposition 2. *Suppose Assumptions 1 and 3 hold, let $\alpha \in (0, 1)$, and for some $b \in \mathbb{R}$ and scalar-valued $\mathcal{T}(\cdot)$ let $T = \mathcal{T}(g, y - bx, w)$ and $T^* = \mathcal{T}(g^*, y - bx, w)$, where g^* is distributed according to $G(\cdot \mid w)$, independently of (g, x, y), conditionally on w. Under the null of $\beta = b$,*

$$Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \geq 1 - \alpha, \tag{9}$$

where the acceptance region is constructed for a given b as

$$T_{\alpha/2} = \sup\left\{t \in \mathbb{R} \cup \{-\infty\} : \ Pr(T^* < t \mid y, x, w) \leq \frac{\alpha}{2}\right\} \tag{10}$$

$$T_{1-\alpha/2} = \inf\left\{t \in \mathbb{R} \cup \{+\infty\} : \ Pr(T^* \geq t \mid y, x, w) \leq \frac{\alpha}{2}\right\}. \tag{11}$$

Equation (9) further holds with equality when $T^* \mid (y, x, w)$ is continuously distributed under the null.

Proof. See Appendix C.2. □

This result shows that when shocks are as-good-as-randomly assigned, a test of $\beta = b$ which rejects when $T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]$ has size of exactly α in finite samples when the test statistic is conditionally continuously distributed under the null. When this distribution is not continuous, the test is still guaranteed to be conservative with a rejection rate of no greater than α.[8] The lower- and upper-bounds of the test region, $T_{\alpha/2}$ and $T_{1-\alpha/2}$, are given by knowledge of the shock assignment process (Assumption 3) and represent the lower- and upper $\frac{\alpha}{2}$th percentile tails of the known conditional distribution of T^*. With exchangeable shocks, for example, $T_{\alpha/2}$ and $T_{1-\alpha/2}$ are given by the tails of the permutation distribution of $\mathcal{T}(g^*, y - bx, w)$ where $g^* = \pi(g)$ for random permutations $\pi(\cdot) \in \Pi$, holding (y, x, w) fixed. These tails can be computed from all permutations or from a random sample of them (Lehmann and Romano, 2006, p. 636).[9] We note that while the previous intuition for such a testing procedure conditioned on ε and w, Proposition 2 establishes correct unconditional

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

[9] When the realized g is added to the set of random permutations, the test remains exact (or slightly conservative because of discreteness) even if the number of random permutations does not grow with L (Lehmann and Romano (2006, p. 636), Hemerik and Goeman (2018)). In contrast to identification (i.e. footnote 7) randomness of permutations is important for inference; see Southworth et al. (2009) for an interesting counterexample.

coverage of the test. This follows by the law of iterated expectations: the unconditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right)$ is the expectation, across realizations of ε and w, of the controlled conditional coverage $Pr\left(T \in \left[T_{\alpha/2}, T_{1-\alpha/2}\right] \mid \varepsilon, w\right)$.[10]

It follows from Proposition 2 that one can construct confidence intervals for β with correct coverage in finite samples under Assumptions 1 and 3. Formally, we have the following result:

Corollary 1. *Suppose Assumptions 1 and 3 hold and let CI denote the set of $b \in \mathbb{R}$ that are not rejected by the test in Proposition 2. Then $Pr\left(\beta \in CI\right) \geq 1 - \alpha$, with equality if $T^* \mid (y, x, w)$ is continuously distributed.*

Proof. Follows from Proposition 2 by the standard logic of test inversion. □

In some settings, the confidence interval (or, more precisely, confidence set) CI obtained from inverting randomization tests may be infinite on one or both sides or empty, with the latter providing evidence against correct specification (i.e. Assumptions 1 and 3) (Imbens and Rosenbaum, 2005).

Different test statistics $\mathcal{T}(\cdot)$ will lead to different confidence intervals of correct size under the null, though they may differ in their power against alternative hypotheses. As usual with randomization-based inference procedures, little can be said about the relative power of different statistics in general. Instead, given our focus on the recentered SEIV estimator, we follow Rosenbaum (2002) and Imbens and Rosenbaum (2005) in proposing a T that induces it as the associated Hodges-Lehmann estimator (Hodges and Lehmann, 1963), as follows:

Proposition 3. *For a given statistic $T = \mathcal{T}\left(g, y - bx, w\right)$ and corresponding T^* defined in Proposition 2, define the Hodges-Lehmann estimator as the $b \in \mathbb{R}$ that solve $T = \mathbb{E}\left[T^* \mid y, x, w\right]$. Then the recentered SEIV estimator in Proposition 1 is the Hodges-Lehmann estimator associated with the statistic $T = \frac{1}{L}\sum_\ell f_\ell(g, w)\left(y_\ell - bx_\ell\right)$.*

Proof. See Appendix C.3. □

This result shows that the recentered SEIV estimate of β rationalizes the sample covariance between the instrument z_ℓ and implied residual $y_\ell - bx_\ell$ as typical under its null $\beta = b$.[11] A similar result, also proved in Appendix C.3, shows that the test statistic which induces the alternative μ_ℓ-controlled SEIV estimator is a Hodges-Lehmann estimator for the residualized covariance: $T = \frac{1}{L}\sum_\ell f_\ell(g, w)\left(y_\ell^\perp - bx_\ell^\perp\right)$. Notably, the same tests and confidence intervals are obtained by using the recentered $\tilde{z}_\ell$ in place of z_ℓ in both formulas. This holds because even without residualization

[10] It is instructive to highlight how exactly the knowledge of the shock assignment process matters in Proposition 2. Suppose that g is incorrectly assumed to be exchangeable, i.e. a uniform distribution is imposed over the $N!$ elements of g's permutation class. By construction, the test is guaranteed to reject the true β in some set of at most $\alpha \cdot N!$ permutations regardless of the true assignment process. However, unless the true conditional distribution of g is uniform, the probability of the realized shocks g being in the true rejection set need not be α, leading to size distortions.

[11] This definition of the Hodges-Lehmann estimator follows Rosenbaum (2002) and Imbens and Rosenbaum (2005). Hodges and Lehmann (1963) originally defines it as 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.

of $y_\ell - bx_\ell$ recentering shifts T and T^* by the same value $\frac{1}{L}\sum_\ell \mu_\ell (y_\ell - bx_\ell)$ and leaves the randomization test unaffected. In this sense, randomization inference performs the recentering needed for identification of β automatically.[12]

Statistics chosen on the basis of corresponding Hodges-Lehmann estimators are expected to inherit power properties of those estimators. While we are not aware of a general result, in Section 4.1 we show that these randomization tests are generally consistent, in the sense of having power against any fixed alternative that asymptotically increases to one, under the conditions which make the recentered or μ_ℓ-controlled SEIV estimator consistent.[13] In the conventional IV setting, Imbens and Rosenbaum (2005) show by Monte-Carlo simulation that randomization inference with such statistics has good finite-sample power properties, both with strong and weak instruments (while keeping the guarantee of finite-sample size, as any RI statistic would).[14]

We conclude this section by noting that randomization inference can also be used to indirectly validate the quasi-experimental model (i.e., equation (2) and Assumptions 1, maintaining Assumption 3) by conventional covariate balance and falsification tests. Given an observed r_ℓ thought to proxy for some component of ε_ℓ, such as a lagged outcome or predetermined covariate, one may test that $g \perp\!\!\!\perp r \mid w$ by seeing whether a given statistic $T = \mathcal{T}(g, r, w)$ is in the tails of its conditional-on-(r, w) distribution. The natural choice of the test statistic is $\frac{1}{L}\sum_\ell f_\ell(g, w) r_\ell$, as it induces the conventional balance regression coefficient from regressing r_ℓ on x_ℓ instrumented by $\tilde{z}_\ell$ as a Hodges-Lehmann estimator. Even absent a proxy for ε_ℓ, one may test correct specification of the shock assignment process (Assumption 3) by re-randomizing shocks. Note that $g \perp\!\!\!\perp \mu \mid w$ trivially, as each μ_ℓ is a deterministic functions of w. Thus the previous logic extends for the case of $r_\ell = w_\ell$; if, for example, $\frac{1}{L}\sum_\ell f_\ell(g, w)\mu_\ell$ is in the tails of its implied conditional-on-w distribution, we may have grounds to reject correct specification of $G(g \mid w)$. An even simpler specification test sets $r_\ell = 1$ and $T = \frac{1}{L}\sum_\ell f_\ell(g, w)$.

4 Consistency and Asymptotic Efficiency

We next consider the asymptotic properties of recentered shock-exposure instruments. While the results of the previous section yield tools for valid finite-sample SEIV inference on β, they do not en-

[12] One might think to compute confidence intervals from the distribution of the recentered SEIV estimator itself with re-randomized shocks g^*. This idea fails 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 it is centered around zero rather than β.

[13] In cases where the test statistic is asymptotically normal with estimable asymptotic variance (e.g., the shift-share IV example discussed in Section 5), DiCiccio and Romano (2017) recommend "studentizing" randomization-based test statistics by subtracting from T an estimate of its asymptotic mean and dividing by an estimate of its asymptotic standard deviation. This does not invalidate the finite-sample size of the test, but may yield large-sample robustness. For example, tests based on a model for $g \mid w$ in which the g_n are *iid* may be studentized to be asymptotically valid when shocks are instead heteroskedastic.

[14] Robustness of RI inference to weak IV follows for the same reason as robustness of the Anderson and Rubin (1950) approach: valid testing can be performed solely based on the properties of $\frac{1}{L}\sum_\ell z_\ell (y_\ell - bx_\ell)$ under the null, regardless of the first stage, and confidence intervals are constructed by test inversion.

sure that the recentered and μ_ℓ-controlled SEIV estimates are likely to be close to β in large samples. Establishing consistency is nontrivial here, as the inherent data dependencies of the SEIV setting invalidate conventional sampling-based asymptotics. A related question is asymptotic efficiency: while any shock-exposure instrument can be made valid by the appropriate recentering, the choice of instrument construction from the set of possible $\{f_\ell(\cdot)\}_{\ell=1}^{L}$ will generally matter for large-sample power.

4.1 SEIV Consistency

We first study consistency of the recentered shock-exposure IV estimator

$$\tilde{\beta} = \beta + \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell}. \tag{12}$$

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

Assumption 4. *(Relevance):* $\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell \xrightarrow{p} M \neq 0$.

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

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

Assumption 5. *(Regularity):* $\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right] \leq B$ *for finite* B.

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

Assumption 6. *(Weak SEIV dependence):* $\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0$.

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

Proposition 4.

(i) *Suppose Assumptions 2–6 hold. Then* $\tilde{\beta} \xrightarrow{p} \beta$.

(ii) *Suppose Assumptions 1 and 3–6 hold with* $\mathbb{E}\left[x_\ell^2 \mid w\right]$ *and* $\mathbb{E}\left[x_\ell \varepsilon_\ell \mid w\right]$ *uniformly bounded. Then the randomization test of Proposition 2 with* $T = \frac{1}{L}\sum_\ell f_\ell\left(g, w\right)\left(y_\ell - b x_\ell\right)$ *is consistent, i.e. for any* $b \neq \beta$ *we have* $Pr\left(T \notin \left[T_{\alpha/2}, T_{1-\alpha/2}\right]\right) \to 1$.

Proof. See Appendix C.4. □

The key condition of weak SEIV dependence states that the average absolute value of mutual covariances of the recentered instrument $\tilde{z}_\ell$ converges to zero as L grows. Typically, this would require the number of shocks N to grow with L, so that only a small fraction of observation pairs are most exposed to the same shocks in g. When this condition holds, Proposition 4 shows that $\tilde{\beta}$ is consistent even when unobserved shocks affect observations jointly (through ε_ℓ), in an unspecified manner.[15] While Proposition 4 applies to recentered SEIV, Appendix A.1 extends it to the μ_ℓ-controlled SEIV regression (Proposition 6(v)).

Our two sufficient conditions for Assumption 6 are non-nested:

Lemma 2.

(i) *Suppose* $\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost surely for all* ℓ *and* m*, and Assumption 1 holds. Then Assumption 6 holds if* $\mathrm{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell\right] \to 0$*. Moreover, if* $f_\ell(g; w)$ *is weakly monotone in* g *for all* ℓ*, and components of* g *are jointly independent conditionally on* w*, then* $\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right] \geq 0$ *almost surely.*

(ii) *Suppose* $G_\ell \subseteq \{1, \ldots, N\}$ *is such that* $f_\ell\left(\cdot; w\right)$ *does not depend on* g_n *for any* $n \notin G_\ell$ *almost surely. Then Assumption 6 holds if* $\frac{1}{L^2}\sum_{\ell,m} \mathbf{1}\left[G_\ell \cap G_m \neq \varnothing\right] \to 0$*, the components of* g *are jointly independent conditionally on* w*, and* $\mathbb{E}\left[\tilde{z}_\ell^2 \mid w\right]$ *is uniformly bounded.*

Proof. See Appendix C.4. □

The first condition applies to the setting when all shocks affect all observations in the same direction, but to different extents. This holds, for example, for shift-share instruments with non-negative exposure weights. More generally nonlinear $f_\ell(\cdot)$ may also be monotone in the shock vector; for example each transportation infrastructure upgrade may weakly improve market access everywhere. In these cases, the recentered SEIV estimator is consistent when the first-stage covariance converges to a non-zero constant M and the average instrument $\frac{1}{L}\sum_\ell \tilde{z}_\ell$ converges to its expectation of zero in the l_2 norm. For linear shift-share IV this extra condition requires the number of shocks to grow with L with the average exposure to each individual shock becoming vanishingly small, as in Borusyak et al. (2019) and Adão et al. (2019b). The assumption of independent shocks can be weakened, for instance to allow for shocks that are independent across many clusters. The second condition in Lemma 2

[15] We note that the recentering of z_ℓ is key for this result. Otherwise, when the ε_ℓ are mutually correlated (because of, for example, unobserved shocks), the non-recentered SEIV may not converge to β even when z_ℓ is valid in the sense of $\mathbb{E}\left[\frac{1}{L}\sum_\ell z_\ell \varepsilon_\ell\right] = 0$; see Lee and Ogburn (2019) for a related example.

follows Aronow and Samii (2017) in assuming that for most pairs of observations the two instruments $\tilde{z}_\ell$ and $\tilde{z}_m$ rely on non-overlapping sets of shocks g. This would be the case, for example, when each observation receives its own quasi-experimental shock, and $f_\ell(\cdot)$ only depends on ℓ's shock and those of its neighbors up to a fixed network distance.[16]

4.2 Optimal Shock-Exposure Instruments

The results of Section 3 imply that, given shock assignment knowledge, valid instruments can based on any instrument constructions $f_\ell(\cdot)$. However, the choice of $f_\ell(\cdot)$ will generally affect the power of null hypothesis tests and lengths of associated confidence intervals. Analytical results on the relative power of randomization inference tests are difficult to obtain in general. Instead, we next characterize the $f_\ell(\cdot)$ that minimize the asymptotic variance of $\tilde{\beta}$ and thus in large samples efficiently use the quasi-experimental variation in shocks. This result extends the classic result of Chamberlain (1987) to the shock-exposure setting. Constructing these optimal instruments may not be feasible in practice, and typically requires an economic model for the dependence of treatment on shocks. Our characterization nevertheless provides guidance for shock-exposure instrument construction by explaining what researchers should strive for when choosing between alternative SEIVs.

Formally, we have the following result:

Proposition 5. *Suppose Assumption 2 holds.*
(i) $z^* = \mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1}\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)$ *minimizes* $V = \mathrm{Var}\left[\tilde{z}'\varepsilon\right]/\mathbb{E}\left[\tilde{z}'x\right]^2$ *over all recentered SEIVs* $\tilde{z}$*, i.e. those satisfying* $\mathbb{E}\left[\tilde{z} \mid w\right] = 0$*. The corresponding value of* V *is given by*

$$\mathrm{Var}\left[z^{*\prime}\varepsilon\right] = \mathbb{E}\left[z^{*\prime}x\right]^{-1} = \mathbb{E}\left[\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)'\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]^{-1}\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right]^{-1}. \quad (13)$$

(ii) *Suppose further* $\tilde{z}^*$ *is consistent and has an asymptotic first stage, i.e.* $\frac{1}{L}z^{*\prime}x \xrightarrow{p} M$ *for some* $M \neq 0$*. Then any recentered SEIV* $\tilde{z}$ *that is also consistent and has an asymptotic first stage is asymptotically weakly less efficient than* z^**. That is, for* $\tilde{\beta} = \frac{1}{L}\tilde{z}'\varepsilon/\frac{1}{L}\tilde{z}'x$*,* $\beta^* = \frac{1}{L}z^{*\prime}\varepsilon/\frac{1}{L}z^{*\prime}x$ *and for any* $\delta > 0$*,* $\mathrm{Var}\left[\tilde{\beta}\right]/\mathrm{Var}\left[\beta^*\right] > 1 - \delta$ *for large enough* L*.*

Proof. See Appendix C.5. □

Proposition 5 relates to the optimal instrument result of Chamberlain (1987), as the optimal instrument reweights a conditional expectation of treatment given shocks and w. A key difference is the subtraction of $\mathbb{E}\left[x \mid w\right]$, corresponding to the expected instrument recentering that follows from the unconventional conditional moment restriction of Assumption 2(i). Another difference is that when

[16] Assumption 6 and Lemma 2 may be more difficult to apply when $w = (\tilde{w}, \Pi(g))$ includes the permutation class of shocks. Even if shock components g_n are *iid* conditionally on $\tilde{w}$, they are dependent conditionally on the permutation class (and, in the scalar g_n case negatively correlated). Appendix C.8 proves a version of Proposition 4 that applies in that case, with similar conditions but some of them applied with $\tilde{w}$ and others to w.

$\mathbb{E}[\varepsilon \mid w] \neq 0$, the $\mathbb{E}[\varepsilon\varepsilon' \mid w]^{-1}$ reweighting is based not only on the residual variance but also the residual mean. If $(\varepsilon_\ell, g_\ell, x_\ell)$ are *iid* and w is constant, so no adjustment is necessary and $\mathbb{E}[\varepsilon \mid w] = 0$, the optimal shock-exposure instrument is $\tilde{z}^*_\ell = \frac{\mathbb{E}[x_\ell|g_\ell]-\mathbb{E}[x_\ell]}{\mathrm{Var}[\varepsilon_\ell]}$, as in Chamberlain (1987).

Practical implementation of the optimal instrument typically requires a model for treatment. The key ingredient of z^* is the best prediction for each observation's x_ℓ from g and w. With a large and *iid* sample this can potentially be estimated non-parametrically (e.g. Newey (1990)), though the curse of dimensionality may make it difficult in finite samples. However, when x_ℓ is jointly determined by all shocks g, observables w and some unobservable shocks u (e.g., $x_\ell = x_\ell(g, w, \eta)$ with $\eta \perp\!\!\!\perp g \mid w$ and $\mathbb{E}[u] = 0$), *a priori* restrictions will typically be needed on this mapping, such as those based on economic theory. The reweighting term $\mathbb{E}[\varepsilon\varepsilon' \mid w]^{-1}$ need not be implied even by such models.[17]

Some instruments used in practice can be viewed as approximating $\mathbb{E}[x \mid g, w]$. Shift-share instruments, for example, have been derived from economic models as first-order approximations to how treatment changes in response to shocks g (Kovak, 2013; Adão et al., 2019b). Moreover, some papers (e.g., Donaldson and Hornbeck (2016) and Tsivanidis (2017)) that do not impose linearity instead use a model to predict the counterfactual change in treatment in the scenario where g is the only shock, while all other exogenous variables are kept at their original equilibrium values. That is if the model implies $x_\ell = x_\ell(g, w, u)$ for known $x_\ell(\cdot)$ and unobserved mean-zero u, the instrument is given by $x_\ell(g, w, 0)$. This can be viewed as a convenient approximation to $\mathbb{E}[x_\ell(g, w, u) \mid g, w]$: instead of averaging across all values of unobserved shocks, which would require additional knowledge of their distributions, x_ℓ is evaluated at the means of those shocks (see Berry et al. (1999), for the same idea in an entirely different context). These papers, however, do not recenter by $\mathbb{E}[x_\ell(g, w, 0) \mid w]$ which is generally required when w is endogenous.

5 Practical Implications and Applications

We now discuss implications of the general SEIV framework for some common examples of shock-exposure instruments, found in various economic fields. These examples include instruments derived from shocks to networks (such as transportation upgrades), simulated policy eligibility instruments, and shift-share or "Bartik" instruments. Bringing our general approach to these settings reveals various ways researchers can ensure valid and powerful inference with conceptually straightforward adjustments to existing methods. We illustrate some of these insights in three empirical applications.

[17] One can in fact improve on (13) given a model for the conditional distribution of $\varepsilon \mid w$. Replacing y with $\check{y} = y - \mathbb{E}[\varepsilon \mid w]$ and thus ε with $\check{\varepsilon} = \varepsilon - \mathbb{E}[\varepsilon \mid w]$ (while not adjusting x) implies $\mathbb{E}[\check{\varepsilon} \mid w] = 0$. In that case any function $f_\ell(g, w)$ is a valid instrument, and the expected instrument adjustment—which isolates the variation in g but reduces power—is not necessary. By analogy with Proposition 5, $\check{z} = \mathbb{E}[\check{\varepsilon}\check{\varepsilon}' \mid w]^{-1}\mathbb{E}[x \mid g, w]$ is then optimal, and generally differs from the optimal recentered instrument $\tilde{z}^*$.

5.1 Network and Transportation Instruments

We interpret equation (2) as a simple model of network effects. Suppose observations ℓ represent nodes in a network (of people, firms, regions, etc.), β denotes the causal effect of some treatment x_ℓ, and this treatment is affected by quasi-experimental shocks to the network nodes or links. A typical case is when x_ℓ captures relevant spillovers at node ℓ from random shocks to other connected nodes. For example, we may be interested in the effects of a co-inventor's death on an inventor's productivity (Jaravel et al., 2018), of having a direct supplier or supplier's supplier hit by a natural disaster on a firm's growth (Carvalho et al., 2016), or of having more "dewormed" students at neighboring schools on a student's test scores (Miguel and Kremer, 2004). We also consider settings in which x_ℓ measures node ℓ's network centrality itself, after a shock to the network structure. For example, we may be interested in the effects of regional market access, as a theoretically founded measure of centrality, after quasi-experimental changes to transportation networks (Donaldson and Hornbeck, 2016). As in our general framework we assume that the treatment has been well-specified, in that it captures all relevant channels by which the shocks affect node ℓ. This exclusion restriction may be given by a particular model for spillovers or general equilibrium effects. Our setup nests reduced-form regressions where the instrument z_ℓ coincides with x_ℓ (as in the deworming example), as well as IV regressions where treatment is affected by both quasi-experimental and other shocks, while the instrument captures only the former .

Our general framework (Section 3) shows that in general these network effect regressions may suffer from omitted variable bias. Here OVB arises from the fact that different nodes face systematically different values of the instrument due to their network position, even when the underlying shocks are as-good-as-randomly assigned. This problem has been previously noted by Aronow (2012): *"Simple randomization of treatment to individuals does not imply simple randomization of proximity to treated units."* . Whether OVB arises, and whether it can be solved by simple low-dimensional controls, will generally depend on two factors: how z_ℓ is constructed from the shocks (what Aronow and Samii (2017) refer to as the exposure mapping) and the shock assignment process.

We first illustrate the OVB problem with the simple case of Miguel and Kremer (2004). In this reduced-form study, spillovers are specified as $z_\ell = x_\ell = N_\ell^T$ where N_ℓ^T denotes the number of treated (dewormed) students at neighboring schools. This N_ℓ^T is a thus a function of which students are dewormed g (which is randomly assigned with a known assignment process) and the existing network of students and schools. It is easy to see that in this case students who live in denser areas will systematically have higher values of N_ℓ^T, regardless of which students are experimentally dewormed; these areas may in turn have different latent health outcomes and trends. Miguel and Kremer address this OVB problem by controlling for the total number of eligible students in neighboring schools, N_ℓ. This is indeed what our approach recommends when all students have equal chances of being dewormed, q: the expected instrument $\mu_\ell = \mathbb{E}\left[N_\ell^T\right] = qN_\ell$ is then proportional to N_ℓ, and Lemma 1 shows

that including this control is enough to purge OVB. This is true both with a simple randomization of deworming at the student level and with Miguel and Kremer's clustered randomization scheme, first at the school level and then at the student level.

This simple solution to OVB generally fails for more complex shock assignment processes. In the Miguel and Kremer (2004) setting, imagine a stratified random assignment system which makes deworming more likely for some students or schools (e.g. depending on student gender or school size). This would make μ_ℓ no longer proportional to N_ℓ. Our framework continues to make the solution to OVB straightforward whenever the shock assignment process is known; μ_ℓ can be computed analytically or simulated by reassigning shocks consistently with the experimental protocol, and again controlled for. Outside of the experimental ideal, one might assume partial exchangeability of shocks conditional on observables and use this to compute an appropriate μ_ℓ. The approach of Jaravel et al. (2018), for example, in which deceased and non-deceased co-inventors are matched based on age and other characteristics, can be viewed as implicitly leveraging such an exchangeability assumption. More generally the shock assignment process may be assumed to be known up to a set of parameters, which the researcher may estimate in a zeroth stage.[18]

Including simple controls may also be insufficient to prevent OVB when spillover effects are nonlinear in shocks, even for simple assignment processes. If, for example, Miguel and Kremer had studied the effects of having at least one treated neighbor, $\mathbf{1}\left[N_\ell^T > 0\right]$, the appropriate μ_ℓ control would have been the student-specific probability of having such neighbor. When deworming is simply randomized across students, this probability is a function of the total number of peers N_ℓ, and flexible controls for N_ℓ would solve the OVB problem. This is no longer true for clustered randomization: the probability of having at least one treated neighbor is much lower due to the correlation structure of treatment and is not summarized by N_ℓ. As before, our framework yields a simple way to compute the relevant μ_ℓ in this case: simulating the shock assignment process.

An additional layer of complexity arises when shocks are allowed to propagate through the entire network, and not just to immediate neighbors. Carvalho et al. (2016), for example, study the effects of network distance from firm ℓ to the nearest firm in the area of a natural disaster (specifically, the 2011 Tōhoku earthquake) in the full firm-to-firm supplier network. This specification of network spillovers inherently makes μ_ℓ more complex: for a firm with a given number of suppliers, the probability of having at least one supplier's supplier in the randomly assigned earthquake zone (i.e. at distance of two) also depends on how connected and how spatially dispersed its suppliers are. Once again this probability can be computed by simulating the earthquake from the knowledge of its assignment process. For Carvalho et al., geological knowledge may be useful for specifying the risk of earthquakes affecting different parts of the network.

[18] See Appendix A.5 on shock assignment processes with unknown parameters. Jaravel et al. (2018) avoid the need to simulate the μ_ℓ because deaths are rare events and the same inventor is very unlikely to have more than one co-inventor who is deceased or matched to a deceased one. However, our solution based on simulating the process where deceased inventors are randomly chosen from matched pairs applies generally.

Market access regressions of D-H, introduced in Section 2, is perhaps the most elaborate network example. Not only do transportation upgrade shocks propagate through the entire network, but their effect on a region's market access depends nonlinearly on other transportation links as well as the distribution of population across regions (and potentially other factors affecting transportation costs). Again, however, the expected growth of regional market access may be readily simulated from knowledge of the upgrade assignment process.

The processes which determine transportation upgrades are all but simple in reality, making it challenging to obtain valid counterfactual maps and apply our framework. We discuss two possible approaches here. One natural approach is to use network upgrade plans and appeal to the quasi-randomness of which subset of the plan materializes, as of some time period or ever. Notably, this use of plans contrasts with the typical one in the literature: rather than viewing the planned network as exogenous (which does not explicitly yield counterfactuals), one may consider the realized network as exogenous, conditional on the plan.[19] Donaldson (2018) and Berger and Enflo (2017) take a step in this direction, using planned but unbuilt railroads in a placebo exercise, as does Lin (2017) by exploiting engineering problems that slowed down construction of certain lines in a robustness check. We illustrate the construction of counterfactual networks based on alternative subsets of planned railroads in Lin's setting below.[20]

A different approach is to specify, based on the economics or politics of the upgrade, the key criteria that the upgrade had to satisfy and then find alternative networks which also satisfy these criteria. One of the placebo analyses by Ahlfeldt and Feddersen (2018) follows this logic: they note that the new railway line connected two major cities in Germany at distance around 160km and had three intermediate stops. They then construct 1,000 random placebo lines that satisfy the same description. Dell and Olken (2018) develop a similar strategy in a different context: they list main factors that affected locations of sugar processing factories in 19th-century Java, such as proximity to a river and distance from each other, and prepare 1,000 counterfactual maps that are similar to the realized ones on these dimensions.

We acknowledge that in most of the described situations there is no "true" randomness of infrastructure construction, in contrast with some of the examples considered above. Still, specifying counterfactuals makes the assumption of upgrade exogeneity explicit, transparent, and testable.[21] Moreover, probing expected instrument controls implied by some reasonable counterfactuals may be useful for robustness analysis: if the μ_ℓ that corresponds to the unknown true shock assignment process is cross-sectionally uncorrelated with the residual and thus not necessary to control for, coefficient

[19] Historical routes can be used in place of the plan if the researcher is willing to assume that some of them decayed for random reasons while others continued to be in use. This again contrasts with the typical use of historical routes (e.g. Duranton et al. (2013)).

[20] This strategy can be compared to the use of accepted vs. rejected sites for plants in Greenstone et al. (2010), but in a setting with spatial interdependence.

[21] Absent assumptions on shock assignment, strong assumptions have to be imposed on the error term: it should be unrelated to all geographic features that may be correlated with unobserved μ_ℓ.

estimates should be robust to including proxies for μ_ℓ in the regression.

Besides providing a solution to the OVB problem, the SEIV framework helps address well-known challenges of network inference. The standard approach in the literature is to assume that $z_\ell \varepsilon_\ell$ is uncorrelated beyond a small geographical or network distance (Conley, 1999; see Acemoglu et al., 2015 for a network example). This may work well when z_ℓ captures local spillovers. However, as discussed in Section 2, when both observed and unobserved shocks may propagate further through the network, implied standard errors may be too small. For example, long railway lines, such as the Pacific Railroad in the 19th-century U.S. or Lanzhou-Xinjiang line in present-day China, generate correlation in connectivity and market access between cities far away from each other, and allow unobserved shocks to be correlated between those cities. As usual, randomization inference is valid regardless of the correlation structure of unobserved shocks, relying only on the correlation structure of the SEIV that is implied by knowledge of the shock assignment process.

We finally note that the same ideas apply more broadly than to regressions estimated at the node level. For example for regressions that relate dyadic (node pair-level) outcomes to the network distance between them, exploiting a shock to the network structure. In this vein, Allen et al. (2019) study how migration between locations in the U.S. and Mexico is affected by log travel time between them, as affected by the wall in some parts of the border between the two countries. One may expect that the change in travel time is correlated with the distance between locations (and thus potentially with the error term): regardless of which sections of the wall are built, places far away from the border will be affected less, as there would always be alternative routes between them unaffected by the wall. Taking an explicit stand on the counterfactual distribution of walls and performing our expected instrument recentering is guaranteed to isolate quasi-experimental variation. Our approach further allows for inference that takes into account that the same segment of the wall may simultaneously affect pairs of locations that are far away from each other—a form of fuzzy clustering by the border segment.

Application We illustrate the potential importance of SEIV recentering in the market access case by estimating the reduced-form effect of predicted log market access growth $z_{\ell t} = \Delta \log \widehat{MA}_{\ell t}$, constructed from the expansion of high-speed rail (HSR) in China over t =2002–2016 while holding population fixed, on the growth of regional outcomes $\Delta \log V_{\ell t}$ across 306 mainland Chinese prefecture-level cities ℓ.[22] As mentioned above, Lin (2017) previously studied the effects of this expansion with a sophisticated market access measure. For the purposes of illustration, we employ a simple predicted market access construction of $\widehat{MA}_{\ell t} = \sum_k \tau_{\ell k t}^{-\theta} P_{k,2000}$ where $P_{k,2000}$ is the initial population of region k in the year 2000, the $\tau_{\ell k t}$ denote predicted travel costs between each region ℓ and k at time t, and $\theta = 1$ (such that $\widehat{MA}_{\ell t}$ aligns with the notion of "market potential" (Harris, 1954)). Transportation

[22] We focus on the reduced form here because of data constraints: annual population growth is not well measured for many of the cities in our sample. This is not an issue for the predicted market access instrument, which holds fixed city population in a pre-period (here, 2000).

costs are themselves simple functions of both geographic distance and completed HSR lines by time t. We focus on the effects of market access on log urban employment growth. Appendix B.1 describes the sample construction and data definitions for this application in detail.

Panels A-E of Figure 2 show the evolution of new HSR construction in China from 2003 (when the first line opened) to 2016, with Panel F showing the full set of lines including those not yet completed but planned as of 2016. The shading in this figure indicates the change in market access (relative to 2000) in each year. The placement of HSR is clearly non-random, unlike in our stylized example in Section 2: major cities such as Beijing and Shanghai get connected while less developed areas do not. Appendix Figure A1 shows that the growth in market access over this period correlates with an increase in rail ridership, suggesting scope for real economic effects.

Panel A of Figure 3 shows that the growth in log market access between 2002 and 2016 is correlated with the growth in regional employment, with a positive regression slope. This slope implies an elasticity of $\beta = 0.57$ that is statistically significant at conventional levels (using spatial-clustered standard errors that allow for correlation of cities within a 500km distance). Panel B furthermore shows that the 2002-2016 growth in log market access is uncorrelated with employment pre-trends (log employment growth over 1999-2002). Such falsification tests are often used to gauge the plausibility of a research design, supporting a causal interpretation of Panel A. Our Proposition 1 shows, however, that this interpretation is only valid when each city's average exposure to quasi-experimental network upgrades is uncorrelated with its potential employment trends.

Without knowing the true China HSR design, we investigate the potential for omitted variables bias in Figure 3A by two recentering strategies. We first suppose that Assumption 1 holds with w containing the plan for eventual HSR construction but not the realized opening dates, which we assume are exchangeable across lines. This allows us to compute the expected instrument $\mu_{\ell t}$ for each city ℓ in each period t by randomly shuffling line opening dates. A second more restrictive assumption includes the date at which each line began construction in w, while instead assuming the construction time is exchangeable across lines. This would be more plausible if, as one would expect, construction projects near major cities were prioritized, but that idiosyncratic shocks determine how long each project took to complete.

Panel A of Figure 4 plots the expected instrument under the initial "opening date" design assumption in 2016, along with the actual HSR construction by this year (see Appendix Figure A2 for a corresponding plot for the "random construction time" assumption). As expected, cities with nearby lines tend to have higher expected market access growth. At the same time, some cities have nearby HSR lines but low expected instruments, while others have high expected $\mu_{\ell t}$ but low $z_{\ell t}$. As a result, there remains variation in the recentered market access measure plotted in Panel B of Figure 4. This variation, of market access growth in 2016 that was higher or lower than expected given a city's location, is exogenous if the HSR assignment process is correctly specified.

Panel A of Figure 5 suggests that the variation in the expected instrument $\mu_{\ell t}$ is not idiosyncratic, such that omitting it from the non-recentered instrument is likely to cause OVB in estimates of β. Variation in expected market access growth in 2002-2016 (again computed under the "random opening date" assumption) is strongly correlated with employment pre-trends, with a coefficient of 0.56 that is quite similar to the simple estimate of β in Figure 3. Panel B of Figure 5 furthermore shows that the estimated effect from Figure 3 is significantly reduced when log market access growth is recentered by the "opening date" expected instrument.[23] The coefficient falls by two thirds, to 0.20, and becomes statistically insignificant at conventional levels with spatial-clustered standard errors. Panel A of Appendix Figure A4 shows a further decline when the market access instrument is recentered under the "construction time" design assumption.

Table 1 confirms this finding in panel regressions of log employment on the market access measure with city and year fixed effects, including all years in 2000-2016. Column 1, which does not recenter market access growth, shows a statistically significant panel estimate of $\beta = 0.25$, with a spatial-clustered confidence interval of $[0.06, 0.44]$. Column 2 shows that recentering by the simple "random opening date" expected market access measure in Panel A or controlling for the expected instrument in Panel B substantially reduces the point estimate while also yielding conventional 95% confidence intervals which include zero. Column 3 shows that the more stringent "random construction time" assignment process for HSR shocks yields similar results. Permutation-based 95% confidence intervals, which leverage the corresponding line assignment assumptions, are also shown to be substantially wider in Table 1 than the conventional spatial-clustered confidence intervals. This suggests that the residual variation in log employment growth are is correlated across geographically distant cities located near to the same HSR lines.[24] Finally, Panel C tests both of the simple design assumptions by randomization inference, using both $T = \frac{1}{L}\sum_\ell \tilde{z}_\ell \mu_\ell$ and $T = \frac{1}{L}\sum_\ell \tilde{z}_\ell$ as test statistics. While both tests reject the null of random line opening dates specification in column 2, random construction time is not rejected at the 5% significance level in column 3.

5.2 Simulated Eligibility Instruments

Our framework also brings insights to settings in which quasi-experimental policy variation differentially affects the eligibility of individuals for a government program, such as Medicaid or unemployment insurance.[25] Currie and Gruber (1996a; 1996b, henceforth CG) famously construct "simulated instruments" for eligibility x_ℓ in regressions of individual outcomes, such as program participation, health status or educational attainment. These instruments are designed to address the endogeneity of el-

[23] Appendix Figure A3 shows that the recentered instrument under both exogeneity assumptions does not systematically correlate with employment pre-trends.

[24] Appendix Table A1 repeats these specifications for an outcome of log GDP growth. In contrast to log employment, market access growth does not appear to predict local GDP growth even absent the expected instrument recentering. This remains true under our two HSR design assumptions.

[25] We thank Paul Goldsmith-Pinkham for helping us make this connection.

igibility, which depends both on policies and on non-random individual characteristics that may be correlated with potential outcomes. The CG solution is to leverage the plausible exogeneity of policies themselves, by constructing an instrument that only varies across an individual's state-of-residence.[26]

The general SEIV approach provides a way to construct valid instruments from quasi-experimental eligibility policy variation that preserve individual-level exposure to such policies. By leveraging more variation, and better predicting true eligibility, these instruments can potentially yield substantial power gains.

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

When state policies are as-good-as-randomly assigned with respect to potential outcomes or trends ε_ℓ (i.e., the rows of g are exchangeable), the CG instrument is valid without further adjustment; in our framework the expected instrument $\mathbb{E}\left[f_\ell(g; w) \mid w\right] = \frac{1}{50}\sum_n h(g_n)$ is constant. Per Section 4, valid tests of $\beta = b$ can be constructed by comparing the z_ℓ^{CG} IV statistic $\frac{1}{L}\sum_\ell z_\ell^{CG}(y_\ell - bx_\ell)$ to the tails of its distribution under state policy permutations. Since z_ℓ^{CG} only varies at the state level, conventional state-clustered standard errors may also be asymptotically valid provided eligibility policy is independent across states.

As in Section 4.2, more powerful tests and shorter confidence intervals may be obtained by replacing the CG instrument with a shock-exposure instrument that better predicts individual eligibility. In practice, all individual characteristics relevant to eligibility determination—such as income, family structure, or employment status—may be observed and contained in w_ℓ. In such cases $\mathbb{E}\left[x_\ell \mid g, w\right] = x_\ell$, or an individual's eligibility itself. Thus apart from the $\mathbb{E}\left[\varepsilon\varepsilon' \mid w\right]$ adjustment, the optimal instrument (**??**) is $\tilde{z}_\ell^* = x_\ell - \mathbb{E}\left[x_\ell \mid w\right]$, and the efficient IV procedure is asymptotically equivalent to a regression of y_ℓ on $\tilde{z}_\ell^*$. Our framework then motivates a regression of outcomes on eligibility, recentered by or controlling for the individual's average eligibility over random reallocations of policies. This contrasts with the CG approach: while they apply ℓ's state policy to random individuals to construct the instrument, we apply random states' policies to ℓ to construct the control.

There are two additional considerations motivating the use of the CG instrument which can be incorporated in our approach. First, in some settings not all determinants of eligibility are observed in w_ℓ. For example Cohodes et al. (2016) study the long-term effects of Medicaid eligibility on children

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

without observing its key determinant (parental income). They use a simulated eligibility instrument which assigns to each individual of a given age, race, and birth year the average eligibility among such individuals nationally if they were subject to her state-of-residence's policy. Our framework instead motivates instrumenting by the average eligibility of individuals with similar demographics actually residing in each state, while adjusting for the average value of this instrument over permutations of state policies. By virtue of Proposition 4 this will again improve efficiency, to the extent the non-observed determinants of eligibility are systematically different across states.

Second, even when all eligibility determinants are observed, a researcher may not wish to include them in r_ℓ. This would be the case when, for example, parental income responds endogenously to the state policy, violating Assumption 1. Indeed, Currie and Gruber (1996a) discuss this as one of the motivations for their original simulated instrument construction. In such cases exogenous predictors of such determinants, such as parental income from before a state policy change, may be used instead. East and Kuka (2015) use a similar approach to augment simulated instrument construction in evaluating the effects of unemployment insurance eligibility.

A related literature on unemployment insurance eligibility effects (Cullen and Gruber, 2000; East and Kuka, 2015) has taken a different approach than CG: regressing outcomes on true or predicted eligibility, while flexibly controlling for individual's observables v_ℓ. This approach is also justified within our SEIV framework since the expected instrument is by necessity some function of v_ℓ. Flexible controls for individual characteristics have an additional benefit of potentially predicting variation in the error term, thus improving asymptotic efficiency. However, Gruber (2003) finds this strategy difficult for Medicaid because several demographic variables have nonlinear effects on eligibility, leading to the curse-of-dimensionality. Focusing on the (perhaps implicit) quasi-experiment in such investigations—that of as-good-as-random state-level policy variation—reveals the single control needed for valid causal inference, avoiding the problem while still allowing for other flexible controls to reduce residual variance.

Application We illustrate the power gain yielded by our SEIV approach to simulated instruments, leveraging a plausibly exogenous change in Medicaid eligibility induced by the US Affordable Care Act (ACA) in 2014. A 2012 Supreme Court decision (National Federation of Independent Business v. Sebelius, 567 U.S. 519) permitted state governors to avoid increasing the household income eligibility threshold for Medicaid to 138% of the federal poverty level (FPL) following a protracted legal challenge of this ACA provision. As a result of this decision, only 24 states plus Washington D.C. expanded to this level of Medicaid generosity by January 2014 (Kaiser, 2015), with individuals in the remaining states continuing to face a variety of stricter income-based Medicaid eligibility rules. Frean et al. (2017) previously studied the coverage effects of this expansion in combination with other ACA policy changes using a simulated instrument triple-difference estimation strategy.

We follow Frean et al. (2017) in studying the effect of ACA-induced Medicaid eligibility on both "take-up" of Medicaid coverage and "crowd-out" of private insurance coverage, both having important implications for public insurance provision. To this end, we construct a representative sample of adults (ages 21-64) residing in the 43 states that did not expand Medicaid eligibility as in the ACA prior to January 1, 2014 from a random sample of 1% of the population collected through the American Community Survey (ACS) in 2012-2015; our main analysis studies expansion effects in 2014. The repeated cross-section includes information on annual insurance coverage, household income, and other variables potentially determining Medicaid eligibility such as family structure (e.g., employment and child status). Construction of these data and the Medicaid eligibility rules in each state and year follows Frean et al. (2017) and is discussed in detail in Appendix B.2.

Columns 1–2 of Table 2 start by reporting simple difference-in-difference estimates of 2014 Medicaid eligibility effects, from regressions of the form

$$y_{\ell t} = \alpha + \beta x_{\ell t} Post_t + \tau Post_t + v'_{\ell t}\gamma + \varepsilon_{\ell t} \tag{14}$$

where t is either 2013 or 2014 and $Post_t$ is a 2014 year indicator. Here the outcome $y_{\ell t}$ is an indicator for a type of insurance coverage (Medicaid, private insurance, or other), $x_{\ell t}$ is an indicator for Medicaid eligibility under the realized state policies in 2014, and the vector $v_{\ell t}$ includes the uninteracted $x_{\ell t}$, state fixed effects, and in column 2 also a fourth-order polynomial in household income as a percentage of the federal poverty line (interacted with $Post_t$). OLS estimates of equation (14) thus capture differences in coverage trends, from 2013 to 2014, between individuals residing in states with more or less generous Medicaid eligibility rules following the partial ACA expansion. Following the literature, we estimate this equation with conventional state-clustered standard errors.[27]

Column 1 of Table 2 shows that, absent further controls, 2014 Medicaid eligibility is associated with 5.2 percentage point (pp) higher insurance coverage. This correlation arises almost entirely from higher Medicaid coverage (5.6pp). There is only a small and statistically insignificantly negative association with private insurance coverage, and a similarly small (but significant) positive association with other forms of coverage.[28] Column 2 shows, however, that this lack of crowd-out is likely a spurious finding, driven by the fact that both 2014 Medicaid eligibility and private insurance coverage are highly correlated with household income. Including a fourth-order income polynomial in $v_{\ell t}$ (interacted with $Post_t$) results in a large and statistically significant negative 2.6pp private crowd-out association.

State Medicaid eligibility rules are complex, with nonlinear income schedules that differ by household structure in each year. As a result, the polynomial income controls in column 2 of Table 2 are unlikely to be enough in purging potential OVB from the OLS estimates. In column 3 we isolate the

[27] See Appendix Table A4 for the corresponding table with 2015 outcomes.

[28] Note that the sum of Medicaid, private insurance, and other insurance coverage effects is not mechanically the overall coverage effect, since individuals can have multiple types of insurance coverage.

state-level policy variation by constructing CG-style simulated instruments for $x_{\ell t}$. This $z_{\ell t}^{CG}$ predicts 2014 Medicaid eligibility as the average eligibility in a nationally-representative sample of individuals if they were to reside in the state of individual ℓ in year t. Thus variation in $z_{\ell t}^{CG}$ is purely due to the variation in eligibility policy (including a state's decision to expand eligibility under the ACA, as well as other variation in the eligibility thresholds).[29]

Column 3 reports IV estimates from instrumenting $x_{\ell t}\tau_t$ in equation (14) with $z_{\ell t}^{CG}\tau_t$, retaining the basic controls from column 1. While the overall coverage effect remains similar, at 4.2pp, the simulated eligibility instrument suggests both higher take-up (10.4pp) and private crowd-out (-5.8pp) effects, with again no effect on other insurance coverage. The column also shows that first stage relationship between actual and policy-predicted eligibility is close to one, as is often found in simulated eligibility instrument designs (Frean et al., 2017).

Column 4 of Table 2 isolates just the plausibly exogenous variation in state Medicaid eligibility in this period by using as an instrument the decision to expand eligibility in the state an individual resides in. This can be viewed as a simulated eligibility instrument which takes on two values: the average eligibility of a representative population when residing in expansion and non-expansion states. The estimates in this column are quite similar to those in column 3. When viewing expansion decisions as quasi-experimental, this suggests the OVB from additional policy variation in column 3 is likely to be minimal.

In columns 5-7 of Table 2 we show how replacing $z_{\ell t}^{CG}$ with a SEIV can increase the precision of Medicaid take-up and crowd-out effects, while continuing to leverage quasi-experimental variation in state expansion decisions. Specifically, we construct $z_{\ell t}^{SEIV}$ as the predicted 2014 eligibility of individual ℓ based on her household income, family structure, and state of residence in time t. The SEIV vector w contains information on state eligibility rules as of 2013, while the shock vector g contains indicators for whether each state n expanded eligibility in 2014. Thus, $z_{\ell t}^{SEIV}$ equals one if individual ℓ would have been eligible for Medicaid in 2013 when her state didn't expand coverage in 2014, while it equals one if individual ℓ's family income is less than the maximum of the ACA-mandated 138 FPL threshold and her state's 2013 thresholds when her state did expand coverage. The expected instrument $\mu_{\ell t}$ for this SEIV under the assumption of exchangeable state expansion decisions thus has a particular structure: it equals one (zero) for any individual with income low (high) enough to always (never) be Medicaid-eligible, regardless of state expansion decisions, and equals the probability of Medicaid expansion for one's state at intermediate income levels and household structures. As a result the recentered $z_{\ell t}^{SEIV}$ is mechanically zero for individuals who are always and never eligible for Medicaid across different realizations of g.

Column 5 finds broadly similar coverage, take-up, and crowd-out point estimates as column 4 from replacing the expansion indicator instrument with the recentered $z_{\ell t}^{SEIV}$. These estimates are also

[29] Some states that expanded under the ACA adopted a Medicaid eligibility threshold above 138 FPL. Thus conditional on a state's expansion decision eligibility policy differs in both the expanded and non-expanded group.

similar in column 6, which controls for expected eligibility while instrumenting for actual eligibility by the column 3 instrument. Consistent with the above theoretical discussion, however, these estimates are more precise (and the first stage is again close to one). The conventional state-clustered standard errors in column 6, which are in this case appropriate for the state-clustered SEIV, are 20-33% smaller than the corresponding standard errors in columns 3 and 4. Further precision gains, with 40-50% smaller standard errors than in columns 3-4, are found in column 7 which restricts estimation to the instrument "sharp sample"—individuals for which there is variation in the recentered $z_{\ell t}^{SEIV}$.

The estimates in Table 2 are suggestive of power gains from applying the general SEIV framework. To more formally study these gains, we conduct a Monte Carlo simulation of the relative power of estimates obtained via the state expansion instrument, recentered SEIV, and sharp sample SEIV. These simulations are calibrated to the estimates in column 5 of Table 2: for both the takeup (has Medicaid) and crowd-out (has private) analysis, we generate outcomes from applying the estimated second-stage coefficients to simulated Medicaid eligibility across 500 permutations of realized state expansion decisions, holding fixed all controls and the residuals estimated in Table 2. In each simulation draw we then apply each of the three IV estimators and test both the true null hypothesis and a range of alternatives at the 95% confidence level, using conventional state-clustered standard errors.

Figure 6 plots simulated power curves for each of the three IV estimators and two outcomes. These confirm substantial power gains from the recentered and sharp sample SEIV approach, relative to the simulated eligibility instrument that just uses variation in state expansion decisions. For the null of no crowd-out effect, for example, both recentered and sharp sample SEIV estimators have rejection rates of essentially one, while the expansion IV rejects in only 60% of simulations. Interestingly the simulations also reveal size distortions in the expansion IV and recentered SEIV approach for the crowd-out analysis; the sharp sample SEIV which is most powerful in this case also has correct size, rejecting the true null in 5% of simulations.

Appendix Tables A2 and A3 illustrate additional theoretical results from the above discussion related to the validity of $z_{\ell t}^{SEIV}$.[30] Table A2 reports pre-trend estimates for simulated eligibility and shock-exposure instruments in columns 1-2, and in column 3 replaces actual household income with a prediction of household income based on individual demographics (see Appendix B.2 for details). These come from versions of equation (14) that replace 2014 outcomes with 2012 outcomes, while still using 2014 to define eligibility. Most coverage pre-trends are small and statistically insignificant across all instruments, lending credibility to the quasi-experimental assignment of state policies. Notably the only significant pre-trend on Medicaid coverage is eliminated by using predicted income in column 3, while the recentered sharp-sample estimates in column 4 of Table 3 which uses this SEIV are quite similar to those in column 6 of Table 2.

Table A3 shows how the simple assignment process assumption of as-good-as-random ACA ex-

[30] Appendix tables A5 and A6 show corresponding results for 2015 outcomes.

pansion decisions across states can be tested and relaxed in our framework. Panel A shows SEIV estimates obtained under different models for the distribution of g given w, estimated by including additional state covariates in w and estimating populated-weighted state-level probit models. These models imply different $\mu_{\ell t}$ among individuals in the sharp sample, which we use to recenter $z_{\ell t}^{SEIV}$. Panel B tests each assignment process assumption by the two restrictions discussed in Section 3.3: that $\mathbb{E}\left[\tilde{z}_{\ell t}^{SEIV}\right] = 0$ and $\mathbb{E}\left[\tilde{z}_{\ell t}^{SEIV}\mu_{\ell t}\right] = 0$. Joint chi-squared test statistics decline as we move from the simple model of exchangeable expansion decisions to allow for the probability of expansion to depend on whether or not the state had a Republican governor in 2013 (column 2), the state's median income in 2012 (column 3), the state's Medicaid enrollment rate in 2012 (column 4), or all covariates simultaneously (column 5), though all tests fail to reject at conventional levels. At the same time, the coefficient estimates in Panel A remain stable with little increase in state-clustered standard errors.

5.3 Shift-Share Instruments

Shift-share instruments average a set of observed shocks with observation-specific shares. In a typical case, a regional instrument z_ℓ is constructed from industry shocks g_n by $z_\ell = \sum_n w_{\ell n} g_n$, where $w_{\ell n}$ measures the industry's share (of, say, employment) in the region. We assume that the shocks are quasi-randomly assigned, whereas the shares measuring regional shock exposure may be correlated with the structural residual, for example because unobserved industry shocks also affect regions via the same shares. Our discussion of shift-share instruments builds on and extends our earlier work (Borusyak et al., 2019, henceforth BHJ); to keep its presentation short, we refer the reader to BHJ for more details and examples of the setup.[31]

Viewing shift-share instruments as SEIVs delivers two main new insights regarding inference and power. For the former, randomization inference provides a new way of constructing confidence intervals that complements the asymptotic standard errors of Adão et al. (2019b) and BHJ. Since randomization inference is exact in finite samples, it is especially valuable in situations when the asymptotic approximation may not be precise, e.g. with relatively few or heavy-tailed shocks.[32] Moreover, in the application below we provide evidence that the power properties of randomization inference dominate those of asymptotic inference, especially with few or heavy-tailed shocks.

Despite good power properties, randomization inference is not meant to replace asymptotic inference because stronger assumptions have to be imposed on the shocks for RI to be valid: exchangeability and full independence from all shares and residuals, compared with a weaker shock orthogonality condition in BHJ. In particular, exchangeability requires homoskedasticity of shocks and is thus violated

[31] The alternative approach of Goldsmith-Pinkham et al. (2019), in which shares are exogenous but shocks need not be, also fits within our SEIV framework, with an exchangeable matrix of shares as g and the vector of shocks as w. The shock-exposure view however does not bring new insights in this case.

[32] Of course, if there are the number of shocks (or the "effective number of shocks" per BHJ) is too small, the quasi-experiment may not have enough power to reject interesting economic hypotheses. However, it is difficult to judge *a priori* what constitutes "too small," and exact confidence intervals provide a valid data-driven answer.

when industry-level shocks are less volatile (perhaps because measured more precisely) for bigger industries; BHJ find this to be the case in the Autor et al. (2013) application. It is straightforward to allow heteroskedasticity of known functional form but even that is restrictive.

However, this limitation can be remedied by using robust randomization inference. Since asymptotic standard errors that allow for heteroskedasticity or clustering in shocks are available from Adão et al. (2019b) and BHJ, they can be used to studentize the test statistic when performing randomization inference, as described in Section 3.3. In this case, the resulting confidence intervals will be asymptotically robust to heteroskedasticity and clustering, while still being exact in finite samples under the stronger RI conditions. That said, we find the power performance of confidence intervals that studentize via heteroskedasticity-robust standard errors to be poor with few shocks (details are available upon request). We also note that both RI and asymptotic confidence intervals depend on the assumption of constant treatment effects; comparing robustness of these methods to treatment effect heterogeneity is left for future research.

Our second insight on the use of SEIVs for shift-share instruments is that the instrument power can be improved by using nonlinear instruments derived from economic theory, while preserving robustness to model misspecification. Kovak (2013) and Adão et al. (2019b) justify shift-share instruments as first-order approximations of certain models of trade. Our results provide a framework for using nonlinear f_ℓ and thus avoiding such approximations. If the shocks are large (as they are in many applications) and the approximation is not precise, a feasible approximation for the optimal instrument from Section 4.2 may deliver a substantial improvement in power. These assumptions are already made (although in a weaker form) when the conventional shift-share instrument is used.

We demonstrate this point in Appendix A.6, by extending the model of Kovak (2013) to allow for imperfectly elastic regional labor supply. We then show that if a quasi-random industry demand shock (e.g., from a trade liberalization) is observed, the model yields an intuitive *nonlinear* shock-exposure instrument for regional employment in the labor supply equation. We then characterize its power advantage compared to the shift-share approximation. Importantly, the recentered shock-exposure instrument remains valid even if the economic model is misspecified: it only relies on the second-stage exclusion restriction and knowledge of the assignment process.

As a final note, we show how one result from BHJ can be derived and understood differently within the shock-exposure framework. BHJ show that when the expected shock is a linear function of industry-level observables q_n, controlling for a share-weighted average of these observables isolates the residual variation in the shocks. That is when $\mathbb{E}[g_n \mid w] = \tau' q_n$ it is enough to control for $\sum_n w_{\ell n} q_n$ in the shift-share regression. In our framework this follows from the fact that the expected instrument is linear in $\sum_n w_{\ell n} q_n$: $\mu_\ell(w) = \mathbb{E}[\sum_n w_{\ell n} g_n \mid w] = \tau' \sum_n w_{\ell n} q_n$. Thus by the equivalence noted in Section 4.1, controlling for $\sum_n w_{\ell n} q_n$ asymptotically yields the expected instrument adjustment.

Application We compare the finite-sample properties of different inference procedures for shift-share IV via Monte-Carlo simulations, as analytical results are difficult to obtain. The simulations are based on the Monte-Carlo analyses in BHJ (their Appendix A.10) and like them use the data from the influential study of Autor et al. (2013), who estimate the effects of import competition with China on U.S. local labor markets. In each simulation we draw 1,500 sets of *iid* industry shocks from some distribution and construct corresponding regional datasets of outcomes and treatments. With the true causal effect $\beta = 0$, we test hypotheses $\beta = \beta_0$ for various levels of β_0 (rescaled so that the IV estimate in the actual data corresponds to $\beta_0 = 1$). The size of the test is obtained for $\beta_0 = 0$, while other values are informative on the power of the tests.

We compare three testing procedures. Two of them rely on the asymptotic results the BHJ: the primary "exposure-robust" inference and its version that imposes the null hypothesis (Adão et al. (2019b) show that this results in better coverage in finite samples). The third procedure is randomization inference based on 1,999 random permutations of the shocks, where the test statistic is the least squares coefficient from the reduced-form regression of the outcome on the shift-share instrument, across regions-by-period observations. To compare power we need to make size comparable; we thus pick nominal coverage differently for each procedure such that the actual finite-sample coverage is always 5%. See Appendix B.3 for data and implementation details.

Figure 7 presents the main findings for three different data-generating processes. Panel A draws industry shocks from a normal distribution for 794 industry-by-period shocks. Consistent with BHJ, exposure-robust inference leads to slight overrejection, as 3.1% nominal coverage implies 5% actual coverage; the other procedures exhibit no significant size distortions. Power of all procedures is also very similar, although slightly larger for randomization inference.

These patterns change drastically when we deviate from many normal shocks, however. Panel B of Figure 7 aggregates industries into SIC2 groups, resulting in 40 industry-by-period shocks, while retaining shock normality. Here overrejection of exposure-robust inference becomes substantial (0.7% nominal rejection corresponds to 5% effective), and randomization power has clearly better power against almost all $\beta_0 \neq 0$. The advantages of randomization inference are even larger in Panel C which keeps 794 shocks (as in Panel A) but makes shocks heavy-tailed, drawing them from the empirical distribution of the Autor et al. (2013) shocks (demeaned within periods). In this simulation exposure-robust inference has even stronger size distortion (0.6% nominal size for 5% effective) and its power is moderately smaller than that of randomization inference for many values of β_0, although mildly higher for some others. Asymptotic inference with the null imposed has only minor size distortions (6.9% nominal for 5% effective) but performs very poorly on power. The power of this procedure remains under 80% for arbitrarily large values of β_0; that is, in over 20% of simulated samples the implied confidence interval is infinite.

5.4 Nonlinear Shift-Share Constructions

A conventional shift-share instrument, defined as $\sum_n w_{\ell n} g_n$, is linear in the shocks. However, a number of recent papers have used nonlinear transformations of the instrument, $z_\ell = h\left(\sum_n w_{\ell n} g_n\right)$. Here we show that even if the untransformed $\sum_n w_{\ell n} g_n$ is valid, z_ℓ need not be and will generally need to be recentered via knowledge of the shock assignment process. The formal argument is straightforward: the expected instrument $\mu_\ell = \mathbb{E}\left[h\left(\sum_n w_{\ell n} g_n\right) \mid w\right]$ may vary across ℓ even if $\mathbb{E}\left[\sum_n w_{\ell n} g_n \mid w\right]$ does not. The latter corresponds to the case considered by BHJ when the exposure shares sum to one, i.e. $\sum_n w_{\ell n} = 1$ for each ℓ, and the shocks are assigned with the same mean $\mathbb{E}\left[g_n \mid w\right] = \alpha$.

To build intuition for the validity problem with nonlinear $h(\cdot)$, first consider the case where the $h(\cdot)$ mapping is concave. Then by Jensen inequality the expectation of z_ℓ is higher when the underlying shift-share term has lower variance, which generally corresponds to observations with more dispersed exposure shares $w_{\ell n}$.[33] Similarly, for a convex $h(\cdot)$ the expected instrument is higher for observations with exposure shares concentrated on a small number of shocks. Observations with dispersed shares may further have systematically different unobservables: for example, regions with diversified economies (in terms of employment across industries) may be more resilient to unobserved regional shocks. This generates omitted variables bias for the nonlinear shift-share instrument, with the direction that depends on the convexity of $h(\cdot)$.

A popular nonlinear shift-share construction is found in panel regressions with fixed effects. Berman et al. (2015), for example, construct an instrument for firm ℓ in period t as

$$z_{\ell t} = \ln\left(\sum_n w_{\ell n} G_{nt}\right), \tag{15}$$

with similar strategies used recently by Berthou et al. (2019) and Costa et al. (2019). Here $w_{\ell n}$ measures the time-invariant share of firm exports to some market n (defined by an industry and destination country), and $G_{nt} > 0$ proxies for the level of demand in that market in period t. This instrument builds on the intuition that panel regressions estimated in logs and with unit fixed effects effectively relies on randomness in the growth rates of the shifters, $\Delta \ln G_{nt}$, but this intuition is erroneous in this case.

For exposition clarity, suppose there are only two periods $t = 0, 1$ and that the regression may equivalently be estimated on first-differenced data with the instrument

$$\Delta z_\ell = \ln \frac{\sum_n w_{\ell n} G_{n1}}{\sum_n w_{\ell n} G_{n0}} = \ln \sum_n \tilde{w}_{\ell n} g_n, \tag{16}$$

where the shares $\tilde{w}_{\ell n} = w_{\ell n} G_{n0} / \sum_{n'} w_{\ell n'} G_{n'0}$ add up to one and the shocks are $g_n = \frac{G_{n1}}{G_{n0}}$. The

[33] This argument is precise for quadratic mappings, $h(x) = x - bx^2$ for $b > 0$, *iid* shocks with some mean α and variance σ^2, and shares adding up to one. Then $\mu_\ell(w) = \tilde{\alpha} - b\sigma^2 \cdot \sum_n w_{\ell n}^2$, for constant $\tilde{\alpha} = \alpha - b\alpha^2$. The expected instrument is higher when the Herfindahl index $\sum_n w_{\ell n}^2$ is lower, i.e. when the shares are dispersed.

nonlinearity concern directly applies, despite the panel dimension of the data: even if the shocks are randomly assigned, the instrument is systematically higher for observations with dispersed $\tilde{w}_{\ell n}$. Moreover, an additional problem is evident: the shares effectively used by the instrument are $\tilde{w}_{\ell n}$ rather than $w_{\ell n}$, which may not always align with the economic intuition behind the instrument.[34]

Both problems are easily avoided by an alternative construction of the instrument. Consider $\check{z}_{\ell t} = \sum_n w_{\ell n} \ln G_{nt}$, which is a conventional shift-share instrument. With two periods, its first difference equals $\Delta \check{z}_\ell = \sum_n w_{\ell n} \ln \frac{G_{n1}}{G_{n0}}$, which again is a shift-share with correct shares $w_{\ell n}$ and shocks which are the growth rates of G_{nt}. This intuition applies more generally: BHJ prove (Section 3.3) that the panel regression with unit fixed effects and instrument $\check{z}_{\ell t}$ can be viewed as leveraging variation in the growth rates of shocks regardless of the number of time periods.

We finally note two other recent examples of nonlinear shift-share constructions. Derenoncourt (2019) uses sample percentiles of a shift-share variable as instrument. This transformation is generally neither convex nor concave, so the simple intuition about share dispersion does not apply. Still, it is nonlinear and thus as-good-as-randomness of shifters does not suffice for validity.

Boustan et al. (2013) use a more complex strategy which goes beyond the $h(\cdot)$ transformation. To instrument for the Gini coefficient corresponding to the income distribution in a region, they construct a predicted income distribution and take its Gini coefficient as the instrument. This prediction uses a shift-share approach, based on the lagged local income distribution (as shares) and national changes in income by income groups (as shocks). The resulting instrument can thus be represented as a nonlinear function of the shares and shocks, i.e. as a general SEIV. Recentering is necessary for the instrument to solely leverage variation in the national income shocks.

5.5 Network Instruments

The shock-exposure setup is well-suited to studying propagation and spillover effects of random shocks through predetermined networks. This section maps identification strategies used in that literature to SEIVs, with an example from Carvalho et al. (2016), and shows how the expected instrument captures the relevant notion of the network position that should be controlled for.

Suppose the researcher specifies a model of spillovers

$$y_\ell = \sum_{k=1}^{K} \beta_k x_{k\ell} + \varepsilon_\ell, \tag{17}$$

where $x_{\ell 1}, \ldots, x_{\ell K}$ capture $K \geq 1$ possible ways how observations (nodes) ℓ can be affected by some set of shocks. This equation is estimated with SEIVs $z_{\ell 1}, \ldots, z_{\ell K}$ which are functions of the shock realizations g that may specify the nodes or links that were affected and by how much, and of the

[34] Approximating Δz_ℓ around $G_{n1} = G_{n0}$ (for small time changes in G_{nt}) one obtains $\Delta z_\ell \approx \sum_n \tilde{w}_{\ell n} \left(\frac{G_{n1}}{G_{n0}} - 1 \right)$: a shift-share instrument based on the growth rates of G_{nt} with shares $\tilde{w}_{\ell n}$.

pre-determined network itself. As usual we assume shocks are exogenous, i.e. satisfy Assumption 1.

For concreteness, consider a stylized version of the Carvalho et al. (2016) setup that studies propagation of the Fukushima earthquake through the network of trading firms. Let y_ℓ denote a firm outcome (e.g. revenue growth) and $x_{k\ell}$ be a set of mutually exclusive dummies for different values of the network distance from the firm to the earthquake, d_ℓ. Specifically, $x_{1\ell}$ switches on if firm ℓ was directly affected by the earthquake ($d_\ell = 1$), $x_{2\ell}$ if it was not affected but had at least one affected supplier ($d_\ell = 2$), $x_{3\ell}$ if the firm instead had at least one supplier whose supplier was hit, and so on.[35] As in Carvalho et al. (2016), suppose for simplicity (5.5) is estimated by OLS, such that $z_{k\ell} = x_{k\ell}$. These instruments qualify as SEIVs, as they can be computed from two pieces of information: the adjacency matrix of the network w and the location of the earthquake, defined as the set of directly affected firms, g.[36]

The OVB problem associated with this strategy derives from differences in network position across firms. The OLS estimator compares firms at different distances from the network, with an omitted category of firms whose supply chains are totally unaffected serving as the control group. However, this comparison need not be causal even if the location of the earthquake is random. In particular, firms which have more suppliers are more likely to have at least one of them hit by the earthquake.[37] The number of connections may be correlated with the residual: well-connected firms are typically larger and they may be more exposed to unobserved shocks in the network or, conversely, more robust to them. In all of these cases, the estimates of β_k will be biased. Heterogeneity by the number of connections is not the only source of the problem. For example, firms that have few direct suppliers but many second-degree suppliers (i.e., their suppliers are very well-connected) are also likely to be close to the earthquake.

The relevant notion of the network position depends on the structure (i.e., the assignment process) of earthquakes. If an earthquake is a shock that hits a random region, then a firm with two suppliers from different regions is systematically more exposed to it than the one with ten suppliers but all located in the same place. If earthquakes concentrate in the north of the country, then firms with supply chains in the north are also more exposed. Conventional measures of network centrality are thus unlikely to capture all systematic variation in the instrument. At the same time, knowledge of the distribution of earthquake locations and structure, perhaps drawn from geological studies, can be readily exploited to measure $\mu_{1\ell}, \ldots, \mu_{K\ell}$—simulated probabilities that firm ℓ is at distance k from a random earthquake. Recentered instruments $\tilde{z}_{k\ell} = z_{k\ell} - \mu_{k\ell}$ can then be used for valid estimation.[38]

[35] More treatments can be specified, for instance allowing for spillover effects from affected customers too. Given the set of included treatments, shock exogeneity applied to (17) requires that no types of spillovers have been omitted.

[36] Market access instruments from Section 5.1 can be viewed as a special case of this discussion too, with the transportation network serving as w.

[37] In line with this view, Carvalho et al. (2016) control for the number of network connections. The following discussion suggests that this is not sufficient for eliminating OVB.

[38] So far we have assumed that spillovers depend on the presence of at least one supplier at a certain network distance. Other work has adopted a more continuous formulation where $x_{k\ell}$ is the share of degree-k suppliers (e.g., by input value) affected by the earthquake. Expected shares also need not be equal across firms, for example if the earthquake

With the earthquake distribution specified, randomization inference provides exact confidence intervals for β. Validity in finite samples is especially useful in this setting for two reasons. First, conventional methods for inference in network situations are difficult and often require strong assumptions. Second, with only one earthquake observed in the data, the asymptotic approach may generally be inapplicable. Even absent spillovers, when only one region is hit by the earthquake it is impossible to get a consistent estimate for β (e.g., in the asymptotic sequence where the number of regions is growing). The causal impact of an earthquake in this scenario cannot be precisely isolated from the effects of unobserved random shocks that hit the same region. The lack of consistency, however, does not preclude informative inference on β_k. For example, if the true causal effects are zero, it is unlikely that unobserved shocks hit exactly the same region where the earthquake randomly happened. Randomization inference-based confidence intervals capture this idea formally.[39]

We finally give another example of a network situation in which the SEIV approach may be useful. Jaravel et al. (2018) study the effects of an inventor's unexpected death on the future productivity of her co-inventors. Recognizing that collaborators of randomly deceased inventors are not representative of the full inventor population (e.g., due to their age distribution and the "friendship paradox"), they construct a control group which matches deceased inventors to "placebo" ones based on observables and look at their surviving collaborators. This strategy implicitly requires deaths to be rare events: if a typical inventor had both deceased and surviving collaborators, the treatment and control groups would largely overlap. However, such overlap can be straightforwardly overcome in the SEIV framework. If, for example, spillovers on surviving inventors are assumed to depend on the share of collaborators who have died, one can recenter by the expected value of this share via redrawing from matched pairs of deceased and placebo inventors. Similarly, while Jaravel et al.'s (2018) clustering of observations by the deceased (or placebo-deceased) inventor relies on the two groups of surviving inventors not overlapping, randomization inference does not require such assumption.

5.6 Other Designs

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

Adão et al. (2019a) develop a spatial general equilibrium model that implies tractable estimating equations for the structural parameters. Then they derive optimal instruments for the endogenous regressors, using the structure of the model. In constructing instruments, Adão et al. (2019a) make two key assumptions. First, they assume that the structural residuals (e.g., unobserved supply-

probabilities vary across space.

[39] We thank Andres Rodriguez-Clare for the conversation on this issue.

shifters) across U.S. local labor markets are mean-zero conditional on both the set of observed shocks g (e.g., China productivity growth across industries, similarly to Autor et al. (2013)) and the original equilibrium w; in our notation, $\mathbb{E}[\varepsilon_\ell \mid g, w] = 0$. Here w includes, for example, regional shares of different industries in a previous period, which are implicitly assumed to be valid instruments (as in Goldsmith-Pinkham et al. (2019)). Our framework clarifies that this can be replaced by a weaker assumption, $\mathbb{E}[\varepsilon_\ell \mid g, w] = \mathbb{E}[\varepsilon_\ell \mid w]$, requiring only exogenous changes in Chinese productivity in Adão et al. (2019a). Permuting observed shocks across industries then yields the expected instrument adjustment for any $z_\ell = f_\ell(g, w)$, with Proposition 4 characterizing the optimal instrument among the recentered IV estimators, which generally does not coincide with the one in Adão et al. (2019a).

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

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

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

The latter might arise when applicants with similar preferences and priorities are similar in other unobserved ways, or because of peer effects.

6 Conclusions

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

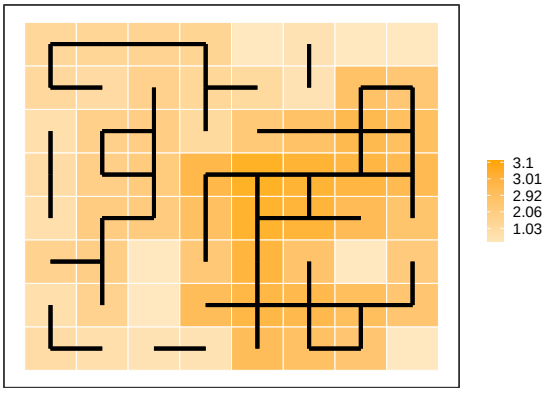
We conclude by noting two important limitations of our current analysis. First, while we characterize the shock-exposure IV estimand under general treatment effect heterogeneity in Appendix A.3, our approach to inference and efficiency is fundamentally based on an assumption of constant effects. That is, the randomization tests we discuss in Section 3.3 are guaranteed to have correct size in testing the strong null hypothesis that treatment has no effect on outcomes for any observations, but not the weaker hypothesis that the average effect is zero. In general exact inference for such null hypotheses is a difficult challenge, though recent progress has been made for ensuring robustness of randomization inference to weaker null hypotheses in much simpler settings (Chung and Romano, 2013). In more complex designs, such as shift-share IV, inference robust to heterogenous treatment effects is challenging even with the traditional asymptotic approach (Adão et al., 2019b).

A second assumption we maintain throughout is *a priori* knowledge of the quasi-experimental design, or shock assignment process. It is enough in our setting to know the vector of shocks is exchangeable, such that different permutations of the observed vector (perhaps within observable groups) are equally likely to arise. As discussed in the shift-share example of Section 5, it is also enough to know the expected instrument is linear in observables. It is not, however, typically sufficient for shocks to be *iid* conditional on observables with an unknown distribution, or for shocks to be clustered by observable groups. In some settings shocks may themselves be estimated from observed data, inducing a complex distribution of the feasible instrument (Borusyak et al., 2019). Characterizing the properties of shock-exposure IVs that use estimated shock assignment processes appears a fruitful area of future study.

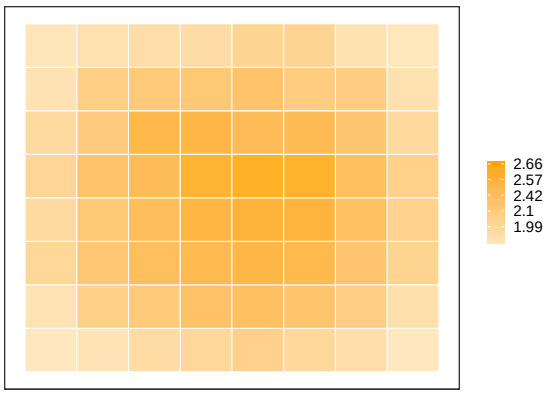
Figures and Tables

Figure 1: Simulated Log Market Access Growth

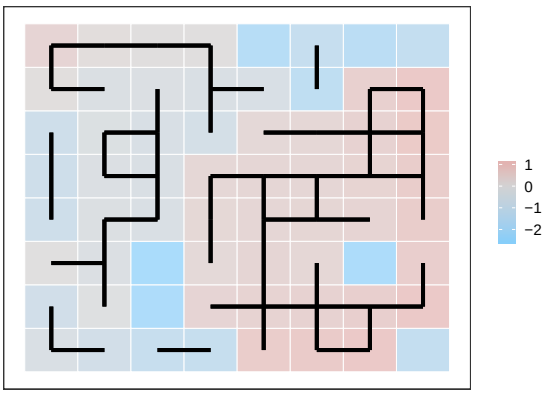
A. Connectivity and Market Access



B. Expected Log Market Access Growth



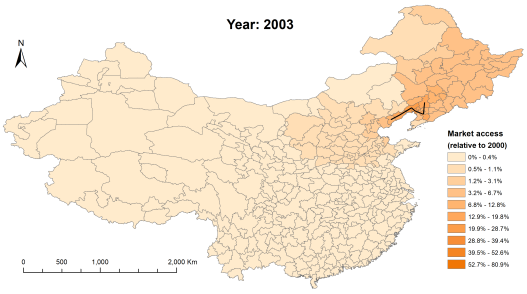
C. Recentered Log Market Access Growth



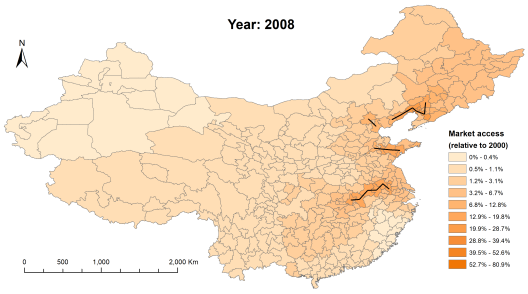
Notes: This figure illustrates the omitted variables bias problem for a simple shock-exposure instrument discussed in Section 2. Panel A shows a random draw of the railroad construction experiment, with lines indicating connected regions and shading indicating implied log market access growth. Panel B shows average log market access growth over many such random draws. The shading in Panel C indicates the recentered log market access instrument which subtracts the expected instrument in Panel B from the realized instrument in Panel A, with the lines again indicating realized connectivity.

Figure 2: Chinese High Speed Rail and Log Market Access Growth

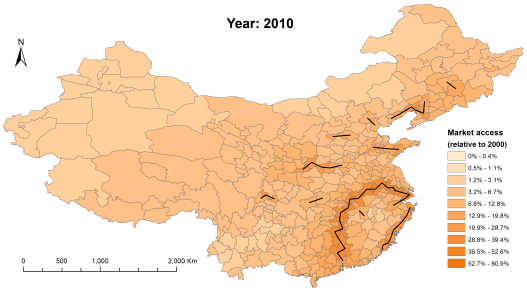
A. Lines Completed by 2003



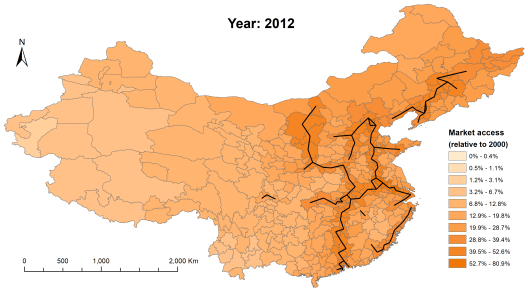
B. Lines Completed by 2008



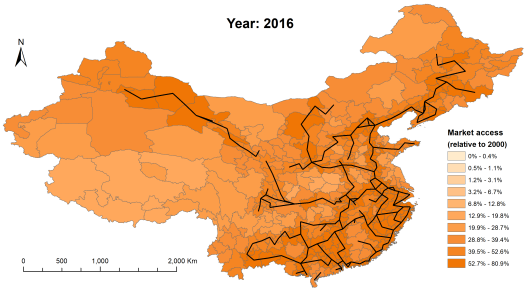
C. Lines Completed by 2010



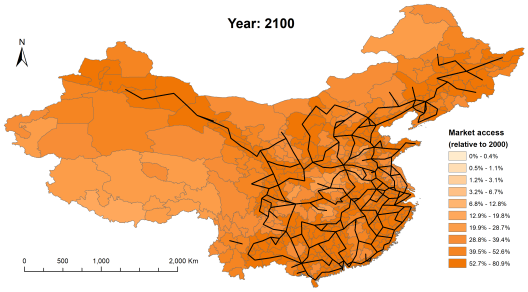
D. Lines Completed by 2012



E. Lines Completed by 2016



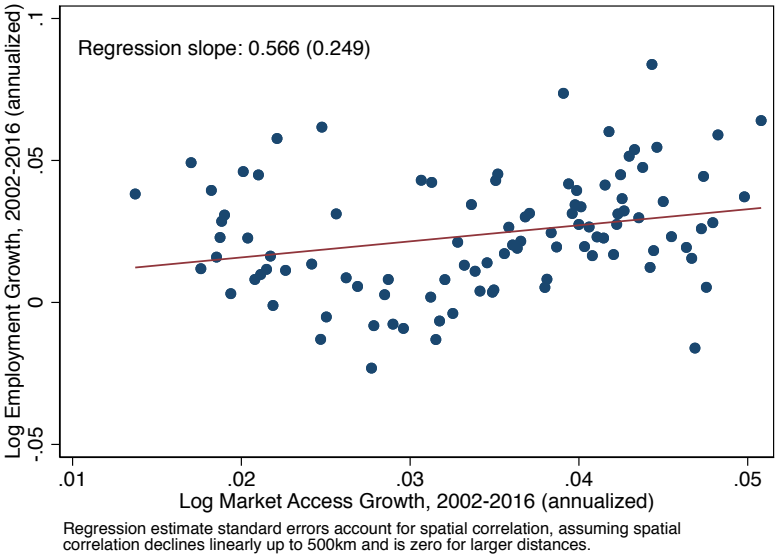
F. All Completed or Planned Lines



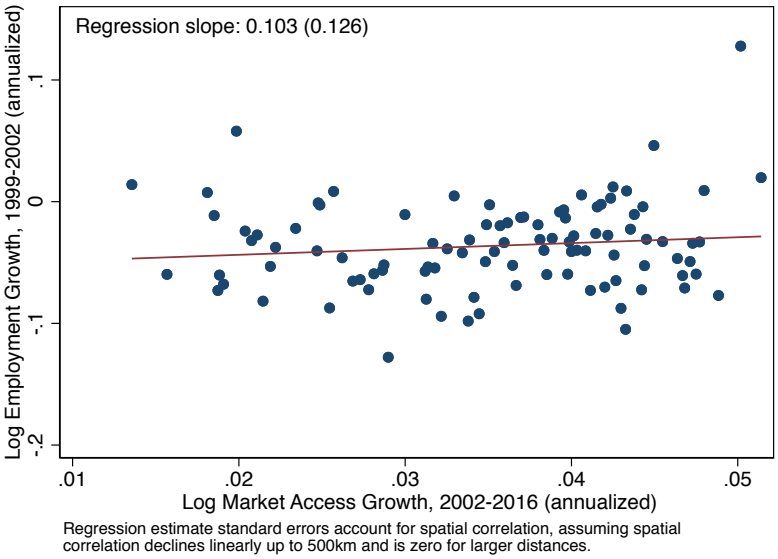
Notes: Panels A–E of this figure illustrate the growth in high-speed rail in China from 2002-2016. Panel F illustrates all completed or planned high-speed rail lines as of 2016. Shading in each panel indicates log market access growth since 2002, computed as described in Section 5.1.

Figure 3: Chinese Market Access Growth, Employment Growth, and Employment Pre-Trends

A. Log Market Access Growth and Employment Growth



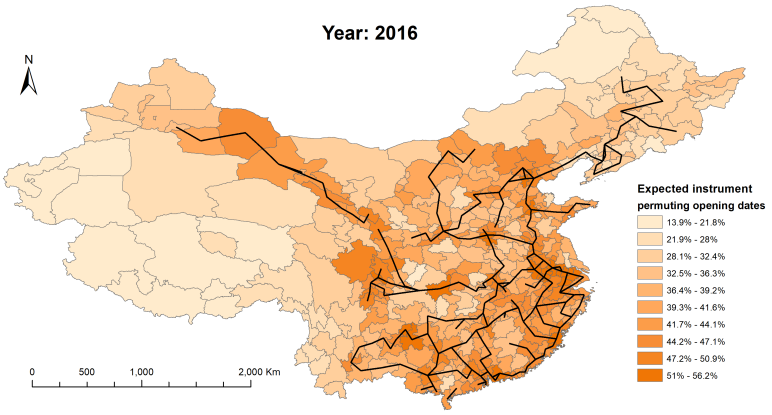
B. Log Market Access Growth and Employment Pre-Trends



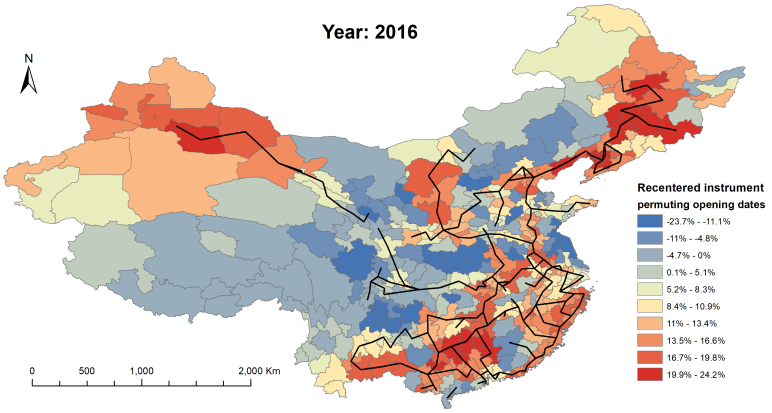
Notes: Panel A of this figure shows a binned scatterplot of annualized log employment growth across 306 cities in China, from 2002 to 2016, against annualized log market access growth in the same period. Panel B shows a binned scatterplot of annualized log employment growth prior to the expansion of high-speed rail, from 1999 to 2002, against annualized market access growth in 2002-2016. Bins are constructed as percentiles of log market access growth. Lines of best fit are indicated in red, with regression coefficients and spatial-clustered standard errors indicated in each panel.

Figure 4: Expected and Recentered Chinese Market Access Growth

A. Expected Log Market Access Growth



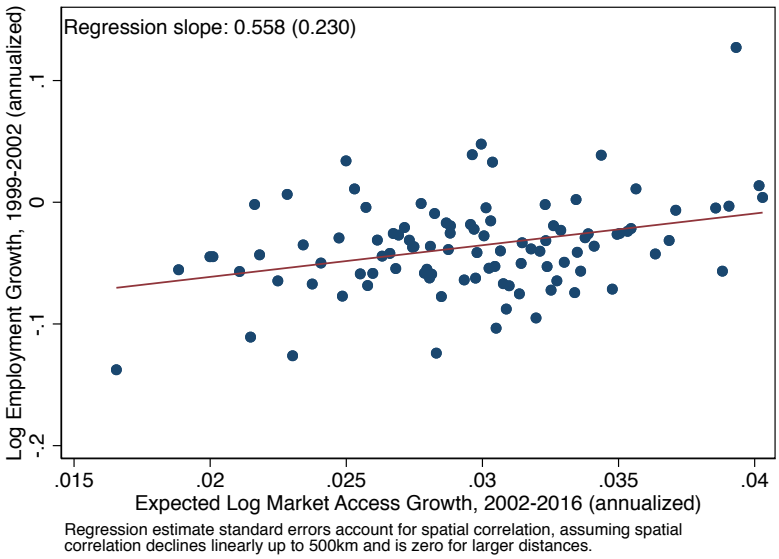
B. Recentered Log Market Access Growth



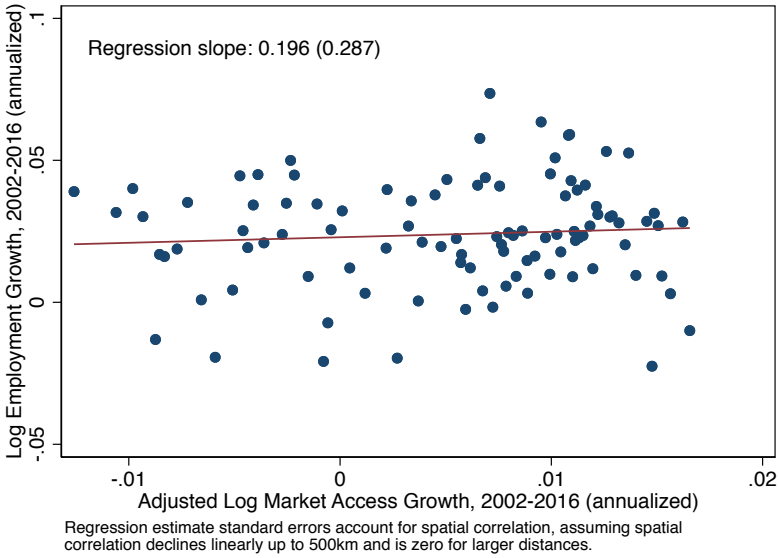
Notes: Panel A of this figure plots the variation in expected log market access growth from 2002 to 2016 in China, along with the high-speed rail lines constructed during this period. Panel B instead plots the variation in recentered log market access growth: the difference between the log market access growth shown in panel E of Figure 2 and panel A of this figure. Expected log market access is simulated by permuting line opening dates, as described in Section 5.1.

Figure 5: Expected and Recentered Market Access Growth and Employment Growth and Pre-Trends

A. Expected Log Market Access Growth and Employment Pre-Trends



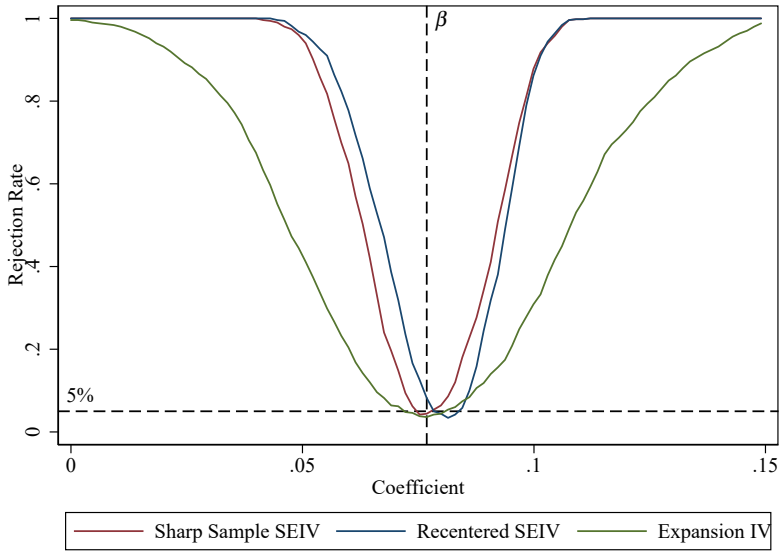
B. Recentered Log Market Access Growth and Employment Growth



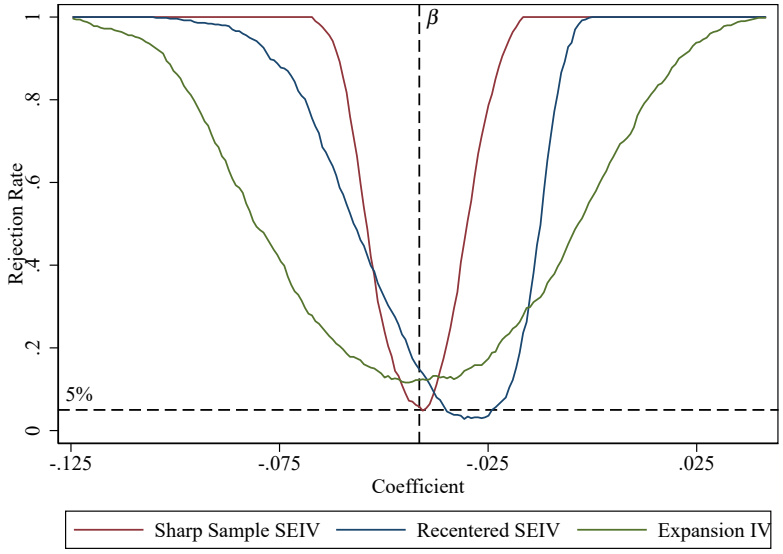
Notes: Panel A of this figure shows a binned scatterplot of annualized log employment growth across 306 cities in China, from 1999 to 2002, against annualized expected log market access growth in 2002 to 2016. Panel B shows a binned scatterplot of annualized log employment growth in 2002 to 2016 against annualized recentered log market access growth in the same period. Bins are constructed as percentiles of log market access growth. Lines of best fit are indicated in red, with regression coefficients and spatial-clustered standard errors indicated in each panel. Expected and recentered log market access are simulated by permuting line opening dates, as described in Section 5.1.

Figure 6: Simulated Coverage and Power of Alternative Medicaid Eligibility Instruments

A. Takeup (Outcome: Has Medicaid)

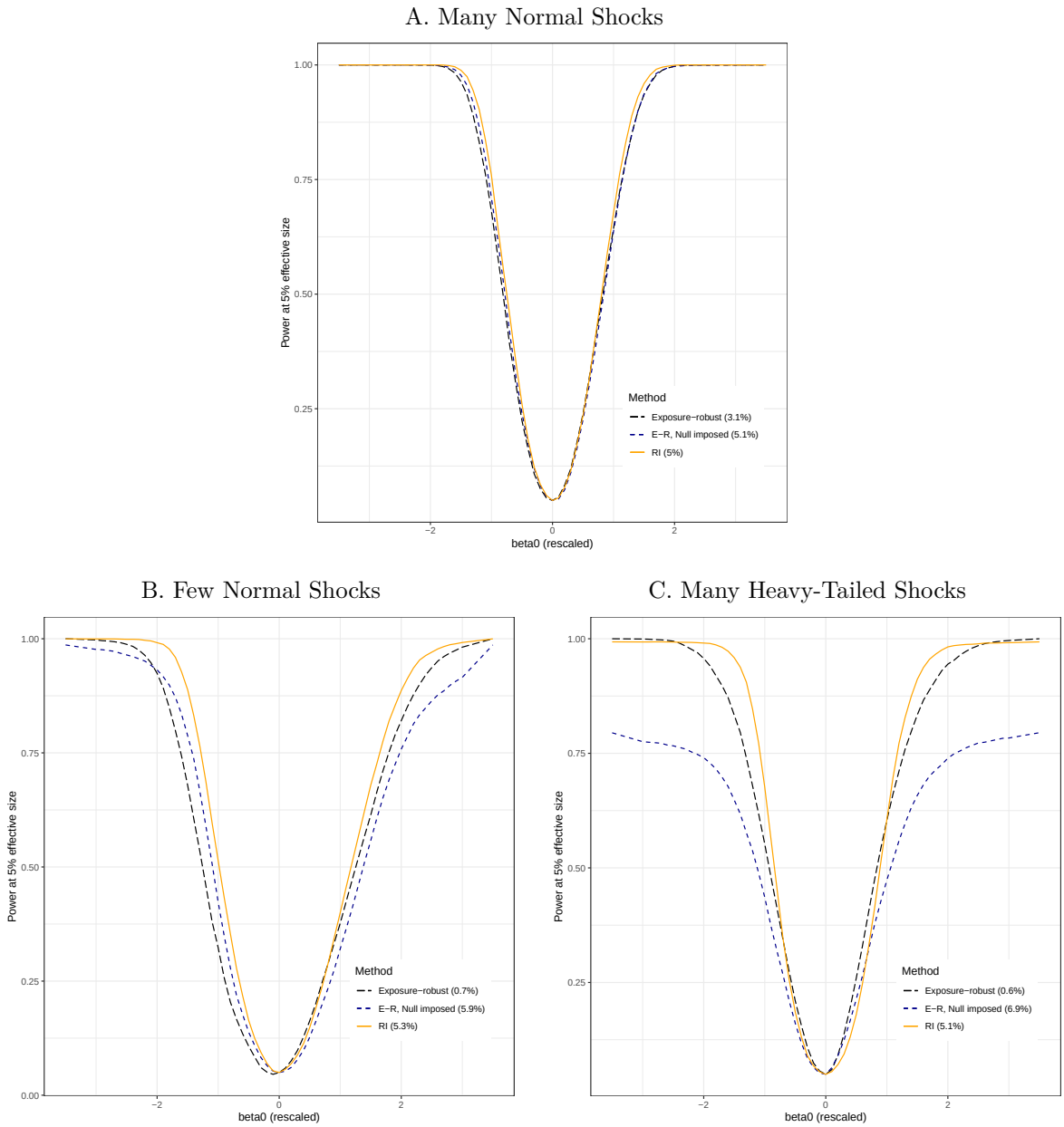


B. Crowd-out (Outcome: Has Private Insurance)



Notes: This figure plots simulated power curves for different valid instruments for Medicaid eligibility when 2014 expansion decisions are randomly assigned across states. The data-generating process comes from the sample and estimates in column 5 of Table 2, for the "has Medicaid" (panel A) and "has private" (panel B) outcomes, and 500 permutations of actual state expansion decisions. Tests are based on state-clustered asymptotic standard errors.

Figure 7: Simulated Coverage and Power of Alternative Shift-Share IV Inference Procedures



Notes: This figure plots simulated power curves for different shift-share IV inference procedures when few or heavy-tailed shocks are random assigned across industries. The data-generating process comes from Autor et al. (2013), as described in Section 5.3.

Table 1: Corrected Panel Estimates and Confidence Intervals - 50% Exclusion Threshold

	Unadjusted Instrument (1)	Permuted opening dates (2)	Permuted construction times (3)
	A. Subtracting off Expected Instrument		
Coefficient	0.249	0.088	−0.014
Conventional 95% CI	[0.062, 0.436]	[-0.101, 0.277]	[-0.249, 0.220]
Permutation 95% CI		[-0.409, 0.567]	[-0.487, 0.321]
	B. Controlling for Expected Instrument		
Coefficient		−0.001	−0.060
Conventional 95% CI		[-0.186, 0.185]	[-0.291, 0.171]
Permutation 95% CI		[-0.452, 0.378]	[-0.511, 0.247]
	C. Specification Tests		
Slope Coefficient		1.002	1.001
Test p-value		0.010	0.066
Mean Difference		0.036	0.012
Test p-value		0.010	0.068
Observations		5,537	5,537

Notes: Panels A and B of this table report OLS estimates of the effect of log market access on log employment in a panel of 308 Chinese cities over the years 2000 to 2016. Column 1 estimates this effect with city and year fixed effects; a 95% confidence interval based on spatial-clustered standard errors is reported in brackets. Columns 2 and 3 of panel A report corresponding estimates and confidence intervals from regressions on recentered log market access. Expected log market access in column 2 is simulated by permuting high speed rail opening dates, while expected log market access in column 3 is simulated by permuting rail construction times. Permutation-based 95% confidence intervals are also included. Panel B instead controls for expected log market access. Panel C reports estimates and permutation-based p-values for tests of correct specification in columns 2 and 3. Slope coefficients come from regressions of log market access on expected log market access, controlling for city and year fixed effects, while mean differences are the average recentered log market access. Test p-values are computed as described in the text.

Table 2: Conventional and Shock-Exposure IV Estimates of Medicaid Eligibility Effects

	Conventional Estimates				SEIV Estimates		
	OLS						
	No Controls	Income Controls	Simulated Instrument IV	Expansion Instrument IV	Recentered	Controlled	Sharp Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Is Insured	0.052***	0.021**	0.042**	0.050**	0.036	0.039**	0.059***
	(0.009)	(0.009)	(0.021)	(0.021)	(0.024)	(0.017)	(0.011)
Has Medicaid	0.056***	0.049***	0.104***	0.109***	0.077***	0.072***	0.085***
	(0.007)	(0.007)	(0.016)	(0.018)	(0.013)	(0.012)	(0.011)
Has Private	−0.002	−0.026***	−0.058***	−0.055***	−0.041**	−0.033***	−0.025***
	(0.004)	(0.005)	(0.016)	(0.016)	(0.018)	(0.010)	(0.006)
Has Other	0.003***	−0.001	0.003	0.003	−0.003	−0.001	0.001
	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
First Stage			0.954***	0.211***	0.765***	0.949***	0.969***
			(0.031)	(0.009)	(0.134)	(0.026)	(0.016)
Individuals	2,397,313	2,397,313	2,397,313	2,397,313	2,397,313	2,397,313	421,042
States	43	43	43	43	43	43	43

Notes: This table reports OLS and IV estimates of the effect of Medicaid eligibility in 2014 on different forms of insurance coverage in 2014. Column 1 estimates a difference-in-differences regression of coverage in 2013 and 2014 on 2014 eligibility, interacted with year, controlling for state and year fixed effects. Column 2 further controls for a fourth-order polynomial in household income (as a percentage of the federal poverty level) interacted with year. Column 3 estimates difference-in-differences IV regressions with a simulated instrument, interacted with year, predicting 2014 eligibility from an individual's state with a nationally representative sample of individuals in 2014. This specification also controls for state and year fixed effects and predicted 2014 eligibility main effects. Columns 5-7 estimate difference-in-differences IV regressions with a shock-exposure instrument, interacted with year, predicting 2014 eligibility from an individual's state's Medicaid expansion decision as described in the text. These specifications also control for state and year fixed effects and predicted 2014 eligibility main effects. The instrument in Column 5 adjusts for an individual's expected eligibility over the state expansion shocks. Column 6 controls for the shock-exposure instrument. Column 7 restricts estimation to observations with a non-degenerate adjusted instrument. Estimates come from a nationally representative sample of individuals in 43 states that did not expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2019): "Sampling-based vs. Design-based Uncertainty in Regression Analysis," *Working Paper*.

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

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

——— (2019): "Breaking Ties: Regression Discontinuity Design Meets Market Design," *Working Paper*.

Acemoglu, D., C. García-Jimeno, and J. A. Robinson (2015): "State Capacity and Economic Development: A Network Approach," *American Economic Review*, 105, 2364–2409.

Adão, R., C. Arkolakis, and F. Esposito (2019a): "General Equilibrium Indirect Effects in Space: Theory and Measurement," *NBER Working Paper 25544*.

Adão, R., M. Kolesár, and E. Morales (2019b): "Shift-Share Designs: Theory and Inference," *Quarterly Journal of Economics*, 134, 1949–2010.

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

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

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

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

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

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

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

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

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

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

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

Berman, N., A. Berthou, and J. Héricourt (2015): "Export Dynamics and Sales at Home," *Journal of International Economics*, 96, 298–310.

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

Berthou, A., J. J.-h. Chung, K. Manova, and C. Sandoz Dit Bragard (2019): "Productivity, (Mis)allocation and Trade," *Working Paper*.

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

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

Borusyak, K., P. Hull, and X. Jaravel (2019): "Quasi-Experimental Shift-Share Research Designs," *NBER Working Paper 24997*.

Boustan, L., F. Ferreira, H. Winkler, and E. 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.

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

Carvalho, V. M., M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi (2016): "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," .

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

Chodorow-Reich, G., P. Nenov, and A. Simsek (2019): "Stock Market Wealth and the Real Economy: A Local Labor Market Approach," *SSRN Electronic Journal*.

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

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

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

Costa, R., S. Dhingra, and S. Machin (2019): "Trade and Worker Deskilling," *Working Paper*.

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

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

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

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

Derenoncourt, E. (2019): "Can you move to opportunity ? Evidence from the Great Migration," *Working Paper*.

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

Ding, P., X. Li, and L. W. Miratrix (2017): "Bridging Finite and Super Population Causal Inference," *Journal of Causal Inference*, 5.

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

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

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

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

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

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

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2019): "Bartik Instruments : What, When, Why, and How," *Working Paper*.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118, 536–598.

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

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

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

Hemerik, J. and J. Goeman (2018): "Exact testing with random permutations," *Test*, 27, 811–825.

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

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

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

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

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

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

Jaravel, X., N. Petkova, and A. Bell (2018): "Team-Specific Capital and Innovation," *American Economic Review*, 108, 1034–1073.

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

Lee, Y. and E. L. Ogburn (2019): "Network Dependence and Confounding by Network Structure Lead to Invalid Inference," 1–29.

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

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

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

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

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

Nyblom, J., S. Borgatti, J. Roslakka, and M. A. Salo (2003): "Statistical analysis of network data - An application to diffusion of innovation," *Social Networks*, 25, 175–195.

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

Rosenbaum, P. 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.

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

Shaikh, A. and P. Toulis (2019): "Randomization Tests in Observational Studies with Staggered Adoption of Treatment," *Working Paper*.

Southworth, L. K., S. K. Kim, and A. B. Owen (2009): "Properties of balanced permutations," *Journal of Computational Biology*, 16, 625–638.

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

A Appendix Results

A.1 Efficiency Controls

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

The following result extends Propositions 1, 2, 3, and 4(i) to the case of efficiency controls:

Proposition 6. *Suppose $g \perp\!\!\!\perp (a, \varepsilon) \mid w$ where a collects the $a_\ell = (a_{\ell 1}, \ldots, a_{\ell r})$. Let $v_\ell^\perp$ denote the sample projection of a variable v_ℓ on a_ℓ: i.e., $v_\ell^\perp = v_\ell - \hat{\alpha}_v' a_\ell$ for $\hat{\alpha}_v = \left(\frac{1}{L}\sum_\ell a_\ell a_\ell'\right)^- \frac{1}{L}\sum_\ell a_\ell v_\ell$ and $(\cdot)^-$ denoting a generalized inverse of a matrix. Then:*

(i) *β is identified by $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell^\perp\right] / \mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp\right]$, assuming $\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp\right] \neq 0$;*

(ii) *The randomization test based on the statistic $T = \frac{1}{L}\sum_\ell z_\ell \left(y_\ell^\perp - b x_\ell^\perp\right)$ is valid;*

(iii) *The Hodges-Lehmann estimator induced by this RI statistic is the recentered SEIV estimator of y_ℓ on x_ℓ instrumented by $\tilde{z}_\ell$ and with the a_ℓ controls, $\tilde{\beta}_\perp = \frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell^\perp / \frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell^\perp$;*

(iv) *Recentering the instrument does not affect the estimator when μ_ℓ is included in a_ℓ.*

(v) *$\hat{\beta}_\perp \xrightarrow{p} \beta$ if Assumptions 3–6 hold, $\mathbb{E}\left[a_{\ell r}^2 \mid w\right] \leq B_a$ almost surely for all ℓ and $r = 1, \ldots, R$, $\frac{1}{L}\sum_\ell a_\ell a_\ell'$ is almost surely invertible (such that $\hat{\alpha}_v$ is unique),$\hat{\alpha}_x = O_p(1)$, and $\hat{\alpha}_\varepsilon = O_p(1)$.*

Proof: See Appendix C.6.

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

A.2 Multiple Instruments or Treatments

In progress.

A.3 Heterogeneous Treatment Effects

This appendix extends the classic result of local average treatment effect identification (Imbens and Angrist, 1994) to the SEIV setting, deriving an appropriate first-stage monotonicity condition under

which recentered SEIV regressions estimate a weighted average of heterogeneous effects. In place of the linear structural equation (2) we suppose outcomes are given by an unrestricted model $y_\ell = y_\ell(x_\ell, \varepsilon)$ of treatment x_ℓ and a set of unobservables ε. For notational simplicity we assume x_ℓ is continuous and of full support, with $y_\ell(x, \varepsilon)$ differentiable in x (these are each straightforward to relax). Treatment effects $\beta_\ell(x, \varepsilon) = \frac{\partial}{\partial x} y_\ell(x, \varepsilon)$ are then unrestricted across ℓ and ε, and we have the following result:

Proposition 7. *Suppose Assumption 1 holds and $Pr(x_\ell \geq x \mid z_\ell = z, \varepsilon, w)$ is almost surely weakly increasing in z for each x. Then the estimand of the recentered SEIV is*

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell) x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int \beta_\ell(x, \varepsilon)\omega_\ell(x, \varepsilon) dx\right], \tag{18}$$

where $\omega_\ell(x, \varepsilon)$ gives a convex weighting: $\omega_\ell(x, \varepsilon) \geq 0$ almost-surely and $\mathbb{E}\left[\frac{1}{L}\sum_\ell \int \omega_\ell(x, \varepsilon) dx\right] = 1$.

Proof. See Appendix C.7. □

In the conventional IV setting, where each observation is assigned its own as-good-as-random shock z_ℓ, one typically defines a triangular system of $y_\ell = y(x_\ell, \varepsilon_\ell)$ and $x_\ell = x(z_\ell, \eta_\ell)$ where $z_\ell \perp\!\!\!\perp (\varepsilon_\ell, \eta_\ell)$ by quasi-random assignment. A linear IV regression then identifies a proper weighted average of treatment effects $\beta_\ell(x, \varepsilon)$ when $x(z, \eta_\ell)$ is almost surely weakly increasing in z (Angrist et al., 2000). The monotonicity condition in Proposition 8 generalizes this to settings where a set of shocks g affect many observations of z_ℓ and x_ℓ jointly, where a causal first stage of $x_\ell = x(z_\ell, \eta_\ell)$ need not exist. For example in the linear shift share case of $z_\ell = \sum_n w_{\ell n} g_n$ we may suppose $x_\ell = \sum_n \pi_{\ell n} g_n + \eta_\ell$ for unobserved $(\pi, \eta) \perp\!\!\!\perp g \mid w$. Proposition 8 shows that the recentered SEIV regression remains causal in this case provided x_ℓ is stochastically increasing in z_ℓ conditional on ε and w. The weights $\omega_\ell(x, \varepsilon)$ are shown in Appendix C.7 to be proportional to the conditional-on-(ε, w) covariance of $\tilde{z}_\ell$ and $\mathbf{1}[x_\ell > x]$, so that more weight is given to treatment effects $\beta_\ell(x, \varepsilon)$ at margins x with a larger first stage response given ε and w. In the shift-share case, such monotonicity holds when the $w_{\ell n}$ and $\pi_{\ell n}$ are almost-surely non-negative and the g_n are mutually independent (Borusyak et al., 2019).

We note that SEIV inference may be challenging when treatment effects vary. For the so-called "strong null" of $\beta_\ell(x, \varepsilon) = 0$, almost surely, the randomization-based tests in Section 3.3 still apply, but may reject under the "weak null" of no average effect (i.e. that the estimand in Proposition is zero). Inverting RI tests to form confidence intervals for β is also no longer sensible with heterogeneous effects. This issue is not specific to RI, as asymptotic inference may also be challenging in this case. For example in the linear shift-share setting, Adão et al. (2019b) derive a necessary central limit theorem only for a reduced-form estimator $\tilde{z}'y/\tilde{z}'x$, under strong conditions. We view this challenge as a potentially fruitful area for future research.

A.4 Consistency with Permutations

This appendix considers a setting where the distribution of the shocks is known conditionally on $w_c = (w, \Pi(g))$ where w is some set of predetermined variables and $\Pi(g)$ is some function $\Pi(g)$ of shocks, such as the permutation class (with all or within-cluster permutations). One may therefore consider two expected instruments: $\mu_\ell^u = \mathbb{E}\left[f_\ell\left(g, w\right) \mid w\right]$ and $\mu_\ell^c = \mathbb{E}\left[f_\ell\left(g, w\right) \mid w_c\right]$, with corresponding recentered instruments $\tilde{z}_\ell^u$ and $\tilde{z}_\ell^c$. Here u and c stand for "unconditional" and "conditional" on $\Pi(g)$, respectively. Similarly, assumptions from the main text can be invoked in two different ways; we will adopt the convention that Assumption 3c is Assumption 3 with w replaced by w_c, and similarly for other assumptions.

We then establish consistency of the feasible conditional estimator $\hat{\beta}^c = \frac{\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c) y_\ell}{\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c) x_\ell}$ with Assumption 6 instead of 6c. The problem this proposition solves is that conditioning on $\Pi(g)$ creates dependencies across shock components, making $\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w_c\right]$ more difficult to bound; for example Corollaries 2 and 3 are not useful conditionally on w_c. We show that when Assumption 6 can be verified, consistency still follows under several regularity conditions. For simplicity we work with the stronger notion of shock exogeneity from Assumption 1.

Proposition 8. *Suppose Assumptions 1, 3c, 4c, 5, and 6 hold. Then the feasible conditional estimator is consistent:*

$$\hat{\beta}^c = \frac{\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right) y_\ell}{\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right) x_\ell} \xrightarrow{p} \beta.$$

Proof. See Appendix C.8. □

A.5 Assignment Processes with Unknown Parameters

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

We consider extensions of Assumption 3 where the distribution of $g \mid w$ is given by by a known function $G\left(g; w, \theta\right)$ of unknown θ. For example, one may assume conditionally independent binary shocks g_n with $Pr\left(g_n = 1 \mid w, \theta\right) = \Lambda\left(r_n' \theta\right)$ for a $K \times 1$ vector of shock-level observables r_n (including a constant) included in w, where $\Lambda\left(\cdot\right) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$ is the logistic function. In this class of models,

θ can be estimated from (g, w) by maximum likelihood (MLE), which is consistent under standard conditions, although other estimators may also be available. Given an estimate $\hat{\theta}$ a recentered SEIV instrument $\hat{z}_\ell = z_\ell - \mu_\ell\left(\hat{\theta}, w\right)$ can be measured, for $\mu_\ell(\theta, w) = \mathbb{E}_\theta[z_\ell \mid w] \equiv \int f_\ell(g, w) dG(g; w, \theta)$. We establish the conditions for large-sample consistency for this plug-in estimator for β below.

Valid, but likely quite conservative confidence intervals for β in such cases can be obtained by a simple extension of the previous randomization inference procedure. Given a value of θ, the randomization test for $\beta = b$ of Proposition 2 applies. Thus using the maximum p-value of this test across all possible values of θ yields a conservative test for β (with a corresponding confidence interval).[40] However, these confidence intervals are likely to be quite wide: even if the observed g is very informative about the precise value of θ, this test still searches through values very far from $\hat{\theta}$.

We propose an alternative two-step approach following Berger and Boos (1994) that is likely to be much less conservative but still valid (see Ding et al. (2017) for another application of this idea to RI). In the first step, a confidence interval CI_θ for θ with coverage $1-\gamma$ is constructed for some $\gamma \in (0, \alpha)$; Berger and Boos (1994) recommend $\gamma = 0.001$. Such tests are easy to build since the distribution of g is fully specified given θ; thus an exact RI-based confidence interval for θ can be constructed from any statistic $S = \mathcal{S}(g; w, \theta)$ by rerandomizing g according to $G(\cdot; w, \theta)$. As usual, the choice of S determines the power of the test and the length of the confidence interval. We propose a statistic that corresponds to the score test, $S = \frac{\partial}{\partial\theta} \log G(g; w, \theta)$, since the Hodges-Lehmann estimator induced by it is the MLE.[41] For vector-valued θ, S can be converted to a scalar LM statistic $S'\mathbb{E}_{\theta_0}[SS' \mid w]^{-1} S$; a value θ_0 is rejected if the LM statistic is in the right tail of its distribution. In the second step, the maximum p-value of the Proposition 2 test is taken across $\theta_0 \in CI_\theta$ only—a much smaller set in large samples than the entire parameter set used in the more conservative procedure. The p-value of the Berger and Boos (1994) test is the obtained maximum plus γ. A value of β is therefore rejected at significance level α if it is rejected under all $\theta_0 \in CI_\theta$ with significance $\alpha - \gamma$.

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

Proposition 9.

(i) *Suppose Assumptions 1 and 3 hold, $\hat{\theta}$ is consistent for θ, $\mu_\ell(\theta_0, w)$ is differentiable with respect to θ_0, and $\frac{\partial \mu_\ell}{\partial \theta}$ is uniformly bounded. Then the plug-in recentered SEIV estimator with instrument $\hat{z}_\ell$ is consistent.*

(ii) *Suppose Assumption 1 holds. Let $p_\beta(\beta; \theta_0)$ be the p-value of the randomization test of Proposition 2 for a given value of θ_0 and let CI_θ denote a confidence interval for θ such that $Pr(\theta \in CI_\theta) \geq 1-\gamma$ for $\gamma < \alpha$. Construct $CI_\beta = \{b \in \mathbb{R}\colon \ \max_{\theta_0 \in CI_\theta} p_\beta(\beta, \theta_0) + \gamma \leq \alpha\}$. Then CI_β is conservative for*

[40] An equivalent view on this procedure is to test joint hypotheses $\beta = b$ and $\theta = \theta_0$ using the test of Proposition 2 and then project the resulting confidence interval on the space of β.

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

β, *i.e.* $Pr\left(\beta \in CI_\beta\right) \leq \alpha$.

Proof. See Appendix C.9 for part (i). Part (ii) follows directly from Berger and Boos (1994). □

Five remarks are due. First, while the Berger and Boos (1994) test is conservative in finite samples only when CI_θ is, using an asymptotic confidence interval for θ will generally yield an asymptotically conservative interval for β. This simplifies computation: constructing the conventional Wald confidence interval for the MLE estimator of θ is much easier than inverting the score-based randomization test. Second, in some cases even simpler RI confidence intervals for β which plug in the estimate of $\hat{\theta}$ as if it was known are asymptotically correct (Shaikh and Toulis, 2019), although general conditions for this are unknown. Third, as discussed in Berger and Boos (1994), in some cases the nuisance parameter θ can be eliminated by using sufficient statistics which also yields a simpler exact confidence interval. In the above binary shocks example, if r_n captures a saturated set of dummy variables then elements of g are exchangeable within the clusters corresponding to them and it is not necessary to know or estimate θ. [42] Fourth, for a consistent $\hat{\theta}$, including $\mu_\ell\left(\hat{\theta}, w\right)$ as a linear control (with an additional coefficient in front of it) may produce a consistent estimator of β, as long as the slope of the auxiliary regression of z_ℓ on $\mu_\ell\left(\hat{\theta}, w\right)$ converges. This is because $\text{Cov}\left[z_\ell, \mu_\ell\left(\theta, w\right)\right] = \text{Var}\left[\mu_\ell\left(\theta, w\right)\right]$ by definition of $\mu_\ell\left(\theta, w\right)$, such that the slope coefficient will converge to one and the regression will asymptotically use the recentered $\tilde{z}_\ell$ as an instrument (by the Frisch-Waugh-Lovell theorem).[43]

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

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

[42] Rosenbaum (1984) shows how this idea can be extended in the logit model with arbitrary discrete observables r_n. He exploits the property of logit that, regardless of θ, $G\left(g \mid w\right)$ is the same for any binary vector g that yields the same vector $\sum_n g_n r_n$.

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

step, the expected instrument given θ is constructed by the following simulation: $\tilde{g}_n$ are randomly permuted to get $\tilde{g}_n^*$ and $g_n^* = \rho_n(\theta, w) + \tilde{g}_n^*$ is then used in constructing $z_\ell^* = f_\ell\left((g_n^*)_{n=1}^N, w\right)$.

The second case is heteroskedasticity, and for simplicity we assume that shocks are known to have a constant mean. One may therefore be willing to assume that $\tilde{g}_n = g_n/\sigma_n(\theta, w)$ is exchangeable; in this case a multiplicative constant is redundant in the formulation of the shock conditional variance, $\zeta\sigma_n^2(\theta, w)$. As usual, a variety of RI statistics can be used, and one reasonable choice is $T_\theta = \frac{1}{N}\sum_n \tilde{g}_n^2 \sigma_n^2 \frac{\partial \sigma_n^2}{\partial \theta}$ as it induces the Hodges-Lehmann estimator that corresponds to the moment of nonlinear least squares estimation for the model $g_n^2 = \zeta^2\sigma_n^2(\theta, w) + u_n$.[44] With an estimate of θ, recentering is performed by permuting $\tilde{g}_n^*$ and simulating $g_n = \tilde{g}_n^*\sigma_n\left(\hat{\theta}, w\right)$, and the Berger and Boos (1994) confidence interval for β is obtained similarly.

A.6 Theory-Consistent Nonlinear Shift-Share Instruments

In progress.

B Data Appendix

B.1 Section 5.1

In progress.

B.2 Section 5.2

In progress.

B.3 Section 5.3

In progress.

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

C Proofs of Propositions

C.1 Proof of Proposition 1

For the recentered SEIV regression,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_\ell\varepsilon_\ell\right] &= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_\ell\mathbb{E}\left[\varepsilon_\ell \mid g, w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_\ell\mathbb{E}\left[\varepsilon_\ell \mid w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}\mathbb{E}\left[\tilde{z}_\ell \mid w\right]\mathbb{E}\left[\varepsilon_\ell \mid w\right]\right] \\
&= 0. \qquad (19)
\end{aligned}$$

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

The alternative approach that regression-adjusts by μ_ℓ while using the uncentered z_ℓ as an instrument identifies β when

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_\ell\varepsilon_\ell^\perp\right] = \mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_\ell y_\ell^\perp\right] - \beta\cdot\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_\ell x_\ell^\perp\right] = 0, \qquad (20)$$

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

$$\begin{aligned}
\mathbb{E}\left[\varepsilon^\perp \mid g, w\right] &= (I - P_\mu)\,\mathbb{E}\left[\varepsilon \mid g, w\right] \\
&= (I - P_\mu)\,\mathbb{E}\left[\varepsilon \mid w\right] \\
&= \mathbb{E}\left[\varepsilon^\perp \mid w\right], \qquad (21)
\end{aligned}$$

where P_μ denotes the sample projection matrix for μ_ℓ and a constant (which is fixed conditional on w). Following the same steps as before, we thus have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{L}\sum_{\ell}z_\ell\varepsilon_\ell^\perp\right] &= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell)\mathbb{E}\left[\varepsilon_\ell^\perp \mid g, w\right]\right] \\
&= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}\mathbb{E}\left[\tilde{z}_\ell \mid w\right]\mathbb{E}\left[\varepsilon_\ell^\perp \mid w\right]\right] \\
&= 0, \qquad (22)
\end{aligned}$$

showing that the alternative μ_ℓ-controlled regression also identifies β.

C.2 Proof of Proposition 2

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

$$Pr\left(T^* \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid y, x, w\right) \geq 1 - \alpha \tag{23}$$

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

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

$$Pr\left(T \in R \mid \varepsilon, w\right) = Pr\left(T^* \in R \mid \varepsilon, w\right) \geq 1 - \alpha. \tag{24}$$

C.3 Proof of Proposition 3

The Hodges-Lehmann estimator of interest solves:

$$\begin{aligned}\frac{1}{L}\sum_\ell f_\ell\left(g, w\right)\left(y_\ell - bx_\ell\right) &= \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell\left(g^*, w\right)\left(y_\ell - bx_\ell\right) \mid w, y, x\right] \\ &= \frac{1}{L}\sum_\ell \mu_\ell\left(y_\ell - bx_\ell\right),\end{aligned} \tag{25}$$

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

$$\hat{\beta} = \frac{\frac{1}{L}\sum_\ell \left(f_\ell(g, w) - \mu_\ell\right) y_\ell}{\frac{1}{L}\sum_\ell \left(f_\ell(g, w) - \mu_\ell\right) x_\ell}, \tag{26}$$

which coincides with the recentered SEIV estimator.

For the statistic that uses the μ_ℓ-residualized outcome and treatment the result follows similarly:

$$\begin{aligned}\frac{1}{L}\sum_\ell f_\ell\left(g, w\right)\left(y_\ell^\perp - bx_\ell^\perp\right) &= \mathbb{E}\left[\frac{1}{L}\sum_\ell f_\ell\left(g^*, w\right)\left(y_\ell^\perp - bx_\ell^\perp\right) \mid w, y, x\right] \\ &= \frac{1}{L}\sum_\ell \mu_\ell\left(y_\ell^\perp - bx_\ell^\perp\right).\end{aligned} \tag{27}$$

The resulting estimator $\frac{\frac{1}{L}\sum_\ell (f_\ell(g,w)-\mu_\ell)y_\ell^\perp}{\frac{1}{L}\sum_\ell (f_\ell(g,w)-\mu_\ell)x_\ell^\perp} = \frac{\frac{1}{L}\sum_\ell f_\ell(g,w)y_\ell^\perp}{\frac{1}{L}\sum_\ell f_\ell(g,w)x_\ell^\perp}$ equals the SEIV estimator with the instrument z_ℓ and controlling for μ_ℓ, as in the Appendix C.1 proof.

C.4 Proof of Proposition 4

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

$$\begin{aligned}\tilde{\beta} - \beta &= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell} \\ &= \frac{\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell}{M}\left(1 + o_p(1)\right) \end{aligned} \tag{28}$$

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

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right] &= \mathbb{E}\left[\left(\frac{1}{L}\sum_\ell \tilde{z}_\ell \varepsilon_\ell\right)^2\right] \\ &= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \varepsilon_\ell \varepsilon_m\right] \\ &= \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\mathbb{E}\left[\varepsilon_\ell \varepsilon_m \mid w\right]\right] \\ &\leq \frac{1}{L^2}\sum_{\ell,m} \mathbb{E}\left[\left|\mathbb{E}\left[\tilde{z}_\ell \tilde{z}_m \mid w\right]\right| \sqrt{\mathbb{E}\left[\varepsilon_\ell^2 \mid w\right]\mathbb{E}\left[\varepsilon_m^2 \mid w\right]}\right] \\ &\leq B\mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m}\left|\mathrm{Cov}\left[\tilde{z}_\ell, \tilde{z}_m \mid w\right]\right|\right] \to 0 \end{aligned} \tag{29}$$

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

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

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

A sufficient condition for test consistency is then that $\frac{1}{L}\sum_\ell \tilde{z}_\ell^*(y_\ell - b x_\ell) \xrightarrow{p} 0$ for $\tilde{z}_\ell^* = f_\ell\left(g^*, w\right) - \mu_\ell$. For any b,

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

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

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

where the last line because the distributions of z^* and z are the same conditionally on w and for $C(b)$ such that $\mathbb{E}\left[(y_\ell - bx_\ell)^2 \mid w\right] \leq C(b)$ uniformly across w and ℓ. Such a bound is constructed from

$$\begin{aligned}
\mathbb{E}\left[\left(y_\ell - bx_\ell\right)^2 \mid w\right] &= \mathbb{E}\left[\varepsilon_\ell^2 + 2\left(\beta - b\right)x_\ell\varepsilon_\ell + \left(\beta - b\right)^2 x_\ell^2 \mid w\right] \\
&\leq B + 2\left|\beta - b\right| \cdot \left|\mathbb{E}\left[x_\ell\varepsilon_\ell \mid w\right]\right| + \left(\beta - b\right)^2\mathbb{E}\left[x_\ell^2 \mid w\right]
\end{aligned}$$

using the bounds for $\mathbb{E}\left[x_\ell\varepsilon_\ell \mid w\right]$ and $\mathbb{E}\left[x_\ell^2 \mid w\right]$.

Proof of Lemma 2(i). For the first statement of the lemma, we have

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

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

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

Lemma 3. If $h\colon \mathbb{R}^N \to \mathbb{R}$ is monotone and random variables $g_1, \ldots, g_N$ are independent, then for any $k \in \{1, \ldots, N-1\}$ the conditional expectation $\mathbb{E}\left[h\left(g_1, \ldots, g_N \mid g_1, \ldots g_k\right)\right]$ is monotone as well.

Proof: Denote the cumulative distribution function of g_n by $G_n(\cdot)$ and consider $g' = (g_1', \ldots, g_k', g_{k+1}, \ldots, g_N)$,

with $g'_n \geq g_n$ for $n \leq k$. Then $h(g') \geq h(g)$ by monotonicity. Therefore,

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

as required.

Lemma 4. For any monotone $h_1, h_2 \colon \mathbb{R}^N \to \mathbb{R}$, $\mathrm{Cov}\left[h_1\left(g\right), h_2\left(g\right)\right] \geq 0$ for $g = (g_1, \ldots, g_n)$ with independent components.

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

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

The first term is the expectation of a covariance between two monotone functions of $N-1$ variables, where monotonicity follows by Lemma 3. The second term, again by Lemma 3, is a covariance of two monotone functions of random scalars. Thus both of the terms are non-negative.

Applying Lemma 4 to $\tilde{z}_\ell = f_\ell\left(g, w\right) - \mu_\ell\left(w\right)$ and $\tilde{z}_m = f_m\left(g, w\right) - \mu_m\left(w\right)$ and conditioning on w everywhere, we obtain the second result of Lemma 2(i).

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

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

C.5 Proof of Proposition 5

Since, under the assumptions,

$$\mathrm{Var}\left[k_L\check{\beta}\right] = \frac{k_L^2 \mathrm{Var}\left[\check{z}'\varepsilon\right]}{\mathbb{E}\left[\check{z}'x\right]^2}(1+o(1)) \tag{30}$$

for any $\check{\beta}$ induced by a $\check{z}$ in this class, the proof follows from showing that $\mathrm{Var}\left[\tilde{z}'\varepsilon\right]/\mathbb{E}\left[\tilde{z}'x\right]^2 \leq \mathrm{Var}\left[\check{z}'\varepsilon\right]/\mathbb{E}\left[\check{z}'x\right]^2$. First note that by the law of iterated expectations and Assumption 2,

$$\begin{aligned}
\mathbb{E}\left[\check{z}'\varepsilon\varepsilon'\tilde{z}\right] &= \mathbb{E}\left[\mathbb{E}\left[\check{z}'\varepsilon\varepsilon'\tilde{z} \mid g, w\right]\right] \\
&= \mathbb{E}\left[\check{z}'\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right] \\
&= \mathbb{E}\left[\check{z}'\mathbb{E}\left[x \mid g, w\right]\right] \\
&= \mathbb{E}\left[\check{z}'x\right], \qquad (31)
\end{aligned}$$

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

$$\frac{k_L^2 \mathrm{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} = k_L^2 \mathrm{Var}\left[\tilde{z}'\varepsilon\right]^{-1} = k_L^2 \mathbb{E}\left[\tilde{z}'x\right]^{-1} \tag{32}$$

$$= k_L^2 \mathbb{E}\left[\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)' \mathbb{E}\left[\varepsilon\varepsilon' \mid g, w\right]^{-1}\left(\mathbb{E}\left[x \mid g, w\right] - \mathbb{E}\left[x \mid w\right]\right)\right]^{-1}. \tag{33}$$

It also shows that with

$$U = \frac{\check{z}'\varepsilon}{\mathbb{E}\left[\check{z}'x\right]} - \frac{\tilde{z}'\varepsilon.}{\mathbb{E}\left[\tilde{z}'x\right]} \tag{34}$$

we have

$$\begin{aligned}
\frac{\mathrm{Var}\left[\check{z}'\varepsilon\right]}{\mathbb{E}\left[\check{z}'x\right]^2} - \frac{\mathrm{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} &= \frac{\mathrm{Var}\left[\check{z}'\varepsilon\right]}{\mathbb{E}\left[\check{z}'x\right]^2} - 2\frac{\mathbb{E}\left[\check{z}'\varepsilon\varepsilon'\tilde{z}\right]}{\mathbb{E}\left[\check{z}'x\right]\mathbb{E}\left[\tilde{z}'x\right]} + \frac{\mathrm{Var}\left[\tilde{z}'\varepsilon\right]}{\mathbb{E}\left[\tilde{z}'x\right]^2} \\
&= \mathbb{E}\left[U^2\right] \\
&\geq 0.
\end{aligned}$$

C.6 Proof of Proposition 6

For part (i) observe that $g \perp\!\!\!\perp \varepsilon^{\perp} \mid w$ because $g \perp\!\!\!\perp (a, \varepsilon) \mid w$. Therefore, $\mathbb{E}\left[\frac{1}{L}\sum_{\ell}\tilde{z}_{\ell}\varepsilon_{\ell}^{\perp}\right] = 0$ by the law of iterated expectations, yielding identification. (A proof under a weaker exogeneity assumption $\mathbb{E}\left[\varepsilon_{\ell} \mid g, a, w\right] = \mathbb{E}\left[\varepsilon_{\ell} \mid a, w\right]$ can be constructed along the lines of Proposition 1, see equation (21)).

Part (ii) follows because under the null the distribution of $g \mid \varepsilon^{\perp}, w$ is the same as $g \mid w$, by

independence established in part (i).

Part (iii) is analogous to the proof of Proposition 3 for the μ_ℓ-controlled regression, see Appendix C.3.

Part (iv) follows from the fact that for any variable v_ℓ, $\frac{1}{L}\sum_\ell z_\ell v_\ell^\perp = \frac{1}{L}\sum_\ell \tilde{z}_\ell v_\ell^\perp$ because $\frac{1}{L}\sum_\ell \mu_\ell v_\ell^\perp = 0$ by the properties of projection.

Finally, for part (v) we write $\tilde{\beta}_\perp - \beta = \frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell / \frac{1}{L}\sum_\ell x_\ell^\perp \tilde{z}_\ell$. We first show that the numerator converges to zero in probability. We have:

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

By Proposition 4(i), $\frac{1}{L}\sum_\ell \varepsilon_\ell \tilde{z}_\ell = o_p(1)$. Moreover, using $\mathbb{E}\left[a_{\ell r}^2 \mid w\right] \leq B_a$, $g \perp\!\!\!\perp a \mid w$, and Assumption 6 and applying the proof of Proposition 4(i) with $a_{\ell r}$ in place of ε_ℓ yields $\frac{1}{L}\sum_\ell a_{\ell r}\tilde{z}_\ell = o_p(1)$ for each $r = 1, \ldots, R$. Since $\hat{\alpha}_\varepsilon = O_p(1)$, we have $\frac{1}{L}\sum_\ell \varepsilon_\ell^\perp \tilde{z}_\ell = o_p(1)$.

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

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

where $\frac{1}{L}\sum_\ell x_\ell \tilde{z}_\ell = M + o_p(1)$ by Assumption 4 and $\hat{\alpha}_x = O_p(1)$. Therefore, $\tilde{\beta}_\perp \xrightarrow{p} \beta$.

C.7 Proof of Proposition 7

Letting $\kappa_\ell(\varepsilon) = \lim_{x\to-\infty} y_\ell(x,\varepsilon)$, we have $y_\ell = \kappa_\ell(\varepsilon) + \int_{-\infty}^{x_\ell} \beta_\ell(x,\varepsilon)dx$. Note that $\mathbb{E}\left[\tilde{z}_\ell \kappa_\ell(\varepsilon)\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{z}_\ell \kappa_\ell(\varepsilon) \mid w\right]\right] = 0$ by the law of iterated expectations and Assumption 1. Thus,

$$\begin{aligned}
\mathbb{E}\left[\tilde{z}_\ell y_\ell\right] &= \mathbb{E}\left[\tilde{z}_\ell \int_{-\infty}^{x_\ell} \beta_\ell(x,\varepsilon)dx\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{x_\ell} \beta_\ell(x,\varepsilon)\tilde{z}_\ell dx \mid \varepsilon, w\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{\infty} \beta_\ell(x,\varepsilon)\tilde{z}_\ell \mathbf{1}\left[x_\ell \geq x\right] dx \mid \varepsilon, w\right]\right] \\
&= \mathbb{E}\left[\int_{-\infty}^{\infty} \beta_\ell\left(x,\varepsilon\right)\phi_\ell(x,\varepsilon)dx\right] \qquad (35)
\end{aligned}$$

where $\phi_\ell(x,\varepsilon) = \mathbb{E}\left[\tilde{z}_\ell \mathbf{1}\left[x_\ell \geq x\right] \mid \varepsilon, w\right]$, again by the law of iterated expectations. By similar steps we can write $\mathbb{E}\left[\tilde{z}_\ell x_\ell\right] = \mathbb{E}\left[\int_{-\infty}^{\infty} \phi_\ell(x,\varepsilon)dx\right]$. Note that

$$\begin{aligned}
\phi_\ell(x,\varepsilon) &= \mathbb{E}\left[\tilde{z}_\ell \mathbb{E}\left[\mathbf{1}\left[x_\ell \geq x \mid z_\ell, \varepsilon, w\right]\right] \mid \varepsilon, w\right] \\
&= \mathrm{Cov}\left[\tilde{z}_\ell, Pr\left(x_\ell \geq x \mid z_\ell, \varepsilon, w\right) \mid \varepsilon, w\right], \qquad (36)
\end{aligned}$$

yet again by the law of iterated expectations and the fact that $\mathbb{E}\left[\tilde{z}_\ell \mid \varepsilon, w\right] = 0$ under Assumption 1. Thus when $Pr\left(x_\ell \geq x \mid z_\ell = z, \varepsilon, w\right)$ is weakly increasing in z almost-surely, $\phi_\ell(x,\varepsilon) \geq 0$ almost-surely and

$$\frac{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell y_\ell\right]}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \tilde{z}_\ell x_\ell\right]} = \mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \beta_\ell\left(x,\varepsilon\right)\omega_\ell(x,\varepsilon)dx\right], \tag{37}$$

where

$$\omega_\ell(x,\varepsilon) = \frac{\phi_\ell(x,\varepsilon)}{\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \phi_\ell(x,\varepsilon)dx\right]} \tag{38}$$

gives a convex weighting scheme satisfying $\omega_\ell(x,\varepsilon) \geq 0$ almost-surely and $\mathbb{E}\left[\frac{1}{L}\sum_\ell \int_{-\infty}^{\infty} \omega_\ell(x,\varepsilon)dx\right] = 1$.

C.8 Proof of Proposition 8

The denominator of $\hat{\beta}^c - \beta = \frac{\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)\varepsilon_\ell}{\frac{1}{L}\sum_\ell (z_\ell - \mu_\ell^c)x_\ell}$ converges to $M \neq 0$ by Assumption 4c, so we focus on the numerator. Because $\Pi(g)$ is a function of g, Assumption 1 implies Assumption 1c ($g \perp\!\!\!\perp \varepsilon \mid (w, \Pi(g))$), so $\mathbb{E}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right)\varepsilon_\ell \mid w_c\right] = 0$ by the law of iterated expectations. Consider the variance now, conditionally on w:

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right)\varepsilon_\ell \mid w\right] &= \mathbb{E}\left[\mathrm{Var}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right)\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\overset{a.s.}{\leq} \mathbb{E}\left[\mathrm{Var}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^c\right)\varepsilon_\ell + \frac{1}{L}\sum_\ell \left(\mu_\ell^c - \mu_\ell^u\right)\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&= \mathbb{E}\left[\mathrm{Var}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^u\right)\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\overset{a.s.}{\leq} \mathrm{Var}\left[\frac{1}{L}\sum_\ell \left(z_\ell - \mu_\ell^u\right)\varepsilon_\ell \mid w\right] \\
&\overset{a.s.}{\leq} B \cdot \mathbb{E}\left[\frac{1}{L^2}\sum_{\ell,m} \left|\mathrm{Cov}\left[z_\ell, z_m \mid w\right]\right| \mid w\right] \xrightarrow{p} 0 \text{ a.s.}
\end{aligned} \tag{39}$$

Here the first line follows by the law of total variance since the conditional expectation is zero. The second line follows because

$$\begin{aligned}
\operatorname{Cov}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^c)\,\varepsilon_\ell,\frac{1}{L}\sum_{m}(\mu_m^c-\mu_m^u)\,\varepsilon_\ell \mid w_c\right] &= \mathbb{E}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^c)\,\varepsilon_\ell\cdot\frac{1}{L}\sum_{m}(\mu_m^c-\mu_m^u)\,\varepsilon_\ell \mid w_c\right] \\
&= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[(z_\ell-\mu_\ell^c)\,\varepsilon_\ell\varepsilon_m \mid w_c\right]\cdot(\mu_m^u-\mu_m^c) \\
&= \frac{1}{L^2}\sum_{\ell,m}\mathbb{E}\left[z_\ell-\mu_\ell^c \mid w_c\right]\cdot\mathbb{E}\left[\varepsilon_\ell\varepsilon_m \mid w_c\right](\mu_m^c-\mu_m^u) \\
&= 0.
\end{aligned}$$

When two random variables are uncorrelated, the variance of the sum exceeds the variance of one. The fourth line of (39) again follows by the law of total variance, specifically that

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Var}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\qquad = \operatorname{Var}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^u)\,\varepsilon_\ell \mid w\right] - \operatorname{Var}\left[\mathbb{E}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^u)\,\varepsilon_\ell \mid w_c\right] \mid w\right] \\
&\qquad\qquad \leq \operatorname{Var}\left[\frac{1}{L}\sum_{\ell}(z_\ell-\mu_\ell^u)\,\varepsilon_\ell \mid w\right].
\end{aligned}$$

Finally, the last line of (39) directly follows from the proof of Proposition 4 (equation (29), conditionally on w) using Assumptions 1, 5, and 6.

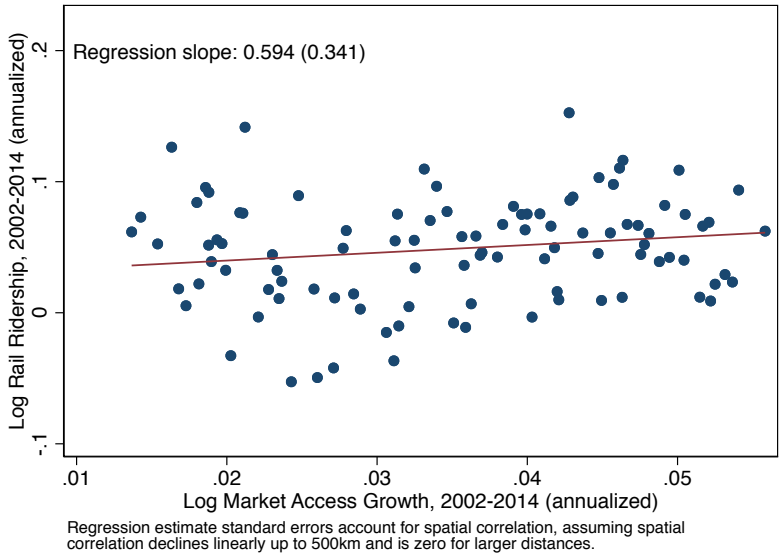
Since $\mathbb{E}\left[\frac{1}{L}\sum_\ell(z_\ell-\mu_\ell^c)\,\varepsilon_\ell \mid w\right]=0$, (39) implies the unconditional $\operatorname{Var}\left[\frac{1}{L}\sum_\ell(z_\ell-\mu_\ell^c)\,\varepsilon_\ell\right]$ converges to zero as well, yielding consistency of $\hat{\beta}^c$.

C.9 Proof of Proposition 9

In progress.

Appendix Figures and Tables

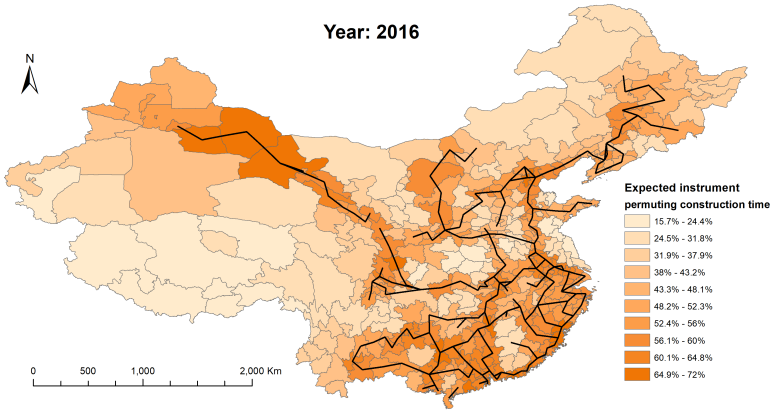
Figure A1: Chinese Log Market Access and Rail Ridership Growth



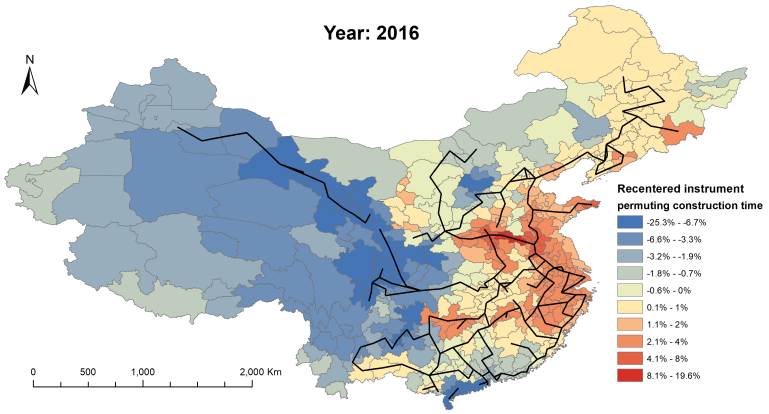
Notes: This figure shows a binned scatterplot of annualized log rail ridership across 306 cities in China, from 2002 to 2014, against annualized log market access growth in the same period. Bins are constructed as percentiles of market access growth. Lines of best fit are indicated in red, with regression coefficients and spatial-clustered standard errors indicated in each panel.

Figure A2: Expected and Recentered Chinese Log Market Access Growth, Permuted Construction Times

A. Expected Instrument



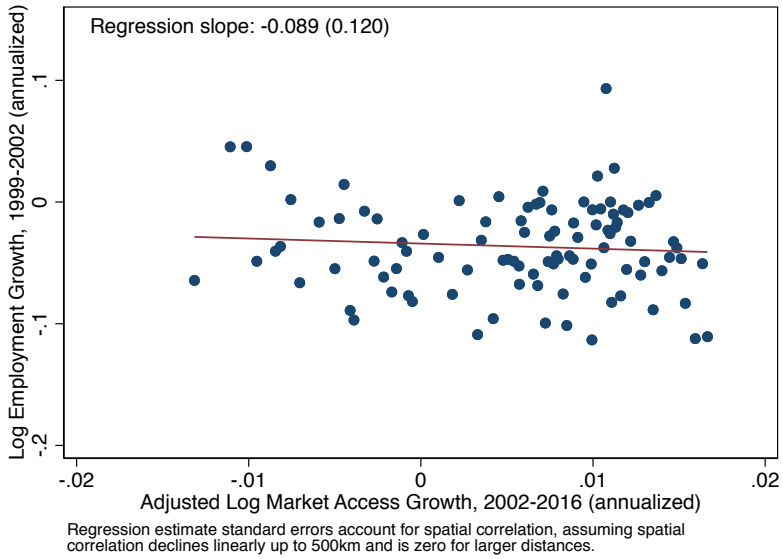
B. Recentered Instrument



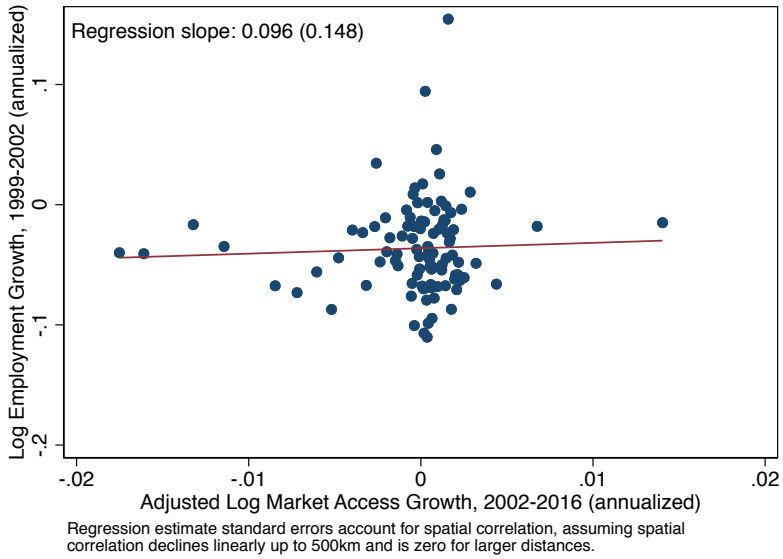
Notes: Panel A of this figure plots the variation in expected log market access growth from 2002 to 2016 in China, along with the high-speed rail lines constructed during this period. Panel B instead plots the variation in the recentered log market access instrument: the difference between the market access growth shown in panel E of Figure 2 and panel A of this figure. Expected market access is simulated by permuting line construction time, as described in Section 5.1.

Figure A3: Recentered Log Market Access Growth and Employment Pre-Trends

A. Permuted Opening Dates



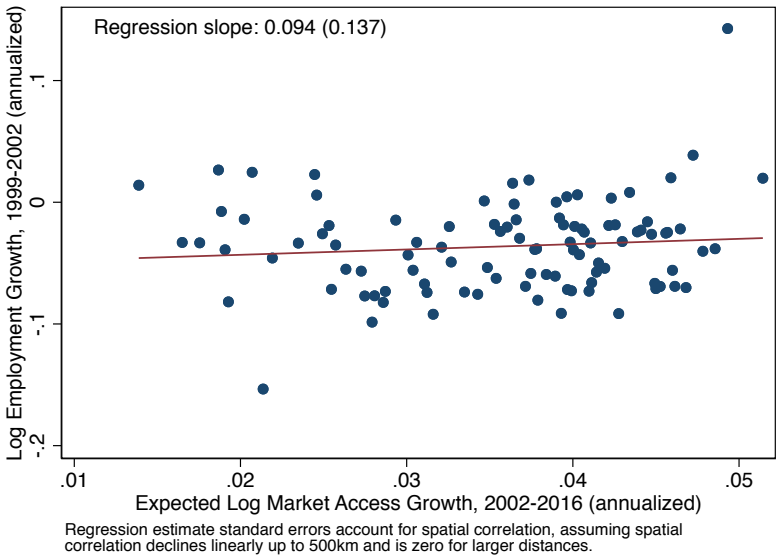
B. Permuted Construction Times



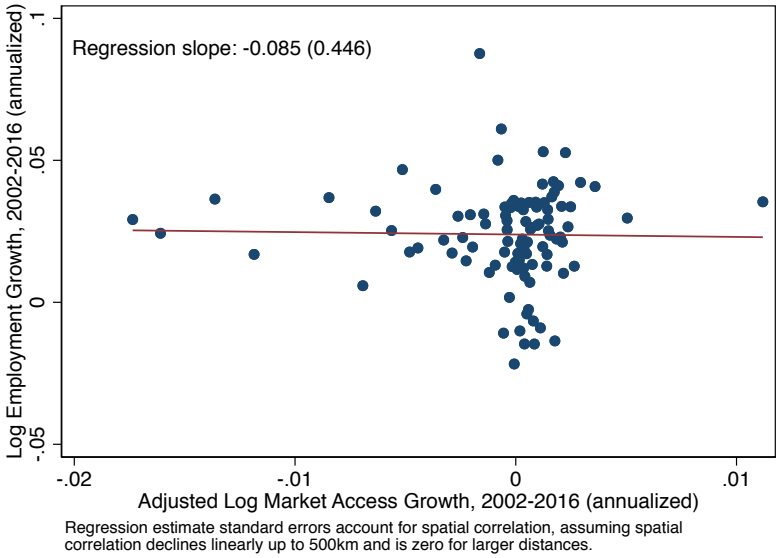
Notes: This figure shows binned scatterplots of annualized log employment growth across 306 cities in China, from 1999 to 2002, against annualized recentered log market access growth in 2002 to 2016. Panel A simulates the expected instrument correction by permuting line opening dates, while Panel B simulates the expected instrument correction by permuting line construction time. Bins are constructed as percentiles of recentered log market access growth. Lines of best fit are indicated in red, with regression coefficients and spatial-clustered standard errors indicated in each panel.

Figure A4: Expected and Recentered Market Access Growth and Employment, Permuted Construction Times

A. Expected Log Market Access Growth and Employment Pre-Trends



B. Recentered Log Market Access Growth and Employment Growth



Notes: Panel A of this figure shows a binned scatterplot of annualized log employment growth across 306 cities in China, from 1999 to 2002, against annualized log expected market access growth in 2002 to 2016. Panel B shows a binned scatterplot of annualized log employment growth in 2002 to 2016 against annualized recentered log market access growth in the same period. Bins are constructed as percentiles of log market access growth. Lines of best fit are indicated in red, with regression coefficients and spatial-clustered standard errors indicated in each panel. Expected and recentered market access are simulated by permuting line construction times, as described in Section 5.1.

Table A1: Log GDP Corrected Panel Estimates and Confidence Intervals - 50% Exclusion Threshold

	Unadjusted Instrument (1)	Permuted opening dates (2)	Permuted construction times (3)
	A. Subtracting off Expected Instrument		
Coefficient	−0.057	−0.129	−0.074
Conventional 95% CI	[-0.217, 0.104]	[-0.295, 0.038]	[-0.264, 0.116]
Permutation 95% CI		[-0.405, 0.577]	[-0.446, 0.364]
	B. Controlling for Expected Instrument		
Coefficient		−0.149	−0.069
Conventional 95% CI		[-0.318, 0.020]	[-0.259, 0.121]
Permutation 95% CI		[-0.460, 0.390]	[-0.487, 0.278]

N=5,282; includes city and year fixed effects; conventional CIs cluster by 373 cities.

Table A2: Conventional and Shock-Exposure IV Pre-Trend and Effect Estimates

	Pre-trends			Estimates
	Simulated Instrument IV (1)	SEIV (Actual Income) (2)	SEIV (Predicted Income) (3)	SEIV (Predicted Income) (4)
Is Insured	0.001	0.006	0.003	0.078***
	(0.009)	(0.005)	(0.006)	(0.014)
Has Medicaid	0.015***	0.015***	0.007	0.098***
	(0.005)	(0.004)	(0.005)	(0.016)
Has Private	−0.011	−0.007*	−0.001	−0.023**
	(0.011)	(0.004)	(0.004)	(0.010)
Has Other	−0.000	−0.000	0.002	0.002
	(0.004)	(0.002)	(0.002)	(0.003)
First Stage				0.538***
				(0.019)
Individuals	2,400,142	425,112	319,870	320,044
States	43	43	43	43

Notes: Column 1 of this table estimates a difference-in-differences regression of coverage in 2012 and 2013 on a simulated instrument, interacted with year, predicting 2014 eligibility from an individual's state with a nationally representative sample of individuals in 2014. This specification also controls for state and year fixed effects and predicted 2014 eligibility main effects. Columns 2-4 estimate similar pre-trend specifications with a shock-exposure instrument predicting 2014 eligibility from an individual's state's Medicaid expansion decision as described in the text. These specifications restrict to observations with a non-degenerate adjusted instrument. The instrument in column 2 is constructed from actual household income, as a percentage of the federal poverty level (FPL), while column 3 predicts income FPL percentages from a model, estimated in 2013, that regresses the log of income FPL percentages on individuals' family type, age, educational attainment, quadratic terms for age and education, and indicator variables for sex, white, married, and labor force participation. Column 4 estimates a difference-in-differences IV regression with this predicted income shock-exposure instrument, as in column 6 of Table 1. All estimates come from a nationally representative sample of individuals in 43 states that did not previously expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.

Table A3: Shock-Exposure IV Estimates of Medicaid Eligibility Effects, Alternative Expansion Designs

		State Covariates			
	No Covariates (1)	Republican Governor (2)	Median Income (3)	Medicaid Rate (4)	All Covariates (5)
Panel A. IV Estimates					
Is Insured	0.059***	0.051***	0.066***	0.066***	0.060***
	(0.011)	(0.009)	(0.014)	(0.014)	(0.013)
Has Medicaid	0.085***	0.074***	0.092***	0.087***	0.078***
	(0.011)	(0.010)	(0.014)	(0.013)	(0.011)
Has Private	−0.025***	−0.023***	−0.024***	−0.022***	−0.020**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)
Has Other	0.001	0.003	0.002	0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
First Stage	0.969***	1.007***	0.917***	0.971***	0.969***
	(0.016)	(0.099)	(0.064)	(0.094)	(0.134)
Panel B. Specification Tests					
Expected Instrument Slope Coefficient		1.110	0.784	1.104	1.056
		(0.284)	(0.471)	(0.261)	(0.124)
Expected Instrument Mean Difference	−0.151*	−0.078	−0.070	−0.035	−0.008
	(0.090)	(0.072)	(0.088)	(0.069)	(0.055)
χ^2 Test: Slope=1, Mean Diff.=0	2.794	1.223	0.784	0.797	0.294
	[0.102]	[0.547]	[0.678]	[0.674]	[0.864]
Individuals	421,042	421,042	320,044	421,042	421,042
States	43	43	43	43	43

Notes: Panel A of this table estimates difference-in-difference shock-exposure IV regressions, as in column 6 of Table 1, with different predicted 2014 eligibility expected instrument controls based on different models for state expansion decision. Coverage effects are measured in 2014. Column 1 repeats the baseline estimates where expansion decisions are assumed random across states, while columns 2-4 model the expansion decision as a function of 2012 state characteristics with a probit regression (weighted by 2012 state population). Column 5 includes all of these state covariates in a single weighted probit model. Panel B reports slope coefficients from regressing predicted 2014 eligibility on the different expected instruments along with the difference in the average instrument and the average expected instrument. Test statistics for the null hypotheses of correct specification (that the slope coefficient is one and the difference in means is zero) are reported, with p-values in brackets. All estimates come from a nationally representative sample of individuals in 43 states that did not previously expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.

Table A4: Conventional and Shock-Exposure IV Estimates of Medicaid Eligibility Effects, 2015 Outcomes

	Conventional Estimates				SEIV Estimates		
	OLS						
	No Controls	Income Controls	Simulated Instrument IV	Expansion Instrument IV	Recentered	Controlled	Sharp Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Is Insured	0.090***	0.029**	0.036	0.075*	0.059	0.067*	0.109***
	(0.012)	(0.013)	(0.035)	(0.043)	(0.051)	(0.035)	(0.018)
Has Medicaid	0.096***	0.082***	0.175***	0.218***	0.170***	0.151***	0.173***
	(0.011)	(0.011)	(0.021)	(0.031)	(0.026)	(0.024)	(0.018)
Has Private	−0.002	−0.051***	−0.128***	−0.129***	−0.106**	−0.079***	−0.058***
	(0.006)	(0.007)	(0.026)	(0.026)	(0.047)	(0.020)	(0.009)
Has Other	0.005***	0.000	0.004	0.003	−0.006	−0.002	−0.000
	(0.001)	(0.001)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
First Stage			0.935***	0.184***	0.618***	0.810***	0.854***
			(0.031)	(0.020)	(0.179)	(0.103)	(0.084)
Individuals	2,400,730	2,400,730	2,400,730	2,400,730	2,400,730	2,400,730	418,852
States	43	43	43	43	43	43	43

Notes: This table reports OLS and IV estimates of the effect of Medicaid eligibility in 2015 on different forms of insurance coverage in 2015. Column 1 estimates a difference-in-differences regression of coverage in 2013 and 2015 on 2015 eligibility, interacted with year, controlling for state and year fixed effects. Column 2 further controls for a fourth-order polynomial in household income (as a percentage of the federal poverty level) interacted with year. Column 3 estimates difference-in-differences IV regressions with a simulated instrument, interacted with year, predicting 2014 eligibility from an individual's state with a nationally representative sample of individuals in 2014. This specification also controls for state and year fixed effects and predicted 2014 eligibility main effects. Columns 5-7 estimate difference-in-differences IV regressions with a shock-exposure instrument, interacted with year, predicting 2014 eligibility from an individual's state's Medicaid expansion decision as described in the text. These specifications also control for state and year fixed effects and predicted 2014 eligibility main effects. The instrument in Column 5 adjusts for an individual's expected eligibility over the state expansion shocks. Column 6 controls for the shock-exposure instrument. Column 7 restricts estimation to observations with a non-degenerate adjusted instrument. Estimates come from a nationally representative sample of individuals in 43 states that did not expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.

Table A5: Conventional and Shock-Exposure IV Pre-Trend and Effect Estimates, 2015 Outcomes

	Pre-trends			Estimates
	Simulated Instrument IV (1)	SEIV (Actual Income) (2)	SEIV (Predicted Income) (3)	SEIV (Predicted Income) (4)
Is Insured	0.001	0.006	0.003	0.137***
	(0.009)	(0.005)	(0.006)	(0.026)
Has Medicaid	0.015***	0.015***	0.007	0.203***
	(0.005)	(0.004)	(0.005)	(0.027)
Has Private	−0.011	−0.007*	−0.001	−0.057***
	(0.011)	(0.004)	(0.004)	(0.012)
Has Other	−0.000	−0.000	0.002	−0.003
	(0.004)	(0.002)	(0.002)	(0.005)
First Stage				0.473***
				(0.053)
Individuals	2,400,142	425,112	319,870	320,569
States	43	43	43	43

Notes: Column 1 of this table estimates a difference-in-differences regression of coverage in 2012 and 2013 on a simulated instrument, interacted with year, predicting 2014 eligibility from an individual's state with a nationally representative sample of individuals in 2015. This specification also controls for state and year fixed effects and predicted 2014 eligibility main effects. Columns 2-4 estimate similar pre-trend specifications with a shock-exposure instrument predicting 2014 eligibility from an individual's state's Medicaid expansion decision as described in the text. These specifications restrict to observations with a non-degenerate adjusted instrument. The instrument in column 2 is constructed from actual household income, as a percentage of the federal poverty level (FPL), while column 3 predicts income FPL percentages from a model, estimated in 2013, that regresses the log of income FPL percentages on individuals' family type, age, educational attainment, quadratic terms for age and education, and indicator variables for sex, white, married, and labor force participation. Column 4 estimates a difference-in-differences IV regression with this predicted income shock-exposure instrument, as in column 6 of Table 1. All estimates come from a nationally representative sample of individuals in 43 states that did not previously expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.

Table A6: Shock-Exposure IV Estimates of Medicaid Eligibility Effects, Alternative Expansion Designs, 2015 Outcomes

		State Covariates			
	No Covariates (1)	Republican Governor (2)	Median Income (3)	Medicaid Rate (4)	All Covariates (5)
Panel A. IV Estimates					
Is Insured	0.109***	0.097***	0.125***	0.112***	0.104***
	(0.018)	(0.012)	(0.021)	(0.026)	(0.021)
Has Medicaid	0.173***	0.159***	0.193***	0.177***	0.172***
	(0.018)	(0.014)	(0.024)	(0.026)	(0.025)
Has Private	−0.058***	−0.057***	−0.059***	−0.060***	−0.063***
	(0.009)	(0.008)	(0.011)	(0.012)	(0.012)
Has Other	−0.000	0.001	0.002	−0.000	0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
First Stage	0.854***	0.903***	0.780***	0.816***	0.798***
	(0.084)	(0.112)	(0.110)	(0.142)	(0.165)
Panel B. Specification Tests					
Expected Instrument Slope Coefficient		1.111	0.783	1.104	1.054
		(0.283)	(0.471)	(0.258)	(0.124)
Expected Instrument Mean Difference	0.072	−0.098	−0.030	−0.051	−0.015
	(0.090)	(0.096)	(0.118)	(0.062)	(0.053)
χ^2 Test: Slope=1, Mean Diff.=0	0.630	1.219	0.793	0.762	0.261
	[0.432]	[0.548]	[0.675]	[0.685]	[0.878]
Individuals	418,852	418,852	320,569	418,852	418,852
States	43	43	43	43	43

Notes: Panel A of this table estimates difference-in-difference shock-exposure IV regressions, as in column 6 of Table 1, with different predicted 2014 eligibility expected instrument controls based on different models for state expansion decision. Coverage effects are measured in 2015. Column 1 repeats the baseline estimates where expansion decisions are assumed random across states, while columns 2-4 model the expansion decision as a function of 2012 state characteristics with a probit regression (weighted by 2012 state population). Column 5 includes all of these state covariates in a single weighted probit model. Panel B reports slope coefficients from regressing predicted 2014 eligibility on the different expected instruments along with the difference in the average instrument and the average expected instrument. Test statistics for the null hypotheses of correct specification (that the slope coefficient is one and the difference in means is zero) are reported, with p-values in brackets. All estimates come from a nationally representative sample of individuals in 43 states that did not previously expand Medicaid eligibility as in the ACA prior to 2014. Robust standard errors, clustered by state, are reported in parentheses. Asterisks denote significance at the 10% (*), 5% (**), and 1% (***) levels.